

AMPLIACIÓN DE LA PRUEBA DE CONCEPTO DEL ATLAS ENERGÉTICO DE BIOMASA RESIDUAL AGRÍCOLA EN COLOMBIA

(Convenio de asociación N°: CO1.PCCNTR.5527461)

(Fecha de inicio: 23-11-23 y Fecha fin: 31-12-23)

Entregable 2: Ampliación de la prueba de concepto del atlas de biomasa residual para la identificación de palma de aceite a nivel nacional

Documento 3: Metodología del desarrollo de la aplicación y la valoración de la pertinencia del modelo de identificación de palma de aceite

Bogotá D.C. 20 de diciembre de 2023

Equipo de trabajo

Universidad de los Andes

Directora de proyecto:

PhD. Rocío Sierra

Investigadores:

PhD. Guillermo Jiménez

PhD. Fernando Jiménez

PhD. Haydemar Núñez

PhD. Luis Cruz

MSc. Andrés Calderón

MSc. Nicolás Díaz

MSc. Norymar Becerra

Ing. Yafar Solano

Daniel Beltrán

Kevin Gámez

Soporte técnico

David Jiménez

Unidad de Planeación Minero-Energética (UPME)

Supervisor de proyecto:

Ing. Esp. John Alejandro
Barrios

Equipo técnico:

Ing. Manuel Sierra

Esp. Mónica Castañeda

Esp. Catalina Londoño

Equipo administrativo:

Esp. Diego Vanegas

Esp. María Alejandra Acosta

Esp. Olga Lucía Carranza

Tabla de contenido

Resumen Ejecutivo	5
1. Antecedentes	6
1.1. Proyecto de Cooperación Triangular Alemania-Chile-Colombia	6
1.2. Estado del arte: Modelos de inteligencia artificial en la identificación de cultivos	8
2. Valoración de la pertinencia del modelo de identificación de palma de aceite a nivel nacional	11
2.1. Metodología aplicada para el desarrollo del modelo de inteligencia artificial que identifica palma de aceite	11
2.2. Resultados de la aplicación del modelo de inteligencia artificial para la identificación de palma de aceite a nivel nacional.	14
2.2.1 Validación del modelo y resultados obtenidos	15
3. Metodología del desarrollo de la aplicación de la herramienta del atlas de biomasa residual para la identificación de palma de aceite a nivel nacional	17
4. Conclusiones y recomendaciones	26
5. Trabajo futuro	27
6. Referencias	28

Tabla de figuras

Figura 1. Metodología aplicada a la prueba de concepto del Atlas de biomasa residual en departamentos de Bolívar y Cesar	7
Figura 2. Arquitectura encoder-decoder de DeepLabv3+	10
Figura 3. Metodología para la construcción del modelo de palma de aceite	11
Figura 4. Metodología para la construcción del modelo de palma de aceite	14
Figura 5. Metodología de procesamiento de imágenes	15
Figura 6. Resultados de la aplicación de modelos de identificación de palma de aceite	16
Figura 7. Arquitectura general de la aplicación.	17
Figura 8. Diagrama entidad-relación para la aplicación	19
Figura 9. Vista inicial de la aplicación	21
Figura 10. Vista general de la aplicación - histórico de producción.	22
Figura 11. Vista de generación de la aplicación - selección de residuos.	23
Figura 12. Vista de generación - composición másica en base seca.	24

Figura 13. Vista de generación - selección de tecnología de conversión.

25

Figura 14. Vista de generación - resultados de la estimación.

26

Tabla de Tablas

Tabla 1. Puntos finales disponibles en la capa de lógica.

20

Resumen Ejecutivo

La Universidad de los Andes en conjunto con la Unidad de Planeación Minero-Energética ejecutaron una primera fase de un proyecto orientado al desarrollo de un *Atlas energético de biomasa residual agrícola en Colombia*, con información actualizada de fuentes oficiales y mediante la aplicación de modelos de inteligencia artificial para la identificación de coberturas agrícolas. En esta primera fase se hizo la ampliación de la prueba de concepto del atlas energético de biomasa residual agrícola en Colombia, a partir de la experiencia adquirida por Uniandes en el desarrollo del trabajo previo en el Proyecto de Cooperación Triangular Alemania-Chile-Colombia. En este proyecto previo se desarrolló una prueba de concepto, para evaluar el potencial energético de la biomasa residual agrícola, correspondiente a cultivos de palma de aceite en municipios vulnerables de los departamentos del Bolívar y del Cesar. Para el desarrollo de la prueba de concepto se aplicó una metodología con un enfoque novedoso, basado en modelos de inteligencia artificial (IA) entrenados a partir de información georreferenciada del cultivo y de imágenes satelitales, para la identificación de palma de aceite en el área de estudio.

La primera fase del presente proyecto tiene dos productos asociados, uno enfocado a hacer una contextualización de la información georreferenciada disponible actualmente de los cultivos de interés de café, palma de aceite, arroz, maíz, plátano, caña y cacao. El segundo a la ampliación de la prueba de concepto del atlas de biomasa residual para la identificación de palma de aceite a nivel nacional.

El presente informe corresponde a la ampliación de la prueba de concepto del atlas de biomasa residual para la identificación de palma de aceite a nivel nacional (Documento 3), para el cual se hizo: i) la valoración de la pertinencia del modelo de identificación de palma de aceite a nivel nacional ; y ii) explicación de la metodología aplicada para el desarrollo del aplicativo del atlas de biomasa residual para la identificación de palma de aceite a nivel nacional y estimación del potencial energético de los residuos de palma.

Los resultados de la valoración de la pertinencia del modelo de identificación de palma de aceite a nivel nacional arrojaron que, se requiere un refinamiento adicional para solventar los cambios observados en la reflectancia en regiones alejadas al área de estudio, donde originalmente se entrenó el modelo. Se observa que es necesario incluir información adicional que dé más soporte a la construcción del modelo y el entrenamiento de adicionales modelos que busquen capturar las peculiaridades de la diversidad de regiones del territorio colombiano.

1. Antecedentes

1.1. Proyecto de Cooperación Triangular Alemania-Chile-Colombia

La Universidad de los Andes en colaboración con la Universidad de Chile y con el respaldo financiero de la Sociedad Alemana para la Cooperación Internacional (GIZ), desarrollaron un Proyecto de Cooperación Triangular basado en pruebas de concepto¹ (Sierra, y otros, 2023), de las cuales una fue para evaluar el potencial energético de la biomasa residual agrícola, correspondiente a cultivos de palma de aceite y café en municipios vulnerables de los departamentos del Bolívar y del Cesar en Colombia. Para el desarrollo de la prueba de concepto se aplicó una metodología con un enfoque novedoso, en el cual se emplearon tanto imágenes satelitales históricas como actuales, así como diversos conjuntos de datos. Esta valiosa información fue sometida a un análisis y procesamiento mediante el uso de un modelo de inteligencia artificial (IA) que permite identificar cultivos de palma de aceite en la zona de estudio.

Para la prueba de concepto se seleccionó el cultivo de palma de aceite de acuerdo a: i) la información disponible de área sembrada, productividad y rendimiento del cultivo, tomada de las Evaluaciones Agropecuarias Municipales (EVA)²; ii) la georreferenciación del cultivo usando como guía el Mapa de Coberturas de la Tierra, del IDEAM (2023)³ de escala de 1:100.000; iii) la disponibilidad de polígonos suministrados por el Instituto Geográfico Agustín Codazzi (IGAC)⁴ de la zona de estudio; iv) la preparación de los datos generados a partir de los polígonos con base en el análisis de la firma espectral del cultivo; e v) información de un modelo digital de elevación. La información mencionada anteriormente se trabajó con imágenes satelitales de la zona de estudio en los departamentos del Cesar y del Bolívar, y se construyó un conjunto de datos para el cultivo de palma de aceite que permitió desarrollar un modelo de inteligencia artificial en la identificación del cultivo y la estimación de su área, con un 95% de exactitud. Por medio de un modelo de regresión lineal ($R^2 = 0,87$) construido de bases de datos históricas se pudo estimar la productividad del cultivo, a partir del área sembrada determinada por el modelo de inteligencia artificial. Por medio de la productividad del cultivo de palma y conocido los porcentajes de residuos, se estimó la biomasa residual que puede ser aprovechada en la generación de energía, ya sea por procesos de incineración (tratamiento térmico) o generación de biogás (digestión anaerobia). Se desarrolló un modelo para la estimación de potencial energético de biomasa residual, con el cual se obtuvieron ecuaciones que permiten estimar el poder calorífico inferior (LHV) a partir del análisis último de la biomasa (contenido de C,H,O,N,S y cenizas). Para la potencia generada (W) se estimaron las ecuaciones por un ajuste de un modelo de regresión múltiple, que varió de acuerdo a la tecnología de aprovechamiento. En el caso de la incineración la ecuación obtenida tiene un $R^2=0,84$ y para biogás un $R^2=0,99$ (Sierra, y otros, 2023). En la Figura 1 se presenta un esquema de las secciones principales de la metodología descrita previamente

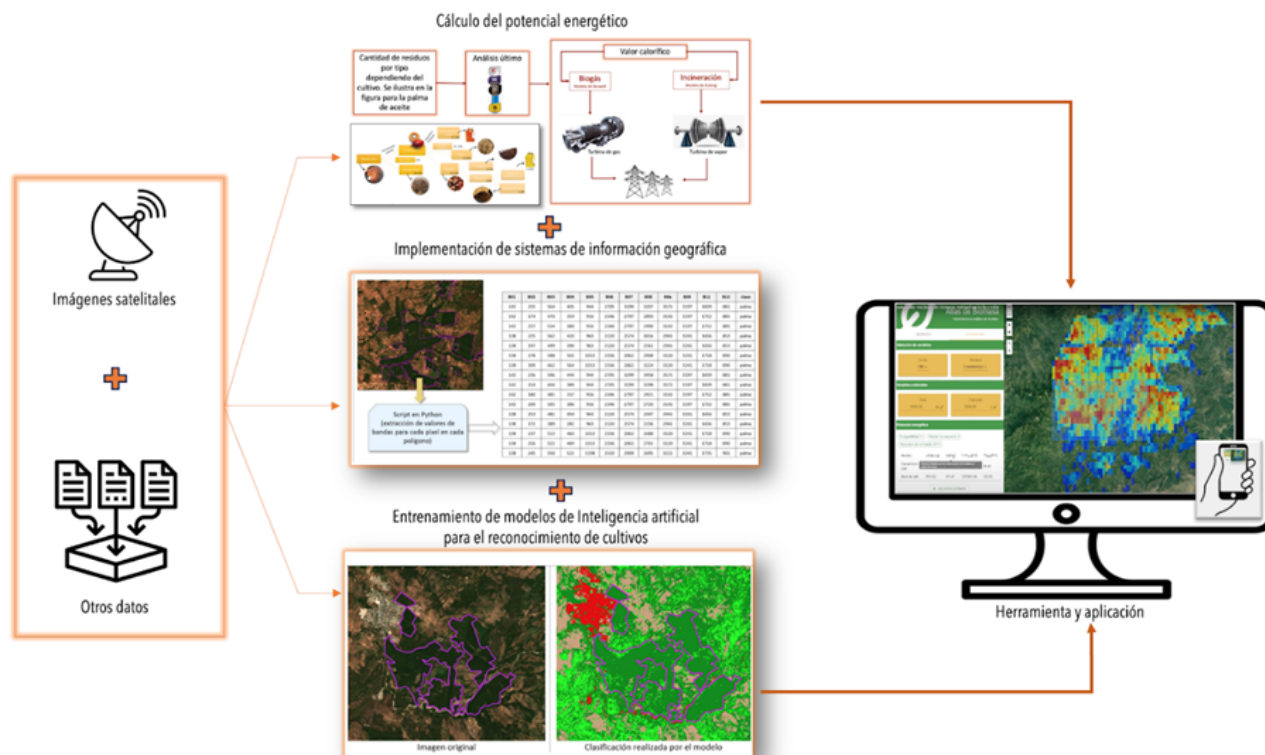
¹ [Informe Proyecto de Cooperación Triangular](#)

² [Evaluaciones Agropecuarias Municipales](#)

³ [Mapa de cobertura de tierras del IDEAM](#)

⁴ Polígonos suministrados por el IGAC, específicamente por la Subdirección de Agrología en el marco de un estudio de uso y cobertura con interpretación visual de los cultivos de palma de aceite y café para los municipios PDET del Departamento del Cesar (Agustín Codazzi, Becerril, La Jagua de Ibirico, Manaure Balcón de Cesar, Pueblo Bello, La Paz, San Diego) el cual se está ejecutando en el año 2023

Figura 1. Metodología aplicada a la prueba de concepto del Atlas de biomasa residual en departamentos de Bolívar y Cesar



Nota: Figura de elaboración propia

Gracias a estos avances en el desarrollo de los modelos mencionados anteriormente, ahora es posible determinar el potencial energético de la biomasa residual de palma de aceite en municipios vulnerables de Bolívar y Cesar. Esta tecnología está disponible a través de una aplicación que permite al usuario calcular el potencial energético eléctrico, dentro de los municipios que formaron parte de este estudio. Para obtener más información sobre cómo utilizar esta aplicación, puede consultar el tutorial que se encuentra disponible haciendo clic en este enlace [prueba de concepto](#).

Además de la evaluación del potencial energético de biomasa, este trabajo también dio lugar al desarrollo del "SenecAtlas". Esta plataforma integra diversos indicadores socioeconómicos, el explorador de potencial solar y de biomasa y otra información relevante en una única interfaz. Esto brinda a los usuarios la capacidad de generar datos y realizar cálculos que pueden ser fundamentales en la toma de decisiones. Si desea aprender más acerca de esta herramienta integrada, puede consultar el tutorial correspondiente haciendo [clic aquí](#).

Con el desarrollo de este proyecto se evidenció la utilidad de imágenes satelitales, de información georreferenciada y de modelos de inteligencia artificial en la identificación y estimación de áreas de cultivos de palma de aceite, que a su vez sirvieron de insumo para la estimación del potencial energético de la biomasa residual agrícola de palma, en una zona de estudio en los departamentos de Bolívar y del Cesar. Con los resultados del proyecto se generaron precedentes para desarrollar futuras pruebas de concepto, en el

desarrollo de un atlas de potencial energético de biomasa residual en Colombia, que abarque los principales cultivos del país y que pueda estar actualizado al usar como entrada imágenes satelitales vigentes, sobre las cuales se apliquen los modelos de inteligencia artificial para la identificación de un cultivo específico.

1.2. Estado del arte: Modelos de inteligencia artificial en la identificación de cultivos

La Teledetección ha emergido como una herramienta invaluable en la observación y comprensión de nuestro entorno terrestre a través de la captura y análisis de datos provenientes de sensores remotos (Campbell, 2023; Chuvieco, 2016; Lillesand et al., 2015). En este contexto, el avance vertiginoso de la Inteligencia Artificial (IA), específicamente el Machine Learning (ML) y el Deep Learning (DL), ha revolucionado la capacidad de procesamiento y extracción de información a partir de imágenes satelitales, ofreciendo posibilidades sin precedentes en una amplia gama de aplicaciones como la detección de cambios en el uso del suelo, la predicción de rendimientos agrícolas, la monitorización ambiental, gestión de recursos naturales, la mitigación de desastres naturales y la toma de decisiones en políticas medioambientales (Jeon, 2023; Li et al, 2018; Thapa et al. 2023) .

En el sector agrícola, la conjunción de estas dos tecnologías ha impulsado diversas aplicaciones prácticas en este campo, entre las que podemos citar: seguimiento dinámico de los cultivos, predicción de la productividad agrícola, segmentación y clasificación de cultivos, así como la detección de enfermedades y estrés en las plantaciones (Benos et al., 2021; Teixeira et al., 2023). Además, facilita una gestión óptima de la distribución de los cultivos y el aprovechamiento del suelo, entre otros usos posibles.

Especialmente en la detección y monitoreo de áreas de cultivo, una tarea crítica para la gestión agrícola y la seguridad alimentaria a nivel global, las técnicas de ML y DL han revelado un potencial extraordinario al posibilitar la automatización de la identificación de plantaciones mediante el análisis de imágenes satelitales. A continuación, se presentan algunas de estas aplicaciones centradas en cultivos relevantes para este trabajo como la caña de azúcar, el maíz, el arroz, el café, el cacao y la palma de aceite:

Caña de azúcar. La anticipación precisa del rendimiento de los cultivos resulta fundamental para planificar la cosecha y tomar decisiones estratégicas en cuanto a su comercialización. Esta necesidad ha impulsado múltiples investigaciones, como el trabajo de (Shendryk et al., 2021), donde proponen un modelo de machine learning sobre imágenes Sentinel-1 y Sentinel-2, combinadas con datos climáticos, de suelo y de elevación, para prever el rendimiento de la caña de azúcar a nivel de campo en diversas zonas. Por otro lado, (Kai et al., 2022) presenta una aplicación para el monitoreo de este cultivo que emplea técnicas de deep learning para diferenciar entre cuatro variedades de caña de azúcar, a partir del análisis de las bandas espectrales e índices de vegetación.

Maíz. En (Alves et al., 2022) se describe otra aplicación de cómo las técnicas de machine learning pueden ser utilizadas para la predicción de la productividad y rendimiento de cultivos utilizando información espectral. Los autores experimentan con diferentes algoritmos demostrando que es posible predecir el rendimiento del grano de maíz antes de la cosecha. Por su parte, en (Murguía-Cozar et al., 2021) se utilizan métodos de clasificación

supervisada para determinar los principales indicadores y características que permiten discriminar las etapas fenológicas del maíz en imágenes satelitales Sentinel 2.

Arroz. Un método para extraer áreas de plantaciones de arroz a partir de series temporales de imágenes de radar de Sentinel-1 e imágenes multiespectrales de Sentinel-2 se propone en (Saadat et al., 2022). Este enfoque utiliza una red convolucional, lo que resulta en una mejora significativa en la cartografía de este cultivo. Por otro lado, en el trabajo de (Zhang et al., 2020), se lleva a cabo una comparación entre estos modelos y técnicas más convencionales, demostrando la eficacia de estas metodologías en la identificación precisa de las áreas de cultivo de arroz en las imágenes capturadas por Sentinel.

Café. Un enfoque para la predicción del rendimiento del café que combina imágenes de satélite de alta resolución y datos de nutrientes foliares se describe en (Martello et al., 2024). Los resultados muestran que es posible estimar el rendimiento hasta nueve meses antes de la cosecha, lo cual resulta crucial para planificar este cultivo y potenciar la rentabilidad de la industria cafetera. La aplicación de algoritmos de machine learning y deep learning a imágenes satelitales también permite identificar los patrones de salud en las plantaciones de café, facilitando la detección de áreas con problemas de crecimiento, según señalan Rocha y Carvalho (2020). Asimismo, otro estudio propone un modelo de deep learning U-Net para clasificar las regiones de plantación de café utilizando datos de Sentinel-2, lo que contribuye al seguimiento anual de los cambios en las áreas cultivadas (Toan et al. 2022).

Cacao. En el trabajo de Batista y colaboradores (2022), se emplearon datos satelitales y técnicas de machine learning para identificar y cartografiar áreas de cultivo de cacao. Este análisis permitió distinguir sus características específicas y precisar las zonas donde se encuentra cultivado. El propósito fundamental es que este conocimiento, acerca de la extensión y ubicación de las plantaciones, sea aprovechado para respaldar tanto la gestión comercial como la conservación medioambiental. Otro estudio se describe en (Numbisi et al., 2019), enfocado en delimitar las áreas de cultivo de agro bosques de cacao, con el objetivo de medir la contribución de la expansión de la actividad agrícola a la deforestación tropical. Para este fin, se emplearon imágenes SAR de Sentinel 1, combinadas con operadores de texturas y técnicas de machine learning.

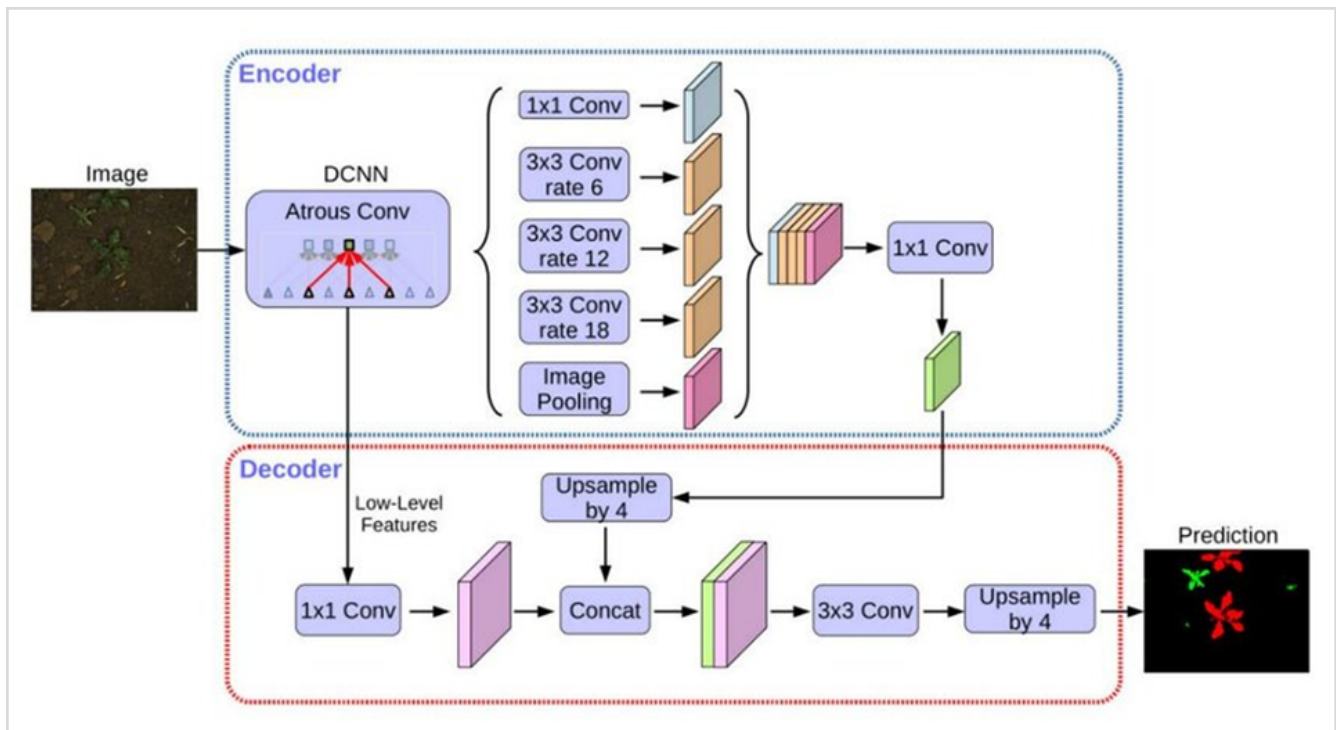
Palma de aceite. Para el caso del cultivo de palma de aceite, las imágenes de satélite, junto con las técnicas de machine learning, se han utilizado para apoyar la toma de decisiones en diferentes procesos asociados con, por ejemplo: cuantificación de plantaciones, con el fin de construir una cartografía más precisa que permita estimar impacto ambiental (Rodríguez *et al.*, 2021; Xu *et al.*, 2023) y económico (Rosyidy *et al.*, 2023); predicción del rendimiento del cultivo, a partir de datos históricos de imágenes de satélite y registros de rendimientos pasados. Esta información puede ser valiosa para la planificación de la cosecha y la gestión de la producción (Watson *et al.*, 2022), y monitoreo del estado fenológico de la planta, lo cual puede ayudar a los agricultores a optimizar las prácticas de manejo y tomar decisiones informadas sobre la aplicación de fertilizantes y otros insumos (Hernawati *et al.*, 2022).

Mención especial merece el trabajo de Descals y colaboradores (2021)⁵, el cual construye un mapa de plantaciones de palma de aceite a escala global. El modelo se basa en DeepLabv3+ (Chen et al, 2018), una red

⁵ <https://essd.copernicus.org/articles/13/1211/2021/>

neuronal convolucional (CNN) para segmentación semántica ⁶, la cual fue entrenada para clasificar imágenes Sentinel-1 y Sentinel-2 en un mapa de cobertura terrestre de palma de aceite. Esta red pudo distinguir entre plantaciones industriales y de pequeños agricultores a nivel global con una precisión general del 98.52%, superando a conjuntos de datos regionales existentes que utilizan algoritmos convencionales de machine learning. La arquitectura de este modelo de segmentación se muestra en la Figura. 2.

Figura 2. Arquitectura encoder-decoder de Deeplabv3+



Nota: Figura tomada de Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation (2018), Chen, LC., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. (2018).

Con base en este modelo se construyó el proyecto BIOPAMA (*Industrial and smallholder oil palm plantations mapping*), en cuyo sitio⁷ es posible visualizar el mapa mundial de la palma de aceite con distintos niveles de detalle: Nivel-1 (probabilidad de plantación de palma aceitera madura), Nivel-2 (mapa temático con tres clases: palma aceitera madura industrial, palma aceitera madura de pequeños propietarios, y otros usos de la tierra), y Nivel-3 (mapa temático con cinco clases: palma aceitera joven y madura industrial, palma aceitera joven y madura de pequeños propietarios, y otros usos de la tierra).

⁶ La segmentación de imágenes tiene como objetivo vincular cada píxel de una imagen a una etiqueta de clase. Por lo tanto, la segmentación semántica es el análogo de los algoritmos estándares de machine learning que se utilizan en la teledetección para la clasificación de imágenes. La diferencia es que la segmentación semántica, como cualquier modelo basado en una red neuronal convolucional, aprende y explota automáticamente los patrones espaciales dentro de la imagen mediante el ajuste de los parámetros de diferentes operaciones convolucionales.

⁷ <https://rris.biopama.org/oilpalm>

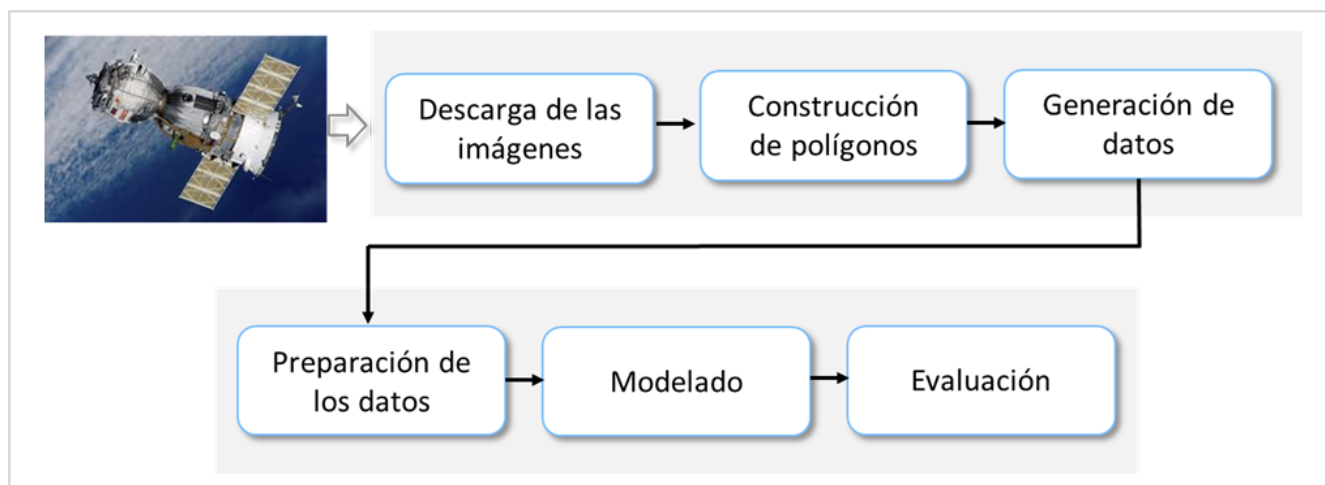
2. Valoración de la pertinencia del modelo de identificación de palma de aceite a nivel nacional

En esta sección se hace un resumen de la metodología aplicada para el desarrollo del modelo de inteligencia artificial que identifica palma de aceite, previamente desarrollado en el Proyecto de Cooperación Triangular Alemania-Chile- Colombia (Sierra, y otros, 2023), así como, los resultados de la aplicación de este modelo para la identificación de palma a nivel nacional en Colombia.

2.1. Metodología aplicada para el desarrollo del modelo de inteligencia artificial que identifica palma de aceite

Para el desarrollo del modelo de clasificación de palma de aceite sobre imágenes satelitales se siguió la metodología ilustrada en la Figura 3, con base en el ciclo de machine learning (Geron, 2002). A continuación, se explican cada uno de los pasos realizados.

Figura 3. Metodología para la construcción del modelo de palma de aceite



Nota: Figura de elaboración propia

1. Descarga de las imágenes de la zona de estudio

A través del portal de datos abiertos de Copernicus se realizó la descarga de las imágenes de Sentinel-2 de las zonas de interés, empleando para ello información de coordenadas y rango de fechas. Es importante resaltar que las imágenes recolectadas cubren solo los municipios de los departamentos del Cesar y Bolívar.

2. Construcción de polígonos de coberturas

Los polígonos de áreas de cultivos de palma de aceite sobre imágenes Sentinel 2 en los departamentos del Cesar y Bolívar fueron suministrados por el Instituto Geográfico Agustín Codazzi (IGAC)⁸. Por otra parte, en este trabajo se consideró utilizar un aprendizaje supervisado para la construcción de los modelos de clasificación, lo que conlleva la identificación de polígonos relacionados con otras coberturas. Esto debido a que estos algoritmos de aprendizaje requieren, además de la información de la clase de interés, datos que representen lo que no es palma de aceite. Así, fueron construidos polígonos asociados a otras coberturas para las siguientes cuatro clases: suelo desnudo, bosque y área seminatural, zona urbanizada y superficie de agua. A tal efecto, se utilizó la herramienta de creación de polígonos provista por el programa QGIS⁹.

3. Generación de los datos para el aprendizaje

Para poder aplicar los algoritmos de aprendizaje es necesario obtener una representación estructurada a partir de las imágenes satelitales. Esta representación es una matriz o tabla de dos dimensiones, en la cual cada banda representará un atributo (columna) y cada píxel será una fila, caracterizada por los valores de las diferentes bandas. Además, será necesario atribuir a cada píxel su clase (tipo de cobertura), que será una columna adicional.

Para transformar los valores de bandas de los polígonos a un formato tabular se construyó un script en el lenguaje de programación Python con apoyo de la librería GDAL¹⁰, el cual extrae la información de cada píxel en cada una de las bandas. En la tabla, cada registro corresponde a un píxel etiquetado con una de las clases.

4. Preparación de los datos.

El procesamiento de datos consistió, inicialmente, en la eliminación de todos los registros que tuvieran valores en cero, ya que los rangos de valores válidos para las bandas B01 a B12 inician desde uno y, por lo tanto, se asumió que estos podrían ser errores en el conjunto. Luego, se añadieron dos capas adicionales relacionadas con un modelo digital de elevación (DEM) de la zona de estudio y la pendiente (SLOPE), que representa la tasa de cambio de elevación para cada celda del modelo digital de elevación. Para esta tarea se descargó el ALOS

⁸ <https://www.igac.gov.co/>

⁹ QGIS (también conocido como Quantum GIS) es un Sistema de Información Geográfica (SIG) de código abierto y multiplataforma, que permite visualizar, editar y analizar datos geoespaciales en diferentes formatos. QGIS se utiliza ampliamente en disciplinas como la cartografía, la planificación urbana, la gestión de recursos naturales, la agricultura, la geología y muchos otros campos relacionados con la geografía.

¹⁰ La librería GDAL (*Geospatial Data Abstraction Library*) es una biblioteca de software de código abierto utilizada para leer, escribir y manipular datos geoespaciales en diversos formatos, incluyendo imágenes satelitales. Proporciona una interfaz común para acceder a datos geoespaciales y facilita el desarrollo de aplicaciones, el intercambio de datos y el procesamiento de información geográfica.

PALSAR DEM, recurso cartográfico disponible dentro de los productos del satélite ALOS¹¹. Por último, se hizo un análisis de la firma espectral de todas las coberturas con base en los datos. Con base en el comportamiento observado se decidió aplicar un filtro para depurar las anomalías al interior de cada una de las clases. Para tal fin, se calculó el promedio y la desviación estándar de todas las bandas, junto con las capas de pendiente y el modelo digital de elevación. El promedio se asumió como la representación de la firma espectral para la clase y se generó un rango de 1.5 desviaciones estándar alrededor de esta. Con los límites superior e inferior de dicho rango se generaron filtros que fueron utilizados para depurar (excluir) todos aquellos píxeles que no estuvieran dentro del rango establecido.

5. Construcción del modelo de clasificación.

Se realizaron experimentos con cinco algoritmos de clasificación implementados en las librerías de Python Scikit-Learn¹² y XGBoost¹³: regresión logística, árboles de decisión, Random Forest, máquinas de vectores de soporte y XGBoost. Para todos ellos se realizó una búsqueda de hiperparámetros con un conjunto de validación con el fin de obtener la mejor configuración con base en las métricas de rendimiento utilizadas.

6. Evaluación del modelo.

Para la validación se utilizaron las métricas de desempeño disponibles en las librerías utilizadas para el entrenamiento. Particularmente, se utilizaron métricas como la exactitud o *accuracy*, la sensibilidad o *recall*, la precisión y el F1-Score, calculadas para cada una de las clases y para el desempeño general del modelo.

En la Tabla 1 pueden apreciarse los valores de las métricas de rendimiento que miden el poder de predicción del modelo. Estos fueron obtenidos sobre el conjunto de test, el cual incluye píxeles que no fueron utilizados para la construcción del modelo. La Tabla 2 muestra el rendimiento sobre la clase de interés

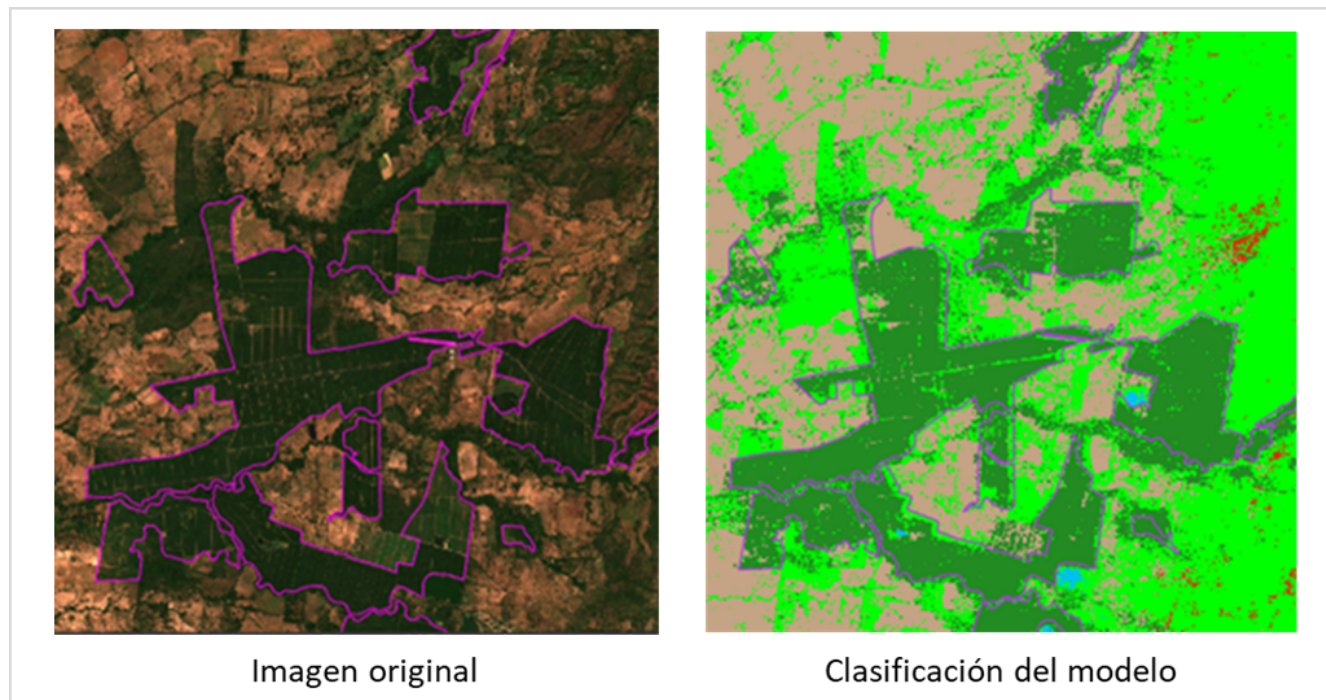
Para visualizar las predicciones de los modelos de clasificación se construyó un script en lenguaje de programación Python el cual “colorea” una nueva imagen en la que se indica cada clase estimada para cada píxel. La Figura 4 muestra la interpretación que hace el modelo de las diferentes coberturas para una región de la zona de estudio.

¹¹ El Satélite Avanzado de Observación Terrestre (ALOS), también conocido como DAICHI, fue una misión satelital japonesa que operó de 2006 a 2011. Llevaba tres instrumentos, incluido el radar de apertura sintética de banda L tipo Phased Array (PALSAR), que se utilizó para obtener observaciones detalladas de la superficie de la Tierra, de día y de noche, en cualquier condición meteorológica. Los datos PALSAR podrían adquirirse en múltiples modos, con diferentes polarizaciones, resoluciones, anchos de franja y ángulos fuera del nadir lo que lo constituía como una herramienta adecuada para la generación de modelos digitales de elevación. PALSAR se utilizó para una variedad de aplicaciones, incluida la cartografía, la observación precisa de la cobertura del suelo a escala regional, el monitoreo de desastres y el estudio de recursos. Fue una herramienta valiosa para científicos e investigadores, y sus datos siguen utilizándose para diversos fines en la actualidad. (<https://asf.alaska.edu/data-sets/sar-data-sets/alos-palsar/alos-palsar-about/>)

¹² <https://scikit-learn.org/stable/>

¹³ <https://xgboost.readthedocs.io/en/stable/>

Figura 4. Metodología para la construcción del modelo de palma de aceite



Nota: Figura tomada de *Informe 4,5,6,7: Exploradores del potencial solar y biomasa residual-Proyecto de Cooperación Triangular Alemania-Chile-Colombia* (2023). Sierra, R., Jiménez, J., Jiménez, G., Núñez, H., Bressan, M., Giraldo, L, & Narváez, G.

2.2. Resultados de la aplicación del modelo de inteligencia artificial para la identificación de palma de aceite a nivel nacional.

Para la aplicación del modelo obtenido en el área de estudio anteriormente descrita, se procedió a la descarga de imágenes satelitales que cubrieran la extensión del territorio continental colombiano. En total se descargaron 1270 escenas obtenidas por el satélite Sentinel 2 cubriendo las bandas principales del espectro visible y el cercano y medio infrarrojo. Dichas escenas se descargaron usando el protocolo STAC¹⁴ y consultando directamente los repositorios abiertos accesibles desde la plataforma de Planetary Computer¹⁵.

Para satisfacer las entradas del modelo ya entrenado también se descargaron 149 imágenes de modelos digitales de elevación de la campaña ALOS PALSAR desde los repositorios de Alaska Satellite Facility. A partir de estos modelos se construyó una capa adicional referente a las pendientes de terreno.

En total se descargaron alrededor de 13000 imágenes satelitales que corresponden a más de 1.2Tb de

¹⁴ <https://stac-utils.github.io/qgis-stac-plugin/user-guide/>

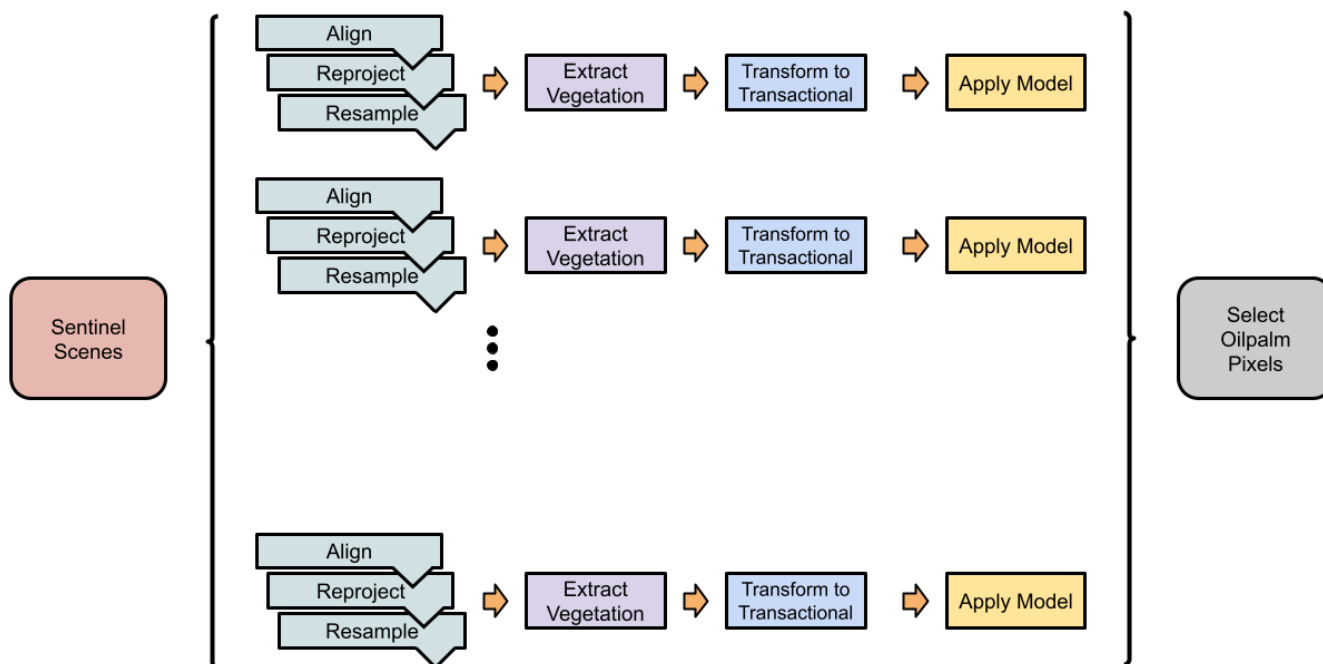
¹⁵ <https://planetarycomputer.microsoft.com/>

almacenamiento en disco y que significa analizar alrededor de 500.000 millones de píxeles.

Una vez descargadas las imágenes se procedió a procesar su contenido con el objetivo de transformarlas en el formato requerido para ser leídas por el modelo. Entre los procedimientos aplicados a cada una de las escenas están el alineamiento, reproyección y reajuste de todas las imágenes dado que estas provenían de diferentes orígenes, con distintos sistemas de coordenadas y bajo diversas resoluciones espaciales. Igualmente, se requirió extraer solo aquellos píxeles correspondientes a vegetación y eliminar aquellos afectados por nubes. Finalmente, se extrajo el valor de cada píxel y se transformó a una tabla transaccional como es requerido como entrada para la aplicación del modelo.

Para gestionar tal volumen de información se programaron scripts que procesaban las imágenes de cada escena de manera distribuida y en paralelo en el servidor dispuesto para tal fin. La Figura 5 muestra de manera esquemática los pasos del procesamiento de cada una de las escenas. Al final del proceso, se recogen únicamente los píxeles clasificados por el modelo como palma de aceite y se utiliza dicho conjunto de datos para la visualización y como insumo de entrada para la aplicación propuesta.

Figura 5. Metodología de procesamiento de imágenes



Nota: Figura de elaboración propia

2.2.1 Validación del modelo y resultados obtenidos

Una vez obtenido el conjunto de datos con la ubicación de las posibles zonas cultivadas con palma de aceite, se procedió a una evaluación y validación de resultados de manera visual para contrastar la consistencia de los resultados.

Para tal fin, se seleccionaron tres zonas por fuera del área de estudio donde se entrenó el modelo y se comparó

con información visual del área así como con un modelo de referencia que también clasificó palma de aceite en la zona y a nivel nacional. Para tal fin se utilizó el modelo BIOPAMA descrito en la Sección 1.2 de este informe.

Las figuras 6A, 6B y 6C ilustran las características del terreno de cada zona (a la izquierda) donde se señalan áreas de cultivo de palma con flechas amarillas. Posteriormente se muestran las ubicaciones clasificadas como cultivos de palma por el modelo evaluado (en rojo) y por el modelo BIOPAMA (en morado). A la derecha de cada figura se muestra un solapamiento de ambos modelos.

Figura 6. Resultados de la aplicación de modelos de identificación de palma de aceite

Figura 6A. Zona 1 de comparación de modelos.

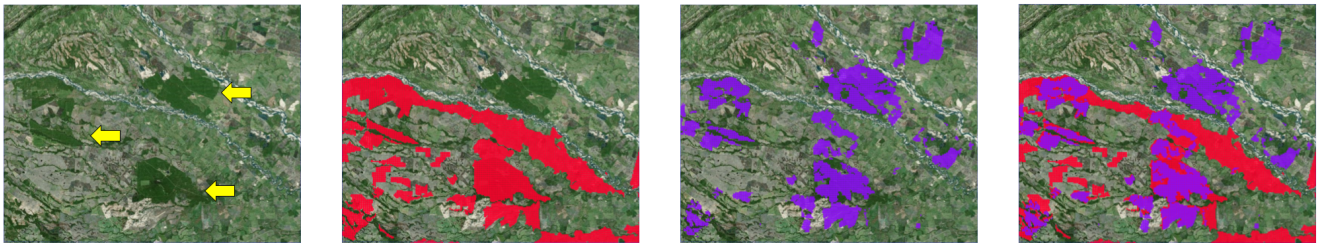


Figura 6B. Zona 2 de comparación de modelos.

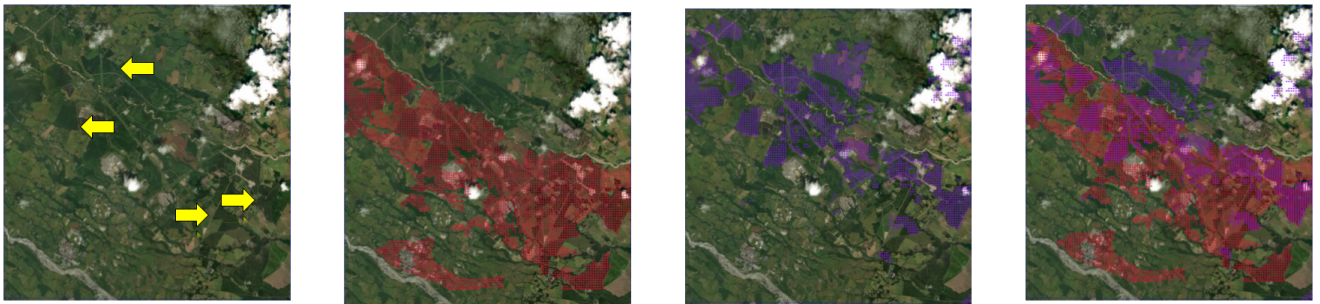
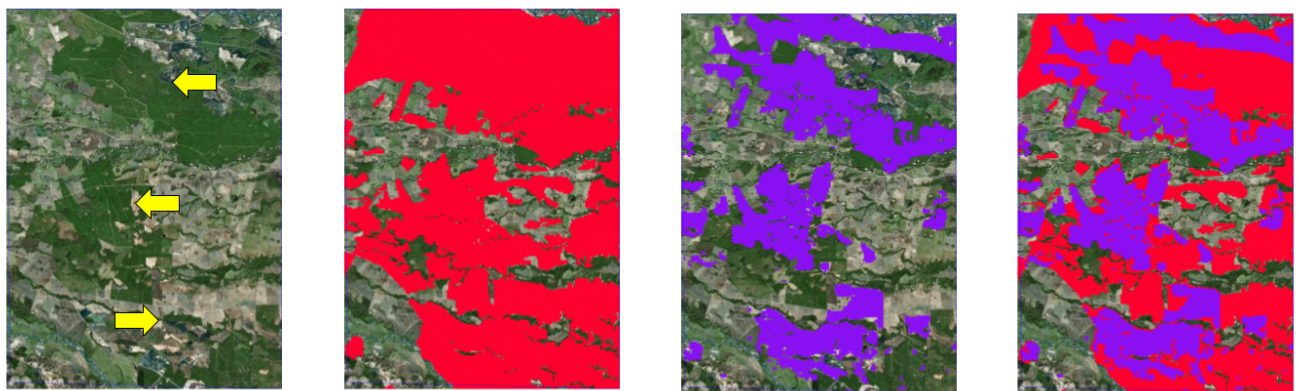


Figura 6C. Zona 3 de comparación de modelos.



Nota: Figura de elaboración propia con los resultados de la aplicación de modelos de identificación de palma de aceite

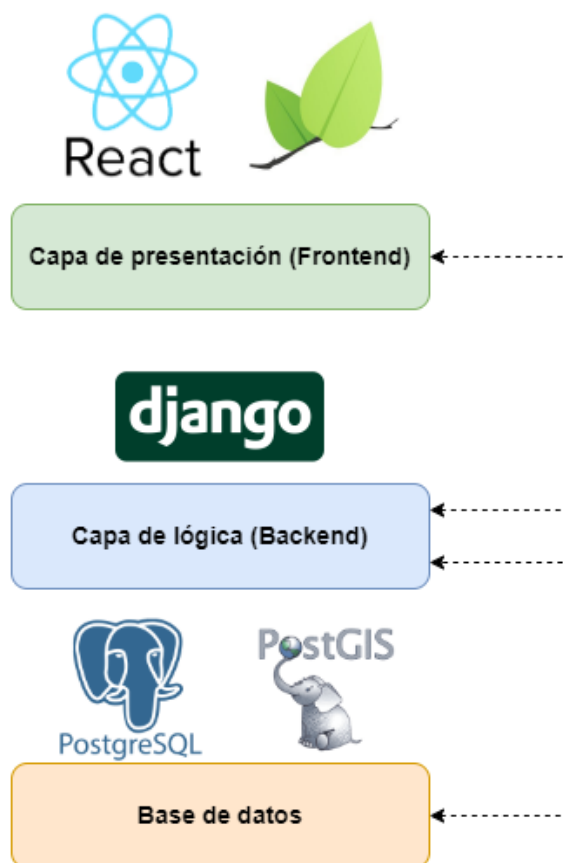
Desafortunadamente, a partir de la inspección visual se puede distinguir un claro overfitting por parte del

modelo evaluado clasificando áreas que no corresponden a la cobertura en cuestión como si se trataran de palma de aceite. Al carecer de polígonos de prueba por fuera del área de estudio inicial donde se entrenó el modelo resulta imposible cuantificar la precisión del modelo evaluado en estas zonas. Por otra parte, el desempeño del modelo de referencia resulta satisfactorio en las áreas comparadas. Como se menciona al final de la sección 1.2, el modelo BIOPAMA cuenta con una precisión global superior al 98% y dentro de las áreas comparadas su probabilidad intraclase se acerca al 95%. En el modelo evaluado la probabilidad intraclase se encuentra alrededor del 55%.

3. Metodología del desarrollo de la aplicación de la herramienta del atlas de biomasa residual para la identificación de palma de aceite a nivel nacional

La herramienta del atlas de biomasa residual para la identificación de palma de aceite permite a los usuarios interactuar con un mapa a nivel nacional para calcular, a partir de un perímetro seleccionado, el potencial de área cultivada, producción de palma de aceite, cantidad de residuos agrícolas y el potencial energético de los residuos al aplicar diferentes técnicas. A continuación se muestra la arquitectura diseñada para la construcción de la aplicación en la Figura 7:

Figura 7. Arquitectura general de la aplicación.



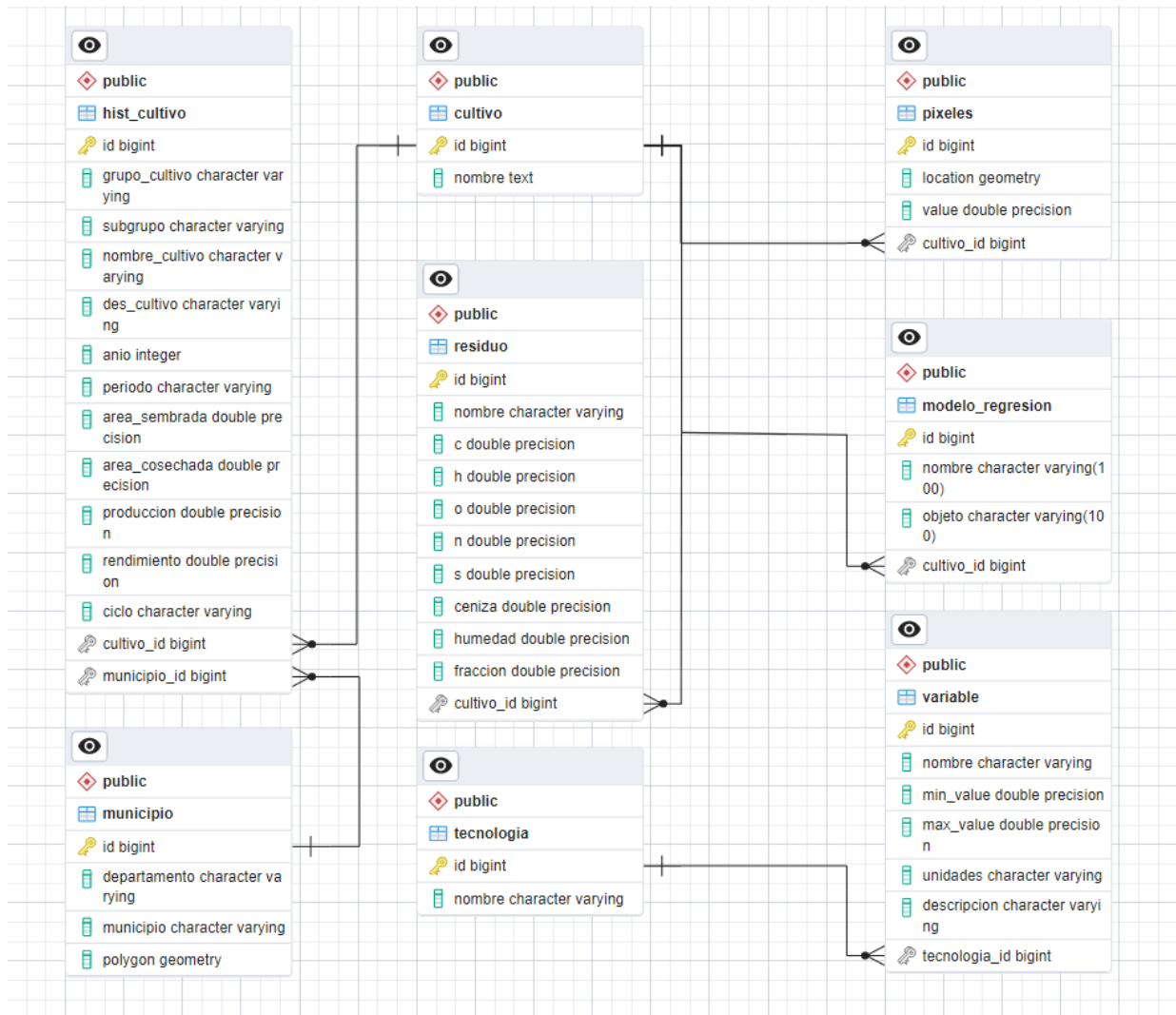
Nota: Figura de elaboración propia

La arquitectura sigue un diseño de tres capas:

1. **Base de datos:** basada en el sistema de gestión de bases de datos relacionales PostgreSQL y su extensión espacial PostGIS. PostgreSQL se encarga de la creación y almacenamiento de tablas que almacenan la información utilizada en la aplicación, haciendo uso de la extensión PostGIS para trabajar eficientemente con objetos espaciales como polígonos y puntos geográficamente referenciados.
2. **Capa de lógica:** la capa de lógica o backend es la encargada de comunicarse con la capa de datos para extraer y procesar la información relevante de acuerdo con las peticiones realizadas por el usuario desde la capa de presentación, sirviendo el resultado correspondiente. En ese orden de ideas, funciona como un puente entre la capa de datos y la capa de presentación. Esta capa está soportada por el marco de trabajo Django, basado en el lenguaje de programación Python y popularmente utilizado en el desarrollo de aplicaciones web.
3. **Capa de presentación:** la capa de presentación o frontend es la encargada de presentar la interfaz de usuario de la herramienta. Esta capa procesa las peticiones del usuario y las envía a la capa de lógica, recibiendo el resultado y mostrándolo al usuario. Específicamente, esta capa utiliza la librería ReactJS para el procesamiento de las entradas del usuario durante su interacción con la aplicación, así como la librería Leaflet para la gestión y visualización de mapas en la web.

Específicamente para la base de datos, se almacenaron los detalles geográficos de la extensión del cultivo de palma de aceite mediante puntos, cada uno referenciado mediante coordenadas. Igualmente, para la identificación de municipios, se almacenaron polígonos que delimitan sus extensiones geográficas. El sistema de coordenadas utilizado es el EPSG 9377 Magna-Sirgas / Origen Nacional para el almacenamiento de los datos, aunque se realizó la proyección al sistema de coordenadas EPSG 3857 Pseudo-mercator al momento de visualizar los mapas en la capa de presentación. Con respecto a la información de residuos, tecnologías e información histórica, se siguió una estructura tabular convencional utilizando tipos de datos genéricos ofrecidos por PostgreSQL. A continuación, en la Figura 8 se muestra el diagrama Entidad-Relación de la base de datos:

Figura 8. Diagrama entidad-relación para la aplicación



Nota: Figura de elaboración propia

Es importante resaltar que la estructura de la base de datos permite la adición de nuevos cultivos, es decir, no está limitada a palma de aceite. En ese sentido, actualmente se tiene información histórica obtenida de las Evaluaciones Agropecuarias Municipales para una mayor cantidad de cultivos. Esto puede ser beneficioso si se desea ampliar el alcance de la herramienta en un futuro, permitiendo la visualización de múltiples cultivos en mapas, así como la definición de residuos, tecnologías y variables para la estimación del potencial.

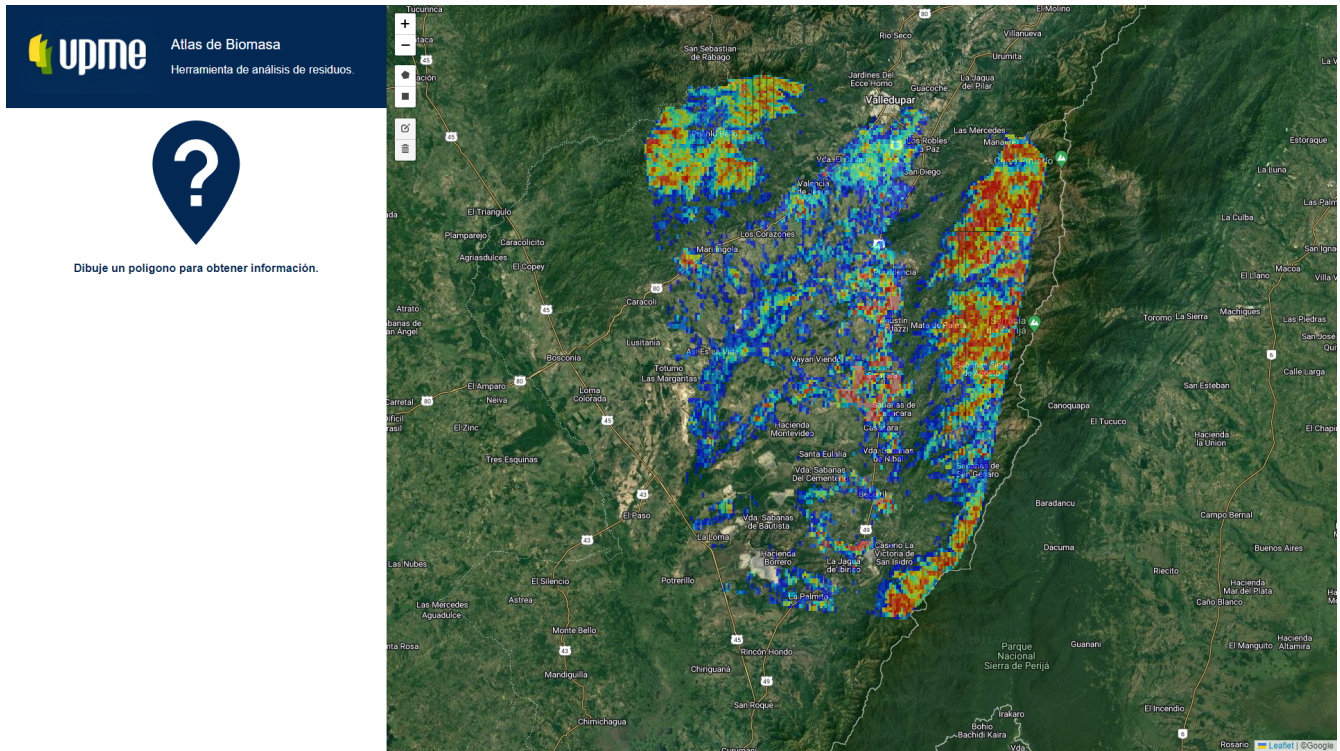
Con respecto a la capa de lógica, se utiliza el patrón MVC (Modelo-Vista-Controlador), permitiendo la representación de objetos de bases de datos (mediante el modelo) y la interacción por parte del usuario (mediante la vista y el controlador). En ese sentido, el servidor expone una interfaz de programación de aplicaciones o API, en la que se encuentran puntos finales o *endpoints* que permiten responder a peticiones del usuario. Estas peticiones pueden ser la consulta de información de la base de datos, e incluso el uso de modelos de aprendizaje automático para realizar predicciones. En ese orden de ideas, el servidor expone los siguientes puntos finales para su uso en la aplicación descritos en la Tabla 1:

Tabla 1. Puntos finales disponibles en la capa de lógica.

Endpoint	Descripción
mapas/pixeles	Retorna puntos en formato GeoJSON para su visualización
biomasa/municipios	Retorna polígonos en formato GeoJSON. Permite consultar si un punto pertenece a un municipio.
biomasa/cultivos	Retorna una lista de cultivos analizados por la aplicación.
biomasa/cultivos/hist	Retorna una lista de registros históricos de un cultivo particular.
biomasa/residuos	Retorna una lista de residuos por cultivo.
biomasa/tecnologias	Retorna una lista de tecnologías contempladas en la aplicación.
biomasa/variables	Retorna una lista de variables utilizadas según la tecnología.
biomasa/inference	Retorna una predicción de la producción de acuerdo con el área sembrada del cultivo.

Finalmente, la capa de presentación muestra una interfaz gráfica construida haciendo uso del lenguaje de programación Javascript, específicamente de la librería React. Esta librería facilita el desarrollo de aplicaciones de una sola página, haciendo el papel de la vista en el patrón MVC, pues la capa de lógica solamente se encarga de implementar modelos y controladores. A continuación, se presenta una descripción general de la interfaz gráfica de la aplicación:

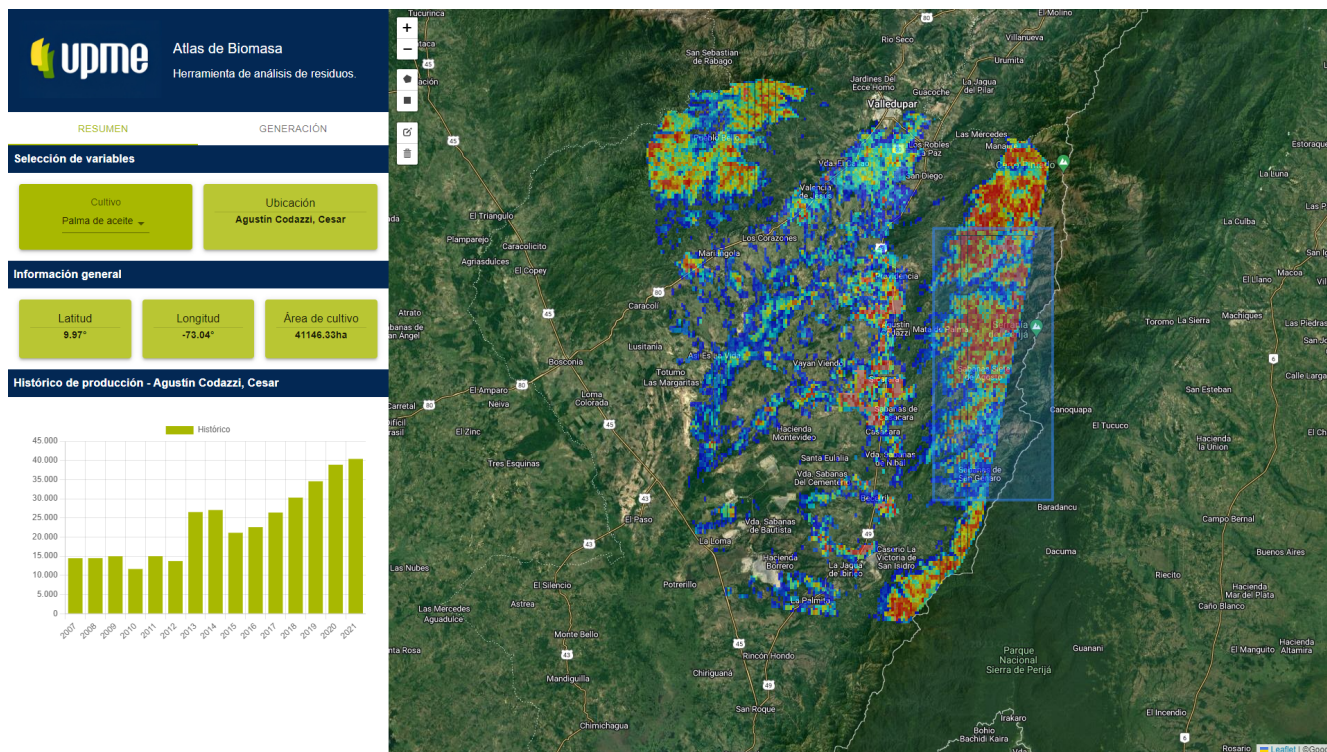
Figura 9. Vista inicial de la aplicación



Nota: Figura tomada de la aplicación de atlas de biomasa residual

La vista principal presenta un mapa con la clasificación realizada por el modelo de aprendizaje automático para la identificación de cultivos de palma de aceite en la (ver Figura 9). En este mapa, el usuario es capaz de seleccionar una herramienta para el dibujo de polígonos como se observa en la Figura 10, lo que a su vez permite obtener información del área y estimar el potencial energético generado.

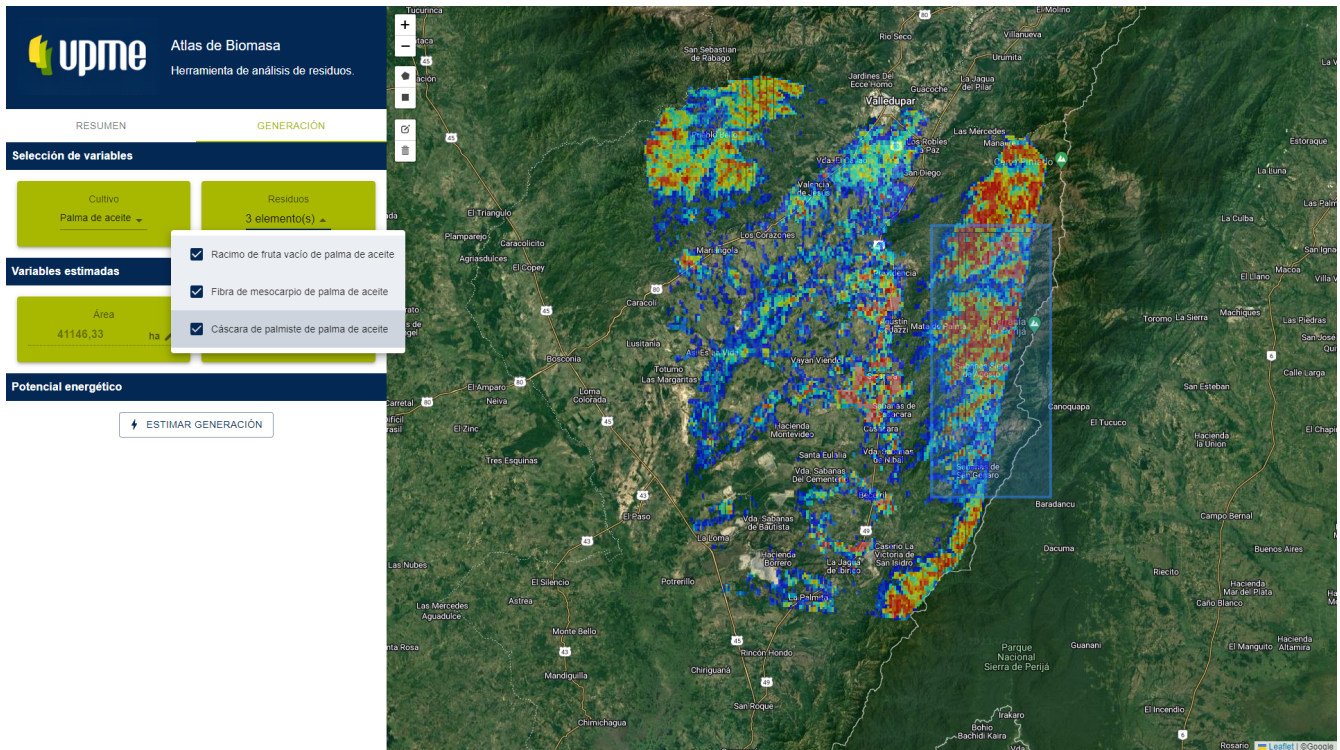
Figura 10. Vista general de la aplicación - histórico de producción.



Nota: Figura tomada de la aplicación de atlas de biomasa residual

Al dibujar un polígono, la aplicación automáticamente determina el municipio en el que se encuentra el área y retorna una gráfica con información histórica obtenida de las Evaluaciones Agropecuarias Municipales. Si se hace clic en la pestaña Generación, la herramienta muestra el menú de selección de residuos.

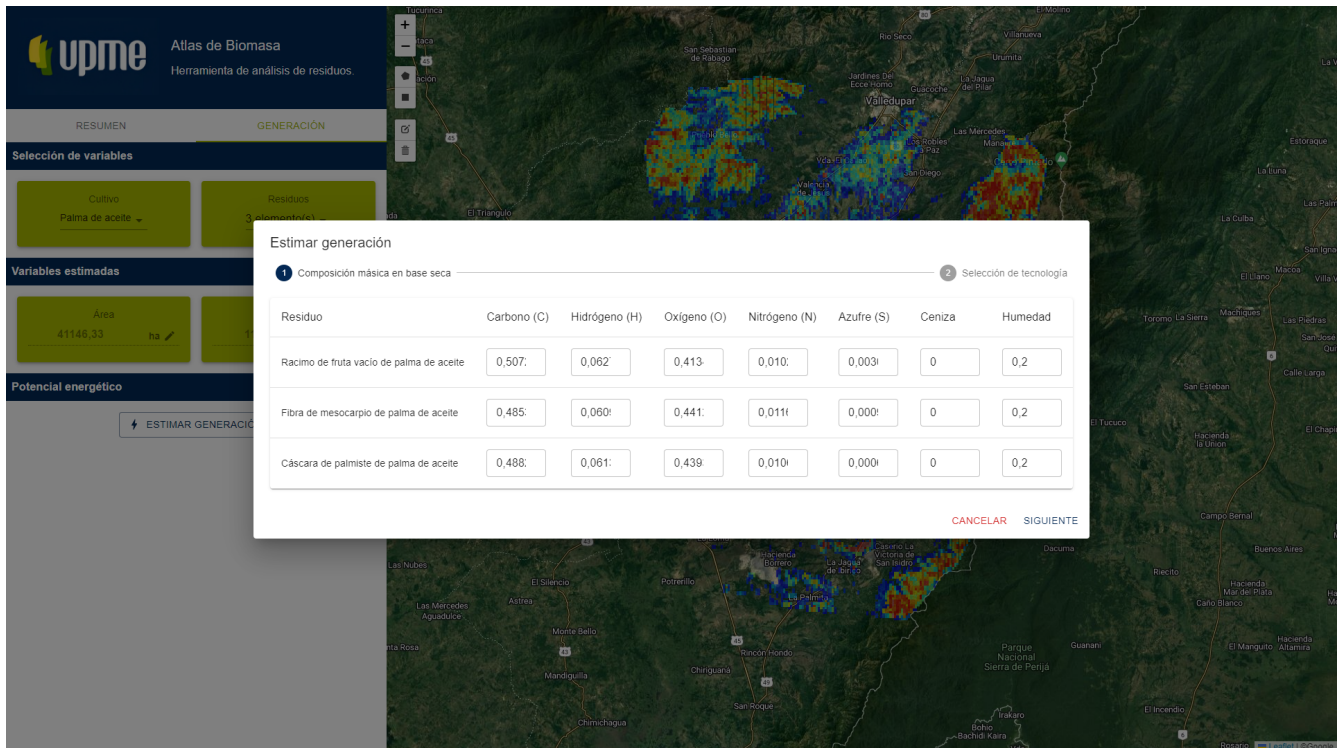
Figura 11. Vista de generación de la aplicación - selección de residuos.



Nota: Figura tomada de la aplicación de atlas de biomasa residual

El usuario puede seleccionar un subconjunto de los residuos disponibles para el cultivo de palma de aceite, así como modificar tanto el área estimada como la producción en toneladas calculada por el modelo de aprendizaje automático, según Figura 11. Una vez se tenga la información deseada, el usuario puede hacer clic sobre el botón de Estimar Generación, lo que abrirá un diálogo con un formulario como se observa en la Figura 12.

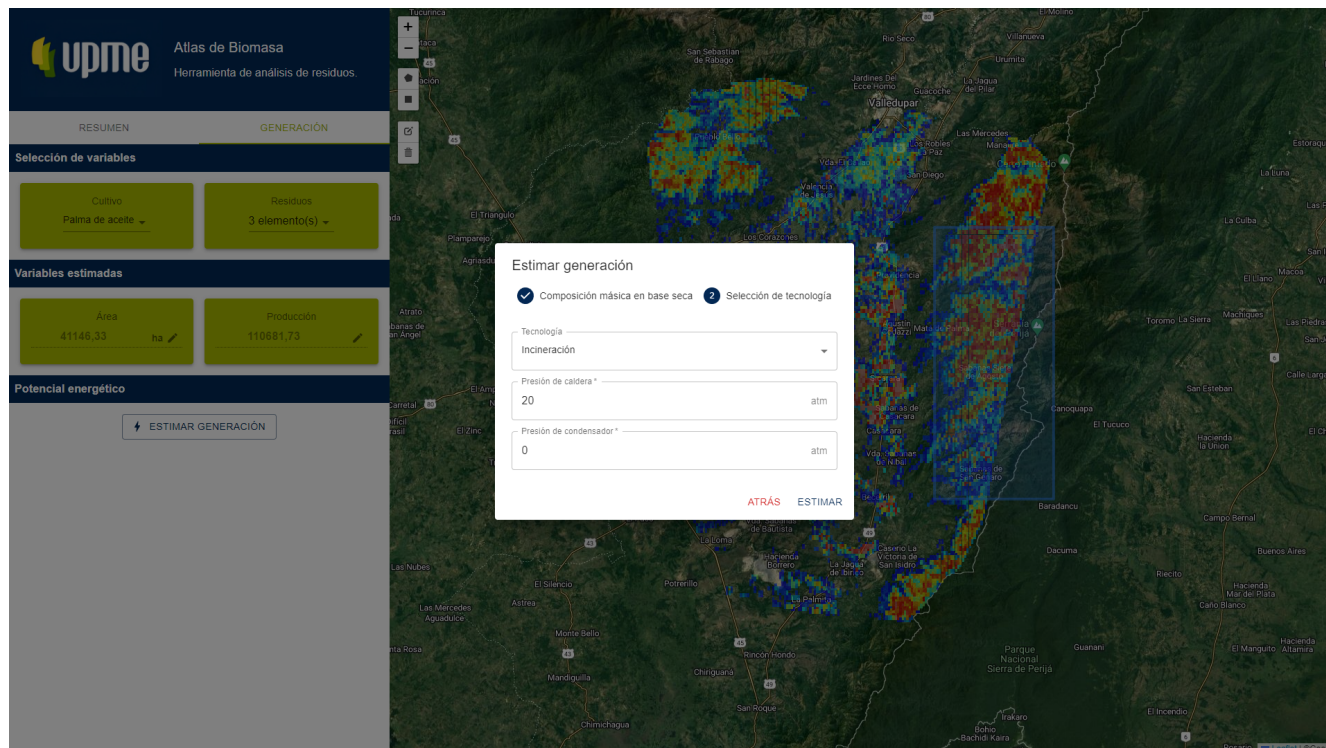
Figura 12. Vista de generación - composición másica en base seca.



Nota: Figura tomada de la aplicación de atlas de biomasa residual

En la primera página, el usuario puede seleccionar la fracción másica en base seca para los elementos carbono, hidrógeno, oxígeno, nitrógeno y azufre, así como la fracción de cenizas y humedad por cada residuo seleccionado como se observa en la Figura 12. Los valores por defecto son obtenidos desde la base de datos. Al hacer clic en el botón Siguiente, se muestra el formulario de selección de tecnologías.

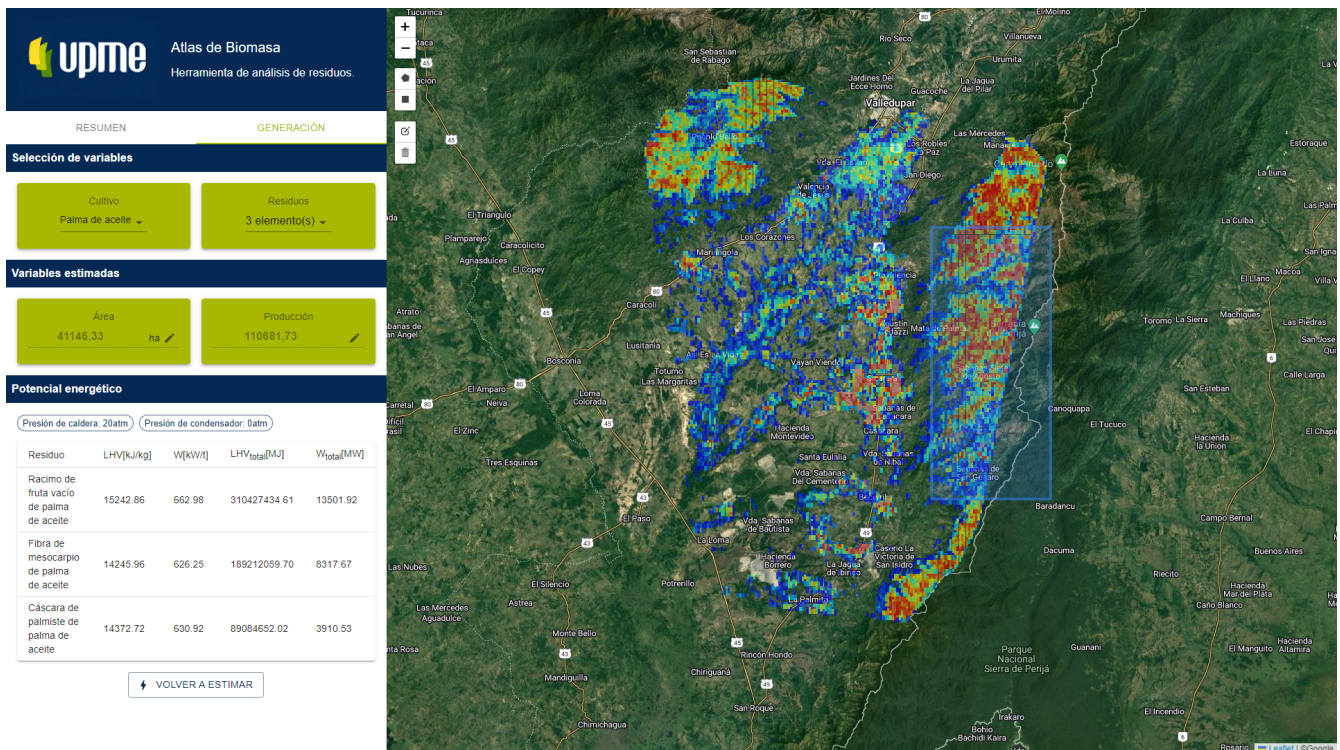
Figura 13. Vista de generación - selección de tecnología de conversión.



Nota: Figura tomada de la aplicación de atlas de biomasa residual

En este formulario, el usuario puede seleccionar el método de conversión de los residuos entre incineración y biogás como se observa en la Figura 13. Cada tecnología tiene asociadas variables y, a su vez, cada variable tiene valores por defecto que provienen de la información almacenada en la base de datos. Al igual que en la pantalla anterior, el usuario tiene la capacidad de modificar los valores por defecto. Finalmente, al hacer clic en el botón Estimar, se modifica la vista principal con un resumen del potencial energético calculado para la producción, residuos y tecnología seleccionados como se observa en la Figura 14.

Figura 14. Vista de generación - resultados de la estimación.



Nota: Figura tomada de la aplicación de atlas de biomasa residual

Debajo de la tabla con los resultados, el usuario puede hacer clic en el botón de Volver a estimar para utilizar nuevos valores y rehacer el cálculo del potencial.

4. Conclusiones y recomendaciones

Tras completar la ampliación de la prueba de concepto del Atlas Energético de Biomasa Residual Agrícola en Colombia, se pueden extraer las siguientes conclusiones significativas:

- La aplicación del modelo de inteligencia artificial, entrenado exclusivamente con datos provenientes de dos departamentos, a nivel nacional, reveló un sesgo de muestra significativo. Este fenómeno se manifiesta cuando el conjunto de datos utilizado para capacitar el modelo no refleja adecuadamente la diversidad y complejidad de la totalidad de la población. En nuestro caso, el modelo pudo haber capturado patrones específicos y particularidades presentes en estos dos departamentos, pero al extender su aplicación al conjunto total del país, se evidencia su falta de diversidad para generalizar patrones y características representativas de toda la extensión nacional. Esta limitación en la representación de datos puede resultar en predicciones inexactas o poco confiables al aplicar el modelo a áreas geográficas no incluidas en la fase de entrenamiento, ya que no se han considerado las variaciones y particularidades presentes en otras regiones.

- Los modelos de Deep Learning, particularmente las redes neuronales convolucionales (CNN), sobresalen en la identificación de cultivos en imágenes satelitales al capturar no solo características visuales, sino también las relaciones espaciales y el contexto geográfico. Estos modelos utilizan capas convolucionales para reconocer patrones visuales locales, pero también aprenden a capturar relaciones espaciales al considerar la disposición y la interacción entre los píxeles vecinos. La capacidad de las CNN para aprender jerarquías de características permite la extracción de información a diferentes escalas, lo que resulta fundamental al identificar patrones de cultivos que pueden variar en tamaño y distribución geográfica. Además, al procesar múltiples imágenes de manera simultánea, estos modelos pueden capturar el contexto geográfico más amplio, incorporando datos de terreno, climáticos y otros factores que influyen en la distribución de los cultivos, lo que enriquece su capacidad predictiva y de identificación.
- Por otra parte, el modelo evaluado puede verse beneficiado al incluir nueva información geográfica como pueden ser capas de radar también disponibles en repositorios abiertos. Particularmente, las capas SAR (Synthetic Aperture Radar) distribuidas por el satélite Sentinel 1 pueden ofrecer importantes características morfológicas de los cultivos al ser capaz de brindar información sobre la forma de la planta y la delimitación de la parcela. Igualmente, incluir capas referentes a variables bioclimáticas puede enriquecer la caracterización de cultivos al identificar las necesidades propias de humedad, temperatura o calidad del suelo de cada uno de ellos.
- Finalmente, la aplicación desarrollada demostró ser modular y flexible al poder integrar diversas capas de entrada provenientes de la implementación de distintos modelos de inteligencia artificial. Dicha flexibilidad fue importante a la hora de acoplarse a un nuevo modelo con mejor precisión una vez fueron identificadas las falencias presentes en el modelo entrenado en el área de estudio. Las funcionalidades ofrecidas para los procesos del cálculo de producción y porcentaje de desechos así como el cómputo del potencial energético a partir de dichas estimaciones no se vio afectado por el cambio de modelo.

5. Trabajo futuro

Resolver el sesgo de muestra en un modelo de machine learning requiere un enfoque integral que considere tanto la calidad y la representatividad de los datos como la adaptación adecuada del modelo a la diversidad del territorio nacional. Se propone entonces extender el conjunto de datos para abarcar información de múltiples zonas de Colombia para entrenar modelos de identificación de cultivos que puedan capturar las diversas características geográficas y agrícolas presentes en el país. Colombia exhibe una amplia gama de climas, altitudes, tipos de suelo y prácticas agrícolas, lo que se traduce en una gran variabilidad en la distribución de cultivos a lo largo de las diferentes regiones. Al incluir datos de múltiples zonas geográficas en el conjunto de entrenamiento, se logra una representación más completa y diversa de las características únicas de cada región. Esto permite que los modelos capturen patrones específicos de cultivos, adaptándose mejor a las particularidades de cada área, desde las regiones cafetaleras de alta altitud hasta las zonas tropicales donde se cultivan frutas tropicales.

Otra estrategia que puede resultar eficaz es entrenar modelos regionales para capturar la variabilidad local y luego emplear técnicas de transferencia de conocimiento o ensembles para fusionar la información de los

modelos regionales y crear un modelo más amplio y generalizable a nivel nacional. El entrenamiento de modelos que se ajusten a las particularidades de la geografía colombiana implica, no sólo la adquisición de datos geoespaciales de distintas áreas, sino también la consideración de variables como la altitud, la pluviometría, la temperatura y otros factores ambientales que influyen en el crecimiento de los cultivos. Al integrar esta información diversa en el proceso de entrenamiento, los modelos pueden aprender a distinguir patrones específicos que reflejen las condiciones únicas de cada región, permitiendo una identificación más precisa y contextualizada de los cultivos en un país tan heterogéneo como Colombia.

Por último, considerar para investigaciones futuras la exploración de modelos más avanzados destinados a la segmentación de imágenes satelitales, como aquellos fundamentados en la arquitectura Transformer. Estos modelos han demostrado habilidades sobresalientes en comprender y extraer rasgos complejos en imágenes de alta resolución. Su aplicación en imágenes satelitales puede significativamente mejorar la precisión de la segmentación, permitiendo una identificación más detallada y precisa de áreas de interés. Un ejemplo concreto de este enfoque se encuentra en (Prado et al., 2023), donde se emplea el Modelo Segment Anything (SAM)¹⁶ desarrollado por Meta AI. SAM se distingue por su excepcional capacidad de generalización y aprendizaje sin entrenamiento, lo que lo convierte en una alternativa prometedora para el procesamiento de imágenes aéreas y orbitales en diversos contextos

6. Referencias

- Alves, A., Zerbato, C., de Souza, G., Rodrigues, M., Vieira da Silva, L., Porto de Oliveira, R. (2022). Corn grain yield forecasting by satellite remote sensing and machine learning models. *Agronomy Journal*. 114(5):2956-2968. <https://doi.org/10.1002/agj2.21141>.
- Batista, J., Rodrigues, N., Cabral, A., Vasconcelos, M., Venturieri, A., Silva, L., Silva, S. (2022). Optical time series for the separation of land cover types with similar spectral signatures: cocoa agroforest and forest. *International Journal of Remote Sensing*. 43(9): 3298-3319. DOI: 10.1080/01431161.2022.2089540.
- Benos, L., Tagarakis, A.C., Dolias, G., Berruto, R., Kateris, D., Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors*, 21, 3758. <https://doi.org/10.3390/s211137>
- Campbell, J. (2023). *Introduction to remote sensing*. The Guilford Pres.
- Chen, LC., Zhu, Y., Papandreou, G., Schroff, F., Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) *Computer Vision, ECCV 2018*. Lecture Notes in Computer Science. Vol 11211. Springer, Cham. https://doi.org/10.1007/978-3-030-01234-2_49
- Chuvieco, E. (2016). *Fundamentals of Satellite Remote Sensing. An Environmental Approach, Second Edition*. CRC Press.

¹⁶ <https://segment-anything.com/>

Descals, A., Wich, S., Meijaard, E., Gaveau, D. L. A., Peedell, S., and Szantoi, Z. (2021). High-resolution global map of smallholder and industrial closed-canopy oil palm plantations- *Earth Syst. Sci. Data*, 13:1211–1231. <https://doi.org/10.5194/essd-13-1211-2021>, 2021.

Geron, A. (2022). "Hands-On Machine Learning with Scikit-Learn and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems ". O'Reilly Media.

Hernawati, R., Wikantika, K., Darmawan, S. (2022). Modeling of oil palm phenology based on remote sensing data: opportunities and challenges. *Journal of Applied Remote Sensing*, 16. DOI: 10.1117/1.JRS.16.021501.

Jeon, G. (2023). Advanced Machine Learning and Deep Learning Approaches for Remote Sensing. *Remote Sensing*. 15(11):2876. <https://doi.org/10.3390/rs15112876>.

Kai, P.M., de Oliveira, B.M., de Costa, R.M. (2022). Deep Learning-Based Method for Classification of Sugarcane Varieties. *Agronomy*. 12, 2722. <https://doi.org/10.3390/>.

Li, Y., Zhang, H., Xue, X., Jiang, Y., Shen, Q. (2018). Deep learning for remote sensing image classification: A survey. *WIREs Data Mining and Knowledge Discovery*. 8(6). <https://doi.org/10.1002/widm.1264>

Lillesand, T., Kiefer, R., Chipman, J. (2015). *Remote Sensing and Image Interpretation*. Wiley.

Martello, M., Molin, J. P., Chan Fu Wei, M., Canal Filho, R., & Moreira Nicoletti, J. V. (2024). Coffee-Yield Estimation Using High-Resolution Time-Series Satellite Images and Machine Learning. *Remote Sensing Applications: Society and Environment*. 33. <https://doi.org/10.1016/j.rsase.2023.10109>.

Murguia-Cozar, A., Macedo-Cruz, A., Fernandez-Reynoso, D., Salgado Transito, J., (2021). Recognition of Maize Phenology in Sentinel Images with Machine Learning. *Sensors (Basel)*. 22(1):94. doi: 10.3390/s22010094.

Numbisi, F., Van Coillie, F., De Wulf, R. (2019). Delineation of Cocoa Agroforests Using Multiseason Sentinel-1 SAR Images: A Low Grey Level Range Reduces Uncertainties in GLCM Texture-Based Mapping. *International Journal of Geo-information*. 8(4), 179; <https://doi.org/10.3390/ijgi8040179>.

Prado, L., Wu, Q., Lopes de Lemos, E., Nunes, W., Marques, A., Li, J., Marcato, J. (2023). The Segment Anything Model (SAM) for remote sensing applications: From zero to one shot. *International Journal of Applied Earth Observation and Geoinformation*. 124. <https://doi.org/10.1016/j.jag.2023.103540>.

Rocha, J., Carvalho, M. (2020). The use of machine learning in digital processing of satellite images applied to coffee crop. *CABI Reviews*. 15(045):1-10. DOI: 10.1079/PAVSNNR202015045.

Rodríguez, A., D'Aronco, S., Schindler, K., Wegner, J. (2021). Mapping oil palm density at country scale: An active learning approach. *Remote Sensing of Environment*. *Remote Sensing of Environment*, 261. <https://doi.org/10.1016/j.rse.2021.112479>.

Saadat, M., Seydi, S., Hasanlou, M., Homayouni, S. (2022). A Convolutional Neural Network Method for Rice Mapping Using Time-Series of Sentinel-1 and Sentinel-2 Imagery. *Agriculture*, 12. <https://doi.org/10.3390/agriculture12122083>.

Sierra, R., Jiménez, J., Jiménez, G., Núñez, H., Bressan, M., Giraldo, L., & Narváez, G. (2023). *Informe 4,5,6,7: Exploradores del potencial solar y biomasa residual-Proyecto de Cooperación Triangular Alemania-Chile-Colombia*.

Shendryk, Y., Davy, R., Thorburn, P. (2021). Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. *Field Crops Research*, 260. <https://doi.org/10.1016/j.fcr.2020.107984>.

Thapa, A., Horanont, T., Neupane, B., Aryal, J. (2023). Deep Learning for Remote Sensing Image Scene Classification: A Review and Meta-Analysis. *Remote Sensing*, 15(19), 4804; <https://doi.org/10.3390/rs15194804>.

Teixeira, I., Morais, R., Sousa, J., Cunha, A. (2023). Deep Learning Models for the Classification of Crops in Aerial Imagery: A Review. *Agriculture*. Special Issue "Computer Vision for Intelligent Crop Identification and Crop Protection". 13(5), 965; <https://doi.org/10.3390/agriculture13050965>.

Toan, Q., Dang, K., Giang, K., Al Tong, T., Nguyen, V., Linch, T., Yasir, M. (2022). Deep Learning Model Development for Detecting Coffee Tree Changes Based on Sentinel-2 Imagery in Vietnam. *I EEE Access*, vol. 10, pp. 109097-109107, 2022, doi: 10.1109/ACCESS.2022.3203405.

Watson-Hernández, F., Gómez-Calderón, N., da Silva, R. (2022). Oil Palm Yield Estimation Based on Vegetation and Humidity Indices Generated from Satellite Images and Machine Learning Techniques. *Agriculture Engineering*, 4(1):279-291. <https://doi.org/10.3390/agriengineering4010019>.

Xu, K., Qian, J., Hu, Z., Duan, Z., Chen, C., Liu, J., Sun, J., Wei, S., Xing, X. (2021). A New Machine Learning Approach in Detecting the Oil Palm Plantations Using Remote Sensing Data. *Remote Sensing*, 13(2). <https://doi.org/10.3390/rs13020236>.

Zhang, W., Liu, H., Wu, W., Zhan, L., Wei, J. (2020). Mapping Rice Paddy Based on Machine Learning with Sentinel-2 Multi-Temporal Data: Model Comparison and Transferability. *Remote Sensing*, 12(10), 1620. <https://doi.org/10.3390/rs12101620>.