

CLASIFICACIÓN DE LULC EMPLEANDO TÉCNICAS DE AGREGACIÓN DE CARACTERÍSTICAS EXTRAÍDAS DE MÚLTIPLES SENSORES REMOTOS

VERONICA HERRERA RUIZ

MAESTRÍA EN AUTOMATIZACIÓN Y CONTROL INDUSTRIAL

DIRECTORES:

MARÍA CONSTANZA TORRES MADROÑERO- JUAN DAVID MARTÍNEZ VARGAS-
JUAN CARLOS GONZÁLEZ VÉLEZ



Institución Universitaria

INSTITUTO TECNOLÓGICO METROPOLITANO
FACULTAD DE INGENIERÍAS
MEDELLÍN, 2024

CLASIFICACIÓN DE LULC EMPLEANDO TÉCNICAS DE AGREGACIÓN DE CARACTERÍSTICAS EXTRAÍDAS DE MÚLTIPLES SENSORES REMOTOS

Autor: Veronica Herrera Ruiz

Directores: María Constanza Torres Madroño, Juan David Martínez Vargas, Juan Carlos González Vélez

Facultad: Facultad de Ingenierías

Titulación: Maestría en Automatización y Control Industrial

Palabras clave: LULC, Clasificación supervisada, Teledetección, Agregación de características, Fusión de datos

Resumen

El agotamiento de los recursos naturales en busca de satisfacer las necesidades básicas de la población ha tenido un impacto devastador en el uso del suelo y la cobertura terrestre. Estudios sobre esto han generado extensas bases de datos que desempeñan un papel fundamental en los modelos ecológicos y socioeconómicos a nivel global. Los datos procedentes de plataformas espaciales, principalmente de tecnologías ópticas, pueden verse afectados por las condiciones atmosféricas, especialmente en áreas con una alta incidencia de nubes.

Ante este problema, surge la pregunta: ¿Cómo integrar datos capturados desde sensores ópticos y de radar para mejorar la estimación de LULC? Para abordar este desafío, en esta investigación se compararon dos enfoques de aprendizaje automático supervisado, Random Forest y XGBoost. Ambos modelos se aplicaron para clasificar el uso del suelo y la cobertura terrestre a partir de imágenes ópticas y de radar, probando diferentes escenarios de fusión de datos y niveles de nubosidad. Las pruebas se realizaron en un área de estudio específica en Colombia, un país que presenta altos niveles de nubosidad debido a su ubicación geográfica.

Los resultados mostraron que el modelo Random Forest, al combinar datos ópticos, radar e índices espectrales, alcanzó una mayor precisión en la clasificación de LULC. El mejor desempeño se obtuvo en el escenario con un 0-30 %

de nubosidad, logrando una precisión del 81.92 %, un F1-Score de 80.57 % y un coeficiente Kappa de 0.687. Estos hallazgos refuerzan la relevancia de integrar datos multisensoriales para mejorar la clasificación en condiciones atmosféricas adversas.

LULC CLASSIFICATION USING FEATURE AGGREGATION TECHNIQUES EXTRACTED FROM MULTIPLE REMOTE SENSORS

Author: Veronica Herrera Ruiz

Supervisor: María Constanza Torres Madroñero, Juan David Martínez Vargas,
Juan Carlos González Vélez

Department: Faculty of Engineering

Degree: Master's Degree in Automation and Industrial Control

Keywords: LULC, Supervised classification, Remote sensing, Feature aggregation, Data fusion

Abstract

The depletion of natural resources in the quest to meet the basic needs of the population has had a devastating impact on land use and land cover (LULC). Studies on this issue have generated extensive databases that play a fundamental role in global ecological and socioeconomic models. Data from space platforms, primarily optical technologies, can be affected by atmospheric conditions, especially in areas with a high incidence of clouds.

Given this problem, the question arises: How can data captured from optical and radar sensors be integrated to improve LULC estimation? To address this challenge, this research compared two supervised machine learning approaches, Random Forest and XGBoost. Both models were applied to classify land use and land cover using optical and radar images, testing different data fusion scenarios and cloud cover levels. The tests were conducted in a specific study area in Colombia, a country that experiences high cloud cover levels due to its geographical location.

The results showed that the Random Forest model, by combining optical data, radar data, and spectral indices, achieved higher accuracy in LULC classification. The best performance was obtained in the scenario with 0-30% cloud cover, achieving an accuracy of 81.92%, an F1-Score of 80.57%, and a Kappa coefficient of 0.687. These findings reinforce the relevance of integrating multisensor data to improve classification in adverse atmospheric conditions.

“Daría todo lo que sé por la mitad de lo que ignoro”

René Descartes

Agradecimientos

Al ver el resultado logrado de este proyecto, solo se me ocurre una palabra: ¡Gracias!

En primer lugar, a Dios.

A mis padres, quienes con cariño me han impulsado a perseguir mis metas, brindándome su apoyo incondicional para cumplir todos mis objetivos personales y académicos, superándome día a día.

A mi querido Jhei, quien con su fe inquebrantable en mí, incluso en los momentos más difíciles, ha sido el pilar de este logro. Por todo el amor que me das y la inspiración que me generas, gracias.

A Vale, mi incondicional hermana del alma, gracias por existir.

A mis amigos Sebas y Eider, quienes compartieron conmigo momentos de estrés y alegría durante este largo y retador camino.

A los profesores que, con su experiencia, paciencia y correcciones precisas, contribuyeron a mi formación en el complejo y gratificante camino de la investigación.

Gracias infinitas a cada uno de ustedes que, de una manera u otra, aportaron a mi fortaleza y ánimo, y, nuevamente, a Dios por ponerlos en mi camino.

Acrónimos

Asc	Ascendente
CLC	Corine Land Cover
CNN	Redes Neuronales Convolucionales
Desc	Descendente
ESA	Agencia Espacial Europea
GEE	Google Earth Engine
GRD	Ground Range Detected
IA	Inteligencia Artificial
IDEAM	Instituto de Hidrología, Meteorología y Estudios Ambientales
LULC	Uso y Cobertura del Suelo / Land Use Land Cover
ML	Machine Learning
NDBI	Índice de Diferencia Normalizada Edificada
NDVI	Índice de Vegetación de Diferencia Normalizada
NBRI	Índice de Calcinación Normalizado
NIR	Infrarrojo Cercano
NMDI	Índice de Humedad de Diferencia Normalizada
RF	Random Forest
SAR	Radar de Apertura Sintética
SVM	Máquinas de Vectores de Soporte

SWIR	Infrarrojo de Onda Corta
VV	Polarización Vertical de Transmisión y Recepción
VH	Polarización Vertical de Transmisión y Horizontal de Recepción
XGBoost	Extreme Gradient Boosting

Tabla de Contenido

Resumen	II
Abstract	IV
Agradecimientos	VI
Acrónimos	VII
1 Introducción	1
1.1 Justificación	1
1.2 Problema de Investigación	4
1.3 Objetivos	5
1.3.1 General	5
1.3.2 Específicos	5
1.4 Estructura del documento	5
1.5 Contribuciones	6
1.5.1 Publicaciones	6
1.5.2 Ponencias	6
2 Estado del Arte y Marco Teórico	7
2.1 Uso y cobertura del suelo / Land Use Land Cover	7
2.2 Teledetección y Sensores	8
2.3 Atributos en la adquisición	12
2.4 Técnicas de fusión	14

2.5	Procesamiento de imágenes	16
2.6	Modelos de Clasificación	18
3	Metodología	21
3.1	Dataset	23
3.1.1	Área de estudio	23
3.1.2	Selección de imágenes	24
3.2	Extracción y Selección de Características	29
3.2.1	Índices Ópticos	30
3.2.2	Índices SAR	32
3.3	Preprocesamiento	33
3.3.1	Normalización	33
3.3.2	Fusión de Datos	34
3.3.3	Validación	35
3.4	Clasificación Supervisada en Imágenes Ópticas y de Radar	36
3.5	Evaluación de la Clasificación LULC	37
3.6	Métricas para validación del rendimiento	38
4	Resultados y Discusión	42
4.1	Dataset	42
4.2	Clasificación Supervisada en Imágenes Ópticas y de Radar	44
4.3	Evaluación de la Clasificación LULC	50
	Conclusiones y Trabajo Futuro	60
	Bibliografía	80

Índice de figuras

2.1	Teledetección y Sensores. Fuente: Creado por la autora.	11
2.2	Fusión por nivel de pixel. Fuente: Creado por la autora.	15
2.3	Fusión por nivel de características. Fuente: Creado por la autora. .	15
2.4	Fusión por nivel de decisión. Fuente: Creado por la autora.	15
3.1	Metodología general. Fuente: Creado por la autora.	22
3.2	Área de estudio. Fuente: Creado por la autora.	24
3.3	Mapa y distribución de las clases en el CLC. Fuente: Creado por la autora.	28
3.4	Mapa y distribución de las clases en el WorldCover. Fuente: Creado por la autora.	29
3.5	Modalidad de Evaluación de la Clasificación LULC. Fuente: Creado por la autora.	38
3.6	Matriz de Confusión. Fuente: Creado por la autora.	39
4.1	Distribución del set de entrenamiento y validación por clases y mapa de referencia. Fuente: Creado por la autora.	44
4.2	Matriz de Confusión de Modelo Random Forest y conjunto de datos Óptico + SAR + Índices Completos. Fuente: Creado por la autora.	46
4.3	Comparativo de la Imagen real del área de estudio óptica en un rango del 0 al 30 % de nubes, con el mapa de referencia CLC y la imagen de la clasificación del modelo. Fuente: Creado por la autora.	48
4.4	Comparativo con acercamiento en dos sectores de la imagen real del área de estudio óptica en un rango del 0 al 30 % de nubes, con el mapa de referencia CLC y la imagen de la clasificación del modelo. Fuente: Creado por la autora.	49

4.5	Matriz de Confusión de Modelo Random Forest y conjunto de datos Óptico + SAR + Índices Completos en escenario de 0 - 30 % de nubes. Fuente: Creado por la autora.	53
4.6	Matriz de Confusión de Modelo Random Forest y conjunto de datos Óptico + SAR + Índices Completos en escenario de 70 - 90 % de nubes. Fuente: Creado por la autora.	54
4.7	Importancia de las Características para el Modelo Random Forest y conjunto de datos Óptico vs Óptico + SAR + Índices Completos en escenario de 0 - 30 % de nubes. Fuente: Creado por la autora.	56
4.8	Importancia de las Características para el Modelo Random Forest y conjunto de datos Óptico vs Óptico + SAR + Índices Completos en escenario de 70 - 90 % de nubes. Fuente: Creado por la autora.	57
4.9	Comparativo de la imagen real del área de estudio óptica en un rango del 70 al 90 % de nubes, con el mapa de referencia World-Cover y la imagen de la clasificación del modelo. Fuente: Creado por la autora.	58

Índice de Tablas

2.1	Características y especificaciones de diferentes sensores	10
2.2	Índices ópticos y SAR utilizados en aplicaciones de clasificación de LULC.	18
3.1	Atributos de las bandas de Sentinel 2	25
3.2	Detalles de los datos Ópticos y SAR	26
3.3	Reclasificación del Nivel 2 de CLC	27
3.4	Composición de los conjuntos de datos	35
3.5	Parámetros utilizados para los modelos	37
4.1	Base de datos según las bandas de las imágenes y el número de píxeles	43
4.2	Resultados de Clasificación Supervisada con Imágenes SAR y Ópticas del 0 al 30% de nubes y mapa CLC	45
4.3	Resultados de Clasificación Supervisada con modelo RF e Imágenes SAR y Ópticas en diferentes rangos de nubes	51

Capítulo 1

Introducción

Contenido

1.1	Justificación	1
1.2	Problema de Investigación	4
1.3	Objetivos	5
1.3.1	General	5
1.3.2	Específicos	5
1.4	Estructura del documento	5
1.5	Contribuciones	6
1.5.1	Publicaciones	6
1.5.2	Ponencias	6

1.1. Justificación

Los recursos naturales en el mundo son y han sido aprovechados por el ser humano de tal manera que se han convertido en una fuente de vida y desarrollo, con el fin de satisfacer las necesidades de subsistencia, como alimento, salud, economía y ocio [1]. Sin embargo, durante los últimos años, la tendencia de derroche y sobreexplotación de los recursos naturales ha generado graves consecuencias para el planeta. El desarrollo económico ha impuesto un alto costo ambiental sobre la capacidad y la salud del mundo natural, debido a las presiones generadas por las actividades humanas sobre el territorio, sus atributos y recursos [2]. Un claro ejemplo de esto son los bosques, que se enfrentan a numerosas perturbaciones que pueden afectar negativamente su vitalidad y reducir

su capacidad para proporcionar una amplia gama de bienes y servicios ecosistémicos [3]. Algunas de las actividades humanas que han llevado al deterioro de los bosques han sido la expansión de la frontera agrícola, la ganadería, la infraestructura y la minería; además de otras causas indirectas asociadas a los cambios sociales, políticos y económicos [4].

La región de América Latina y el Caribe alberga 23 % de los bosques del mundo [5]. Cuenta con una relación entre el grado de cobertura boscosa y el territorio de los países sumamente variable según su tamaño y las condiciones ecosistémicas propias, por lo que al año 2020, diecisiete países de esta región contaban con más del 40 % de su superficie cubierta por bosques [6]. Sin embargo, esta región también ha sufrido el impacto de la deforestación y los cambios en la cobertura del suelo, lo que ha contribuido a la actual pérdida de 2,6 millones de hectáreas de bosque cada año en el período 2010-2020, posicionándola como la segunda tasa más alta del mundo después de África.

En el 2021, Colombia contaba con una extensión de bosques de 59,8 millones de hectáreas, gran parte de esta superficie se encuentra en la región Amazónica, además de una presencia significativa en las regiones Andina y Pacífico [7,8]. En cuanto a las consecuencias negativas, se ha registrado que uno de los picos de deforestación más elevados en Colombia ocurrió en 2017, cuando las cifras del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) señalaron la pérdida de casi 220 mil hectáreas en el país. En el balance más reciente presentado por el Ministerio de Ambiente se mostró un nuevo incremento del 8 % en el monitoreo del 2020, donde el 79 % de esta deforestación se concentró en los departamentos de Meta, Caquetá, Guaviare, Putumayo y Antioquia [9]. El apoderamiento de tierras, las malas prácticas de ganadería extensiva, la infraestructura de transporte no planificada, los cultivos de uso ilícito, la extracción ilícita de minerales, la tala ilegal y la ampliación de la frontera agrícola en áreas no permitidas son las principales causas de este desastre [10]. Esto plantea la necesidad de monitorear y controlar estas variaciones en el uso del suelo.

De esta manera, el cuidado y evolución de los recursos naturales en el mundo ha despertado un gran interés en la comunidad científica en la generación de información y conocimiento sobre el uso del suelo, la cobertura terrestre y sus cambios a través del tiempo [11]. Es así como en la actualidad se encuentran iniciativas que han impulsado a evidenciar estos cambios como un factor de suma importancia, con el fin de generar conocimiento sobre el estado actual del territorio, realizar evaluaciones y producir diagnósticos de los cambios que también se pueden esperar en el futuro [11, 12]. La dinámica del uso de la tierra se presenta como resultado de la interacción de los factores y las consecuencias en los ecosistemas y la sociedad, destacándose como un indicador clave para

desarrollar herramientas que respalden y apoyen el análisis de escenarios para la planificación y políticas de estos sectores [13, 14].

Los cambios en el uso y la cobertura del suelo (LULC, por sus siglas en inglés Land Use Land Cover) tienen un impacto significativo en la interacción entre la superficie terrestre y la atmósfera, lo que resulta en alteraciones que afectan a más del 70 % de la superficie terrestre [15]. Estos cambios desencadenan diversos procesos antropogénicos que pueden tener consecuencias ambientales y climáticas significativas. Por esta razón, es fundamental desarrollar metodologías que permitan detectar y monitorear estos cambios a lo largo del tiempo. Los análisis multitemporales de la cobertura del suelo y las métricas del patrón espacial son herramientas clave en este sentido [4, 16].

Los análisis multitemporales brindan información sumamente relevante sobre las etapas fenotípicas de la vegetación, lo que es útil para procesos de clasificación, especialmente en áreas propensas a la variación estacional [14]. En análisis como este, donde se realiza una comparación entre las coberturas interpretadas por imágenes satelitales, fotografías aéreas o mapas de un sector en diferentes periodos de tiempo, el potencial de estas metodologías se puede ver seriamente afectado por la calidad y confiabilidad de los datos [17]. Ya que esta información se adquiere por medio de sensores ópticos, es importante mencionar que estos tienen algunas limitaciones, en particular, la recopilación de datos puede verse afectada por las condiciones atmosféricas en el sector y el momento de la toma, especialmente en áreas que tienen nubes la mayor parte del tiempo [11, 18].

Múltiples estudios han evaluado la integración de diferentes técnicas con las que es posible garantizar que este factor de riesgo pueda ser mitigado. Una de las técnicas que permite mejoras sustanciales en los datos es el uso de radares, los cuales pueden tomar imágenes de la superficie de la Tierra sin importar las circunstancias atmosféricas, operando en condiciones nubladas e incluso en la oscuridad [19–21]. La necesidad de implementar esto en las metodologías actuales es imperativa, ya que puede llenar los vacíos de información que en estos momentos no permiten mejorar los resultados en la precisión de los monitoreos y la clasificación de LULC. Adicional a esto, las imágenes de radar específicamente los de apertura sintética, pueden proporcionar información sobre diferentes parámetros biofísicos y geofísicos de los objetivos de cobertura terrestre, ampliando las posibilidades de recopilación de datos para múltiples propósitos, dando como resultado en la fusión de estas técnicas una percepción más completa gracias a las variaciones en resolución espectral, espacial y temporal [22].

1.2. Problema de Investigación

A medida que las investigaciones en LULC han tomado relevancia, se han aprovechado en gran medida los abundantes conjuntos de datos que han surgido hasta el momento. Estos contribuyen esencialmente en el desarrollo de modelos ecológicos y socioeconómicos a escala local, regional y mundial. Cada conjunto de datos se elabora utilizando diferentes fuentes, algoritmos de clasificación, metodologías e incluso características espaciales [18]. En la mayoría de los casos se presentan algunas áreas de incertidumbre relacionadas con diversas condiciones. Mediante el uso de sensores ópticos, la recopilación de datos puede verse obstaculizada por las condiciones atmosféricas en el sector y momento de la toma, especialmente en áreas que tienen nubes la mayor parte del tiempo, como lo es Colombia [23].

Buscando mitigar los factores de riesgo mediante los beneficios ofrecidos por otros sistemas, han surgido nuevas investigaciones donde se plantea la integración de diversas técnicas. En consecuencia, se ha explorado el uso de datos de radar, que tienen la capacidad de captar información de la superficie de la Tierra bajo condiciones atmosféricas específicas, a diferencia de otros sistemas con los que esto no es posible [19–21]. Sin embargo, estas exploraciones no han tenido gran relevancia en nuestro país, a pesar de contar con un gran volumen de datos y la evidente necesidad de renovar nuestra metodología actual basada en la clasificación visual y manual de imágenes satelitales multitemporales liderada por el IDEAM [24].

Es así como, contemplando la revisión bibliográfica y la ecuación de búsqueda realizada en mayo de 2024 en la base de datos SCOPUS: ((hyperspectral OR multispectral) AND (remote AND sensing) AND (lulc OR land AND use OR land AND cover)), se encontró un total de 3619 documentos relacionados. Sin embargo, al agregar la condición (AND (colombia)), este resultado disminuye drásticamente a un total de 5 documentos, ratificando que en Colombia no se han registrado actualmente un número significativo de estudios relacionados con esta necesidad. Por lo tanto, es de suma importancia iniciar la exploración de las técnicas de fusión que faciliten la creación de nuevos mapas de clasificación y aplicaciones para el monitoreo de LULC en el país.

Atendiendo a la problemática planteada anteriormente, la pregunta de investigación es: ¿Cómo integrar datos capturados desde sensores ópticos y de radar para mejorar la estimación de LULC?

1.3. Objetivos

1.3.1. General

Desarrollar una metodología de integración de imágenes capturadas por múltiples sensores remotos, ópticos y radar, empleando técnicas de agregación de características para la clasificación de LULC.

1.3.2. Específicos

- Diseñar una metodología para extracción y selección de características para imágenes remotas – ópticas y radar – que permitan la clasificación de LULC.
- Establecer una metodología de clasificación supervisada de imágenes remotas – ópticas y radar – que permita mejorar la identificación de LULC.
- Evaluar la metodología de integración de imágenes remotas por extracción de características para la clasificación de LULC empleando imágenes satelitales sobre territorio Colombiano.

1.4. Estructura del documento

En el Capítulo 2 de este documento, se presenta el estado del arte y el marco teórico relacionado con los fundamentos de la teledetección, los modelos de clasificación LULC, las técnicas de fusión de imágenes y los algoritmos propuestos de machine learning. Se realiza una revisión exhaustiva de los conceptos de fusión de datos y se analizan algunas arquitecturas de fusión de datos para la clasificación de uso y cobertura del suelo.

El Capítulo 3 proporciona un enfoque detallado de la metodología de esta investigación, abordando aspectos cruciales como la descripción del área de estudio, la base de datos utilizada, el preprocesamiento con los procesos de caracterización y fusión de datos, la clasificación de LULC, la evaluación de la clasificación y finalmente las métricas para la validación del rendimiento.

Además, un Capítulo 4 que se centra en la presentación y análisis de los resultados obtenidos a partir de la aplicación de la metodología. Finalmente, el documento se concluye con una sección que presenta el resumen de los hallazgos

clave, discusiones sobre las implicaciones de la investigación y recomendaciones para futuros estudios en esta área.

1.5. Contribuciones

1.5.1. Publicaciones

- *Fusion of optical and radar data by aggregation into a single feature space for LULC classification* - Communications in Computer and Information Science - Estado: Publicado - Categoría Publindex C - V. Herrera-Ruiz, J. Perez-Guerra, JD. Martínez-Vargas, JC. Gonzalez-Velez, and MC. Torres-Madronero.
- *Land Cover Classification Using Remote Sensing and Supervised Convolutional Neural Networks* - Communications in Computer and Information Science - Estado: Publicado - 2023 - Categoría Publindex C - J. Perez-Guerra, V. Herrera-Ruiz, JD. Martínez-Vargas, JC. Gonzalez-Velez, and MC. Torres-Madronero.

1.5.2. Ponencias

- *Fusion of optical and radar data by aggregation into a single feature space for LULC classification* - 17 Congreso Colombiano de Computación – Medellín, Colombia - Agosto 2023.
- *Supervised machine learning approaches for Land Use/Land Cover classification using fusion of multispectral image and radar data* - XVI Congreso Nacional AEET - Almería, España - Octubre 2023.
- *Land Cover Classification Using Remote Sensing and Supervised Convolutional Neural Networks* - 17 Congreso Colombiano de Computación – Medellín, Colombia - Agosto 2023.
- *Land Use/Land Cover Classification Using Multi-Temporal Multispectral Imagery and Supervised Convolutional Neural Networks* - XVI Congreso Nacional AEET - Almería, España - Octubre 2023.

Capítulo 2

Estado del Arte y Marco Teórico

Contenido

2.1	Uso y cobertura del suelo / Land Use Land Cover	7
2.2	Teledetección y Sensores	8
2.3	Atributos en la adquisición	12
2.4	Técnicas de fusión	14
2.5	Procesamiento de imágenes	16
2.6	Modelos de Clasificación	18

2.1. Uso y cobertura del suelo / Land Use Land Cover

El uso de la tierra y la cobertura de la tierra, son dos términos fuertemente relacionados, pero con diferentes significados. La cobertura terrestre hace referencia a los elementos físicos que cubren la superficie en un lugar determinado, ya sea vegetación, infraestructura urbana, agua, suelo desnudo u otros. Por el contrario, el uso de la tierra se refiere al propósito al que sirve esta, como en recreación, hábitat de vida silvestre o agricultura, siendo además una forma de descripción de cómo las personas utilizan la tierra para las actividades socio-económicas [25, 26]. La información que se puede obtener sobre LULC es sumamente valiosa, ya que permite realizar análisis detallados incluso en regiones muy extensas. Sin embargo, la divergencia entre ambos términos es crucial para la teledetección, ya que los satélites proporcionan imágenes de la cobertura del

suelo, mientras que la información sobre el uso del suelo se basa principalmente en la interpretación humana adicional [27].

Cuando se usan en conjunto, generalmente se refiere a la categorización o clasificación de las actividades humanas y los elementos naturales en el paisaje dentro de un marco de tiempo específico basado en métodos científicos y estadísticos. Esto hace que la caracterización de los recursos naturales y su dinámica sea la forma más particular para dar respuesta a la mayoría de preocupaciones medioambientales de la humanidad en la actualidad, como el cambio climático, la pérdida de biodiversidad y la contaminación [28, 29]. Convirtiéndose en necesidades vitales que exigen un mayor esfuerzo en la creación de mapas de cobertura del suelo de alta resolución espacial, precisos y oportunos [30]. Ratificando que la información recolectada por medio de diferentes técnicas es ahora vital para la conservación de los ecosistemas y el medio ambiente, en particular, si se piensa en la reducción de las emisiones derivadas de la deforestación y la degradación de los bosques, como principal ejemplo típico de una iniciativa que ilustra la necesidad de disponer de información LULC precisa [31].

Un factor estimulante para continuar mejorando las investigaciones relacionadas con LULC es la gran evolución de los últimos años en relación con las capacidades de almacenamiento, transmisión, integración y gestión de datos, cambiando radicalmente la forma de observar la Tierra [32]. Por lo tanto, la disponibilidad de datos de LULC ha aumentado tanto que se ha convertido en una fuente de información sin precedentes, permitiendo obtener una imagen más completa del estado del planeta [26], además de convertirse en una gran oportunidad para nuevas investigaciones en una infinidad de aplicaciones que acompañan nuevos retos en materia de investigación científica.

2.2. Teledetección y Sensores

La teledetección es una tecnología que utiliza sensores montados en plataformas aéreas o satelitales para capturar información sobre la superficie terrestre sin necesidad de contacto físico, estos sensores son instrumentos clave que permiten la observación y análisis detallado de la superficie desde el espacio [33]. La aplicación de datos de teledetección en la clasificación de LULC se remonta a mediados de los años 40, cuando se inició la cartografía de Estados Unidos, donde se asociaron los usos del suelo a la superficie de la tierra mediante fotografía aérea [26]. Transitando a la actualidad, los sistemas de teledetección continúan siendo un recurso de suma importancia para el seguimiento de cambios en las coberturas y usos del suelo a escala global [30, 34].

Las imágenes satelitales se consideran una representación visual de la información obtenida por un sensor remoto, mediante la captación de la radiación electromagnética emitida o reflejada por un cuerpo, que posteriormente se transmite a estaciones terrestres para su visualización, procesamiento y análisis [11, 35]. Estas imágenes proporcionadas por los sensores de percepción remota cuentan con una amplia gama de características que permiten identificar su potencial uso, dentro de estas características es importante resaltar la resolución de cada sensor, que puede definirse como, resolución espacial, temporal, espectral y radiométrica [36]. En términos generales, la resolución espacial se refiere al tamaño del píxel; la resolución temporal es la frecuencia con la que puede capturar imágenes de un mismo objeto o superficie; la resolución espectral indica el número de bandas y el ancho de bandas espectrales que capta el instrumento; y la resolución radiométrica es la cantidad de niveles de grises que puede ser almacenada en las bandas del instrumento del sensor [37].

Ahora, estas imágenes pueden obtenerse desde múltiples plataformas y sitios web, especialmente de libre acceso, que se han desarrollado a lo largo del tiempo para facilitar el almacenamiento y la descarga de datos satelitales según el propósito para el cual se necesiten [38]. Entre estos se encuentra el programa de observación de la tierra de la Unión Europea Copernicus. Este fue constituido en gran medida por la Agencia Espacial Europea (ESA), organización internacional dedicada a la exploración espacial que ha desarrollado una nueva familia de misiones de observación de la tierra para atender las necesidades operacionales del programa [39]. Copernicus Open Access Hub está diseñado para proporcionar un acceso completo, gratuito y abierto a los productos de misiones como Sentinel 1, Sentinel 2, Sentinel 3 y Sentinel 5P [40]. Cada misión Sentinel está formada por una constelación de dos satélites, lo que permite cumplir los requisitos de cobertura y frecuencia de observación para ofrecer un conjunto robusto de datos [39].

Por otro lado, Google Earth Engine (GEE) es otra de las plataformas relevantes para la visualización y descarga de imágenes satelitales [41]. GEE utiliza infraestructura de Google basada en la nube para facilitar el acceso y procesamiento de datos geospaciales, proporcionando valiosas fuentes de información al mundo académico y de la investigación, lo que representa una de sus características más destacadas [42]. Esta plataforma permite acceder a un extenso catálogo de datos listos para su análisis, combinado con un servicio computacional inherentemente paralelo y de alto rendimiento [43]. Su funcionamiento se basa en el uso de una interfaz de programación de aplicaciones (API) accesible a través de Internet y cuenta con un entorno de desarrollo interactivo (IDE) basado en web, que facilita la creación rápida de prototipos y la visualización de resultados [44]. Adicionalmente, es de suma importancia mencionar que es-

ta plataforma ofrece tanto datos sin procesar como preprocesados, y productos disponibles a escala global, nacional y local, contando con datos provenientes de diferentes sensores como Landsat, MODIS, NOAA, Sentinel y ALOS [45].

En la actualidad, existe una enorme cantidad de repositorios con información satelital accesible de forma gratuita a escala global, que en muchos casos permanece sin explotar [30,34]. Por lo tanto, la alta disponibilidad de datos satelitales y la elección de plataformas de análisis adecuadas añade otro nivel de desafío al procesamiento de imágenes de teledetección [46]. Sin embargo, para superar estos desafíos es esencial conocer las características de los datos disponibles en estos repositorios, incluyendo conjuntos de datos y productos según áreas de estudio, precisiones de clasificación, tipo y resolución de sensores y sus diversas aplicaciones [47]. En la Tabla 2.1 se pueden observar algunas características y especificaciones de diferentes sensores ya mencionados.

Tabla 2.1: Características y especificaciones de diferentes sensores

Tipo de Sensor	Sensor	Resolución (m)	Altitud (km)	Frecuencia (días)	País de Operación
Óptico	Landsat-8	15	705	16	USA
	Sentinel 2	10-60	785	10	Unión Europea
	SPOT	10-20	822	26	Francia
SAR	Sentinel 1	5	693	10	Unión Europea
	TerraSAR-X	1	514	11	Alemania
	Radarsat-2	3-100	798	—	Canadá

El uso de sensores activos o pasivos para captar y analizar información es uno de los principales fundamentos de la teledetección [48]. Los sensores pasivos dependen de la luz solar reflejada por los objetos en la superficie, a partir de la cual, capturan la información en diversas bandas del espectro electromagnético [49]. Los sensores pasivos son ampliamente utilizados para aplicaciones como la cartografía, la evaluación de la vegetación, identificación de incendios, zonas inundadas, la calidad del agua, entre otros [50–52]. Estos sensores son sensibles a una gama espectral que va del espectro visible al infrarrojo medio de las radiaciones emitidas, y produce imágenes multiespectrales o hiperespectrales [26]. Las imágenes multiespectrales constan de varios canales, cubriendo frecuencias espectrales entre 2 y 13 cortes en el espectro electromagnético,

mientras que las imágenes hiperespectrales pueden abarcar cientos de canales de información [53]. Dentro de los sensores pasivos, también encontramos los sensores térmicos, estos captan las radiaciones térmicas de la superficie terrestre sobre longitudes de onda infrarrojas de onda media a onda larga [26].

Por otro lado, los sensores activos, como los radares de apertura sintética (SAR), emiten su propia energía y registran las señales reflejadas por la superficie terrestre, lo que les permite operar independientemente de las condiciones de iluminación, atravesar nubes y vegetación densa, y proporcionar datos cruciales en condiciones atmosféricas adversas [54]. Otra ventaja de la teledetección basada en el SAR es la capacidad de capturar datos de día y noche, lo que permite obtener conjuntos de datos de series temporales largas que pueden utilizarse para supervisar los cambios de la cobertura terrestre [31, 34]. A diferencia de los sensores térmicos y ópticos que funcionan de forma pasiva, el sensor SAR es un instrumento activo de microondas que ilumina el suelo dispersando las radiaciones de microondas y capta las ondas reflejadas [55]. Finalmente, es importante mencionar que considerando esa cantidad de dispersión y el tiempo de viaje, es posible estimar la distancia a la que se encuentra el objetivo.

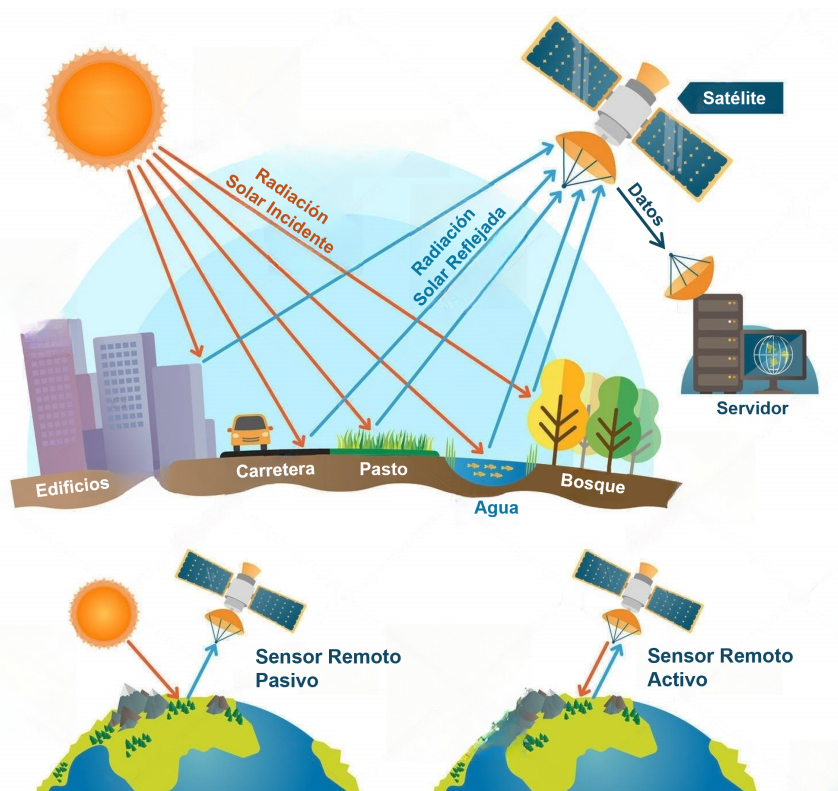


Figura 2.1: Teledetección y Sensores. Fuente: Creado por la autora.

2.3. Atributos en la adquisición

La adquisición de datos para los mapas de clasificación de LULC puede realizarse manualmente a través de salidas de campo. Sin embargo, esta metodología presenta numerosos obstáculos y limitaciones, ya que es laborioso y dificulta la actualización cartográfica debido al tiempo requerido y los elevados costos económicos implicados [56]. Ahora, el uso de datos de teledetección ha permitido desarrollar sistemas automatizados con mayor precisión, lo que a su vez ha sido posible gracias a las mejoras y evolución en el rendimiento de máquinas computacionales, sensores a bordo de los satélites y la inteligencia artificial (IA), facilitando solventar las limitantes de la adquisición manual [26, 31, 57].

Dado que la teledetección proporciona métodos alternativos al tradicional para una adquisición con menor costo y ciclos de revisión más frecuentes, se ratifica que hoy en día es una forma conveniente de monitorear y evaluar diversos factores ambientales en el mundo [58]. Por lo tanto, es indispensable entender que, en la aplicación de la teledetección, se analiza la reflectancia en partes individuales del espectro electromagnético de una cubierta terrestre observada [59]. Cada cobertura y sus propiedades dan lugar a un conjunto único de valores en las partes observadas del espectro, representados mediante firmas espectrales. Por lo cual, para detectar las propiedades y los cambios en la cobertura, se realiza un análisis de las firmas de una misma zona en diferentes periodos de tiempo [60].

Los sistemas de teledetección constituyen lo que hoy se denomina como misiones espaciales, distinguidas por su amplia cobertura de imágenes y su capacidad para integrarse con varios tipos de sensores compatibles [61]. Sin embargo, estas misiones difieren principalmente en función de las longitudes de onda del espectro electromagnético a las que son sensibles sus sensores [60]. La distancia al suelo y el movimiento de las plataformas satelitales pueden ser responsables de la notable reducción en la resolución espacial de las imágenes [62], por lo cual, la incompatibilidad de escala dificulta inevitablemente la explotación de las mediciones adquiridas, evidenciándose como uno de los principales obstáculos entre los métodos de adquisición [63]. No obstante, las misiones espaciales pueden proporcionar imágenes con tasas de muestreo temporal significativamente altas, puesto que, cada vez hay más de estas plataformas alrededor de la tierra [30].

La ESA ha desarrollado una familia de misiones satelitales ampliamente utilizadas para la adquisición de datos de teledetección [64]. Una de ellas es Sentinel 2, esta es una misión multiespectral de alta resolución y gran amplitud en órbita polar, diseñada para monitorizar la superficie de la tierra, estudiar la cubierta vegetal, los usos del terreno, las aguas, los cauces navegables, zonas costeras,

entre otros [65]. Sentinel 2 está formada por una constelación de dos satélites de órbita polar situados en la misma órbita sincrónica al sol, desfasados 180° entre sí [64]. Su gran anchura de barrido de 290 km y su tiempo de revisita de 10 días en el ecuador con un satélite, y 5 días con 2 satélites en condiciones de ausencia de nubes, permiten vigilar con gran destreza los cambios de la superficie terrestre [66].

Estos sensores llevan una carga de instrumentos ópticos que muestrean trece bandas espectrales [67]. Las trece bandas se adquieren en tres niveles de resolución espacial diferentes: cuatro bandas están a 10 m, seis bandas están a 20 m y tres bandas están a 60 m [68]. En contraste con las bandas de 10 m, las bandas con resolución de 60 m, presentan una estructura espacial menos detallada de las características del terreno. Así mismo, las bandas de resolución más baja proporcionan respuestas espectrales únicas que pueden ser útiles en dominios específicos. Por ejemplo, las bandas de borde rojo de 20 m contienen información sobre el contenido de clorofila de la vegetación y su estado de crecimiento, mientras que la banda de aerosoles de 60 m es útil para la corrección atmosférica y la detección de nubes [69].

Por otro lado, Sentinel 1 es una misión SAR en órbita polar, capaz de tomar imágenes en cualquier condición meteorológica, de día o de noche, para estudiar la superficie de la tierra y los océanos [70]. La misión está compuesta por los satélites Sentinel 1A y Sentinel 1B, que comparten el mismo plano orbital. Incluye imágenes en banda C que operan en cuatro modos exclusivos con una resolución variable de hasta 5 m y un área de cobertura de hasta 400 km. También, ofrece capacidad de doble polarización, tiempos de revisita muy cortos y una rápida entrega de productos [66]. Asimismo, Sentinel 1 está diseñado para trabajar en modo de operación preprogramado y libre de conflictos, garantizando la fiabilidad del servicio y un archivo de datos consistente a largo plazo construido para aplicaciones basadas en series temporales largas [71].

Finalmente, de manera general entre las configuraciones y características de los SAR existen misiones que operan en diferentes bandas del espectro electromagnético, entre las más comunes se encuentran la banda X, C y L. En términos generales, los sistemas de banda X tienen una longitud de onda cercana a los 3 cm, banda C cercana a los 5 cm y banda L cercana a los 23 cm en el espectro electromagnético [67]. La banda se elige según la información que se desea estudiar y está estrechamente relacionada con el modo de operación, donde se diferencia el rango de cobertura y la resolución espacial de la adquisición [64]. Otra característica importante es la polarización, haciendo referencia a la orientación de la transmisión y recepción de onda. Esta puede ser horizontal o vertical, además, los objetivos en el suelo tienen firmas de polarización distintivas que

reflejan diferentes polarizaciones e intensidades [64, 72].

2.4. Técnicas de fusión

Para comprender plenamente las características y los cambios en el uso y cobertura del suelo, es necesario aprovechar exhaustivamente las múltiples ventajas que ofrecen los diferentes sensores satelitales [73]. Por lo tanto, la fusión de datos de distintas plataformas de teledetección puede proporcionar información diversa y complementaria, lo que mejora significativamente la precisión y el valor de la información extraída [74]. De esta manera, en los últimos años las técnicas de fusión de datos han ganado gran relevancia en estudios relacionados con la vigilancia de los recursos naturales, así como en el análisis y clasificación de LULC. [75, 76].

Las técnicas de fusión de datos aprovechan las estructuras espaciales, la información espectral, la independencia de las condiciones meteorológicas, las propiedades dieléctricas, entre otras características para ofrecer una visión unificada y completa de cada fuente [77, 78]. Entre las aplicaciones más relevantes de estas técnicas de fusión se encuentra la integración de datos ópticos y SAR, utilizadas en una gran variedad de investigaciones de teledetección con excelentes resultados [79, 80]. Esta combinación puede producir resultados significativamente mejorados, ya que los datos SAR miden las características físicas de los objetos terrestres y son independientes de las condiciones climáticas, mientras que los datos ópticos generalmente abarcan las longitudes de onda del infrarrojo cercano, el visible y el infrarrojo corto [79, 81].

Los métodos de integración de datos se pueden clasificar en tres niveles diferentes: nivel de pixel, nivel de características y nivel de decisión [82, 83]. En la primera categoría, como se puede ver en la Figura 2.2 se combinan directamente los datos de los sensores, que luego se utilizan en los métodos clásicos de detección y estimación, el prerequisite aquí es que las imágenes deben estar estrictamente registradas antes de la fusión, por lo que requiere que los datos estén debidamente georeferenciados y cuenten con la misma resolución espacial [84].

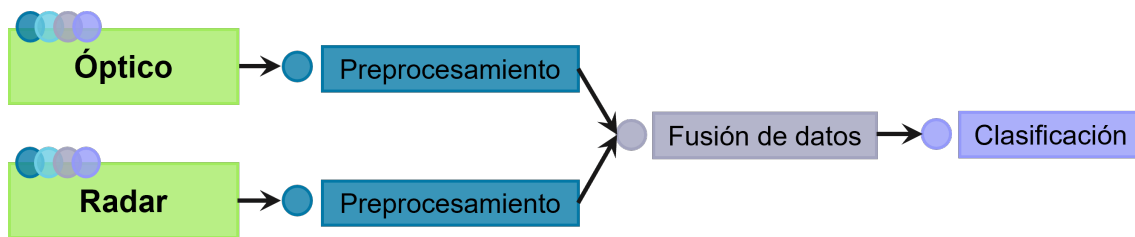


Figura 2.2: Fusión por nivel de píxel. Fuente: Creado por la autora.

En la segunda categoría, como se puede observar en la Figura 2.3 se extraen las características más destacadas de diferentes imágenes y se produce un mapa de características combinado. Este mapa luego se utiliza en algoritmos de agrupación y enfoques de reconocimiento de patrones, permitiendo que se puedan tomar decisiones más precisas [82].

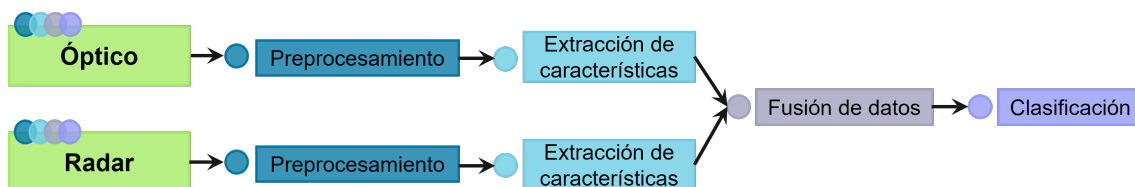


Figura 2.3: Fusión por nivel de características. Fuente: Creado por la autora.

Finalmente, en la Figura 2.4 se evidencia el proceso de fusión a nivel de decisión, donde las imágenes se preprocesan y clasifican por separado. La información obtenida de las imágenes individuales clasificadas se combina y se refina. Esta se conoce como una fusión al más alto nivel, por lo que implica la fusión de la información de los sensores después de que se haya realizado una determinación preliminar de la clasificación de forma independiente [85].

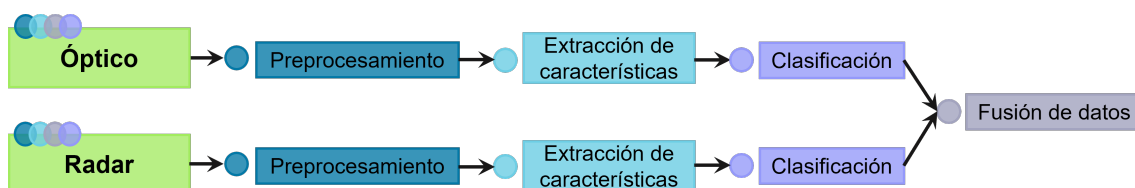


Figura 2.4: Fusión por nivel de decisión. Fuente: Creado por la autora.

En general, estas técnicas de fusión ofrecen flexibilidad en las aplicaciones de teledetección, permitiendo seleccionar el enfoque más adecuado según las características específicas de las fuentes de datos y los objetivos del análisis.

2.5. Procesamiento de imágenes

El procesamiento automatizado de imágenes ha disminuido el esfuerzo humano y se ha convertido en un resultado codiciado, permitiendo aumentar la eficiencia de la extracción de información [86]. No obstante, un requisito previo al procesamiento y análisis de imágenes es el preprocesamiento. Para los datos de teledetección este representa un conjunto de correcciones, hecho que tiende a ser específico para cada misión individual, sin embargo, se suelen llevar a cabo múltiples técnicas generales para solucionarlas [60]. Estas correcciones son necesarias debido a que la toma de datos puede contener ruido y otras deficiencias derivadas de los sensores o de los procesos de transferencia.

Dichas operaciones consisten generalmente en procesos destinados a la corrección geométrica, radiométrica y atmosférica, permitiendo mejorar la capacidad de interpretación de los componentes de la imagen cualitativa y cuantitativamente [87]. Es así como se considera que los fallos en el sensor que generan píxeles incorrectos, requieren de una corrección radiométrica, las alteraciones en el movimiento del satélite y el mecanismo de captación generan distorsiones globales necesitando una corrección geométrica y la interferencia atmosférica, que altera los valores de los píxeles, precisa de una corrección atmosférica [88]. Otra técnica importante es la ortorrectificación, con la cual se puede garantizar un posicionamiento espacial preciso entre las imágenes multitemporales y de diversas fuentes, además de eliminar las distorsiones geométricas y definir todos los conjuntos de datos en un sistema de proyección común [89].

Una vez finalizado el preprocesamiento se puede continuar con la etapa del procesamiento y análisis de los datos. Esta puede dividirse regularmente en tres aspectos: visual, cuantitativo y cualitativo [90]. La interpretación visual permite mostrar las características del terreno específicas para la combinación seleccionada de bandas espectrales [91]. El enfoque cuantitativo se basa principalmente en la aplicación de imágenes multiespectrales, combinando los valores de los píxeles de diferentes bandas. La propiedad seleccionada del área observada se cuantifica mediante la aplicación de índices espectrales, donde los valores de esa propiedad son proporcionales a los valores del índice espectral calculado [92]. Por último, la evaluación cualitativa se basa en la clasificación de los valores espectrales de los píxeles u objetos de la imagen en un número predefinido de clases [60].

En la continuación de un buen procesamiento se tiene la extracción de características, esta es la parte más utilizada en esta etapa. En la práctica, se emplean dos métodos de extracción: tipo I y tipo II [86]. Los métodos de extracción de características tipo I se centran principalmente en la búsqueda de

parámetros originales desde cero, mientras que el método tipo II se utiliza para optimizar la precisión de un conjunto mediante la eliminación de características inconsistentes [93]. En esta fase, se consideran dos enfoques distintos: el enfoque basado en píxeles, que utiliza exclusivamente la información espectral de los píxeles para extraer características de la superficie, y el enfoque basado en objetos, que se emplea para obtener información relevante de la superficie mediante la combinación de información espectral, espacial y la textura y contexto de la imagen [94, 95].

Un factor fundamental en la extracción de características son los índices de vegetación. Estos índices son mediciones radiométricas obtenidas mediante combinaciones matemáticas de los diferentes espectros de cada sensor [96], además, se contempla que los índices han sido ideados para resaltar características específicas de las imágenes adquiridas, permitiendo discernir entre diferentes tipos de cobertura terrestre [97]. La combinación de ciertos índices y bandas espectrales también permite establecer relaciones entre los datos, aportando información sustancial en los procesos de teledetección [98]. Consecuentemente, se han desarrollado una variedad de índices para la caracterización de diferentes propiedades del terreno, tanto para sensores pasivos como activos. En la Tabla 2.2 se presentan algunos de los índices más destacados en el estado del arte de LULC.

Tabla 2.2: Índices ópticos y SAR utilizados en aplicaciones de clasificación de LULC.

Índice	Abreviatura	Ecuación
Índice de Vegetación de Diferencia Normalizada	NDVI	$NDVI = \frac{NIR_1 - RED}{NIR_1 + RED}$
Índice de Diferencia Normalizada Acumulada	NDBI	$NDBI = \frac{SWIR_1 - NIR_1}{SWIR_1 + NIR_1}$
Índice de Calcinación Normalizado	NBRI	$NBRI = \frac{NIR_1 - SWIR_2}{NIR_1 + SWIR_2}$
Índice de humedad de Diferencia Normalizada	NDMI	$NDMI = \frac{NIR_1 - SWIR_1}{NIR_1 + SWIR_1}$
Índice Diferencial de Agua Normalizado Modificado	MNDWI	$MNDWI = \frac{GREEN - SWIR_1}{GREEN + SWIR_1}$
Índice Diferencial de Agua Normalizado	NDWI	$NDWI = \frac{GREEN - NIR_1}{GREEN + NIR_1}$
Índice de Vegetación de Diferencia Normalizada Verde	GNDVI	$GNDVI = \frac{NIR_1 - GREEN}{NIR_1 + GREEN}$
Índice de Suelo Desnudo	BSI	$BSI = \frac{(RED + SWIR_1) - (NIR_1 + BLUE)}{(RED + SWIR_1) + (NIR_1 + BLUE)}$
Índice de Vegetación Mejorado	EVI	$EVI = 2,5 \times \frac{NIR_1 - RED}{NIR_1 + 6 \times RED + 7,5 \times BLUE + 1}$
Índice Diferencial Normalizado de Nieve	NDSI	$NDSI = \frac{GREEN - SWIR_1}{GREEN + SWIR_1}$
Índice de Vegetación Simétrico de Polarización Dual	DPSVI	$DPSVI = \frac{VV - VH}{VV + VH}$
Índice Mejorado de Máxima Razón de Respuesta de Vegetación	RVIm	$RVIm = \frac{VV_{t2} - VV_{t1}}{VV_{t1}}$
Ratio de Polarización	Pol	$Pol = \frac{VH}{VV}$
Índice de Contraste de Relación de Volumen de VV por VH	CRI	$CRI = \frac{Coherence_{VV}}{Coherence_{VH}}$

2.6. Modelos de Clasificación

Durante décadas, se ha llevado a cabo una amplia investigación para desarrollar y perfeccionar algoritmos capaces de procesar y clasificar el uso del suelo y la cobertura terrestre. En este contexto, el aprendizaje automático (Machine

Learning, ML) ha surgido como uno de los enfoques más destacados y prometedores [99]. La evolución de ML ha permitido avances significativos en la precisión y eficiencia de la clasificación de la cobertura terrestre, facilitando la creación de modelos más robustos y adaptativos que pueden manejar grandes volúmenes de datos con alta variabilidad. ML, como disciplina, se centra en extraer conocimiento de los datos y se esfuerza por encontrar modelos que no solo identifiquen patrones, sino que también mejoren la precisión y confiabilidad de las clasificaciones resultantes [11].

La gran mayoría se basan en métodos supervisados o no supervisados [100]. Los supervisados necesitan siempre algunas áreas de entrenamiento, mientras que, los resultados de los métodos no supervisados dependen en gran medida del contenido de la imagen y necesitan un etiquetado semántico posterior [101]. Aun así, las muestras etiquetadas son escasas e insuficientes en comparación con la dimensión del espacio de datos, además de que las muestras no etiquetadas suelen estar disponibles en cantidades importantes [102]. Dentro de los métodos supervisados, se destacan las máquinas de vectores de soporte (SVM) y Random Forest (RF) como dos de los más populares y efectivos [103]. Las SVM son reconocidas por su capacidad para lograr una alta precisión de clasificación, velocidad de aprendizaje y adaptabilidad, incluso en conjuntos de datos de entrenamiento limitados [30, 104]. Por otro lado, RF como metaclasificador basado en árboles de decisión, es conocido por su capacidad para manejar grandes volúmenes de datos, estimar y normalizar valores atípicos y proporcionar una medida de importancia de las variables [11, 105].

Las redes neuronales artificiales (ANN), han sido uno de los principales algoritmos de ML, estos han demostrado rendimientos robustos en el área de clasificación supervisada de imágenes satelitales [25]. Las ANN son una estructura de neuronas unidas por nodos que transmiten información a otras neuronas, las cuales dan un resultado mediante funciones matemáticas. Estas aprenden de la información existente a través de un proceso de entrenamiento, mediante el que se ajustan sus parámetros para proporcionar una salida aproximada a la deseada [106]. A pesar de que estas han sido sumamente estudiadas, su aplicabilidad se ha visto cuestionada por la complejidad de su parametrización, ya que, hay que tener en cuenta una serie de factores, como los datos de entrada, las muestras de entrenamiento, los ajustes de salida, la arquitectura de las redes, los pesos iniciales y los parámetros de entrenamiento [107], sin embargo, esto no ha impedido que se consideren entre las más capaces de mejorar la precisión de la clasificación automatizada.

Adicionalmente, se han realizado análisis comparativos entre métodos supervisados y algoritmos de aprendizaje profundo. En el estudio de [56], se comparó

el rendimiento de modelos de RF, SVM y Redes Neuronales Convolucionales (CNN). Los resultados mostraron que la CNN propuesta tiene un mejor desempeño general en comparación con los algoritmos supervisados, y que SVM es el segundo clasificador más potente. Sin embargo, el rendimiento de estas arquitecturas de aprendizaje profundo está estrechamente relacionado con la cantidad de datos etiquetados disponibles para su entrenamiento, lo cual representa un desafío significativo [108].

También se han realizado diversos estudios comparativos de clasificación para datos multisensor entre SAR, ópticos y sus fusiones [109]. Por ejemplo, en [110], se realizó un metaanálisis de familias de clasificadores que mostró que los algoritmos SVM son superiores a cualquier otra familia de clasificadores. En un estudio en la selva brasileña realizado por [111], donde se utilizaron imágenes SAR para la clasificación de la cobertura del suelo, demostrando que los algoritmos SVM y RF proporcionan los mejores resultados, con SVM obteniendo la mayor precisión global. En contraste, en un estudio similar del bosque mediterráneo realizado por [103], se mostró que el clasificador RF logra una mayor precisión en comparación con el clasificador SVM, lo cual se puede explicar por el desequilibrio en el número de muestras entre las clases en este estudio específico [34].

Finalmente, en estudios más recientes se ha analizado y comparado la implementación del algoritmo Extreme Gradient Boosting (XGBoost) en aplicaciones de teledetección y fusión de datos multisensor [112]. Este algoritmo construye un clasificador utilizando el método de gradient boosting, que añade iterativamente nuevas predicciones basadas en árboles débiles, mejorando constantemente el error de los clasificadores anteriores para crear un clasificador robusto [113]. XGBoost se destaca por proporcionar una resolución más precisa de problemas a gran escala [79], y diversos estudios respaldan esta premisa. Por ejemplo, en el estudio de [114], se compararon los clasificadores Random Forest y XGBoost utilizando datos ópticos y SAR integrados, obteniendo una precisión general del 81 % para XGBoost y del 77 % para RF, demostrando que XGBoost superó a RF en precisión de clasificación. Además, en otras aplicaciones de clasificación realizadas por [82], se obtuvo una precisión global del 86 % con un enfoque propuesto de fusión de datos y un clasificador XGBoost, demostrando su amplio potencial para nuevas aplicaciones y perspectivas en la cartografía.

Capítulo 3

Metodología

Contenido

3.1 Dataset	23
3.1.1 Área de estudio	23
3.1.2 Selección de imágenes	24
3.2 Extracción y Selección de Características	29
3.2.1 Índices Ópticos	30
3.2.2 Índices SAR	32
3.3 Preprocesamiento	33
3.3.1 Normalización	33
3.3.2 Fusión de Datos	34
3.3.3 Validación	35
3.4 Clasificación Supervisada en Imágenes Ópticas y de Radar .	36
3.5 Evaluación de la Clasificación LULC	37
3.6 Métricas para validación del rendimiento	38

Sinopsis

En esta investigación, se implementaron múltiples etapas dentro de la metodología general. Como se muestra en la Figura 3.1, se aprecian los procesos ejecutados para la clasificación y el análisis de la cobertura terrestre a partir de datos de teledetección multisensor ópticos y de radar.

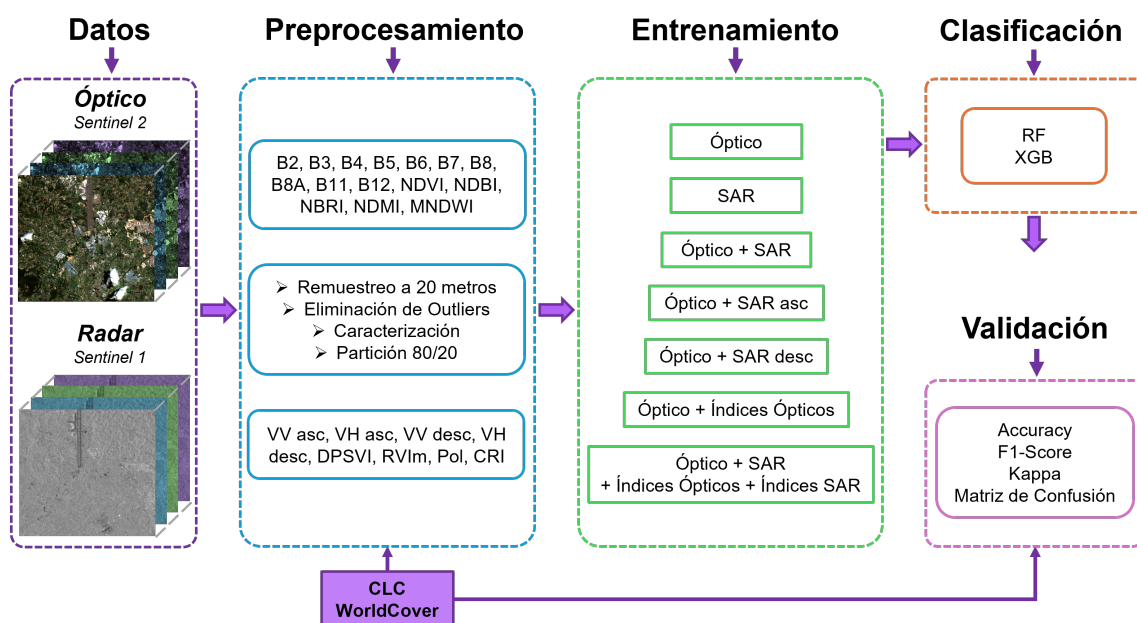


Figura 3.1: Metodología general. Fuente: Creado por la autora.

Inicialmente, se recopilaron los datos ópticos del sensor Sentinel 2, que incluye diversas bandas espectrales como B2, B3, B4, B5, B6, B7, B8, B8A, B11 y B12. Simultáneamente, se obtienen datos de radar de Sentinel 1, utilizando las polarizaciones VV y VH en órbita ascendente y descendente. Se integró una fase de preprocesamiento en la cual, las imágenes fueron remuestreadas a una resolución espacial de 20 metros para asegurar la consistencia en el análisis. Se eliminan los valores atípicos que podrían distorsionar los resultados y se realiza una caracterización detallada de los datos preprocesados mediante el cálculo de índices derivados de los datos ópticos y SAR.

Posteriormente, los datos fueron divididos en conjunto de entrenamiento y un conjunto de pruebas con una proporción de 20/80, respectivamente. Para la clasificación, se emplearon dos tipos de modelos: Radom Forest y XGBoost utilizando diferentes fusiones de datos que incluye únicamente datos ópticos, datos de radar y combinaciones de ambos. Además, se consideran combinaciones que incluyen índices ópticos y SAR, junto a 2 mapas de coberturas Corine Land Cover (CLC) y WorldCover. Finalmente, incluimos las métricas de precisión (Accuracy), F1-Score, Kappa en conjunto con matrices de confusión como método de evaluación.

3.1. Dataset

3.1.1. Área de estudio

El área de estudio seleccionada para esta investigación, como se muestra en la Figura 3.2, se encuentra en la región de Antioquia, Colombia, ubicada en el extremo noroccidental del país. Esta área abarca aproximadamente 190,000 hectáreas de territorio diversificado, conocida como la subregión del Oriente Antioqueño. Se extiende entre los municipios de Guarne, Guatapé, La Ceja y Coconá. Esta región se caracteriza por su alto nivel de pluviosidad, con una media anual de 2,500 mm, y presenta una gran variedad climática debido a su amplio rango altitudinal.

El territorio del Oriente Antioqueño se compone de una topografía montañosa y valles fértiles, lo que lo convierte en un área geográficamente diversa. La altitud varía desde las zonas bajas hasta picos de montaña que superan los 2,000 metros sobre el nivel del mar, lo cual genera una amplia gama de microclimas, que van desde cálidos y húmedos en las tierras bajas hasta frescos y templados en las zonas más elevadas [115].

A su vez, el crecimiento poblacional y la diversificación de las actividades productivas en el Oriente Antioqueño presentan una gran oportunidad de desarrollo para la subregión. Sin embargo, el cambio demográfico, la transición en la estructura empresarial y los procesos masivos de urbanización también plantean nuevos retos. Estas transformaciones del territorio, marcadas por el desarrollo de la infraestructura vial, la expansión urbana y la intensificación de la agricultura, han sido significativas en las últimas décadas [116]. Estas modificaciones en el uso del suelo y la cobertura terrestre han generado la necesidad de monitorear y comprender la evolución de la composición territorial a lo largo del tiempo.

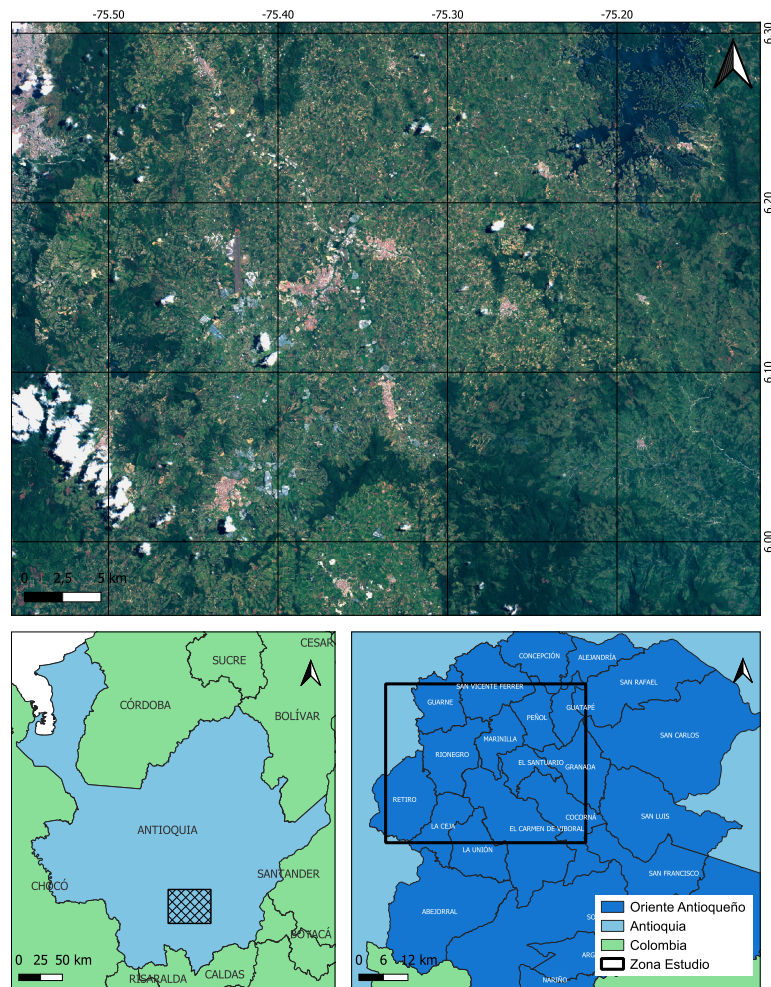


Figura 3.2: Área de estudio. Fuente: Creado por la autora.

3.1.2. Selección de imágenes

Imágenes de percepción remota

La base de datos utilizada en esta investigación comprende imágenes ópticas del sensor Sentinel-2 e imágenes SAR del sensor Sentinel-1, capturadas a partir del mes de agosto de 2019 a través de la plataforma Google Earth Engine (GEE). La cual es reconocida por proporcionar acceso gratuito a una extensa base de datos de imágenes satelitales, representándola como una herramienta importante para la investigación y el análisis geoespacial. Además, su capacidad de procesamiento en la nube permitió realizar estas tareas de manera rápida y escalable.

Para los datos de Sentinel-2, se seleccionaron las bandas con una resolución espacial inferior a 20 metros, como se puede observar en la Tabla 3.1. Optando por utilizar el nivel 2A, lo que implica que se usaron imágenes ortorrectificadas con niveles de reflectancia corregidos por efectos atmosféricos. Esta selección garantiza una mayor precisión en la representación de las características del terreno al eliminar las distorsiones causadas por la atmósfera.

Tabla 3.1: Atributos de las bandas de Sentinel 2

Banda	Nombre de banda	Longitud de onda (nm)	Resolución (m)
B2	Blue	490	10
B3	Green	560	10
B4	Red	665	10
B5	Red Edge 1	705	20
B6	Red Edge 2	740	20
B7	Red Edge 3	783	20
B8	NIR 1	842	10
B8A	NIR 2	865	20
B11	SWIR 1	1610	20
B12	SWIR 2	2190	20

En cuanto a los datos de Sentinel-1, son un producto de datos SAR que recopila imágenes de banda C que operan a una frecuencia central de 5,405 GHz. Se recopilaron datos de la colección Ground Range Detected (GRD) adquiridos en el modo exclusivo de franja interferométrica ancha (IW por sus siglas en inglés Interferometric Wide), lo que permite combinar una franja de gran anchura de 250 km con una resolución geométrica moderada de 5 m x 20 m. Fueron tomados en polarización vertical de transmisión y recepción (VV) y polarización vertical de transmisión y horizontal de recepción (VH), incluyendo datos de órbita ascendentes (asc) y descendentes (desc), destinados a aplicaciones terrestres. La inclusión de múltiples polarizaciones y órbitas permitió una visión más completa del área de estudio, proporcionando información detallada sobre las propiedades físicas de la superficie terrestre independientemente de las condiciones climáticas.

Además, para los datos ópticos de Sentinel-2, se consideraron cuatro rangos de nubosidad equivalentes a diferentes porcentajes para evaluar el comportamiento de la presencia de nubes en comparación con los datos SAR. Esta evaluación es crucial para entender cómo la nubosidad afecta la calidad de las

imágenes ópticas y cómo los datos SAR pueden complementar esta información. Los tiempos de captura de estas imágenes son variables, pero se mantienen dentro del mismo año y mes, lo cual se detalla en la Tabla 3.2. Esta sincronización temporal asegura la coherencia en el análisis comparativo de los datos.

Tabla 3.2: Detalles de los datos Ópticos y SAR

Sensor	Píxeles, Bandas	Porcentaje de nubes	Fecha	Id de la imagen
Sentinel 2	4810748, 10	0 - 30 %	27/08/2019	20190827T152649_20190827T152643_T18NVM
		30 - 50 %	22/08/2019	20190822T152641_20190822T152641_T18NVM
		50 - 70 %	02/08/2019	20190802T152641_20190802T152643_T18NVM
		70 - 90 %	17/08/2019	20190817T152649_20190817T152645_T18NVM
Sentinel 1	4810748, 4	N/A	20/08/2019	S1B_IW_GRDH_1SDV_20190820T105010_20190820T105035_017668_0213D5_OC35

Mapas de referencia

Es relevante destacar que se utilizaron dos mapas diferentes como datos de referencia para validar el proceso de clasificación. Además, se llevó a cabo un proceso de remuestreo para los conjuntos de datos y los mapas de referencia utilizando la herramienta GEE. Este remuestreo facilitó la estandarización de la resolución de todas las imágenes a 20 metros, preparando adecuadamente los datos para su análisis.

■ Mapa Corine Land Cover (CLC)

El mapa CLC, liderado actualmente por el IDEAM en Colombia, proporciona clasificaciones detalladas de las coberturas del suelo que van desde el nivel 1 hasta el nivel 6. Estas clasificaciones se obtienen a través de la interpretación visual de cartógrafos expertos y ofrecen una base de comparación para evaluar la precisión de las clasificaciones resultantes [117]. Para esta investigación, se optó por trabajar con el nivel 2, aunque se obtuvieron solo 7 categorías de las 15 definidas en este nivel. Esto debido a que la interpretación visual realizada por los cartógrafos conlleva un alto error en la categorización de las clases en zonas muy extensas, además de la ausencia de algunas categorías definidas en el nivel 2 del CLC en el área de estudio.

La reclasificación se llevó a cabo mediante un análisis visual de la imagen en la zona de estudio utilizando el sistema de información geográfica QGIS Desktop

3.36.2 (<https://qgis.org/download/>). Se tomaron una a una las clases identificadas por el CLC en el nivel 2 y se reclasificaron en los casos donde se evidenció que no pertenecían a la clase asignada inicialmente. En la Tabla 3.3 se muestra el detalle de la recategorización de las clases realizada.

Tabla 3.3: Reclasificación del Nivel 2 de CLC

Nivel 2	Aplica / No aplica	Nivel 2 Reclasificado
Zonas urbanas	✓	
Zonas industriales, comerciales y de transporte	✓	Zonas urbanizadas
Zonas de extracción minera, vertederos y de construcción	✓	
Zonas verdes artificiales, no agrícolas	✓	
Tierras de labor	X	–
Cultivos permanentes	✓	Cultivos permanentes
Prados y praderas	✓	Pastos
Zonas agrícolas heterogéneas	✓	Áreas agrícolas y heterogéneas
Bosques	✓	Bosques
Espacios de vegetación arbustiva y/o herbácea	✓	Áreas con vegetación herbácea y/o arbustiva
Espacios abiertos con poca o sin vegetación	X	–
Zonas húmedas continentales	X	–
Zonas húmedas litorales	X	–
Aguas continentales	✓	Aguas continentales
Aguas marinas	X	–

Finalmente, la distribución de este mapa se puede observar en la Figura 3.3, con las clases definidas como: Zonas Urbanizadas, Cultivos Permanentes, Pastos, Áreas Agrícolas y Heterogéneas, Bosques, Áreas con Vegetación Herbácea y/o Arbustiva y Aguas Continentales.

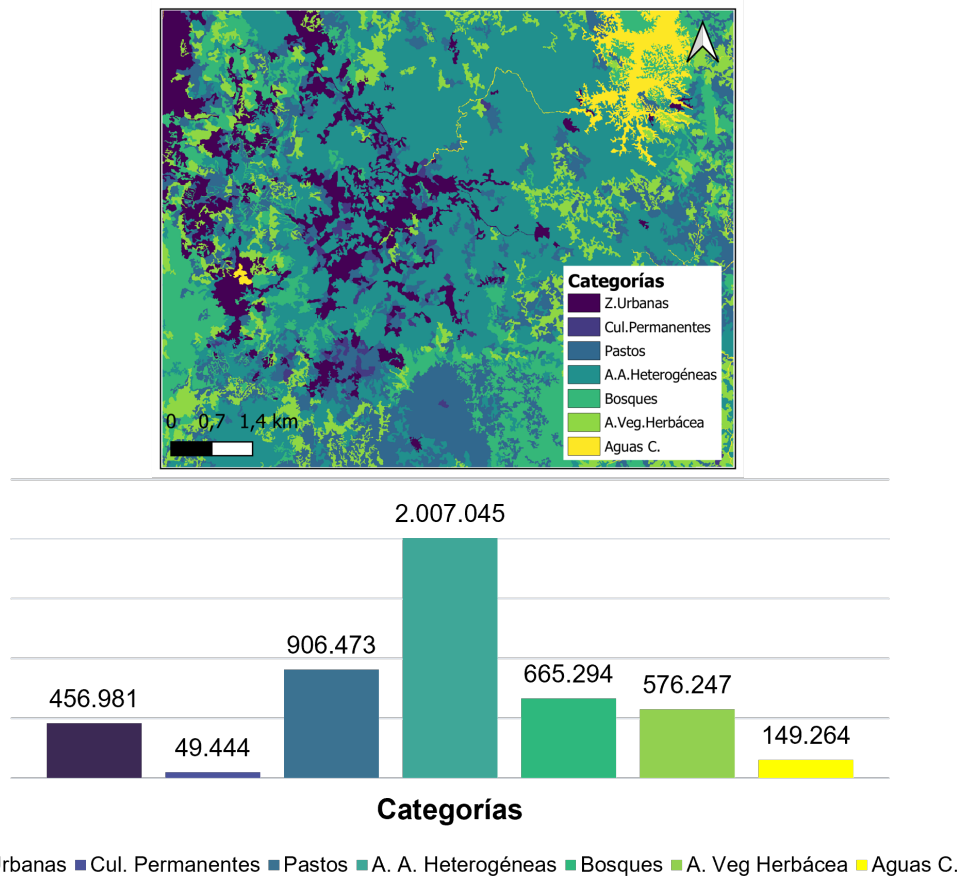


Figura 3.3: Mapa y distribución de las clases en el CLC. Fuente: Creado por la autora.

■ Mapa WorldCover

Considerando los inconvenientes presentados con el mapa CLC debido a su alto error en la clasificación de algunas áreas, se decidió incorporar en una segunda etapa el producto WorldCover 10 m 2020 de la ESA. Este producto proporciona un mapa global de la cubierta terrestre para 2020 con una resolución de 10 metros, basado en datos de Sentinel-1 y Sentinel-2. El mapa WorldCover incluye 11 clases de cobertura del suelo; sin embargo, para esta investigación, se utilizaron solo 7 clases que se encontraban presentes en el área de estudio: Cubierta arbórea, Matorrales, Praderas, Cultivos, Construido, Vegetación desnuda/escasa y Masas de agua permanentes [118]. Las categorías y su distribución se pueden apreciar en la Figura 3.4.

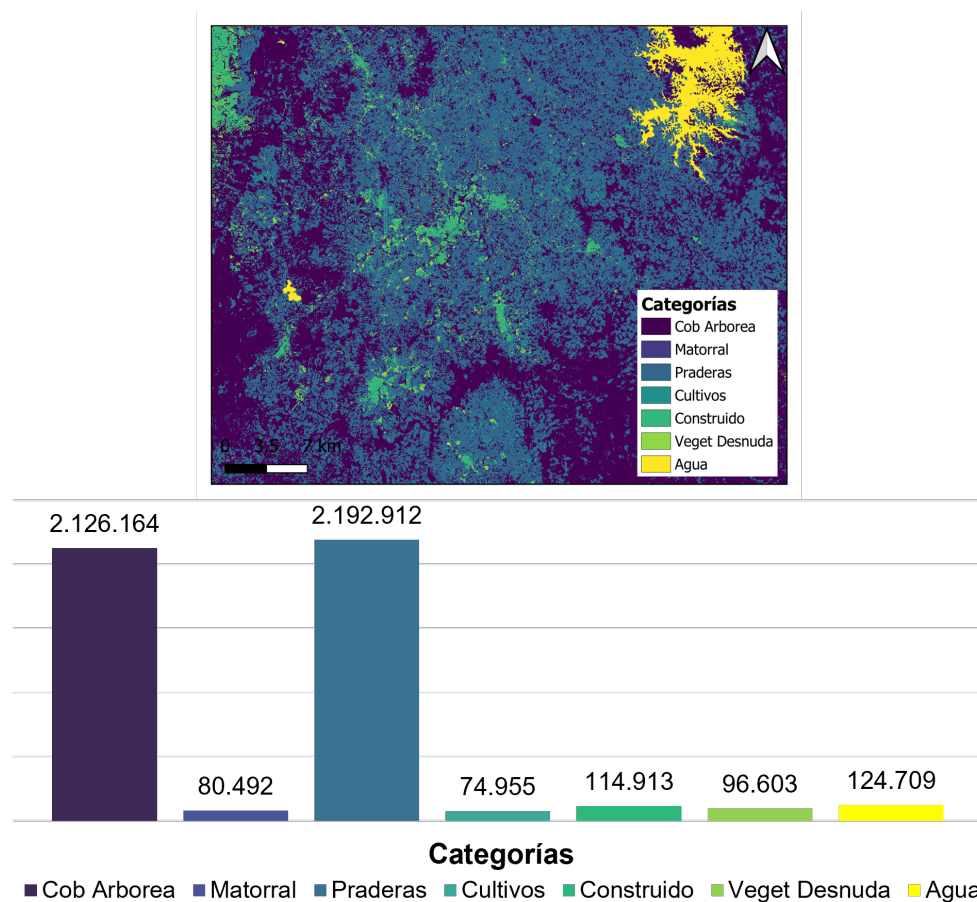


Figura 3.4: Mapa y distribución de las clases en el WorldCover. Fuente: Creado por la autora.

3.2. Extracción y Selección de Características

Los índices enriquecen significativamente el proceso de clasificación al desempeñar un papel crucial en la extracción de características relevantes de las imágenes, siendo fundamentales para una clasificación precisa del uso y cobertura del suelo. Por ello, para la selección y extracción de características en este estudio, se llevó a cabo un análisis exhaustivo de múltiples investigaciones relacionadas con la clasificación de usos y coberturas terrestres que integran datos multisensor e índices como características fundamentales de las imágenes [119–121]. A partir de este análisis, se seleccionaron y calcularon cuidadosamente los siguientes índices que permitieron capturar aspectos específicos

del espectro que proporcionan información crucial sobre las condiciones de diversas coberturas existentes en el área de estudio, como la vegetación, las áreas urbanas, los suelos expuestos y los cuerpos de agua.

3.2.1. Índices Ópticos

■ Índice de Vegetación de Diferencia Normalizada (NDVI)

El NDVI es un índice ampliamente utilizado para estimar la salud y la densidad de la vegetación. Se calcula utilizando las bandas de reflectancia en el infrarrojo cercano (NIR_1) y el rojo (RED) de las imágenes [122]. La fórmula para el cálculo es la siguiente:

$$NDVI = \frac{(NIR_1 - RED)}{(NIR_1 + RED)} \quad (3.1)$$

El NDVI varía de -1 a 1, donde valores más altos indican vegetación saludable, mientras que valores más bajos pueden indicar áreas sin vegetación o superficies artificiales.

■ Índice de Diferencia Normalizada Edificada (NDBI)

El NDBI permite llevar a cabo la estimación de zonas con superficies edificadas o en desarrollo de construcción frente a las habituales zonas naturalizadas con vegetación o desnudas. Este índice se calcula utilizando las bandas de reflectancia NIR_1 y el infrarrojo de onda corta (SWIR_1) para enfatizar las áreas urbanizadas fabricadas de las imágenes [123]. La fórmula para el cálculo es la siguiente:

$$NDBI = \frac{(SWIR_1 - NIR_1)}{(SWIR_1 + NIR_1)} \quad (3.2)$$

■ Índice de Calcinación Normalizado (NBRI)

El NBRI es otro índice que se utiliza para evaluar la salud y la densidad de la vegetación. Este aprovecha las bandas espectrales de infrarrojo cercano e infrarrojo de onda corta, que son sensibles a los cambios en la vegetación, para detectar áreas quemadas y monitorear la recuperación del ecosistema [124]. La fórmula para el cálculo es la siguiente:

$$NBRI = \frac{(NIR_1 - SWIR_2)}{(NIR_1 + SWIR_2)} \quad (3.3)$$

El NBRI se centra en destacar las áreas calcinadas a la vez que mitiga las diferencias en la iluminación y los efectos atmosféricos.

■ Índice de humedad de Diferencia Normalizada (NDMI)

El NDMI detecta los niveles de humedad en la vegetación utilizando una combinación de bandas espectrales del infrarrojo cercano NIR_1 y del infrarrojo de onda corta SWIR_1. La fórmula para el cálculo es la siguiente:

$$NDMI = \frac{(NIR_1 - SWIR_1)}{(NIR_1 + SWIR_1)} \quad (3.4)$$

El NDMI es un gran indicador de información sobre el estrés hídrico en los cultivos en una región.

■ Índice de Diferencia Normalizada de Agua y Material Seco (MNDWI)

El MNDWI se utiliza para resaltar características acuáticas, ya que el agua tiende a tener una reflectancia baja en la banda del verde y alta en la banda del SWIR_1. Al calcularlo se resaltan los cuerpos de agua, lo que facilita su detección y análisis en imágenes satelitales [125]. La fórmula para el cálculo es la siguiente:

$$MNDWI = \frac{Green - SWIR_1}{Green + SWIR_1} \quad (3.5)$$

3.2.2. Índices SAR

Para el cálculo de los índices SAR los valores de retrodispersión de las imágenes se convirtieron a decibeles (dB) con el fin de manejar un amplio rango de los datos y facilitar su análisis a escala logarítmica. Dicha transformación se realizó utilizando la Ecuación 3.6.

$$\sigma_0(dB) = 10 * \log_{10}(\sigma_0) \quad (3.6)$$

■ Índice de Vegetación Simétrico de Polarización Dual (DPSVI)

El DPSVI es un índice simétrico de doble polarización utilizado para la detección de la vegetación en imágenes SAR. Es útil para resaltar la vegetación y suprimir el ruido y otras características del terreno [126]. La fórmula de este se define como:

$$DPSVI = \frac{\sigma_{VV}^0 - \sigma_{VH}^0}{\sqrt{2}} \quad (3.7)$$

VV : es la intensidad de retrodispersión en la polarización vertical.

VH : es la intensidad de retrodispersión en la polarización horizontal.

Valores positivos de DPSVI indican áreas de vegetación, mientras que valores negativos pueden sugerir otras características del terreno.

■ Índice de Máxima Razón de Respuesta de Vegetación (RVIm)

El RVIm es un índice de vegetación basado en el coeficiente de retrodispersión. Es útil para detectar cambios en la vegetación a lo largo del tiempo [127]. La fórmula de este se define como:

$$RVIm = \frac{4 * \sigma_{VH}^0}{\sigma_{VH}^0 + \sigma_{VV}^0} \quad (3.8)$$

Un aumento en los valores de RVIm puede indicar un aumento en la vegetación entre los dos momentos temporales.

■ Ratio de Polarización (Pol)

El índice de polarización es una medida de la distribución angular de retrodispersión, que puede proporcionar información sobre la estructura y composición del terreno [128]. La fórmula de este se define como:

$$Pol = \frac{\sigma_{VH}^0 - \sigma_{VV}^0}{\sigma_{VH}^0 + \sigma_{VV}^0} \quad (3.9)$$

Valores altos de Pol pueden indicar superficies lisas o acuosas, mientras que valores bajos pueden sugerir superficies rugosas o terreno vegetado.

■ Índice de Contraste de Relación de Volumen de VV por VH (CRI)

El CRI es un índice que evalúa la coherencia entre las imágenes SAR en las polarizaciones vertical y horizontal. Es útil para identificar áreas con características de cambio o inestabilidad [129]. La fórmula de este se define como:

$$CRI = \frac{\sigma_{VV}^0}{\sigma_{VH}^0} \quad (3.10)$$

Valores altos de este pueden indicar áreas con cambios significativos entre las polarizaciones vertical y horizontal, mientras que valores bajos pueden sugerir áreas estables o menos propensas al cambio.

Finalmente, la combinación de estos índices permitió una caracterización más completa y precisa del paisaje, ya que cada tipo de índice aporta información complementaria sobre sus características.

3.3. Preprocesamiento

3.3.1. Normalización

Se organizaron las matrices de datos de las imágenes ópticas y SAR, asegurando que estuvieran uniformes en longitud y orden de bandas según los sensores. Los datos fueron sometidos a una limpieza exhaustiva para eliminar valores nulos o atípicos. Los valores atípicos se identificaron considerando los datos por encima del percentil 95 y por debajo del percentil 5, eliminándolos para evitar sesgos en el análisis. Posteriormente, los datos faltantes fueron imputados utilizando la función SimpleImputer de la librería Scikit-Learn, asignándoles el valor

de la media de un vecindario de 5x5 del pixel correspondiente. Este proceso fue crucial para asegurar la integridad y coherencia de los conjuntos de datos resultantes.

Posteriormente, se normalizaron todos los datos utilizando la función Standard Scaler de la librería Scikit-Learn, la cual estandariza los datos, convirtiendo la media igual a 0 y escalando los datos de forma que su varianza sea igual a 1, como se puede observar en la siguiente ecuación:

$$X_n = \frac{X - u}{s} \quad (3.11)$$

X : corresponde a cada pixel de la imagen.

u : es la media de los datos.

s : es la desviación estándar.

3.3.2. Fusión de Datos

La integración de datos ópticos y SAR en el proceso de clasificación proporciona una mejor discriminación entre las clases de interés, mejora la detección de cambios y aumenta la precisión general de los modelos de clasificación [78, 79]. Es así como la metodología de esta investigación incluye un enfoque integral para evaluar la fusión de datos usando 2 niveles de fusión: nivel de pixel y nivel de características.

Para esto, se crearon 7 conjuntos de datos: Óptico, SAR, Óptico + SAR, Óptico + SAR Ascendente, Óptico + SAR Descendente, Óptico + Índices Ópticos y Óptico + SAR + Índices Ópticos y SAR, Tabla 3.4. Para el conjunto de datos óptico, se utilizan las bandas B2, B3, B4, B5, B6, B7, B8, B8A, B11 y B12, con un nivel de fusión a nivel de píxel. En el caso del SAR, se emplean bandas VV ascendente y descendente, y VH ascendente y descendente, también con fusión a nivel de píxel. Las combinaciones de datos ópticos y SAR mantienen la fusión a nivel de píxel. Mientras que al integrar índices ópticos y SAR, la fusión se realiza a nivel de características.

Tabla 3.4: Composición de los conjuntos de datos

Dataset	Bandas	Nivel de fusión
Óptico	B2, B3, B4, B5, B6, B7, B8, B8A, B11 y B12	Pixel
SAR	VV asc, VV desc, VH asc y VH desc	Pixel
Óptico + SAR	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, VV asc, VH asc, VV desc y VH desc	Pixel
Óptico + SAR asc	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, VV asc y VH asc	Pixel
Óptico + SAR desc	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, VV desc y VH desc	Pixel
Óptico + Índices Ópticos	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, NDVI, NDBI, NBRI, NDMI y MNDWI	Características
Óptico + SAR + Índices Ópticos y SAR	B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12, VV asc, VV desc, VH asc, VH desc, NDVI, NDBI, NBRI, NDMI, MNDWI, DPSVI, RVIm, Pol y CRI	Características

3.3.3. Validación

Seguido a esto se realizó la partición de los datos en conjuntos de entrenamiento y prueba. Este procedimiento implicó la selección aleatoria del 20 % de los registros para entrenamiento, mientras que el 80 % restante se asignó al conjunto de pruebas, asegurando una representación adecuada de cada valor único en ambos conjuntos de datos.

La decisión de dividir los datos con un 20 % para el entrenamiento y un 80 % para las pruebas, aunque contraria a las proporciones habituales en el entrenamiento de modelos de aprendizaje automático, responde a la realidad del problema abordado. En este caso, la dificultad de obtener grandes volúmenes de datos etiquetados y libres de nubes, especialmente para imágenes satelitales, hace necesario optimizar el uso de los datos disponibles. Por lo cual, se considera prudente emplear una menor cantidad de datos para el entrenamiento, enfocándose en maximizar el volumen de datos para las pruebas. Este enfoque permite evaluar de manera más robusta la capacidad del modelo para generalizar en condiciones reales, donde la disponibilidad de datos etiquetados es restringida.

Además, utilizar un conjunto reducido de datos de entrenamiento somete al modelo a aprender de manera más eficiente a partir de patrones limitados, evitan-

do sobreajustes. De esta manera, el modelo es expuesto a una mayor cantidad de datos en la fase de prueba, permitiendo una evaluación exhaustiva y representativa de su capacidad para clasificar correctamente nuevas observaciones bajo condiciones atmosféricas diversas. Este enfoque permite que el modelo sea robusto frente a la escasez de datos óptimos y minimice el sesgo hacia clases mayoritarias, promoviendo una mejor generalización.

3.4. Clasificación Supervisada en Imágenes Ópticas y de Radar

En esta investigación implementamos dos modelos de aprendizaje automático: Random Forest y XGBoost. Estos modelos fueron seleccionados por su capacidad de manejar grandes volúmenes de datos, su precisión en la clasificación, y su habilidad para integrar múltiples fuentes de datos de teledetección [105, 112].

El modelo Random Forest consiste en una colección de clasificadores en forma de árboles de decisión, donde cada árbol se genera a partir de un subconjunto aleatorio de datos y características. Este enfoque introduce variabilidad entre los árboles, lo que mejora la robustez del modelo frente al sobreajuste. Cada árbol contribuye con una predicción, y el resultado final del modelo se basa en el voto mayoritario de los árboles individuales en el caso de una tarea de clasificación. La aleatoriedad en la selección de datos y características, junto con la combinación de los votos de múltiples árboles, permite que Random Forest logre una clasificación más estable y precisa [130, 131]. El modelo de Random Forest fue entrenado utilizando la librería scikit-learn, específicamente con la función `RandomForestClassifier`. Este modelo se configuró con los siguientes parámetros: profundidad máxima de los árboles `max_depth=100` para controlar el sobreajuste, el parámetro `random_state=0` garantiza la reproducibilidad de los resultados, y `n_jobs=-1` permite el uso de todos los procesadores disponibles para acelerar el entrenamiento.

Por otro lado, XGBoost se basa en el método de gradient boosting, donde se construye un clasificador de manera aditiva. En cada iteración, un nuevo árbol débil es entrenado para predecir las residuales de los clasificadores anteriores, mejorando constantemente el error acumulado. Este proceso iterativo y aditivo refina la capacidad del modelo para realizar predicciones precisas y robustas, aumentando significativamente la precisión general del clasificador final [113]. El modelo XGBoost fue entrenado utilizando la librería xgboost, específicamente con la función `XGBClassifier`. Este modelo se configuró con los siguientes parámetros: el parámetro `objective="multi:softprob"` se utiliza para especificar que el

objetivo es la clasificación multiclase, donde el modelo predice las probabilidades de cada clase, mientras que `random_state=42` garantiza la reproducibilidad de los resultados.

Tabla 3.5: Parámetros utilizados para los modelos

Modelo	Parámetro
Random Forest	<code>max_depth=100</code>
	<code>random_state=0</code>
	<code>n_jobs=-1</code>
	<code>criterion=gini</code>
	<code>n_estimators=100</code>
	<code>max_features=sqrt</code>
XGBoost	<code>objective="multi:softprob"</code>
	<code>random_state=42</code>
	<code>n_estimators=100</code>
	<code>loss=log_loss</code>
	<code>learning_rate=0.1</code>

Para evaluar el rendimiento de los modelos mencionados, se llevaron a cabo pruebas utilizando los siete subconjuntos de datos descritos anteriormente en la Tabla 3.4. Estos subconjuntos incluyen combinaciones de datos ópticos, SAR y diversos índices espectrales, con niveles de fusión tanto a nivel de píxel como de características. Los datos utilizados para estas pruebas consisten en la fusión de imágenes SAR con imágenes ópticas que tienen un 0-30 % de cobertura nubosa. Además, se empleó el mapa CLC como referencia para validar las clasificaciones. Este enfoque facilita analizar la efectividad de cada subconjunto de datos y de cada modelo en diferentes condiciones de datos, proporcionando una evaluación exhaustiva de su desempeño en la clasificación LULC.

3.5. Evaluación de la Clasificación LULC

Con el fin de explorar la integración de datos SAR con imágenes ópticas para evaluar su impacto en condiciones de alta nubosidad, se propuso el siguiente escenario:

Al identificar el modelo con mejor desempeño en la etapa anterior y evaluar su rendimiento utilizando los siete subconjuntos de datos descritos en la Tabla 3.4.

Continúa esta evaluación, empleando imágenes SAR combinadas con diferentes imágenes ópticas, clasificadas según su porcentaje de cobertura nubosa en tres rangos: 30-50 %, 50-70 % y 70-90 %. La validación de las clasificaciones se llevó a cabo utilizando el mapa CLC y, adicionalmente, el mapa WorldCover de la ESA. Este enfoque metodológico asegura que el modelo seleccionado no solo sea preciso bajo condiciones ideales, sino que también aprovecha las ventajas de cada sensor para superar las limitaciones impuestas por la nubosidad, mejorando así la capacidad de monitoreo y clasificación de LULC.

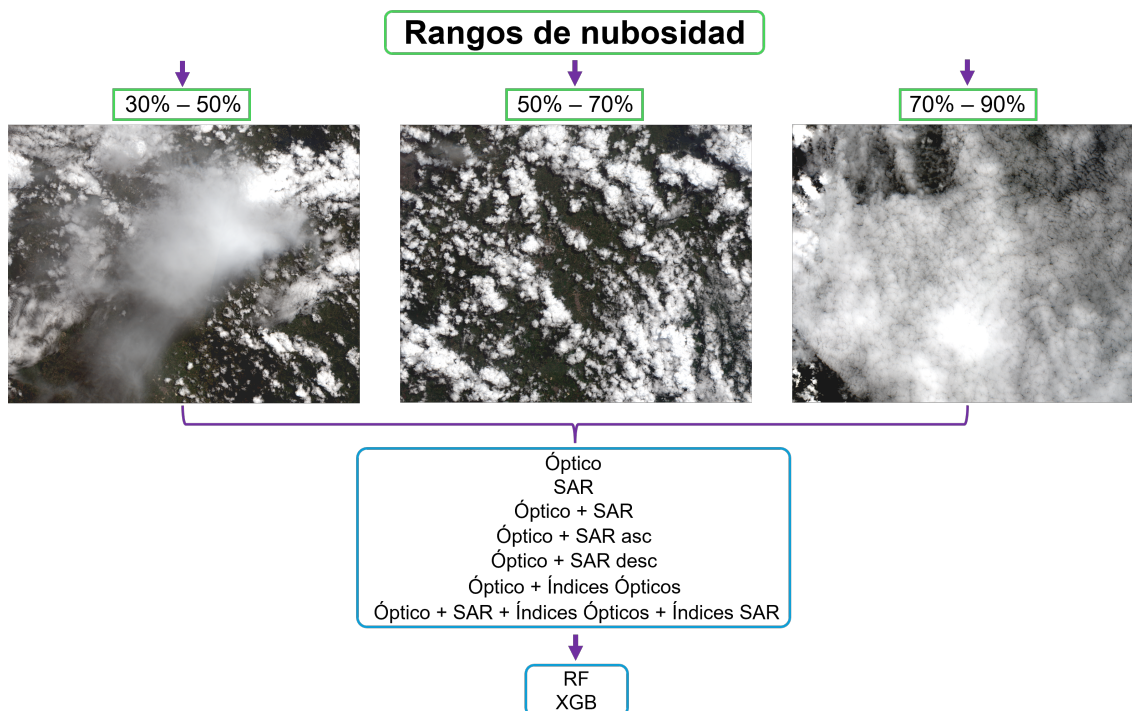


Figura 3.5: Modalidad de Evaluación de la Clasificación LULC. Fuente: Creado por la autora.

3.6. Métricas para validación del rendimiento

En esta sección, se detalla el proceso con las métricas utilizadas para evaluar el rendimiento de los modelos de clasificación de uso y cobertura del suelo. Por lo cual, esta validación es esencial para garantizar la precisión y la confiabilidad de los resultados obtenidos. Las métricas utilizadas en este estudio son las siguientes:

■ Matriz de Confusión

Con la matriz de confusión se muestra la relación entre las clasificaciones reales y las clasificaciones predichas por el modelo [132]. Su resultado se organiza en forma de tabla incluyendo cuatro valores importantes:

Verdaderos Positivos (VP): Representa la cantidad de instancias correctamente clasificadas como positivas.

Verdaderos Negativos (VN): Indica la cantidad de instancias correctamente clasificadas como negativas.

Falsos Positivos (FP): Representa la cantidad de instancias que fueron incorrectamente clasificadas como positivas cuando en realidad son negativas (error tipo I).

Falsos Negativos (FN): Indica la cantidad de instancias que fueron incorrectamente clasificadas como negativas cuando en realidad son positivas (error tipo II).

		Predicción	
		Positivos	Negativos
Actual	Positivos	VP	FN
	Negativos	FP	VN

Figura 3.6: Matriz de Confusión. Fuente: Creado por la autora.

■ Exactitud (Accuracy)

Se calculó la exactitud para medir la proporción de clasificaciones correctas realizadas por el modelo en relación con el total de clasificaciones. Obtener un valor alto de exactitud indica que el modelo ha realizado la mayoría de las clasificaciones correctamente [133]. Se utilizó la siguiente fórmula:

$$Accuracy = \frac{(VP + VN)}{(VP + VN + FP + FN)} \quad (3.12)$$

■ Puntaje F1 (F1-score)

F1-score es una métrica que proporciona una medida equilibrada al combinar Precision y Recall en una sola medida. Es de gran utilidad cuando se busca un equilibrio entre ambas métricas [99]. Para el cálculo de esta se utilizó la siguiente fórmula:

$$F1 - score = \frac{2 \cdot (Precision \cdot Recall)}{(Precision + Recall)} \quad (3.13)$$

Precision $(VP)/(VP + FP)$: es la proporción de todas las clasificaciones positivas del modelo que son realmente positivas.

Recall $(VP)/(VP + FN)$: es la proporción de todos los positivos reales que se clasificaron correctamente como positivos.

■ Índice Kappa

Para el cálculo del índice Kappa, también conocido como coeficiente de concordancia de Cohen, se utilizó la siguiente ecuación como medida estadística para evaluar la fiabilidad o precisión de los métodos de clasificación en relación con un estándar de referencia [134]:

$$k = \frac{(P_o - P_e)}{(1 - P_e)} \quad (3.14)$$

P_o : es la proporción de concordancia observada.

P_e : es la proporción de concordancia esperada.

El valor de Kappa varía entre -1 y 1. Un valor de 1 indica un acuerdo perfecto, mientras que un valor de 0 indica que el acuerdo observado es equivalente al que se esperaría por azar. Valores negativos indican un acuerdo inferior al que se esperaría por azar.

Finalmente, es importante mencionar la participación del F1-Score y el coeficiente Kappa como métricas esenciales para evaluar el rendimiento de los modelos cuando se trabaja con datos desbalanceados. A diferencia del Accuracy, que simplemente mide la proporción de predicciones correctas, el F1-Score combina la precisión y el recall en una única métrica, lo que permite capturar mejor el rendimiento en escenarios donde una clase puede dominar sobre las otras. Por otro lado, el coeficiente Kappa considera el acuerdo entre las predicciones del modelo y las observaciones reales, ajustando los resultados en función del azar. Esto lo convierte en una herramienta útil para evaluar modelos en conjuntos de datos desbalanceados, ya que da una visión más completa del verdadero rendimiento del modelo, tomando en cuenta la distribución desigual de las clases.

Capítulo 4

Resultados y Discusión

Contenido

4.1 Dataset	42
4.2 Clasificación Supervisada en Imágenes Ópticas y de Radar .	44
4.3 Evaluación de la Clasificación LULC	50

4.1. Dataset

Como resultado de la metodología aplicada para la base de datos de esta investigación, se obtuvieron un total de siete conjuntos de datos. Cada conjunto fue seleccionado en función de criterios específicos relacionados con el aporte de información según el tipo de imagen, garantizando así la relevancia y la calidad de la información recopilada. Estos conjuntos están descritos en la Tabla 4.1, donde, se proporciona la información específica sobre las bandas de las imágenes utilizadas y el número de píxeles correspondientes por conjunto de datos según el número de bandas asociado.

Tabla 4.1: Base de datos según las bandas de las imágenes y el número de píxeles

Dataset	Bandas	Total píxeles	Entrenamiento	Validación
Óptico	10			
SAR	4			
Óptico + SAR	14			
Óptico + SAR asc	12	4810748	962146	3848602
Óptico + SAR desc	12			
Óptico + Índices Ópticos	15			
Óptico + SAR + Índices Ópticos y SAR	23			

Cada conjunto de datos incluye imágenes con un número específico de bandas, lo cual es crucial para el análisis espectral y la clasificación de las imágenes. La fusión de datos ópticos y SAR a nivel de píxel y características, ha permitido obtener una amplia gama de datos con un número de bandas que varía de 4 a 23, incrementando la complejidad y la riqueza de la información contenida en cada imagen. Asimismo, la variedad de bandas de imágenes incluidas en cada conjunto permite un análisis multifacético, considerando diferentes espectros y características de las imágenes. En consecuencia, la metodología aplicada ha resultado en la obtención de una base de datos robusta y variada que cumple con los requisitos necesarios para llevar a cabo los análisis propuestos en los objetivos de esta investigación.

Después de implementar la etapa de preprocesamiento, en la cual se realizó la división de los datos en una proporción aleatoria del 20 % para entrenamiento y 80 % para pruebas, se obtuvo una visualización detallada de la distribución de datos por clases, proporcionando una comprensión clara de la variedad de las clases dentro de los conjuntos de entrenamiento y prueba (Figura 4.1). Específicamente, la Figura 4.1a presenta la distribución de los datos con el mapa de referencia CLC, mientras que la Figura 4.1b muestra la distribución empleando el mapa de referencia WorldCover.

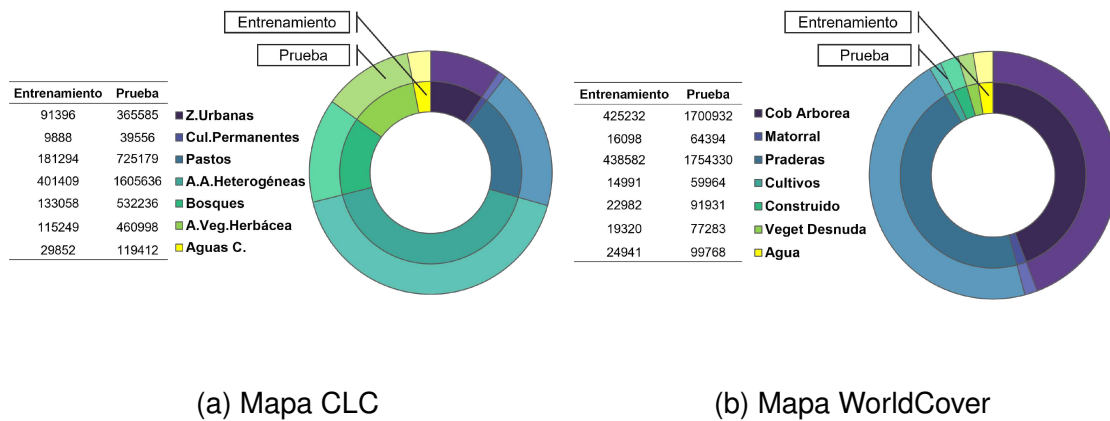


Figura 4.1: Distribución del set de entrenamiento y validación por clases y mapa de referencia. Fuente: Creado por la autora.

4.2. Clasificación Supervisada en Imágenes Ópticas y de Radar

La Tabla 4.2 presenta los resultados obtenidos para los modelos Random Forest y XGBoost en términos de precisión (Accuracy), F1-Score y coeficiente Kappa. Estos resultados corresponden a los siete conjuntos de datos descritos en la Tabla 4.1, los cuales combinan imágenes ópticas con un rango de nubosidad del 0 al 30 % y SAR. Siendo evaluados en esta etapa utilizando únicamente el mapa de referencia del CLC.

Tabla 4.2: Resultados de Clasificación Supervisada con Imágenes SAR y Ópticas del 0 al 30 % de nubes y mapa CLC

Conjunto de datos	Modelo	Accuracy (%)	F1-Score (%)	Kappa
Óptico	RF	59,69	57,41	0,424
	XGB	59,17	56,31	0,412
SAR	RF	42,80	36,79	0,132
	XGB	45,99	35,85	0,135
Óptico + SAR	RF	60,16	57,62	0,428
	XGB	59,69	57,04	0,421
Óptico + SAR asc	RF	59,85	57,48	0,425
	XGB	59,35	56,68	0,416
Óptico + SAR desc	RF	59,91	57,59	0,427
	XGB	59,33	56,59	0,416
Óptico + Índices Ópticos	RF	59,62	57,40	0,424
	XGB	59,43	56,89	0,419
Óptico + SAR + Índices Ópticos y SAR	RF	61,54	59,42	0,451
	XGB	60, 91	58,68	0,441

Los resultados indican que la combinación de datos ópticos y SAR, junto con índices ópticos y SAR, resaltado en color verde, proporcionó los mejores resultados de clasificación. Específicamente, el modelo Random Forest alcanzó una precisión del 61.54 %, un F1-Score del 59.42 % y un coeficiente Kappa de 0.451, superando a los demás conjuntos de datos y modelos evaluados.

En comparación, los conjuntos de datos que utilizaron solo datos ópticos o SAR obtuvieron resultados significativamente inferiores, resaltados en color rojo. Por ejemplo, el uso exclusivo de datos SAR con el modelo RF resultó en una precisión del 42.80 %, un F1-Score del 36.79 % y un coeficiente Kappa de 0.132, lo que destaca la importancia de la fusión de datos para mejorar la precisión de las clasificaciones.

La literatura existente respalda estos hallazgos. Según estudios como [79, 135], la combinación de imágenes Sentinel 1 y Sentinel 2 también ha demostrado mejorar la clasificación general en comparación con el uso de las imágenes de forma independiente. Ya que la fusión de estas imágenes contribuye a mejorar la precisión geométrica, el enriquecimiento de características y la nitidez de la imagen.

Por lo cual, los mejores resultados que fueron resaltados en color verde en la Tabla 4.2 destacan la superioridad de la metodología implementada. Ya que, estos resultados demuestran que la integración de múltiples fuentes de datos y la utilización de índices espectrales adicionales pueden mejorar significativamente el rendimiento de los modelos de clasificación LULC. Los índices proporcionan una mayor cantidad de información espectral y espacial, lo que contribuye a una mejor discriminación entre clases. Además, la validación con el mapa CLC proporciona una evaluación exhaustiva del desempeño en diversas condiciones de datos para este enfoque metodológico.

Ahora, la matriz de confusión presentada en la Figura 4.2 muestra los resultados de la clasificación utilizando el modelo Random Forest con imágenes ópticas en un rango de nubosidad del 0-30 %, combinadas con datos SAR y diversos índices espectrales. Este modelo, que fue el mejor resultado obtenido en esta etapa, proporciona una visión detallada del desempeño de la clasificación para cada clase. La validación de estos resultados se realizó con respecto al mapa de referencia del CLC, el cual presenta algunos problemas en sus clasificaciones debido a la resolución y a la precisión de las etiquetas.

Realidad	Z.Urbanas	35.3% 128904/365585	0.3% 935	2.5% 9115	59.7% 218268	1.6% 5718	0.5% 1960	0.2% 685
	Cul.Permanentes	11.2% 4430	25.3% 9996/39556	3.4% 1337	53.7% 21236	2.8% 1108	3.0% 1167	0.7% 282
	Pastos	1.2% 8684	0.1% 578	34.5% 250497/725179	56.8% 412080	3.2% 23097	4.0% 28849	0.2% 1394
	A.A.Heterogéneas	1.9% 30782	0.1% 1835	6.1% 97589	83.9% 1346672/1605636	3.4% 54998	4.4% 69914	0.2% 3846
	Bosques	0.6% 2981	0.0% 116	1.5% 8105	16.5% 88060	67.3% 358076/532236	12.9% 68513	1.2% 6385
	A.Veg.Herbácea	0.6% 2562	0.0% 48	3.2% 14908	35.3% 162943	23.1% 106656	37.6% 173405/460998	0.1% 476
	Aguas C.	0.9% 1033	0.1% 86	0.5% 620	9.9% 11843	3.6% 4323	0.5% 643	84.5% 100864/119412
		Z.Urbanas	Cul.Permanentes	Pastos	A.A.Heterogéneas	Bosques	A.Veg.Herbácea	Aguas C.
		Predicción						

Figura 4.2: Matriz de Confusión de Modelo Random Forest y conjunto de datos Óptico + SAR + Índices Completos. Fuente: Creado por la autora.

En la diagonal principal de la matriz se encuentran los valores correctamente clasificados para cada clase. Por ejemplo, la clase 'Aguas Continentales' presenta la tasa más alta de precisión con el 84.5%, lo que indica un excelente desempeño del modelo para esta clase específica. De manera similar, la clase 'Áreas Agrícolas Heterogéneas' presenta una precisión notable con el 83.9% de los píxeles correctamente clasificados.

Sin embargo, algunas clases presentan confusiones significativas. Por ejemplo, las 'Áreas de Vegetación Herbácea' tienen una tasa de precisión del 37.6% y una alta confusión con la clase 'Áreas Agrícolas Heterogéneas' (35.3%) y 'Bosques' (23.1%). Este patrón de confusión indica la necesidad de un análisis más profundo para mejorar la discriminación entre estas clases vegetativas. Además, las 'Zonas Urbanas' también presentan desafíos en la clasificación, con una precisión del 35.3% y una notable cantidad de píxeles clasificados erróneamente como 'Áreas Agrícolas Heterogéneas' (59.7%). Esto sugiere que la distinción entre estas clases es problemática, posiblemente debido a características espectrales similares.

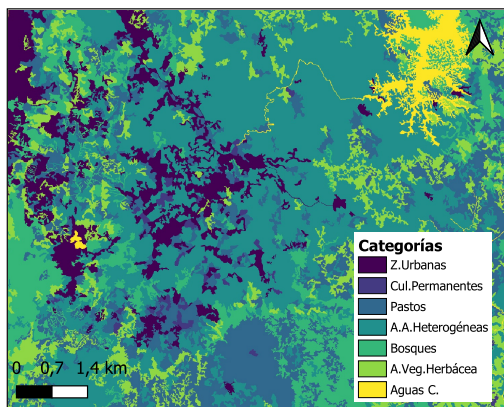
La validación utilizando el mapa de referencia del CLC revela no solo la capacidad del modelo para clasificar diversas clases, sino también las limitaciones inherentes al propio mapa de referencia. Aunque el CLC es una herramienta valiosa para la validación, presenta problemas en sus clasificaciones que pueden influir en la evaluación del rendimiento del modelo. La resolución relativamente baja del CLC y la posible inexactitud en la asignación de etiquetas pueden contribuir a las confusiones observadas en la matriz de confusión. Como se indica en [136], esta es una metodología que requiere ajustes para mejorar la evaluación de coberturas a escalas más detalladas.

En resumen, la matriz de confusión revela que, aunque el modelo Random Forest muestra un rendimiento robusto en general, especialmente para clases como 'Áreas Agrícolas Heterogéneas' y 'Aguas Continentales', existen áreas donde la precisión puede mejorarse. Las confusiones entre clases relacionadas con vegetación y zonas urbanas destacan la complejidad de la clasificación en entornos mixtos y la influencia de la presencia de nubes en la precisión de los datos ópticos.

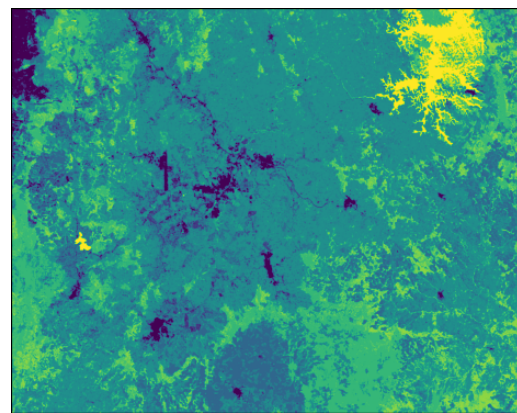
La Figura 4.3 presenta una comparación visual entre la imagen óptica con un 0-30% de nubosidad 4.3a, el mapa de referencia del CLC 4.3b y la clasificación obtenida con el mejor modelo 4.3c para este escenario.



(a) Imagen Óptica 0-30 % de nubes



(b) Mapa de referencia CLC



(c) Clasificación

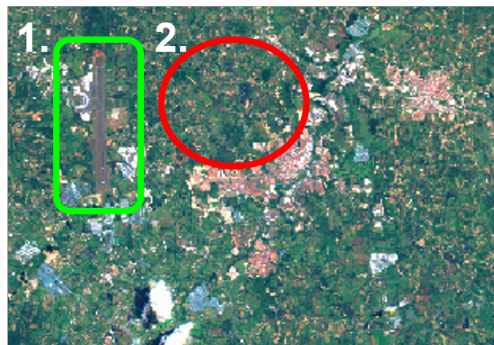
Figura 4.3: Comparativo de la Imagen real del área de estudio óptica en un rango del 0 al 30 % de nubes, con el mapa de referencia CLC y la imagen de la clasificación del modelo. Fuente: Creado por la autora.

La imagen 4.3a muestra una captura óptica de la zona de estudio, donde se puede observar una cobertura nubosa parcial que oscila entre el 0 y el 30 %. La presencia de nubes es evidente en algunas áreas, lo que influye negativamente en la calidad de los datos y, en consecuencia, en la precisión de la clasificación. La imagen 4.3b representa el mapa de referencia del CLC, que clasifica el territorio en diversas categorías de uso y cobertura del suelo. La imagen 4.3c muestra la clasificación realizada por el modelo Random Forest, entrenado con datos ópticos, SAR y diversos índices.

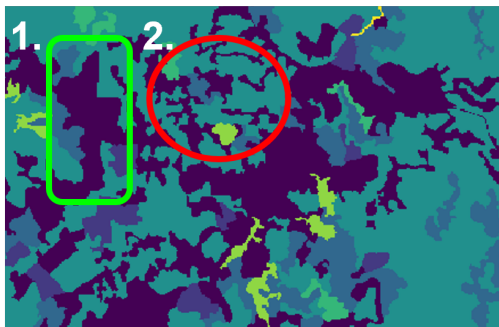
Comparando las imágenes 4.3b y 4.3c, se observa que la clasificación del modelo Random Forest se alinea de manera efectiva con las categorías del CLC en muchas áreas. No obstante, existen discrepancias en ciertas clases, como

las Zonas Urbanas y Áreas de Vegetación Herbácea, donde la precisión puede ser afectada por las limitaciones inherentes al mapa de referencia del CLC y las características espectrales similares de diferentes clases.

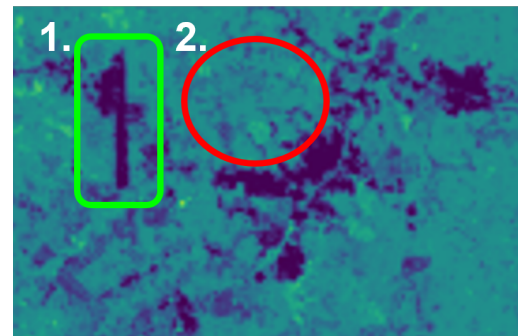
La comparación visual entre la clasificación y el mapa de referencia del CLC revela que el modelo logra una alta correspondencia en varias clases, a pesar de la presencia de nubosidad en la imagen óptica. Por lo cual, la incorporación de datos SAR en la clasificación ha demostrado ser crucial para mitigar los efectos adversos de la nubosidad en las imágenes ópticas. Los atributos SAR, que son insensibles a las condiciones atmosféricas y pueden penetrar a través de las nubes, proporcionan información complementaria que mejora significativamente la precisión de la clasificación. Esto se refleja en la capacidad del modelo para identificar correctamente varias categorías, incluso en áreas donde la nubosidad puede haber dificultado la interpretación óptica.



(a) Zoom de la imagen Óptica 0-30 % de nubes



(b) Zoom del mapa de referencia CLC



(c) Zoom de la Clasificación

Figura 4.4: Comparativo con acercamiento en dos sectores de la imagen real del área de estudio óptica en un rango del 0 al 30 % de nubes, con el mapa de referencia CLC y la imagen de la clasificación del modelo. Fuente: Creado por la autora.

Observando los detalles de las imágenes en la Figura 4.4, podemos notar varios puntos importantes. En la sección señalada con el círculo rojo, se aprecia que en el mapa de referencia CLC 4.4b se encuentran predominantemente clases de 'Zonas Urbanas' y 'Áreas Agrícolas Heterogéneas'. Sin embargo, al validar esta misma área en la imagen óptica real 4.4a, estas categorías y delimitaciones no coinciden de manera precisa con las definidas en el mapa de referencia.

Al comparar esta sección con la imagen clasificada por el modelo 4.4c, observamos que el modelo parece tener una mayor concordancia con la realidad, a pesar de que la validación contra el mapa de referencia penalice la precisión del modelo. Esto se debe a que, como se muestra en la matriz de confusión, hay una discordancia significativa entre las clases de 'Zonas Urbanas' y 'Áreas Agrícolas Heterogéneas' en el mapa de referencia.

En la sección demarcada con el rectángulo verde, donde se encuentra ubicado el aeropuerto José María Córdova, podemos observar que este tiene una forma bien definida en la imagen óptica real. Comparando con lo definido en el mapa de referencia CLC, notamos que la demarcación es poco precisa. No obstante, en la imagen de la clasificación del modelo, el aeropuerto está delimitado de manera más minuciosa y correctamente clasificado en su categoría correspondiente.

Estos resultados indican que el modelo de clasificación podría estar proporcionando una representación más precisa de ciertas áreas específicas en comparación con el mapa de referencia CLC, a pesar de las discrepancias en otras áreas que afectan la precisión general del modelo.

4.3. Evaluación de la Clasificación LULC

La Tabla 4.3 presenta los resultados de la segunda etapa del proceso de clasificación supervisada utilizando el mejor modelo resultante de la primera etapa, que es Random Forest. En esta fase, se evaluaron solo cuatro de los siete conjuntos de datos iniciales en diferentes combinaciones de imágenes ópticas, SAR y sus índices, validados en este caso con el mapa de referencia WorldCover. Este mapa proporciona información mucho más precisa, basada en los mismos sensores utilizados para esta investigación y en un periodo de tiempo más reciente en comparación con el CLC. Los conjuntos de datos fueron evaluados bajo cuatro escenarios de rangos de nubosidad: 0-30 %, 30-50 %, 50-70 %, y 70-90 %.

Tabla 4.3: Resultados de Clasificación Supervisada con modelo RF e Imágenes SAR y Ópticas en diferentes rangos de nubes

Porcentaje de nubes	Conjunto de datos	Accuracy (%)	F1-Score (%)	Kappa
0 - 30 %	Óptico	80,61	79,24	0,664
	SAR	62,96	61,15	0,346
	Óptico + SAR	81,21	79,78	0,674
	Óptico + SAR + Índices Ópticos y SAR	81,92	80,57	0,687
30 - 50 %	Óptico	68,37	66,77	0,443
	SAR	62,96	61,15	0,346
	Óptico + SAR	71,15	69,27	0,492
	Óptico + SAR + Índices Ópticos y SAR	74,11	72,38	0,546
50 - 70 %	Óptico	60,85	59,35	0,307
	SAR	62,96	61,15	0,346
	Óptico + SAR	68,39	66,62	0,444
	Óptico + SAR + Índices Ópticos y SAR	71,38	69,68	0,498
70 - 90 %	Óptico	63,40	61,36	0,353
	SAR	62,96	61,15	0,346
	Óptico + SAR	68,78	66,85	0,450
	Óptico + SAR + Índices Ópticos y SAR	71,71	69,85	0,502

Para el rango de nubosidad del 0-30 %, se observa que el mejor resultado se obtuvo con el conjunto de datos que combina imágenes ópticas, SAR e índices espectrales de ambos, alcanzando una precisión del 81.92 %, un F1-Score del 80.57 % y un coeficiente Kappa de 0.687, resaltados en verde en la tabla.

En contraste, el rango de nubosidad del 30-50 % muestra que, aunque los resultados disminuyen en general, el conjunto de datos que incluye SAR e índices espectrales sigue proporcionando mejores resultados comparados con solo datos ópticos o la combinación simple de óptico y SAR. Aquí, la precisión más alta es del 74.11 %, el F1-Score del 72.38 % y el coeficiente Kappa de 0.546, nuevamente con el conjunto de datos que incluye índices, lo que subraya la importancia de estos índices en mejorar la clasificación en presencia de mayor nubosidad.

Para el rango de nubosidad del 50-70 %, los resultados más bajos, resaltados en rojo, indican que la clasificación se vuelve más desafiante a medida que aumenta la presencia de nubes. En este caso, el conjunto de datos solo óptico obtiene la precisión más baja, incluso por debajo de los datos solo SAR, con una precisión de 60.85 %, un F1-Score de 59.35 % y un coeficiente Kappa de 0.307, reflejando la dificultad de mantener una alta precisión sin el apoyo de datos SAR.

Sin embargo, la combinación de datos ópticos y SAR, junto con los índices, sigue mejorando los resultados, alcanzando una precisión del 71.38 % y un F1-Score de 69.68 %, lo que demuestra la utilidad continua de los datos SAR y los índices espectrales.

En el rango de nubosidad del 70-90 %, se observa que, aunque los resultados generales son bajos, superan los del escenario anterior a pesar de la presencia extrema de nubes. El conjunto de datos que incluye SAR e índices espectrales sigue proporcionando el mejor desempeño, con una precisión de 71.71 %, un F1-Score de 69.85 % y un coeficiente Kappa de 0.502. Este resultado, aunque inferior a los de rangos con menor nubosidad, sigue siendo significativo y refuerza la importancia del uso de datos SAR para mejorar la robustez de la clasificación bajo condiciones adversas.

Consistente con estos resultados, el estudio de [137] presenta un análisis en el que los autores dividen los estados de nubosidad en niveles, cada uno con rangos de porcentajes de nubes diferentes. Destacan que esta segmentación es esencial para abordar las variaciones de interferencia en los diferentes niveles de nubosidad y mitigar su impacto. La conclusión del estudio es que, al tener en cuenta los diferentes niveles de nubosidad y al aplicar estrategias específicas para cada uno de ellos, se logró mejorar de manera significativa la capacidad de discriminación bajo condiciones de nubosidad.

En la Tabla 4.3 también se presentan los resultados del F1-Score, una métrica relevante para evaluar la precisión y robustez del modelo de clasificación. Para destacar mejor las diferencias entre los resultados obtenidos con datos ideales (porcentajes de nubosidad muy bajos) y datos más realistas (porcentajes de nubosidad muy altos), se analizan los rangos de nubosidad de 0-30 % y 70-90 %. En el rango de 0-30 %, el F1-Score para el conjunto de datos de solo imágenes ópticas es del 79.24 %. Al incorporar datos SAR, el F1-Score aumenta ligeramente a 79.78 %. Sin embargo, la mayor mejora se observa cuando también se añaden los índices, alcanzando un F1-Score de 80.57 %. Este incremento de poco más del 1.3 %, aunque pequeño, resalta la importancia de los índices y los datos SAR en mejorar la precisión de la clasificación, incluso en condiciones de nubosidad leve.

En el rango de 70-90 % de nubosidad, el impacto de los datos SAR es aún más evidente. El F1-Score para el conjunto de datos de solo imágenes ópticas es significativamente más bajo, con un 61.36 %. En la fusión de datos ópticos y SAR mejora el F1-Score a 66.85 %. Y la incorporación de índices eleva el F1-Score a 69.85 %. Estos resultados subrayan la relevancia de los datos SAR y los índices espectrales para mitigar los efectos adversos de la alta nubosidad, proporcionando una mejora sustancial de un poco más del 8 % entre el conjunto

de datos solo ópticos y la fusión con datos SAR e índices.

El F1-Score es una métrica fundamental en la evaluación de modelos de clasificación, ya que considera tanto la precisión como el recall del modelo, proporcionando una visión equilibrada del rendimiento [114, 138]. En el contexto de esta evaluación, los resultados demuestran que la combinación de imágenes ópticas con datos SAR y sus índices ofrece una ventaja significativa en la clasificación de uso del suelo y cobertura vegetal, especialmente en condiciones de nubosidad extrema. Los datos SAR son cruciales porque pueden penetrar a través de las nubes y proporcionar información complementaria que no está disponible en las imágenes ópticas afectadas por las nubes.

Matriz de confusión con nubosidad del 0-30 %

La siguiente matriz de confusión (Figura 4.5) muestra los resultados de la clasificación obtenida con el modelo Random Forest en el escenario que combina datos ópticos, SAR e índices espectrales, con un rango de nubosidad del 0-30 % y utilizando el mapa de referencia WorldCover.

	Cob Arborea	Matorral	Pradera	Cultivos	Construido	Veget Desnuda	Agua
Cob Arborea	82.9% 1410617/1700932	0.1% 1558	16.5% 279830	0.0% 475	0.3% 4831	0.1% 2045	0.1% 1576
Matorral	69.3% 44625	3.8% 2468/64394	26.9% 17292	0.0% 1	0.0% 5	0.0% 3	
Pradera	10.3% 179821	0.1% 1269	88.6% 1555179/1754330	0.1% 2447	0.2% 4175	0.3% 5927	0.3% 5512
Cultivos	8.2% 4940		73.7% 44193	11.0% 6596/59964	1.3% 794	5.2% 3090	0.6% 351
Construido	6.8% 6223		14.1% 12986	0.3% 301	74.2% 68250/91931	4.5% 4158	0.0% 13
Veget Desnuda	10.1% 7829		53.0% 40965	1.6% 1217	14.0% 10848	19.8% 15323/77283	1.4% 1101
Agua	1.3% 1337		3.9% 3855	0.0% 10	0.1% 141	0.2% 155	94.5% 94270/99768
	Cob Arborea	Matorral	Pradera	Cultivos	Construido	Veget Desnuda	Agua

Figura 4.5: Matriz de Confusión de Modelo Random Forest y conjunto de datos Óptico + SAR + Índices Completos en escenario de 0 - 30 % de nubes. Fuente: Creado por la autora.

Uno de los aspectos más destacados es la alta precisión en la clase 'Cobertura Arbórea', con un 82.9% de píxeles correctamente clasificados. Esta alta precisión subraya la efectividad del modelo para identificar áreas de esta categoría, lo cual es relevante para el monitoreo forestal. La clase 'Pradera' también muestra una notable precisión del 88.6%, lo que indica una fuerte capacidad del modelo para discriminar esta clase, probablemente debido a la combinación de

datos SAR e índices espectrales que proporcionan información adicional sobre la textura y la estructura de la vegetación.

En contraste, la clase 'Vegetación Desnuda' presenta una menor precisión relativa, con un 19.8 %. Esta menor precisión puede deberse a la similitud espectral con otras clases, como 'Pradera', con la que muestra cierta confusión. La clase 'Matorral', con una precisión del 3.8 %, también muestra una notable confusión con 'Cobertura Arbórea' y 'Pradera'. Como se menciona en [139], existen clases como cultivos arbóreos y bosques, o suelos desnudos y tejido urbano, que presentan comportamientos o características espectrales similares, lo que puede generar confusión en la clasificación.

Finalmente, la clase 'Agua' se destaca por su alta precisión del 94.5 %, lo que demuestra la capacidad del modelo para identificar cuerpos de agua, una tarea que se beneficia significativamente de la penetración del radar SAR a través de nubes y su capacidad para distinguir superficies acuáticas de manera efectiva. Por lo cual, estos resultados validan la metodología de integrar múltiples fuentes de datos para lograr una clasificación más precisa y confiable en diversas condiciones atmosféricas y de terreno.

Matriz de confusión con nubosidad del 70-90 %

Esta matriz de confusión (Figura 4.6) muestra los resultados de la clasificación obtenida con el mismo experimento anterior, pero en un escenario con rango de nubosidad del 70-90 %.

Realidad	Cob Arborea	73.9% 1257022/1700932	0.0% 10	25.8% 438169	0.0% 166	0.2% 4188	0.0% 76	0.1% 1301
	Matorral	67.4% 43382	0.1% 51/64394	32.5% 20943		0.0% 16		0.0% 2
	Pradera	21.8% 383318	0.0% 29	77.7% 1362894/1754330	0.0% 666	0.1% 1574	0.0% 204	0.3% 5645
	Cultivos	13.8% 8246		80.8% 48452	4.0% 2380/59964	0.5% 315	0.3% 201	0.6% 370
	Construido	29.2% 26882		22.9% 21013	0.1% 134	47.4% 43583/91931	0.3% 283	0.0% 36
	Veget Desnuda	26.3% 20308		65.0% 50210	0.5% 392	5.5% 4249	1.6% 1210/77283	1.2% 914
	Agua	1.4% 1399		5.6% 5571	0.0% 14	0.1% 53	0.0% 25	92.9% 92706/99768
		Cob Arborea	Matorral	Pradera	Cultivos	Construido	Veget Desnuda	Agua
		Predicción						

Figura 4.6: Matriz de Confusión de Modelo Random Forest y conjunto de datos Óptico + SAR + Índices Completos en escenario de 70 - 90 % de nubes. Fuente: Creado por la autora.

En este escenario de alta nubosidad, la clase 'Cobertura Arbórea' muestra una precisión del 73.9%. Aunque sigue siendo alta, es notablemente menor que la precisión de 82.9% observada en condiciones de menor nubosidad. Para la clase 'Pradera' que presenta una precisión del 77.7%, también es inferior al 88.6% obtenido con menor nubosidad. Lo cual indica que la mayor presencia de nubes impacta negativamente la precisión de clasificación, aunque la combinación de datos SAR y ópticos sigue proporcionando una mejora considerable.

Por otro lado, la clase 'Agua' mantiene una alta precisión de 92.9%, similar al 94.5% obtenido anteriormente. Esto sugiere que la clasificación de cuerpos de agua es menos afectada por la nubosidad, beneficiándose de la capacidad del SAR para penetrar las nubes. Sin embargo, la clase 'Construido' muestra una precisión del 47.4%, significativamente menor que el 74.2% obtenido con menor nubosidad. Las clases 'Vegetación Desnuda' y 'Matorral' muestran precisiones de 1.6% y 0.1% respectivamente en este último resultado, es decir, también disminuyeron drásticamente con respecto al escenario anterior que tuvo resultados de 'Vegetación Desnuda' con una precisión del 19.8% y 'Matorral' del 3.8%. Esta severa disminución indica que la alta presencia de nubes complica la correcta identificación de estas clases, posiblemente debido a la mayor heterogeneidad espectral y la similitud con otras clases en estas condiciones.

Finalmente, comparando ambos escenarios, se ratifica que la alta nubosidad impacta negativamente la precisión de varias clases, aunque la incorporación de datos SAR y sus índices sigue proporcionando mejoras significativas en la clasificación. Las diferencias más marcadas se observan en la precisión de las clases 'Cobertura Arbórea' y 'Construido', mientras que las clases 'Vegetación Desnuda' y 'Matorral' muestran una disminución significativa. Estos resultados validan la robustez del modelo en condiciones desafiantes, aunque subrayan la necesidad de seguir optimizando la integración de múltiples fuentes de datos para mitigar los efectos adversos de la nubosidad.

Importancia de Características con nubosidad del 0-30 %

La Figura 4.7 muestra la importancia de las características para el modelo Random Forest en dos conjuntos de datos: uno con solo datos ópticos y otro que combina datos ópticos, SAR e índices, en un escenario con un 0-30% de nubosidad.

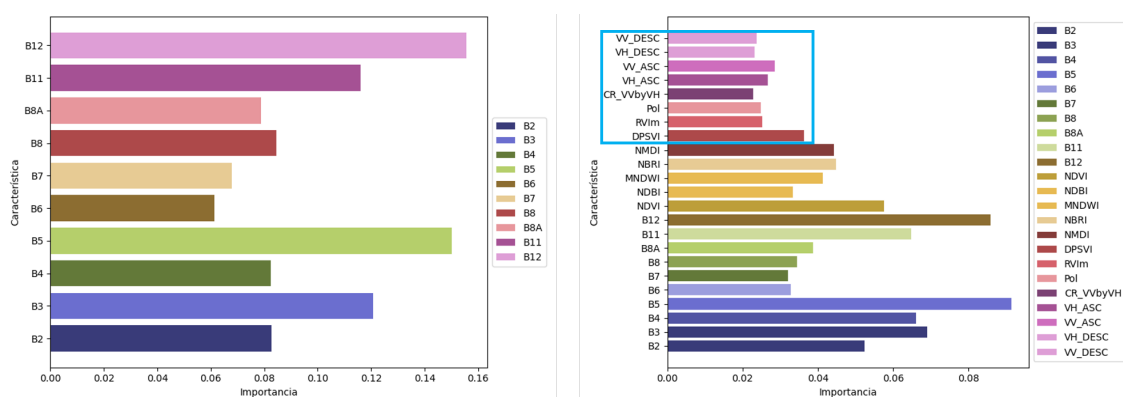


Figura 4.7: Importancia de las Características para el Modelo Random Forest y conjunto de datos Óptico vs Óptico + SAR + Índices Completos en escenario de 0 - 30 % de nubes. Fuente: Creado por la autora.

En el gráfico de la izquierda, correspondiente al conjunto de datos ópticos, las bandas espectrales B12, B5 y B3 son las más importantes. La banda B12 permite identificar humedad en la vegetación y el suelo. Es útil para monitorear el estrés hídrico en las plantas y para aplicaciones en agricultura y gestión del agua. B5 es fundamental para monitorear la salud y el vigor de las plantas. Se utiliza en estudios de vegetación, agricultura de precisión y monitoreo de bosques. Y la banda B3 ayuda a identificar vegetación saludable, cuerpos de agua y diferencias en la cobertura del suelo. Esencial para estudios de uso del suelo y análisis de la cobertura vegetal.

El gráfico de la derecha, que incluye datos ópticos, SAR e índices, muestra una mayor diversificación en la importancia de las características. Los índices espectrales junto con las características SAR son poco relevantes en este escenario. Por lo cual en términos generales para este caso, a pesar de que los datos SAR aportan cierta información de importancia, los datos ópticos siguen siendo los de mayor relevancia, incluyendo algunos de sus índices espectrales.

Importancia de Características con nubosidad del 70-90 %

La Figura 4.8 muestra la importancia de las características en este caso para el escenario con un 70-90 % de nubosidad: uno con solo datos ópticos y otro que combina datos ópticos, SAR e índices.

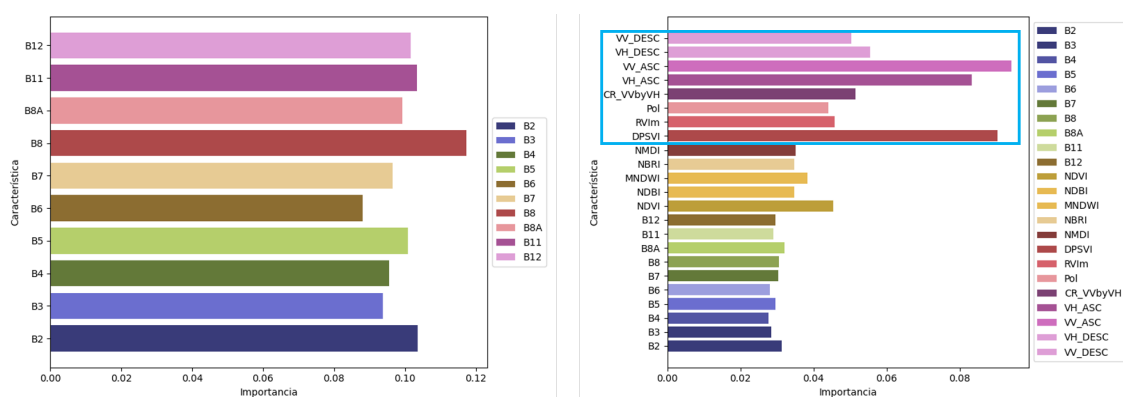
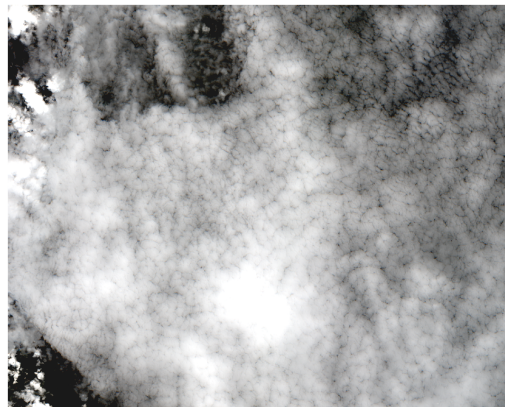


Figura 4.8: Importancia de las Características para el Modelo Random Forest y conjunto de datos Óptico vs Óptico + SAR + Índices Completos en escenario de 70 - 90% de nubes. Fuente: Creado por la autora.

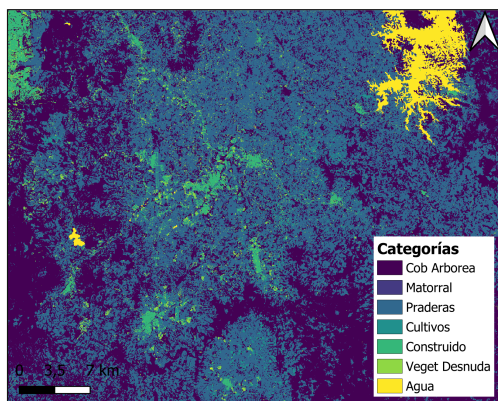
En el gráfico de la izquierda, correspondiente a los datos ópticos, las bandas espectrales B8, B2 y B11 son las más importantes. B8 es un indicador clave de la salud y vigor de la vegetación, utilizada en estudios de vegetación, monitoreo agrícola y gestión de recursos naturales. La banda B2 es ampliamente utilizada en estudios de calidad del agua, ya que es útil para observar cuerpos de agua, evaluar la turbidez y monitorear su color. Mientras que B11 es similar a B12, ya que permite el análisis de la humedad en suelos y vegetación, valiosa para estudios geológicos y de suelos desnudos.

Para el gráfico de la derecha, que incluye datos ópticos, SAR e índices, podemos observar que las características SAR, como VV_ASC, DPSVI y VH_ASC, ganan en relevancia. Por lo cual, para escenarios con alta presencia de nubes podemos ver como las características de los datos SAR pasan a ser dominantes. Esto resalta la capacidad de los datos SAR para penetrar la nubosidad y proporcionar información complementaria sobre la estructura y la textura del terreno, lo que es crucial para mantener la precisión de la clasificación en estas condiciones. A pesar de que las bandas ópticas continúan aportando un poco de información, es evidente que la inclusión de características SAR e índices proporcionan una mejora significativa en la robustez y precisión del modelo en escenarios de nubosidad extrema.

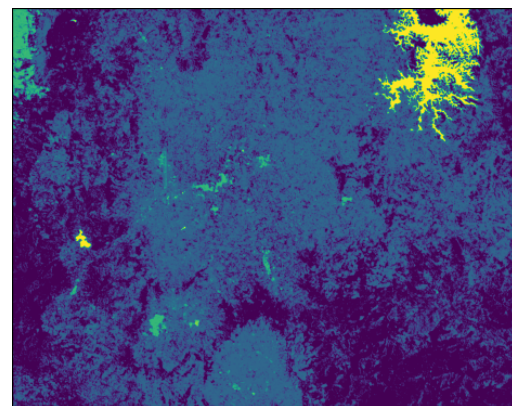
La Figura 4.9 presenta una comparación visual entre la imagen óptica con un 70-90% de nubosidad 4.9a, el mapa de referencia WorldCover 4.9b y la clasificación obtenida con el mejor modelo 4.9c.



(a) Imagen Óptica 70-90 % de nubes



(b) Mapa de referencia WorldCover



(c) Clasificación

Figura 4.9: Comparativo de la imagen real del área de estudio óptica en un rango del 70 al 90 % de nubes, con el mapa de referencia WorldCover y la imagen de la clasificación del modelo. Fuente: Creado por la autora.

La imagen 4.9a muestra una densa cobertura nubosa que oscurece gran parte del área de estudio, dificultando la interpretación del terreno con solo datos ópticos. Esto resalta la necesidad de incorporar datos SAR para mejorar la clasificación en tales condiciones. El mapa de referencia WorldCover 4.9b ilustra las diferentes categorías de uso del suelo y cobertura vegetal, sirviendo como referencia para validar la precisión del modelo.

La clasificación del modelo Random Forest 4.9c demuestra la integración efectiva de datos ópticos, SAR e índices. A pesar de la alta nubosidad, el modelo identifica correctamente la gran mayoría de categorías de uso del suelo. Por ejemplo, la clase 'Agua' se distingue claramente, mostrando la capacidad del modelo para manejar la nubosidad mediante la inclusión de datos SAR en ca-

tegorías de este tipo. Sin embargo, se observan discrepancias en clases como 'Matorral' y 'Vegetación Desnuda', indicando que a pesar del gran aporte de los datos SAR se continúa afectando la precisión en categorías que probablemente tienen gran heterogeneidad espectral y similitud con otras clases.

Finalmente, la clasificación obtenida por el modelo muestra una correspondencia visual excelente con el mapa de referencia de WorldCover. Esto es notable incluso en condiciones de nubosidad extrema, donde la complejidad para lograr una reconstrucción precisa de la imagen es evidente al observar solo la imagen óptica. Las discrepancias observadas resaltan ciertas limitaciones que podrían abordarse mediante procesos más detallados y exhaustivos, especialmente para las clases más afectadas. Sin embargo, los resultados obtenidos en estos escenarios extremos, bajo la metodología implementada, reafirman la importancia de desarrollar y utilizar la integración de datos de múltiples sensores para lograr clasificaciones muy precisas en territorios altamente impactados por condiciones climáticas.

Conclusiones y Trabajo Futuro

Esta investigación ha demostrado la eficiencia de fusionar imágenes ópticas y SAR para mejorar la clasificación de LULC mediante la integración de técnicas avanzadas de agregación de características y métodos de clasificación supervisados. Se ha evaluado la precisión de los mapas obtenidos con diferentes conjuntos de datos y en condiciones de alta nubosidad, destacando que el enfoque de fusión multisensorial es una solución efectiva para superar algunas limitaciones tradicionales de la teledetección, especialmente en condiciones atmosféricas adversas.

La metodología implementada permitió superar varios obstáculos asociados con la variabilidad de condiciones atmosféricas y la diversidad de coberturas terrestres. La elección de un enfoque de fusión de datos multisensor, combinado con la selección cuidadosa de índices espectrales y de radar, facilitó una captura más completa de las características del terreno, lo cual es esencial para mejorar la precisión de las clasificaciones.

En términos generales, el modelo Random Forest alcanzó la mayor precisión en todos los experimentos realizados en cada escenario, destacándose como el mejor enfoque. Específicamente, en la primera etapa de clasificación en un escenario de imágenes ópticas con presencia de nubes en un porcentaje de 0 al 30 % y la evaluación con el mapa de referencia CLC, este modelo obtuvo una precisión del 61.54 %, un F1-Score del 59.42 % y un coeficiente Kappa de 0.451 al utilizar el conjunto de datos compuesto por la fusión de imágenes ópticas, SAR e índices de ambos. Demostrando que la combinación de múltiples fuentes de datos proporciona una mejor discriminación de las diferentes clases, mitigando las limitaciones impuestas por la nubosidad.

Sin embargo, los resultados obtenidos en la primera etapa demostraron el impacto de las deficiencias en el mapa de referencia del CLC debido a la resolución y a la precisión de las etiquetas en su clasificación. Esto resalta la importancia de utilizar mapas de referencia y conjuntos de datos robustos y bien balanceados, donde además la distribución equitativa en los conjuntos de entrenamiento

y prueba validen la capacidad del modelo para generalizar a nuevas observaciones, reflejando su potencial para aplicaciones en escenarios del mundo real.

En la segunda etapa se utilizó exclusivamente el modelo Random Forest para la clasificación en múltiples escenarios con imágenes ópticas en diferentes rangos de porcentaje de nubes y la evaluación adicional con el mapa de referencia WorldCover que proporciona información más precisa, basada en los mismos sensores utilizados para esta investigación y en un periodo de tiempo más reciente en comparación con el CLC. Los conjuntos de datos fueron evaluados bajo cuatro escenarios de rangos de nubosidad: 0-30 %, 30-50 %, 50-70 % y 70-90 %. En esta etapa, el mejor resultado obtenido fue una precisión del 81.92 %, un F1-Score del 80.57 % y un coeficiente Kappa de 0.687, también con el conjunto de datos de imágenes ópticas, SAR e índices de ambos en el escenario con 0-30 % de nubes.

A lo largo de los diferentes escenarios de nubosidad, se confirmó la superioridad de la fusión de datos ópticos, SAR e índices, especialmente en la clasificación de coberturas acuáticas, aunque los desafíos persistieron en clases vegetativas con mayor heterogeneidad espectral y bajo condiciones de nubosidad extrema. Esto sugiere que, si bien el enfoque multisensorial mejora el rendimiento, aún existen áreas donde se puede optimizar la metodología. El análisis realizado sobre la importancia de las características dentro de la clasificación de LULC en diferentes escenarios refleja que la diversificación en la importancia de las características denota la capacidad de estos modelos para adaptarse a diferentes niveles de nubosidad, aumentando la relevancia de la información aportada por los datos SAR en circunstancias extremas y optimizando el desempeño final de la clasificación.

Si bien esta investigación ha alcanzado avances importantes, abre numerosas oportunidades para desarrollos futuros en la clasificación de LULC, que podrían ampliarse con la incorporación de nuevos tipos de datos y técnicas. Futuras investigaciones podrían explorar la inclusión de datos hiperespectrales y de satélites con mayor resolución temporal y espacial, como Sentinel-3 o Landsat-9, lo que permitiría mejorar la capacidad de distinguir clases de vegetación y coberturas terrestres con mayor precisión, especialmente en condiciones de alta nubosidad.

Asimismo, otra área de investigación futura se podría centrar en la optimización de los métodos de fusión de datos. En lugar de utilizar enfoques de fusión estáticos, otros estudios podrían beneficiarse de la implementación de algoritmos que ajusten dinámicamente los hiperparámetros, optimizando la adaptabilidad del modelo a distintos niveles de nubosidad y diversidad espectral. Esto permitiría que los modelos sean más flexibles y efectivos en una variedad de entornos

atmosféricos y de cobertura terrestre.

Otra posibilidad para la continuidad de esta línea de investigación es aplicar la metodología en regiones geográficas con alta vulnerabilidad a fenómenos climáticos, como zonas costeras o áreas forestales amenazadas por la deforestación. Evaluar el enfoque en escenarios reales ofrecería una valiosa validación de la robustez de la metodología en contextos críticos para la toma de decisiones sobre planificación territorial y gestión de recursos naturales.

Adicionalmente, realizar las validaciones con mapas de referencia más recientes y de mayor resolución, como se hizo con el mapa WorldCover, que ya ha demostrado una mejora significativa en comparación con el CLC. Sin embargo, la creación de nuevos mapas de referencia con etiquetas más precisas y con cobertura más extensa podría ofrecer una base más sólida para futuros entrenamientos y evaluaciones de los modelos, lo que mejoraría aún más su desempeño en la clasificación de LULC en entornos del mundo real.

Finalmente, ya que se sugiere una evaluación crítica de la idoneidad de los productos LULC globales para diversas aplicaciones, podemos destacar dos estudios recientes que relacionan estas sugerencias. Por ejemplo, el estudio de Zander S. Venter presenta una comparación cruzada y una evaluación de la precisión de tres diferentes productos con una resolución de 10 m, con el fin de informar la adopción y aplicación de estos mapas en el futuro, subrayando la precisión y sesgos de cada uno de los mapas en diferentes clases de cobertura, como agua, superficie construida y vegetación. Este concluye con una recomendación general de evaluar críticamente cada producto LULC con referencia al propósito de la aplicación deseada [140]. De igual forma, el segundo estudio de Cidália C. Fonte que basados en la gran disponibilidad de imágenes de alta resolución realizan una comparación de seis productos LULC, utilizando imágenes de Sentinel-1 y 2 con una resolución espacial de 10 m, centrándose en la precisión temática y las similitudes cartográficas. Donde concluyeron que la selección de un mapa para usar en una aplicación particular no es tarea fácil y se hace necesario evaluar los productos disponibles para determinar cuáles proporcionan la información necesaria [141].

Bibliografía

- [1] Jordy Alexander Orellana Salas and Tatiana Del Cisne Lalvay Portilla. Uso e importancia de los recursos naturales y su incidencia en el desarrollo turístico. Caso Cantón Chilla, El Oro, Ecuador. *Revista interamericana de ambiente y turismo*, 14:65 – 79, 06 2018. URL: http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-235X2018000100065&nrm=iso.
- [2] Jeannette Sánchez and Mauricio León. Recursos naturales y desarrollo sostenible: Propuestas teóricas en el contexto de América Latina y el Caribe. 2024. URL: <https://www.issuu.com/publicacionescepal/stacks>.
- [3] FAO. Los bosques de América Latina y el Caribe son claves para la sostenibilidad ambiental y la seguridad alimentaria del planeta, 2021. URL: <https://www.fao.org/americas/noticias/ver/es/c/1439482/>.
- [4] María Constanza and Dolores Armenteras. Uso del suelo y estructura de la vegetación en paisajes fragmentados en la Amazonia, Colombia. *Colombia Forestal*, 21(2):205–223, 2018. URL: <https://doi.org/10.14483/2256201X.12330>.
- [5] FAO. Informe de la Evaluación de los recursos forestales mundiales 2020, 2021. URL: <https://www.fao.org/forest-resources-assessment/2020/es>.
- [6] CEPAL. La pérdida de los bosques de América Latina y el Caribe 1990–2020: evidencia estadística. *Temas estadísticos de la CEPAL*, (2):1–9, 2021. URL: <https://hdl.handle.net/11362/47151>.
- [7] Visión Amazonía. En 2021, el Inventario Nacional Forestal retomará labores en campo, 2021. URL: <https://visionamazonia.minambiente.gov.co/news/en-2021-el-inventario-nacional-forestal-retomara-labores-en-campo/>.

- [8] Danny Sandoval, Natalia Triana-Ángel, and Stefan Burkart. Relaciones directas y subyacentes de la deforestación: Análisis de caso para el departamento del caquetá, Colombia, 2024. URL: <https://hdl.handle.net/10568/141514>.
- [9] Visión Amazonía. Informe anual 2021-2022, 2022. URL: <https://visionamazonia.minambiente.gov.co/content/uploads/2023/05/Informe-anual-2021-2022-.pdf>.
- [10] Santiago Luque. Crece la deforestación en Colombia: más de 171 mil hectáreas se perdieron en el 2020, 2021. URL: <https://es.mongabay.com/2021/07/crece-deforestacion-colombia-2020/>.
- [11] Mancera Florez Juan Ricardo. Evaluación de imágenes de radar Sentinel-1a e imágenes multiespectrales Sentinel-2a en la clasificación de cobertura del suelo en diferentes niveles de detalle, 2019. URL: <https://repositorio.unal.edu.co/handle/unal/76982>.
- [12] Edison Fabian and Vargas Becerra. Revisión bibliométrica de las principales metodologías para la evaluación del uso y cobertura del suelo, 2024.
- [13] Daniel G. Brown, Peter H. Verburg, Robert Gilmore Pontius, and Mark D. Lange. Opportunities to improve impact, integration, and evaluation of land change models. *Current Opinion in Environmental Sustainability*, 5(5):452–457, 2013. URL: <http://dx.doi.org/10.1016/j.cosust.2013.07.012>, doi:10.1016/j.cosust.2013.07.012.
- [14] R. Yonaba, M. Koïta, L. A. Mounirou, F. Tazen, P. Queloz, A. C. Biaou, D. Niang, C. Zouré, H. Karambiri, and H. Yacouba. Spatial and transient modelling of land use/land cover (LULC) dynamics in a Sahelian landscape under semi-arid climate in northern Burkina Faso. *Land Use Policy*, 103(July 2019), 2021. doi:10.1016/j.landusepol.2021.105305.
- [15] Cassia Brocca Caballero, Anderson Ruhoff, and Trent Biggs. Land use and land cover changes and their impacts on surface-atmosphere interactions in Brazil: A systematic review. *Science of the Total Environment*, 808:152134, 2022. doi:10.1016/j.scitotenv.2021.152134.
- [16] Bo Sun and Qiming Zhou. Expressing the spatio-temporal pattern of farmland change in arid lands using landscape metrics. *Journal of Arid Environments*, 124:118–127, 2016. URL: <http://dx.doi.org/10.1016/j.jaridenv.2015.08.007>, doi:10.1016/j.jaridenv.2015.08.007.

- [17] Abdulqadeer Rash, Yaseen Mustafa, and Rahel Hamad. Quantitative assessment of land use/land cover changes in a developing region using machine learning algorithms: A case study in the kurdistan region, iraq. *Helvion*, 9, 11 2023. doi:10.1016/j.helivion.2023.e21253.
- [18] Alexander Cass, George P. Petropoulos, Konstantinos P. Ferentinos, Andrew Pavlides, and Prashant K. Srivastava. Exploring the synergy between Landsat and ASAR towards improving thematic mapping accuracy of optical EO data. *Applied Geomatics (AGMJ)*, 2019. URL: <https://doi.org/10.1007/s12518-019-00258-7>.
- [19] S Abdikan, F B Sanli, M Ustuner, Synthetic Aperture Radar, and Support Vector Machines. Land cover mapping using sentinel-1 sar data. *XLI(July):757–761*, 2016. doi:10.5194/isprsarchives-XLI-B7-757-2016.
- [20] Heiko Balzter, Beth Cole, Christian Thiel, and Christiane Schmillius. Mapping CORINE Land Cover from Sentinel-1A SAR and SRTM Digital Elevation Model Data using Random Forests. *Remote Sensing*, (December 2014):14876–14898, 2015. doi:10.3390/rs71114876.
- [21] Xin Niu and Yifang Ban. Multitemporal RADARSAT-2 Polarimetric SAR Data for Urban Land Cover Classification Using Support Vector Machine. *Remote Sensing for Science*, 2010. URL: <https://doi.org/10.1080/01431161.2012.700133>.
- [22] J. Soria Ruiz, Y. Fernandez Ordoñez, and I. H. Woodhouse. Land-cover classification using radar and optical images: A case study in Central Mexico. *International Journal of Remote Sensing*, 31(12):3291–3305, 2010. doi:10.1080/01431160903160777.
- [23] Juan Ricardo Mancera-flórez and Ivan Lizarazo. Land cover classification at three different levels of detail from optical and radar Sentinel SAR data : a case study in Cundinamarca. *DYNA*, 87(215):136–145, 2020. URL: <https://doi.org/10.15446/dyna.v87n215.84915>.
- [24] Juan C González Vélez, Juan D Martínez Vargas, and Maria C Torres Madronero. Land cover classification using CNN and semantic segmentation: A case of study in Antioquia, Colombia. In *Communications in Computer and Information Science*, page 14. 2022. URL: <https://www.itm.edu.co>.
- [25] Sohaib K.M. Abujayyab and Ismail Rakip Karas. Employing neural networks algorithm for LULC mapping. *Baltic Journal of Modern Computing*, 8(2):370–378, 2020. doi:10.22364/BJMC.2020.8.2.12.

- [26] Ava Vali, Sara Comai, and Matteo Matteucci. Deep Learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sensing*, 12(15), 2020. doi:10.3390/RS12152495.
- [27] Yashwant Singh. Significance Of Land Use / Land Cover (LULC) Maps, 2017. URL: <https://www.satpalda.com/blogs/significance-of-land-use-land-cover-lulc-maps>.
- [28] Birhanu Tadesa Edosa and Milkessa Dangia Nagasa. Spatiotemporal assessment of land use land cover change, driving forces, and consequences using geospatial techniques: The case of naqamte city and hinterland, western ethiopia. *Environmental Challenges*, 14, 1 2024. doi:10.1016/j.envc.2023.100830.
- [29] María Perevochtchikova. La evaluación del impacto ambiental y la importancia de los indicadores ambientales, 2012. URL: https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-10792013000200001.
- [30] Konstantinos Karalas, Grigorios Tsagkatakis, Michalis Zervakis, and Panagiotis Tsakalides. Land Classification Using Remotely Sensed Data: Going Multi-Label. In *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–16. 2016. doi:10.1109/TGRS.2016.2520203.
- [31] Tomohiro Shiraishi, Takeshi Motohka, Rajesh Bahadur Thapa, and Manabu Watanabe. Comparative Assessment of Supervised Classifiers for Land Use – Land Cover Classification in a Tropical Region Using Time-Series PALSAR Mosaic Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4):1186–1199, 2014. doi:10.1109/JSTARS.2014.2313572.
- [32] Yanzhao Wang, Yonghua Sun, Xuyue Cao, Yihan Wang, Wangkuan Zhang, and Xinglu Cheng. A review of regional and global scale land use/land cover (lulc) mapping products generated from satellite remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 206:311–334, 2023. URL: <https://www.sciencedirect.com/science/article/pii/S0924271623003209>, doi:10.1016/j.isprsjprs.2023.11.014.
- [33] Allisson Stephanie Ube Moran. Análisis multitemporal del crecimiento urbano, superficie agrícola en el cantón durá, provincia del guayas, 2024. URL: <http://dspace.ups.edu.ec/handle/123456789/27670>.

- [34] Rajat Garg, Anil Kumar, Manish Prateek, Kamal Pandey, and Shashi Kumar. Land Cover classification of spaceborne multifrequency SAR and optical multispectral data using machine learning. *Advances in Space Research*, 69(4):1726–1742, 2022. doi:10.1016/j.asr.2021.06.028.
- [35] F. D. Van der Meer, H. M.A. van der Werff, and F. J.A. van Ruitenbeek. Potential of ESA’s Sentinel-2 for geological applications. *Remote Sensing of Environment*, 148:124–133, 2014. URL: <http://dx.doi.org/10.1016/j.rse.2014.03.022>, doi:10.1016/j.rse.2014.03.022.
- [36] Ziqi Li Qiqi Zhu, Xi Guo and Deren Li. A review of multi-class change detection for satellite remote sensing imagery. *Geo-spatial Information Science*, 27(1):1–15, 2024. doi:10.1080/10095020.2022.2128902.
- [37] Juan Luis Jacobo Cuéllar, Heriberto Guevara Rojo, and Óscar Cruz Álvarez. Teledetección y automatización en la agricultura del futuro. *Elementos* 133, 2024. URL: <https://doi.org/10.14350/rig.59929>.
- [38] Pablo Ariel Paccioletti, Franca Giannini Kurina, Franco Marcelo Suarez, Carlos Matias Scavuzzo, Vanina Alemandri, Brenda Emiliana Gómez Montenegro, and Monica Graciela Balzarini. Protocolo para automatizar la descarga de datos climáticos desde la nube y generar indicadores biometeorológicos para el monitoreo epidemiológico de cultivos. 2023. URL: <https://doi.org/10.31047/1668.298x.v40.n1.39619>.
- [39] ESA. Missions - Sentinel Online - Sentinel Online, 2022. URL: <https://sentinels.copernicus.eu/web/sentinel/missions>.
- [40] Copernicus. Open Access Hub. <https://scihub.copernicus.eu/>, 2022.
- [41] Andrés Velastegui-Montoya, Néstor Montalván-Burbano, Paúl Carrión-Mero, Hugo Rivera-Torres, Luís Sadeck, and Marcos Adami. Google earth engine: A global analysis and future trends. *Remote Sensing*, 15(14), 2023. URL: <https://www.mdpi.com/2072-4292/15/14/3675>, doi:10.3390/rs15143675.
- [42] Lalit Kumar and Onesimo Mutanga. Google earth engine applications since inception: Usage, trends, and potential. *Remote Sensing*, 10(10), 2018. URL: <https://www.mdpi.com/2072-4292/10/10/1509>, doi:10.3390/rs10101509.
- [43] Haifa Tamiminia, Bahram Salehi, Masoud Mahdianpari, Lindi Quackenbush, Sarina Adeli, and Brian Brisco. Google earth engine for geo-big data applications: A meta-analysis and systematic review.

- ISPRS Journal of Photogrammetry and Remote Sensing*, 164:152–170, 2020. URL: <https://www.sciencedirect.com/science/article/pii/S0924271620300927>, doi:10.1016/j.isprsjprs.2020.04.001.
- [44] Pedro Pérez-Cutillas, Alberto Pérez-Navarro, Carmelo Conesa-García, Demetrio Antonio Zema, and Jesús Pilar Amado-Álvarez. What is going on within google earth engine? a systematic review and meta-analysis. *Remote Sensing Applications: Society and Environment*, 29:100907, 2023. URL: <https://www.sciencedirect.com/science/article/pii/S2352938522002154>, doi:10.1016/j.rsase.2022.100907.
- [45] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202:18–27, 2017. Big Remotely Sensed Data: tools, applications and experiences. URL: <https://www.sciencedirect.com/science/article/pii/S0034425717302900>, doi:10.1016/j.rse.2017.06.031.
- [46] Leandro Parente, Evandro Taquary, Ana Paula Silva, Carlos Souza, and Laerte Ferreira. Next generation mapping: Combining deep learning, cloud computing, and big remote sensing data. *Remote Sensing*, 11(23), 2019. URL: <https://www.mdpi.com/2072-4292/11/23/2881>, doi:10.3390/rs11232881.
- [47] Meisam Amani, Arsalan Ghorbanian, Seyed Ali Ahmadi, Mohammad Kakooei, Armin Moghimi, S. Mohammad Mirmazloumi, Sayyed Hamed Alizadeh Moghaddam, Sahel Mahdavi, Masoud Ghahremanloo, Saeid Parsian, Qiusheng Wu, and Brian Brisco. Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:5326–5350, 2020. doi:10.1109/JSTARS.2020.3021052.
- [48] Jianxin Jia, Haibin Sun, Changhui Jiang, Kirsi Karila, Mika Karjalainen, Eero Ahokas, Ehsan Khoramshahi, Peilun Hu, Chen Chen, Tianru Xue, Tinghuai Wang, Yuwei Chen, and Juha Hyypä. Review on active and passive remote sensing techniques for road extraction. *Remote Sensing*, 13(21), 2021. URL: <https://www.mdpi.com/2072-4292/13/21/4235>, doi:10.3390/rs13214235.
- [49] Zhujun Gu and Maimai Zeng. The use of artificial intelligence and satellite remote sensing in land cover change detection: Review and perspectives. *Sustainability*, 16(1), 2024. URL: <https://www.mdpi.com/2071-1050/16/1/274>, doi:10.3390/su16010274.

- [50] Aumed Rahman M Amen, Andam Mustafa, Dalshad Ahmed Kareem, Hasan Mohammed Hameed, Ayub Anwar Mirza, Michał Szydłowski, and Bala Kawa M. Saleem. Mapping of flood-prone areas utilizing gis techniques and remote sensing: A case study of duhok, kurdistan region of iraq. *Remote Sensing*, 15(4), 2023. URL: <https://www.mdpi.com/2072-4292/15/4/1102>, doi:10.3390/rs15041102.
- [51] S.M. Sohel Rana, SM Ahsan Habib, M Nur Hossain Sharifee, Nasrin Sultana, and Syed Hafizur Rahman. Flood risk mapping of the flood-prone rangpur division of bangladesh using remote sensing and multi-criteria analysis. *Natural Hazards Research*, 4(1):20–31, 2024. URL: <https://www.sciencedirect.com/science/article/pii/S266659212300094X>, doi:10.1016/j.nhres.2023.09.012.
- [52] Saeed Mhanna, Landon J.S. Halloran, François Zwahlen, Ahmed Haj Asaad, and Philip Brunner. Using machine learning and remote sensing to track land use/land cover changes due to armed conflict. *Science of The Total Environment*, 898:165600, 2023. URL: <https://www.sciencedirect.com/science/article/pii/S0048969723042237>, doi:10.1016/j.scitotenv.2023.165600.
- [53] Emilio Chuvieco. *Fundamentals of Satellite Remote Sensing*. CRC Press, 3rd edition, 2020. doi:10.1201/9780429506482.
- [54] Petro Kogut. Teledetección satelital: Tipos, usos y aplicaciones, 2021. URL: <https://eos.com/es/blog/teledeteccion/>.
- [55] Liping Di and Eugene Yu. *Remote Sensing*, pages 17–43. Springer International Publishing, Cham, 2023. doi:10.1007/978-3-031-33932-5_2.
- [56] Manuel Carranza García, Jorge García Gutiérrez, and José C. Riquelme. A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*, 11(3), 2019. doi:10.3390/rs11030274.
- [57] Federico Raspini, Silvia Bianchini, Andrea Ciampalini, Matteo Del Soldato, Lorenzo Solari, Fabrizio Novali, Sara Del Conte, Alessio Rucci, Alessandro Ferretti, and Nicola Casagli. Continuous, semi - automatic monitoring of ground deformation using Sentinel-1 satellites. *Scientific Reports*, 8(1), dec 2018. doi:10.1038/S41598-018-25369-W.
- [58] Maruf Billah, A.K.M. Saiful Islam, Wasif Bin Mamoon, and Mohammad Rezaur Rahman. Random forest classifications for landuse map-

- ping to assess rapid flood damage using sentinel-1 and sentinel-2 data. *Remote Sensing Applications: Society and Environment*, 30:100947, 2023. URL: <https://www.sciencedirect.com/science/article/pii/S2352938523000290>, doi:10.1016/j.rsase.2023.100947.
- [59] Freda Patricia Baquedano Alonzo and Perla Carolina Brenes Gutiérrez. Diseño de un sistema de recepción de imágenes satelitales, climáticas y ambientales, 2012. URL: <http://repositorio.uca.edu.ni/id/eprint/563>.
- [60] Dorijan Radočaj, Jasmina Obhodaš, Mladen Jurišić, and Mateo Gašparovi. Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review. *Land*, pages 1–23, 2020. doi:10.3390/land9110402.
- [61] Kathiravan Thangavel, Roberto Sabatini, Alessandro Gardi, Kavindu Ranasinghe, Samuel Hilton, Pablo Servidia, and Dario Spiller. Artificial intelligence for trusted autonomous satellite operations. *Progress in Aerospace Sciences*, 144:100960, 2024. URL: <https://www.sciencedirect.com/science/article/pii/S0376042123000763>, doi:10.1016/j.paerosci.2023.100960.
- [62] Juan Peña Llopis, Rosa María Poveda Martínez, Andreu Bonet Jornet, Juan Francisco Bellot Abad, and Antonio Escarré Esteve. Cartografía de las coberturas y usos del suelo de la marina baixa (alicante) para 1956, 1978 y 2000. *Investigaciones Geográficas*, (37):93–107, sep 2005. URL: <https://www.investigacionesgeograficas.com/article/view/2005-n37-cartografia-de-las-coberturas-y-usos-del-suelo-de-la-ma>, doi:10.14198/INGEO2005.37.08.
- [63] Andrea E Pizarro. *Análisis de coberturas y usos del suelo mediante técnicas de teledetección e integración en un Sistema de Información Geográfico en Barreal, Calingasta, San Juan*. PhD thesis, Universidad Nacional de Luján, 2023. URL: <https://ri.unlu.edu.ar/xmlui/handle/rediunlu/2364>.
- [64] ESA. Sentinel-2 - Missions - Sentinel Online - Sentinel Online, 2022. URL: <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2>.
- [65] S.S. Dagne, H.H. Hirpha, A.T. Tekoye, et al. Fusion of sentinel-1 sar and sentinel-2 msi data for accurate urban land use-land cover classification in gondar city, ethiopia. *Environmental Systems Research*, 12(40), 2023. doi:10.1186/s40068-023-00324-5.

- [66] Carmen Valdivieso-Ros, Francisco Alonso-Sarria, and Francisco Gomariz-Castillo. Effect of the synergetic use of sentinel-1, sentinel-2, lidar and derived data in land cover classification of a semiarid mediterranean area using machine learning algorithms. *Remote Sensing*, 15(2), 2023. URL: <https://www.mdpi.com/2072-4292/15/2/312>, doi:10.3390/rs15020312.
- [67] Shivam Pande. Land use/land cover classification of fused sentinel-1 and sentinel-2 imageries using ensembles of random forests, 2018. doi:10.48550/arXiv.2312.10798.
- [68] Tomasz Tarasiewicz, Jakub Nalepa, Reuben A. Farrugia, Gianluca Valentino, Mang Chen, Johann A. Briffa, and Michal Kawulok. Multitemporal and multispectral data fusion for super-resolution of sentinel-2 images. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–19, 2023. URL: <http://dx.doi.org/10.1109/TGRS.2023.3311622>, doi:10.1109/tgrs.2023.3311622.
- [69] Jingan Wu, Liupeng Lin, Chi Zhang, Tongwen Li, Xiao Cheng, and Fang Nan. Generating sentinel-2 all-band 10-m data by sharpening 20/60-m bands: A hierarchical fusion network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196:16–31, 2023. URL: <https://www.sciencedirect.com/science/article/pii/S0924271622003331>, doi:10.1016/j.isprsjprs.2022.12.017.
- [70] Sara Mastrorosa, Mattia Crespi, Luca Congedo, and Michele Munafò. Land consumption classification using sentinel 1 data: A systematic review. *Land*, 12(4), 2023. URL: <https://www.mdpi.com/2073-445X/12/4/932>, doi:10.3390/land12040932.
- [71] ESA. Sentinel-1 - Missions - Sentinel Online - Sentinel Online, 2022. URL: <https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1>.
- [72] María Mata Muñoz. *Monitorización de inundaciones mediante Radar SAR (SENTINEL-1)*. PhD thesis, 2019. URL: <https://ebuah.uah.es/dspace/handle/10017/41791>.
- [73] Wenmei Li, Jiaqi Wu, Qing Liu, Yu Zhang, Bin Cui, Yan Jia, and Guan Gui. An effective multimodel fusion method for sar and optical remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:5881–5892, 2023. doi:10.1109/JSTARS.2023.3288143.

- [74] Jarrett Adrian, Vasit Sagan, and Maitiniyazi Maimaitijiang. Sentinel sar-optical fusion for crop type mapping using deep learning and google earth engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175:215–235, 2021. URL: <https://www.sciencedirect.com/science/article/pii/S0924271621000599>, doi:10.1016/j.isprsjprs.2021.02.018.
- [75] Jiaxin Li, Danfeng Hong, Lianru Gao, Jing Yao, Ke Zheng, Bing Zhang, and Jocelyn Chanussot. Deep learning in multimodal remote sensing data fusion: A comprehensive review. *International Journal of Applied Earth Observation and Geoinformation*, 112:102926, 2022. URL: <https://www.sciencedirect.com/science/article/pii/S1569843222001248>, doi:10.1016/j.jag.2022.102926.
- [76] Armugha Khan, Himanshu Govil, Gaurav Kumar, and Rucha Dave. Synergistic use of sentinel-1 and sentinel-2 for improved lulc mapping with special reference to bad land class: a case study for yamuna river floodplain, india. *Spatial Information Research*, 28:669–681, 12 2020. doi:10.1007/s41324-020-00325-x.
- [77] Zongjia Zhang, Yiping Zeng, Zhejun Huang, Junguo Liu, and Lili Yang. Multi-source data fusion and hydrodynamics for urban waterlogging risk identification. *International Journal of Environmental Research and Public Health*, 20(3), 2023. URL: <https://www.mdpi.com/1660-4601/20/3/2528>, doi:10.3390/ijerph20032528.
- [78] Filiz Sunar Saygin Abdikan, Fusun Balik Sanli and Manfred Ehlers. A comparative data-fusion analysis of multi-sensor satellite images. *International Journal of Digital Earth*, 7(8):671–687, 2014. doi:10.1080/17538947.2012.748846.
- [79] Achala Shakya, Mantosh Biswas, and Mahesh Pal. Fusion and Classification of SAR and Optical Data Using Multi-Image Color Components with Differential Gradients. *Remote Sensing*, 15(1), 2023. doi:10.3390/rs15010274.
- [80] Richard Dein D. Altarez, Armando Apan, and Tek Maraseni. Deep learning u-net classification of sentinel-1 and 2 fusions effectively demarcates tropical montane forest's deforestation. *Remote Sensing Applications: Society and Environment*, 29, 1 2023. doi:10.1016/j.rsase.2022.100887.
- [81] Jingliang Hu, Rong Liu, Danfeng Hong, Andrés Camero, Jing Yao, Mathias Schneider, Franz Kurz, Karl Segl, and Xiao Xiang Zhu. MDAS : A New Multimodal Benchmark Dataset for Remote Sensing. *Earth System Science Data*, (August):1–26, 2022. doi:10.5194/essd-15-113-2023.

- [82] Muhammad Nasar Ahmad, Zhenfeng Shao, Akib Javed, Israr Ahmad, Fakhru Islam, Hariklia D. Skilodimou, and George D. Bathrellos. Optical–sar data fusion based on simple layer stacking and the xgboost algorithm to extract urban impervious surfaces in global alpha cities. *Remote Sensing*, 16(5), 2024. URL: <https://www.mdpi.com/2072-4292/16/5/873>, doi:10.3390/rs16050873.
- [83] Jinjin Li, Jiacheng Zhang, Chao Yang, Huiyu Liu, Yangang Zhao, and Yuanxin Ye. Comparative analysis of pixel-level fusion algorithms and a new high-resolution dataset for sar and optical image fusion. *Remote Sensing*, 15(23), 2023. URL: <https://www.mdpi.com/2072-4292/15/23/5514>, doi:10.3390/rs15235514.
- [84] Samadhan C. Kulkarni and Priti P. Rege. Pixel level fusion techniques for SAR and optical images: A review. *Information Fusion*, 59(July 2019):13–29, 2020. doi:10.1016/j.inffus.2020.01.003.
- [85] Maria Constanza Torres Madronero. *Development of the hyperspectral coastal image analysis toolbox (HYCIAT) with a focus on hyperspectral and lidar data fusion*. PhD thesis, 2008. URL: <https://scholar.uprm.edu/handle/20.500.11801/2359>.
- [86] M. Mazhar U. Rathore, Awais Ahmad, Anand Paul, and Jiaji Wu. Real-time continuous feature extraction in large size satellite images. *Journal of Systems Architecture*, 64:122–132, 2016. doi:10.1016/j.sysarc.2015.11.006.
- [87] Alaa Adnan Musleh and Hussein Sabah Jaber. Comparative Analysis of Feature Extraction and Pixel-based Classification of High-Resolution Satellite Images Using Geospatial Techniques. In *E3S Web of Conferences*, volume 318, page 04007. 2021. doi:10.1051/e3sconf/202131804007.
- [88] D. R. Sowmya, Deepa Shenoy P., and Venugopal. Remote Sensing Satellite Image Processing Techniques for Image Classification: A Comprehensive Survey. *International Journal of Computer Applications*, 161(11):24–37, 2017. doi:10.5120/ijca2017913306.
- [89] Paria Ettehadi Osgouei, Elif Sertel, and M. Erdem Kabadayi. Integrated usage of historical geospatial data and modern satellite images reveal long-term land use/cover changes in Bursa/Turkey, 1858–2020. *Scientific Reports*, 12(1):1–17, 2022. doi:10.1038/s41598-022-11396-1.
- [90] Darius Phiri, Justin Morgenroth, Cong Xu, and Txomin Hermosilla. Effects of pre-processing methods on landsat oli-8 land cover classification using

- obia and random forests classifier. *International Journal of Applied Earth Observation and Geoinformation*, 73:170–178, 2018. URL: <https://www.sciencedirect.com/science/article/pii/S0303243418303453>, doi:10.1016/j.jag.2018.06.014.
- [91] Hyun-Ju Ban, Young-Joo Kwon, Hayan Shin, Han-Sol Ryu, and Sungwook Hong. Flood monitoring using satellite-based rgb composite imagery and refractive index retrieval in visible and near-infrared bands. *Remote Sensing*, 9(4), 2017. URL: <https://www.mdpi.com/2072-4292/9/4/313>, doi:10.3390/rs9040313.
- [92] Bang Nguyen Tran, Mihai A. Tanase, Lauren T. Bennett, and Cristina Apon-te. Evaluation of spectral indices for assessing fire severity in australian temperate forests. *Remote Sensing*, 10(11), 2018. URL: <https://www.mdpi.com/2072-4292/10/11/1680>, doi:10.3390/rs10111680.
- [93] Joshua Adams, Damon Woodard, Gerry Dozier, Philip Miller, George Glenn, and Kelvin Bryant. Gefe: genetic & evolutionary feature extraction for periocular-based biometric recognition. In *Proceedings of the 48th Annual Southeast Regional Conference, ACM SE '10*, New York, NY, USA, 2010. Association for Computing Machinery. doi:10.1145/1900008.1900069.
- [94] Nikita Aggarwal, Mohit Srivastava, and Maitreyee Dutta. Comparative Analysis of Pixel-Based and Object-Based Classification of High Resolution Remote Sensing Images – A Review. *International Journal of Engineering Trends and Technology*, 38(1):5–11, 2016. doi:10.14445/22315381/ijett-v38p202.
- [95] Neha Gupta and H S Bhadauria. Object based Information Extraction from High Resolution Satellite Imagery using eCognition. *IJCSI International Journal of Computer Science Issues*, 11(3):139–144, 2014. URL: <file:///C:/Users/Jefferson/Downloads/71755beb95626b29efb02e6a102d49f4.pdf>.
- [96] Marília Isabelle Oliveira da Silva Emanuel Araújo Silva José Jorge Monteiro Junior Vanessa Sousa da Silva, Gabriela Salami and Elisiane Alba. Methodological evaluation of vegetation indexes in land use and land cover (lulc) classification. *Geology, Ecology, and Landscapes*, 4(2):159–169, 2020. doi:10.1080/24749508.2019.1608409.
- [97] Suzanna Cuypers, Andrea Nascetti, and Maarten Vergauwen. Land use and land cover mapping with vhr and multi-temporal sentinel-2 imagery.

- Remote Sensing*, 15(10), 2023. URL: <https://www.mdpi.com/2072-4292/15/10/2501>, doi:10.3390/rs15102501.
- [98] Jagannath Aryal, Chiranjibi Sitaula, and Alejandro César Frery. Land use and land cover (lulc) performance modeling using machine learning algorithms: a case study of the city of melbourne, australia. *Scientific Reports*, 13, 2023. URL: <https://api.semanticscholar.org/CorpusID:261029588>.
- [99] Shengyu Zhao, Kaiwen Tu, Shutong Ye, Hao Tang, Yaocong Hu, and Chao Xie. Land use and land cover classification meets deep learning: A review. *Sensors*, 23(21), 2023. URL: <https://www.mdpi.com/1424-8220/23/21/8966>, doi:10.3390/s23218966.
- [100] J Ronald Eastman. Idrisi selva. *Worcester, MA: Clark University*, 2012.
- [101] Thomas Krauß. Preprocessing of satellite data for urban object extraction. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3W2):115–120, 2015. doi:10.5194/isprsarchives-XL-3-W2-115-2015.
- [102] Germain Forestier and Cédric Wemmert. Semi-supervised learning using multiple clusterings with limited labeled data. *Information Sciences*, 361-362:48–65, 2016. doi:10.1016/j.ins.2016.04.040.
- [103] Alessandro Lapini, Simone Pettinato, Emanuele Santi, Simonetta Paloscia, Giacomo Fontanelli, and Andrea Garzelli. Comparison of machine learning methods applied to SAR images for forest classification in mediterranean areas. *Remote Sensing*, 12(3), 2020. doi:10.3390/rs12030369.
- [104] Jian Guo, Ji Xian Zhang, Yong Hong Zhang, and Yin Xuan Cao. Study of the comparison of land cover classification for multitemporal MODIS images. *2008 International Workshop on Earth Observation and Remote Sensing Applications*, 38(1):88–92, 2008. doi:{10.1109/EORSA.2008.4620305}.
- [105] Afifuddin, Siti Arfah, Dionysius Bryan Sencaki, Zilda Dona Okta Permata, Mega Novetrishka Putri, and Hartanto Sanjaya. Land cover change in Badung Regency 2016-2020 : An approach using machine learning method: Random Forest Extreme Gradient Boost (XGB). *IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology, AGERS 2021*, pages 38–45, 2021. doi:10.1109/AGERS53903.2021.9617249.

- [106] Mónica Bocco, Gustavo Ovando, Silvina Sayago, and Enrique Willington. Neural network models for land cover classification from satellite images. *Agricultura Tecnica*, 67(4):414–421, 2007. doi:10.4067/s0365-28072007000400009.
- [107] L. Zhou and X. Yang. Use Of neural networks for land cover classification from remotely sensed imagery. *Remote Sensing and Spatial Information Sciences*, 37, 2008. doi:10.1109/IGARSS.2018.8518619.
- [108] Jose Luis Holgado Alvarez, Mahdyar Ravanbakhsh, Begüm Demir, and Faculty. S2-CGAN: Self-Supervised Adversarial Representation Learning for Binary Change Detection in Multispectral Images. *International Geoscience and Remote Sensing Symposium (IGARSS 2020)*, pages 2515–2518, 2020. doi:10.48550/arXiv.2007.02565.
- [109] Anuvi Rawat, Anil Kumar, Priyadarshi Upadhyay, and Shashi Kumar. Deep learning-based models for temporal satellite data processing: Classification of paddy transplanted fields. *Ecological Informatics*, 61(November 2020):101214, 2021. doi:10.1016/j.ecoinf.2021.101214.
- [110] Reza Khatami, Giorgos Mountrakis, and Stephen V. Stehman. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment*, 177:89–100, 2016. URL: <http://dx.doi.org/10.1016/j.rse.2016.02.028>, doi:10.1016/j.rse.2016.02.028.
- [111] Flávio F. Camargo, Edson E. Sano, Cláudia M. Almeida, José C. Mura, and Tati Almeida. A comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian tropical savanna using ALOS-2/PALSAR-2 polarimetric images. *Remote Sensing*, 11(13), 2019. doi:10.3390/rs11131600.
- [112] Yuzhen Zhang, Jingjing Liu, and Wenjuan Shen. A review of ensemble learning algorithms used in remote sensing applications. *Applied Sciences*, 12(17), 2022. URL: <https://www.mdpi.com/2076-3417/12/17/8654>, doi:10.3390/app12178654.
- [113] Stefanos Georganos, Tais Grippa, Sabine Vanhuysse, Moritz Lennert, Michal Shimoni, and Eleonore Wolff. Very High Resolution Object-Based Land Use-Land Cover Urban Classification Using Extreme Gradient Boosting. *IEEE Geoscience and Remote Sensing Letters*, 15(4):607–611, 2018. doi:10.1109/LGRS.2018.2803259.

- [114] Zhenfeng Shao, Muhammad Nasar Ahmad, and Akib Javed. Comparison of random forest and xgboost classifiers using integrated optical and sar features for mapping urban impervious surface. *Remote Sensing*, 16(4), 2024. URL: <https://www.mdpi.com/2072-4292/16/4/665>, doi:10.3390/rs16040665.
- [115] CORNARE. Plan de crecimiento verde y desarrollo compatible con el clima para el oriente antioqueño, 2015. URL: <https://www.cornare.gov.co/plan-de-crecimiento-verde-y-desarrollo-compatible-con-el-clima/>.
- [116] Cámara de Comercio del Oriente Antioqueño. Competitividad en el oriente antioqueño, 2023. URL: <https://ccoa.org.co/wp-content/uploads/2023/06/COMPETITIVIDAD-EN-EL-ORIENTE-ANTIOQUEÑO-2023-vf.pdf>.
- [117] IDEAM. Mapas de coberturas de la tierra, 2018. URL: <http://www.ideam.gov.co/web/ecosistemas/coberturas-nacionales>.
- [118] Daniele Zanaga, Ruben Van De Kerchove, Wanda De Keersmaecker, Niels Souverijns, Carsten Brockmann, Ralf Quast, Jan Wevers, Alex Grosu, Audrey Paccini, Sylvain Vergnaud, Oliver Cartus, Maurizio Santoro, Steffen Fritz, Ivelina Georgieva, Myroslava Lesiv, Sarah Carter, Martin Herold, Linlin Li, Nandin-Erdene Tsendbazar, Fabrizio Ramoino, and Olivier Arino. Esa worldcover 10 m 2020 v100, October 2021. doi:10.5281/zenodo.5571936.
- [119] René Tobar-Díaz, Yan Gao, Jean Francois Mas, and Víctor Hugo Cambrón-Sandoval. Clasificación de uso y cobertura del suelo a través de algoritmos de aprendizaje automático: revisión bibliográfica. *Revista de Teledetección*, (62):1–19, 2023. doi:10.4995/raet.2023.19014.
- [120] Edison Fabian Vargas Becerra. Revisión bibliométrica de las principales metodologías para la evaluación del uso y cobertura del suelo. 2024.
- [121] Cristian Muñoz Ordóñez, Julián Muñoz Ordóñez, Leonairo Pencue Fierro, and Apolinar Figueroa Casas. Mapeo de cobertura terrestre utilizando aprendizaje máquina. *Investigación e Innovación en Ingenierías*, 8(3):85–101, 2020. doi:10.17081/invinno.8.3.4706.
- [122] Yongguang Hu, Ali Raza, Neyha Rubab Syed, Siham Acharki, Ram L. Ray, Sajjad Hussain, Hossein Dehghanisanij, Muhammad Zubair, and Ahmed Elbeltagi. Land use/land cover change detection and ndvi estimation in pakistan's southern punjab province. *Sustainability*, 15(4), 2023. doi:10.3390/su15043572.

- [123] Yong Zha, Jay Gao, and Shaoxiang Ni. Use of normalized difference built-up index in automatically mapping urban areas from tm imagery. *International journal of remote sensing*, 24(3):583–594, 2003. doi:10.1080/01431160304987.
- [124] Salvatore Di Fazio Giandomenico De Luca, João M. N. Silva and Giuseppe Modica. Integrated use of sentinel-1 and sentinel-2 data and open-source machine learning algorithms for land cover mapping in a mediterranean region. *European Journal of Remote Sensing*, 55(1):52–70, 2022. arXiv:<https://doi.org/10.1080/22797254.2021.2018667>, doi:10.1080/22797254.2021.2018667.
- [125] Hanqiu Xu. Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery. *International journal of remote sensing*, 27(14):3025–3033, 2006. doi:10.1080/01431160600589179.
- [126] Ren David Da Silva Erli Pinto dos Santos and Cibele Hummel do Amaral. Vegetation cover monitoring in tropical regions using sar-c dual-polarization index: seasonal and spatial influences. *International Journal of Remote Sensing*, 42(19):7581–7609, 2021. doi:10.1080/01431161.2021.1959955.
- [127] Erli Pinto dos Santos, Michel Castro Moreira, Elpídio Inácio Fernandes-Filho, José Alexandre M. Demattê, Emily Ane Dionizio, Ren David da Silva, Renata Ranielly Pedroza Cruz, Jean Michel Moura-Bueno, Uemeson José dos Santos, and Marcos Heil Costa. Sentinel-1 imagery used for estimation of soil organic carbon by dual-polarization sar vegetation indices. *Remote Sensing*, 15(23), 2023. URL: <https://www.mdpi.com/2072-4292/15/23/5464>, doi:10.3390/rs15235464.
- [128] Jennifer N. Hird, Evan R. DeLancey, Gregory J. McDermid, and Jahan Kariyeva. Google earth engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. *Remote Sensing*, 9(12), 2017. URL: <https://www.mdpi.com/2072-4292/9/12/1315>, doi:10.3390/rs9121315.
- [129] Pierre-Louis Frison, Bénédicte Fruneau, Syrine Kmiha, Kamel Soudani, Eric Dufrêne, Thuy Le Toan, Thierry Koleck, Ludovic Villard, Eric Mougín, and Jean-Paul Rudant. Potential of sentinel-1 data for monitoring temperate mixed forest phenology. *Remote Sensing*, 10(12), 2018. URL: <https://www.mdpi.com/2072-4292/10/12/2049>, doi:10.3390/rs10122049.

- [130] David. Nhemaphuki, Kiran. Thapa Chetri, and Sanjeevan. Shrestha. Fusion of Radar and Optical Data for Land Cover Classification Using Machine Learning Approach. *Journal on Geoinformatics*, 20(1):39–45, 2020. doi: 10.1080/10106040208542232.
- [131] Swapan Talukdar, Pankaj Singha, Susanta Mahato, Shahfahad, Swades Pal, Yuei An Liou, and Atiqur Rahman. Land-use land-cover classification by machine learning classifiers for satellite observations-A review. *Remote Sensing*, 12(7), 2020. doi:10.3390/rs12071135.
- [132] Muhammad Junaid, Jianguo Sun, Amir Iqbal, Mohammad Sohail, Shahzad Zafar, and Azhar Khan. Mapping lulc dynamics and its potential implication on forest cover in malam jabba region with landsat time series imagery and random forest classification. *Sustainability*, 15(3), 2023. URL: <https://www.mdpi.com/2071-1050/15/3/1858>, doi:10.3390/su15031858.
- [133] Ricardo Borja Robalino, Antonio Monleón Getino, and José Rodellar. Estandarizacion de metricas de rendimiento para clasificadores Machine y Deep Learning. *Revista Ibérica de Sistemas y Tecnologías de Información*, pages 184–196, 2020.
- [134] Ayesha Shafique, Guo Cao, Zia Khan, Muhammad Asad, and Muhammad Aslam. Deep Learning-Based Change Detection in Remote Sensing Images: A Review. *Remote Sensing*, 14, 2022. doi:10.3390/rs14040871.
- [135] Nicola Clerici, Cesar Augusto Valbuena Calderón, and Juan Manuel Posada. Fusion of sentinel-1a and sentinel-2a data for land cover mapping: a case study in the lower magdalena region, colombia. *Journal of Maps*, 13(2):718–726, 2017. doi:10.1080/17445647.2017.1372316.
- [136] Karen V Suárez-Parra, Germán E Cély-Reyes, and Fabio E Forero-Ulloa. Validación de la metodología corine land cover (clc) para determinación espacio-temporal de coberturas: caso microcuenca de la quebrada mecha (cómbita, boyacá), colombia. *Biota Colombiana*, 17(1):1–15, 2016. doi: 10.21068/C2016v17r01a01.
- [137] Jing Ling and Hongsheng Zhang. Wcdl: A weighted cloud dictionary learning method for fusing cloud-contaminated optical and sar images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:2931–2941, 2023. doi:10.1109/JSTARS.2023.3259469.
- [138] Mulugeta Yikuno Lilay and Gizatie Desalegn Taye. Semantic segmentation model for land cover classification from satellite images in gambia

- Ila national park, ethiopia. *SN Applied Sciences*, 5(3):76, 2023. doi: 10.1007/s42452-023-05280-4.
- [139] Jesús A Anaya, Susana Rodríguez-Buriticá, and Maria C Londoño. Clasificación de cobertura vegetal con resolución espacial de 10 metros en bosques del caribe colombiano basado en misiones sentinel 1 y 2. *Revista de Teledetección*, (61):29–41. doi:10.4995/raet.2023.17655.
- [140] Zander S. Venter, David N. Barton, Tirthankar Chakraborty, Trond Simensen, and Geethen Singh. Global 10 m land use land cover datasets: A comparison of dynamic world, world cover and esri land cover. *Remote Sensing*, 14(16), 2022. URL: <https://www.mdpi.com/2072-4292/14/16/4101>, doi:10.3390/rs14164101.
- [141] Cidália C. Fonte, Diogo Duarte, Ismael Jesus, Hugo Costa, Pedro Benevides, Francisco Moreira, and Mário Gaetano. Accuracy assessment and comparison of national, european and global land use land cover maps at the national scale—case study: Portugal. *Remote Sensing*, 16(9), 2024. URL: <https://www.mdpi.com/2072-4292/16/9/1504>, doi:10.3390/rs16091504.

