



Deep Learning en agricultura: conceptos y aplicaciones en la identificación de cultivos sobre imágenes satelitales

Kevin Steven Gamez Abril

Universidad de los Andes
Departamento de Ingeniería de Sistemas y Computación
Ingeniería de Sistemas y Computación
Bogotá, Colombia
2023

Deep Learning en agricultura: conceptos y aplicaciones en la identificación de cultivos sobre imágenes satelitales

Kevin Steven Gamez Abril

Tesis presentada como requisito parcial para optar al título de:
Ingeniero de Sistemas y Computación

Directores:

Ph.D. Haydemar Núñez Castro
Ph.D(C) Andrés Oswaldo Calderón Romero

Grupo de Investigación:
COMIT

Universidad de los Andes
Departamento de Ingeniería de Sistemas y Computación
Ingeniería de Sistemas y Computación
Bogotá, Colombia
2023

Resumen

Esta tesis aborda el desafío de optimizar la agricultura utilizando técnicas de Deep Learning para la clasificación de cultivos en imágenes satelitales de Sentinel 1 y Sentinel 2, enfocándose en la región de Colombia. Frente a la creciente demanda de alimentos y el cambio climático, se desarrolla un modelo avanzado para mejorar la gestión de recursos agrícolas y la productividad. El estudio comienza con un análisis exhaustivo del estado del arte en la identificación de zonas cultivadas mediante Deep Learning, destacando la necesidad de optimizar el uso de recursos como agua y fertilizantes. Se implementa un modelo basado en la arquitectura DeepLab V3+ con ajustes en MobileNet V3 y ResNet50, para clasificar cultivos de palma de aceite en Colombia. Se emplean estrategias como transferencia de aprendizaje y ajuste fino, utilizando modelos pre entrenados para acelerar el entrenamiento. El diseño metodológico se divide en tres fases: modelado, implementación y validación. Se evalúan métricas de desempeño y se analizan limitaciones, especialmente en la detección de áreas recién despejadas y plantaciones jóvenes. Los resultados muestran una alta precisión del modelo en la clasificación de cultivos de palma de aceite, destacando su eficacia y potencial para aplicaciones prácticas en la gestión agrícola. Se concluye la superioridad del Deep Learning sobre métodos tradicionales, sugiriendo futuras mejoras y ampliación del estudio a otros cultivos y condiciones geográficas.

Palabras clave: Deep Learning, imágenes satelitales, clasificación de cultivos, agricultura, Sentinel, gestión de recursos, productividad agrícola, Sensores remotos.

Abstract

This thesis addresses the challenge of optimizing agriculture through the use of Deep Learning techniques for crop classification in Sentinel 1 and Sentinel 2 satellite images, focusing on the region of Colombia. In response to the increasing demand for food and climate change, an advanced model is developed to improve agricultural resource management and productivity. The study begins with a comprehensive analysis of the state of the art in identifying cultivated areas using Deep Learning, highlighting the need to optimize the use of resources such as water and fertilizers. A model based on the DeepLab V3+ architecture with adjustments in MobileNet V3 and ResNet50 is implemented to classify oil palm crops in Colombia. Strategies such as transfer learning and fine-tuning are employed, using pretrained models to accelerate training. The methodological design is divided into three phases: modeling, implementation, and validation. Performance metrics are evaluated, and limitations are analyzed, especially in detecting newly cleared areas and young plantations. The results show the model's high accuracy in classifying oil palm crops, underscoring its effectiveness and potential for practical applications in agricultural management. The study concludes with the superiority of Deep Learning over traditional methods and suggests future improvements and expansion of the study to other crops and geographical conditions.

Keywords: Deep Learning, satellite images, crop classification, agriculture, Sentinel, resource management, agricultural productivity, Remote sensing.

Contenido

	Pág.
1. Contexto y planteamiento del problema.....	3
2. Revisión de la literatura.....	4
2.1 Optimización de la Agricultura con Deep Learning	4
2.2 Casos de Estudio en Agricultura con Deep Learning	6
2.3 Tendencias Futuras y Potencial del Deep Learning en Agricultura.....	11
3. Propuesta de solución.....	13
4. Objetivos	14
4.1.1 Objetivo general	14
4.1.2 Objetivos específicos.....	14
5. Marco teórico	15
5.1 Satélites Sentinel 1 y Sentinel 2	15
5.2 Deep Learning.....	17
5.2.1 Segmentación semántica	18
5.2.2 Redes Neuronales.....	19
5.2.3 Redes Neuronales Convolucionales	20
5.2.1 Convolución Atrous	21
5.2.2 Arquitectura Encoder-Decoder.....	23
5.2.1 Backbone.....	24
5.2.2 Modelos pre entrenados.....	24
5.3 MobileNet	25
5.4 DeepLab	26
5.5 Métricas de desempeño.....	29
6. Desarrollo de la solución.....	31
6.1 Ambiente de desarrollo, datos y restricciones.....	31
6.2 Construcción del conjunto de datos.....	31
6.2.1 Preprocesamiento de Imágenes	31
6.2.2 Etiquetado de Imágenes	33
6.2.3 Aumento de Datos.....	34
6.3 Implementación del modelo	34
6.4 Validación del modelo	35
6.4.1 Métricas	35
6.4.2 Validación de los resultados.....	35
7. Pruebas y resultados	37

8. Conclusiones y trabajo futuro	43
8.1 Conclusiones.....	43
8.2 Trabajo futuro	44

Lista de figuras

	Pág.
Figura 5-1: <i>Inteligencia artificial, machine Learning y Deep Learning</i>	17
Figura 5-2: <i>Ejemplo de segmentación semántica</i>	19
Figura 5-3: <i>Filtro de una convolución 3x3</i>	22
Figura 5-4: <i>Filtro de una convolución atrous (dilatada) 3x3</i>	22
Figura 5-5: <i>Arquitectura DeepLabV3+</i>	27
Figura 7-1: <i>Ejemplo de predicción del modelo sobre un área con cultivos</i>	38
Figura 7-2: <i>Ejemplo de predicción del modelo sobre un área con pocos cultivos</i>	38
Figura 7-3: <i>Ejemplo de predicción del modelo sobre área sin cultivos</i>	39
Figura 7-4: <i>Predicción del modelo sobre toda la zona seleccionada</i>	40
Figura 7-5: <i>Gráficas de pérdida (loss) y exactitud (accuracy)</i>	40
Figura 7-6: <i>Curva ROC del modelo</i>	41

Lista de tablas

Pág.

Tabla 2-1:	<i>Resumen de casos de estudio de la agricultura con Deep Learning.</i>	10
Tabla 5-1:	<i>Bandas de las misiones satelitales Sentinel 1 y Sentinel 2.</i>	16
Tabla 5-2:	<i>Matriz de confusión.....</i>	29

Introducción

La agricultura moderna se enfrenta a varios retos, entre ellos está la creciente demanda de alimentos como consecuencia del aumento de la población mundial, sumado a esto, el cambio climático junto con el agotamiento de recursos naturales (Thayer et al., 2020). Esto ha creado la necesidad de optimizar el uso de la tierra para la producción de cultivos con el fin de disminuir los impactos al medio ambiente y aumentar el impacto socioeconómico. Se ha demostrado que los datos satelitales permiten realizar mapas agrícolas y globales, los cuales permiten una gestión más eficiente de recursos críticos como el agua, fertilizantes y pesticidas. Además, permite el monitoreo de la salud de las plantas y un aumento de la productividad al optimizar la planificación de la siembra y la cosecha (Boryan et al., 2011).

Históricamente, la identificación de cultivos y el monitoreo de su evolución se basaron en observaciones de campo, un proceso laborioso y propenso a errores humanos. Para reducir este esfuerzo, en tiempos recientes se han utilizado imágenes aéreas de campos de cultivo, que se adquieren y procesan desde vehículos aéreos no tripulados. Esta capacidad ha abierto las puertas a la agricultura, permitiendo a los agricultores y a los responsables de la toma de decisiones acceder a datos valiosos sobre el estado de los cultivos, la utilización de la tierra y la gestión de recursos (Hu et al., 2021). El uso de imágenes satelitales para la clasificación de diferentes tipos de cultivos se ha abordado en estudios anteriores debido a la gran cantidad de información que estas imágenes contienen.

Esta investigación está centrada en desarrollar un modelo de clasificación avanzado utilizando técnicas de Deep Learning para identificar y clasificar cultivos a partir de una composición de bandas de imágenes satelitales. Esta tarea es vital en el contexto agrícola, especialmente para optimizar el uso de recursos y mejorar las prácticas de gestión de cultivos. Inicialmente, se llevará a cabo un exhaustivo análisis del estado del arte, evaluando cómo se han aplicado previamente técnicas de Deep Learning en la

identificación de zonas cultivadas. Esto nos permitirá entender mejor los enfoques actuales y las posibles áreas de mejora.

El enfoque de este trabajo también incluye identificar las características más relevantes en las imágenes satelitales que permiten clasificar eficientemente los diferentes tipos de cultivos. Esta fase es crucial para asegurar la precisión y relevancia del modelo propuesto. Además, investigaremos y aplicaremos estrategias de transferencia de aprendizaje y ajuste fino (Fine-tuning) para aprovechar los modelos de Deep Learning pre entrenados. Esta aproximación tiene como objetivo acelerar y mejorar el proceso de entrenamiento de nuestro modelo.

Luego, aplicaremos el modelo desarrollado en una zona geográfica específica de Colombia. Se compara su rendimiento con un modelo de clasificación basados en índices de vegetación convencionales, buscando evidenciar las ventajas y mejoras que el enfoque del Deep Learning puede aportar en el campo de la clasificación de cultivos mediante imágenes satelitales. Para desarrollar el modelo de clasificación se planteó una metodología dividida en tres fases. La primera es el modelado donde se destaca la selección de la arquitectura del modelo y el preprocesamiento de las imágenes satelitales de Sentinel 1 y Sentinel 2 junto con la composición de estas mismas realizada por Descals et al. (2021). Luego en la etapa de implementación se centra en entrenamiento del modelo, así como la selección y prueba de modelo pre entrenados. Por último, la validación del modelo se realiza específicamente en imágenes de zonas seleccionadas, representativas del territorio colombiano. En esta última etapa se evalúan las métricas de desempeño del modelo obtenido junto con el análisis de las limitaciones de este.

1.Contexto y planteamiento del problema

En Colombia, la agricultura fue un sector determinante en el desarrollo del país durante la mayor parte del siglo pasado, destacándose por el crecimiento de la producción de café en diferentes regiones (Bejarano, 1980). Sin embargo, desde la década de 2000, la agricultura ha disminuido su participación en la economía colombiana. Según datos del Banco Mundial, el DANE y el Departamento Nacional de Planeación (DNP), la contribución de la agricultura al PIB de Colombia pasó del 25% en 1965 al 22,30% en 1990 y llegó solo al 6,30% en 2017 (Junguito et al., 2014). Esta pérdida en la contribución de la agricultura y la ganadería al saldo comercial colombiano se debe, por un lado, al mayor crecimiento relativo de otros sectores y, por otro lado, a la baja productividad del propio sector agrícola. Esto también genera menos desarrollo, especialmente en áreas donde la agricultura ha sido considerada la principal vocación económica. En respuesta a la situación que prevalece en diversas regiones del país, se han implementado programas de desarrollo destinados a reducir la pobreza y mejorar las condiciones de vida de los agricultores (DANE, 2014).

En el contexto de la disminución de la contribución de la agricultura al PIB de Colombia y su baja productividad, el uso de la clasificación de imágenes satelitales con Deep Learning emerge como una solución crucial. La capacidad de procesar grandes volúmenes de datos, mejorar la precisión en la identificación de cultivos plantados, detectar patrones complejos y automatizar tareas esencialmente cambia la dinámica de la gestión agrícola (Hu et al., 2021). Esto no solo revitalizaría el sector agrícola colombiano, sino que también promovería prácticas agrícolas sostenibles y contribuiría a la seguridad alimentaria. El Deep Learning se presenta como una herramienta precisa y eficaz para abordar los desafíos actuales y mejorar la eficiencia en la identificación de cultivos y así mejorar el desarrollo de productivo en la agricultura de Colombia.Revisión de la literatura

2.Revisión de la literatura

El desarrollo de las redes neuronales, en particular la aplicación de técnicas de Deep Learning para el análisis de imágenes satelitales, ha revolucionado la agricultura en los últimos años. Estas tecnologías se han convertido en herramientas fundamentales para la gestión y planificación agrícola, permitiendo una comprensión y predicción más precisas del rendimiento de los cultivos y la biomasa. Esta transformación marca una nueva era en la monitorización y gestión del rendimiento de los cultivos, ofreciendo un gran potencial para la toma de decisiones basadas en datos y la optimización de recursos.

Un análisis exhaustivo de 150 estudios sobre el uso de Deep Learning en imágenes satelitales para la agricultura presentado por Victor et al. (2022), en particular identifican cinco tareas comunes: uso y cobertura del suelo, salud del suelo, fisiología de las plantas, daño a los cultivos y predicción del rendimiento. Se descubrió que los métodos modernos de Deep Learning superan consistentemente a los métodos tradicionales de aprendizaje automático en la mayoría de las tareas, con la excepción de la predicción del rendimiento con redes neuronales recurrentes de memoria a corto y largo plazo (LSTM), que no superan consistentemente a los modelos basados en Random Forest. En el estudio se resalta la necesidad de conjuntos de datos de referencia públicos para comparar adecuadamente los resultados entre estudios.

2.1 Optimización de la Agricultura con Deep Learning

Un ejemplo innovador de esta aplicación es la estimación del rendimiento del café en Brasil. Aquí, el Deep Learning, combinado con modelos de regresión y análisis de imágenes satelitales, ha logrado reproducir con éxito la variabilidad espacial de los mapas de rendimiento con un año de anticipación, lo que es crucial para la planificación y gestión de la producción de café (Martello et al., 2022). Finalmente, la escalabilidad y flexibilidad del Deep Learning han sido destacadas en un estudio realizado en Brasil sobre la

estimación de rendimientos de cultivos. Utilizando datos geográficos, meteorológicos y calendarios de cultivo para cinco cultivos principales, este enfoque subraya la capacidad del Deep Learning para adaptarse a grandes regiones con diversas condiciones agrícolas (Cunha & Silva, 2020).

En cuanto a las imágenes satelitales, se enfrenta al desafío de tener una resolución espacial a menudo más baja que la escala de los objetos observados. No obstante, la resolución espectral de estas imágenes es generalmente superior, lo que proporciona una abundancia de información por píxel. Esto ha llevado al uso de algoritmos per-píxel, en contraposición a los enfoques tradicionales de visión por computadora que dependen de patrones espaciales de múltiples píxeles. El Deep Learning tiene ventajas distintivas sobre el aprendizaje automático tradicional. Puede construir modelos más complejos y robustos, especialmente con datos ruidosos, y aprende sus propias características, eliminando la necesidad de extracción manual. Sin embargo, esto requiere conjuntos de datos más grandes para un rendimiento óptimo (Victor et al., 2022).

Los enfoques de interpretación de datos en el Deep Learning son variados. Algunos estudios tratan la información espectral como secuencias para aplicar algoritmos de análisis de secuencias, mientras que otros consideran los datos temporales como "imágenes" espectrales/temporales, permitiendo el uso de redes neuronales convolucionales (CNNs) en dimensiones más altas. Un desafío significativo en el uso de imágenes satelitales es la cobertura de nubes y otros efectos atmosféricos. Para manejar esto, los estudios excluyen a menudo píxeles o imágenes afectadas por nubes, o utilizan compuestos de imágenes tomadas en diferentes momentos. El uso de imágenes SAR de satélites como Sentinel-1, que pueden penetrar la cobertura de nubes, ha mejorado los resultados en comparación con el uso exclusivo de imágenes ópticas.

Finalmente, para estudios del uso y ocupación del suelo, los datos más comúnmente utilizados provienen de satélites como Sentinel-1/Sentinel-2/Landsat-8, que ofrecen imágenes SAR y multiespectrales ópticas. La combinación de estas fuentes de datos ha demostrado mejorar el rendimiento en comparación con el uso de una sola fuente.

2.2 Casos de Estudio en Agricultura con Deep Learning

- **Coffee-Yield Estimation Using High-Resolution Time-Series Satellite Images and Machine Learning:** En la investigación de Martello et al. (2022) se centra en la estimación del rendimiento del café utilizando imágenes satelitales de alta resolución y aprendizaje automático. El estudio fue llevado a cabo en una zona comercial en Minas Gerais, Brasil, y utilizó un monitor de rendimiento en una cosechadora para medir el volumen de café cosechado. Se emplearon imágenes de la plataforma PlanetScope y se ajustaron modelos de Random Forest y regresión lineal múltiple a diferentes conjuntos de datos compuestos por rendimiento de café y series temporales de datos de imágenes satelitales. Los resultados mostraron que el modelo implementado con Random Forest utilizando bandas espectrales proporcionó los mejores resultados, alcanzando un R^2 de 0.93 para el conjunto de validación y los errores más bajos de predicción. Estos hallazgos pueden ayudar a monitorear la variabilidad espacial y temporal de los campos y guiar las prácticas de gestión basadas en los principios de la agricultura de precisión. Este caso de estudio demuestra el potencial del Deep Learning en la agricultura, especialmente en la estimación precisa del rendimiento del café, un cultivo de gran importancia económica en Brasil. La combinación de datos de teledetección con técnicas avanzadas de aprendizaje automático puede proporcionar diversos puntos de vista valiosos para la toma de decisiones en la gestión agrícola.
- **Corn Biomass Estimation by Integrating Remote Sensing and Long-Term Observation Data Based on Machine Learning Techniques:** Este artículo presentado por Geng et al. (2021) se centra en la estimación de la biomasa del maíz utilizando datos de teledetección y observación a largo plazo basados en técnicas de aprendizaje automático. El estudio se realizó en la cuenca del río Heihe en China, utilizando datos de reflectancia diaria del MODIS entre 2012 y 2019. Se adoptaron cuatro modelos de aprendizaje automático: Random Forest (RF), Support Vector Machine (SVM), Redes Neuronales Artificiales (ANN) y Extreme Gradient Boosting (XGBoost) para estimar la biomasa estacional del maíz. Se seleccionaron nueve variables, incluyendo índices de vegetación y bandas espectrales, para influir en la biomasa del maíz. Los resultados mostraron que el modelo XGBoost tuvo un rendimiento superior ($R^2 = 0.78$, RMSE = 2.86 t/ha, MAE = 1.86 t/ha), seguido de cerca

por el modelo RF ($R^2 = 0.77$, RMSE = 2.91 t/ha, MAE = 1.91 t/ha). Este estudio proporciona una referencia importante para la estimación de la biomasa de los cultivos y demuestra el potencial de las técnicas de aprendizaje automático para lograr una estimación relativamente precisa de la biomasa diaria del maíz a gran escala.

- **Estimating crop yields with remote sensing and Deep Learning:** El caso de estudio presentado por Cunha & Silva. (2020) se enfoca en la estimación de rendimientos de cultivos usando aprendizaje profundo con datos de fácil acceso. El estudio se llevó a cabo en Brasil, centrado en cinco cultivos principales: soja, maíz, arroz, caña de azúcar y algodón. El modelo propuesto utiliza coordenadas geográficas, datos meteorológicos y de suelo, y calendarios de cultivo para realizar estimaciones de rendimiento. Se utilizó un enfoque innovador que no requiere el procesamiento de datos de NDVI, lo que hace que la solución sea fácilmente escalable y rápida. Los resultados mostraron que el modelo podía realizar predicciones precisas para grandes regiones, utilizando una cantidad limitada de datos de entrada. Este enfoque es especialmente relevante para la toma de decisiones en la gestión agrícola y gubernamental, como la planificación de alquiler de maquinaria, la negociación de precios y la logística.
- **Forecasting crop yield with deep learning-based ensemble mode:** Este artículo presentado por Divakar et al. (2022) se centra en la predicción del rendimiento de cultivos como la soja y el arroz utilizando un modelo de ensamble basado en Deep Learning. El estudio se llevó a cabo en los Estados Unidos e India, utilizando datos de rendimiento de cultivos a nivel de condado y distrito, respectivamente. Se implementó un modelo de ensamble que combina la Memoria a Largo y Corto Plazo (LSTM) y la LSTM convolucional (ConvLSTM) para procesar datos de reflectancia espectral superficial, temperatura de la superficie terrestre y humedad del suelo. Estos datos fueron obtenidos de los instrumentos MODIS del satélite Terra y datos de humedad del suelo de la NASA-USDA. El modelo demostró un rendimiento comparable a otros modelos existentes, logrando un valor R^2 de 0.73 para la soja y 0.49 para el arroz, lo cual es significativo considerando que se utilizaron menos parámetros en el modelo propuesto.
- **Forecasting sunflower grain yield using remote sensing data:** El estudio presentado por Debaeke et al. (2023) se centra en la predicción del rendimiento de

granos de girasol utilizando datos de teledetección y modelos estadísticos. Este estudio se realizó en el suroeste de Francia, cubriendo los departamentos de Haute-Garonne y Gers, y se llevó a cabo entre 2014 y 2016 en 359 campos de girasol. Se emplearon diferentes procedimientos de modelado estadístico, incluyendo regresión lineal, regresión polinómica de segundo grado, Random Forest y proceso gaussiano, para predecir el rendimiento de granos de girasol a nivel de campo. Las variables clave utilizadas en los modelos fueron el Índice de Área Verde (GAI) máximo y la Duración del Área Verde (GAD), calculados a partir de datos de teledetección satelital. Los resultados indicaron que los modelos basados en la combinación de GAD y GAI_{max} o solo GAD proporcionaron mejores predicciones. Este caso de estudio demuestra el valor del Deep Learning y la teledetección para mejorar la precisión en la predicción del rendimiento de cultivos a gran escala, lo cual es vital para una gestión eficiente de los recursos en la agricultura.

- **A systematic review of the use of Deep Learning in Satellite Imagery for Agriculture:** El documento presentado por Victor et al. (2022) ofrece una revisión sistemática de 150 estudios que utilizan el aprendizaje profundo en imágenes satelitales para aplicaciones agrícolas. Como ya se mencionó identifica cinco tareas comunes en la investigación agrícola utilizando imágenes satelitales: clasificación de uso del suelo y cobertura vegetal, monitoreo de la salud del suelo, fisiología de las plantas, detección de daños en cultivos y predicción de rendimientos. La revisión destaca que los métodos modernos de aprendizaje profundo superan consistentemente a otros métodos para la mayoría de estas tareas.

Uno de los hallazgos clave es que, mientras que los métodos de aprendizaje profundo superan a los métodos tradicionales en casi todas las tareas, no siempre superan a los métodos de Random Forest para la predicción de rendimientos. Se subraya la necesidad de benchmarks públicos para comparar adecuadamente los resultados entre estudios. Además, se observa que algunos estudios aprovechan la resolución espectral extra disponible en las imágenes satelitales, pero otras propiedades divergentes de las imágenes satelitales, como las escalas espaciales significativamente diferentes, no se están aprovechando en los estudios revisados.

- **High-resolution global map of smallholder and industrial closed-canopy oil palm plantations:** El artículo presentado por Descals et al. (2021) se enfoca en la creación de un mapa global de alta resolución de plantaciones de palma aceitera de dos tipos: pequeños propietarios y plantaciones industriales. El estudio utiliza el modelo DeepLabv3+, una red neuronal convolucional (CNN) para la segmentación semántica, y fue entrenado para clasificar imágenes de los satélites Sentinel-1 y Sentinel-2 en un mapa de cobertura de tierra de palma aceitera. El estudio destaca por su precisión en la diferenciación entre plantaciones industriales y de pequeños propietarios a nivel global, con una exactitud general del $98.52 \pm 0.20\%$. La capacidad del modelo para aprender patrones espaciales como las redes de carreteras de cosecha permitió esta distinción precisa, superando en precisión a los conjuntos de datos regionales de palma aceitera existentes que utilizaban algoritmos de aprendizaje automático convencionales. Además, el estudio revela que las plantaciones de palma aceitera con dosel cerrado se encuentran en 49 países, cubriendo un área mapeada de 19.60 Mha. Para esta investigación realizaron una composición de imágenes satelitales de Sentinel 1 y Sentinel 2. Sentinel-1: Se utilizó por su capacidad de proporcionar datos de radar de apertura sintética (SAR), lo que permite obtener imágenes independientemente de las condiciones climáticas y de la luz. Este satélite ofrece una respuesta de retrodispersión característica de las palmas, con valores bajos de transmitancia vertical y recepción vertical (VV) y altos valores de transmitancia vertical y recepción horizontal (VH). Sentinel-2: Proporciona imágenes multiespectrales que son útiles para capturar la reflectancia en diferentes bandas. En este estudio, se destacó la relevancia de la banda 4 (banda roja; longitud de onda central = 665 nm) para predecir plantaciones industriales y de palma aceitera. La banda 4 se seleccionó por su capacidad para mostrar los caminos en las plantaciones industriales debido al alto contraste en términos de reflectancia entre el camino y las palmas circundantes.

La razón por la que se seleccionaron estas bandas de Sentinel 1 es por su respuesta de retrodispersión. La estructura de la copa de las palmas produce una respuesta de retrodispersión distintiva, que es crucial para detectar plantaciones de palma aceitera, especialmente aquellas con dosel cerrado. Para la elección de la banda 4 de Sentinel 2 se basó en su efectividad para resaltar características estructurales dentro de las plantaciones, como los sistemas de caminos, que son fundamentales para diferenciar entre tipos de plantaciones.

El estudio demuestra cómo la combinación de datos SAR y ópticos, junto con técnicas avanzadas de Deep Learning, puede proporcionar un mapeo detallado y preciso de cultivos específicos a gran escala, superando los métodos de aprendizaje automático convencionales. Este enfoque no solo mejora la precisión en la clasificación de tipos de plantaciones de palma aceitera, sino que también proporciona información valiosa para debates sobre el impacto ambiental y la planificación agrícola sostenible.

La **Tabla 2-1** resume los artículos presentados junto con la información de a que cultivo pertenece, su año de publicación y los resultados más destacables.

Tabla 2-1: *Resumen de casos de estudio de la agricultura con Deep Learning.*

Titulo investigación	Ubicación	Cultivo(s)	Metodologia	Resultados Clave	Año de publicación
Coffee-Yield Estimation Using High-Resolution Time-Series Satellite Images and Machine Learning	Minas Gerais, Brasil	Café	Imágenes de PlanetScope, modelos RF y MLR	RF con bandas espectrales alcanzó un R^2 de 0.93	2022
Corn Biomass Estimation by Integrating Remote Sensing and Long-Term Observation Data Based on Machine Learning Techniques	Cuenca del río Heihe, China	Maíz	Datos de MODIS, modelos RF y XGBoost	XGBoost tuvo un rendimiento superior, $R^2 = 0.78$	2021
Estimating crop yields with remote sensing and Deep Learning	Brasil	Varios	Coordenadas geográficas, datos meteorológicos y de suelo	Predicciones precisas para grandes regiones con datos limitados	2020
Forecasting crop yield with deep learning based ensemble model	Estados Unidos e India	Soja, Arroz	Modelo de ensamble (LSTM y ConvLSTM), datos de MODIS y NASA-USDA	Rendimiento comparable con modelos existentes, $R^2 = 0.73$ para soja	2022
Forecasting sunflower grain yield using remote sensing data and statistical models	Suroeste de Francia	Girasol	Modelos estadísticos, datos de teledetección, GAI y GAD	Mejores predicciones con la combinación de GAD y GAlmax o solo GAD	2023
A systematic review of the use of Deep Learning in Satellite Imagery for Agriculture	Revisión global	Varios	Revisión de 150 estudios sobre Deep Learning y teledetección	Identifica tareas comunes en la investigación agrícola y evalúa la eficacia	2022
High-resolution global map of smallholder and industrial closed-canopy oil palm plantations	Global	Palma Aceitera	DeepLabv3+ CNN para clasificación de imágenes Sentinel-1 y Sentinel-2	Distingue entre plantaciones industriales y de pequeños propietarios con un $98.52 \pm 0.20\%$ de exactitud	2021

2.3 Tendencias Futuras y Potencial del Deep Learning en Agricultura

El campo de la agricultura ha experimentado una transformación notable con la incorporación del Deep Learning, especialmente en la estimación de rendimientos y biomasa de cultivos como café, maíz, caña de azúcar, girasol y la palma de aceite. La integración de datos de teledetección con algoritmos avanzados de aprendizaje profundo, como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN), ha mejorado significativamente la precisión de estas estimaciones. Estos avances han permitido a los agricultores y científicos obtener predicciones más precisas y oportunas, lo que es crucial para la planificación y gestión eficaz de los recursos agrícolas. El uso de imágenes satelitales para el monitoreo agrícola ha ganado una importancia considerable. La habilidad del Deep Learning para procesar y analizar estas imágenes permite aprovechar los índices espectrales para mejorar la eficiencia en la agricultura de precisión. Esta tecnología facilita el monitoreo de la salud del suelo y la fisiología de las plantas, así como la evaluación de los daños a los cultivos, ofreciendo una visión integral y actualizada de las condiciones agrícolas.

La necesidad de establecer benchmarks públicos y métodos estandarizados de evaluación ha sido ampliamente reconocida en la comunidad científica. Estos benchmarks permitirán a los investigadores comparar efectivamente los resultados de diferentes estudios y validar sus modelos con mayor eficacia. La creación de estos conjuntos de datos de referencia es un paso crucial para el avance y la adopción generalizada de soluciones de Deep Learning en la agricultura. El Deep Learning también está abordando desafíos clave relacionados con la calidad y la utilidad de las imágenes satelitales, como la resolución espacial y las interferencias atmosféricas. Se están desarrollando métodos para manejar estas limitaciones de manera eficiente, lo que permite maximizar el uso de los datos disponibles y mejorar las aplicaciones agrícolas basadas en teledetección. Se espera que las innovaciones continuas en el campo del Deep Learning impulsen aplicaciones prácticas en la agricultura, como la detección temprana de enfermedades de cultivos, la optimización del uso del agua y la gestión eficiente de nutrientes. Estas aplicaciones no solo mejorarán la eficiencia y productividad agrícola, sino que también contribuirán a la sostenibilidad y al menor impacto medioambiental de la agricultura. La colaboración entre expertos en

diferentes campos, como agrónomos, científicos de datos y especialistas en teledetección, está emergiendo como una tendencia clave. Estas colaboraciones interdisciplinarias son fundamentales para el desarrollo de soluciones de Deep Learning más integradas y efectivas en la agricultura, facilitando el intercambio de conocimientos y la innovación en este campo en constante evolución.

3.Propuesta de solución

Se propone entrenar un modelo de Deep Learning que se encargue de clasificar cultivos de palma de aceite sobre de Sentinel 1 y Sentinel 2 construidas a partir de la banda 4 de Sentinel 2 basándonos en el trabajo realizado por Descals et al. (2021). En donde se usa DeepLab V3+, una red convolucional para la segmentación semántica junto con un backbone basado en el modelo pre entrenado MobileNet V2. Realizando algunas mejoras como la prueba de redes más actuales como MobileNet V3 y ResNet50. La metodología propuesta se divide en tres fases clave: Modelado, Implementación y Evaluación del modelo de clasificación basado en técnicas de Deep Learning. En la etapa de Modelado, se destaca la selección de la arquitectura de red neuronal DeepLab V3+ y el preprocesamiento de datos de las misiones Sentinel-1 y Sentinel-2. La Implementación se centra en el entrenamiento del modelo, destacando la elección de MobileNet V3 como arquitectura óptima. La validación se realiza específicamente en imágenes de zonas seleccionadas, representativas del territorio colombiano. La Evaluación de Desempeño incluye el cálculo de métricas clave, mientras que el Análisis de Limitaciones aborda aspectos críticos como la detección de áreas recién despejadas y plantaciones jóvenes. Este enfoque metodológico integral garantiza la robustez y eficacia del modelo en la identificación de cultivos, especialmente en el contexto de las plantaciones de palma de aceite en Colombia.

4. Objetivos

4.1.1 Objetivo general

Desarrollar un modelo de clasificación para la identificación de cultivos sobre imágenes satelitales utilizando técnicas de Deep Learning.

4.1.2 Objetivos específicos

- Realizar un análisis del estado del arte con relación al uso de técnicas de Deep Learning para identificar zonas cultivadas.
- Identificar las características relevantes en la clasificación de cultivos a partir de imágenes satelitales.
- Investigar enfoques de transferencia de aprendizaje y ajuste (Fine-tuning) para aprovechar modelos pre entrenados en conjuntos de datos relacionados y acelerar el proceso de entrenamiento del modelo.
- Aplicar el modelo desarrollado en una zona geográfica específica de Colombia y comparar sus resultados con un modelo de clasificación basado en índices de vegetación convencionales.

5.Marco teórico

El aprendizaje automático es una rama de la inteligencia artificial y la informática que se centra en usar datos y algoritmos para imitar la forma en que los humanos aprenden a hacer cierta tarea, mejorando progresivamente su precisión. El Deep Learning es un subconjunto de la inteligencia artificial basado en redes neuronales artificiales con aprendizaje de representación (Vargas et al., 2017).

5.1 Satélites Sentinel 1 y Sentinel 2

Las misiones Sentinel son un conjunto de programas de observación de la Tierra desarrollados por la Agencia Espacial Europea (ESA) bajo el Programa Copérnico. Este proyecto multi-satélite está diseñado para proporcionar una comprensión detallada y actualizada del planeta, cubriendo aspectos terrestres, oceánicos y atmosféricos. Cada misión Sentinel tiene un enfoque y objetivos específicos. La misión Sentinel-1, parte de la iniciativa Copernicus desarrollada por la ESA, consiste en una constelación de dos satélites en órbita polar, capaces de operar día y noche mediante imágenes de radar de apertura sintética (SAR) en banda C. Esta tecnología les permite adquirir imágenes independientemente de las condiciones climáticas, lo cual es crucial para la monitorización continua y precisa de la superficie terrestre. La misión Sentinel-1 tiene como propósito principal la creación de un archivo de datos a largo plazo, coherente y confiable, indispensable para aplicaciones que se basan en series temporales extensas. Al ser la primera de las cinco misiones desarrolladas para Copernicus, Sentinel-1 sienta las bases para un enfoque sistemático y avanzado en la observación de la Tierra (*Earth Engine Data Catalog: Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected, log scaling*, 2014).

Por otro lado, la misión Sentinel-2 complementa y amplía las capacidades de observación terrestre. También compuesta por dos satélites en órbita polar sincronizada con el sol,

Sentinel-2 se enfoca en el monitoreo de la variabilidad de las condiciones de la superficie terrestre. Con una amplitud de barrido de 290 km y un tiempo de revisita altamente eficiente 10 días en el ecuador con un satélite y 5 días con dos, bajo condiciones sin nubes, Sentinel-2 es especialmente apto para el seguimiento de cambios en la superficie de la Tierra. Su instrumento principal, el MultiSpectral Instrument (MSI), es fundamental para aplicaciones como el monitoreo de la tierra y el estudio del cambio climático (European Space Agency, 2017).

A continuación, se presenta una tabla con las bandas de las misiones Sentinel 1 radar de apertura sintética (SAR) Alcance terrestre detectado (GRD) y Sentinel 2 Level-2A, junto con su resolución y su nombre:

Tabla 5-1: *Bandas de las misiones satelitales Sentinel 1 y Sentinel 2.*

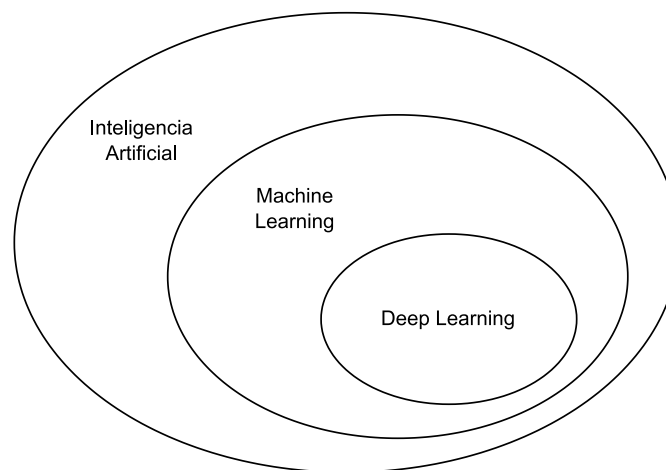
Satélite	Banda	Descripción	Tamaño de píxel
Sentinel-1	HH	Horizontal Transmit - Horizontal Receive Polarisation	10 m
	HV	Vertical Transmit-Horizontal Receive Polarisation	10 m
	VV	Vertical Transmit-Vertical Receive Polarisation	10 m
	VH	Horizontal Transmit - Vertical Receive Polarisation	10 m
	B1	Coastal aerosol	60 m
Sentinel-2	B2	Color azul	10 m
	B3	Color verde	10 m
	B4	Color rojo	10 m
	B5	Vegetation red edge	20 m
	B6	Vegetation red edge	20 m
	B7	Vegetation red edge	20 m
	B8	NIR	10 m
	B8a	Narrow NIR	20 m
	B9	Water vapour	60 m
	B10	SWIR - cirrus	60 m
	B11	SWIR	20 m
	B12	SWIR	20 m

Ambas misiones, Sentinel-1 y Sentinel-2, son fundamentales para la observación de la Tierra, cada una con un enfoque específico y complementario: Sentinel-1 con su capacidad de radar para la obtención de imágenes en cualquier condición climática, y Sentinel-2 con su enfoque en el monitoreo de alta resolución de la superficie terrestre.

5.2 Deep Learning

La inteligencia artificial (IA) es un área de la informática que surgió en los años 50. Con el paso del tiempo, la IA evolucionó para abordar problemas más complejos como el reconocimiento de voz o la traducción de lenguaje natural, lo que llevó a un enfoque en el aprendizaje automático. La siguiente es una representación del panorama de la inteligencia artificial, donde se centran cada uno de sus subcampos:

Figura 5-1: *Inteligencia artificial, machine Learning y Deep Learning*



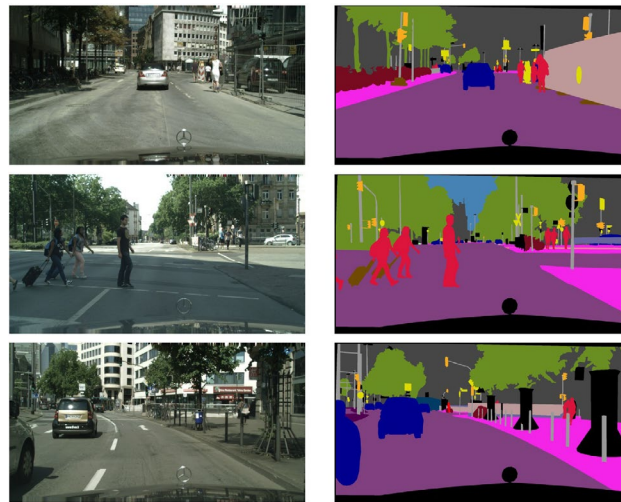
El aprendizaje profundo es un subcampo del aprendizaje automático que se enfoca en construir modelos a través de múltiples capas de procesamiento para aprender representaciones complejas de los datos. A diferencia del aprendizaje automático tradicional, que suele trabajar con una o dos capas de representaciones, el aprendizaje profundo utiliza muchas capas, lo que le permite captar patrones y características más abstractas y profundos de los datos (Vargas et al., 2017). Estos modelos usan como base las redes neuronales, cuyo diseño se inspira superficialmente en la biología neuronal, pero son fundamentalmente modelos matemáticos para procesar y transformar datos. A través

de múltiples capas, el aprendizaje profundo efectúa una especie de filtración progresiva de la información, permitiendo que el modelo realice tareas complejas como el reconocimiento de imágenes y el procesamiento del lenguaje natural de manera muy efectiva (Vargas et al., 2017).

5.2.1 Segmentación semántica

La segmentación semántica hace parte del campo de la visión por computadora y el procesamiento de imágenes. Su objetivo es dividir una imagen en partes o segmentos, donde cada segmento corresponde a un objeto o clase de objetos diferentes. A diferencia de la segmentación tradicional, que podría dividir una imagen basándose únicamente en diferencias de color o textura, la segmentación semántica busca entender y etiquetar cada parte de la imagen con una categoría significativa en términos de contenido (Guo et al., 2018). Por ejemplo, como se puede observar en la **Figura 5-2**, una imagen de una calle, la segmentación semántica no solo distinguiría entre el cielo, los edificios y el suelo, sino que también identificaría y etiquetarían estos elementos como cielo, edificio y suelo respectivamente. La segmentación semántica se logra generalmente a través de técnicas de aprendizaje profundo, donde las redes neuronales convolucionales (CNN) son entrenadas con grandes cantidades de imágenes anotadas para aprender a clasificar y segmentar diferentes partes de las imágenes. Estas redes pueden identificar patrones y características visuales complejas, permitiendo una segmentación precisa y detallada (Guo et al., 2018).

Figura 5-2: *Ejemplo de segmentación semántica*



Nota. Adaptado de (L.-C. Chen et al., 2016)

5.2.2 Redes Neuronales

Son modelos matemáticos que se utilizan en el campo del aprendizaje automático. Las redes neuronales se inspiran en la estructura del cerebro humano dado que nuestro cerebro, las neuronas individuales reciben señales, las procesan y transmiten señales a otras neuronas. De manera similar, en las redes neuronales artificiales, las "neuronas" (unidades de procesamiento) reciben datos de entrada, los procesan y pasan el resultado a otras unidades (Courses Teach, 2023). Una red neuronal puede tener muchas capas, donde hay tres tipos de capas:

- **Capa de Entrada:** Recibe los datos iniciales.
- **Capas Ocultas:** Realizan cálculos complejos a través de sus neuronas. Puede haber múltiples capas ocultas que extraen características progresivamente más abstractas de los datos.
- **Capa de Salida:** Produce el resultado final, como una clasificación o una predicción.

Cada neurona en una red realiza un cálculo simple. Toma varios valores de entrada, los multiplica por pesos, que son parámetros ajustables, suma estos productos y luego aplica una función de activación. La función de activación introduce no linealidad, permitiendo a la red aprender patrones complejos. El aprendizaje en una red neuronal se realiza

ajustando los pesos. Esto se hace generalmente a través de un proceso llamado backpropagation y algoritmos de optimización como el descenso de gradiente (Courses Teach, 2023). Durante el entrenamiento, la red intenta minimizar la diferencia entre su salida y la salida deseada, el error, ajustando repetidamente los pesos en función del error. Las redes neuronales son excelentes para tareas que son fáciles para los humanos, pero difíciles de describir formalmente. Algunos casos de uso son el reconocimiento de imágenes, la traducción de idiomas, el juego de videojuegos y la conducción autónoma.

5.2.3 Redes Neuronales Convolucionales

Una red neuronal convolucional (CNN), por sus siglas en inglés "Convolutional Neural Network" es un tipo especial de red neuronal diseñada para procesar datos que tienen una estructura de rejilla, como las imágenes. Estas redes son especialmente eficaces en tareas de visión por computadora, como la clasificación y detección de objetos en imágenes, debido a su habilidad para capturar patrones espaciales y temporales en los datos (Courses Teach, 2023). Las CNN se diferencian de las redes neuronales tradicionales en varios aspectos clave como lo son:

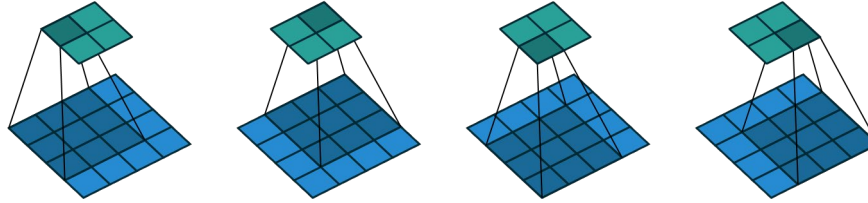
- **Convoluciones:** El componente fundamental de una CNN es la capa convolucional, que aplica una operación matemática llamada convolución. Esta operación consiste en deslizar un filtro o kernel sobre la entrada y realizar el producto punto entre los valores del filtro y la entrada local. Esto permite a la red detectar características como bordes, texturas y patrones en diferentes partes de la entrada.
- **Localidad y compartición de pesos:** En una CNN, cada filtro se enfoca en una pequeña región espacial de la entrada, capturando características locales. Además, el mismo filtro se aplica a toda la entrada, lo que reduce significativamente la cantidad de parámetros a aprender en comparación con una red neuronal densa.
- **Pooling:** Las capas de pooling se utilizan para reducir la dimensión espacial de las representaciones aprendidas, aumentando así la eficiencia computacional y la robustez de la red. Una operación común en estas capas es el max pooling, que toma el valor máximo de una región de la entrada.

- **Capas Densas:** Al final de una CNN, después de varias capas convolucionales y de pooling, se suelen incluir una o más capas densas para realizar la clasificación o la regresión, dependiendo del problema.
- **Captura de Patrones Jerárquicos:** Las CNN son capaces de aprender patrones jerárquicos en los datos. Las primeras capas convolucionales suelen detectar características simples como bordes, mientras que las capas más profundas pueden identificar patrones más complejos basados en las características simples capturadas por las capas anteriores.

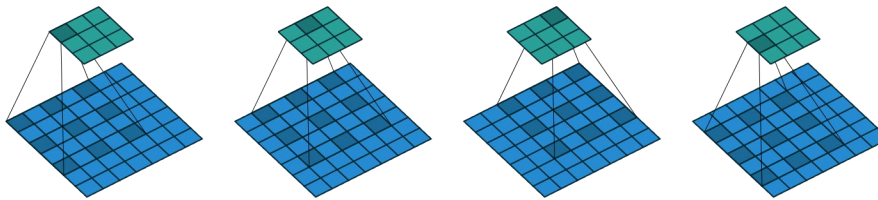
Las CNN han revolucionado el campo de la visión por computadora, ofreciendo un rendimiento significativamente mejorado en comparación con los enfoques tradicionales en tareas como el reconocimiento de imágenes, la detección de objetos, y la segmentación semántica. Su eficacia en la captura de características espaciales también las ha hecho útiles en otros dominios, como el procesamiento del lenguaje natural para el análisis de secuencias de texto (Courses Teach, 2023).

5.2.1 Convolución Atrous

La convolución atrous, también conocida como convolución dilatada, es una variante de la convolución tradicional utilizada en el procesamiento de imágenes y en redes neuronales convolucionales (CNNs). La diferencia principal entre una convolución atrous y una convolución estándar radica en la inserción de espacios entre los elementos del kernel de la convolución (L.-C. Chen et al., 2017). En una convolución tradicional, un kernel se desliza sobre la imagen o el mapa de características de entrada, realizando una multiplicación y suma de elementos en cada posición, así como se puede ver en la **Figura 5-3**. Esto permite extraer características locales de la entrada. La convolución atrous modifica este proceso introduciendo un parámetro de tasa de dilatación que define el espaciamiento entre los elementos del kernel.

Figura 5-3: *Filtro de una convolución 3x3*

Nota. (No padding, unit strides) Convolving a 3 X 3 kernel over a 4 X 4 input using unit strides (i.e., $i = 4$, $k = 3$, $s = 1$ and $p = 0$). Tomado de (Dumoulin et al., 2018)

Figura 5-4: *Filtro de una convolución atrous (dilatada) 3x3*

Nota. (Dilated convolution) Convolving a 3 X 3 kernel over a 7 X 7 input with a dilation factor of 2 (i.e., $i = 7$, $k = 3$, $d = 2$, $s = 1$ and $p = 0$). Tomado de (Dumoulin et al., 2018)

Por ejemplo, en una convolución atrous con una tasa de dilatación de 2, habría espacios entre los elementos del kernel, véase la **Figura 5-4**. Esto aumenta el campo receptivo del filtro, permitiéndole abarcar una región más amplia de la entrada sin aumentar el número de parámetros. Este enfoque es útil para capturar características a una escala más amplia y es especialmente beneficioso en tareas como la segmentación semántica, donde se necesita una visión más global de la entrada. La convolución atrous permite una extracción de características más densa en comparación con las convoluciones estándar, especialmente en las capas más profundas de una red, donde la reducción de la resolución espacial (debido a las operaciones de pooling o convoluciones con stride) puede ser un problema. Al utilizar una tasa de dilatación mayor, la convolución atrous puede abarcar información a diferentes escalas sin perder la resolución espacial, lo que es crucial para algunas aplicaciones como la segmentación de imágenes detallada (Chollet, 2016).

5.2.2 Arquitectura Encoder-Decoder

La arquitectura de codificador-decodificador en redes neuronales, es una estructura fundamental para el procesamiento de datos, especialmente en aplicaciones como la segmentación de imágenes. Esta arquitectura se divide en dos componentes principales: el codificador (encoder) y el decodificador (decoder), cada uno con funciones específicas y complementarias (Cho et al., 2014).

Encoder:

- **Procesamiento de Entrada de Alta Dimensión:** El codificador recibe una entrada de alta dimensión, como una imagen, y comienza el proceso de transformación. En el contexto de la segmentación de imágenes, la entrada podría ser una imagen compleja que necesita ser analizada y segmentada en diferentes partes o categorías.
- **Extracción de Características:** A través de varias capas computacionales y de submuestreo, el codificador procesa la entrada para extraer características esenciales. Estas capas incluyen convoluciones, que ayudan a identificar patrones y elementos clave en la imagen, y operaciones de agrupamiento (pooling), que reducen las dimensiones de los datos manteniendo la información importante.
- **Reducción de Dimensionalidad:** El objetivo principal del codificador es comprimir la información de alta dimensión en un conjunto de características de baja dimensión. Esta compresión es crucial para identificar los aspectos más significativos de los datos, facilitando así la tarea de procesamiento y análisis posteriores.

Decoder

- **Reconstrucción de Datos:** El decodificador toma las características de baja dimensión producidas por el codificador y comienza el proceso de reconstruir la salida deseada. En el caso de la segmentación de imágenes, esto implica crear una imagen segmentada clara a partir de las características abstractas extraídas por el codificador.
- **Proceso Inverso al Codificador:** Mientras que el codificador reduce la dimensionalidad, el decodificador realiza la operación inversa. Utiliza capas que incrementan gradualmente las dimensiones de los datos, a menudo a través de operaciones como "up-sampling" o "transposed convolutions", para reconstruir una salida de alta dimensión que se asemeje a la estructura original de los datos de entrada.

- **Generación de Resultados Finales:** Finalmente, el decodificador produce la salida final, que, en el contexto de la segmentación de imágenes, sería una imagen con distintas regiones claramente definidas y clasificadas según las características aprendidas.

5.2.1 Backbone

En el contexto de Deep Learning se refiere a una red neuronal conocida y pre-entrenada que se utiliza para la extracción de características en tareas de visión por computadora y otros modelos de aprendizaje profundo. Estos backbones son esenciales para capturar información relevante de los datos de entrada (como imágenes o vídeos) y se utilizan comúnmente al principio de los modelos de Deep Learning (Elharrouss et al., 2022). En los modelos de tipo encoder-decoder, el backbone generalmente forma parte del encoder. La función del encoder es comprimir la entrada en una representación de menor dimensión que captura la información esencial. El backbone, como parte del encoder, ayuda en este proceso al extraer características significativas de los datos de entrada. Estas características son luego utilizadas por el decoder para realizar la tarea específica del modelo, como puede ser la generación de imágenes, la traducción de idiomas, etc.

Por ejemplo, en tareas de visión por computadora, un backbone puede ser una red convolucional pre entrenada como VGG, ResNet, o MobileNet, que se encarga de extraer patrones visuales complejos (como bordes, texturas, formas) de las imágenes. Esta información condensada es luego utilizada por el decoder para realizar tareas como la clasificación, detección de objetos, segmentación semántica, entre otras.

5.2.2 Modelos pre entrenados

Los modelos pre entrenados en el contexto de la inteligencia artificial y el aprendizaje automático son modelos que ya han sido entrenados en un gran conjunto de datos para realizar una tarea específica. Estos modelos suelen ser desarrollados y entrenados utilizando cantidades masivas de datos, a menudo en tareas generales como el reconocimiento de imágenes o el procesamiento del lenguaje natural. Una vez entrenados, estos modelos pueden ser reutilizados o adaptados para tareas similares o relacionadas, lo que se conoce como aprendizaje por transferencia (H. Chen et al., 2021). El uso de

estos modelos tiene varios beneficios como el ahorro de tiempo y recursos computacionales, mejorar el rendimiento a comparación de un modelo entrenado desde cero, gracias al conocimiento previo que ya poseen y versatilidad en diferentes tareas, algunos modelos son lo suficientemente generales para ser aplicados a una variedad de tareas.

5.3 MobileNet

MobileNet¹ es una arquitectura de red neuronal diseñada para proporcionar modelos eficientes y efectivos para aplicaciones móviles. El enfoque central del desarrollo de MobileNet es optimizar la relación entre la precisión y la eficiencia, especialmente en términos de latencia y uso de recursos, lo cual es crucial para dispositivos con recursos limitados como teléfonos móviles. El concepto fundamental detrás de MobileNet es el uso de convoluciones separables en profundidad, que dividen una convolución estándar en dos capas: una capa de convolución de profundidad que aplica un solo filtro por canal de entrada, y una capa de convolución 1x1 que combina las salidas de la capa de profundidad. Esta factorización reduce significativamente la cantidad de operaciones y parámetros, permitiendo que la red sea mucho más ligera y rápida sin una gran pérdida de precisión (Elharrouss et al., 2022).

Una de las contribuciones clave de MobileNet es su adaptabilidad a diferentes requisitos de recursos y escenarios de uso. Esto se logra mediante hiperparámetros ajustables, como el ancho del multiplicador y la resolución del multiplicador, que permiten a los usuarios crear un equilibrio entre la precisión y la eficiencia que mejor se adapte a sus necesidades específicas. Con estos ajustes, MobileNet puede adaptarse para ofrecer un rendimiento óptimo en una variedad de dispositivos y aplicaciones, desde teléfonos móviles de gama alta hasta dispositivos IoT con restricciones severas de recursos (Elharrouss et al., 2022). Además, las iteraciones posteriores de MobileNet han introducido mejoras y refinamientos adicionales. Por ejemplo, MobileNetV2 introduce bloques residuales con inversiones

¹ MobileNet Model Reference: Howard et al, (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. ArXiv, abs/1704.04861. Recuperado de <https://arxiv.org/abs/1704.04861>

lineales y expansiones de cuello de botella, lo que mejora la eficiencia sin comprometer la capacidad representativa de la red. También se introducen enfoques de búsqueda de arquitectura de red y mejoras en la arquitectura para adaptar y optimizar aún más los modelos para dispositivos móviles.

5.4 DeepLab

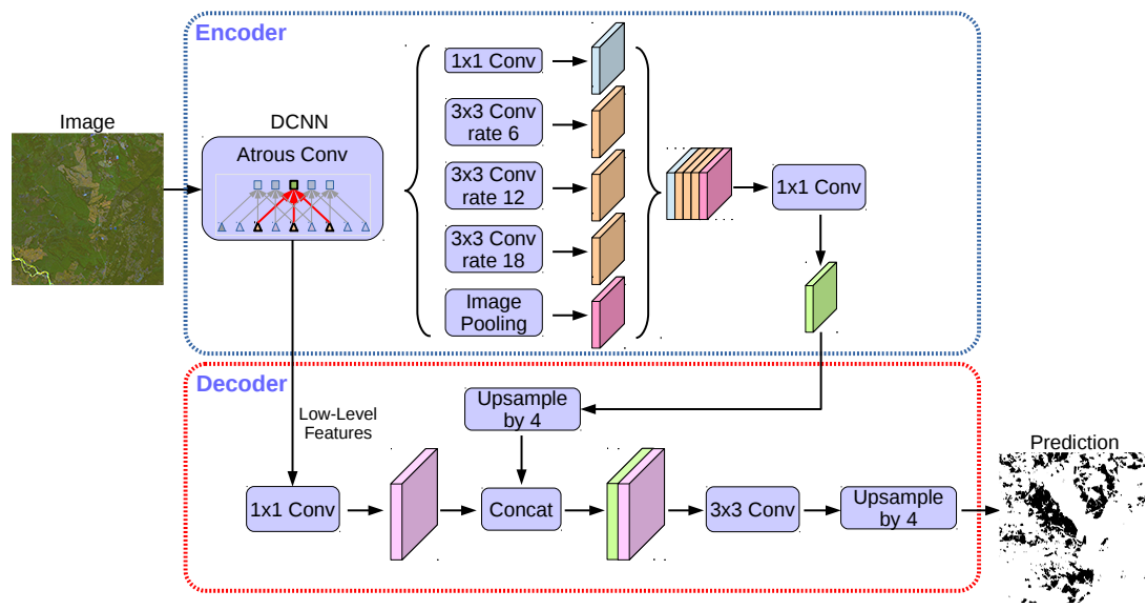
DeepLab es un modelo innovador en el campo del aprendizaje profundo, especialmente diseñado para la segmentación semántica, una tarea que implica la clasificación de cada píxel de una imagen en categorías semánticas. Lo que distingue a DeepLab de otros modelos es su capacidad para capturar detalles finos y entender el contexto de la imagen a un nivel más profundo (L.-C. Chen et al., 2016). Uno de los componentes clave de DeepLab² es la convolución atrous. Esta técnica permite al modelo capturar información a diferentes escalas y ampliar su campo receptivo sin aumentar el número de parámetros. Esta característica es crucial para entender los detalles más finos de la imagen, así como para capturar información contextual más amplia, que es fundamental para la segmentación precisa.

Para aprovechar aún más las capacidades de la convolución atrous, DeepLab incorpora Atrous Spatial Pyramid Pooling (ASPP). ASPP utiliza convoluciones atrous a diferentes tasas de dilatación, aplicadas en paralelo, para capturar objetos y características en múltiples escalas. Esta estrategia permite al modelo ser efectivo en la segmentación de objetos de diferentes tamaños, lo que es un desafío común en la segmentación semántica. DeepLab también se beneficia de una estructura de codificador-decodificador (L.-C. Chen et al., 2016). En esta estructura, el codificador reduce la dimensión de la imagen y extrae características clave, mientras que el decodificador se enfoca en recuperar la resolución espacial y los detalles finos. Este enfoque garantiza que, mientras el modelo comprime la imagen para análisis, también es capaz de reconstruir los detalles necesarios para una

² DeepLab Model Reference: Chen et al, (2017). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. ArXiv, abs/1606.00915. Recuperado de <https://arxiv.org/abs/1606.00915>

segmentación precisa. A continuación, se presenta una ilustración de la arquitectura de DeepLabV3+³:

Figura 5-5: *Arquitectura DeepLabV3+*



Nota. Adaptado de (L.-C. Chen et al., 2018)

Como se mencionó previamente la arquitectura se basa en un codificador poderoso que utiliza convolución atrous, una técnica que ha sido heredada de su predecesor DeepLabV3⁴. Un avance notable en DeepLabV3+ es la introducción de convoluciones atrous separables. Esta innovación se traduce en una reducción significativa en el número de cálculos necesarios para el procesamiento de imágenes, sin comprometer la eficacia del modelo. Esta optimización se logra descomponiendo las convoluciones en dos partes: una convolución atrous seguida de una convolución punto a punto (1x1), lo que mejora tanto la eficiencia del cálculo como la precisión de la segmentación.

³ DeepLab v3+ Model Reference: Chen et al, (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. ArXiv, abs/1802.02611. Recuperado de <https://arxiv.org/abs/1802.02611>

⁴ Código Fuente de DeepLab v3+: <https://github.com/google-research/deeplab2>

Por otra parte, el decodificador en DeepLabV3+ juega un papel fundamental en la mejora de la precisión de la segmentación, especialmente en los límites de los objetos. Trabaja para recuperar la información espacial que se pierde durante el proceso de codificación, asegurando así que los detalles finos y las fronteras de los objetos se capturen con precisión. Esta capacidad de refinamiento es lo que diferencia a DeepLabV3+ de otros modelos, permitiéndole producir resultados de segmentación más precisos y detallados. La aplicación de convoluciones separables en profundidad tanto en el módulo ASPP como en el decodificador no solo mejora la precisión, sino que también aumenta la eficiencia computacional del modelo (L.-C. Chen et al., 2018).

Además, DeepLab a menudo incluye pasos adicionales para mejorar la precisión en los límites de los objetos. Esto se logra mediante la aplicación de técnicas avanzadas en la etapa de decodificación, asegurando que las fronteras de los objetos se mantengan nítidas y bien definidas. El resultado final es una predicción de píxeles altamente precisa, donde cada píxel en la imagen se asigna a una categoría específica, resultando en una segmentación detallada y precisa.

Otro aspecto importante de las redes backbone en DeepLabV3+ es su adaptabilidad a diferentes resoluciones de entrada. Esto permite que el modelo se ajuste a diferentes presupuestos de recursos computacionales, lo que es crucial en aplicaciones donde se necesita un equilibrio entre precisión y eficiencia. La capacidad de trabajar con diversas resoluciones asegura que DeepLabV3+ sea aplicable en una amplia gama de escenarios, desde dispositivos móviles con recursos limitados hasta potentes servidores de cómputo.

Finalmente, la integración de la red backbone con el módulo del decodificador en DeepLabV3+ es un factor clave para su éxito. Mientras que la red backbone actúa como un codificador, el módulo del decodificador refina los resultados de segmentación. Este refinamiento es especialmente crucial a lo largo de los límites de los objetos, donde el decodificador utiliza características de bajo nivel de la red backbone para mejorar la precisión. Además, el modelo utiliza convoluciones separables atrous en ambos módulos, ASPP y decodificador, para mejorar la eficiencia computacional sin sacrificar el rendimiento.

5.5 Métricas de desempeño

En el ámbito del aprendizaje automático, la evaluación de modelos de clasificación es esencial para asegurar su precisión y eficacia. Esta sección se enfoca en las métricas más relevantes para dicha evaluación basadas en el artículo de Goyal (2021):

- **Precisión de Clasificación (Accuracy):** Esta métrica es la más simple y se calcula como la proporción de predicciones correctas frente al total de predicciones realizadas. Es útil cuando las clases objetivo están bien balanceadas. Sin embargo, no es la mejor elección cuando se trabaja con clases desbalanceadas, ya que puede dar una impresión erróneamente positiva del desempeño del modelo.

$$Accuracy = \frac{\text{Número de predicciones correctas}}{\text{Total de predicciones hechas}} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Matriz de Confusión:** Esta matriz muestra el número de predicciones correctas e incorrectas realizadas por el modelo en comparación con las clasificaciones reales en el conjunto de prueba. Ayuda a entender qué tipos de errores está cometiendo el modelo y es especialmente útil para visualizar el rendimiento en problemas de clasificación binaria o multiclase.

Tabla 5-2: Matriz de confusión

		Valores reales	
		Verdaderos	Falsos
Predicciones	Verdaderos	TP: verdaderos positivos	FP: Falsos positivos
	Falsos	FN: Falsos negativos	TN: Verdaderos negativos

- **Precisión:** Se calcula como la proporción de verdaderos positivos entre todas las predicciones positivas hechas por el modelo. Esta métrica es importante en conjuntos de datos sesgados o desbalanceados y se enfoca en los falsos positivos.

$$Precisión = \frac{TP}{TP + FP}$$

- **Recall:** Es la proporción de verdaderos positivos frente a todos los positivos reales en el conjunto de datos. Mide la capacidad del modelo para detectar muestras positivas, siendo especialmente importante cuando los falsos negativos son críticos.

$$Recall = \frac{TP}{TP + FN}$$

- **F1 Score:** Combina la precisión y la recuperación en una sola métrica, proporcionando un balance entre ambas. Es útil cuando se desea un equilibrio entre evitar falsos positivos y falsos negativos.

$$F1 = \frac{2 * Precisión * Recall}{Precisión + Recall}$$

- **Curva ROC:** Muestra la relación entre la Tasa de Verdaderos Positivos (Sensibilidad) y la Tasa de Falsos Positivos a diferentes umbrales de decisión. Es una herramienta visual para comparar el rendimiento de varios modelos de clasificación.
- **Área Bajo la Curva ROC (AUC - ROC):** Resume la curva ROC en un solo valor, proporcionando una medida agregada del rendimiento a lo largo de todos los posibles umbrales de clasificación. Un AUC más alto indica un mejor rendimiento del modelo en términos de distinguir entre las clases positivas y negativas.

6.Desarrollo de la solución

En esta sección, se presenta un panorama completo del desarrollo de una solución para clasificar cultivos de palma de aceite en Colombia sobre imágenes satelitales de Sentinel 1 y Sentinel 2. Se aborda desde el establecimiento del ambiente de desarrollo y las restricciones técnicas, hasta la construcción detallada del conjunto de datos, incluyendo la selección y procesamiento de imágenes satelitales y su etiquetado. También se describe el enriquecimiento del conjunto de datos mediante técnicas de aumentación y la implementación y validación del modelo de aprendizaje profundo, enfatizando en las estrategias empleadas para garantizar la precisión y efectividad del sistema en la identificación y clasificación de plantaciones de palma aceitera en Colombia.

6.1 Ambiente de desarrollo, datos y restricciones

El ambiente de desarrollo consistió en una máquina con sistema operativo Ubuntu, equipada con una tarjeta gráfica de 16 GB de VRAM y 16 GB de RAM. Este entorno fue elegido por su compatibilidad y rendimiento con las herramientas de Deep Learning. Se utilizó Python como lenguaje de programación principal, especialmente con la librería PyTorch para el desarrollo y entrenamiento de los modelos. Las restricciones encontradas incluyeron limitaciones de memoria y tiempo de procesamiento, lo cual fue mitigado mediante técnicas de optimización de modelos y uso eficiente de recursos.

6.2 Construcción del conjunto de datos

6.2.1 Preprocesamiento de Imágenes

El preprocesamiento de las imágenes de los satélites Sentinel-1 y Sentinel-2 fue un paso crucial en la construcción del conjunto de datos. Este proceso se basó en la adaptación de la composición presentada por Descals et al, (2021). El proceso comenzó con la selección y descarga de imágenes de estos satélites desde Google Earth Engine (GEE). Para el

Sentinel-1, se utilizó el radar de apertura sintética (SAR) en modo Ground Range Detected (GRD), el cual proporciona una resolución temporal de 12 días en órbitas ascendentes y descendentes. Las imágenes utilizadas correspondían a la modalidad de Interferometric Wide Swath, procesadas a una resolución espacial de 10 metros.

Un aspecto importante del preprocesamiento fue la corrección del ángulo de incidencia local (LIA), que es fundamental para ajustar las imágenes de radar para variaciones topográficas y de orientación. Una vez realizada esta corrección, se calculó el valor mediano de las escenas de radar para la segunda mitad del año 2019, tanto para las órbitas ascendentes como las descendentes. Finalmente, se creó un compuesto final a partir del promedio de estos dos conjuntos de imágenes compuestas, representando tanto las órbitas ascendentes como las descendentes.

Este meticuloso proceso de preprocesamiento de las imágenes de Sentinel-1 y Sentinel-2 fue esencial para asegurar la calidad y la precisión de los datos utilizados en el estudio, permitiendo una representación precisa de las plantaciones de palma aceitera a nivel mundial. Se utilizaron específicamente la banda 4 del satélite Sentinel-2 y el radar de apertura sintética (SAR) del Sentinel-1. Para el Sentinel-2, se empleó la banda 4, también conocida como la banda roja, que tiene una longitud de onda central de 665 nm y proporciona datos de reflectancia de la superficie en una resolución de 10 metros. Esta banda fue seleccionada debido a su alta utilidad en la identificación de plantaciones de palma aceitera, particularmente en las plantaciones industriales. La banda 4 destaca por ofrecer un alto contraste en términos de reflectancia entre los caminos dentro de las plantaciones y las áreas circundantes de palma aceitera. Este contraste facilita la identificación de los caminos en las imágenes satelitales, un elemento clave para diferenciar entre plantaciones industriales y de pequeños propietarios. Otro factor que respalda la elección de esta banda es que la dispersión de luz alta en el espectro del infrarrojo cercano hace que la identificación de caminos sea menos factible en otras bandas, como la banda 8 que también opera en el infrarrojo cercano.

Por otro lado, para el Sentinel-1, se utilizó el modo Ground Range Detected (GRD) del radar de apertura sintética (SAR). Este modo se procesó en la modalidad de Interferometric Wide Swath, con una resolución espacial de 10 metros. Los datos del Sentinel-1, que

trabajan en la banda C del espectro de microondas, son especialmente valiosos para la detección y clasificación de características en la superficie terrestre, incluyendo la vegetación. Este modo de radar es capaz de proporcionar datos confiables independientemente de las condiciones climáticas o la luz del día, lo que lo hace muy útil para el monitoreo constante de áreas extensas como las plantaciones de palma aceitera. Además, los datos de Sentinel-1 fueron procesados con una corrección del ángulo de incidencia local (LIA), y se calculó el valor mediano para las escenas ascendentes y descendentes por separado durante la segunda mitad de 2019, creando así un compuesto final a partir del promedio de estas dos órbitas.

6.2.2 Etiquetado de Imágenes

El proceso de etiquetado de imágenes fue un paso crucial para el entrenamiento y la predicción utilizando modelos de segmentación semántica. Este proceso implicó varios pasos detallados a continuación:

- **Establecimiento del Tamaño de las Imágenes:** Las imágenes utilizadas para el etiquetado y el entrenamiento del modelo necesitaban tener un tamaño constante. Siguiendo el trabajo realizado por Descals et al. (2021), se estableció un tamaño de imagen de entrada de 1000×1000 píxeles. Este tamaño corresponde a un área de 10×10 km en una imagen con una resolución de 10 metros.
- **Selección de Imágenes de Satélite:** Para el etiquetado, se utilizaron la composición que se describe en el preprocesamiento en un periodo de tiempo semestral de Sentinel-1 y Sentinel-2.
- **Etiquetado:** Para el etiquetado de las imágenes se usaron los datos de Descals et al. (2021) en las zonas seleccionadas. Adicionalmente, se cambiaron las máscaras de etiquetado de pequeños agricultores y los industriales a un solo tipo de categoría.

6.2.3 Aumento de Datos

El proceso de aumentación de datos se centró en generar un conjunto de datos de entrenamiento más diverso a partir del conjunto original. Esta técnica es especialmente útil cuando el tamaño del conjunto de datos de entrenamiento es limitado. En el caso de este estudio, se aplicaron transformaciones afines a los datos de entrenamiento originales. La técnica específica utilizada fue la rotación de imágenes. Se rotaron las imágenes 90 grados en el sentido de las agujas del reloj.

El uso de transformaciones afines como la rotación es una práctica común en estudios de teledetección y aprendizaje profundo. Estas técnicas han demostrado mejorar la precisión de los modelos de aprendizaje profundo, ya que ayudan a crear un conjunto de datos de entrenamiento que representa de manera más efectiva la variabilidad y las diferentes orientaciones que pueden encontrarse en imágenes reales.

6.3 Implementación del modelo

El modelo en cuestión acepta como entrada imágenes en formato PNG. La salida generada por el modelo son igualmente imágenes en formato PNG, con una resolución de 1000x1000 píxeles. Tras la generación de la salida, se procede a la creación de una imagen en formato TIFF. Esta imagen se caracteriza por incorporar georreferenciación y una máscara de salida proporcionada por el modelo. Para la implementación de la arquitectura DeepLabV3+, inicialmente propuesta por Chen et al. (2018), se ha optado por el uso de la biblioteca `segmentation_models.pytorch` de PyTorch. El modelo se configuró utilizando el encoder `timm-mobilenetv3_large_100`, el cual fue preentrenado con el dataset ImageNet. Dado el objetivo del modelo de clasificar la presencia o ausencia de palma de aceite, se establecieron dos clases y se optó por la función de activación sigmoid. La función de pérdida seleccionada fue la Entropía Cruzada Binaria (Binary Cross Entropy - BCE), considerada idónea para tareas de clasificación binaria.

El entrenamiento del modelo se desarrolló a lo largo de 20 épocas. Durante cada época, se evaluaron y registraron las métricas especificadas en la sección 6.4.1. Se procedió a guardar el modelo que demostraba el mejor rendimiento en cada instancia. Se definió un

tamaño de lote (batch size) de 7, optimizando el uso de los recursos computacionales disponibles. El conjunto de datos empleado para el entrenamiento y la validación del modelo consistió en 270 imágenes. De estas, el 70% (189 imágenes) se asignaron al conjunto de entrenamiento y el 30% restante (81 imágenes) al conjunto de validación, asegurando así una distribución equilibrada de los datos, lo que facilitó un entrenamiento eficiente y una validación precisa del modelo.

6.4 Validación del modelo

6.4.1 Métricas

Para una evaluación integral y precisa del modelo, se seleccionaron cuatro métricas clave: precisión (accuracy), sensibilidad (recall), precisión (precision) y la puntuación F1 (f1-score). Cada una de estas métricas aporta una perspectiva única en la evaluación del modelo, permitiendo un análisis exhaustivo de su rendimiento. La combinación de estas métricas proporciona una visión holística y equilibrada del rendimiento del modelo, asegurando no solo su capacidad para clasificar con exactitud, sino también su eficiencia en la identificación de casos relevantes y la minimización de falsos positivos.

6.4.2 Validación de los resultados

Para la validación del modelo, se seleccionaron específicamente varias zonas dentro del territorio colombiano. Estas áreas fueron elegidas en función de su relevancia agrícola y la presencia diversificada de cultivos, incluyendo, de manera prominente, plantaciones de palma de aceite. La elección de estas zonas permite una evaluación efectiva del modelo en diferentes escenarios y condiciones ambientales, lo cual es crucial para asegurar su robustez y adaptabilidad. Como parte del proceso de validación, se realizaron evaluaciones visuales utilizando el software QGIS⁵. Esta herramienta de Sistemas de Información Geográfica (SIG) permitió una inspección detallada y minuciosa de las

⁵ QGIS es un Sistema de Información Geográfica libre y de código abierto que permite crear, editar, visualizar, analizar y publicar información geoespacial. Además, proporciona bibliotecas para desarrolladores. Para más información, puede consultar el sitio web en [QGIS](https://qgis.org/).

predicciones realizadas por el modelo sobre los diferentes cultivos. La evaluación visual en QGIS implicó la comparación de las imágenes de salida del modelo con las imágenes satelitales reales y los datos existentes sobre la ubicación y extensión de los cultivos en las zonas de estudio. Este enfoque permitió identificar no solo la precisión general del modelo, sino también comprender mejor sus limitaciones y áreas de mejora potencial, como la diferenciación entre tipos de cultivo y la precisión en la detección de los límites de las plantaciones.

7.Pruebas y resultados

El modelo seleccionado, que demostró la mayor eficacia, se basó en la arquitectura MobileNet V3. Esta red neuronal convolucional fue elegida como la red troncal del modelo debido a su capacidad para operar eficientemente en dispositivos con recursos limitados, sin comprometer significativamente el rendimiento. Los resultados obtenidos son altamente prometedores, como se evidencia en las métricas de rendimiento del modelo. Se alcanzó una exactitud del 98.37%, lo cual implica que el modelo es capaz de clasificar correctamente el 98.37% de las imágenes probadas. Esta métrica es particularmente importante en aplicaciones prácticas donde la fiabilidad de la clasificación es crítica para el análisis posterior.

El recall obtenido fue de 0.9914, lo que refleja la alta sensibilidad del modelo en la detección de las regiones con presencia de palma de aceite. En términos prácticos, esto significa que el modelo identificó correctamente el 99.14% de todas las áreas relevantes en el conjunto de datos. Esta tasa de detección es crucial en estudios de cobertura de tierra donde la omisión de áreas significativas puede llevar a conclusiones erróneas.

En cuanto a la precisión, el modelo alcanzó un 99.04%, indicando que, de las regiones clasificadas como palma de aceite, el 99.04% efectivamente correspondían a dichos cultivos. Esta alta precisión es indispensable para minimizar los falsos positivos, una métrica especialmente importante cuando se toman decisiones de gestión de tierras basadas en los datos de clasificación.

Finalmente, el F1-score del modelo fue de 0.9909, revelando un equilibrio casi perfecto entre precisión y recall. Esto sugiere que el modelo no solo es altamente confiable en la identificación de áreas de palma de aceite, sino que también mantiene una baja tasa de

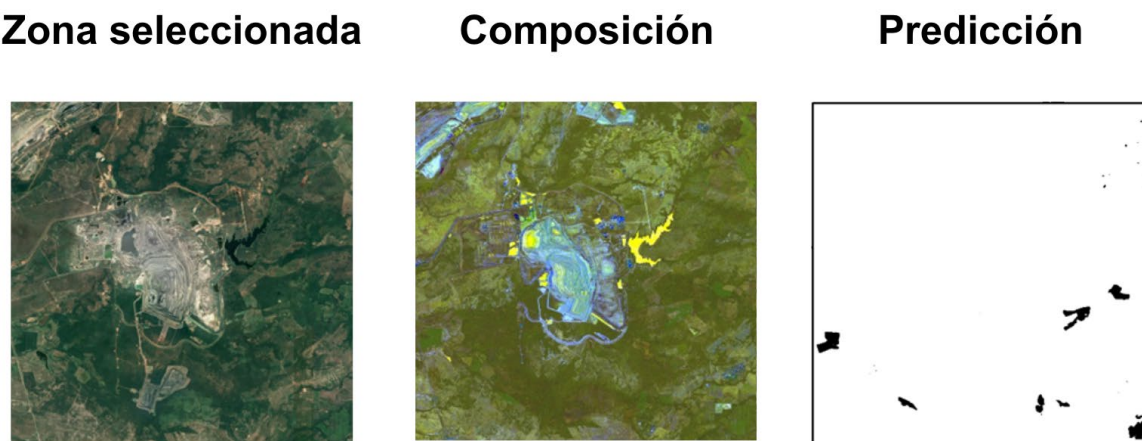
falsos positivos y negativos, un aspecto fundamental para la integridad del análisis espacial. Estos resultados subrayan la eficacia del modelo MobileNet V3 para la detección de palma de aceite y su aplicabilidad potencial en la monitorización a gran escala de la agricultura y la gestión del uso de la tierra.

Figura 7-1: *Ejemplo de predicción del modelo sobre un área con cultivos*



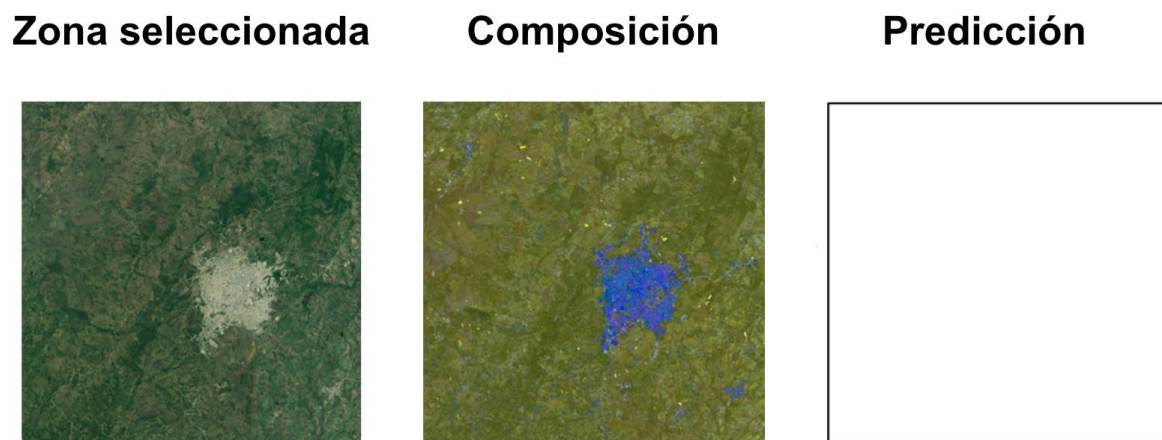
En la **Figura 7-1** podemos ver una zona seleccionada de 10x10 km de Colombia, seguido de la composición de Sentinel 1 y Sentinel 2 y por último tenemos la salida del modelo la en la cual la clase 1 es la que esta en negro donde el modelo clasifica que hay palma de aceite mientras que donde en la mascara donde esta en blanco el modelo indica que no hay palma de aceite en esta imagen.

Figura 7-2: *Ejemplo de predicción del modelo sobre un área con pocos cultivos*

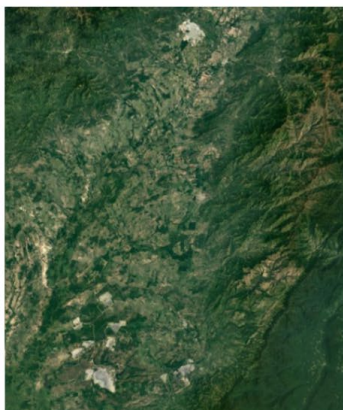
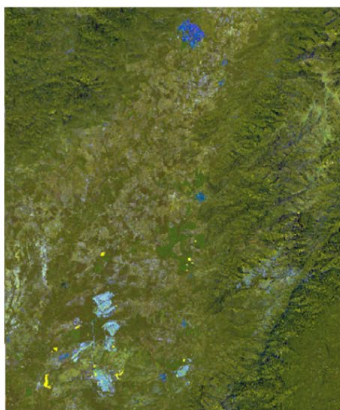


En la **Figura 7-2**, podemos observar una zona de explotación minera que, en la composición, adquiere un color celeste. Además, se identifica un cuerpo de agua que adopta un color amarillo en la composición. Por otra parte, tenemos las plantaciones de palma de aceite, las cuales presentan un color verde distintivo. En esta zona seleccionada, los cultivos de palma de aceite son escasos, motivo por el cual el color verde no es tan predominante en la imagen. En la máscara predicha por el modelo, podemos observar que se clasifican correctamente las zonas que contienen plantaciones de palma de aceite.

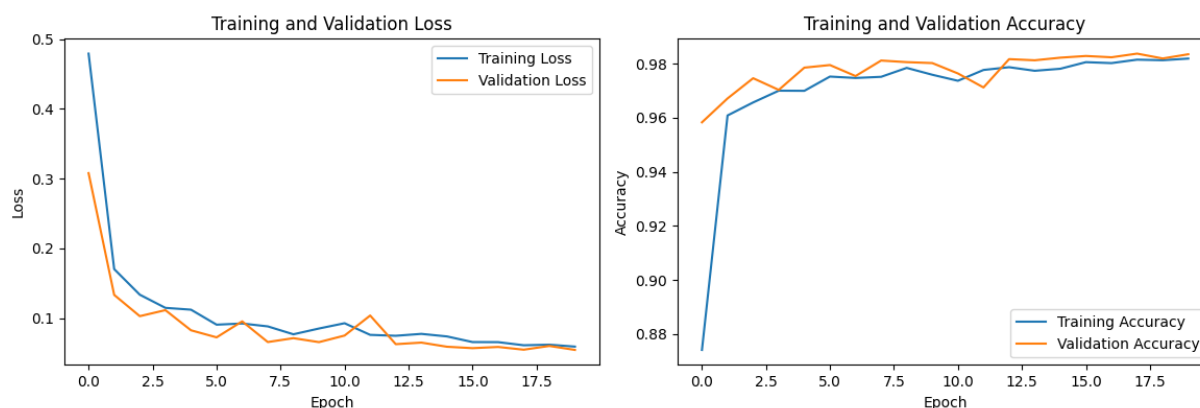
Figura 7-3: *Ejemplo de predicción del modelo sobre área sin cultivos*



En la **Figura 7-3** se seleccionó una zona donde no hay presencia de palma de aceite y, además, existe una área metropolitana. En este caso, se puede observar que el modelo devuelve una máscara completamente blanca, indicando que el modelo clasifica correctamente la ausencia de cultivos de palma de aceite en esta zona.

Figura 7-4: *Predicción del modelo sobre toda la zona seleccionada***Zona seleccionada****Composición****Predicción**

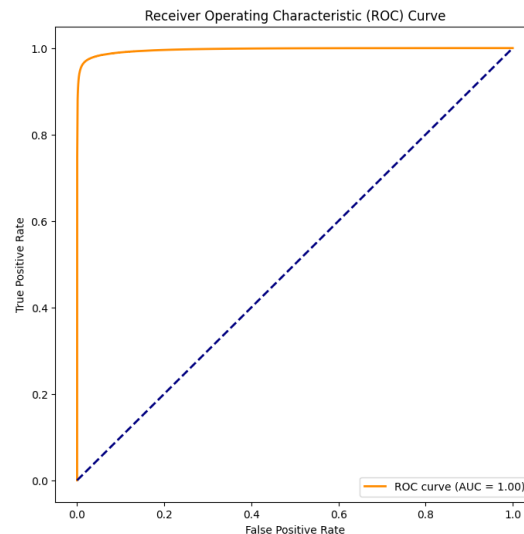
En la **Figura 7-4** se presenta la zona seleccionada en su totalidad, donde se encuentran áreas metropolitanas, bosques, diversas plantaciones de otros cultivos, además de zonas rocosas y una gran diversidad de flora. A continuación, se muestra la predicción, en la que las áreas en negro representan las regiones que el modelo de clasificación ha identificado como contenedoras de palma de aceite. El color blanco representa áreas donde el modelo predice la ausencia de palma de aceite.

Figura 7-5: *Gráficas de pérdida (loss) y exactitud (accuracy)*

En el gráfico de la izquierda, las líneas azul y naranja indican la pérdida de entrenamiento y validación, respectivamente. Se observa una disminución rápida de la pérdida en las primeras épocas, seguida de una estabilización, lo que sugiere que el modelo rápidamente

alcanzó un punto de convergencia. En el gráfico de la derecha, las líneas representan la precisión de entrenamiento y validación, ambas alcanzando valores superiores al 90% y manteniéndose relativamente constantes después de las primeras épocas, lo que indica un buen ajuste del modelo.

Figura 7-6: *Curva ROC del modelo*



La figura **Figura 7-6** representa la Curva Característica de Operación del Receptor (ROC) con un área bajo la curva (AUC) de 1.00. La curva ROC es una línea naranja que se adhiere perfectamente al borde superior izquierdo del gráfico, lo cual es indicativo de un rendimiento excepcional del modelo, ya que demuestra una tasa de verdaderos positivos del 100% a lo largo de todos los umbrales de clasificación, con una tasa de falsos positivos que se mantiene en cero. El gráfico también incluye una línea punteada azul que representa el rendimiento de un clasificador aleatorio, lo que resalta aún más la superioridad del modelo en cuestión.

Por último, haciendo una comparación con el modelo presentado por (Díaz, 2023) en la que se utilizaron una variedad de algoritmos de aprendizaje automático más tradicionales, como regresión logística, Random Forest, SVM y XGBoost. Aunque estos métodos son eficaces y versátiles, pueden no ser tan específicos como los modelos de Deep Learning para tareas detalladas de segmentación de imágenes. Además, Díaz incorporó índices de vegetación y variables de textura en su análisis, lo que indica un enfoque en aspectos específicos de la vegetación y el terreno. Aunque esta metodología es útil, posiblemente

no captura la gama completa de características que se pueden obtener con el enfoque del usuario. La precisión del modelo de Díaz fue ligeramente inferior, con un 96%, lo que sugiere que los enfoques más tradicionales de aprendizaje automático pueden seguir siendo efectivos para tareas de clasificación detalladas en comparación con métodos avanzados de Deep Learning.

8. Conclusiones y trabajo futuro

8.1 Conclusiones

La investigación desarrollada en esta tesis ha culminado en una serie de hallazgos significativos en el ámbito de la clasificación de cultivos mediante técnicas de Deep Learning, utilizando imágenes satelitales de Sentinel 1 y Sentinel 2. Los resultados obtenidos han demostrado que el modelo propuesto, basado en técnicas avanzadas de aprendizaje profundo, es altamente eficaz en la identificación de plantaciones de palma de aceite, alcanzando una precisión global del 98%. Esta alta precisión subraya la capacidad del modelo para integrar eficientemente datos de radar y ópticos, lo que resalta el potencial de las imágenes SAR de Sentinel 1 y las ópticas de Sentinel 2 en la clasificación precisa de este tipo de cultivos.

Sin embargo, el estudio también ha identificado ciertas limitaciones en el modelo, especialmente en la detección de áreas recién despejadas y plantaciones jóvenes. Este hallazgo enfatiza la necesidad de incorporar datos adicionales que reflejen la edad y el estado de desarrollo de las plantaciones para una evaluación más precisa y detallada. Además, se observó que la aplicación de técnicas de aumento de datos, como la rotación de imágenes, ha sido fundamental para mejorar la eficacia del modelo. Estas técnicas han probado ser particularmente valiosas en regiones con una cantidad limitada de datos de entrenamiento, demostrando la importancia de un conjunto de datos diversificado y enriquecido en estudios de teledetección.

Al comparar estos resultados con el enfoque más tradicional de aprendizaje automático empleado por Diaz (2023), que utilizó una combinación de algoritmos convencionales y análisis de índices de vegetación, se observa que, aunque estos métodos tradicionales son efectivos, el modelo basado en Deep Learning destaca por su especialización y mayor precisión en la tarea específica de clasificar cultivos de palma de aceite. Este contraste

pone en un pequeño relieve la superioridad de las técnicas de Deep Learning en aplicaciones especializadas de teledetección.

8.2 Trabajo futuro

Se sugiere entrenar el modelo con más datos, en especial imágenes de plantaciones jóvenes y áreas recién despejadas. Además, explorar el uso de técnicas más avanzadas de procesamiento de imágenes, como redes neuronales generativas, para mejorar la clasificación en condiciones de baja calidad de imagen. Además, se propone extender la investigación a otros tipos de cultivos y condiciones geográficas para evaluar la adaptabilidad del modelo. Un área de interés particular podría ser el análisis de cambios temporales en las plantaciones, utilizando series de tiempo de imágenes satelitales para monitorear el desarrollo y salud de los cultivos. También sería valioso investigar el impacto de diferentes condiciones climáticas y estacionales en la precisión del modelo. Por último, un enfoque en la interpretación y explicabilidad de los modelos de Deep Learning podría proporcionar puntos de vista más profundos sobre cómo y por qué el modelo toma ciertas decisiones, lo cual es crucial para su implementación y confianza en escenarios de la vida real.

Bibliografía

- Bejarano, J. A. (1980). Los estudios sobre la historia del café en Colombia. *Cuadernos de Economía*, 1(2), 115–140.
<https://revistas.unal.edu.co/index.php/ceconomia/article/view/19063>
- Boryan, C., Yang, Z., Mueller, R., & Craig, M. (2011). Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. <http://dx.doi.org/10.1080/10106049.2011.562309>, 26(5), 341–358. <https://doi.org/10.1080/10106049.2011.562309>
- Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., & Gao, W. (2021). *Pre-Trained Image Processing Transformer* (pp. 12299–12310). <https://github.com>.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2016). *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*. <http://arxiv.org/abs/1606.00915>
- Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). *Rethinking Atrous Convolution for Semantic Image Segmentation*. <https://arxiv.org/abs/1706.05587v3>
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*. <http://arxiv.org/abs/1802.02611>
- Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *Proceedings of SSST 2014 - 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111. <https://doi.org/10.3115/v1/w14-4012>
- Chollet, F. (2016). *Xception: Deep Learning with Depthwise Separable Convolutions*. <http://arxiv.org/abs/1610.02357>
- Courses Teach. (2023, julio). *Deep Learning (Part 1): Understanding Basic Neural Networks*. <https://medium.com/@Coursesteach/deep-learning-part-1-86757cf5a0c3>
- Cunha, R. L. F., & Silva, B. (2020). ESTIMATING CROP YIELDS WITH REMOTE SENSING AND DEEP LEARNING. *ISPRS Annals of the Photogrammetry*,

- Remote Sensing and Spatial Information Sciences, IV-3/W2-2020*, 59–64.
<https://doi.org/10.5194/isprs-annals-iv-3-w2-2020-59-2020>
- DANE. (2014). *3 er Censo Nacional Agropecuario*.
- Debaeke, P., Attia, F., Champolivier, L., Dejoux, J. F., Micheneau, A., Bitar, A. Al, & Trépos, R. (2023). Forecasting sunflower grain yield using remote sensing data and statistical models. *European Journal of Agronomy*, 142.
<https://doi.org/10.1016/j.eja.2022.126677>
- Descals, A., Wich, S., Meijaard, E., Gaveau, D. L. A., Peedell, S., & Szantoi, Z. (2021). High-resolution global map of smallholder and industrial closed-canopy oil palm plantations. *Earth System Science Data*, 13(3), 1211–1231.
<https://doi.org/10.5194/ESSD-13-1211-2021>
- Diaz, N. (2023). *Herramienta para la estimación del potencial bioenergético en municipios vulnerables de Colombia mediante imágenes satelitales y machine learning*.
- Divakar, M. S., Elayidom, M. S., & Rajesh, R. (2022). Forecasting crop yield with deep learning based ensemble model. *Materials Today: Proceedings*, 58, 256–259. <https://doi.org/10.1016/j.matpr.2022.02.109>
- Dumoulin, V., Visin, F., & Box, G. E. P. (2018). *A guide to convolution arithmetic for deep learning*. <http://ethanschoonover.com/solarized>
- Earth Engine Data Catalog: Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected, log scaling*. (2014).
- Elharrouss, O., Akbari, Y., Almaadeed, N., & Al-Maadeed, S. (2022). *Backbones-Review: Feature Extraction Networks for Deep Learning and Deep Reinforcement Learning Approaches*. <https://arxiv.org/abs/2206.08016v1>
- European Space Agency. (2017). *Earth Engine Data Catalog: Sentinel-2 MSI: MultiSpectral Instrument, Level-2A*. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR.
- Geng, L., Che, T., Ma, M., Tan, J., & Wang, H. (2021). Corn biomass estimation by integrating remote sensing and long-term observation data based on machine learning techniques. *Remote Sensing*, 13(12).
<https://doi.org/10.3390/rs13122352>
- Goyal S. (2021). *Evaluation Metrics for Classification Models*.
<https://medium.com/analytics-vidhya/evaluation-metrics-for-classification-models-e2f0d8009d69>

- Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval*, 7(2), 87–93. <https://doi.org/10.1007/S13735-017-0141-Z/FIGURES/3>
- Hu, K., Wang, Z., Coleman, G., Bender, A., Yao, T., Zeng, S., Song, D., Schumann, A., & Walsh, M. (2021). *Deep Learning Techniques for In-Crop Weed Identification: A Review*.
- Junguito, R., Perfetti, J. J., & Becerra, A. (2014). Desarrollo de la agricultura colombiana. *Cuadernos de Fedesarrollo*. <https://ideas.repec.org/p/col/000439/011565.html>
- Martello, M., Molin, J. P., Wei, M. C. F., Canal Filho, R., & Nicoletti, J. V. M. (2022). Coffee-Yield Estimation Using High-Resolution Time-Series Satellite Images and Machine Learning. *AgriEngineering*, 4(4), 888–902. <https://doi.org/10.3390/agriengineering4040057>
- Thayer, A. W., Vargas, A., Castellanos, A. A., Lafon, C. W., McCarl, B. A., Roelke, D. L., Winemiller, K. O., & Lacher, T. E. (2020). Integrating Agriculture and Ecosystems to Find Suitable Adaptations to Climate Change. *Climate 2020*, Vol. 8, Page 10, 8(1), 10. <https://doi.org/10.3390/CLI8010010>
- Vargas, R., Mosavi, A., & Ruiz, R. (2017). *Deep learning: A review*.
- Victor, B., He, Z., & Nibali, A. (2022). *A systematic review of the use of Deep Learning in Satellite Imagery for Agriculture*. <https://arxiv.org/abs/2210.01272v2>