

Clasificación de Plantaciones de Palma de Aceite en Colombia utilizando Deep Learning y Datos Satelitales de Sentinel-1 y Sentinel-2

Kevin Steven Gamez Abril

Departamento de Ingeniería de Sistemas y Computación

Universidad de los Andes

Bogotá, Colombia

ks.gamez@uniandes.edu.co

Haydemar María Nuñez Castro

Departamento de Ingeniería de Sistemas y Computación

Universidad de los Andes

Bogotá, Colombia

h.nunez@uniandes.edu.co

Andrés Oswaldo Calderón Romero

Department of Computer Science and Engineering

University of California

Riverside, USA

acald013@ucr.edu

Rocío Sierra Ramírez

Departamento de Ingeniería Química y de Alimentos

Universidad de los Andes

Bogotá, Colombia

rsierra@uniandes.edu.co

Abstract—Este estudio presenta un modelo avanzado basado en Deep Learning para la clasificación de plantaciones de palma de aceite en Colombia, utilizando datos satelitales de Sentinel-1 y Sentinel-2. Ante la creciente demanda de alimentos y los desafíos del cambio climático, el enfoque propuesto mejora la gestión agrícola mediante la integración de datos radar y ópticos. El modelo implementado, basado en la arquitectura DeepLabV3+ con MobileNet V3 y ResNet50, emplea estrategias de transferencia de aprendizaje y ajuste fino para optimizar su rendimiento. La metodología comprende tres fases: modelado, implementación y validación, destacándose por una precisión global del 98.37%. Este trabajo subraya la eficacia del Deep Learning en comparación con métodos tradicionales, superando limitaciones como la cobertura de nubes y resolviendo tareas complejas de segmentación de imágenes. Los resultados demuestran su aplicabilidad práctica en la gestión de cultivos, sugiriendo futuras mejoras para incluir plantaciones jóvenes y explorar otros tipos de cultivos y regiones geográficas.

Index Terms—Deep Learning, Clasificación de cultivos, Imágenes satelitales, Sentinel-1 y Sentinel-2, Agricultura de precisión.

I. INTRODUCCIÓN

La agricultura moderna enfrenta retos como el aumento de la demanda alimentaria debido al crecimiento poblacional, el cambio climático y el agotamiento de recursos naturales [1]. Esto resalta la necesidad de optimizar el uso de la tierra para reducir el impacto ambiental y maximizar los beneficios socioeconómicos. Los datos satelitales permiten mapear terrenos agrícolas, gestionar eficientemente recursos críticos como el agua y los fertilizantes, y monitorear la salud de los cultivos, mejorando así la productividad [2].

Tradicionalmente, la identificación y el monitoreo de cultivos dependían de observaciones de campo, un proceso intensivo y propenso a errores. Recientemente, las imágenes aéreas capturadas por vehículos no tripulados han facilitado

este proceso, proporcionando datos valiosos para la toma de decisiones agrícolas [3]. Las imágenes satelitales se han utilizado ampliamente para clasificar cultivos debido a su riqueza informativa.

Este trabajo propone un modelo de clasificación de cultivos basado en técnicas de Deep Learning, utilizando composiciones de bandas de imágenes satelitales. En primer lugar, se analizará el estado del arte sobre el uso de Deep Learning para la identificación de zonas cultivadas, evaluando enfoques existentes y áreas de mejora. Además, se identificarán características relevantes de las imágenes que contribuyan a mejorar la clasificación. Se implementarán estrategias de transferencia de aprendizaje y ajuste fino (Fine-tuning) para aprovechar modelos preentrenados.

El modelo se aplicará a una región específica de Colombia y se comparará con métodos tradicionales basados en índices de vegetación, destacando las ventajas del enfoque de Deep Learning. La metodología se divide en tres fases: modelado, que incluye la selección de la arquitectura y el preprocesamiento de imágenes de Sentinel-1 y Sentinel-2 [4]; implementación, centrada en el entrenamiento y prueba de modelos preentrenados; y validación, en la cual se evaluarán métricas de desempeño y se analizarán las limitaciones del modelo aplicadas a imágenes representativas del territorio colombiano.

II. REVISIÓN DE LA LITERATURA

El uso de técnicas de Deep Learning en el análisis de imágenes satelitales ha revolucionado la agricultura, convirtiéndola en una herramienta clave para la gestión y planificación agrícola. Estas tecnologías han mejorado significativamente la precisión en la predicción del rendimiento de

cultivos y biomasa, facilitando decisiones basadas en datos y optimizando recursos.

Un análisis exhaustivo de 150 estudios [5] identificó cinco tareas principales donde la aplicación de Deep Learning podría traer beneficios: uso y cobertura del suelo, salud del suelo, fisiología de las plantas, daño a los cultivos y predicción del rendimiento. El estudio concluye que los métodos de Deep Learning superan a los tradicionales en la mayoría de las tareas, salvo en la predicción del rendimiento, donde los modelos LSTM muestran un rendimiento comparable a Random Forest. Este hallazgo subraya la importancia de disponer de conjuntos de datos de referencia públicos para una comparación efectiva entre estudios.

A. Optimización de la Agricultura con Deep Learning

El Deep Learning ha emergido como una herramienta clave para optimizar prácticas agrícolas mediante el análisis avanzado de datos satelitales. Su capacidad para procesar grandes volúmenes de información y abordar problemas complejos ha permitido mejorar la precisión en la estimación de rendimientos, superar limitaciones tradicionales y adaptarse a diversas condiciones geográficas. A continuación, se destacan ejemplos representativos que ilustran cómo estas técnicas están transformando la agricultura.

Por ejemplo, la estimación del rendimiento del café en Brasil mediante Deep Learning, combinado con modelos de regresión y análisis de imágenes satelitales, ha demostrado ser eficaz para anticipar la variabilidad espacial en mapas de rendimiento, apoyando la planificación agrícola [6]. Además, la escalabilidad del Deep Learning destacó en un estudio que utilizó datos geográficos y meteorológicos para estimar rendimientos de cinco cultivos principales en Brasil, mostrando su adaptabilidad a diversas condiciones agrícolas [7].

Aunque las imágenes satelitales suelen tener limitaciones en su resolución espacial, su resolución espectral superior permite el uso de algoritmos *per-píxel*, mejorando el rendimiento en comparación con enfoques tradicionales [5]. Métodos como las redes neuronales convolucionales (CNNs), aplicados a datos temporales y secuenciales, han abordado desafíos como la cobertura de nubes. En este contexto, las imágenes SAR de satélites como Sentinel-1 han mostrado mejores resultados al penetrar nubes en comparación con imágenes ópticas.

Para estudios sobre el uso del suelo, la combinación de datos de Sentinel-1, Sentinel-2 y Landsat-8 ha demostrado mejorar significativamente el rendimiento en comparación con el uso de una sola fuente.

B. Casos de Estudio en Agricultura con Deep Learning

A continuación, se presentan diversos casos de estudio que aplican técnicas de Deep Learning para clasificar diferentes tipos de cultivos. La Tabla I resume los principales hallazgos de cada caso.

- *Coffee-Yield Estimation Using High-Resolution Time-Series Satellite Images and Machine Learning:* En este estudio [6], se utilizó aprendizaje automático y datos de

imágenes de alta resolución (PlanetScope) para predecir el rendimiento del café en Minas Gerais, Brasil. Se implementaron modelos Random Forest y regresión lineal múltiple, logrando un coeficiente de determinación (R^2) de 0.93 para la validación, destacando la precisión en el monitoreo de la variabilidad espacial.

- *Corn Biomass Estimation by Integrating Remote Sensing and Long-Term Observation Data Based on Machine Learning Techniques:* En [8] se estimó la biomasa del maíz en la cuenca del río Heihe, China, utilizando datos MODIS y modelos como Random Forest, SVM y XG-Boost. El modelo XGBoost obtuvo el mejor rendimiento ($R^2 = 0.78$, RMSE = 2.86 t/ha).
- *Estimating Crop Yields with Remote Sensing and Deep Learning:* [7] propusieron un modelo basado en Deep Learning para predecir rendimientos de cultivos en Brasil. Este modelo utilizó coordenadas geográficas, datos meteorológicos y del suelo, mostrando alta escalabilidad con datos limitados.
- *Forecasting Crop Yield with Deep Learning-Based Ensemble Model:* [9] emplearon un modelo basado en LSTM y ConvLSTM para predecir el rendimiento de cultivos en Estados Unidos e India. Con un coeficiente R^2 de 0.73 para la soja y 0.49 para el arroz, destacaron la eficiencia del modelo con menos parámetros.
- *Forecasting Sunflower Grain Yield Using Remote Sensing Data:* [10] utilizaron datos de teledetección en Francia para predecir el rendimiento del girasol, combinando índices de vegetación como GAI¹ y GAD² en modelos como Random Forest y regresiones polinómicas, logrando mejorar las predicciones a nivel de campo.
- *High-Resolution Global Map of Smallholder and Industrial Closed-Canopy Oil Palm Plantations:* [4] mapearon plantaciones de palma aceitera usando modelos avanzados de aprendizaje profundo diseñados para segmentación semántica aplicado a imágenes Sentinel, obteniendo una precisión superior 98%.
- *A Systematic Review of the Use of Deep Learning in Satellite Imagery for Agriculture:* Se destaca [5] donde se revisaron 150 estudios de Deep Learning aplicados a imágenes satelitales, identificando tareas clave como clasificación del uso del suelo, fisiología de plantas y predicción de rendimientos.

C. Tendencias Futuras y Potencial del Deep Learning en Agricultura

El Deep Learning ha transformado significativamente la agricultura, mejorando la precisión en la estimación de rendimientos y biomasa de cultivos como café, maíz y palma de aceite. La integración de algoritmos avanzados, como CNN y RNN, con datos de teledetección ha facilitado predicciones oportunas y eficientes para la gestión de recursos agrícolas.

El uso de imágenes satelitales se ha consolidado como una herramienta clave, permitiendo monitorear la salud del suelo,

¹Green Area Index

²Growth Accumulation Dynamics

TABLE I
RESUMEN DE CASOS DE ESTUDIO EN AGRICULTURA CON DEEP LEARNING

| Referencia | Ubicación | Métodos y Modelos | Resultados Clave |
|------------|-----------------------------|--|---|
| [6] | Minas Gerais, Brasil | Random Forest, Regresión Lineal Múltiple, imágenes PlanetScope | $R^2 = 0.93$, alta precisión en variabilidad espacial |
| [8] | Cuenca del río Heihe, China | MODIS, Random Forest, SVM, XG-Boost | XGBoost: $R^2 = 0.78$, RMSE = 2.86 t/ha |
| [7] | Brasil | Coordinadas geográficas, datos meteorológicos y del suelo | Modelo escalable con datos limitados |
| [9] | EE.UU. e India | LSTM, ConvLSTM, datos MODIS y NASA-USDA | Soja: $R^2 = 0.73$, Arroz: $R^2 = 0.49$ |
| [10] | Francia | GAI, GAD, Random Forest, regresiones polinómicas | Mejoras significativas en predicciones de rendimiento |
| [4] | Global | DeepLabv3+, imágenes Sentinel-1 y Sentinel-2 | Precisión del 98.52% en clasificación de palma aceitera |
| [5] | Global | Revisión de 150 estudios de Deep Learning aplicados a imágenes satelitales | Identifica tareas clave: uso del suelo, fisiología de plantas, predicción de rendimientos |

la fisiología de las plantas y evaluar daños en cultivos mediante índices espectrales. La comunidad científica destaca la necesidad de benchmarks públicos y métodos estandarizados para validar modelos y comparar resultados.

Para superar desafíos como la baja resolución espacial y las interferencias atmosféricas, se están desarrollando técnicas que maximizan el uso de los datos disponibles. Las innovaciones en Deep Learning prometen aplicaciones prácticas como la detección temprana de enfermedades, la optimización del uso del agua y la gestión de nutrientes, contribuyendo así a la sostenibilidad agrícola [11].

La colaboración interdisciplinaria entre agrónomos, científicos de datos y especialistas en teledetección es esencial para desarrollar soluciones integradas que impulsen la innovación en este campo en constante evolución [12].

D. Ventajas y Desafíos

El uso combinado de imágenes de Sentinel-1 (radar SAR) y Sentinel-2 (multi-espectrales) mejora significativamente la clasificación de cultivos, superando limitaciones como la cobertura de nubes. Sin embargo, persisten desafíos relacionados con la resolución espacial y la necesidad de benchmarks públicos que faciliten la comparación entre estudios.

III. PRELIMINARES

A. Convolución Atrous

La convolución atrous o dilatada es una variante de la convolución tradicional en redes neuronales convolucionales (CNNs) que inserta espacios entre los elementos del kernel, ampliando así el campo receptivo sin incrementar los parámetros [13]. Esto permite capturar características a mayor escala, siendo útil para tareas como la segmentación semántica, donde se requiere una visión global de la entrada. Además, la convolución atrous facilita una extracción de características más densa, especialmente en las capas profundas, sin comprometer la resolución espacial. Esto se logra ajustando la tasa de dilatación, lo que la hace ideal para aplicaciones que requieren un alto nivel de detalle, como la segmentación de imágenes [14].

B. Arquitectura Encoder-Decoder

La arquitectura de codificador-decodificador es una estructura fundamental en redes neuronales, utilizada principalmente en aplicaciones como la segmentación de imágenes. Esta arquitectura consta de dos componentes principales que trabajan de manera complementaria: el *codificador (encoder)* y el *decodificador (decoder)*.

Codificador (Encoder):

- 1) *Procesamiento de Entrada:* El codificador recibe datos de alta dimensión, como una imagen compleja, que necesitan ser analizados y procesados. Su objetivo es transformar esta entrada en un formato más manejable y significativo.
- 2) *Extracción de Características:* Utiliza múltiples capas computacionales para identificar patrones clave. Las convoluciones ayudan a detectar características como bordes, texturas o formas, mientras que las operaciones de agrupamiento (pooling) reducen las dimensiones de los datos, conservando información relevante y minimizando el ruido.
- 3) *Reducción de Dimensionalidad:* La información procesada se comprime en un conjunto de características de baja dimensión que representan los aspectos más significativos de los datos. Este paso es crucial para simplificar el problema y facilitar la reconstrucción posterior en el decodificador.

Decodificador (Decoder)

- 1) *Reconstrucción de Datos:* A partir de las características de baja dimensión generadas por el codificador, el decodificador reconstruye la salida deseada. En segmentación de imágenes, esto implica generar una representación visual segmentada y detallada de la entrada.
- 2) *Proceso Inverso al Codificador:* Mientras el codificador reduce las dimensiones de los datos, el decodificador realiza la operación inversa. Emplea técnicas como "up-sampling" (ampliación de datos) o convoluciones transpuestas para incrementar gradualmente la dimensionalidad y recuperar el formato original de la entrada.

- 3) *Generación de Resultados Finales:* El decodificador produce la salida final. En el caso de segmentación de imágenes, genera una imagen con regiones claramente segmentadas y clasificadas según las características aprendidas durante el proceso.

Importancia de la Arquitectura: El codificador se centra en extraer información relevante y comprimirla, mientras que el decodificador la utiliza para reconstruir una salida interpretativa. Este diseño permite procesar datos complejos de manera eficiente, manteniendo un equilibrio entre la reducción de dimensionalidad y la reconstrucción precisa, lo que lo convierte en una herramienta poderosa en tareas de procesamiento de datos avanzados, como la segmentación, traducción automática y análisis de series temporales.

C. Modelos pre-entrenados

1) *Backbone:* En Deep Learning, un backbone es una red neuronal pre-entrenada utilizada para la extracción de características en tareas como visión por computadora y otros modelos de aprendizaje profundo. Forma parte del codificador (encoder) en arquitecturas encoder-decoder, donde su función principal es comprimir los datos de entrada en una representación de menor dimensión que conserva la información esencial. Posteriormente, el decodificador (decoder) utiliza esta información para realizar tareas específicas, como generación de imágenes, segmentación semántica, o detección de objetos.

Ejemplos comunes de backbones son redes convolucionales pre-entrenadas como VGG, ResNet o MobileNet, que identifican patrones visuales complejos (bordes, texturas, formas) en imágenes. Estas características son clave para tareas de clasificación o segmentación, ahorrando tiempo y recursos al aprovechar modelos ya optimizados.

2) *MobileNet:* MobileNet [15] 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 [16].

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

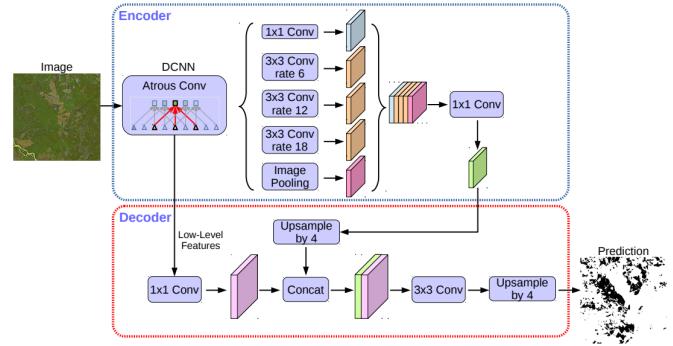


Fig. 1. Arquitectura DeepLabV3+ (Adaptado de [18])

hasta dispositivos IoT con restricciones severas de recursos [16]. Además, las iteraciones posteriores de MobileNet han introducido mejoras y refinamientos adicionales. Por ejemplo, MobileNetV2 introduce bloques residuales con inversiones 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.

3) *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 [17]. Uno de los componentes clave de DeepLab [13] es la convolución atriosa. 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 atriosa, DeepLab incorpora Atrous Spatial Pyramid Pooling (ASPP). ASPP utiliza convoluciones atriosas 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 [17]. 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 segmentación precisa. La Figura 1 muestra una ilustración de la arquitectura de DeepLabV3+ [18].

Como se mencionó previamente la arquitectura se basa en un codificador poderoso que utiliza convolución atriosa, 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.

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 [18].

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.

IV. 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.

A. 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.

B. Construcción del conjunto de datos

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.

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 x 1000 píxeles. Este tamaño corresponde a un área de 10 x 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.

3) *Aumento de Datos:* El proceso de aumento 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.

C. 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.

D. Validación del modelo

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.

E. 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 QGIS [19]. Esta herramienta de Sistemas de Información Geográfica (SIG) permitió una inspección detallada y minuciosa de las 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.

V. 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. El modelo alcanzó una precisión global del 98.37%, superando métodos tradicionales como SVM y Random Forest (precisión del 96% según Díaz, 2023). Esto implica que el modelo es capaz de clasificar correctamente el mas del 98% 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.

En la Figura 2 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.

En la Figura 3, 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.

En la Figura 4 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.

En la Figura 5 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.

En el gráfico que se muestra en la Figura 6, se muestra la pérdida de entrenamiento y validación. 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. De manera similar, en el gráfico 7, 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.

La figura Figura 8 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%



Fig. 2. Ejemplo de predicción del modelo sobre un área con cultivos

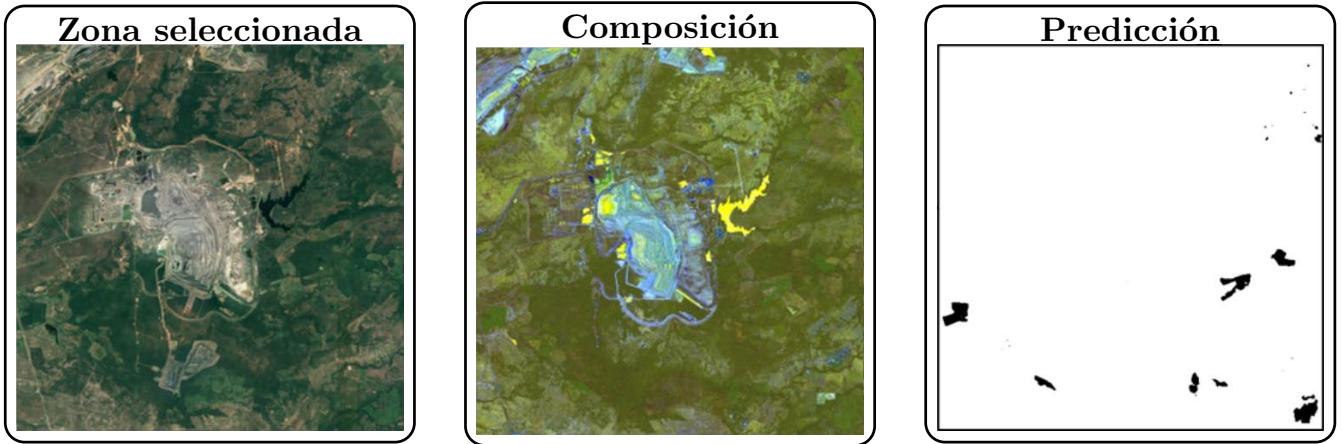


Fig. 3. Ejemplo de predicción del modelo sobre un área con pocos cultivos

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 [20] 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.

VI. CONCLUSIONES Y TRABAJOS FUTUROS

Este estudio demuestra la eficacia del uso de técnicas avanzadas de Deep Learning para la clasificación de cultivos de palma de aceite en Colombia, integrando datos satelitales de Sentinel-1 y Sentinel-2. El modelo propuesto, basado en DeepLabV3+ y MobileNet V3, alcanzó una precisión global del 98.37%, superando significativamente los métodos tradicionales. Estos resultados destacan la capacidad del enfoque para manejar desafíos complejos como la integración de datos radar y ópticos, y su robustez ante condiciones como la cobertura de nubes.

Sin embargo, se identificaron limitaciones, particularmente en la detección de áreas recientemente despejadas y plantaciones jóvenes. Esto sugiere la necesidad de incorporar datos adicionales que reflejen mejor la variabilidad en las etapas

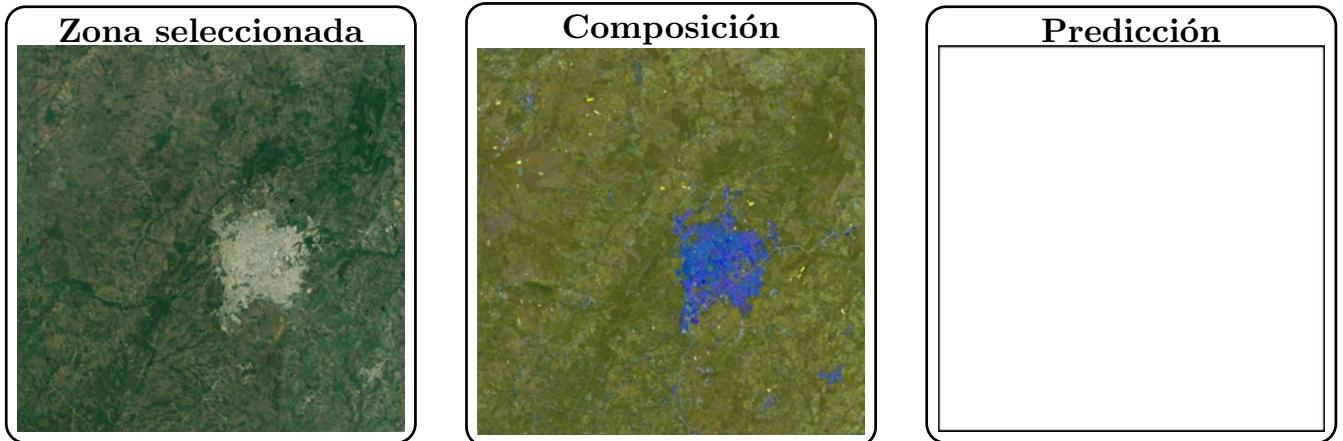


Fig. 4. Ejemplo de predicción del modelo sobre área sin cultivos

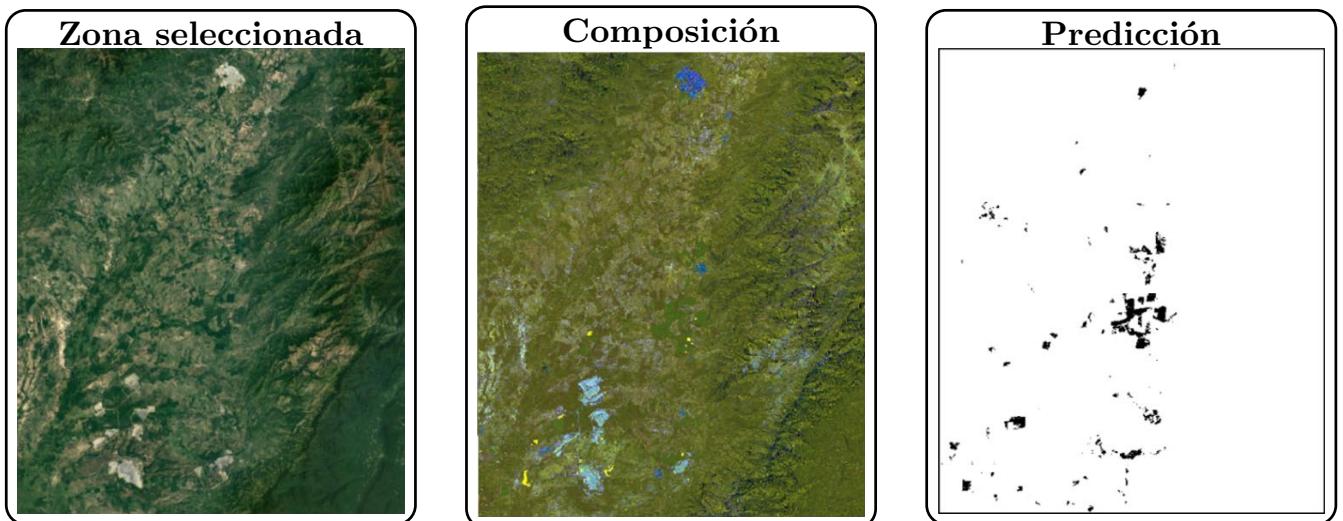


Fig. 5. Predicción del modelo sobre toda la zona seleccionada

de desarrollo de las plantaciones. Además, técnicas como la rotación de imágenes han demostrado ser esenciales para enriquecer los datos de entrenamiento y mejorar la eficacia del modelo, especialmente en regiones con conjuntos de datos limitados.

En comparación con enfoques tradicionales, como los basados en índices de vegetación y algoritmos de aprendizaje automático, el modelo de Deep Learning exhibe una superioridad notable en tareas específicas de clasificación y segmentación de imágenes agrícolas. Este trabajo subraya el potencial de estas herramientas para transformar la agricultura de precisión, no solo en Colombia, sino también en otros contextos globales.

Para futuros estudios, se recomienda ampliar el conjunto de datos, incluir más variedades de cultivos y explorar análisis temporales utilizando series de imágenes satelitales. Asimismo, sería valioso investigar cómo las condiciones

climáticas y estacionales afectan la precisión del modelo, además de priorizar la interpretabilidad de los resultados para fomentar su adopción en aplicaciones del mundo real. Estas iniciativas fortalecerán aún más el impacto de las tecnologías de Deep Learning en la sostenibilidad y gestión agrícola.

REFERENCES

- [1] A. W. Thayer, A. Vargas, A. A. Castellanos, C. W. Lafon, B. A. McCarl, D. L. Roelke, K. O. Winemiller, and T. E. Lacher, “Integrating agriculture and ecosystems to find suitable adaptations to climate change,” *Climate*, vol. 8, no. 1, p. 10, 2020.
- [2] C. Boryan, Z. Yang, R. Mueller, and M. Craig, “Monitoring us agriculture: the us department of agriculture, national agricultural statistics service, cropland data layer program,” *International Journal of Remote Sensing*, vol. 26, no. 5, pp. 341–358, 2011.
- [3] K. Hu, Z. Wang, G. Coleman, A. Bender, T. Yao, S. Zeng, D. Song, A. Schumann, and M. Walsh, “Deep learning techniques for in-crop weed identification: A review,” 2021.

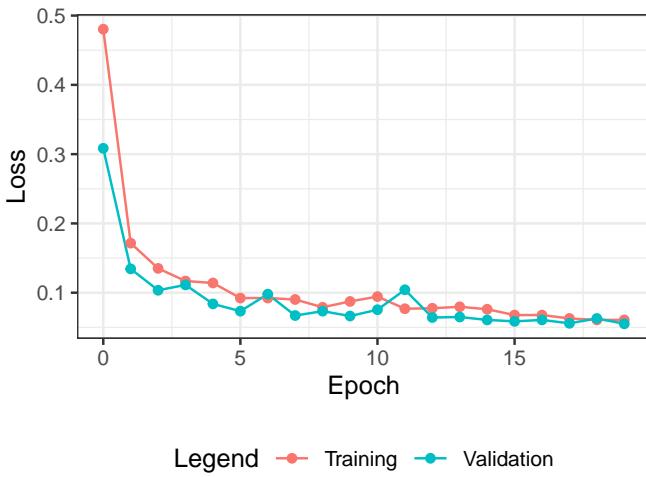


Fig. 6. Gráfica de perdida (loss)

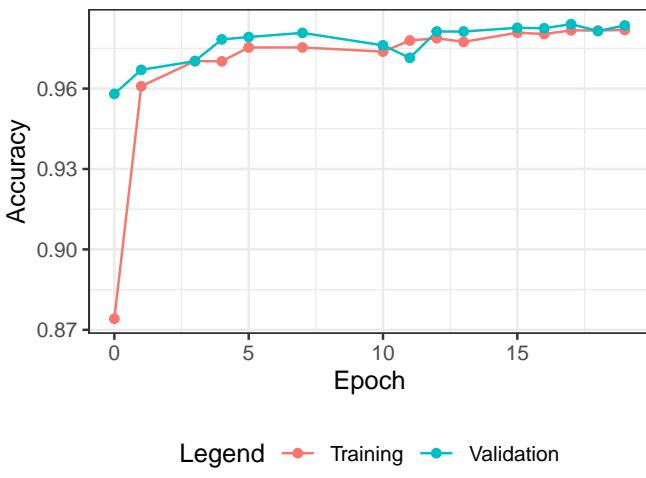


Fig. 7. Gráfica de exactitud (accuracy)

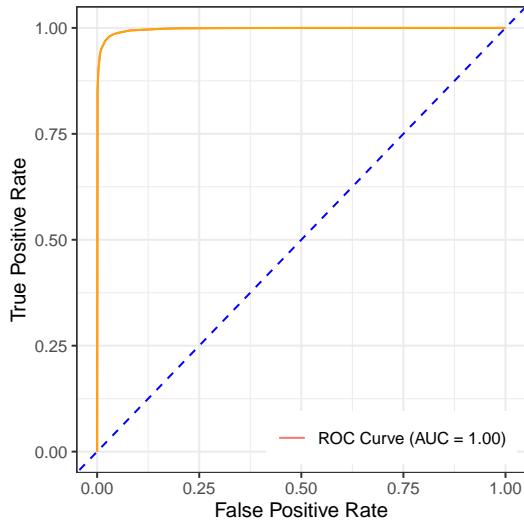


Fig. 8. Curva ROC del modelo

- [4] A. Descals, S. Wich, E. Meijaard, D. L. A. Gaveau, S. Peedell, and Z. Szantoi, "High-resolution global map of smallholder and industrial closed-canopy oil palm plantations," *Earth System Science Data*, vol. 13, no. 3, pp. 1211–1231, 2021.
- [5] B. Victor, Z. He, and A. Nibali, "A systematic review of the use of deep learning in satellite imagery for agriculture," vol. 12, pp. 345–367, 2022. [Online]. Available: <https://arxiv.org/abs/2210.01272v2>
- [6] M. Martello, J. P. Molin, M. C. F. Wei, R. Canal Filho, and J. V. M. Nicoletti, "Coffee-yield estimation using high-resolution time-series satellite images and machine learning," *AgriEngineering*, vol. 4, no. 4, pp. 888–902, 2022.
- [7] R. L. F. Cunha and B. Silva, "Estimating crop yields with remote sensing and deep learning," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-3/W2-2020, 2020, pp. 59–64.
- [8] L. Geng, T. Che, M. Ma, J. Tan, and H. Wang, "Corn biomass estimation by integrating remote sensing and long-term observation data based on machine learning techniques," *Remote Sensing*, vol. 13, no. 12, 2021.
- [9] M. S. Divakar, M. S. Elayidom, and R. Rajesh, "Forecasting crop yield with deep learning based ensemble model," *Materials Today: Proceedings*, vol. 58, pp. 256–259, 2022.
- [10] P. Debaeke, F. Attia, L. Champolivier, J. F. Dejoux, A. Micheneau, A. Bitar, and R. Trépos, "Forecasting sunflower grain yield using remote sensing data and statistical models," *European Journal of Agronomy*, vol. 142, 2023.
- [11] Y. Sepúlveda Casadiego, "Machine learning in agriculture, review and applications. machine learning en la agricultura, revisión y aplicaciones," 09 2020.
- [12] NASA Applied Sciences, "ARSET - Crop Mapping using Synthetic Aperture Radar (SAR) and Optical Remote Sensing. NASA Applied Remote Sensing Training Program (ARSET)," 2023. [Online]. Available: <http://appliedsciences.nasa.gov/get-involved/training/english/arset-crop-mapping-using-synthetic-aperture-radar-sar-and-optical>
- [13] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," 2017. [Online]. Available: <https://arxiv.org/abs/1606.00915>
- [14] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2016. [Online]. Available: <http://arxiv.org/abs/1610.02357>
- [15] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017. [Online]. Available: <https://arxiv.org/abs/1704.04861>
- [16] O. Elharrouss, Y. Akbari, N. Almaadeed, and S. Al-Maadeed, "Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches," 2022. [Online]. Available: <https://arxiv.org/abs/2206.08016v1>
- [17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," 2016. [Online]. Available: <https://arxiv.org/abs/1606.00915>
- [18] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," 2018. [Online]. Available: <https://arxiv.org/abs/1802.02611>
- [19] QGIS Development Team, *QGIS Geographic Information System*, <https://www.qgis.org>, QGIS Association, 2024.
- [20] N. Diaz, "Herramienta para la estimación del potencial bioenergético en municipios vulnerables de colombia mediante imágenes satelitales y machine learning," 2023.