

ESCUELA SUPERIOR POLITECNICA DEL LITORAL
MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

Clasificación de Cultivos en Ecuador Utilizando Técnicas de Aprendizaje

Profundo con escasos datos

NOMBRE DEL TUTOR:

José Córdova-García, Ph. D.

EVALUADORES:

Colon Enrique Pelaez Jarrin, Ph. D.

Ing. JOCELYN MARIE LUNA GONZÁLEZ

24 de julio de 2025

espol

Dedicatoria

A mis padres y familia, quienes con su sacrificio y sabiduría me enseñaron que el conocimiento es la herramienta más poderosa para transformar el mundo.

A mi pareja, por su paciencia infinita durante las largas noches de programación y análisis, por creer en mis sueños cuando yo dudaba, y por ser mi compañero en esta travesía académica.

Y finalmente, a todos los investigadores que trabajan incansablemente para que la tecnología sea una herramienta de progreso social y desarrollo sostenible.

Declaración Expresa

Yo, Jocelyn Marie Luna González, acuerdo y reconozco que: La titularidad de los derechos patrimoniales de autor (derechos de autor) del proyecto de graduación corresponderá al autor o autores, sin perjuicio de lo cual la ESPOL recibe en este acto una licencia gratuita de plazo indefinido para el uso no comercial y comercial de la obra con facultad de sublicenciar, incluyendo la autorización para su divulgación, así como para la creación y uso de obras derivadas. En el caso de usos comerciales se respetará el porcentaje de participación en beneficios que corresponda a favor del autor o autores. El o los estudiantes deberán procurar, en cualquier caso de cesión de sus derechos patrimoniales, incluir una cláusula en la cesión que proteja la vigencia de la licencia aquí concedida a la ESPOL.

La titularidad total y exclusiva sobre los derechos patrimoniales de patente de invención, modelo de utilidad, diseño industrial, secreto industrial, secreto empresarial, derechos patrimoniales de autor sobre software o información no divulgada que corresponda o pueda corresponder respecto de cualquier investigación, desarrollo tecnológico o invención realizada por mí durante el desarrollo del proyecto de graduación, pertenecerán de forma total, exclusiva e indivisible a la ESPOL, sin perjuicio del porcentaje que me corresponda de los beneficios económicos que la ESPOL reciba por la explotación de mi innovación, de ser el caso.

En los casos donde la Oficina de Transferencia de Resultados de Investigación (OTRI) de la ESPOL comunique al autor que existe una innovación potencialmente patentable sobre los resultados del proyecto de graduación, no se realizará publicación o divulgación alguna, sin la autorización expresa y previa de la ESPOL.

Guayaquil, 24 de julio de 2025.

Jocelyn Marie Luna González

Comité Evaluador

José Córdova-García, Ph. D.

Tutor de Proyecto

Colón Enrique Peláez Jarrín, Ph. D.

Profesor de Maestría

Reconocimiento

Deseo expresar mi reconocimiento a las siguientes personas que hicieron posible la realización de esta investigación:

Al Dr. José Córdova, director de esta tesis, por su orientación académica, valiosas sugerencias metodológicas y apoyo constante durante el desarrollo de este trabajo de investigación.

Al Dr. Enrique Peláez, revisor de esta tesis, por su tiempo, expertise y observaciones constructivas que mejoraron la calidad científica de este estudio.

A la comunidad científica internacional por hacer disponibles de forma pública los conjuntos de datos y metodologías que sirvieron como base para esta investigación.

Resumen

En Ecuador, la escasez de datos etiquetados sobre cultivos agrícolas representa una barrera significativa para la planificación territorial, la gestión de recursos y el desarrollo de políticas públicas a largo plazo. Esta falta de información dificulta la generación de mapas confiables de uso del suelo, especialmente en un contexto donde se requiere conocer la distribución y evolución de los principales cultivos del país. Esta tesis aborda dicho problema mediante la evaluación de distintas estrategias de clasificación de cultivos a partir de imágenes satelitales, considerando escenarios con datos limitados. Se comparan tres enfoques: modelos tradicionales (como Random Forest), redes neuronales convolucionales (CNN) y el modelo Transformador de teledetección preentrenado (Presto por sus siglas en inglés), recientemente propuesto en la literatura. Se utilizan diferentes combinaciones de entradas espectrales (RGB, 10 bandas e índices) y tipos de clasificación (binaria y multiclase), evaluando su desempeño mediante métricas como el F1-score macro. Los resultados muestran que, en contextos de escasez de datos, los modelos tradicionales siguen siendo una alternativa sólida, mientras que las CNN aportan mejoras moderadas. El modelo Presto ofrece un balance prometedor entre precisión y eficiencia computacional, generando mapas útiles para el monitoreo agrícola. Este trabajo demuestra que, incluso en países con recursos limitados, es posible adaptar estrategias de aprendizaje automático para mejorar la representación espacial de cultivos y apoyar la toma de decisiones en el sector agrícola.

Abstract

In Ecuador, the lack of labeled agricultural data is a major obstacle to long-term planning, land management, and the development of evidence-based public policies. This data scarcity hinders the generation of reliable land use maps, especially when aiming to understand the spatial distribution and temporal dynamics of key crops. This thesis tackles this challenge by evaluating different crop classification strategies using satellite imagery, with a focus on low-data scenarios. Three main approaches are compared: traditional models (e.g., Random Forest), Convolutional Neural Networks (CNNs), and Presto, a recent model in the literature. Various combinations of spectral inputs (RGB, 10 bands, and indices) and classification types (binary and multiclass) are tested and evaluated using metrics such as the macro F1-score. The results show that traditional models remain robust in low-resource settings, while CNNs offer modest improvements. Presto demonstrates a promising trade-off between accuracy and computational efficiency, enabling the generation of informative maps for agricultural monitoring. This work highlights that even in data-constrained environments, it is possible to adapt machine learning methods to improve crop mapping and support decision-making in the agricultural sector.

Índice general

Resumen	vi
Abstract	vii
Índice de figuras	x
Índice de Tablas	xi
Lista de abreviaciones	xii
1 Introducción	1
1.1 Antecedentes	2
1.2 Justificación	2
1.3 Objetivos	2
1.3.1 Objetivo general	2
1.3.2 Objetivos específicos	2
2 Datos	3
2.1 Datasets	5
2.1.1 Datos Satelitales	5
2.1.2 Índices Espectrales	5
2.1.3 Datos Etiquetados	7
2.2 Datos y Tareas de Teledetección para la Agricultura	7
2.3 Limitaciones de Adquisición de Datos	8
2.4 Datos Personalizados para Clasificación de Cultivos	9
3 Técnicas de Aprendizaje Profundo	10
3.1 Clasificación Mediante Aprendizaje por Transferencia	15
3.2 LCLU en Ecuador	17
3.3 Criterios de Evaluación y Optimización en DL	18
3.3.1 Función de Pérdida	18
3.3.2 Métricas	20
3.4 Preprocesamiento de imágenes satelitales	21

3.4.1	Manejo de nubes	21
3.4.2	Normalización espectral	22
4	Metodología	23
4.1	Selección de Datos	24
4.2	Procesamientos de Datos	26
4.2.1	Manejo de nubes	26
4.2.2	Preparación y formateo de datos	27
4.2.3	División de datos	28
4.2.4	Obtención de índices espectrales	29
4.2.5	Escalado de datos	30
4.2.6	Representación de máscaras	32
4.3	Caso de Estudios	33
4.3.1	Modelos de Clasificación	34
4.3.2	Características de los datos	34
4.3.3	Tipos de Clasificación	35
4.4	Estrategias de entrenamiento	36
4.4.1	Transferencia de aprendizaje	36
4.4.2	Ajuste de Hiperparámetros	36
4.5	Ambiente de Implementación	37
5	Resultados	39
5.1	Selección de la función de pérdida	39
5.1.1	Clasificación binaria	39
5.1.2	Clasificación categórica	40
5.2	Configuración de modelos	42
5.3	Resultados de los experimentos	43
5.3.1	Modelos Tradicionales	43
5.3.2	Modelos basados en CNN	44
5.3.3	Modelos complejos específicos	48
5.4	Comparativa de modelos y limitaciones	50
6	Conclusión	53
6.1	Trabajo futuro	54
Referencias		55
Archivos adjuntos		59
Anexo A		60

Índice de figuras

2.1	Tipos de Datos Satelitales	4
2.2	Tipos de índices Espectrales	6
3.1	Mecanismo de algoritmo Random Forest	11
3.2	Mecanismo de una neurona	11
3.3	Capas de red neuronal	12
3.4	Arquitectura de red neuronal convolucional	13
3.5	Arquitectura encoder-decoder	13
3.6	Arquitectura U-net	14
3.7	Ejemplo de Trasferencia de aprendizaje usando dos conjuntos de datos	16
3.8	Arquitectura de Presto	17
3.9	Matriz de confusión	21
4.1	Diagrama de Flujo de Metodología	23
4.2	Ejemplo de etiquetas de cultivos del 2020	24
4.3	Zonas de datos de Ecuador	25
4.4	Imágenes del conjunto de datos Ghana	26
4.5	Procesamiento de datos de Ecuador	28
4.6	Comparación de etiquetas de cultivo	29
4.7	Diagrama de cajas de datos de Ghana	31
4.8	Diagrama de cajas de datos de Ecuador	32
4.9	Codificación de máscaras	33
5.1	Pérdida y F1-score de función de pérdida dice loss para clasificación binaria	40
5.2	Pérdida y F1-score de función de pérdida categorical crossentropy loss para clasificación categórica	41
5.3	Clasificación de cultivos usando el modelo Presto	50

Índice de Tablas

4.1	Cantidad de imágenes obtenidas por zona	27
4.2	Distribución de etiquetas en los conjuntos de entrenamiento, validación y prueba para cada escenario de clasificación.	30
4.3	Combinación de casos de estudio	36
5.1	Resultados de experimentos binarios de función de pérdida	40
5.2	Resultados de experimentos categóricos de función de pérdida	41
5.3	Configuración de modelos base	42
5.4	Resultados de escenarios de modelo Random Forest (accuracy)	43
5.5	Resultados de escenarios de modelo Random Forest (F1-score macro)	44
5.6	Comparación de F1-score por escenario de modelo Random Forest	45
5.7	Resultados de los escenarios con modelos U-net entrenados en Ecuador (F1-score macro)	46
5.8	Resultados de escenarios de modelos U-net preentrenados con datos de Ghana (F1-score macro)	47
5.9	Resultados de escenarios de modelo Presto (F1-score macro)	48
5.10	Comparación de F1-score por escenario de modelo Presto	49
5.11	Tiempo de entrenamiento por modelo (minutos)	52

Lista de abreviaciones

BSI Bare Soil Index.

CNN Convolutional Neural Network.

CRS Coordinates Reference System.

ESA European Space Agency.

GEE Google Earth Engine.

IoU Intersection over Union.

LCLU Land Cover/Land Use.

MAG Ministerio de Agricultura y Ganadería.

MAML Model-Agnostic Meta-Learning.

NDVI Normalized Difference Vegetation Index.

NDWI Normalized difference water index.

Presto Pretrained Remote Sensing Transformer.

ReLU Rectified Linear Unit.

RF Random Forest.

RS Remote Sensing.

SAVI Soil Adjusted Vegetation Index.

SMOTE Synthetic Minority Oversampling Technique.

SSL semi-supervised Learning.

TIML Task-Informed Meta-Learning for Agriculture.

TL Transfer Learning.

UAV Unmanned aerial vehicle.

1

Introducción

La productividad agrícola es vital para la seguridad alimentaria y la estabilidad económica, especialmente en países de Latinoamérica, donde el comercio depende de productos agrícolas como materia prima [1]. Por ello, una clasificación precisa de cultivos mejora significativamente la toma de decisiones en la agricultura, lo que conduce a una gestión más eficiente de recursos, aumento de rendimiento agrícola y beneficios económicos [2].

En Ecuador, la estimación de cultivos se realiza principalmente mediante métodos manuales como los empleados por el Ministerio de Agricultura y Ganadería (MAG) [3], que se basan en interpretaciones visuales de imágenes satelitales. Este enfoque, aunque valioso, requiere una considerable inversión de tiempo y recursos, generando apenas uno a dos mapas anuales con limitada capacidad de respuesta ante cambios estacionales o emergencias agrícolas.

Si bien técnicas de aprendizaje automático (ML) como Random Forest se usan globalmente para la clasificación de cultivos [4], su adopción en Ecuador enfrenta una limitante crítica: la escasez de datos etiquetados [2]. Estos datos son esenciales para entrenar modelos y, si son escasos, esto restringe su aplicación a zonas geográficas específicas y pequeñas, imposibilitando clasificaciones a nivel nacional. Este mismo desafío afecta a técnicas más avanzadas como el aprendizaje profundo (DL), que requiere volúmenes aún mayores de datos para alcanzar buenos rendimientos [5].

El DL emerge como solución para generar mapas generalizados y de alta frecuencia, mejorando prácticas agrícolas mediante optimización de recursos y predicción de rendimientos [5]. El DL permite incorporar estrategias como la transferencia de aprendizaje (TL) -reutilizar modelos preentrenados en regiones con datos abundantes- mitigando parcialmente la escasez de datos etiquetados. Sin embargo, persisten obstáculos clave: los datos del MAG, pese a su calidad, son limitados y se actualizan anualmente [2], mientras que la variabilidad en cultivos, suelos y climas exige modelos altamente adaptativos.

Otra alternativa reside en el uso directo de imágenes satelitales, pero aquí surge una limitante: aunque se tiene acceso a estas imágenes para el territorio ecuatoriano, la falta de grandes conjuntos etiquetados dificulta entrenar modelos robustos [6]. Además, la variabilidad en los tipos de cultivos, las condiciones ambientales y las prácticas agrícolas complica aún más la creación de modelos universales adecuados para diversos escenarios agrícolas [7].

En este trabajo se plantea una metodología para generar mapas de estimación de cultivos a nivel nacional con mayor periodicidad, haciendo uso de modelos de DL. Para abordar la problemática de la falta de datos etiquetados en el territorio ecuatoriano, se proponen diferentes enfoques, como utilizar datos de zonas geográficas similares a las ecuatorianas y el uso

1 Introducción

de modelos entrenados para tareas similares, con el fin de mejorar la generación de mapas de estimación agrícola.

1.1 Antecedentes

Actualmente, el MAG ha realizado un arduo trabajo para generar mapas anuales de Ecuador que contienen información sobre la cobertura y el uso de la tierra, así como los sistemas productivos agropecuarios a nivel nacional [3]. Estos mapas incluyen datos sobre el tipo de suelo, la zonificación de la tierra y la estimación de cultivos anuales. Sin embargo, la elaboración de estos productos es un proceso laborioso que puede tomar entre uno y dos años desde la recolección de datos hasta la publicación del mapa final, y requiere la intervención de un gran número de técnicos para la obtención, validación y registro de los diferentes usos de la tierra. Por esta razón, estos mapas solo se generan después de largos períodos, siendo el último de este tipo producido en 2021 [3].

La falta de actualización temporal de estos mapas los hace inefficientes para la toma de decisiones que requieren analizar la evolución de los cultivos agrícolas a lo largo del tiempo, incluyendo su crecimiento y las distintas cosechas que se producen anualmente.

1.2 Justificación

El proyecto busca resolver el problema de la escasez de datos etiquetados en Ecuador mediante la aplicación de técnicas avanzadas de aprendizaje profundo que permiten la clasificación precisa de cultivos con conjuntos de datos limitados. Utilizando enfoques como el TL y el aumento de datos, se puede mitigar la falta de datos y generar modelos robustos que mejoren la capacidad de decisión en la gestión agrícola. De esta manera, se espera contribuir a mejorar los rendimientos y la gestión de recursos en la agricultura ecuatoriana, optimizando prácticas locales a través de la innovación tecnológica.

1.3 Objetivos

1.3.1 Objetivo general

Desarrollar un modelo de clasificación de cultivos para Ecuador utilizando técnicas de aprendizaje profundo para la mejora de la toma de decisiones agrícolas que supere la limitante de grandes cantidades de datos etiquetados.

1.3.2 Objetivos específicos

- Investigar técnicas de aumento de datos y transferencia de aprendizaje que puedan aplicarse a conjuntos de datos limitados.
- Desarrollar un modelo de clasificación de cultivos adaptado a las condiciones locales de Ecuador.
- Evaluar la precisión y efectividad del modelo en diferentes escenarios agrícolas con datos heterogéneos.
- Desarrollar una aplicación que clasifique cultivos en Ecuador a partir de imágenes satelitales, utilizando modelos de aprendizaje profundo desarrollados.

2

Datos

En este capítulo, se describe a profundidad la problemática de los datos, estrategias y demás limitaciones que podemos encontrar al momento de proponer una metodología para la clasificación de cultivos y generación de mapas para tipos de cultivos.

La teledetección (Remote Sensing, RS) se utiliza ampliamente para la clasificación y monitoreo de cultivos, así como para el análisis del uso y la cobertura del suelo (Land Cover Land Use, LCLU) en general [8, 9]. Este campo se centra en la recolección, análisis y obtención de información a partir de imágenes capturadas por satélites, drones y otras plataformas aéreas, permitiendo observar y monitorear la superficie terrestre y su entorno [8]. Las imágenes espectrales y de alta resolución generadas por estos sistemas se emplean en la clasificación y categorización de diferentes coberturas y usos del suelo, abordando problemas en agricultura, gestión de recursos y observación ambiental.

Antes de profundizar en la problemática de clasificación de cultivos, es fundamental entender qué datos tenemos disponibles y qué se puede hacer con ellos. Al buscar datos que puedan ser útiles para resolver un problema, es necesario indagar en las diversas fuentes y formatos que podamos encontrar. En el caso de problemas relacionados con LCLU, se requieren dos grandes grupos de datos: información sobre el suelo y las etiquetas correspondientes.

La información del suelo se refiere a los datos visuales o sensoriales que se pueden registrar del área que se desea estudiar. Estos datos pueden incluir desde una imagen a color (o RGB) hasta información más compleja, como lecturas de radar, topografía, datos meteorológicos, entre otros.

La mayoría de estos datos suelen almacenarse en formato raster, es decir, en mapas de bits o matrices, donde cada celda (o píxel) representa un valor correspondiente a una ubicación geográfica específica. La dimensión de cada celda dependerá de la resolución espacial de la imagen, que puede variar desde varios kilómetros (baja resolución) hasta pocos metros o incluso centímetros en imágenes de alta resolución.

El formato de archivo TIFF es comúnmente utilizado para almacenar mapas de bits de imágenes satelitales, ya que permite almacenar información adicional como la resolución, el sistema de coordenadas y las dimensiones del archivo, necesarios para ubicar geográficamente la imagen. En estos archivos, es posible encontrar matrices de hasta cuatro dimensiones. Las dos primeras dimensiones corresponden a la latitud y longitud de los datos. La tercera dimensión depende de la cantidad de tipos de datos disponibles. Por ejemplo, al contar con una imagen RGB, podemos tener tres matrices diferentes, una para cada color. De esta manera, a medida que se agregan más capas de información de una imagen, estas pueden apilarse.

2 Datos

Estas matrices de información se conocen como “*bandas*”. La última dimensión corresponde a la temporalidad. Una vez que tenemos un bloque de datos para una zona (latitud y longitud) con diferentes características (bandas), también podemos almacenarlos en diferentes momentos en el tiempo. Esto permite organizar la información de manera estructurada sin necesidad de varios archivos.

El otro tipo de datos es la etiqueta. Estos datos son necesarios dependiendo del tipo de tarea que se desea realizar. Por ejemplo, la tarea de clasificación (desde la perspectiva del aprendizaje automático) requiere principalmente datos etiquetados. Dado que la necesidad de datos etiquetados varía según la tarea, este aspecto se discutirá en la sección 2.2. Una vez que tenemos los grupos de información, es fundamental definir qué queremos clasificar. Las etiquetas o clases dependerán del tipo de problema que se busque resolver, pudiendo abarcar desde la clasificación de tipos de cultivos hasta la identificación de tipos de suelo, entre otros.

Las zonas etiquetadas suelen estar delimitadas por polígonos, aunque también pueden representarse mediante puntos o líneas, dependiendo de la entidad geográfica que se desea estudiar. Un formato comúnmente utilizado para almacenar estos datos geoespaciales es el shapefile [10], en el cual se registran las coordenadas de uno o varios puntos (ubicaciones específicas), líneas (trayectorias o límites lineales) o polígonos (áreas cerradas que representan superficies). La complejidad y tamaño de estas formas dependen de la cantidad de coordenadas que contengan.

Una opción para mapear estas etiquetas en forma de polígonos con los datos de mapas de bits es mediante máscaras. Una máscara es simplemente una matriz con la misma dimensión en latitud y longitud que el mapa de bits. Al utilizar máscaras, los polígonos se transforman en una matriz de bits donde se asignan valores numéricos a las etiquetas correspondientes. Si se trabaja con una sola clase (por ejemplo, identificar si una zona pertenece o no a una categoría específica), se emplean máscaras binarias, donde el valor 1 indica la presencia de la clase y el valor 0 su ausencia. En caso de querer identificar múltiples clases, se utilizan máscaras multiclase, asignando un número distinto a cada clase.

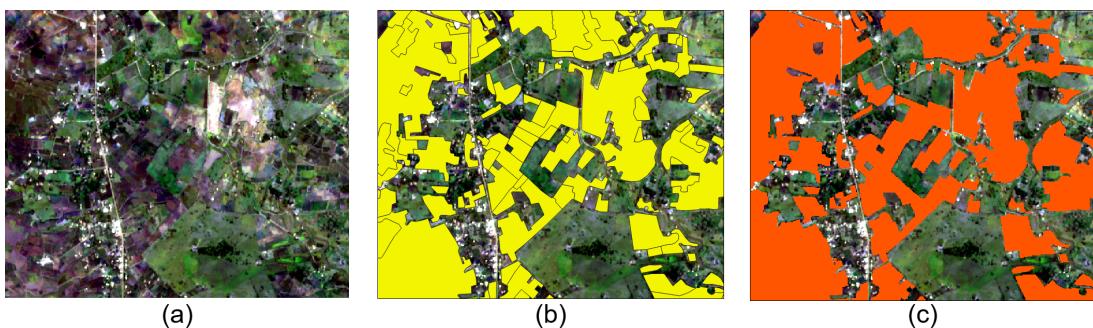


Figura 2.1: Tipos de Datos Satelitales: (a) Imagen RGB satelital, (b) etiquetas en forma de polígonos, (c) etiquetas en forma de mapa de bits.

En la Figura 2.1 se pueden observar los diferentes tipos de formatos y la información obtenida de cada uno. La Fig. 2.1.a muestra cómo una imagen puede representarse en mapas de bits utilizando las bandas RGB (rojo, verde y azul). La Fig. 2.1.b es un ejemplo de una clase de polígonos, en este caso, correspondiente a cultivos de maíz en el territorio ecuatoriano. Por último, la Fig. 2.1.c presenta la máscara binaria generada tras procesar la etiqueta de cultivo de maíz como un mapa de bits.

En las siguientes secciones se presentan ejemplos de datos y se analiza cómo la necesidad de cada tipo de dato varía en

2 Datos

función de la tarea a realizar y la disponibilidad de información.

2.1 Datasets

En esta sección se describen los diferentes tipos de datos utilizados en problemas relacionados con la teledetección, como datos satelitales, índices espectrales y datos etiquetados.

2.1.1 Datos Satelitales

Podemos encontrar diferentes tipos de datos satelitales. Los satélites tienen diferentes sensores para capturar diferentes frecuencias del espectro electromagnético. Dependiendo de la cantidad de sensores, frecuencia de recorrido y resolución del sensor, tendremos diferentes datos con diferentes resoluciones y temporalidad. Entre los diferentes datos abiertos satelitales tenemos el dataset Sentinel 1 y 2, imágenes satelitales obtenidas del programa Copernicus [11], impulsado por la Agencia Espacial Europea, ESA en sus siglas en inglés. El dataset Sentinel 1 se trata de datos obtenidos por radar, y Sentinel 2 tiene una colección de bandas multiespectrales. Ambos datasets han sido ampliamente usados en diferentes proyectos relacionados con RS. En este proyecto, nos centraremos en el dataset de Sentinel 2.

Los datos satelitales adquiridos fueron mejorados mediante un proceso de preprocesamiento, ya que las imágenes originales obtenidas directamente del satélite requieren correcciones antes de ser usadas. El dataset Sentinel 2-2A ha sido procesado mediante algoritmos de corrección atmosférica y procesos de corrección para reducir los errores en la recolección de datos desde el satélite [12].

El conjunto de datos cuenta con 13 bandas espectrales, incluyendo las bandas RGB. Entre las bandas comúnmente utilizadas se encuentran: azul (B2), verde (B3), rojo (B4), borde rojo 1 (B5), borde rojo 2 (B6), borde rojo 3 (B7), infrarrojo cercano NIR (B8), borde rojo 4 (B8A), infrarrojo de longitud de onda corta SWIR 1 (B11), y SWIR 2 (B12). Estas bandas pueden encontrarse en diferentes satélites y en diferentes resoluciones. En el caso de Sentinel 2-2A, la resolución de las bandas varía entre 10 m, 20 m y 60 m.

2.1.2 Índices Espectrales

Otra forma de obtener datos adicionales es mediante índices espectrales. Estos índices se utilizan para realzar características específicas de la superficie terrestre a partir de las bandas espectrales de las imágenes de satélite [13]. Los índices más comúnmente usados son los siguientes:

Índice de Vegetación de Diferencia Normalizada (NDVI), el cual resalta las zonas de vegetación usando la siguiente ecuación:

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \quad (2.1)$$

Donde *NIR* corresponde a la reflectancia en la banda del infrarrojo cercano (banda B8 de Sentinel-2) y *Red* a la reflectancia en la banda del rojo (banda B4).

2 Datos

Índice de Vegetación Ajustado al Suelo (SAVI), se utiliza para minimizar la influencia del suelo en áreas con baja densidad vegetal. Se calcula utilizando la siguiente ecuación:

$$SAVI = \frac{(NIR - Red) \times (1+L)}{(NIR + Red + L)} \quad (2.2)$$

Donde L es un factor de corrección, comúnmente establecido en $L = 0.5$.

Índice de Agua de Diferencia Normalizada (NDWI), el cual resalta los cuerpos de agua usando las bandas NIR (B8) y SWIR1 (B11), ya que la reflectancia del agua es baja en el infrarrojo cercano y alta en SWIR. Se calcula utilizando la siguiente ecuación:

$$NDWI = \frac{(NIR - SWIR1)}{(NIR + SWIR1)} \quad (2.3)$$

Índice de Suelo Desnudo (BSI), para identificar partes del suelo que se encuentren llanos o en construcción. De esta forma se pueden distinguir las áreas urbanas y suelos de lugares con mayor vegetación o cuerpos de agua.

$$BSI = \frac{(SWIR1 + Red) - (NIR + Blue)}{(SWIR1 + Red) + (NIR + Blue)} \quad (2.4)$$

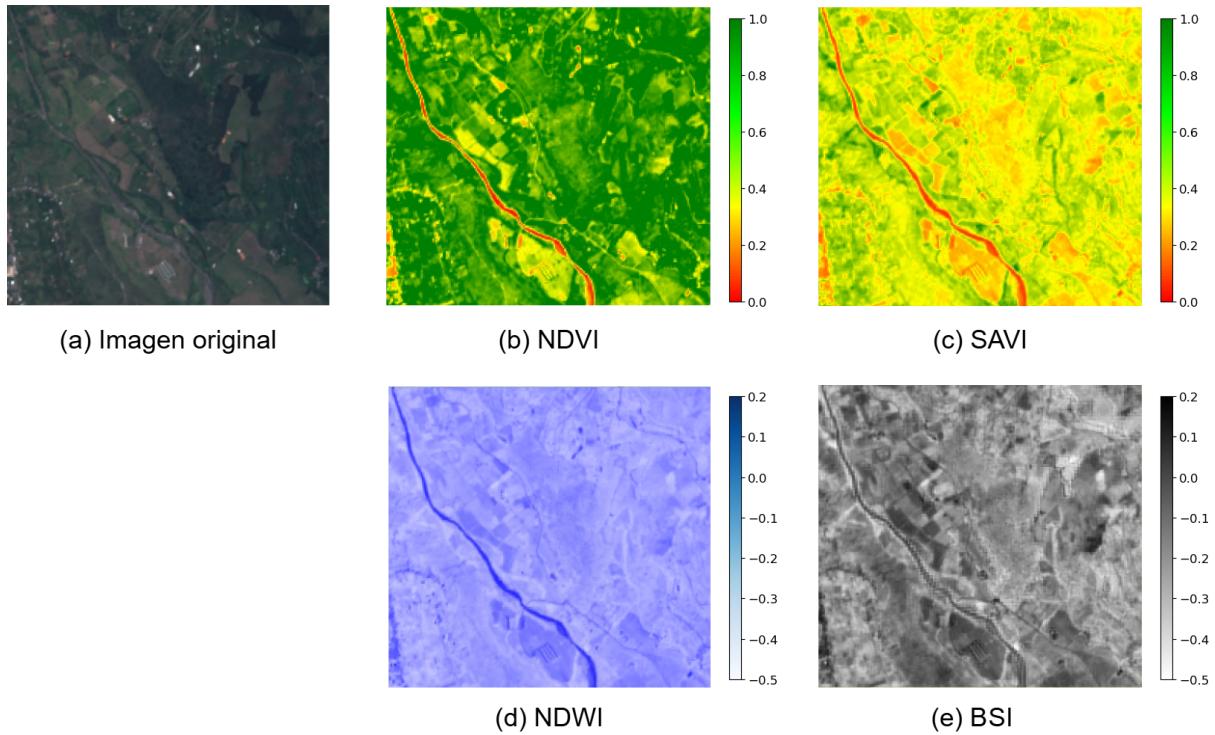


Figura 2.2: Tipos de índices Espectrales: (a) Imagen RGB satelital, (b) NDVI, (c) SAVI, (d) NDWI y (e) BSI.

En la Figura 2.2 se presentan los distintos índices mencionados anteriormente. La Fig. 2.2.a muestra la imagen satelital en RGB; las Figuras 2.2.b y 2.2.c ilustran índices de vegetación que realzan las áreas con cobertura vegetal y atenúan las zonas áridas y los cuerpos de agua. En este caso, el SAVI evidencia con mayor precisión las diferencias entre diversas densidades

2 Datos

de vegetación, mientras que el NDVI resalta de manera general la presencia de vegetación. La Fig. 2.2.d destaca los cuerpos de agua, y la Fig. 2.2.e pone de relieve el suelo desnudo, abarcando tanto áreas de cultivos recogidos como suelos áridos.

2.1.3 Datos Etiquetados

En relación a datos etiquetados, es fundamental definir el tipo de etiquetas necesarias para abordar el problema que se desea solucionar. En el ámbito agrícola, especialmente en regiones tropicales, podemos encontrar datasets de etiquetado de polígonos correspondientes a cultivos como banano, arroz y soya [3].

Existen diversos datasets etiquetados que facilitan el estudio y monitoreo de cultivos. Por ejemplo, en [14] se genera una colección de datos, tanto datos satelitales como etiquetados, de cultivos en África. Estos datos se obtuvieron específicamente en sectores de Sudán y Ghana, abarcando etiquetas de 23 tipos diferentes de cultivos. Asimismo, iniciativas como el Experimento Conjunto para Evaluación y Monitoreo de Cultivos [15], o en sus siglas en inglés JECAM, proporcionan colecciones de polígonos etiquetados de varios países, incluyendo países de América Latina como Brasil y Argentina.

Estos datasets suelen centrarse en proporcionar etiquetas o máscaras, requiriendo la recolección e integración con datos adicionales, con imágenes satelitales, con información detallada del suelo. Programas como Sentinel-2 ofrecen imágenes que complementan estos conjuntos de datos etiquetados.

Por otro lado, iniciativas como CropHarvest [6] han desarrollado conjuntos de datos que integran etiquetas de cultivos de diferentes fuentes, resultando en un total de 95,186 puntos de datos estandarizados. Cada punto de datos incluye etiquetas georreferenciadas considerando latitud, longitud, elevación del suelo, etiquetas de tipo de cultivo y temporalidad. El tipo de etiqueta varía según el país, incluyendo etiquetas binarias (cultivo y no cultivo) en países como Kenia y Brasil, así como conjuntos de datos multiclasificadas con etiquetas para cultivos específicos —como maíz— y características del tipo de suelo, por ejemplo, pastizales y cuerpos de agua.

Además, alrededor del 74% de estas etiquetas están emparejadas con datos de teledetección y datos meteorológicos, incluyendo datos satelitales como Sentinel-2 (imágenes ópticas multiespectrales), Sentinel-1 (radar), el Modelo Digital de Elevación SRTM y datos climáticos ERA5. Estos conjuntos de datos reducen el tiempo de procesamiento al integrar diferentes tipos de datos, permitiendo enfocarse en el desarrollo de modelos especializados para diversas regiones geográficas.

2.2 Datos y Tareas de Teledetección para la Agricultura

En el ámbito de la teledetección, se distinguen tres grandes tareas: detección, clasificación y segmentación.

- **Detección:** se refiere a la identificación de la presencia y localización de características específicas, objetos o cambios en la superficie terrestre a lo largo del tiempo [16, 17].
- **Clasificación:** consiste en asignar etiquetas a imágenes para identificar diferentes tipos de coberturas, como áreas urbanas, vegetación, cuerpos de agua y cultivos [9].
- **Segmentación:** implica dividir las imágenes en regiones homogéneas de píxeles con características similares, agrupando áreas de interés [18]. Además, la segmentación también puede usarse como un paso previo a la asignación de

2 Datos

etiquetas a las áreas que agrupa cuando no se dispone de datos etiquetados o estos son insuficientes.

La problemática abordada en este proyecto se centra en la clasificación del uso y cobertura del suelo (LCLU), enfocada en etiquetas de agricultura. Para realizar una clasificación precisa en el ámbito agrícola, es esencial disponer de datos multispectrales con alta resolución temporal y espacial. La selección de estos datos depende del tipo de cultivo que se desea identificar, ya que cada uno posee diferentes patrones de crecimiento, con variaciones en los períodos de siembra, cosecha y rotación [19]. Por lo tanto, la información sobre los cultivos influye directamente en los requisitos temporales y espaciales de los datos. Además, la incorporación de índices espectrales, como el NDVI, y datos medidos en campo, como la calidad del suelo y la temperatura, enriquece el análisis y contribuye a una clasificación más precisa [20]. Por ello, la adecuada selección de datos y la frecuencia de su recolección son fundamentales para una clasificación eficaz.

La segmentación en teledetección de cultivos es fundamental para mapear y monitorear parcelas agrícolas a nivel de campo. Este proceso implica la agrupación de píxeles en imágenes de alta resolución, permitiendo distinguir claramente los límites entre diferentes grupos, en este caso, parcelas y tipos de cultivo [21]. A diferencia de la clasificación, la segmentación puede no requerir etiquetas, ya que su objetivo es agrupar características similares en la imagen. Posteriormente, estos grupos pueden ser clasificados utilizando un menor volumen de datos etiquetados en comparación con la clasificación directa, lo que optimiza el proceso y reduce la necesidad de grandes conjuntos de datos etiquetados [19].

Sin embargo, aunque la segmentación puede facilitar la reducción de datos etiquetados necesarios, no resuelve completamente el problema, ya que la calidad y representatividad de los datos de entrenamiento siguen siendo factores críticos. Además, ciertas técnicas de segmentación requieren imágenes previamente segmentadas para el entrenamiento de modelos avanzados, lo que puede limitar su aplicabilidad en escenarios con escasez de datos anotados. Por lo tanto, persisten desafíos importantes en la obtención y uso eficiente de datos para la segmentación y clasificación en teledetección agrícola.

2.3 Limitaciones de Adquisición de Datos

Los datos requeridos para un estudio específico dependen de la zona que se pretende clasificar. Estas zonas, denominadas “áreas de estudio”, pueden categorizarse según su extensión.

- **Estudios por sectores:** Estos se centran generalmente en asentamientos puntuales. Requieren imágenes de alta resolución, habitualmente obtenidas mediante drones o vehículos aéreos no tripulados (UAV, por sus siglas en inglés). Estos dispositivos permiten capturar imágenes detalladas a baja altitud [22].
- **Estudios por ciudad:** Estas clasificaciones establecen su área de estudio a una ciudad o áreas específicas. Esto permite un análisis de cultivos en una zona con similitudes en clima y propiedades de suelo, aumentando el área de clasificación sin perder exactitud. La resolución requerida de las imágenes espectrales y la complejidad de los modelos de clasificación pueden disminuir debido a la homogeneidad de las zonas [23, 24].
- **Estudios por país:** Mientras aumenta el área que se desea clasificar, surgen más problemas. Un área mayor requiere imágenes nítidas. En varios estudios [25], se han encontrado dificultades para obtener imágenes sin nubes y falta de datos etiquetados. En países europeos, donde existe un mayor acceso a datos, el problema radica en la heterogeneidad de los mismos. Para solucionar esto, se han creado conjuntos de datos como Eurocrops [26] y CropHarvest [6], que

2 Datos

buscan ofrecer alternativas estandarizadas para encontrar datos etiquetados.

Para Ecuador, el MAG ha publicado diversos datos del sector agropecuario en el Geoportal del Agro Ecuatoriano [3]. Este geoportal proporciona información sobre estimaciones de cultivos permanentes como banano, caña de azúcar y palma africana desde 2018 hasta 2022; y cultivos anuales como arroz, maíz y soya desde 2020 hasta 2021. Sin embargo, solo se ha podido recopilar un mapa de etiquetas por año.

Además, en lugares tropicales que presentan diferentes condiciones climáticas, se enfrentan desafíos como la dificultad para obtener datos nítidos sin la presencia de nubes temporales. Esto se debe a que las nubes pueden ocultar la superficie terrestre, lo que complica la captura de imágenes satelitales claras [25]. Ambas cuestiones dificultan la obtención de datos adecuados para el entrenamiento de clasificadores de cultivos.

2.4 Datos Personalizados para Clasificación de Cultivos

La clasificación de cultivos requiere acceso a datos precisos y personalizados, especialmente imágenes satelitales de alta calidad en el área de estudio requerida. A continuación, se presentan algunas herramientas que permiten la obtención y procesamiento de estos datos:

- **Sentinel Hub [27]**: Proporciona acceso a imágenes espectrales, incluyendo Sentinel-2 y Sentinel-1 y herramientas para obtenerlas de forma fácil y personalizada.
- **Copernicus Open Access Hub [28]**: Similar a Sentinel Hub, ofrece acceso a imágenes de Sentinel-2 y Sentinel-1, pero a través de la plataforma oficial de Copernicus.
- **Geosat [29]**: Adicionalmente, existen empresas que se dedican a ofrecer servicios de imágenes satelitales personalizadas a demanda para monitoreo y detección.
- **Google Earth Engine (GEE) [30]**: GEE brinda acceso a una vasta colección de imágenes satelitales históricas y actuales, incluyendo Landsat y Sentinel-2. Permite procesar grandes volúmenes de datos de forma remota en la nube y es ampliamente utilizado en aplicaciones de teledetección y clasificación de cultivos.

Estas herramientas permiten a los investigadores y agricultores obtener y procesar datos personalizados para mejorar la precisión en la clasificación de cultivos, lo que a su vez ayuda a optimizar las prácticas agrícolas y mejorar la productividad.

Sin embargo, aunque plataformas como GEE facilitan el acceso y procesamiento eficiente de grandes conjuntos de datos —actualizándose diariamente y permitiendo la automatización de procesos—, existen limitaciones importantes. Por ejemplo, la calidad y resolución espacial de las imágenes pueden no ser suficientes para todas las aplicaciones agrícolas, especialmente cuando se requiere identificar cultivos a nivel de parcela. Además, la disponibilidad de datos etiquetados y la adaptación de los algoritmos a condiciones locales siguen siendo desafíos relevantes. Por tanto, la tesis aborda la problemática de cómo superar estas limitaciones para lograr una clasificación de cultivos más precisa y adaptable a contextos específicos, aprovechando al máximo las capacidades de las plataformas existentes, pero reconociendo sus restricciones.

3

Técnicas de Aprendizaje Profundo

En este capítulo se describen las técnicas de aprendizaje profundo y se analizan las que se emplean con mayor frecuencia en la problemática de clasificación de Uso y Cobertura del Suelo (LCLU), con un enfoque particular en la clasificación de cultivos. Además, se detallan las arquitecturas de los modelos utilizados en el estado del arte, así como las metodologías propuestas en estudios realizados en el territorio ecuatoriano.

La clasificación en LCLU ha sido objeto de mucha investigación [4, 2, 26, 31, 32, 33, 34], debido a la importancia de comprender las características que nos permiten describir la cobertura de la tierra enfocada en las áreas de toma de decisiones de la agricultura. A lo largo de los años, la clasificación LCLU ha incorporado tecnologías como el aprendizaje automático (ML) y las imágenes satelitales para recopilar datos geoespaciales de áreas de interés [31].

Entre los algoritmos de ML empleados para la clasificación de uso y cobertura del suelo (LCLU), el método Random Forest (RF), una técnica supervisada que requiere datos etiquetados, destaca por su capacidad para manejar datos complejos y multidimensionales de manera eficiente [4]. Este algoritmo construye múltiples árboles de decisión durante el entrenamiento, donde cada árbol clasifica individualmente cada píxel en función de las características proporcionadas usando las etiquetas reales.

En la Figura 3.1.a se ilustra cómo RF se compone de diferentes árboles, cada uno de los cuales utiliza umbrales determinados automáticamente por el algoritmo de aprendizaje para dividir los datos y asignar las posibles clases a las que puede pertenecer una muestra. La clasificación final de cada píxel se determina mediante una votación mayoritaria entre todos los árboles, produciendo así mapas de segmentación, como se muestra en la Fig. 3.1.b, donde cada píxel es asignado a la clase más votada por el conjunto de árboles.

RF es especialmente efectivo para conjuntos de datos grandes, con ruido o con correlaciones entre características, gracias a su mecanismo de aleatorización. En cada división de un árbol, el algoritmo selecciona aleatoriamente un subconjunto de las características disponibles (por ejemplo, los 18 o N atributos espectrales de cada píxel satelital), lo que reduce la correlación entre los árboles y mejora la generalización del modelo. Esta aleatorización, junto con la sumarización por medio de voto mayoritario, permite que RF sea robusto frente a datos complejos y variables.

Sin embargo, a medida que aumenta el área de estudio y la complejidad de las características (por ejemplo, información temporal y áreas de estudio por país), los algoritmos simples como RF pueden volverse menos efectivos. Estas limitaciones, junto con la naturaleza tanto de los datos satelitales como de la tarea de clasificación, han impulsado el desarrollo de

3 Técnicas de Aprendizaje Profundo

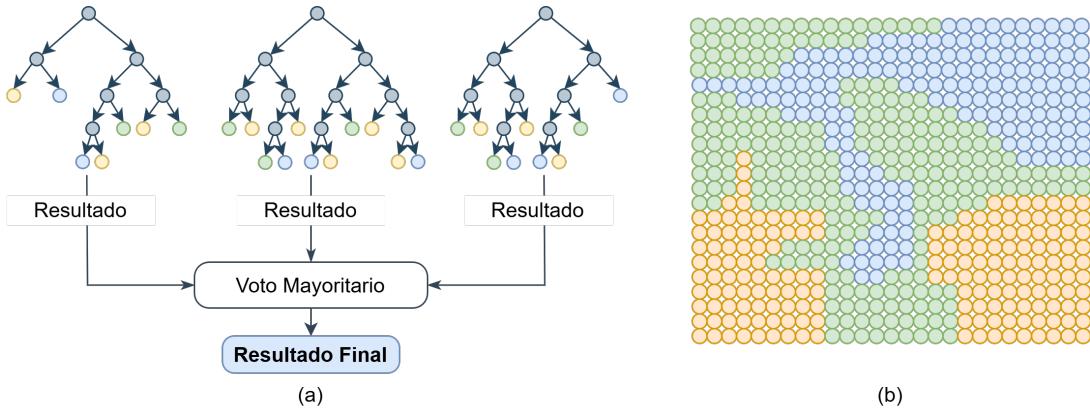


Figura 3.1: Mecanismo de algoritmo Random Forest: (a) Clasificación de píxeles por votación mayoritaria. (b) representación del mapa de la clasificación de píxeles.

modelos capaces de manejar grandes volúmenes de información y extraer características más complejas para mejorar la precisión en LCLU.

El aprendizaje profundo (DL), una subárea del ML, se basa en el uso de redes neuronales artificiales. La unidad básica de estas redes es la neurona, como se ilustra en la Figura 3.2. Cada neurona recibe información de múltiples unidades de la capa anterior o directamente de los datos de entrada del modelo. Esta información se combina mediante una suma ponderada, donde cada entrada x_i se multiplica por un peso individual w_i , y se añade un sesgo b :

$$z = \sum_{i=1}^N w_i x_i + b \quad (3.1)$$

Posteriormente, el valor resultante z se transforma mediante una función de activación $f(z)$, que introduce no linealidad al modelo y permite a la red aprender patrones complejos. De esta forma, la salida de la neurona es a , la cual se transmite a las neuronas de la siguiente capa [35]. Este mecanismo permite a las redes neuronales aprender patrones complejos y realizar tareas avanzadas como clasificación, predicción y reconocimiento.

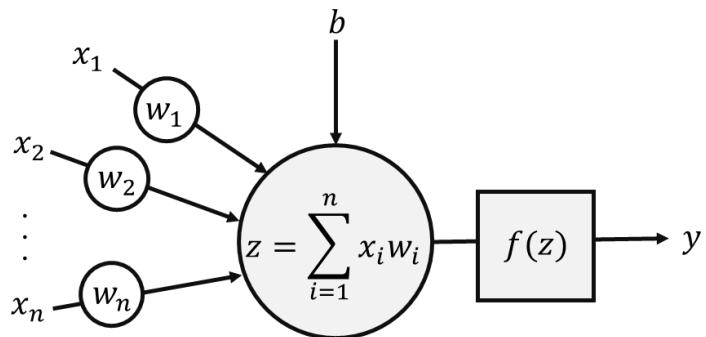


Figura 3.2: Mecanismo de una neurona.

Una red neuronal consta de tres partes principales: la capa de entrada, las capas ocultas y la capa de salida, como se muestra

3 Técnicas de Aprendizaje Profundo

en la Figura 3.3. El número de neuronas y las conexiones entre capas se establecerán según la complejidad del problema que se busca resolver. Cada capa de la red está representada por un vector o matriz. Cada elemento de este vector puede interpretarse como una unidad que desempeña el papel de una neurona [35].

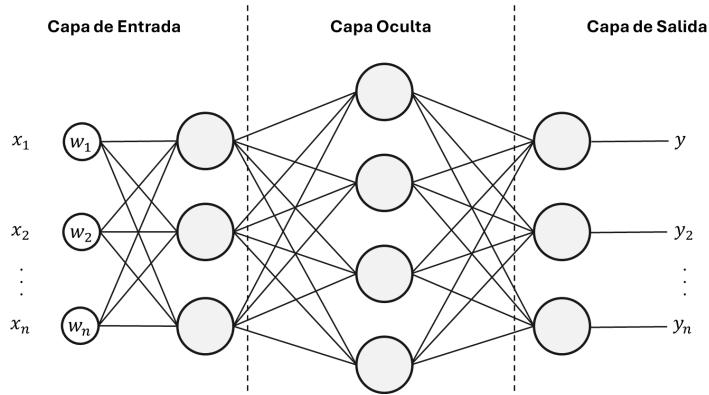


Figura 3.3: Capas de red neuronal: Capa de entrada, oculta y salida.

En la clasificación de LCLU, se utilizan imágenes espectrales como datos de entrada para redes neuronales. Estas imágenes se representan como matrices tridimensionales que contienen información sobre latitud, longitud y bandas espectrales. Las redes neuronales especializadas en el procesamiento de datos matriciales se conocen como redes neuronales convolucionales (CNN, por sus siglas en inglés) [35]. A diferencia de las redes tradicionales, las CNN emplean la operación de convolución.

Las configuraciones de las capas de convolución y las dimensiones de las matrices pueden variar, lo que permite diseñar redes más eficientes según la naturaleza de los datos. Las redes convolucionales surgieron en el campo del DL como una herramienta especializada para el análisis de datos matriciales, destacando particularmente en la identificación y extracción de patrones complejos en imágenes [35].

En la Figura 3.4 se ilustra el funcionamiento de las redes convolucionales. Cada capa de convolución, seguida de una función de activación como ReLU (Rectified Linear Unit), actúa en conjunto como un filtro no lineal que permite extraer características relevantes de la imagen. Es importante destacar que no es únicamente la función de activación la que opera como filtro, sino la combinación de la operación de convolución y la activación la que posibilita la detección de patrones complejos y la representación jerárquica de las características.

Posteriormente, una capa de pooling reduce la cantidad de parámetros y el tamaño espacial de la representación generada por la capa de convolución, disminuyendo así la dimensionalidad y mejorando la eficiencia computacional. Finalmente, se incluye una capa completamente conectada (fully connected) o una capa específica de salida, encargada de clasificar las características extraídas en las etapas anteriores.

Las CNN suelen enfocarse en clasificar imágenes completas, asignando una clase o categoría a toda la imagen. Sin embargo, en tareas como la identificación o segmentación de cultivos a gran escala, es necesario clasificar cada píxel de una imagen para generar un mapa detallado de segmentación. Para lograr esto, se emplea una arquitectura especializada llamada encoder-decoder (ver Figura 3.5). Esta arquitectura está diseñada para transformar una entrada, como una imagen, en una salida estructurada, como un mapa de segmentación donde cada píxel tiene asignada una clase [37]. Consta de dos

3 Técnicas de Aprendizaje Profundo

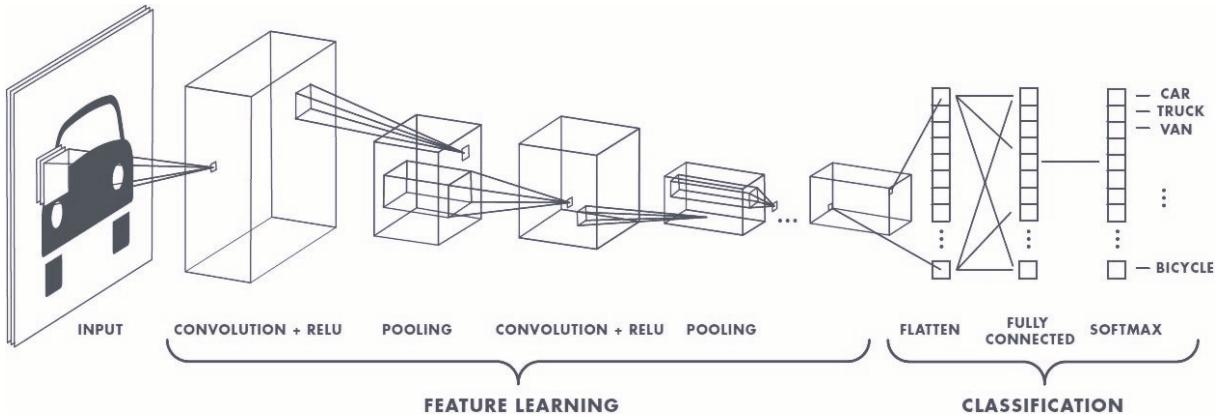


Figura 3.4: Arquitectura de red neuronal convolucional. Obtenido de [36].

componentes principales:

Las CNN suelen enfocarse en clasificar imágenes completas, asignando una clase o categoría a toda la imagen. Sin embargo, en tareas como la identificación o segmentación de cultivos a gran escala, es necesario clasificar cada pixel de una imagen para generar un mapa detallado de segmentación. Para lograr esto, se emplea una arquitectura especializada llamada **encoder-decoder** (ver Figura 3.5). La arquitectura encoder-decoder está diseñada para transformar una entrada, como una imagen, en una salida estructurada, como un mapa de segmentación donde cada pixel está etiquetado con su clase correspondiente [37]. Esta arquitectura consta de dos componentes principales:

- **Encoder (Codificador):** El encoder toma la imagen de entrada y la convierte en una representación comprimida, extrayendo características relevantes mientras reduce la resolución espacial.
- **Decoder (Decodificador):** A partir de esta representación comprimida, el decoder reconstruye un mapa a la resolución original de la imagen, prediciendo una distribución de probabilidad por clase para cada pixel. La clase final se asigna seleccionando aquella con la mayor probabilidad.

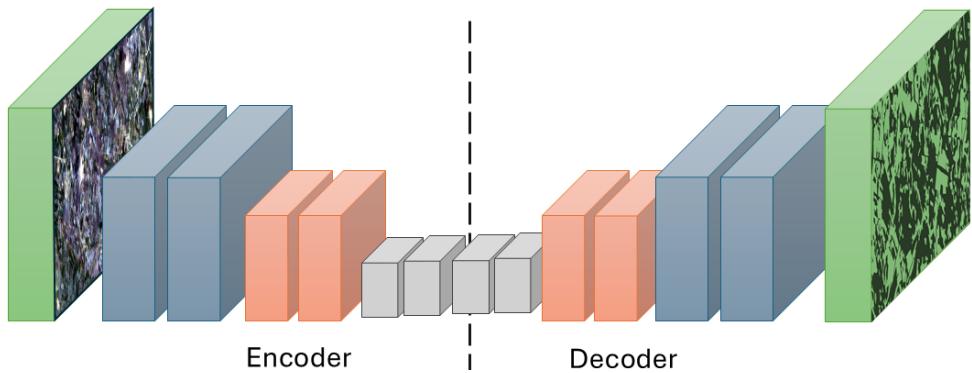


Figura 3.5: Arquitectura encoder-decoder.

La flexibilidad de esta arquitectura permite que tanto el encoder como el decoder utilicen diferentes mecanismos para procesar los datos. Por ejemplo, las capas convolucionales son comúnmente empleadas para reducir la dimensionalidad y

3 Técnicas de Aprendizaje Profundo

extraer características de las imágenes. Además, esta arquitectura puede incorporar información adicional, como metadatos o semántica contextual, para mejorar la precisión de la clasificación.

Una arquitectura variante que utiliza capas convolucionales es **U-Net** [38]. Como se muestra en la Figura 3.6, U-Net se asemeja al mecanismo encoder-decoder, donde la primera parte (encoder) reduce la resolución espacial mediante convoluciones y pooling, mientras que la segunda parte (decoder) la expande mediante capas convolucionales transpuestas (también conocidas como deconvoluciones), que permiten aumentar la resolución espacial. Sin embargo, U-Net introduce una mejora: durante la expansión, concatena los mapas de características de las capas correspondientes del encoder con las del decoder mediante conexiones de salto (skip connections). Esto explica su forma simétrica en "U" y su nombre.

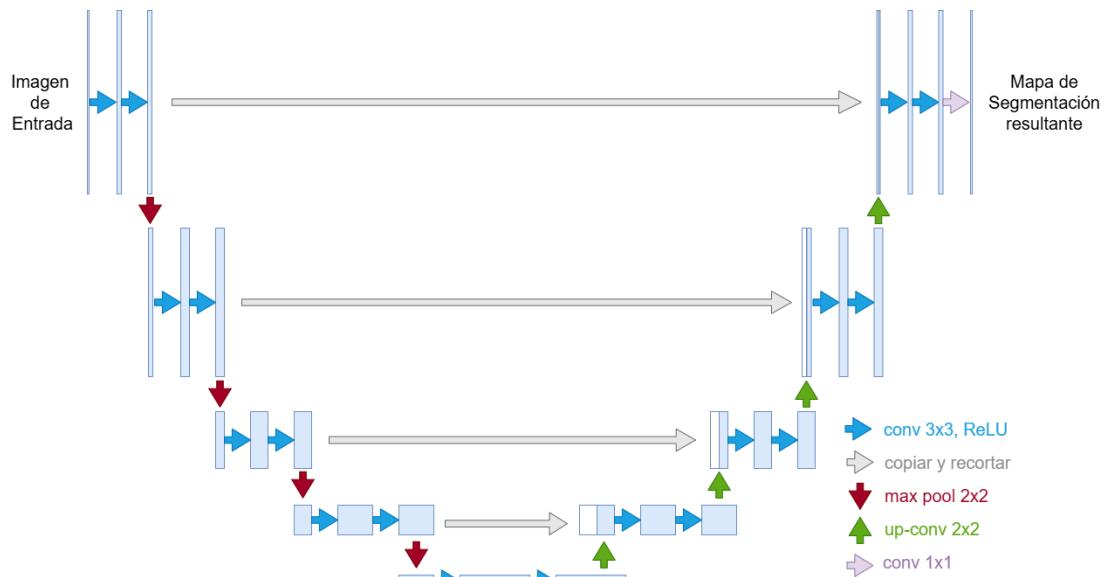


Figura 3.6: Arquitectura U-net.

Estas conexiones permiten recuperar información espacial detallada, como bordes y texturas, que se pierden durante la reducción de resolución en el encoder. Por ejemplo, en imágenes de cultivos, esto es necesario principalmente para segmentar con precisión los límites entre cultivos o tipos de suelo. En la parte del encoder, U-Net puede incorporar configuraciones de modelos CNN. Algunos ejemplos específicos aplicados a la clasificación de cultivos se detallarán en secciones posteriores.

Tanto las arquitecturas basadas en CNN y U-Net han demostrado un gran potencial para la clasificación y segmentación de cultivos en imágenes de teledetección; sin embargo, presentan limitaciones importantes. Entre las principales restricciones destaca la necesidad de grandes volúmenes de datos etiquetados de alta calidad, los cuales suelen ser escasos o costosos de obtener en el ámbito agrícola. Además, estos modelos pueden tener dificultades para generalizar a nuevas regiones o condiciones distintas a las vistas durante el entrenamiento, lo que limita su transferibilidad y utilidad práctica [5, 6].

Asimismo, las CNN y U-Net son sensibles a la variabilidad espectral y espacial de las imágenes satelitales, así como a la presencia de ruido y artefactos, lo que puede afectar la precisión de la segmentación y clasificación. Por ello, es fundamental explorar técnicas complementarias como el aprendizaje transferido, el aprendizaje semi-supervisado o el uso de datos sintéticos para mitigar estos desafíos y mejorar la generalización de los modelos en contextos de datos limitados [5].

3 Técnicas de Aprendizaje Profundo

A continuación, se presentan modelos y técnicas de DL utilizados en la clasificación de cultivos, los cuales implementan diversas estrategias de aprendizaje basadas en redes neuronales.

3.1 Clasificación Mediante Aprendizaje por Transferencia

La clasificación de uso y cobertura del suelo, incluyendo cultivos, se realiza comúnmente mediante arquitecturas de DL, especialmente redes neuronales convolucionales (CNN) como ResNet-50, Inception-V3 y VGG-19 [31, 39, 40]. Estos modelos han demostrado una precisión superior frente a otros métodos tradicionales, gracias a su capacidad para extraer y aprender representaciones complejas de las imágenes satelitales.

Sin embargo, el entrenamiento de estas arquitecturas desde cero requiere grandes volúmenes de datos etiquetados y un considerable poder computacional. Aquí es donde el aprendizaje por transferencia (TL) cobra relevancia: esta técnica permite reutilizar el conocimiento adquirido por modelos preentrenados en tareas relacionadas, reduciendo significativamente el tiempo de entrenamiento y la complejidad computacional, especialmente en contextos con datos geoespaciales y temporales de gran volumen. El TL ha permitido obtener modelos generalizables con menos datos y menor tiempo de entrenamiento, aunque aún se requiere más investigación para explotar su potencial en teledetección [31].

Estudios recientes [31, 41, 39] han reportado altos niveles de exactitud al aplicar TL en tareas de clasificación de uso de suelo. Utilizando pesos preentrenados en grandes conjuntos de imágenes, como ImageNet, y ajustando los modelos a tareas específicas de uso de suelo, se han alcanzado precisiones del 97,29% para ResNet-50, 97,57% para Inception-V3 y 96,86% para VGG-19.

ImageNet, aunque no está centrado en LCLU, contiene alrededor de 1000 clases variadas y ha demostrado ser útil para la identificación de uso de suelo y cultivos [41]. En [31], se empleó TL con pesos preentrenados en ImageNet para clasificar imágenes satelitales de baja altura de cultivos. El conjunto de datos UCM, utilizado para el ajuste fino (fine-tuning), contiene 21 clases de LCLU con 100 imágenes por clase, incluyendo categorías como bosque, agricultura y ríos. Con este enfoque, se lograron las precisiones mencionadas anteriormente.

En la Figura 3.7 se ilustra la diversidad de imágenes de ImageNet (Fig. 3.7.a) y ejemplos del conjunto de datos UCM (Fig. 3.7.b), mostrando las diferencias y similitudes entre las imágenes de preentrenamiento y las de ajuste fino empleadas en la tarea de LCLU.

Otro enfoque utilizado en DL que aprovecha el aprendizaje por transferencia son las redes neuronales convolucionales profundas (DCNN), caracterizadas por su gran número de capas y complejidad. Ejemplos destacados incluyen CaffeNet, GoogleNet y EfficientNet. En el estudio de Scott et al. [39], se explora la aplicación de TL usando estas DCNN, analizando cómo el aumento de datos y el ajuste en la extracción de características afectan el rendimiento. Para evaluar los modelos, utilizaron dos conjuntos de datos de referencia ampliamente empleados en estudios de clasificación: UCM [42], que contiene imágenes de uso de suelo de alta resolución, y RSD [43], enfocado en imágenes de teledetección. Los resultados mostraron que el aumento de datos mejora significativamente la exactitud, con ResNet-50 alcanzando un $98,5 \pm 1,4\%$ en UCM y un $97,8 \pm 2,7\%$ en RSD. Además, estas arquitecturas fueron usadas para clasificar imágenes completas, por lo que esto nos puede proporcionar una idea de los modelos que pueden usarse como base en la parte de codificación si se desea realizar segmentación.

3 Técnicas de Aprendizaje Profundo



Figura 3.7: Ejemplo de Trasferencia de aprendizaje usando dos conjuntos de datos: (a) Conjunto de imágenes ImageNet (+14 millones de imágenes), (b) Conjunto de imágenes UCM (2 100 imágenes).

El TL también se ha utilizado para desarrollar nuevos modelos que requieren mecanismos para reducir el tiempo de procesamiento, ya sea debido a su complejidad o a la falta de datos necesarios en los modelos preentrenados. En [40], se introduce un innovador modelo llamado TIML (Task-Informed Meta-Learning), basado en el metaaprendizaje agnóstico de modelos (MAML) [44], diseñado para aprovechar los metadatos en aplicaciones agrícolas. Este estudio subraya la importancia de que las áreas de preentrenamiento y las áreas objetivo comparten características y tareas similares para una transferencia efectiva del aprendizaje. Un hallazgo significativo fue que, en los puntos de datos de Brasil, el rendimiento del modelo empeoró al aplicar TL. Esto resalta la necesidad de seleccionar cuidadosamente los conjuntos de datos preentrenados para garantizar una transferencia de conocimiento efectiva entre diferentes regiones agrícolas. Esto significa que al usar TL en segmentación de cultivos es primordial que los datos usados para preentrenar sean lo más similares a los datos objetivo.

Por otro lado, el modelo más avanzado para la clasificación de cultivos se presenta en [45]. Este modelo utiliza un Transformador de Teledetección Preentrenado (Presto) y ha demostrado un alto rendimiento en diversos escenarios de imágenes satelitales. La arquitectura facilita la adaptación a diferentes dimensiones de entrada y admite predicciones de series temporales. Como se ilustra en la Figura 3.8, el modelo selecciona un píxel, que incluye las diferentes bandas espectrales, y lo codifica para obtener características de dicho píxel. Posteriormente, extrae más estadísticos para luego realizar un algoritmo de clasificación como Random Forest para clasificar el píxel. Este proceso se repite en toda la imagen hasta conseguir un mapa de segmentación completo.

A pesar de su aparente simplicidad, el modelo ha mostrado resultados prometedores con una exactitud del 95,3%, reduciendo significativamente el tiempo y la memoria requeridos para el entrenamiento y ejecución de los modelos usados para la segmentación. Sin embargo, al ser un desarrollo reciente, requiere mayor análisis para comprender sus capacidades y

3 Técnicas de Aprendizaje Profundo

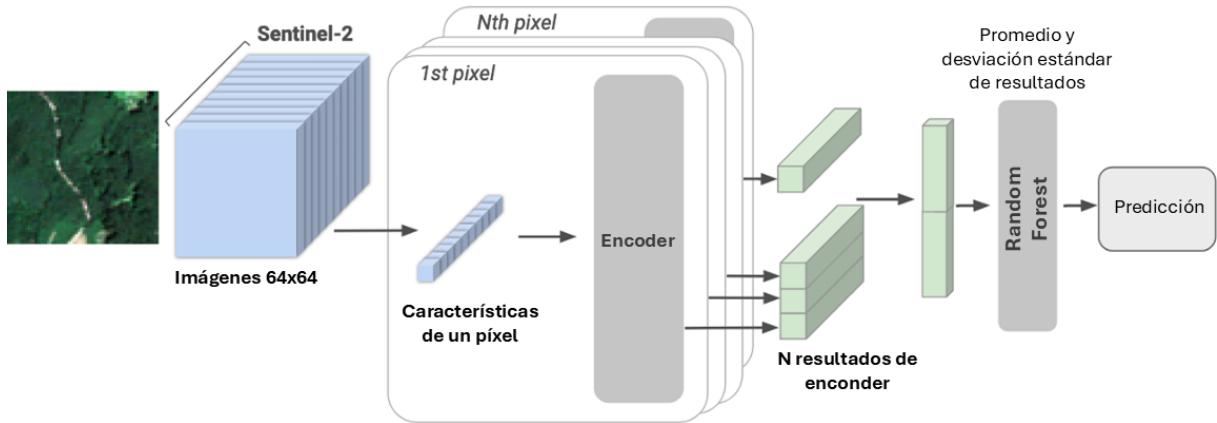


Figura 3.8: Arquitectura de Presto. Obtenido de [45].

limitaciones en diferentes contextos agrícolas.

3.2 LCLU en Ecuador

En Ecuador, la falta de datos detallados sobre el uso y cobertura del suelo representa un desafío significativo para los estudios ambientales y la toma de decisiones en el sector agrícola [46]. Ante esta carencia, los trabajos de monitoreo y detección de uso de suelo suelen apoyarse en fuentes públicas, principalmente imágenes satelitales [46, 47, 2].

Un ejemplo de este enfoque se aprecia en [47], que utilizó imágenes satelitales para identificar patrones de deforestación y cultivos ilegales de coca en la Amazonía ecuatoriana y colombiana. En este caso, se emplearon algoritmos de clasificación no supervisada, como métodos de agrupamiento (clustering), para categorizar áreas forestales, agrícolas, hídricas y urbanizadas a lo largo de varios años, sin necesidad de etiquetas previas. Este tipo de análisis permitió calcular índices de deforestación y detectar posibles zonas de cultivos ilegales.

En la región costera, Cevallos et al. [46] aplicaron imágenes Sentinel-2 para identificar cultivos de caña de azúcar, enfrentando como principal limitación la escasa disponibilidad de imágenes libres de nubosidad, lo que dificultó la obtención de clasificaciones precisas.

En contextos urbanos, los métodos de aprendizaje automático basados en agrupamiento han demostrado ser eficaces para áreas geográficas reducidas. Por ejemplo, en el cantón Quero, provincia de Tungurahua, la clasificación del uso del suelo mediante imágenes satelitales alcanzó una precisión del 83,29% con técnicas tradicionales de ML, y mejoró hasta un 87,43% al aplicar modelos de DL [2].

Sin embargo, a pesar de estos avances, el desarrollo de tecnologías robustas para la detección y monitoreo de la cobertura agrícola en Ecuador sigue viéndose limitado por la escasez de datos detallados y etiquetados. Aunque se han alcanzado altos niveles de precisión en estudios puntuales, como el mencionado, estos resultados no son fácilmente generalizables a regiones agrícolas más amplias y diversas.

Frente a estas limitaciones, el presente proyecto se diferencia al centrarse en el desarrollo y evaluación de metodologías

3 Técnicas de Aprendizaje Profundo

de aprendizaje profundo orientadas a la generación precisa de mapas de cultivos en distintas zonas agrícolas del país, priorizando la superación de los desafíos asociados a la adquisición y el etiquetado manual de datos en escenarios donde estos recursos son escasos. De este modo, la investigación busca avanzar en la aplicación de técnicas de teledetección agrícola adaptadas a la realidad ecuatoriana, ampliando el alcance y la robustez de las soluciones existentes para el monitoreo y gestión eficiente del sector agrícola.

3.3 Criterios de Evaluación y Optimización en DL

Los criterios de evaluación y optimización en modelos de Deep Learning son fundamentales para garantizar la eficacia y generalización de los modelos. Se probaron y evaluaron varias funciones de pérdida para optimizar el entrenamiento de los modelos CNN, y se utilizaron diversas métricas para comparar la eficiencia de los modelos propuestos.

3.3.1 Función de Pérdida

La función de pérdida (loss function) cumple un rol central al cuantificar la discrepancia entre las predicciones del modelo y los valores reales, guiando así el proceso de optimización durante el entrenamiento. La elección de esta función depende directamente del tipo de problema abordado. Para casos de clasificación, se emplean comúnmente las siguientes funciones de pérdida:

Entropía cruzada (Cross-entropy) [48] es una función usada para medir la discrepancia entre distribuciones de probabilidad reales (P) y predicciones (Q). La entropía cruzada se define generalmente de la siguiente forma:

$$H(P, Q) = - \sum_{x \in X} P(x) \log(Q(x)) \quad (3.2)$$

donde $P(x)$ representa la distribución real, es decir, las etiquetas verdaderas, y $Q(x)$ representa la distribución de las predicciones del modelo.

La entropía puede usarse para la clasificación binaria y la clasificación categórica, como se emplea en [31, 49] de la siguiente forma:

Binary cross-entropy [48], usadas en clasificaciones binarias, es decir, dos clases (0,1), la cual se define como:

$$L = -[y \log(p) + (1 - y) \log(1 - p)] \quad (3.3)$$

Donde y representa la etiqueta verdadera (0 o 1), y p , la probabilidad de que el valor predicho sea de la clase 1.

Categorical cross-entropy [48], es usada para clasificaciones multiclas ($C \geq 2$ clases), la cual se define:

$$L = - \sum_{i=1}^C y_i \log(p_i) \quad (3.4)$$

3 Técnicas de Aprendizaje Profundo

Donde y_i representa la etiqueta real en formato one-hot encoding (1 para la clase verdadera, 0 para las demás), este formato se explica en profundidad en la subsección 4.2. La variable p_i es la probabilidad predicha por el modelo para la clase i . En este caso, cada píxel tendrá un conjunto de probabilidades de pertenecer a las diferentes clases.

Binary focal loss [50] es una adaptación de la entropía cruzada binaria para manejar desbalanceo de clases, enfocándose en ejemplos difíciles de clasificar. Esta función de pérdida se define como:

$$\text{Binary focal loss}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

donde: $p_t = \begin{cases} p & \text{si } y = 1 \\ 1 - p & \text{si } y = 0 \end{cases}$ (3.5)

Donde γ (gamma) controla el enfoque en ejemplos difíciles (con valores típicos entre 2 y 5) y α_t es el peso para balancear clases (por ejemplo: $\alpha = 0.25$ para una clase minoritaria).

Categorical focal loss [50] es una extensión multiclasa que generaliza el concepto de binary focal loss.

$$\text{Categorical focal loss}(p, y) = -\sum_{i=1}^K \alpha_i(1 - p_i)^\gamma y_i \log(p_i)$$

donde: $y_i = \begin{cases} 1 & \text{si la clase } i \text{ es verdadera} \\ 0 & \text{para el resto} \end{cases}$ (3.6)

Otras funciones de pérdidas usadas para problemas de segmentación tratan de evaluar la superposición entre la máscara real y la máscara predicha de las diferentes clases. Entre dichas funciones, probaremos usar las siguientes:

Dice loss [50], o función de pérdida Sorensen-Dice mide la superposición entre máscaras de segmentación predichas por el modelo (\hat{Y}) y las reales (Y):

$$Dice(Y, \hat{Y}) = \frac{2|Y \cap \hat{Y}|}{|Y| + |\hat{Y}|},$$

$Dice loss = 1 - Dice$

Donde un resultado de 0 muestra una superposición perfecta. Este tipo de funciones ha demostrado robustez ante desbalanceo de clases.

Jaccard loss [50], o función Jaccard de pérdida viene del índice IoU (intersección sobre la unión), la cual penaliza mayormente los errores en bordes y objetos pequeños en comparación con la función Dice. Esta función se define a continuación:

$$IoU(Y, \hat{Y}) = \frac{|Y \cap \hat{Y}|}{|Y \cup \hat{Y}|},$$

$Jaccard loss = 1 - IoU$

3 Técnicas de Aprendizaje Profundo

Por último, también se pueden proponer funciones de pérdida que combinan funciones previamente mencionadas [50]:

Crossentropy Dice loss, el cual combina la función de entropía cruzada y la función Dice. Según el caso, puede usar la entropía cruzada binaria (BCE) o la entropía cruzada categórica (CCE).

$$\begin{aligned} BCE \text{ Dice loss} &= \text{binary crossentropy} + \text{Dice loss} \\ CCE \text{ Dice loss} &= \text{categorical crossentropy} + \text{Dice loss} \end{aligned} \tag{3.9}$$

Crossentropy Jaccard loss, el cual combina la función de entropía cruzada y la función Jaccard.

$$\begin{aligned} BCE \text{ dice loss} &= \text{binary crossentropy} + \text{Jaccard loss} \\ CCE \text{ dice loss} &= \text{categorical crossentropy} + \text{Jaccard loss} \end{aligned} \tag{3.10}$$

Focal Dice loss, combina la Dice loss (sensibilidad a la superposición de regiones) con la filosofía de focal loss (énfasis en ejemplos difíciles). Es útil cuando hay desbalanceo de clases y necesitas buena superposición de regiones.

$$\begin{aligned} \text{Binary focal Dice loss} &= \text{binary focal loss} + \text{Dice loss} \\ \text{Categorical focal Dice loss} &= \text{categorical focal loss} + \text{Dice loss} \end{aligned} \tag{3.11}$$

Focal Jaccard loss, es una extensión del Jaccard loss (IoU) con parámetro focal para priorizar errores críticos.

$$\begin{aligned} \text{Binary focal Jaccard loss} &= \text{binary focal loss} + \text{Jaccard loss} \\ \text{Categorical focal Jaccard loss} &= \text{categorical focal loss} + \text{Jaccard loss} \end{aligned} \tag{3.12}$$

3.3.2 Métricas

Además de la función de pérdida, el rendimiento de los modelos se evalúa mediante métricas específicas, calculadas mediante la matriz de confusión (ver Figura 3.9). Usaremos las siguientes métricas:

Accuracy, o exactitud que calcula el porcentaje de los píxeles que fueron identificados correctamente en su respectiva clase:

$$\text{accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{3.13}$$

Recall, el cual calcula la proporción de los píxeles que fueron correctamente asignados a su categoría.

$$\text{recall} = \frac{TP}{TP + FN} \tag{3.14}$$

Precisión, la cual representa la proporción de píxeles que fueron identificados correctamente en la clase de cultivos.

3 Técnicas de Aprendizaje Profundo

		Predicción	
		Cultivo	No Cultivo
Real	Cultivo	Aciertos True Positive TP	Fallos False Positive FP
	No cultivo	Fallos False Negative FN	Aciertos True Negative TN

Figura 3.9: Matriz de confusión.

$$precision = \frac{TP}{TP + FP} \quad (3.15)$$

F1-score, el cual es la media armónica de precisión y recall.

$$F1\text{-score} = \frac{2 \times precision \times recall}{precision + recall} \quad (3.16)$$

Estas métricas, combinadas con análisis cualitativos de los mapas generados, proporcionan una visión del desempeño del modelo en escenarios reales, donde factores como la resolución espacial, presencia de nubes y la variabilidad de los cultivos introducen desafíos adicionales. En este tipo de tareas, la evaluación se realiza a nivel de píxel: cada píxel de la imagen clasificada se compara con su etiqueta real, y así se calculan métricas como accuracy, precision, recall y F1-score, lo que permite cuantificar de manera precisa el rendimiento del modelo en la identificación de las distintas coberturas o tipos de cultivo presentes en la imagen.

3.4 Preprocesamiento de imágenes satelitales

El preprocesamiento de imágenes satelitales constituye una etapa crítica en la teledetección moderna, especialmente en aplicaciones agrícolas donde la precisión y consistencia de los datos espectrales son fundamentales para el éxito de modelos de aprendizaje profundo [8]. La literatura reciente destaca que los principales desafíos en este contexto son la presencia de nubes y la heterogeneidadpectral derivada de diferentes sensores, condiciones atmosféricas o regiones geográficas [14].

3.4.1 Manejo de nubes

En la sección 2.3 se mencionan diferentes limitaciones en los datos, entre ellas la presencia de nubes en zonas tropicales y clima variado como el de Ecuador. La gestión de la presencia de nubes es uno de los problemas más recurrentes, debido

3 Técnicas de Aprendizaje Profundo

a su impacto directo sobre la calidad y continuidad de las series temporales [8]. Diversos trabajos recomiendan la utilización de técnicas de composición temporal, como la mediana, para generar imágenes libres de nubes a partir de múltiples observaciones dentro de una ventana temporal definida [14].

Adicionalmente, se destaca la importancia de emplear bandas de calidad con menor presencia de nubes, para lo cual herramientas como GEE permiten seleccionar imágenes más limpias [30]. La combinación de técnicas de filtrado y composición temporal constituye un procedimiento habitual para asegurar la integridad de los datos de entrada y mitigar los efectos de la nubosidad en estudios de clasificación y segmentación agrícola.

3.4.2 Normalización espectral

La normalización espectral es otro problema que debe tratarse durante el preprocesamiento de imágenes satelitales, especialmente relevante cuando se integran datos de diferentes sensores, fechas o regiones [51]. Para este propósito, se recomienda usar el escalado **Min-Max**, definido en la ecuación 3.17:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.17)$$

Donde x representa cada valor individual, x_{min} es el valor mínimo de la banda y x_{max} el valor máximo. Tanto el valor mínimo como el máximo deben determinarse por banda, y posteriormente se aplica la Ecuación 3.17 a cada píxel de forma independiente. Este procedimiento asegura que todas las bandas tengan valores de 0 a 1, lo que ayuda a evitar que los modelos otorguen un peso excesivo a aquellas bandas con valores numéricos más altos por naturaleza. Además, esta normalización es fundamental para facilitar la transferencia de aprendizaje entre dominios geográficos y garantizar la comparabilidad de los resultados.

4

Metodología

Este capítulo se centra en describir el desarrollo de una metodología robusta para la clasificación y segmentación de cultivos en regiones con disponibilidad limitada de datos, utilizando a Ecuador como caso de estudio. Se emplea un enfoque de aprendizaje por transferencia, utilizando un conjunto de datos de mayor volumen y geográficamente similar a Ecuador, para garantizar una transferencia generalizada de conocimiento. Este enfoque busca abordar los desafíos específicos de la región, aprovechando la similitud geográfica para mejorar la precisión en la identificación de patrones agrícolas locales.

A continuación, en la Figura 4.1 se describe la metodología propuesta para el desarrollo de un clasificador de cultivos.

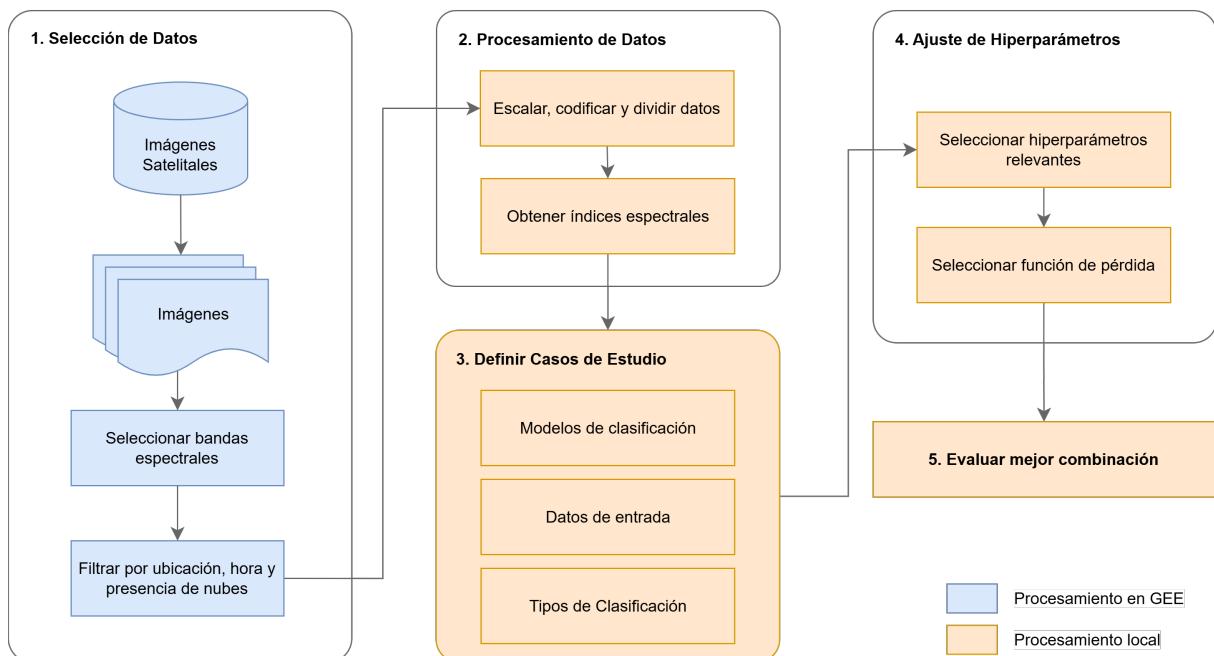


Figura 4.1: Diagrama de Flujo de Metodología.

La metodología está estructurada en cinco etapas clave:

1. **Selección de datos**, tanto de imágenes satelitales mediante la herramienta de GEE como datos etiquetados,
2. **Procesamiento de datos** que incluye procesamiento de imágenes como el cálculo de índices espectrales,

4 Metodología

3. **Definir de casos de estudio**, que incluye la comparación entre modelos de clasificación, datos de entrada y tipos de clasificación,
4. **Ajustes de hiperparámetros**, que contemplan la selección de hiperparámetros de los modelos y estrategias evaluadas, y la elección de una función de pérdida adecuada.
5. **Evaluar mejor combinación** de las opciones proporcionadas en los casos de estudio y ajuste de hiperparámetros.

En las secciones siguientes se profundiza en cada paso de la metodología.

4.1 Selección de Datos

En Ecuador, las etiquetas utilizadas para este estudio fueron obtenidas del Ministerio de Agricultura y Ganadería (MAG), correspondiendo específicamente a los cultivos de maíz, arroz y soya. Al explorar la base de datos del MAG, se identificó que únicamente existían registros etiquetados para estos cultivos en los años 2020 y 2021, hasta la fecha de la revisión. Debido a esta limitación temporal, la recopilación de imágenes satelitales se enfocó exclusivamente en ese periodo y en las zonas agrícolas donde se disponía de etiquetas confiables para maíz, arroz y soya.

Las etiquetas fueron descargadas a través de la plataforma de Datos Abiertos de Ecuador [52], la cual permite el acceso a información georreferenciada de diversos tipos. En este caso, se obtuvieron colecciones de polígonos en formato shapefile, que almacenan tanto las coordenadas como la etiqueta correspondiente a cada cultivo. Los datos etiquetados utilizados corresponden específicamente al segundo periodo de siembra, que abarca los meses de mayo a agosto.

La Figura 4.2 muestra un ejemplo de cómo se visualizan las etiquetas de los cultivos de maíz, arroz y soya para el año 2020 en la totalidad del territorio ecuatoriano. Se observa una alta presencia de polígonos de maíz (41,386), seguidos por arroz (11,559), mientras que la clase minoritaria corresponde a soya (801 polígonos). Esto evidencia un claro desbalance en las clases.

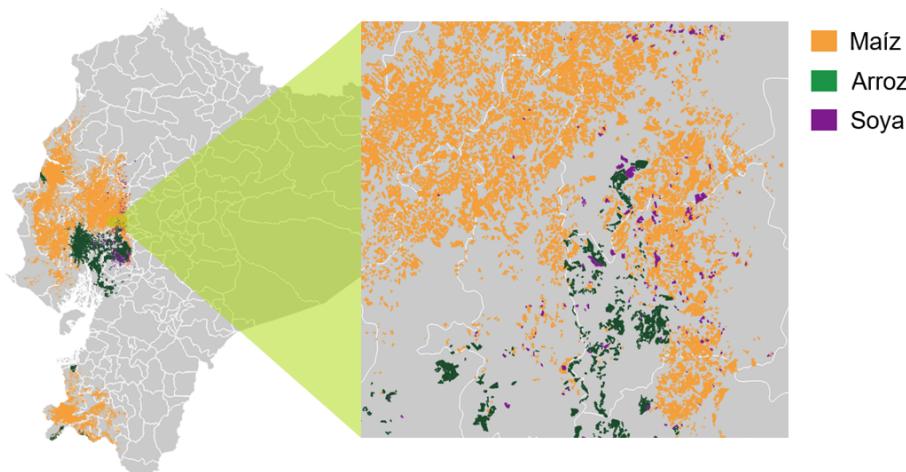


Figura 4.2: Ejemplo de etiquetas de cultivos de maíz, arroz y soya del 2020. Obtenido de MAG [3]

Para obtener los datos espectrales de las zonas de cultivo, se recurrió a imágenes satelitales Sentinel-2 mediante la apli-

4 Metodología

cación GEE, cargando los polígonos etiquetados a esta plataforma. Posteriormente, se seleccionaron tres zonas del mapa ecuatoriano con mayor cantidad de cultivos etiquetados y menor presencia de nubes, como se muestra en la Figura 4.3. Se extrajeron 10 bandas espectrales disponibles en Sentinel-2: azul, verde, rojo, cuatro bandas de borde rojo, infrarrojo cercano (NIR) y dos bandas de infrarrojo de onda corta (SWIR1 y SWIR2), lo que permite una representación espectral detallada de la superficie agrícola.

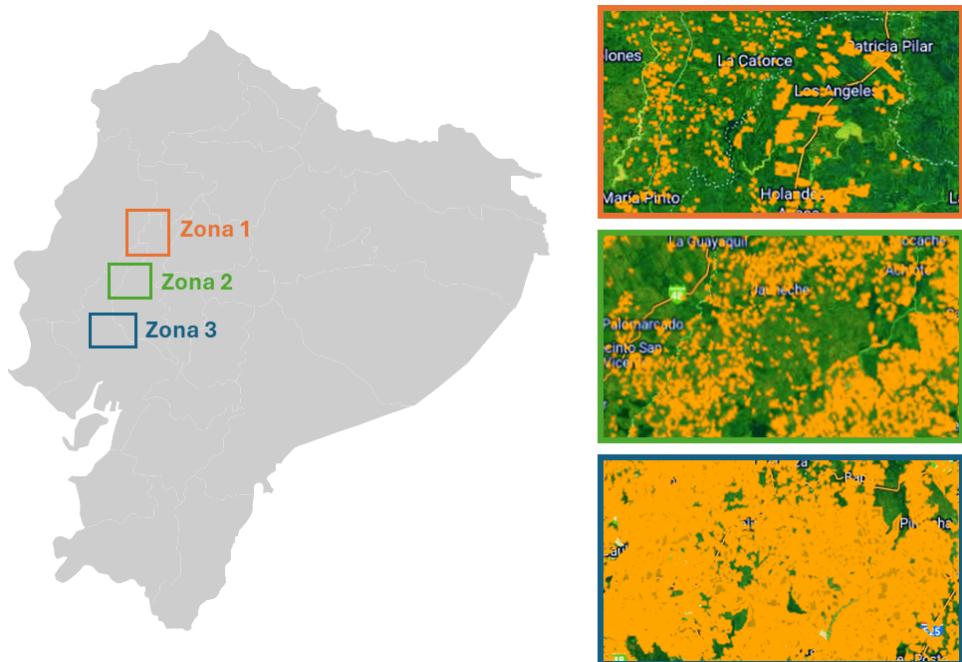


Figura 4.3: Zonas de datos escogidas del mapa de Ecuador, donde la etiqueta amarilla corresponde a los cultivos unificados de maíz, arroz y soya.

Como se explicó en secciones anteriores, la limitada cantidad de datos etiquetados disponibles en Ecuador dificulta el entrenamiento de modelos de aprendizaje profundo realmente robustos. Para superar esta restricción, se optó por incorporar un conjunto de datos internacional que presentara cultivos similares a los de interés local y así aplicar técnicas de TL. En este contexto, El conjunto de datos propuesto por Rustowicz et al.[14] de Ghana fue seleccionado como fuente complementaria, ya que su base de datos satelital incluye maíz, arroz y soya, junto con series temporales extensas y etiquetas validadas en campo. Para mantener la coherencia entre ambos contextos, únicamente se seleccionaron del conjunto de Ghana aquellas muestras correspondientes a los mismos cultivos presentes en las etiquetas ecuatorianas, lo que permite transferir el conocimiento adquirido y adaptar los modelos a las condiciones agrícolas locales.

Este conjunto de datos principal está compuesto por 4,040 series temporales de imágenes Sentinel-2 capturadas en Ghana durante 2016 y 2017, años que corresponden al periodo en el que se realizó el etiquetado de los cultivos en esa región y representan la información disponible y validada. Cada serie temporal corresponde a una ubicación específica y contiene entre 45 y 213 imágenes tomadas en diferentes momentos del ciclo agrícola, variando según la frecuencia de adquisición de Sentinel-2, la presencia de nubes y la disponibilidad de datos válidos en cada sitio.

Las imágenes tienen un tamaño de 64x64 píxeles, con una resolución espacial de 10 metros por píxel. En términos de for-

4 Metodología

mato, cada imagen es un mapa de bits donde cada píxel representa información espectral de una ubicación concreta dentro del recorte regular correspondiente a la zona de interés. Estas imágenes contienen las mismas 10 bandas espectrales que las obtenidas para Ecuador. En la Figura 4.4 se muestran ejemplos del conjunto de imágenes satelitales de Ghana: cada columna corresponde a una ubicación diferente, mientras que las dos primeras filas presentan imágenes de una misma parcela tomadas en distintos momentos del ciclo agrícola. La última fila muestra las etiquetas asociadas a cada parcela, que permanecen constantes para todos los tiempos correspondientes a la misma ubicación.

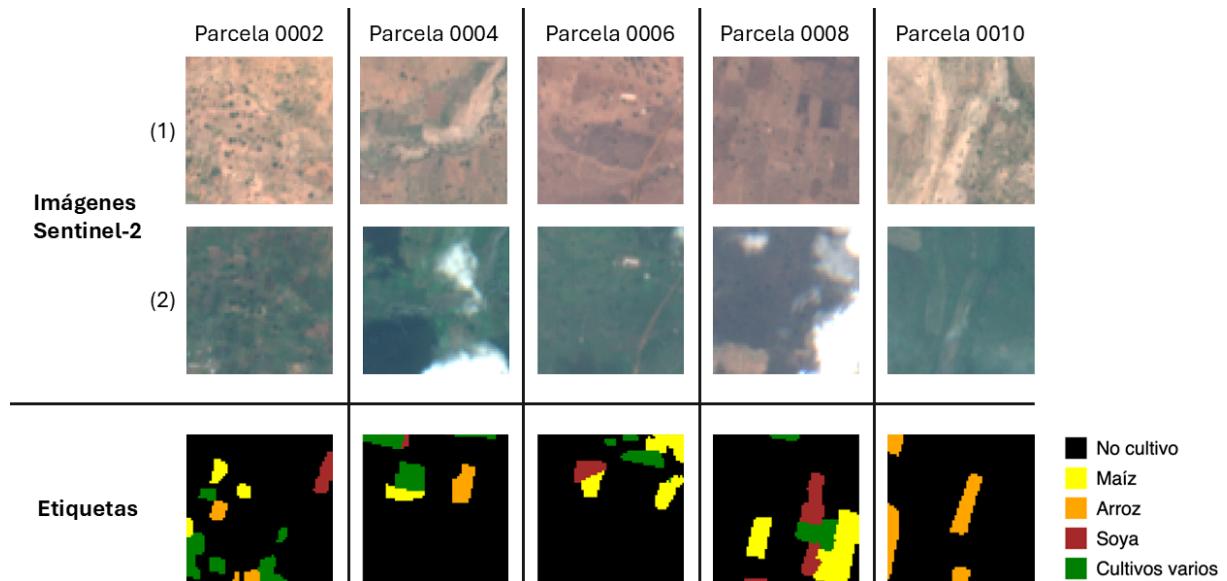


Figura 4.4: Imágenes del conjunto de datos Ghana: Imágenes pertenecientes al satélite Sentinel-2 (S2) de diferentes fechas, con sus respectivas etiquetas.

4.2 Procesamientos de Datos

En la subsección anterior se identificaron algunos desafíos en la aplicación de modelos de aprendizaje supervisado, como la diferencia entre las etiquetas del conjunto de datos de Ghana y los cultivos objetivo en las imágenes ecuatorianas, así como la presencia de nubes en ambos conjuntos. Para abordar estos problemas, se realizó un proceso de preprocesamiento con el fin de adecuar, limpiar y estandarizar los distintos conjuntos de datos propuestos.

4.2.1 Manejo de nubes

Para reducir la presencia de nubes en las imágenes satelitales, se aplicaron estrategias diferenciadas según el conjunto de datos utilizado. En el caso de Ghana, el propio dataset incluía máscaras de nubes, lo que permitió identificar y descartar directamente las zonas afectadas. Se estableció un umbral de calidad y aquellas imágenes que presentaban más del 25% de cobertura nubosa fueron excluidas del análisis, siguiendo prácticas habituales en estudios agrícolas donde se busca un equilibrio entre calidad y disponibilidad de datos. Para las imágenes que cumplieron con el umbral establecido, se utilizaron las máscaras correspondientes para descartar los píxeles nublados antes del entrenamiento, evitando así que estos afectaran el proceso de aprendizaje.

4 Metodología

Para datos ecuatorianos, se empleó el filtro de GEE para obtener imágenes con menos del 25% de presencia de nubes. Una vez obtenidas las imágenes con menor presencia, se realizó un proceso de composición temporal por mes, para reducir aún más los espacios en blanco por nubes. Debido al clima húmedo del segundo periodo de siembra, los meses que se encontraron imágenes con el requerimiento de presencia de nubes fueron mínimos.

4.2.2 Preparación y formateo de datos

Luego de la selección, filtrado y composición de las imágenes ecuatorianas , se definieron tres zonas agrícolas dentro del país (ver Figura 4.3). Estas zonas fueron seleccionadas por su alta densidad de etiquetas de cultivos y una menor presencia de nubes. Cada imagen obtenida de las diferentes zonas corresponden a un mapa de bits tridimensional de dimensiones $(X, Y, 10)$, donde X y Y representan las dimensiones espaciales de la zona (en números de píxeles), y 10 corresponde a las bandas espectrales seleccionadas.

Se intentó obtener una imagen mensual por zona para los años 2020 y 2021 (es decir, hasta ocho imágenes por zona). Sin embargo, debido a las restricciones impuestas por la presencia de nubes (ver sección 4.2.1), solo se obtuvieron entre 2 y 3 imágenes luego del proceso de filtrado y composición mediante mediana temporal.

Las dimensiones espaciales finales de las imágenes compuestas por zona, así como la cantidad de sub-imágenes extraídas a partir de cada una, se detallan en la Tabla 4.1. Las dimensiones indicadas corresponden al ancho, largo y número de bandas (10), y fueron luego divididas en fragmentos de $(64, 64, 10)$ píxeles. Solo se conservaron aquellas sub-imágenes que contenían al menos un pixel etiquetado como cultivo, descartando las que representaban únicamente áreas sin cultivo para reducir el desbalance de clases. Al finalizar este proceso, se obtuvieron 3,944 matrices de dimensión $(64, 64, 10)$, combinando las tres zonas y todos los años en los que se dispuso de datos.

Tabla 4.1: Cantidad de imágenes obtenidas por zona.

Zonas escogidas	Dimensiones (ancho, largo, bandas)	imágenes totales por zona ($64 \times 64 \times 10$)
Zona 1	$(2036, 2736, 10)$	2,080
Zona 2	$(1201, 1105, 10)$	988
Zona 3	$(1939, 1709, 10)$	876
Total		3,944

En la Figura 4.5.a se muestran las tres zonas seleccionadas sobre el mapa de Ecuador. La subfigura 4.5.b muestra un ejemplo del área correspondiente a la zona 2, donde se visualizan las etiquetas de cultivo resaltadas. Finalmente, en la subfigura 4.5.c se presenta una subimagen procesada con el formato final de $(64, 64, 10)$ utilizada como entrada al modelo.

Para generar la máscara correspondiente a cada sub-imagen o matriz de bits, se utilizó el conjunto ecuatoriano de polígonos etiquetados. Estos polígonos contienen información georreferenciada, incluyendo el nombre del cultivo (en la etiqueta `ncu`) y su geometría espacial.

El primer paso consistió en estandarizar el sistema de referencia espacial. En el ámbito de la geodetección, existen distintos sistemas de referencia de coordenadas (CRS en inglés), los cuales constituyen marcos utilizados para medir con precisión

4 Metodología

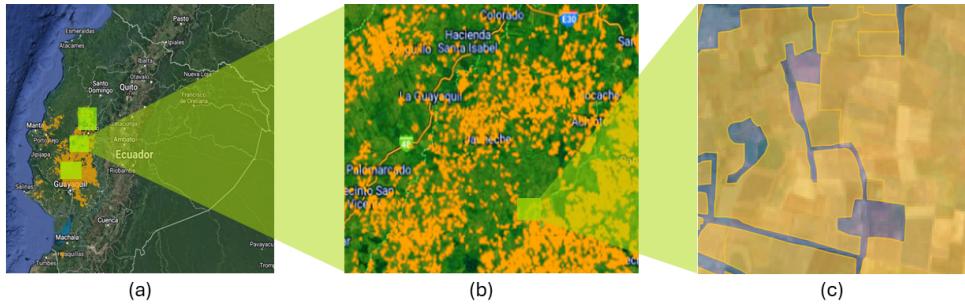


Figura 4.5: Procesamiento de datos de Ecuador: (a) Zonas escogidas del mapa de Ecuador con menor presencia de nubes, (b) Ejemplo de la zona 2 seleccionada, (c) Ejemplo de imagen (64, 64, 10) resultante del procesamiento.

ubicaciones sobre la superficie terrestre y garantizar que diferentes archivos geoespaciales se alineen correctamente. Se adoptó el sistema de referencia EPSG:3857 utilizado por GEE, lo que permitió estandarizar y alinear la georreferenciación entre imágenes y polígonos.

Posteriormente, una vez transformados, los polígonos fueron recortados para que coincidieran con las tres zonas seleccionadas, y posteriormente rasterizados para generar máscaras con las mismas dimensiones espaciales que las imágenes compuestas. En estas máscaras, cada píxel fue codificado con un valor entero según la clasificación. En la sección 4.2.6 se describe a mayor profundidad las máscaras obtenidas de los polígonos.

De forma análoga, las imágenes del conjunto de Ghana fueron procesadas utilizando los mismos criterios de bandas, formato espacial y clases seleccionadas. Este conjunto resultó en 203,102 sub-imágenes con sus respectivas máscaras, todas en el mismo formato de entrada (64, 64, 10). En contraste, el conjunto ecuatoriano logró generar únicamente 3,944 sub-imágenes, reflejando una notable desigualdad en la disponibilidad de datos geoespaciales entre ambos países.

Además de la diferencia en cantidad de muestras, existe una diferencia significativa en el área geográfica cubierta. Mientras que las tres zonas seleccionadas en Ecuador abarcan un total aproximado de 1,021.5 km², el conjunto de Ghana cubre una región estimada en más de 7,000 km², según las dimensiones del dataset público utilizado [14]. Esta diferencia no solo afecta la cantidad de datos generados, sino también la diversidad espacial y espectral disponible para el entrenamiento de modelos.

La Figura 4.6 presenta una comparación visual entre las regiones utilizadas en Ecuador y Ghana, evidenciando el alcance territorial y el volumen de datos disponibles en cada caso. Este contraste subraya uno de los principales desafíos de aplicar aprendizaje profundo en países con escasez de datos satelitales, datos etiquetados y libres de nubes, que restringe el desarrollo de modelos generalizables.

4.2.3 División de datos

Para garantizar una evaluación rigurosa de los modelos de clasificación, los datos se dividieron en tres conjuntos: entrenamiento (70%), validación (10%) y prueba (20%). Esta partición asegura que la evaluación se realice con información no vista durante el entrenamiento, siguiendo mejores prácticas en aprendizaje automático.

En el caso de los datos de Ecuador, la partición se realizó utilizando un criterio espacial estricto. Las zonas 1 y 3 se destinaron

4 Metodología

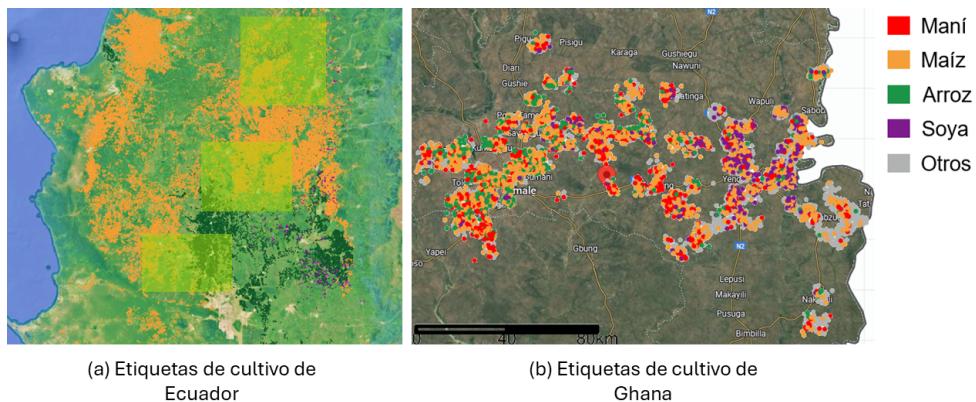


Figura 4.6: Comparación de etiquetas de cultivo: (a) Etiquetas de cultivos en Ecuador, (b) Etiquetas de cultivos en Ghana. Adaptado de [14].

a los conjuntos de entrenamiento y validación, respectivamente, mientras que la zona 2, que contenía una menor cantidad de datos y correspondía a una región geográficamente distinta, se reservó exclusivamente para el conjunto de prueba. De este modo, se asegura que las imágenes del conjunto de prueba provengan de una zona espacialmente independiente, lo que permite evaluar la capacidad de generalización del modelo ante regiones no vistas durante su entrenamiento. En total, el conjunto ecuatoriano quedó conformado por 2.770 imágenes para entrenamiento, 394 para validación y 780 para prueba.

En suma, si consideramos cada píxel como una etiqueta, la cantidad de etiquetas por clase se muestra en la Tabla 4.2, donde se aprecia claramente el desbalance de clases presente en el conjunto de imágenes de Ecuador. La clase mayoritaria corresponde a “no cultivo”, mientras que entre los cultivos objetivo, el maíz es el que cuenta con mayor representación. Esta distribución desigual de etiquetas resalta la importancia de considerar estrategias específicas para abordar el desbalance durante el entrenamiento y evaluación de los modelos de clasificación.

Por otro lado, los datos de Ghana fueron utilizados únicamente para Transferencia de Aprendizaje, por lo que el conjunto fue dividido en entrenamiento y validación en proporciones del 80% y 20%, respectivamente, resultando en 162.482 y 40.620 imágenes.

4.2.4 Obtención de índices espectrales

En la subsección 2.1.2, se mencionaron cuatro índices espectrales comúnmente utilizados en tareas de clasificación de cultivos, debido a su capacidad para resaltar características específicas de la vegetación, el contenido de humedad y las propiedades del suelo. Como parte del proceso de preprocesamiento de datos, se calcularon y añadieron cuatro bandas adicionales a cada matriz de entrada, correspondientes a los índices NDVI, SAVI, NDWI y BSI.

Estas nuevas bandas se integraron a las 10 bandas originales provenientes de Sentinel-2, generando así imágenes de tamaño (64, 64, 14) para ambos conjuntos de datos: Ghana y Ecuador. Esta ampliación del espaciopectral busca mejorar la capacidad de los modelos para distinguir entre distintos tipos de cultivos, al proporcionar información derivada que complementa las bandas originales.

4 Metodología

Tabla 4.2: Distribución de etiquetas en los conjuntos de entrenamiento, validación y prueba para cada escenario de clasificación.

Tipo de Clasificación	Train	Valid	Test
<i>(1) Binario</i>			
Clase 0 (no cultivo)	6,285,063	534,214	2,902,236
Clase 1 (cultivo)	5,060,857	107,9610	292,644
<i>(2) Categórica 1 cultivo (maíz)</i>			
Clase 0 (no cultivo)	6,285,063	534,214	2,902,236
Clase 1 (maiz)	4,352,536	961,898	184,484
Clase 2 (cultivos varios)	708,321	117,712	108,160
<i>(3) Categórica 1 cultivo (arroz)</i>			
Clase 0 (no cultivo)	6,285,063	534,214	2,902,236
Clase 1 (arroz)	609,768	110,468	101,992
Clase 2 (cultivos varios)	4,451,089	969,142	190,652
<i>(4) Categórica 1 cultivo (soya)</i>			
Clase 0 (no cultivo)	6,285,063	534,214	2,902,236
Clase 1 (soya)	98,553	7,244	6,168
Clase 2 (cultivos varios)	4,962,304	1,072,366	286,476
<i>(5) Categórica completa</i>			
Clase 0 (no cultivo)	6,285,063	534,214	2,902,236
Clase 1 (maiz)	4,352,536	961,898	184,484
Clase 2 (arroz)	609,768	110,468	101,992
Clase 3 (soya)	98,553	7,244	6,168

4.2.5 Escalado de datos

Mediante el proceso de preprocesamiento descrito en la subsección anterior, se obtuvieron matrices compuestas por 14 bandas diferentes. Dado que estas bandas provienen de distintos sensores y además incluyen índices espectrales derivados, es probable que presenten rangos de valores distintos. Por lo tanto, es necesario aplicar una función de escalado que ajuste los rangos de las bandas, con el fin de evitar sesgos en el entrenamiento del modelo.

Para solventar esto, las matrices obtenidas fueron escaladas usando la ecuación 3.17 de **escalado Min-Max**, obteniendo bandas sin pesos excesivos. Dado que los datos de Ghana se planean usar para transferencia de aprendizaje (TL), el escalado se realizó considerando los valores mínimos y máximos de dicho conjunto de datos. Es por esto que se aplica un escalado por dominio, donde los datos de Ecuador son transformados utilizando los valores extremos de Ghana para cada banda. Esta

4 Metodología

estrategia busca asegurar la compatibilidad entre ambos conjuntos en cuanto a rango de entrada para el modelo sin filtrar información sobre el área de interés con bajos recursos de datos.

La Figura 4.7 muestra los diagramas de caja de las bandas espectrales e índices del conjunto de datos de Ghana, que sirvió como base para el modelo principal. En esta figura se observan las distribuciones originales de cada variable, las cuales reflejan el comportamiento típico del ambiente y la vegetación en esa región.

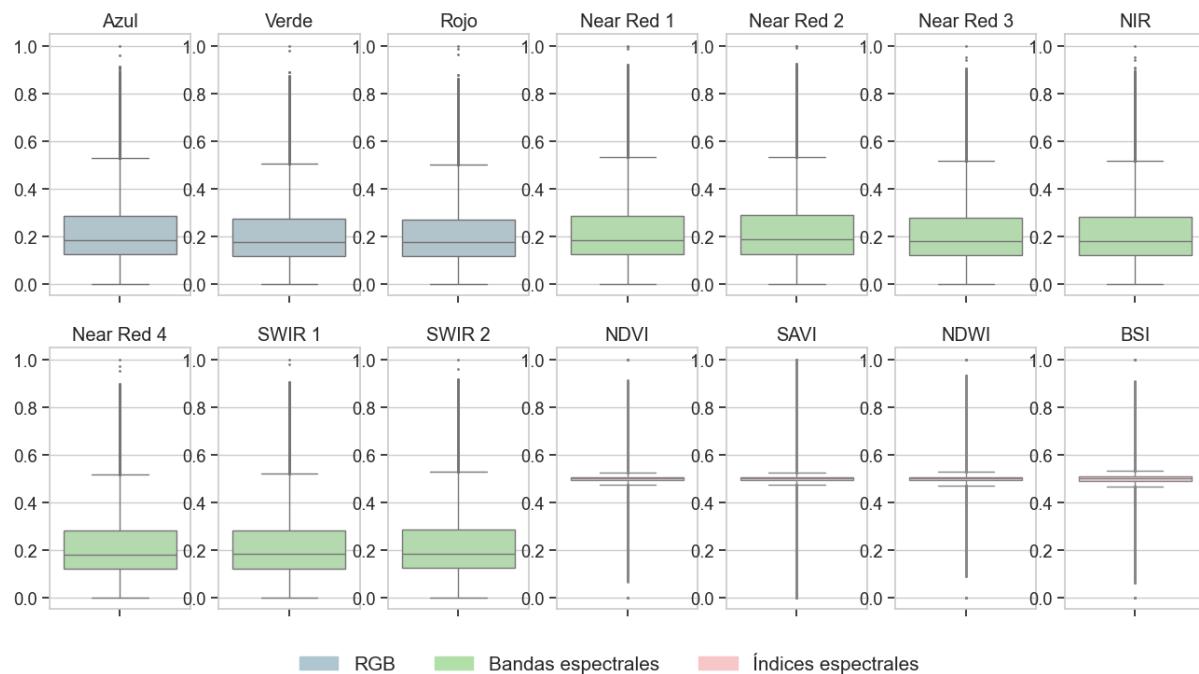


Figura 4.7: Diagrama de cajas de datos de Ghana. Se presentan las 14 bandas: RGB (azul), adicionales (verde), índices (rojo).

Tras aplicar este escalado por dominio, las variables del conjunto de Ecuador comparten el mismo rango de valores que las de Ghana. La Figura 4.8 muestra los diagramas de cajas resultantes de este escalado para Ecuador. Aunque ahora las variables están dentro del mismo rango, se evidencian diferencias relevantes en sus distribuciones. Estas diferencias pueden deberse a variaciones en la cobertura vegetal, el uso del suelo o condiciones atmosféricas distintas entre ambos países.

En términos ecológicos, tanto Ecuador como Ghana comparten la presencia de bosques tropicales húmedos, lo que justifica parcialmente el uso de técnicas de escalado entre dominios. Sin embargo, estudios recientes muestran que Ghana presenta también transiciones hacia sabanas y bosques [53], mientras que en Ecuador se encuentran zonas de bosques secos y variaciones altitudinales marcadas [54]. Estas diferencias en estructura vegetal, estacionalidad y cobertura pueden explicar las discrepancias observadas en las distribuciones espectrales, incluso tras normalizar las variables al mismo rango.

En general, Ecuador muestra menor variabilidad en las bandas RGB y SWIR en comparación con Ghana. Las distribuciones de bandas como Rojo, Verde y Azul están más concentradas en valores bajos, lo que podría reflejar diferencias en las condiciones atmosféricas, el tipo de vegetación o la composición del suelo. En el caso de los índices espectrales, como NDVI y SAVI, Ecuador tiende a mostrar valores más altos y más variables, lo que puede estar asociado a una mayor densidad vegetal. Por otro lado, NDWI muestra valores más bajos en Ecuador, lo que sugiere diferente presencia de cuerpos de agua en las imágenes.

4 Metodología

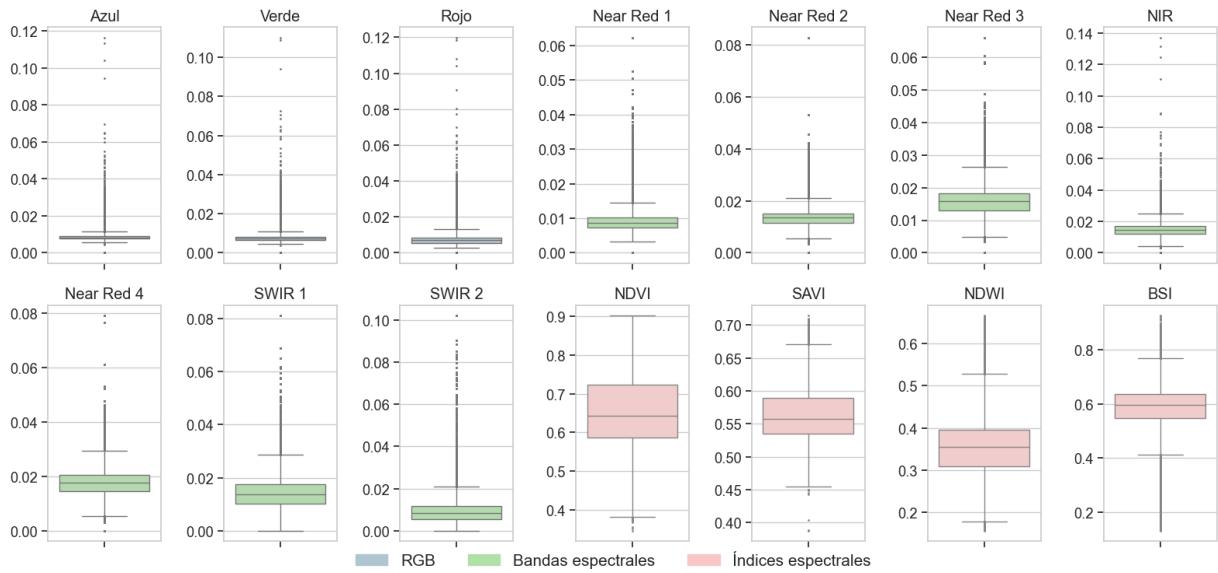


Figura 4.8: Diagrama de cajas de datos de Ecuador. Se presentan las 14 bandas: RGB (azul), adicionales (verde), índices (rojo).

escogidas. BSI, en cambio, mantiene una dispersión similar, aunque con ligeros desplazamientos en los valores centrales.

Estas diferencias indican que, a pesar del escalado por dominio, los datos de Ecuador no son completamente equivalentes en distribución a los de Ghana. Por ello, se implementó una estrategia de fine-tuning, en la que el modelo previamente entrenado con datos de Ghana fue ajustado utilizando un pequeño conjunto de datos ecuatorianos. En particular, se reutilizaron los pesos del encoder entrenado sobre Ghana y se refinaron con las muestras locales, permitiendo que el modelo se adapte a las particularidades espectrales y espaciales de Ecuador sin necesidad de ser entrenado desde cero. Esta técnica ayuda a que el modelo generalice mejor a un nuevo dominio, mitigando el impacto de los desajustes en las distribuciones, incluso si ambos países comparten ciertas condiciones climáticas o ecológicas.

4.2.6 Representación de máscaras

Una vez realizado el escalado, los datos de las imágenes están listos para ser utilizados en los modelos. No obstante, las máscaras de las imágenes requieren un cambio en su forma de representación. Actualmente, estas máscaras contienen valores enteros entre 0 a 4, donde cada número representa una clase: 0 para ausencia de cultivo, 1 para maíz, 2 para arroz, 3 para soya y 4 para cultivos varios. Sin embargo, estas etiquetas no se usan directamente, ya que su representación depende del tipo de tarea de clasificación.

En el caso de clasificación binaria, las máscaras se codifican usando solo dos valores: 0 para “*no cultivo*” y 1 para “*cultivo*”, sin distinguir entre tipos específicos. En la Figura 4.9 se muestra un ejemplo: una pequeña sección de la máscara original (Fig. 4.9.a) es transformada en una máscara binaria (Fig. 4.9.b), donde todos los píxeles correspondientes a cultivos (maíz, arroz, soya y cultivos varios) se agrupan en una sola clase.

Por otro lado, para tareas de clasificación multiclase, se utiliza una codificación **one-hot** (one-hot encoding). Esta codificación transforma cada clase original (0, 1, 2, 3, 4) en un vector binario. Por ejemplo:

4 Metodología

$$0 \rightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad 1 \rightarrow \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad 2 \rightarrow \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad 3 \rightarrow \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad 4 \rightarrow \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Dado que no estamos trabajando con valores individuales, sino con matrices completas, cada clase puede representarse como una máscara independiente. De esta manera, el conjunto de etiquetas puede expresarse como un tensor de dimensiones $(64, 64, C)$, donde ($C = 5$) es el número de clases. En la Figura 4.9.c se muestra cómo una sección de la máscara original se transforma en cinco máscaras binarias, correspondientes a las clases: no cultivo, maíz, arroz, soya y cultivos varios.

En algunos casos específicos, por ejemplo, si se desea realizar una clasificación focalizada en un solo tipo de cultivo, se puede simplificar la codificación usando únicamente tres máscaras: una para “no cultivo”, otra para el “cultivo seleccionado”, y una tercera para “cultivos varios”. Por ejemplo, si el modelo está enfocado únicamente en la detección de maíz, las tres máscaras resultantes representarían: “no cultivo”, “maíz”, y “cultivos varios”. Esta representación se muestra en la Figura 4.9.d, donde se reorganiza la máscara original para facilitar la clasificación focalizada.

Esta última representación (Figura 4.9.d) enfocada en un solo cultivo permite reorganizar las clases de forma que el modelo distinga claramente entre tres categorías: el cultivo de interés, otros cultivos y ausencia de cultivo. Esta simplificación facilita la especialización del modelo para tareas de detección específicas, como monitoreo de maíz o arroz, sin descartar completamente la información de los demás cultivos, agrupándolos bajo una única categoría denominada “cultivos varios”. Además, reduce la complejidad del problema manteniendo una representación multiclasé más manejable.

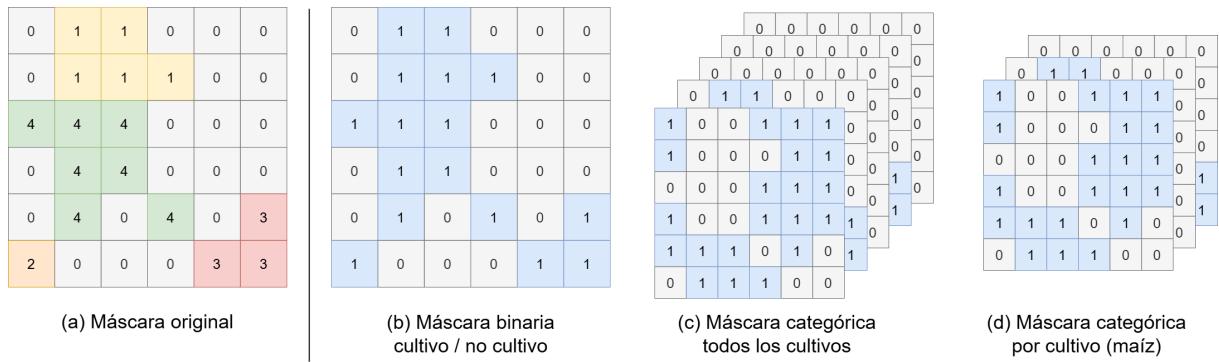


Figura 4.9: Codificación de máscaras. (a) Máscara original con valores de [0 a 4]; (b) Codificación binaria de cultivo / no cultivo; (c) Codificación one-hot para todos los cultivos a la vez, las matrices corresponden a (no cultivo, maíz, arroz, soya y cultivos varios); (d) Codificación one-hot por cultivo, las matrices corresponden a (no cultivo, maíz, cultivos varios).

4.3 Caso de Estudios

Con el objetivo de abordar el problema de clasificación de cultivos en Ecuador en condiciones de escasez de datos, se desarrollaron y evaluaron múltiples casos de estudio que combinaron distintos modelos de clasificación, configuraciones de datos

4 Metodología

de entrada y tipos de clasificación. Estas combinaciones permitieron explorar cómo cada decisión influía en el desempeño de los modelos, así como identificar las estrategias más eficaces para mitigar las limitaciones propias del contexto.

4.3.1 Modelos de Clasificación

Se establecieron tres grupos de modelos de clasificación con distintos niveles de complejidad:

- (a) **Modelos tradicionales:** Se seleccionó el algoritmo Random Forest, ampliamente utilizado en aplicaciones de teledetección debido a su robustez ante ruido, facilidad de interpretación y buen rendimiento en tareas de clasificación con conjuntos de datos limitados. Este modelo sirve como línea base para comparar el desempeño de arquitecturas más complejas.
- (b) **Modelos basados en CNN:** Se utilizó la arquitectura U-Net, la cual fue explicada en el Capítulo 3. Para aumentar la capacidad de representación, se exploraron diferentes backbones en la parte codificadora de la red: ResNet-50, Inception-V3 y EfficientNet-b3. Estos modelos permiten capturar patrones espaciales y espectrales de forma más profunda y son capaces de generalizar mejor en presencia de ruido o escasez de datos, especialmente al aplicar técnicas de transferencia de aprendizaje.
- (c) **Modelos complejos específicos:** Se implementó el modelo Presto, una arquitectura basada en Transformers recientemente propuesta para tareas generales de percepción remota. Presto no se limita únicamente a la clasificación de cultivos, sino que es un modelo preentrenado que puede adaptarse a diferentes tareas como segmentación, detección o clasificación, mediante el ajuste del decodificador correspondiente. En este trabajo, se empleó específicamente para una tarea de clasificación multiclas, aprovechando su capacidad para capturar variaciones espectrales y espaciales presentes en las imágenes satelitales.

4.3.2 Características de los datos

El tipo de entrada del modelo representa una de las variables más críticas en la tarea de clasificación. Por ello, se probaron distintas configuraciones para entender cómo cada banda e índices espectrales impactaron en la capacidad del modelo de discriminar clases. Las configuraciones probadas fueron las siguientes:

- (a) **Bandas Rojo, Verde, Azul (RGB):** Este escenario simula la información mínima disponible, correspondiente a imágenes ópticas estándar. Permite evaluar el desempeño cuando se dispone únicamente de información visual superficial.
- (b) **Bandas espectrales completas (10):** Se incluyen todas las bandas disponibles en los conjuntos de datos (en total, 10 bandas espectrales), con el objetivo de determinar si la inclusión de bandas adicionales (como el infrarrojo cercano, SWIR y red edge) mejora la capacidad discriminativa del modelo.
- (c) **Índices espectrales (index):** Se expandieron los conjuntos de datos calculando índices comúnmente utilizados en teledetección, como NDVI, SAVI, NDWI y BSI, que resaltan información espectral relacionada con la vegetación, la humedad y la estructura del suelo. Esta configuración permite evaluar si una representación compacta de la información espectral es suficiente o incluso preferible frente a las bandas originales.

4 Metodología

- (d) **Combinación de todas las bandas e índices (all)**: Se concatenaron todas las fuentes de información anteriores en una sola entrada, para analizar si una representación completa y rica en variables mejora los resultados o si introduce redundancia y ruido que afecta el entrenamiento.

4.3.3 Tipos de Clasificación

La forma de clasificar los cultivos también fue extendida para analizar cómo influye el enfoque de clasificación en el rendimiento general y la utilidad práctica de los mapas generados. Se realizó clasificación binaria y categórica en diferentes niveles de complejidad:

- (a) **Clasificación binaria**: Donde el modelo debe distinguir únicamente entre etiquetas de "*no cultivo*" y "*cultivo*". Esta tarea simplificada puede ser útil para generar máscaras de cultivo, detectar tierras productivas o filtrar áreas de interés.
- (b) **Clasificación categórica enfocada**: Donde se selecciona un cultivo de interés (por ejemplo, maíz), y el modelo debe distinguir entre "*no cultivo*", "*maíz*" y "*cultivos varios*". Este enfoque permite especializar el modelo en un cultivo a la vez para fines de monitoreo y planificación agrícola específica.
- (c) **Clasificación categórica completa**: Donde el modelo diferencia entre cinco clases: "*no cultivo*", "*maíz*", "*arroz*", "*soya*" y "*cultivos varios*". Este escenario busca obtener un mapa detallado y multifuncional de los cultivos previamente seleccionados.

En suma, se probaron diversas combinaciones entre las configuraciones de los casos de estudio. En la Tabla 4.3 se presentan dichas combinaciones. Cada X representa los cinco escenarios de tipos de clasificación:

1. Clasificación binaria,
2. Clasificación categórica enfocada en maíz,
3. Clasificación categórica enfocada en arroz,
4. Clasificación categórica enfocada en soya, y
5. Clasificación categórica completa.

En el caso del modelo Presto, su uso se limitó a escenarios que emplean únicamente bandas espectrales (RGB y las 10 bandas de Sentinel-2). Aunque el modelo admite la incorporación del índice NDVI como banda adicional [45], no permite la inclusión de otros índices espectrales. En este estudio se utilizaron cuatro índices (NDVI, SAVI, NDWI y BSI), cuya combinación con las bandas espectrales daba lugar a un total de 14 bandas de entrada. Dado que Presto no está diseñado para procesar este tipo de configuraciones extendidas, se excluyó de los escenarios que requerían exclusivamente índices o la combinación completa de bandas e índices.

4 Metodología

Tabla 4.3: Combinación de casos de estudio.

Modelos de Clasificación	RGB	10b	index	all	Nº Casos
Random Forest	X	X	X	X	20
U-Net + ResNet-50	X	X	X	X	20
U-Net + Inception-V3	X	X	X	X	20
U-Net + EfficientNet-b3	X	X	X	X	20
Presto	X	X			10
Total	25	25	20		90

4.4 Estrategias de entrenamiento

El proceso de entrenamiento de los modelos requiere de diferentes decisiones para garantizar la capacidad de generalización y el desempeño en la clasificación de cultivos a partir de imágenes satelitales. En esta sección se describen las estrategias adoptadas para el entrenamiento de los diferentes modelos empleados en el estudio, haciendo especial énfasis en la transferencia de aprendizaje y el ajuste de hiperparámetros.

4.4.1 Transferencia de aprendizaje

Dada la limitada disponibilidad de datos etiquetados en Ecuador, se implementó una estrategia de transferencia de aprendizaje para aprovechar el conocimiento adquirido a partir de un dominio fuente con mayor cantidad de datos (Ghana).

En primer lugar, se preentrenó los modelos base utilizando el conjunto de datos de Ghana, el cual cuenta con más de 200,000 imágenes etiquetadas de cultivos relevantes. Durante este preentrenamiento, el modelo aprendió a identificar patrones espectrales y espaciales generales asociados a la clasificación de cultivos en imágenes satelitales.

Posteriormente, se aplicó una fase de ajuste (fine-tuning) en el dominio objetivo ecuatoriano. Para ello, se reutilizaron los pesos preentrenados en Ghana y se reentrenaron las capas del modelo utilizando exclusivamente los datos ecuatorianos. Este ajuste se realizó con una tasa de aprendizaje reducida, lo que permitió adaptar el modelo a las particularidades fenológicas y espectrales de los cultivos locales, manteniendo al mismo tiempo la información relevante aprendida en el dominio fuente.

4.4.2 Ajuste de Hiperparámetros

Durante el proceso experimental se llevó a cabo un ajuste de hiperparámetros específico para cada uno de los modelos considerados, con el objetivo de mejorar su rendimiento. Si bien se exploraron distintos parámetros como profundidad de red, tasa de aprendizaje y tamaño de lote, se puso especial énfasis en la elección de la función de pérdida, dado su impacto directo en la calidad de la segmentación, especialmente en presencia de clases desbalanceadas o cultivos minoritarios.

4 Metodología

Random Forest

Para el modelo de Random Forest, se ajustaron hiperparámetros como el número de árboles (`n_estimators`) y la profundidad máxima del árbol (`max_depth`).

Modelos U-Net con backbones CNN

Para los modelos de tipo U-Net con distintas arquitecturas de backbone (ResNet-50, Inception V3, EfficientNet-B3), se realizó una búsqueda controlada sobre distintos hiperparámetros: la tasa de aprendizaje (`learning_rate`), el tamaño del lote (`batch_size`), y el número de épocas. Sin embargo, el aspecto más exhaustivo de la exploración fue la selección de la función de pérdida, considerada como un hiperparámetro crítico para tareas de segmentación con clases desbalanceadas.

Se planteó una prueba general para seleccionar una configuración y un backbone base con el fin de comparar las funciones de pérdida propuestas. Una vez seleccionado el mejor backbone, se procedió a evaluar ocho funciones de pérdida diferentes tanto para clasificación binaria como para clasificación categórica.

Las funciones evaluadas para clasificación binaria fueron: binary crossentropy, dice loss, jaccard loss, binary focal loss, bce jaccard loss, bce dice loss, binary focal dice loss y binary focal jaccard loss.

Para clasificación categórica se evaluaron: categorical crossentropy, dice loss, jaccard loss, categorical focal loss, cce jaccard loss, cce dice loss, categorical focal dice loss y categorical focal jaccard loss.

Los resultados de la comparación de las funciones de pérdida se presentan más adelante en la sección 5.1, y las configuraciones seleccionadas para los modelos se detallan en la sección 5.2.

Modelo Presto

El modelo Presto, adaptado de su implementación original propuesta en [45], fue configurado con sus hiperparámetros recomendados y ajustado usando datos ecuatorianos.

4.5 Ambiente de Implementación

La etapa de preprocesamiento, así como los procedimientos de entrenamiento, se llevaron a cabo utilizando las siguientes librerías para Python 3.8.20:

- **Tifffile** [55]: Esta biblioteca especializada en la lectura y escritura de archivos TIFF fue utilizada para cargar y manejar las imágenes multiespectrales en formato geoespacial.
- **TensorFlow** [56]: Un framework de aprendizaje profundo desarrollado por Google, se utilizó como base para construir y entrenar los modelos de redes neuronales convolucionales.
- **Pytorch** [57]: Un framework de código abierto para DL desarrollado por Meta AI. Fue utilizado en este trabajo para implementar modelos avanzados basados en Transformers, como Presto, debido a su flexibilidad, soporte dinámico de grafos computacionales y amplia compatibilidad con bibliotecas modernas de visión por computadora.

4 Metodología

- **Segmentation models** [58]: Esta librería, basada en Keras y TensorFlow, fue empleada para implementar arquitecturas modernas de segmentación semántica, como U-Net.
- **Scikit-learn** [59]: Se utilizó para tareas de evaluación de desempeño, validación cruzada, y procesamiento adicional, como el escalado de variables y la selección de métricas. Además, incluye modelos de aprendizaje automático como random forest.

Para implementar las diferentes arquitecturas y gestionar el procesamiento de datos, se utilizó un servidor con sistema operativo Ubuntu. El servidor cuenta con un procesador Intel(R) Xeon(R) CPU E5-2650 @ 2.30GHz, con 10 núcleos habilitados y 8 módulos de memoria DDR4 de 8 GB cada uno.

5

Resultados

En este capítulo se presentan los resultados obtenidos a partir de los experimentos realizados con los distintos modelos y configuraciones evaluadas. Se analizan los experimentos para seleccionar la mejor función de pérdida para nuestro caso específico de clasificación de cultivos en Ecuador. Además, se evaluó el desempeño de diferentes modelos mediante métricas estándar de clasificación y a través de mapas de predicción generados por los modelos.

Por último, se discute el impacto de las diferentes configuraciones de los casos de estudio, tipos de clasificación y el uso de estrategias como el transfer learning. Los resultados permitieron identificar las combinaciones más efectivas y comprender cómo varía el desempeño del modelo en función de la clase objetivo y sus limitaciones.

5.1 Selección de la función de pérdida

Como se explicó en la sección 4.4.2, se propuso una prueba general para seleccionar una configuración y un backbone base. En este caso, la prueba mostró que el backbone EfficientNet-B3 proporcionó los mejores resultados. Por lo tanto, se utilizó dicho backbone en la U-Net para comparar y seleccionar las mejores funciones de pérdida para clasificación binaria y categórica. A continuación, se detallan los resultados obtenidos.

5.1.1 Clasificación binaria

En la Tabla 5.1 se presentan los resultados obtenidos al comparar distintas funciones de pérdida para la tarea binaria, detalladas previamente en la Subsección 3.3.2. Para cada función, se reportan el número de épocas de entrenamiento (con *early stopping* activado), los valores de *loss* y *F1-score* tanto en el conjunto de entrenamiento como en el de validación, así como las diferencias absolutas entre ambos conjuntos (Δloss y $\Delta\text{F1-score}$).

Los resultados muestran que varias funciones de pérdida combinadas (filas 5 a 8) tienden a presentar mayores discrepancias entre entrenamiento y validación, especialmente en el f1-score. Esto sugiere que estas funciones pueden inducir sobreajuste, como se evidencia en el caso de *bce + dice loss* (5), que alcanza el f1-score más alto en entrenamiento (0.2352), pero cae drásticamente en validación (0.0851), con una diferencia de 0.1501 puntos.

Por otro lado, si bien la *binary crossentropy* (1) presenta la menor diferencia absoluta de f1-score (0.0008), su desempeño general es bajo, con valores de f1 menores a 0.09 en ambos conjuntos. De forma similar, *jaccard loss* (3) presenta buena

5 Resultados

Tabla 5.1: Resultados de experimentos binarios de función de pérdida.

Funciones de pérdida	épocas	train loss	train f1-score	valid loss	valid f1-score	Δ loss	Δ F1-score
(1) binary crossentropy	61	0.2836	0.0859	0.2922	0.0867	0.0086	0.0008
(2) dice loss	43	0.8297	0.1703	0.8329	0.1671	0.0032	0.0032
(3) jaccard loss	43	0.9046	0.1678	0.9062	0.1637	0.0016	0.0041
(4) binary focal loss	34	0.0217	0.1572	0.0305	0.1110	0.0088	0.0462
(5) bce + dice loss	39	1.0399	0.2352	1.2733	0.0851	0.2334	0.1501
(6) bce + jaccard loss	65	1.2291	0.1070	1.2423	0.1033	0.0132	0.0037
(7) binary focal + dice loss	69	0.8856	0.1522	0.9078	0.1297	0.0222	0.0225
(8) binary focal + jaccard loss	42	0.9099	0.2305	1.0045	0.0951	0.0946	0.1354

estabilidad, pero su f1-score es inferior al de otras funciones.

La función que mejor balance ofrece es *dice loss* (2). Esta opción no solo logra uno de los f1-score más altos en validación (0.1671), sino que además mantiene una diferencia mínima con respecto al entrenamiento, tanto en la pérdida (Δ loss = 0.0032) como en el f1-score (Δ F1 = 0.0032). Este resultado indica que el modelo generaliza bien y no sufre de sobreajuste, incluso bajo condiciones de datos limitados. La Figura 5.1 muestra la evolución de los valores de pérdida y f1-score durante el entrenamiento con esta función.

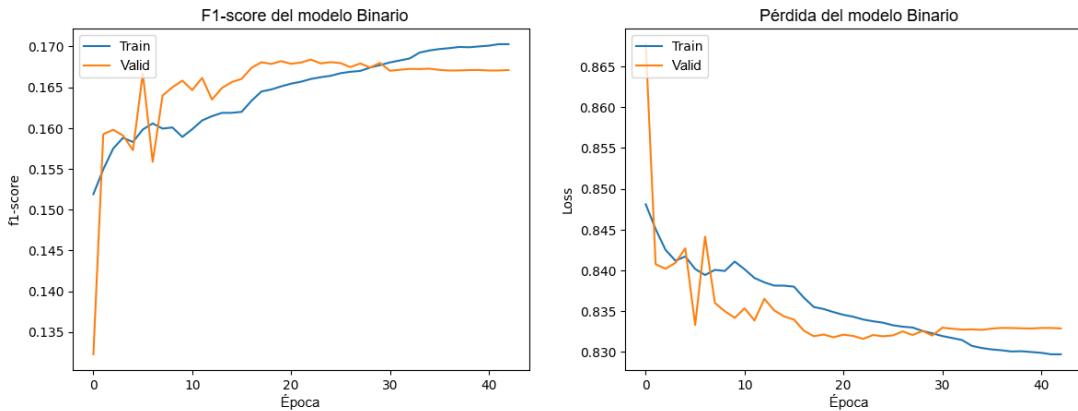


Figura 5.1: Pérdida y F1-score de función de pérdida dice loss para clasificación binaria.

5.1.2 Clasificación categórica

En la Tabla 5.2 se presentan los resultados obtenidos al comparar diferentes funciones de pérdida categóricas, descritas previamente en la Subsección 3.3.2. Al igual que en los experimentos binarios, el objetivo fue identificar la función de pérdida con mejor capacidad de generalización, es decir, aquella que presentara resultados consistentes entre entrenamiento y validación, minimizando el sobreajuste incluso con una cantidad limitada de datos.

Se observa que muchas funciones de pérdida combinadas (filas 5 a 8) presentan una diferencia considerable entre los conjuntos de entrenamiento y validación, en especial en el f1-score, lo que puede indicar una menor capacidad de generalización.

5 Resultados

Tabla 5.2: Resultados de experimentos categóricos de función de pérdida.

Funciones de pérdida	épocas	train loss	train f1-score	valid loss	valid f1-score	Δ loss	Δ F1-score
(1) categorical crossentropy	90	0.1182	0.3324	0.1119	0.3319	0.0063	0.0005
(2) dice loss	48	0.5056	0.4944	0.6674	0.3326	0.1618	0.1618
(3) jaccard loss	42	0.6932	0.3288	0.6997	0.3256	0.0065	0.0032
(4) categorical focal loss	48	0.0111	0.3146	0.0115	0.3079	0.0004	0.0067
(5) cce + dice loss	48	0.6524	0.4722	0.8127	0.3355	0.1603	0.1367
(6) cce + jaccard loss	61	0.8188	0.3388	0.8146	0.3346	0.0042	0.0042
(7) cat. focal + dice loss	39	0.5795	0.4577	0.7058	0.3349	0.1263	0.1228
(8) cat. focal + jaccard loss	39	0.6907	0.3976	0.7278	0.3355	0.0371	0.0621

Por ejemplo, *cce + dice loss* (5) alcanza un f1-score de entrenamiento elevado (0.4722), pero este cae en validación a 0.3355 (Δ F1-score = 0.1367).

Algo similar ocurre con *dice loss* (2), que si bien logra el f1-score más alto en entrenamiento (0.4944), presenta una caída considerable en validación (0.3326), lo que sugiere cierto grado de sobreajuste. Sin embargo, ambas funciones siguen mostrando resultados competitivos en validación, por lo que podrían ser útiles en contextos con mayor disponibilidad de datos o con regularización adicional.

La función que destaca por su estabilidad y buen desempeño general es **categorical crossentropy** (1). Esta mantiene los valores de pérdida y f1-score prácticamente constantes entre entrenamiento y validación, con diferencias mínimas de solo 0.0063 en loss y 0.0005 en f1-score. Este comportamiento evidencia una excelente capacidad de generalización del modelo, incluso con el mayor número de épocas evaluadas (90). A pesar de no ser una función especializada, su rendimiento competitivo demuestra que no siempre es necesario recurrir a funciones más complejas para obtener buenos resultados.

En la Figura 5.2 se presenta la evolución de la pérdida y el f1-score durante el entrenamiento con *categorical crossentropy*. Como se puede observar, la función muestra un aprendizaje progresivo sin signos de sobreajuste.

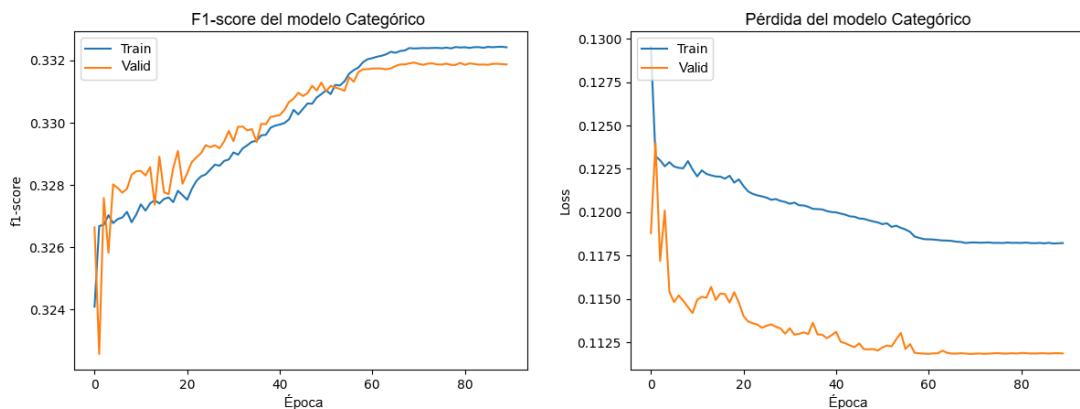


Figura 5.2: Pérdida y F1-score de función de pérdida categorical crossentropy loss para clasificación categórica.

5 Resultados

5.2 Configuración de modelos

En esta subsección se detallan las diferentes configuraciones seleccionadas luego de las pruebas iniciales, la selección de función de pérdida y el estado del arte. La Tabla 5.3 presenta la configuración de los hiperparámetros utilizados tanto para el modelo de Random Forest como para los modelos basados en redes neuronales convolucionales (CNN).

Tabla 5.3: Configuración de hiperparámetros de los modelos base.

Hiperparámetro	Valor asignado
<i>Random Forest</i>	
Número de árboles (<code>n_estimators</code>)	150
Ajuste de pesos de clase (<code>class_weight</code>)	balanced
Profundidad máxima (<code>max_depth</code>)	20
Mínimo de muestras por hoja (<code>min_samples_leaf</code>)	2
Mínimo número de muestras para división (<code>min_samples_split</code>)	2
<i>Modelos basados en CNN</i>	
Optimizador (<code>optimizer</code>)	Adam
Función de activación (<code>activation</code>)	ReLU (capas ocultas) y softmax (capa de salida)
Función de pérdida (<code>loss</code>)	dice loss (binaria) y crossentropy (categórica)
Tamaño de batch (<code>batch_size</code>)	64
Número máximo de épocas (<code>epochs</code>)	100 (con early stopping a 30)
Tasa de aprendizaje (<code>lr</code>)	0.0001

La selección de los hiperparámetros listados en la Tabla 5.3 se basó en una combinación de revisión de literatura y pruebas exploratorias iniciales. En el caso del modelo **Random Forest**, el número de árboles (150) y la profundidad máxima (20) se eligieron tras evaluar distintos valores en experimentos preliminares con datos reducidos, observando una mejora en la estabilidad del modelo y una reducción del sobreajuste. Estos valores también se encuentran dentro de rangos comúnmente reportados en trabajos similares en clasificación de cultivos.

Por otro lado, para los **modelos basados en CNN**, se optó por una tasa de aprendizaje baja (0.0001) y el uso del optimizador Adam, dado su buen desempeño en tareas con conjuntos de datos ruidosos y no balanceados. Estas configuraciones fueron validadas empíricamente mediante experimentación iterativa, con el objetivo de asegurar una convergencia estable sin sacrificar capacidad de generalización. Aunque no se realizó una búsqueda exhaustiva de hiperparámetros, los valores elegidos ofrecieron un buen compromiso entre rendimiento y costo computacional, permitiendo comparaciones justas entre los distintos modelos evaluados.

Finalmente, el modelo **Presto** se compuso de diferentes estrategias. La parte de decodificación, encargada de extraer información de las imágenes Sentinel-2, se mantuvo sin modificaciones, dado que fue preentrenada utilizando millones de imágenes provenientes de diversas fuentes, lo que garantizaba su capacidad para obtener características relevantes. La única parte que se pudo ajustar fue el clasificador Random Forest, el cual utilizó los datos resultantes de la codificación para realizar la clasificación o predicción.

5 Resultados

Como se trataba de un Random Forest, se emplearon los mismos hiperparámetros previamente definidos para asegurar una comparación justa y evaluar si esta configuración mejoraba o empeoraba el desempeño de la clasificación buscada.

5.3 Resultados de los experimentos

En esta sección se presentan los resultados obtenidos mediante el entrenamiento y evaluación de los modelos en los diferentes escenarios de clasificación detallados en secciones previas.

Los análisis se organizan en tres subsecciones: en la Subsección 5.3.1 se detallan los resultados del modelo Random Forest, incluyendo su desempeño en las configuraciones binarias y categóricas; la Subsección 5.3.2 aborda los modelos Unet con transferencia de aprendizaje implementados como una mejora a los modelos tradicionales; y la Subsección 5.3.3 expone los resultados del modelo Presto, el cual ha sido reportado en la literatura como una de las arquitecturas más prometedoras para problemas de teledetección.

Las métricas reportadas incluyen el accuracy y el F1-score, priorizando el F1-score macro. Este último se calcula como el promedio aritmético del F1-score de cada clase, tal como se especifica en la Ecuación 5.1:

$$F1\text{-score}_{macro} = \frac{\sum_{i=1}^n F1\text{-score}_i}{n} \quad (5.1)$$

donde n representa el número de clases del modelo. Esta métrica garantiza una evaluación equitativa entre clases, ya que asigna igual importancia al rendimiento del modelo en cada una de ellas, independientemente del número de muestras disponibles por clase. De esta forma, el F1-score macro permite evaluar si los modelos clasifican correctamente todas las clases, incluso aquellas con menor representación en el conjunto de datos.

5.3.1 Modelos Tradicionales

En esta subsección se detallan los resultados obtenidos al aplicar el modelo Random Forest (RF) para la clasificación de cultivos. La Tabla 5.4 muestra los valores de exactitud (accuracy) alcanzados en los distintos escenarios de clasificación, utilizando diferentes tipos de datos de entrada: bandas RGB, 10 bandas espectrales (10b), índices espectrales (index) y la combinación de todos ellos (all).

Tabla 5.4: Resultados de escenarios de modelo Random Forest (accuracy).

Tipos de Clasificación	RGB	10b	index	all
(1) Binario	0.66	0.82	0.81	0.67
(2) Categórica 1 cultivo (maíz)	0.65	0.82	0.81	0.67
(3) Categórica 1 cultivo (arroz)	0.65	0.82	0.81	0.66
(4) Categórica 1 cultivo (soya)	0.65	0.80	0.81	0.66
(5) Categórica completa	0.65	0.82	0.67	0.67

Aunque los valores de *accuracy* parecen elevados —especialmente al usar datos 10b e index— esta métrica puede resultar engañosa en presencia de clases desbalanceadas, como ocurre en nuestros datos. En estos casos, el modelo tiende a favorecer

5 Resultados

la clase mayoritaria (ausencia de cultivo), enmascarando un bajo desempeño en la identificación de cultivos minoritarios. Por ello, se recurrió al uso del F1-score macro como métrica principal, ya que pondera equitativamente todas las clases sin importar su frecuencia.

Este enfoque está respaldado por trabajos como [14], donde se advierte que la métrica de *accuracy* puede estar sesgada hacia los cultivos dominantes en conjuntos de datos agrícolas desbalanceados. Al comparar las Tablas 5.4 y 5.5, se observa que, aunque los valores de *accuracy* son altos y consistentes en escenarios que utilizan índices espectrales (*index*) y todas las bandas (*all*), el F1-score macro ofrece una visión más realista del rendimiento del modelo: muestra un buen desempeño en la clasificación binaria, pero un rendimiento deficiente en clasificaciones más específicas, particularmente en aquellas que incluyen clases minoritarias, como las configuraciones categóricas.

Tabla 5.5: Resultados de escenarios de modelo Random Forest (F1-score macro).

Tipos de Clasificación	RGB	10b	index	all
(1) Binario	0.64	0.79	0.78	0.40
(2) Categórica 1 cultivo (maíz)	0.42	0.55	0.54	0.32
(3) Categórica 1 cultivo (arroz)	0.42	0.55	0.54	0.32
(4) Categórica 1 cultivo (soya)	0.40	0.53	0.52	0.30
(5) Categórica completa	0.32	0.41	0.20	0.20

La Tabla 5.6 presenta el desglose detallado por clase del F1-score en los escenarios donde la precisión (*accuracy*) fue más alta según la Tabla 5.4. Se observa que el modelo RF logra un buen desempeño en la clase *no cultivo*, pero falla completamente al identificar cultivos como arroz o soya (F1 = 0.00 en varios escenarios). Incluso en configuraciones donde el maíz es parcialmente reconocido, el rendimiento global sigue siendo limitado. El escenario más desafiante —la clasificación categórica completa— evidencia un F1-score macro tan bajo como 0.20 al solo usar índices espectrales (*index*), lo cual indica una incapacidad del modelo para discriminar entre clases específicas.

En suma, estos resultados evidencian las limitaciones del modelo Random Forest ante clases desbalanceadas y múltiples categorías. Si bien el uso de bandas espectrales (10b) mejora levemente el rendimiento en algunas clases, persisten dificultades notorias para distinguir cultivos minoritarios. Esto pone de manifiesto que, para este tipo de tarea, pueden ser necesarias estrategias adicionales o enfoques distintos para mejorar la discriminación entre clases poco representadas.

5.3.2 Modelos basados en CNN

En esta subsección se presentan los resultados de los modelos propuestos basados en U-net. Primero, se intentó entrenar los modelos U-net solo con datos de Ecuador, los cuales muestran un rendimiento moderado en la tarea de clasificación de cultivos, evaluados mediante el F1-score macro (ver Tabla 5.7). En general, los valores obtenidos para la clasificación binaria son bajos, oscilando entre 0.10 y 0.15, independientemente del backbone empleado (EfficientNet B3, Inception V3 o ResNet 50) y del tipo de datos de entrada (RGB, 10 bandas, índices espectrales y todas las bandas combinadas). Esto indica que la cantidad de datos usados en el entrenamiento no es suficiente para que los modelos puedan distinguir correctamente entre presencia y ausencia de cultivos.

En el caso de la clasificación categórica para un solo cultivo, los resultados mejoran notablemente, alcanzando F1-scores

5 Resultados

Tabla 5.6: Comparación de F1-score por escenario de modelo Random Forest.

Tipo de Clasificación	10b	index
<i>(1) Binario</i>		
Clase 0 (no cultivo)	0.88	0.86
Clase 1 (cultivo)	0.77	0.70
f1-score macro global	0.79	0.78
<i>(2) Categórica 1 cultivo (maíz)</i>		
Clase 0 (no cultivo)	0.88	0.86
Clase 1 (maíz)	0.76	0.75
Clase 2 (cultivos varios)	0.00	0.00
f1-score macro global	0.55	0.54
<i>(3) Categórica 1 cultivo (arroz)</i>		
Clase 0 (no cultivo)	0.88	0.86
Clase 1 (arroz)	0.00	0.00
Clase 2 (cultivos varios)	0.76	0.76
f1-score macro global	0.55	0.54
<i>(4) Categórica 1 cultivo (soya)</i>		
Clase 0 (no cultivo)	0.88	0.86
Clase 1 (soya)	0.00	0.00
Clase 2 (cultivos varios)	0.71	0.70
f1-score macro global	0.53	0.52
<i>(5) Categórica completa</i>		
Clase 0 (no cultivo)	0.88	0.80
Clase 1 (maíz)	0.76	0.00
Clase 2 (arroz)	0.00	0.00
Clase 3 (soya)	0.00	0.00
f1-score macro global	0.41	0.20

entre 0.29 y 0.35. El rendimiento es similar entre los distintos backbones, aunque se observa una ligera ventaja al utilizar EfficientNet B3 y ResNet 50, especialmente para la soya, donde se logra el mayor F1-score (0.35). Para la clasificación categórica completa, los valores de F1-score macro disminuyen nuevamente, situándose entre 0.17 y 0.20, lo que sugiere que la complejidad de la tarea afecta negativamente el desempeño del modelo.

En cuanto al uso de diferentes datos de entrada, el uso de las 10 bandas espectrales o la combinación de todas las bandas tiende a mejorar ligeramente los resultados respecto al uso exclusivo de las bandas RGB, aunque las diferencias no son destacables. Esto indica que la información espectral adicional aporta valor, pero no es suficiente por sí sola para superar las limitaciones impuestas por los datos de entrenamiento o la arquitectura del modelo.

Una de las estrategias propuestas en la metodología fue emplear la transferencia de aprendizaje (TL) para mejorar el rendimiento de los modelos utilizando un conjunto de datos de mayor volumen. En este caso, primero se entrenó el modelo con

5 Resultados

Tabla 5.7: Resultados de los escenarios con modelos U-net entrenados en Ecuador (F1-score macro).

Tipos de Clasificación	RGB	10b	index	all
<i>U-net + Efficientnet b3</i>				
(1) Binario	0.15	0.13	0.10	0.11
(2) Categórica 1 cultivo (maíz)	0.30	0.33	0.33	0.33
(3) Categórica 1 cultivo (arroz)	0.30	0.33	0.33	0.33
(4) Categórica 1 cultivo (soya)	0.31	0.35	0.34	0.35
(5) Categórica completa	0.17	0.20	0.20	0.20
<i>U-net + Inception V3</i>				
(1) Binario	0.14	0.10	0.10	0.11
(2) Categórica 1 cultivo (maíz)	0.29	0.34	0.34	0.33
(3) Categórica 1 cultivo (arroz)	0.29	0.33	0.33	0.33
(4) Categórica 1 cultivo (soya)	0.32	0.35	0.34	0.34
(5) Categórica completa	0.18	0.20	0.20	0.20
<i>U-net + Resnet 50</i>				
(1) Binario	0.14	0.11	0.10	0.10
(2) Categórica 1 cultivo (maíz)	0.30	0.34	0.33	0.33
(3) Categórica 1 cultivo (arroz)	0.30	0.33	0.33	0.33
(4) Categórica 1 cultivo (soya)	0.32	0.34	0.34	0.34
(5) Categórica completa	0.18	0.20	0.20	0.20

el conjunto de Ghana y, posteriormente, se realizó un ajuste fino (fine-tuning) usando el conjunto ecuatoriano, que es de menor tamaño. En este proceso, no se congelaron los pesos de ninguna capa ni se modificó la arquitectura de la red; todo el modelo fue ajustado con los datos de Ecuador tal como estaba originalmente. Los resultados de esta estrategia de TL se muestran en la Tabla 5.8, donde se observa una tendencia general ligera de mejora en los F1-scores para la mayoría de las tareas de clasificación, en comparación con entrenar únicamente con datos ecuatorianos.

En el caso de la clasificación binaria, se observa un leve incremento en los valores, alcanzando hasta 0.17, particularmente al utilizar EfficientNet B3 en combinación con todas las bandas (all) como entrada. Este comportamiento sugiere que el preentrenamiento en un dominio distinto puede aportar cierto grado de generalización y robustez al modelo. No obstante, los resultados obtenidos aún no son lo suficientemente sólidos como para ser aplicados en una clasificación a gran escala.

En la clasificación categórica de un solo cultivo, los F1-scores se mantienen estables o ligeramente superiores en comparación con los modelos entrenados exclusivamente con datos de Ecuador, situándose entre 0.31 y 0.34. La diferencia más notable se observa en la clasificación categórica completa, donde los valores de F1-score macro alcanzan hasta 0.20.

Al comparar ambos escenarios, se evidencia que el preentrenamiento con datos de Ghana aporta beneficios modestos, pero consistentes en el desempeño de los modelos, particularmente en tareas de mayor complejidad como la clasificación categórica completa. No obstante, los F1-scores macro obtenidos en todas las configuraciones siguen siendo relativamente bajos. Esto sugiere que, aunque la técnica de transfer learning permitió una mejora, esta no fue significativa. Una posible

5 Resultados

Tabla 5.8: Resultados de escenarios de modelos U-net preentrenados con datos de Ghana (F1-score macro).

Tipos de Clasificación	RGB	10b	index	all
<i>U-net + Efficientnet b3</i>				
(1) Binario	0.16	0.14	0.16	0.17
(2) Categórica 1 cultivo (maíz)	0.33	0.33	0.32	0.33
(3) Categórica 1 cultivo (arroz)	0.32	0.32	0.33	0.32
(4) Categórica 1 cultivo (soya)	0.33	0.33	0.32	0.32
(5) Categórica completa	0.19	0.20	0.20	0.20
<i>U-net + Inception V3</i>				
(1) Binario	0.14	0.17	0.16	0.16
(2) Categórica 1 cultivo (maíz)	0.32	0.32	0.31	0.32
(3) Categórica 1 cultivo (arroz)	0.33	0.33	0.32	0.33
(4) Categórica 1 cultivo (soya)	0.32	0.32	0.32	0.33
(5) Categórica completa	0.20	0.20	0.20	0.20
<i>U-net + Resnet 50</i>				
(1) Binario	0.15	0.14	0.11	0.16
(2) Categórica 1 cultivo (maíz)	0.32	0.33	0.32	0.32
(3) Categórica 1 cultivo (arroz)	0.32	0.33	0.33	0.33
(4) Categórica 1 cultivo (soya)	0.32	0.34	0.32	0.33
(5) Categórica completa	0.19	0.20	0.20	0.20

explicación radica en las diferencias entre los datos de Ghana y Ecuador, que, pese a compartir cultivos y climas similares, presentan variabilidad espectral considerable.

Tal como se detalla en la subsección 4.2.5 sobre el escalado de datos, incluso al aplicar una normalización min-max, se observaron diferencias de tendencia entre dominios. Además, el uso de escalado por dominio (empleando los valores máximos y mínimos de Ghana para normalizar datos de Ecuador) no logró compensar dichas diferencias. A esto se suma la limitada cantidad de datos disponibles para entrenar los modelos U-Net con los backbones propuestos, lo cual afecta su rendimiento.

En el estado actual de los experimentos, el modelo Random Forest (RF) ha demostrado ser claramente superior frente a las arquitecturas U-Net con transferencia de aprendizaje, al menos bajo las condiciones y volúmenes de datos disponibles. Aunque se implementó una estrategia de transferencia de aprendizaje entrenando primero con datos de Ghana y luego ajustando con muestras de Ecuador, las mejoras observadas en los F1-scores fueron apenas marginales y, en general, insuficientes para aplicaciones prácticas exigentes.

En tareas de clasificación binaria, los modelos U-Net apenas alcanzan F1-scores macro de 0.16 a 0.17, mientras que RF logra valores mucho más altos (por ejemplo, 0.79 usando 10 bandas espectrales). Incluso en escenarios categóricos, los U-Net solo llegan a 0.33–0.34, y en la clasificación multiclas completa el F1-score se estanca alrededor de 0.20. Esto sugiere que, bajo condiciones de datos limitados y alta variabilidad espectral entre dominios, los modelos tradicionales como RF siguen siendo más robustos y confiables para la clasificación de cultivos en Ecuador.

5 Resultados

5.3.3 Modelos complejos específicos

A diferencia de las arquitecturas tradicionales basadas en convoluciones (como ResNet, Inception o EfficientNet), Presto introduce el uso de transformers al procesamiento de series temporales en teledetección. Los transformers, originalmente diseñados para tareas de lenguaje natural, destacan por su capacidad para modelar dependencias a largo plazo y relaciones complejas en secuencias de datos. Presto adapta estos principios al contexto de imágenes satelitales multitemporales, permitiendo capturar patrones temporales y espaciales relevantes para la clasificación de cultivos. Además, su diseño ligero y preentrenado facilita la transferencia a nuevos dominios con pocos datos etiquetados, manteniendo una alta eficiencia computacional y tiempos de entrenamiento reducidos. Esta combinación de ventajas lo diferencia de los modelos convolucionales, que suelen requerir mayor cantidad de datos y recursos para alcanzar desempeños comparables en tareas multitemporales y multiclasé.

Los resultados presentados en la Tabla 5.9 evidencian que el modelo Presto supera significativamente el desempeño de los modelos tradicionales en términos de F1-score macro, especialmente cuando se utilizan datos espectrales de 10 bandas. Por ejemplo, en la clasificación binaria, el modelo alcanza un F1-score macro de 0.84 con 10b, un valor que representa una mejora notable frente a los métodos previos. Esta tendencia se mantiene en escenarios más complejos, como la clasificación categórica completa, donde Presto logra un 0.59 con 10 bandas, superando ampliamente los 0.41 o menos que obtuvieron los modelos Random Forest.

Tabla 5.9: Resultados de escenarios de modelo Presto (F1-score macro).

Tipos de Clasificación	RGB	10b
(1) Binario	0.74	0.84
(2) Categórica 1 cultivo (maíz)	0.43	0.61
(3) Categórica 1 cultivo (arroz)	0.55	0.65
(4) Categórica 1 cultivo (soya)	0.55	0.58
(5) Categórica completa	0.34	0.59

La Tabla 5.10, que desglosa el F1-score por clase en cada escenario, revela ciertos patrones sobre el comportamiento del modelo Presto según los cultivos clasificados. En la clasificación binaria, el uso de 10 bandas espectrales no solo mejora la detección de cultivos en general (F1-score de 0.82 para la clase “cultivo”), sino que también mantiene una alta precisión en la clase “no cultivo”(0.86), demostrando una mejora significativa en detectar los cultivos clave seleccionados.

Sin embargo, al analizar escenarios categóricos específicos, existen diferencias críticas. Por ejemplo, en la clasificación de arroz (3), aunque el F1-score global alcanza 0.65 con 10b, la clase objetivo “arroz”solo logra 0.28, mientras que la categoría “cultivos varios”llega a 0.82. Esta diferencia sugiere que el modelo prioriza la agrupación de cultivos no objetivos bajo etiquetas genéricas, posiblemente debido a la baja representatividad del arroz en el conjunto de entrenamiento o a la similitud espectral con otras especies.

En el caso del maíz (2), el modelo muestra un comportamiento más equilibrado: la clase objetivo alcanza 0.82 con 10b, pero la categoría “cultivos varios”cae a 0.16, indicando una especialización efectiva para este cultivo cuando existe suficiente diferenciación espectral. No obstante, el escenario donde solo clasifica soya (4) presenta un escenario diferente: con 10b, la

5 Resultados

Tabla 5.10: Comparación de F1-score por escenario de modelo Presto.

Tipo de Clasificación		RGB	10b
<i>(1) Binario</i>			
Clase 0 (no cultivo)		0.82	0.86
Clase 1 (cultivo)		0.67	0.82
f1-score macro global		0.74	0.84
<i>(2) Categórica 1 cultivo (maíz)</i>			
Clase 0 (no cultivo)		0.68	0.86
Clase 1 (maíz)		0.60	0.82
Clase 2 (cultivos varios)		0.03	0.16
f1-score macro global		0.43	0.61
<i>(3) Categórica 1 cultivo (arroz)</i>			
Clase 0 (no cultivo)		0.68	0.86
Clase 1 (arroz)		0.37	0.28
Clase 2 (cultivos varios)		0.60	0.82
f1-score macro global		0.55	0.65
<i>(4) Categórica 1 cultivo (soya)</i>			
Clase 0 (no cultivo)		0.68	0.86
Clase 1 (soya)		0.36	0.07
Clase 2 (cultivos varios)		0.60	0.82
f1-score macro global		0.55	0.58
<i>(5) Categórica completa</i>			
Clase 0 (no cultivo)		0.69	0.64
Clase 1 (maíz)		0.61	0.55
Clase 2 (arroz)		0.06	0.56
Clase 3 (soya)		0.01	0.59
f1-score macro global		0.34	0.59

clase objetivo obtiene un F1-score de apenas 0.07, mientras que “*cultivos varios*” mantiene 0.82. Este resultado podría deberse a una superposición espectral entre la soya y otros cultivos en las bandas seleccionadas, o a una distribución geográfica particular de este cultivo en los datos de entrenamiento.

La clasificación completa (5) sintetiza estos desafíos: aunque el F1-score macro global mejora a 0.59 con 10b, los resultados por clase son heterogéneos. Mientras el maíz y la soya logran 0.55 y 0.59 respectivamente, el arroz se mantiene en 0.56, y la clase “*no cultivo*” muestra una caída inesperada a 0.64 (frente a 0.69 en RGB). Esta variabilidad subraya la sensibilidad del modelo a la composición espectral de cada cultivo y sugiere que, aunque Presto maneja mejor la complejidad que los métodos tradicionales, su eficacia depende críticamente de la calidad y representatividad de los datos de entrada. Además, la mejora consistente con 10 bandas refuerza la hipótesis de que la información espectral adicional permite al modelo discriminar contextos agrícolas que serían un reto identificar usando solo bandas RGB.

5 Resultados

La capacidad de Presto para generar mapas de clasificación precisos se puede ver en la Figura 5.3, donde se muestran los resultados de una clasificación binaria aplicada a una zona de Ecuador. Al comparar la imagen satelital original (Figura 5.3.a) con las etiquetas reales de presencia de cultivo obtenidas de la rasterización de las etiquetas de MAG (Figura 5.3.b) y las etiquetas predichas por Presto usando datos de 10 bandas (Figura 5.3.c), se aprecia una alta coincidencia visual, especialmente en la identificación de áreas cultivadas, consistente con el rendimiento del clasificador binario que alcanza hasta 84% de macro F1 score. Este resultado muestra que Presto puede detectar de manera efectiva las zonas agrícolas a partir de imágenes satelitales, lo que lo hace muy útil para tareas como el monitoreo y la gestión del uso del suelo en el sector agrícola.

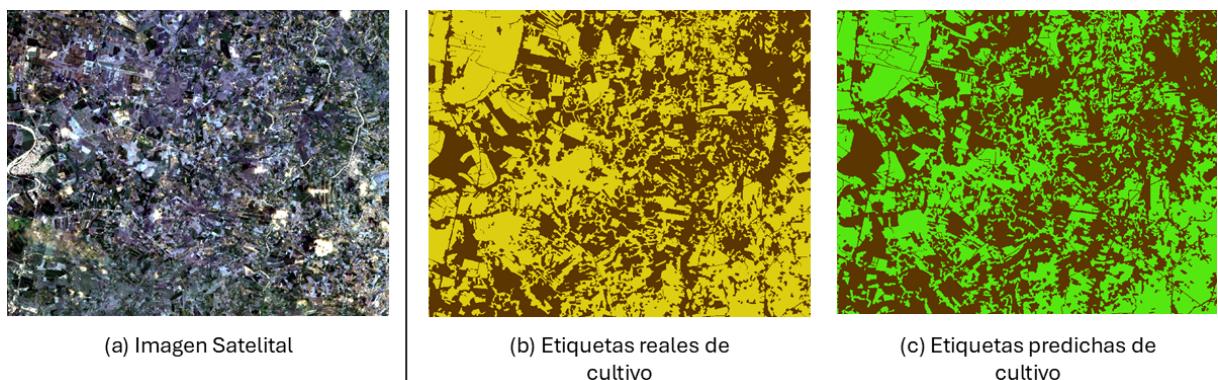


Figura 5.3: Clasificación de cultivos usando el modelo Presto: (a) Imagen Satelital de territorio ecuatoriano, (b) etiquetas reales de clase cultivo, (c) etiquetas predichas usando la clasificación binaria usando 10b.

Aunque el análisis cuantitativo, basado en el F1-score macro, evidenció desafíos en la clasificación de ciertos cultivos, la representación visual resalta la utilidad de Presto como herramienta para la detección general de áreas cultivadas. La clasificación logra identificar correctamente los cultivos, e incluso se observa cómo puede delimitar con mayor precisión las zonas agrícolas en comparación con las etiquetas manuales. Se nota que los cultivos de menor tamaño son los que presentan mayor dificultad para el modelo, mientras que las áreas más extensas son bien identificadas. Por ello, este modelo puede ser útil a nivel general para visualizar la distribución de los asentamientos agrícolas ecuatorianos de estos tres cultivos.

5.4 Comparativa de modelos y limitaciones

Los resultados obtenidos permiten comparar el rendimiento de los distintos modelos aplicados a la clasificación de cultivos, considerando tanto escenarios binarios como categóricos, así como sus requerimientos computacionales.

En primer lugar, el modelo **Random Forest** alcanza sus mejores desempeños en escenarios de clasificación binaria utilizando datos de 10 bandas e índices espectrales, con F1-score macro de 0.79 y 0.78 respectivamente. Sin embargo, su rendimiento disminuye considerablemente al aumentar la complejidad de la tarea, como ocurre en las clasificaciones categóricas, evidenciando limitaciones para diferenciar entre cultivos específicos.

En contraste, los modelos basados en **U-Net** no lograron mejorar ninguna tarea de clasificación, incluso al implementar técnicas de DL como la transferencia de aprendizaje. Esta última técnica solo logró mejorar ligeramente las tareas de clasificación categórica respecto al caso binario sin pre-entrenado usando datos de Ghana, aunque los resultados siguen siendo

5 Resultados

poco satisfactorios en términos generales. Dentro de este grupo, el modelo con mejor rendimiento relativo fue aquel que utilizó **EfficientNet-B3** como *backbone*, si bien las diferencias no fueron sustanciales y en ningún caso superaron a Random Forest.

El modelo **Presto** destaca por alcanzar los F1-score macro más altos entre todos los modelos evaluados, con valores entre 0.74 y 0.84. Este rendimiento es particularmente relevante al compararlo con resultados reportados en Ecuador usando métodos tradicionales y de deep learning: por ejemplo, estudios en la ciudad de Tungurahua reportan precisiones del 83,29% con técnicas de ML tradicionales y mejoras hasta 87,43% al aplicar modelos de DL [2]. Es importante resaltar que, en este trabajo, se observa un patrón y performance similar, pero en áreas mucho más extensas que las reportadas en dichos estudios previos.

Con respecto a los valores y modelo esperado, la literatura indica que valores de F1-score o accuracy superiores al 90% suelen alcanzarse únicamente en contextos con grandes volúmenes de datos balanceados y altamente representativos [32, 6, 41]. En escenarios de datos limitados o desbalanceados, como es común en Ecuador y en muchos países en desarrollo, los valores reportados para tareas multiclasificación suelen situarse entre 70% y 85% [5, 14, 7]. Por tanto, tanto los resultados de Presto como los de Random Forest pueden considerarse competitivos y alineados con la línea base internacional para este tipo de tarea.

En cuanto a la transferencia de aprendizaje (TL), se observa que el uso de modelos preentrenados no solo con miles, sino con millones de imágenes (como ocurre con modelos entrenados en ImageNet [41, 31]), permite una mayor capacidad de generalización. Este es precisamente el caso de Presto, que aprovecha un preentrenamiento masivo y logra adaptarse de forma eficiente a nuevos dominios con cantidades limitadas de datos etiquetados, superando el rendimiento de arquitecturas entrenadas desde cero o con menos datos de base.

En cuanto al uso de recursos, los modelos U-Net requieren una considerable capacidad computacional, tanto en tiempo de entrenamiento como en espacio de memoria, debido a que procesan imágenes como matrices completas. En contraste, los modelos RF y Presto trabajan a nivel de píxel, lo que facilita la partición de los datos en bloques y permite su procesamiento sin necesidad de cargar todo el conjunto en memoria. No obstante, esta estrategia genera un mayor número de muestras, lo que implica más iteraciones durante el entrenamiento.

Desde el punto de vista del tiempo de entrenamiento (ver Tabla 5.11), el modelo Random Forest (RF) resulta ser el más ligero y rápido, lo que se refleja en sus tiempos de entrenamiento que oscilan entre 14.6 minutos para la clasificación binaria con imágenes RGB y 45.6 minutos en el caso de la clasificación categórica completa utilizando todas las bandas. Este comportamiento evidencia que, aunque RF es eficiente, el tiempo de procesamiento aumenta proporcionalmente con la complejidad de la tarea y la cantidad de datos de entrada.

Por su parte, el modelo Presto, que incorpora un paso adicional de extracción de características mediante un codificador pre-entrenado antes de entrenar el Random Forest, presenta tiempos de entrenamiento ligeramente superiores. Por ejemplo, para la tarea binaria con imágenes RGB, Presto requiere 18.56 minutos, mientras que para la categórica completa con 10 bandas el tiempo asciende a 43.21 minutos. Este incremento se debe al procesamiento extra que implica la extracción de características, que en promedio añade alrededor de 4 minutos al tiempo total respecto al uso exclusivo de RF.

En contraste, los modelos basados en U-Net requieren tiempos de entrenamiento considerablemente mayores. Entre las variantes evaluadas, U-Net con Inception V3 destaca como la arquitectura más eficiente para la clasificación binaria con imágenes RGB, con un tiempo de 34.65 minutos. Sin embargo, a medida que aumenta la complejidad de la clasificación y

5 Resultados

Tabla 5.11: Tiempo de entrenamiento por modelo (minutos).

Tipos de Clasificación	RGB	10b	index	all
<i>Random Forest</i>				
(1) Binario	14.6	30.0	15.5	40.9
(2) Categórica 1 cultivo	14.7	36.6	20.8	41.9
(3) Categórica completa	15.2	42.1	36.3	45.6
<i>U-net + Efficientnet b3</i>				
(1) Binario	38.95	44.98	35.95	47.02
(2) Categórica 1 cultivo	45.11	49.57	44.39	53.71
(3) Categórica completa	51.52	53.62	50.44	87.54
<i>U-net + Inception V3</i>				
(1) Binario	34.65	43.51	37.92	79.23
(2) Categórica 1 cultivo	48.96	48.32	39.78	82.76
(3) Categórica completa	44.69	66.82	38.14	102.11
<i>U-net + Resnet 50</i>				
(1) Binario	38.08	36.22	35.61	58.22
(2) Categórica 1 cultivo	40.05	44.49	38.94	61.91
(3) Categórica completa	41.56	56.99	45.19	76.95
<i>Presto</i>				
(1) Binario	18.56	22.74	-	-
(2) Categórica 1 cultivo	18.6	31.76	-	-
(3) Categórica completa	19.5	43.21	-	-

la cantidad de bandas utilizadas, los tiempos pueden superar fácilmente los 80 minutos, como ocurre en la categorización completa con todas las bandas. Aunque los modelos U-Net ofrecen ventajas en precisión y capacidad de generalización, su mayor demanda computacional es un factor a considerar en aplicaciones a gran escala.

En resumen, **Presto** ofrece el mejor rendimiento general, especialmente en tareas binarias. Sin embargo, su desempeño decrece al abordar clasificaciones más detalladas, como las categóricas por cultivo. A pesar de haber sido preentrenado con millones de muestras de píxeles de tareas de teledetección, el modelo presenta restricciones relevantes: su arquitectura limita el uso de ciertos índices y bandas personalizadas.

6

Conclusión

El análisis exhaustivo de los resultados obtenidos con modelos tradicionales, arquitecturas CNN y el modelo Presto para la clasificación de cultivos en Ecuador permite extraer varios hallazgos:

1. La métrica de F1-score macro evidenció que la precisión global (accuracy) de estos modelos enmascara deficiencias críticas en la clasificación usando conjuntos de datos con clases desbalanceadas como entrenamiento. Si bien esto es conocido en tareas de ML de forma general, en el caso de datos escasos se vuelve un desafío más crítico.
2. Los métodos tradicionales como Random Forest muestran un desempeño aceptable en escenarios binarios y cuando se utilizan datos espectrales (10b), pero presentan limitaciones importantes en escenarios multiclase y en la identificación de cultivos minoritarios.
3. Las arquitecturas CNN, aunque aportan mejoras en la capacidad de modelar patrones espaciales y espectrales, requieren recursos computacionales elevados y su rendimiento no siempre compensa este costo, especialmente en contextos con datos limitados o alta variabilidad espectral, el cual fue el caso en este proyecto.
4. El modelo Presto emerge como una alternativa eficiente y competitiva, logrando F1-scores macro superiores en todos los escenarios, en particular cuando se emplean datos de 10 bandas espectrales. Presto demuestra una notable capacidad para mejorar la discriminación entre cultivos y reducir la dependencia de infraestructura computacional avanzada, lo que lo posiciona como una herramienta viable para aplicaciones operativas en agricultura de precisión. Sin embargo, incluso con este modelo, persisten desafíos en la clasificación de cultivos minoritarios y en escenarios categóricos completos, donde la heterogeneidad de los datos y la similitud espectral entre clases afectan la precisión alcanzable.

Estos resultados evidencian que la incorporación de información espectral adicional (más allá de las bandas RGB) es fundamental para mejorar la capacidad de discriminación de los modelos de clasificación, tanto tradicionales como avanzados. Desde una perspectiva práctica, la adopción de modelos eficientes como Presto puede facilitar la implementación de soluciones de monitoreo agrícola en regiones con recursos computacionales limitados, permitiendo una mayor escalabilidad y

6 Conclusión

acceso a tecnologías de agricultura de precisión.

6.1 Trabajo futuro

A partir de estos hallazgos, se identificaron varias líneas de trabajo para investigaciones futuras. En primer lugar, la ampliación y diversificación de los conjuntos de datos de entrenamiento son cruciales para mejorar la generalización y la capacidad de los modelos para identificar cultivos minoritarios. La integración de técnicas de aumento de datos, muestreo balanceado y generación sintética puede contribuir significativamente a mitigar el desbalance de clases y mejorar la robustez de los modelos.

En particular, el desarrollo de estrategias avanzadas de data augmentation, como transformaciones geométricas y enfoques de aprendizaje auto-supervisado[34], ha mostrado ser efectivo en la literatura reciente para incrementar la variabilidad del entrenamiento y reducir el sesgo hacia clases mayoritarias. Asimismo, el uso de modelos híbridos y arquitecturas especializadas continúa siendo una vía prometedora para abordar los desafíos inherentes a escenarios con datos desbalanceados y limitados [60].

Por otro lado, el desarrollo e incorporación de técnicas de explicabilidad e interpretabilidad de modelos (por ejemplo, visualización de mapas de atención o análisis de importancia de variables) permitirá comprender mejor las decisiones de los algoritmos y facilitar su adopción por usuarios finales y lograr mejorar la planificación y toma de decisiones en el sector agrícola.

Referencias

- [1] M. A. Lachaud and B. E. Bravo-Ureta, "Agricultural productivity growth in latin america and the caribbean: an analysis of climatic effects, catch-up and convergence," *Australian Journal of Agricultural and Resource Economics*, vol. 65, no. 1, pp. 143–170, 2021.
- [2] M. Castelo-Cabay, J. A. Piedra-Fernandez, and R. Ayala, "Deep learning for land use and land cover classification from the ecuadorian paramo." *International Journal of Digital Earth*, vol. 15, no. 1, pp. 1001–1017, 2022.
- [3] M. de Agricultura y Ganadería (MAG), "Geoportal del agro ecuatoriano," 2025. [Online]. Available: <http://geoportal.agricultura.gob.ec/index.php>
- [4] P. Prasad, V. J. Loveson, P. Chandra, and M. Kotha, "Evaluation and comparison of the earth observing sensors in land cover/land use studies using machine learning algorithms," *Ecological Informatics*, vol. 68, p. 101522, 2022.
- [5] A. Safanova, G. Ghazaryan, S. Stiller, M. Main-Knorn, C. Nendel, and M. Ryo, "Ten deep learning techniques to address small data problems with remote sensing," *International Journal of Applied Earth Observation and Geoinformation*, vol. 125, p. 103569, 2023.
- [6] G. Tseng, I. Zvonkov, C. L. Nakalembe, and H. Kerner, "CropHarvest: A global dataset for crop-type classification," in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.
- [7] F. Mena, D. Arenas, M. Nuske, and A. Dengel, "A comparative assessment of multi-view fusion learning for crop classification," in *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2023, pp. 5631–5634.
- [8] J. B. Campbell and R. H. Wynne, *Introduction to remote sensing*. Guilford press, 2011.
- [9] M. Mahmood, A. Shahzad, B. Zafar, A. Shabbir, and N. Ali, "Remote sensing image classification: A comprehensive review and applications," *Mathematical problems in engineering*, vol. 2022, no. 1, p. 5880959, 2022.
- [10] Esri, "ShapeFiles—ARCGIS Online | Documentation." [Online]. Available: <https://doc.arcgis.com/en/arcgis-online/reference/shapefiles.htm>
- [11] "About Copernicus | Copernicus." [Online]. Available: <https://www.copernicus.eu/en/about-copernicus>
- [12] "Sentinel 2 s2|processing," 2015. [Online]. Available: <https://sentiwiki.copernicus.eu/web/s2-processing>
- [13] "El apilamiento de índices de la imagen (NDVI, NDWI, NDSI)," 11 2023. [Online]. Available: <https://eos.com/es/make-an-analysis/index-stack/>
- [14] R. M Rustowicz, R. Cheong, L. Wang, S. Ermon, M. Burke, and D. Lobell, "Semantic segmentation of crop type in africa: A novel dataset and analysis of deep learning methods," in *Proceedings of the IEEE/cvf conference on computer vision and pattern recognition workshops*, 2019, pp. 75–82.
- [15] "JECAM – Joint Experiment for Crop Assessment and Monitoring." [Online]. Available: <https://jecam.org/>

6 Referencias

- [16] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep learning-based change detection in remote sensing images: A review," *Remote Sensing*, vol. 14, no. 4, p. 871, 2022.
- [17] T. Bai, L. Wang, D. Yin, K. Sun, Y. Chen, W. Li, and D. Li, "Deep learning for change detection in remote sensing: a review," *Geo-spatial Information Science*, vol. 26, no. 3, pp. 262–288, 2023.
- [18] J. Li, Y. Cai, Q. Li, M. Kou, and T. Zhang, "A review of remote sensing image segmentation by deep learning methods," *International Journal of Digital Earth*, vol. 17, no. 1, p. 2328827, 2024.
- [19] J. Xu, J. Yang, X. Xiong, H. Li, J. Huang, K. Ting, Y. Ying, and T. Lin, "Towards interpreting multi-temporal deep learning models in crop mapping," *Remote Sensing of Environment*, vol. 264, p. 112599, 2021.
- [20] A. Khan, A. D. Vibhute, S. Mali, and C. H. Patil, "A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications," *Ecological Informatics*, vol. 69, p. 101678, 2022.
- [21] O. Csillik, "Fast segmentation and classification of very high resolution remote sensing data using slic superpixels," *Remote Sensing*, vol. 9, no. 3, p. 243, 2017.
- [22] R. Reedha, E. Dericquebourg, R. Canals, and A. Hafiane, "Transformer neural network for weed and crop classification of high resolution uav images," *Remote sensing*, vol. 14, no. 3, p. 592, 2022.
- [23] H. R. Khan, Z. Gillani, M. H. Jamal, A. Athar, M. T. Chaudhry, H. Chao, Y. He, and M. Chen, "Early identification of crop type for smallholder farming systems using deep learning on time-series sentinel-2 imagery," *Sensors*, vol. 23, no. 4, p. 1779, 2023.
- [24] H. Xue, Y. Fan, G. Dong, S. He, Y. Lian, and W. Luan, "Crop classification in the middle reaches of the hei river based on model transfer," *Scientific Reports*, vol. 14, no. 1, p. 28964, 2024.
- [25] M. Choukri, A. Laamrani, and A. Chehbouni, "Use of optical and radar imagery for crop type classification in africa: A review," *Sensors*, vol. 24, no. 11, p. 3618, 2024.
- [26] M. Schneider, A. Broszeit, and M. Körner, "Eurocrops: A pan-european dataset for time series crop type classification," *arXiv preprint arXiv:2106.08151*, 2021.
- [27] P. Labs, "Sentinel hub." [Online]. Available: <https://www.sentinel-hub.com/explore/>
- [28] Copernicus, "Conventional data access hubs | copernicus." [Online]. Available: <https://www.copernicus.eu/en/access-data/conventional-data-access-hubs>
- [29] GEOSAT, "Geosat - observación de la tierra en vhr 2024," Mar 2024. [Online]. Available: <https://geosat.space/es/>
- [30] G. E. Engine, "Google earth engine," 2019. [Online]. Available: <https://earthengine.google.com/>
- [31] A. Alem and S. Kumar, "Transfer learning models for land cover and land use classification in remote sensing image," *Applied Artificial Intelligence*, vol. 36, no. 1, p. 2014192, 2022.

6 Referencias

- [32] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2217–2226, 2019.
- [33] L. P. Pott, T. J. C. Amado, R. A. Schwalbert, G. M. Corassa, and I. A. Ciampitti, "Satellite-based data fusion crop type classification and mapping in rio grande do sul, brazil," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 176, pp. 196–210, 2021.
- [34] Y. Wang, C. M. Albrecht, N. A. A. Braham, L. Mou, and X. X. Zhu, "Self-supervised learning in remote sensing: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 4, pp. 213–247, 2022.
- [35] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.
- [36] R. Prabhu, "Understanding of convolutional neural network (cnn)—deep learning," *Medium. Com*, vol. 1, 2018.
- [37] L. Lei, Q. Yang, L. Yang, T. Shen, R. Wang, and C. Fu, "Deep learning implementation of image segmentation in agricultural applications: A comprehensive review," *Artificial Intelligence Review*, vol. 57, no. 6, p. 149, 2024.
- [38] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.
- [39] G. J. Scott, M. R. England, W. A. Starns, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land-cover classification of high-resolution imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 4, pp. 549–553, 2017.
- [40] G. Tseng, H. Kerner, and D. Rolnick, "Timl: Task-informed meta-learning for agriculture," *arXiv preprint arXiv:2202.02124*, 2022.
- [41] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning earth observation classification using imagenet pretrained networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 105–109, 2015.
- [42] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 2010, pp. 270–279.
- [43] Z. Xiao, Y. Long, D. Li, C. Wei, G. Tang, and J. Liu, "High-resolution remote sensing image retrieval based on cnns from a dimensional perspective," *Remote Sensing*, vol. 9, no. 7, p. 725, 2017.
- [44] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 1126–1135.
- [45] G. Tseng, R. Cartuyvels, I. Zvonkov, M. Purohit, D. Rolnick, and H. Kerner, "Lightweight, pre-trained transformers for remote sensing timeseries," 2023.
- [46] J. C. Cevallos, J. A. Villagomez, and I. Andryshchenko, "Convolutional neural network in the recognition of spatial images of sugarcane crops in the troncal region of the coast of ecuador," *Procedia Computer Science*, vol. 150, pp. 757–763, 2019.

6 Referencias

- [47] A. Viña, F. R. Echavarria, and D. C. Rundquist, "Satellite change detection analysis of deforestation rates and patterns along the colombia–ecuador border," *AMBIO: A Journal of the Human Environment*, vol. 33, no. 3, pp. 118–125, 2004.
- [48] J. Terven, D.-M. Cordova-Esparza, J.-A. Romero-González, A. Ramírez-Pedraza, and E. Chávez-Urbiola, "A comprehensive survey of loss functions and metrics in deep learning," *Artificial Intelligence Review*, vol. 58, no. 7, p. 195, 2025.
- [49] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck, "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," *ISPRS journal of photogrammetry and remote sensing*, vol. 150, pp. 59–69, 2019.
- [50] L. Huang, B. Jiang, S. Lv, Y. Liu, and Y. Fu, "Deep-learning-based semantic segmentation of remote sensing images: A survey," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 8370–8396, 2023.
- [51] N. Liu, Q. Zhao, R. Williams, and B. Barrett, "Enhanced crop classification through integrated optical and sar data: a deep learning approach for multi-source image fusion," *International Journal of Remote Sensing*, vol. 45, no. 19-20, pp. 7605–7633, 2024.
- [52] M. de Telecomunicaciones y de la Sociedad de la Información, "Datos abiertos ecuador." [Online]. Available: <https://www.datosabiertos.gob.ec/>
- [53] L. Amissah, G. M. Mohren, F. Bongers, W. D. Hawthorne, and L. Poorter, "Rainfall and temperature affect tree species distribution in ghana," *Journal of Tropical Ecology*, vol. 30, no. 5, pp. 435–446, 2014.
- [54] X. Haro-Carrión, P. Waylen, and J. Southworth, "Spatiotemporal changes in vegetation greenness across continental ecuador: a pacific-andean-amazonian gradient, 1982–2010," *Journal of Land Use Science*, vol. 16, no. 1, pp. 18–33, 2021.
- [55] C. Gohlke, "tifffile: A library to read and write tiff files," <https://pypi.org/project/tifffile/>, 2022, accessed: 2024-10-01.
- [56] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous systems," <https://www.tensorflow.org/>, 2015, software available from tensorflow.org.
- [57] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," <https://pytorch.org/>, 2019, accessed: 2024-11-10.
- [58] P. Yakubovskiy, "Segmentation models," https://github.com/qubvel/segmentation_models, 2021, accessed: 2024-11-11.
- [59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [60] A. Özdemir, K. Polat, and A. Alhudhaif, "Classification of imbalanced hyperspectral images using smote-based deep learning methods," *Expert Systems with Applications*, vol. 178, p. 114986, 2021.

Archivos adjuntos

Anexo A

Para automatizar la ejecución de los experimentos con los distintos modelos (Random Forest, CNNs y Presto), se diseñaron scripts en Bash que coordinan la llamada a los scripts de Python responsables del entrenamiento, evaluación y generación de resultados. A continuación, se describe el uso de cada uno de estos scripts dentro del flujo de ejecución por lotes.

model_exp_rf.py (modelos tradicionales): Script dedicado a la experimentación con el modelo Random Forest. Permite configurar los siguientes parámetros mediante argumentos por línea de comandos:

- -- dataType : Tipo de datos de entrada (ej: rgb, 10b, index, all)
- -- cropType : Tipo de clasificación (ej: binary, maize, rice, soybean, all)

Durante la ejecución por lotes, este script se invoca repetidamente con distintas combinaciones de parámetros. Por ejemplo:

```
python model_exp_rf.py --dataType rgb --cropType binary
```

model_exp_cnn.py (modelos CNN): Utilizado para entrenar redes convolucionales, este script permite mayor personalización a través de los siguientes argumentos:

- -- modelType : Arquitectura de red (ej: efficientnetb3, inceptionv3, resnet50)
- -- dataType : Tipo de datos de entrada
- -- lossFunc : Función de pérdida (ej: dice_loss, categorical_crossentropy)
- -- cropType : Tipo de clasificación

Al igual que el anterior, este script se ejecuta múltiples veces variando los parámetros. Ejemplo de uso:

```
python model_exp_cnn.py --modelType efficientnetb3 --dataType rgb --  
lossFunc dice_loss --cropType binary
```

model_exp_presto.py (modelo Presto): Pensado para probar el modelo Presto, este script emplea los siguientes parámetros:

- -- dataType : Tipo de entrada (ej: rgb, 10b)
- -- cropType : Tipo de clasificación (ej: binary, maize, rice, soybean, all)

Su ejecución también forma parte del flujo por lotes. Ejemplo:

```
python model_exp_presto.py --dataType rgb --cropType binary
```

En total se ejecutaron 90 escenarios diferentes, combinando tanto las estrategias propuestas (Random Forest, tres CNN y Presto), los diferentes tipos de entrada de los modelos (RGB, 10b, index y all) y el tipo de clasificación.