Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado

MAESTRÍA EN MÉTODOS CUANTITATIVOS PARA LA GESTIÓN Y ANÁLISIS DE DATOS EN ORGANIZACIONES

TRABAJO FINAL DE MAESTRÍA

Detección urbana y aprendizaje profundo

Imágenes de RADAR SAR Sentinel 1 en la ciudad de Bogotá

Urban detection and Deep Learning

Sentinel 1 imagery on Bogotá city

AUTOR: ANDRES LEONARDO VELA FONSECA

DIRECTOR: JAVIER GARCÍA FRONTI

MAYO 2024

Resumen

Gracias a las técnicas de inteligencia artificial ahora es posible procesar con alta frecuencia vastos volúmenes de datos, permitiendo así ajustar las estrategias, planes, programas y proyectos, tomando decisiones basados en la información extraída de los datos (Provost & Fawcett, 2013). Es así como cada día se pretende hacer más uso de tales técnicas en las organizaciones privadas y públicas, no estando ajenas a estas tendencias las entidades administradoras del territorio, las cuales han trabajado con imágenes satelitales durante años, ya que han demostrado ser imprescindibles en la planificación del territorio (Wellmann et al., 2020). Sin embargo, algunas zonas no pueden gozar de este beneficio debido a su alta nubosidad la mayor parte del año, lo que dificulta el uso de dicha fuente de datos. Por lo tanto, en los últimos años se han incorporado cada vez más imágenes de satélite de RADAR SAR en la observación de la tierra, las cuales ofrecen una frecuencia alta en la obtención de datos, haciéndolas propicias para el estudio, monitoreo y seguimiento de desastres naturales. Debido a sus características de adquisición, se puede obtener imágenes de día o noche, con alta nubosidad y en algunos satelitales es posible la adquisición de datos con frecuencia diaria, lo que implica una cantidad muy importante de datos. Sin embargo, estos datos cuentan con un tamaño considerable y una interpretabilidad más compleja, para lo cual la inteligencia artificial se presenta como alternativa para poder procesarlos e identificar cambios en los tejidos urbanos de los territorios, proveyendo información más rápidamente para la toma de decisiones.

Por lo anterior, el propósito de este trabajo fue establecer una alternativa para la identificación y cuantificación de áreas con características urbanas, utilizando imágenes de RADAR SAR Sentinel 1, las cuales no son fuertemente afectadas por las condiciones climáticas y pueden mantener su frecuencia de adquisición sin importar la hora del día, y algoritmos de aprendizaje profundo como modelos de redes neuronales convolucionales, con el fin de poder clasificar y calcular la cantidad de área que posee características urbanas, tomando el caso del área metropolitana de Bogotá. Para tal fin, fue necesario identificar de las diferentes opciones de imágenes ofrecidas por Sentinel 1, cual formato y producto es apropiado para el contexto urbano. También, dado que la interpretación de imágenes de RADAR no es intuitiva, una cadena de procesamiento fue necesaria para resaltar las características urbanas, disminuir el ruido y maximizar la distinción con otros elementos presentes en las imágenes. Además, se probaron y compararon 3 modelos de aprendizaje profundo en la clasificación de áreas

urbanizadas y no urbanizadas para finalmente usar el más destacado para clasificar el área metropolitana de Bogotá entre los años 2017 y 2023, lo cual dejó ver la cantidad y la ubicación del aumento del tejido urbano.

Palabras clave

RADAR SAR, polarización, mapa de coherencia, aprendizaje automático, aprendizaje profundo, Deep learning, red neuronal convolucional, Sentinel 1, Bogotá.

Summary

Thanks to artificial intelligence techniques, it is now possible to process vast volumes of data with high frequency, allowing for the adjustment of strategies, plans, programs, and projects based on information extracted from the data (Provost & Fawcett, 2013). Each day, there is an increasing intention to make more use of such techniques in both private and public organizations. Territorial administrative entities are not exempt from these trends, and they have been working with satellite images for years, proving indispensable in territorial planning (Wellmann et al., 2020). However, some areas cannot benefit from this advantage due to high cloud cover most of the year, hindering the use of such data sources. Therefore, in recent years, there has been a growing incorporation of synthetic aperture RADAR (SAR) satellite images in Earth observation, offering high-frequency data acquisition suitable for the study, monitoring, and tracking of natural disasters. Due to their acquisition characteristics, SAR images can be obtained day or night, in high cloud cover conditions, and some satellites allow daily data acquisition, resulting in a significant amount of data. However, these data have considerable size and more complex interpretability. Artificial intelligence emerges as an alternative to process and identify changes in urban tissues more quickly, providing information for decision-making.

The purpose of this study was to establish an alternative for the identification and quantification of areas with urban characteristics using RADAR SAR Sentinel 1 images. These images are not strongly affected by weather conditions and can maintain their acquisition frequency regardless of the time of day. Deep learning algorithms, specifically convolutional neural network models, were employed to classify and quantify the urban areas, focusing on the metropolitan area of Bogotá. To achieve this objective, it was necessary to identify the most suitable format and product among the options offered by Sentinel 1 for the urban context. Additionally, since the interpretation of RADAR images is not intuitive, a processing chain was implemented to highlight urban characteristics, reduce noise, and maximize distinction from other elements in the images. Furthermore, three deep learning models were tested and compared for the classification of urbanized and non-urbanized areas. The most successful model was then applied to the Bogotá metropolitan area between the years 2017 and 2023, revealing the quantity and location of urban tissue expansion.

Keywords

RADAR SAR, polarization, coherence map, Machine Learning, Deep Learning, convolucional neural network, Sentinel 1, Bogotá.

Índice

Resumen....2
Palabras clave....3
Summary....4
Keywords....5
Índice....6
Lista de imágenes....7
Lista de tablas....8
Abreviaturas....8
Capítulo 1. Imágenes satelitales de RADAR SAR Sentinel 1 en la clasificación del tejido urbano....14
1.1 Tecnología satelital de imágenes de RADAR SAR....14
1.2. Mapa de coherencia....17
1.3. Imágenes Sentinel 1....19
1.4. Características de tejido urbano en imágenes satelitales de RADAR SAR: (patrones, respuestas y texturas)....21
1.5. Imágenes Sentinel 1 en el monitoreo urbano....23
1.6. Uso de inteligencia artificial en la clasificación de imágenes satelitales....25
1.7. Aprendizaje profundo (Deep Learning)....26
1.8. Redes neuronales convolucionales....28
1.9. Métricas de evaluación de redes neuronales convolucionales....31
1.10. Clasificación de imágenes satelitales....33
Capítulo 2. De la selección de productos Sentinel-1 a la clasificación de áreas urbanas....35
2.1. Adquisición de imágenes Sentinel 1....35
2.2. Preparación de imágenes RADAR SAR Sentinel 1....38
2.3. Imágenes de muestra para el entrenamiento....40

2.4. Arquitectura de red neuronal convolucional 42
2.5. Construcción de imagen satelital clasificada 44
Capítulo 3. Hallazgos de procesamiento y resultados de clasificación. 45
3.1. Medidas de tendencia central en la eliminación de ruido de las imágenes 46
3.2. Métricas de la clasificación automática 48
3.3. Comparación visual de resultados 49
3.4. Refinamiento en la clasificación 51
3.5. Comparación temporal de la clasificación de CNN entrenada 53
3.6. Cuantificación de áreas Bogotá 55
4. Conclusiones 58
4.1. Ventajas y desventajas del clasificador 59
4.2. Coherencia en su aplicación 60
4.3. Recomendaciones para estudios posteriores 61
5. Referencias bibliográficas 62

Lista de imágenes

Imagen 1. Órbita polar. 15
Imagen 2. Geometría de formación de imagen RADAR SAR. 16
Imagen 3. Imagen RADAR SAR. 17
Imagen 4. Mapa de coherencia. 18
Imagen 5. Tejido urbano en imagen de RADAR SAR Sentinel 1. 22
Imagen 6. Comparación polarizaciones VV y VH de Sentinel 1. 24
Imagen 7. Estructura general de red neuronal profunda. 27
Imagen 8. Operación kernel. 29
Imagen 9. Filtros en imágenes. 29

Imagen 10. Estructura red neuronal convolucional. 30
Imagen 11. Imagen Sentinel 1 procesada de 3 bandas.....40
Imagen 12. Muestra de área urbanizada. 41
Imagen 13. Arquitectura de red neuronal entrenada. 42
Imagen 14. Predicciones sobrepuestas. 45
Imagen 15. Comparación filtro valor medio de píxel. 47
Imagen 16. Mejora en distinción de elementos tras la aplicación del filtro. 48
Imagen 17. Refinamiento de predicción por votación. 52
Imagen 18. Comparación áreas clasificadas con tejido urbano área metropolitana de Bogotá, años 2017 y 2023. 56
Imagen 19. Comparación de resultados con catastro Bogotá. 58

Lista de tablas

Tabla 1. Matriz de confusión. 31
Tabla 2. Imágenes adquiridas. 37
Tabla 3. Funciones de activación más usadas en CNN. 43
Tabla 4. Métricas de evaluación de modelos. 49
Tabla 5. Comparación gráfica de resultados de los diferentes modelos. 50
Tabla 6. Ejemplos de cambio de clasificación entre años 2017 y 2023 53
Tabla 7. Diferencia porcentual en área con tejido urbano, años 2017 y 2023. 57

Abreviaturas

RADAR	Radio detection and ranging
SAR	Synthetic apertura RADAR

ASF	Alaska Satellite Facility
SDK	Software development kit
ESA	European espacial agency
SLC	Single look complex
GRD	Ground range detected
RTC	Radiometric terrain correction
INSAR	Interferometric SAR
DEM	Digital elevation model
WGS	World geographic system
AI	Artificial inteligence
ML	Machine learning
DL	Deep learning
CNN	Convolutional neural network
ReLU	Rectified linear units
RGB	Red, Green and Blue
ROC	Receiver Operating Characteristics
AUROC	Area Under ROC

Introducción

El rápido crecimiento de los tejidos urbanos es una manifestación del aumento poblacional y puede deberse a la consolidación económica de la zona, a los desplazamientos forzados desde zonas rurales que buscan una fuente de trabajo ya sea ocasional o informal (Freire, 2006). Esto supone muchos retos para las administraciones locales, las cuales deben contar con un diagnóstico actualizado para la creación, modificación y ejecución de programas que permitan el desarrollo de sus territorios. Es así como en las principales ciudades de América latina, como Río de Janeiro, Buenos Aires, Ciudad de México, Bogotá, Lima, Santiago de Chile, entre varias más, se ha visto un aumento de la población que a su vez ha desencadenado una mayor demanda de vivienda, contribuyendo a la generación de viviendas informales o precarizadas, en su mayoría ubicadas en la periferia de la ciudad, en zonas de riesgo o en áreas sin ningún tipo de cobertura de servicios básicos (Boldrini & Malizia, 2017). Es un reto importante para los gobiernos y las entidades administradoras del territorio poder contar con información actualizada que permita acciones preventivas más que correctivas, debido a la naturaleza del problema, dicha expansión implica cada vez mayor cantidad de área para monitorear.

Desde la década de los años 70 se ha utilizado tecnología satelital para estudiar y monitorear la tierra, lo que abrió las puertas a la realización de estudios a nivel regional más fácilmente gracias a la gran cantidad de datos que proveen (resolución espacial, temporal y espectral), mejorando la clasificación de cobertura y su monitoreo, y a un costo reducido en comparación con las técnicas más convencionales de toma de fotografía aérea. Por otra parte, existen dos grandes categorías de imágenes satelitales, las cuales dependen del sensor que las adquiera, estos pueden ser pasivos (imágenes ópticas) y activos (imágenes de RADAR), donde los primeros requieren de la luz solar para la captura de imágenes, mientras que los segundos emiten su propia energía para la captura. Sin embargo, las imágenes satelitales ópticas no pueden asegurar una frecuencia alta de captura, ya que están muy supeditadas a las condiciones climáticas al momento de la toma, es decir que con alta nubosidad no se puede lograr imágenes satelitales, una situación muy frecuente en zonas con alto relieve como zonas de montaña. Los sistemas activos en cambio tienen algunas ventajas, dado que el sistema activo emite su propia fuente de energía, lo hace bajo condiciones favorables para la toma, lo que le permite minimizar los efectos negativos a la señal causados por condiciones atmosféricas, lo que le ofrece la capacidad de atravesar la capa de nubes (Parag et al., 2024). Por tal razón, en zonas con alto

porcentaje de nubosidad, usualmente en zonas de montaña, templadas o con alta humedad gran parte del año, el análisis de imágenes de RADAR toma mayor relevancia, puesto que pueden adquirirse independientemente de la luz solar o de las condiciones climáticas, asegurando así la frecuencia constante en la toma de imágenes.

Las áreas urbanas pueden estar altamente concentradas en una ciudad o pueden encontrarse en áreas metropolitanas regionales con diversidad de paisajes, lo cual deriva en el análisis de áreas extensas, lo que añade complejidad a los estudios, tanto en términos de recursos computacionales como de personal técnico requerido. Dadas tales condiciones, la inteligencia artificial se presenta como una herramienta potencial para agilizar el procesamiento y análisis de diagnósticos del territorio basados en grandes volúmenes de datos de imágenes de RADAR satelital, apuntando a obtener una metodología que contemple dichos retos y provea una alternativa viable a los administradores del territorio para obtener diagnósticos más actualizados para la toma de decisiones más oportunas.

La expansión urbana planificada permite establecer áreas idóneas para la construcción de edificaciones, residenciales, comerciales e industriales. Dichas áreas garantizan el correcto desarrollo y funcionamiento de los servicios públicos, vías de acceso y la dinámica ambiental y económica. Un territorio que planifica su desarrollo y tiene control sobre los cambios en este, puede tomar decisiones que le permitan corregir tendencias y replantear los planes de desarrollo territorial. Sin embargo, pese a la gestión de las organizaciones administradoras del territorio para controlar las mutaciones en este, el flujo de personas de sectores rurales a las zonas urbanizadas, los precios de las viviendas y la informalidad laboral generan asentamientos informales con poca o nula intervención del gobierno (Sandoval & Sarmiento, 2019). Algunas de las tendencias más comunes son la construcción de edificaciones en zonas no habilitadas para el desarrollo urbano, ya sea por condiciones ambientales, geotécnicas, por invadir zonas de producción agrícola, o por interferir en futuras obras de bien común. Dicha situación en muchos casos pone en riesgo a la población, dado que se generan nuevas áreas con edificaciones de vivienda en zonas categorizadas como de riesgo (inundación, derrumbe, remoción en masa, contaminación ambiental, entre otras.), convirtiéndose en un verdadero reto el realizar una buena gestión de riesgo, que en definitiva es menos costoso que reubicar a dicha población a zonas en condiciones de desarrollo habitacional (Lincoln Institute of Land Policy, 2011). Por lo tanto, una rápida identificación puede ayudar a evitar la expansión del problema, conteniendo la urbanización no autorizada y permitiendo tomar decisiones de reubicación más

apropiadas. Sin embargo, las tareas de reconocimiento en terreno en muchas ocasiones no son fáciles, ya sea por motivos de logística, así como por motivos de seguridad para el reconocedor, por lo que técnicas de percepción remota son ideales para analizar grandes extensiones de territorio con alta periodicidad. Dentro de los productos generados por la percepción remota se encuentran las imágenes de satélite, las cuales se clasifican en diferentes categorías de acuerdo con el sensor que adquiere los datos, y pueden ser adquiridas incluso con una periodicidad diaria. Sin embargo, ¿Cuál categoría de imágenes puede ser utilizada para garantizar una frecuencia de adquisición constante? Esto supone el manejo de gran cantidad de datos ya que estas imágenes tienen un tamaño en disco considerable, y en muchos casos una sola imagen puede requerir de más de 4 gigabytes de espacio en disco para su almacenamiento, por lo que al considerar series temporales se introduce un reto importante, ¿Qué técnica se puede utilizar para interpretar grandes cantidades de datos en poco tiempo, de manera que se pueda extraer información para la toma de decisiones oportunas?

Con base en los interrogantes planteados y buscando una metodología que permitiera identificar de manera recurrente las áreas urbanizadas, que aporte al monitoreo del territorio en zonas en las que el alto porcentaje de nubosidad anual no permite una buena adquisición de imágenes satelitales ópticas, y teniendo en cuenta que deberían ser datos de fácil acceso, alta recurrencia y lo más detalladas posibles, se propuso, por una parte, usar como fuente de datos principal imágenes satelitales de RADAR SAR, las cuales son más propicias a mantener una frecuencia de toma constantes gracias a sus características de adquisición, pudiendo adquirir datos de día o noche y con presencia de nubosidad a diferencia de las imágenes ópticas (Pandya & Gajjar, 2014), así como las correspondientes al satélite Sentinel 1, dado que son las únicas imágenes de RADAR SAR de libre uso con cobertura mundial. Por otra parte, se propuso aplicar inteligencia artificial para la clasificación de las imágenes de RADAR SAR, por medio de técnicas de aprendizaje profundo tales como las redes neuronales convolucionales, dado que estas son reconocidas por su buen desempeño en el tratamiento de imágenes y por su capacidad para trabajar con relaciones espaciales y temporales. Como resultado de esta combinación de datos y modelo predictivo, se configuró un procedimiento que permite la identificación del tejido urbano en zonas de alta nubosidad, ofreciendo una herramienta de diagnóstico de bajo costo para las entidades interesadas en el desarrollo del territorio.

Por lo anterior, para la definición de la metodología se trabajó en tres objetivos, el primero enfocado en la identificación del formato de las imágenes y de los productos derivados de estas,

que permitan una identificación del tejido urbano, a modo de ser utilizadas como muestras para el modelo de aprendizaje automático. El segundo objetivo estuvo dedicado a la identificación y establecimiento de los pasos de procesamiento de las imágenes, las cuales derivan en una imagen sintetizadora de la información de 4 fechas de toma consecutivas de la misma escena, destacando las características propias del tejido urbano identificables en estas imágenes. Además, incluyó la definición de la arquitectura de red neuronal propuesta al tipo de dato de entrada, que permita la clasificación de imágenes en urbano o no urbano. El último objetivo estuvo enfocado en la aplicación del procedimiento, el cual aportó una comparación de resultados de la clasificación del área metropolitana de Bogotá Colombia, entre los años 2017 y 2023.

A lo largo del documento se encuentran 3 capítulos, encontrando en el primero la información importante de referencia, como las características de los sistemas de RADAR SAR y las imágenes Sentinel 1, así como características de las áreas urbanizadas vistas en dichas imágenes. Además, se presenta un producto derivado de estas imágenes llamado mapa de coherencia que aporta información importante sobre los elementos persistentes en el tiempo en tierra.

En el segundo capítulo, se presenta la elección de imágenes satelitales de RADAR SAR en los diferentes formatos necesarios y se define su cadena de procesamiento, la cual configuró la creación de la imagen definitiva compuesta por diferentes imágenes de intervalos de tiempo cortos, mejorando y destacando las áreas urbanizadas. También, se presenta la arquitectura de red neuronal convolucional propuesta y se comparan los resultados obtenidos con dos arquitecturas mucho más complejas y robustas previamente entrenadas, las cuales son de libre uso y acceso.

Finalmente, un capítulo más presenta los hallazgos de la aplicación de la metodología, mostrando para las 3 arquitecturas un comparativo de las métricas de rendimiento, y una inspección visual de los resultados de clasificación. Además, con base en los resultados del procedimiento propuesto, se presenta una estimación del crecimiento del tejido urbano para la ciudad de Bogotá entre los años 2017 y 2023.

Capítulo 1. Imágenes satelitales de RADAR SAR Sentinel 1 en la clasificación del tejido urbano

Las imágenes satelitales son una herramienta importante en la observación de la tierra, permite reconstruir la historia y la dinámica del territorio, entender los cambios climáticos y acercarnos a zonas donde es muy difícil llegar. Las imágenes satelitales obtienen datos a través de la teledetección, lo cual significa que se puede obtener información de lugares remotos sin tener contacto con ellos, así como de ofrecer una amplia variedad de alternativas en sensores para capturar datos, los cuales no son perceptibles por el ojo humano, como por ejemplo sensores infrarrojos. Por otra parte, así como las imágenes satelitales tienen la capacidad de revelar el estado de la vegetación, o la composición mineral de los suelos, también ayudan a entender las dinámicas del ser humano y de la transformación del entorno para la creación de espacios de industria, recreación, vivienda, entre otros, lo cual ha facilitado la toma de decisiones y el monitoreo del territorio.

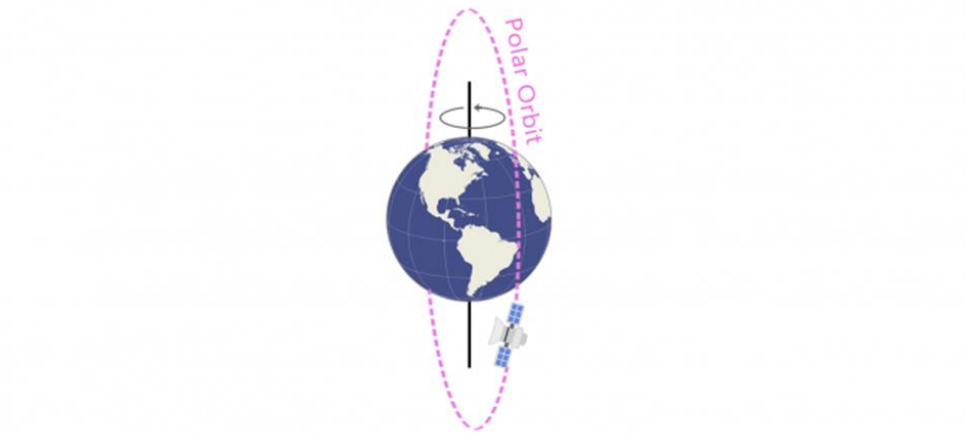
Aunque existen diferentes tipos de imágenes satelitales para el estudio de la tierra, hay una característica que divide este tipo de datos en dos grupos, de sensor activo o pasivo. El primero hace referencia a los sensores que capturan la energía del sol reflejada en la superficie, mientras que la segunda utiliza la energía reflejada de un pulso electromagnético generado por el mismo sensor, como es el caso de las imágenes de RADAR SAR. Este tipo de imágenes tienen características útiles para estudios urbanos, ya que elementos persistentes en el tiempo pueden destacarse fácilmente, además de poder tener imágenes con mayor frecuencia gracias a que no se ven afectadas por condiciones de nubosidad.

1.1 Tecnología satelital de imágenes de RADAR SAR

Un sistema de RADAR trabaja como el flash de una cámara de fotos, pero en vez de flash usando una frecuencia de radio (Chan & Koo, 2008). Es un sistema que utiliza un pulso electromagnético emitido, el cual es enviado en una dirección o área de interés para que interactúe con los objetos que pueda encontrar en su trayectoria. De dicha interacción hay parte de la energía emitida que se absorbe por el objeto, otra parte se refracta o cambia su trayectoria y parte se refleja hacía el emisor. Este sistema requiere de una antena que por definición

conceptual entre mayor sea su longitud mayor es la capacidad de captar energía reflejada. El sistema SAR satelital está compuesto por varias herramientas a bordo de un spacecraft (plataforma espacial), el cual se encuentra a una distancia de alrededor de 700 kilómetros de la superficie terrestre, y viaja con una velocidad aproximada de 27000 kilómetros por hora, siguiendo una órbita casi polar (Imagen 1). El sistema no requiere buenas condiciones de luz solar para adquirir imágenes, ya que, al ser un sistema activo, emite el pulso electromagnético necesario para crear la imagen. Esta es una función muy útil, ya que permite que el sistema funcione indistintamente de día o de noche. Además, la órbita casi polar permite la adquisición de imágenes en forma ascendente o descendente. Estas son algunas razones técnicas que hacen a los sistemas de RADAR satelital de gran ayuda para los estudios de monitoreo de la tierra.

Imagen 1. Órbita polar.



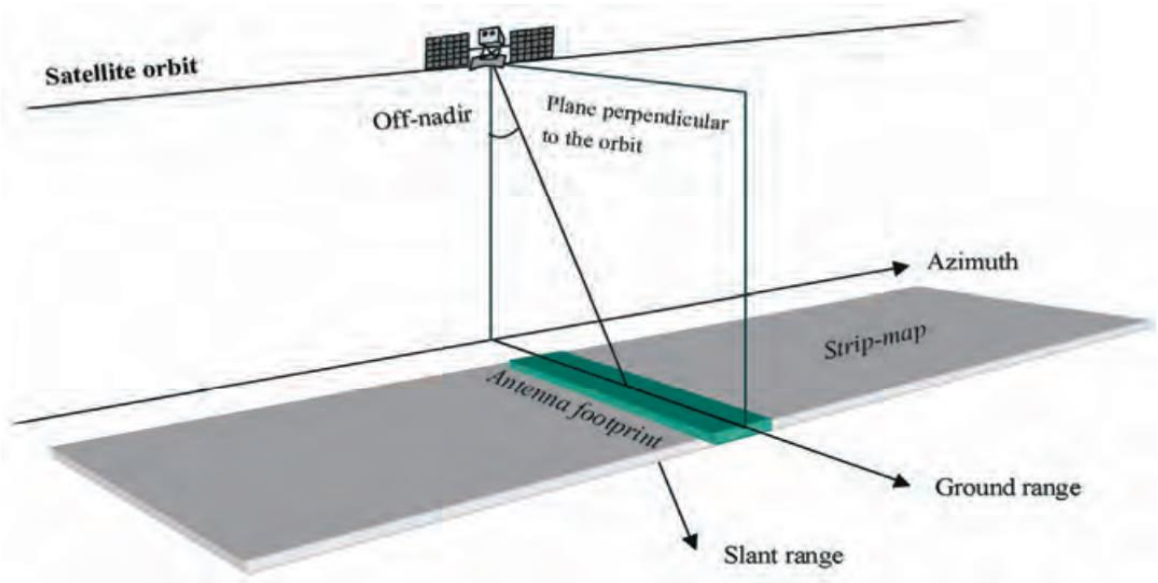
Nota. Adaptado de *What are polar orbits?* [Figura], por SatNow, s.f., SatNow, https://www.satnow.com/community/what-are-polar-orbits

Cuando se conocen las características de la señal emitida, es posible identificar sus cambios una vez que ha regresado a la plataforma (señal devuelta). Por ejemplo, la cantidad de energía emitida puede compararse con la devuelta y usarse para crear una imagen con base en dicha cantidad. Por otra parte, como se conoce la longitud de onda y el pulso electromagnético viaja a la velocidad de la luz, la distancia entre el satélite y la tierra se puede calcular utilizando la siguiente ecuación:

$$c = \lambda * f$$

Donde c es la velocidad de la luz, λ es la longitud de onda y f es la frecuencia del pulso electromagnético. Sin embargo, es importante considerar que la velocidad de la luz es constante en un espacio vacío, mientras que en este caso la señal viaja a través de la atmósfera, esto genera algunas alteraciones que posteriormente deberán ser corregidas para poder usar correctamente las imágenes. La Imagen 2 muestra la geometría de adquisición de las imágenes de RADAR, donde el azimut corresponde a la dirección del movimiento del satélite (Sur - Norte en dirección de paso ascendente), y su línea perpendicular es el rango de tierra (Oeste - Este en dirección de paso ascendente). Esta es una característica interesante porque dependiendo de la dirección de paso, el análisis de los resultados podría cambiar. También es importante hablar de la resolución espacial, entendida como la representación de área de terreno por píxel de imagen. Los sensores de RADAR, de hecho, no siempre muestran una resolución espacial cuadrada (geometría de píxeles cuadrados), a diferencia de los sensores ópticos clásicos (sensores de modo pasivo que adquieren las imágenes utilizando la energía solar reflejada). Esto depende de dos factores principales, el tamaño de la escena y el ángulo fuera del nadir. En una imagen de RADAR, el ángulo fuera del nadir podría estar entre 20 y 50 grados, y podría causar deformaciones en la imagen debido a los efectos de la topografía (superposición y sombra).

Imagen 2. Geometría de formación de imagen RADAR SAR.



Nota. Tomado de *InSAR Principles* (p. A-6), por ESA, 2007, ESA publications.

Imagen 3. Imagen RADAR SAR.

Nota. Tomado de *InSAR Principles* (p. A-7), por ESA, 2007, ESA publications.

Debido a la interacción de múltiples elementos en tierra con la señal del RADAR, se obtiene múltiples respuestas que se superponen en la posición del píxel, lo que genera heterogeneidad con los pixeles vecinos, produciéndose el efecto moteado o de sal y pimienta, incluso cuando el área es homogénea como el caso de un solo tipo de cultivo o vegetación. Esta característica de las imágenes de RADAR SAR impacta negativamente en las tareas de segmentación, sin embargo, el uso de varias imágenes de la misma área o con diferentes ángulos de toma, promediando el valor capturado por cada una de esta, permite disminuir el efecto y crear imágenes más homogéneas. (ESA, 2007).

1.2. Mapa de coherencia

El mapa de coherencia es un producto derivado de las imágenes de RADAR SAR, el cual se utiliza para detectar y medir pequeños cambios en la superficie terrestre. El mapa de coherencia se calcula comparando dos o más imágenes de RADAR SAR adquiridas en diferentes momentos y midiendo la similitud de la señal de retorno en cada píxel de la imagen. Los píxeles con una alta similitud se consideran "coherentes", lo que indica que no ha habido cambios significativos en la superficie terrestre en esa área. Por otro lado, los píxeles con una baja similitud se consideran "incoherentes", lo que indica que ha habido cambios en la superficie terrestre. En la Imagen 4 se aprecia como los píxeles menos coherentes se representan con

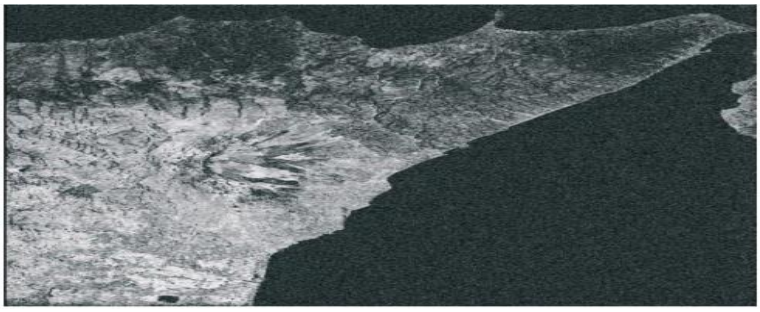
tonos oscuros, tales como el agua o zonas con vegetación, mientras que las áreas con roca o alto relieve se muestran con tonos mucho más claros identificando píxeles más coherentes.

El cálculo del mapa de coherencia requiere la adquisición de dos o más imágenes de RADAR SAR de la misma zona con una diferencia de tiempo conocida. Las imágenes deben tener una geometría similar (es decir, deben estar registradas en la misma geometría espacial y en el mismo sistema de coordenadas), y deben ser adquiridas en condiciones atmosféricas similares para minimizar los efectos de la interferencia atmosférica, lo cual generalmente se logra con intervalos pequeños de tiempo entre adquisición de las imágenes.

Para calcular el mapa de coherencia, se utilizan técnicas de interferometría de RADAR SAR (InSAR), que consisten en comparar la fase de la señal de retorno de las dos imágenes. La fase es una propiedad de la onda electromagnética que indica la posición relativa de la onda en un momento dado. La técnica InSAR compara la fase de la señal de retorno de dos imágenes para calcular la diferencia de fase entre estas. El mapa de coherencia es útil para detectar pequeños cambios en la superficie terrestre, como deformaciones del terreno, movimientos de tierra y cambios en la vegetación. También se utiliza para la clasificación de superficies terrestres, como la detección de áreas urbanas y la identificación de cuerpos de agua.

Sin embargo, para obtener un mapa de coherencia confiable, se debe procurar crearlo con dos imágenes que generen la menor distancia entre el punto de referencia de toma de cada una, lo que se conoce como base perpendicular, así como la menor cantidad de días entre cada toma, lo que se conoce como base temporal, lo que permite contar con condiciones atmosféricas similares.

Imagen 4. Mapa de coherencia.



Nota. Tomado de *InSAR Principles* (p. A-27), por ESA, 2007, ESA publications.

Una vez que se cumplen estos requisitos, se puede generar una imagen de coherencia utilizando técnicas de procesamiento de RADAR SAR, comparando las fases de las señales en las dos imágenes para medir la coherencia. El resultado es una imagen que muestra la coherencia de la superficie terrestre entre las dos fechas de adquisición. El mapa de coherencia oscila en valores de 0 a 1, donde los píxeles incoherentes toman valores bajos y los coherentes valores cercanos a 1. Los píxeles incoherentes pueden deberse a áreas cubiertas por agua que debido a su movimiento natural es difícil que en ambas imágenes se vea similar a nivel de fase de onda, así como la vegetación por los movimientos generados por el viento o por su crecimiento, pero también pueden deberse a cambios en zonas urbanas, como construcción o demolición. Por otra parte, elementos del paisaje como rocas, relieves, construcciones y obras de infraestructura son en muchos casos inmutables entre la toma de dos imágenes, por lo que la respuesta de los píxeles en ambas imágenes es prácticamente idéntica dando lugar a píxeles coherentes con valores altos y muy claros en una escala de grises como usualmente se representa el mapa de coherencia. Algunas de las diferencias entre imágenes pueden estar dadas por ruido inherente del sistema de adquisición, sin embargo, se pueden utilizar técnicas de reducción de ruido que incluye medidas de tendencia central para los mapas de coherencia, mejorando la identificación de los elementos persistentes en el tiempo.

1.3. Imágenes Sentinel 1

La misión Sentinel 1 es la primera del programa Copernicus de la agencia espacial europea (ESA), consta de 2 satélites denominados Sentinel 1A y Sentinel 1B de iguales características, y que proporcionan un sistema de observación de la tierra basado en RADAR, generando imágenes de alta resolución y datos de interferometría de RADAR para una amplia gama de aplicaciones, incluyendo vigilancia marítima, gestión de emergencias, seguridad y gestión del medio ambiente. Estos satélites, llevan a bordo un instrumento de RADAR de apertura sintética (SAR), esto quiere decir que tiene la capacidad de emitir un pulso electromagnético en dirección a la superficie terrestre, el cual interactúa con los elementos en tierra y el resultado de dicha interacción es capturado por una antena de RADAR en el mismo satélite. El pulso que maneja la misión Sentinel 1 tiene como característica una longitud de onda de 5.5 cm, conocida como banda C, y puede ser utilizado para el modo imagen como de interferometría. El modo de imagen opera con la intensidad de la señal, calculando cuanta energía se recibe de la emitida,

permitiendo la observación de la superficie de la tierra bajo cualquier condición meteorológica y en cualquier momento del día o de la noche, mientras que el modo de interferometría opera con la fase de onda del pulso electromagnético, permitiendo medir pequeñas deformaciones de la superficie terrestre con alta precisión. Sin, embargo, esta última es más difícil de interpretar y a menudo no es mostrada como un producto de las imágenes de RADAR SAR, pero si puede ser utilizada para la generación de otros productos derivados a través de técnicas de interferometría. (Thaens & Coman, 2018).

El satélite Sentinel 1 se encuentra en una órbita polar heliosincronica a una altitud de aproximadamente 700 km, lo que le permite cubrir todo el planeta en un ciclo de 12 días (cada 12 días se puede obtener una imagen de la misma escena). Al ser dos satélites que operan en la misma orbita, pero en posición opuesta uno del otro, se puede obtener una imagen desde la misma posición cada 6 días si se opera con ambos satélites.

Las imágenes pueden obtenerse en varios modos de adquisición, con el fin ofrecer el mejor producto para diferentes tipos de aplicaciones. Estos modos de adquisición son:

Modo Stripmap (SM): Este es el modo de adquisición de alta resolución espacial de Sentinel-1, con una resolución de 5x20 metros y una anchura de escena de 80 km. Este modo es adecuado para la detección de cambios pequeños en la superficie terrestre y la cartografía de zonas urbanas.

Modo Interferométrico de Apertura (Interferometric Wide-Swath, IW): Este modo tiene una resolución espacial de 5x20 metros y una anchura de escena de 250 km. Es adecuado para la generación de mapas de coherencia y el monitoreo de movimientos de tierra.

Modo ExtraWide-Swath (EW): Este modo tiene una resolución espacial de 20x20 metros y una anchura de escena de 400 km. Es adecuado para el monitoreo de grandes áreas y la generación de productos de detección de inundaciones.

Modo de Banda Ancha (Wave, WV): Este modo se utiliza para la medición de la altura del océano y la detección de vientos en alta mar.

Además de los modos de adquisición enfocados al tamaño de la escena, también se pueden adquirir imágenes con diferentes polarizaciones y así ampliar el rango de opciones. La polarización se refiere a la dirección del campo eléctrico de las ondas electromagnéticas transmitidas y recibidas por el RADAR. Sentinel 1 puede adquirir imágenes en tres modos de polarización:

Modo de polarización simple (VV o HH): en este modo, el RADAR transmite y recibe ondas electromagnéticas con una única polarización vertical-vertical (VV) u horizontal-horizontal (HH). Estas imágenes son útiles para la detección de características de la superficie terrestre, como la rugosidad del terreno o la vegetación.

Modo de polarización doble (VH o HV): en este modo, el RADAR transmite ondas electromagnéticas con una polarización vertical (V) y recibe ondas con una polarización horizontal (H), o viceversa. Estas imágenes son útiles para la detección de características como la humedad del suelo, la vegetación y el hielo marino.

Modo de polarización completa (VV, VH, HH, HV): en este modo, el RADAR transmite y recibe ondas electromagnéticas con todas las combinaciones posibles de polarización. Las imágenes de polarización completa proporcionan una información más amplia sobre la superficie terrestre y se utilizan para aplicaciones como la clasificación de uso del suelo, la detección de cambios en la superficie terrestre y la cartografía de inundaciones.

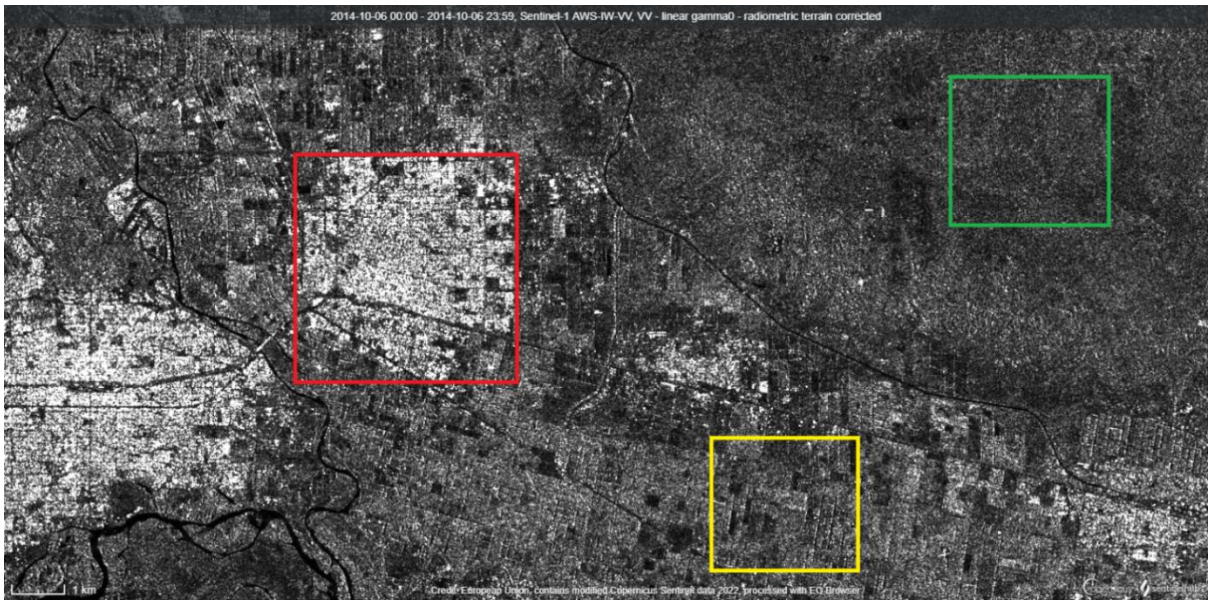
Dadas las diferentes opciones de adquisición, el abanico de aplicaciones es muy amplio, por lo que hay que elegir correctamente el modo de adquisición, polarización, dirección de toma, frecuencia y nivel de procesamiento para cada caso particular.

1.4. Características de tejido urbano en imágenes satelitales de RADAR SAR: (patrones, respuestas y texturas)

Para empezar, conviene recordar que la señal emitida por el sensor de RADAR viaja desde su órbita, atraviesa las condiciones atmosféricas para luego interactuar con los objetos en tierra. Una vez que la interacción se produce la energía emitida por el sensor experimenta refracción y reflexión, así como una dispersión de la energía. Para el caso del tejido urbano, los materiales con los que se edifica proporcionan una fuerte respuesta hacía el RADAR, es decir hay gran parte de la energía reflejada, razón por la cual es característico ver en una imagen de RADAR en escala de grises al tejido urbano en color blanco (European Commision (JRC) & Institute for the Protection and Security of the Citizen (IPSC), 2009) como se aprecia en el recuadro rojo de la Imagen 5. Aunque hay otros materiales en tierra que proporcionan una respuesta similar como el caso de las rocas, las construcciones presentan geometrías más definidas, lo que permite reconocer mejor las zonas urbanizadas. Por otra parte, aunque los cultivos también

presentan geometrías regulares, la respuesta hacía el satélite es mucho más débil, debido a que la vegetación genera mucha más dispersión y por lo tanto la energía reflejada es mucho menor, por lo que, aunque la geometría podría ser similar la respuesta es diferente. Una característica muy común en las imágenes de RADAR es el efecto sal y pimienta (recuadro verde Imagen 5), el cual es un ruido generado en la formación de la imagen, y por lo cual se requiere de técnicas de filtrado o multitemporales para obtener los mejores resultados en su disminución. Este efecto hace que se encuentren respuestas altas y bajas contiguas, lo que dificulta la interpretación, por ejemplo, zonas con respuesta homogénea que presentan valores muy atípicos altos o bajos, los cuales generarían una interpretación o clasificación errónea si estas se basan solo en el valor del píxel (Md Tazmul & Qingmin, 2022), por lo que, se debe tener en cuenta el patrón geométrico en la interpretación y clasificación de las zonas urbanizadas.

Imagen 5. Tejido urbano en imagen de RADAR SAR Sentinel 1.



Nota. Rojo: Tejido urbano, Amarillo: Cultivos, Verde: Suelos desnudos. Adaptado de *EO Browser app*, Sinergise Ltd., (s.f.), https://apps.sentinel-hub.com/eo-browser/. (Contains modified Copernicus Service information [2014] for Copernicus Service Information).

1.5. Imágenes Sentinel 1 en el monitoreo urbano

Las imágenes Sentinel 1 pueden ser usadas para clasificar el territorio y así delimitar las zonas urbanizadas, haciendo uso de un intérprete profesional, utilizando técnicas de clasificación supervisadas o no supervisadas. Sin embargo, con el objetivo de aprovechar al máximo las capacidades de las imágenes satelitales de RADAR SAR, haciendo uso de amplitud y fase, los modos de adquisición que entregan dicha posibilidad son el modo IW y EW. Sin embargo, dado que las imágenes EW son mucho más extensas, es decir abarcan mucha más área de la superficie terrestre que las IW, requieren más capacidad de procesamiento y almacenamiento en disco, por lo que la oferta de imágenes en este modo es muy reducida. Por lo anterior, la mayor disponibilidad de imágenes está en el modo IW, el cual ofrece las mismas prestaciones, pero cubriendo menor área.

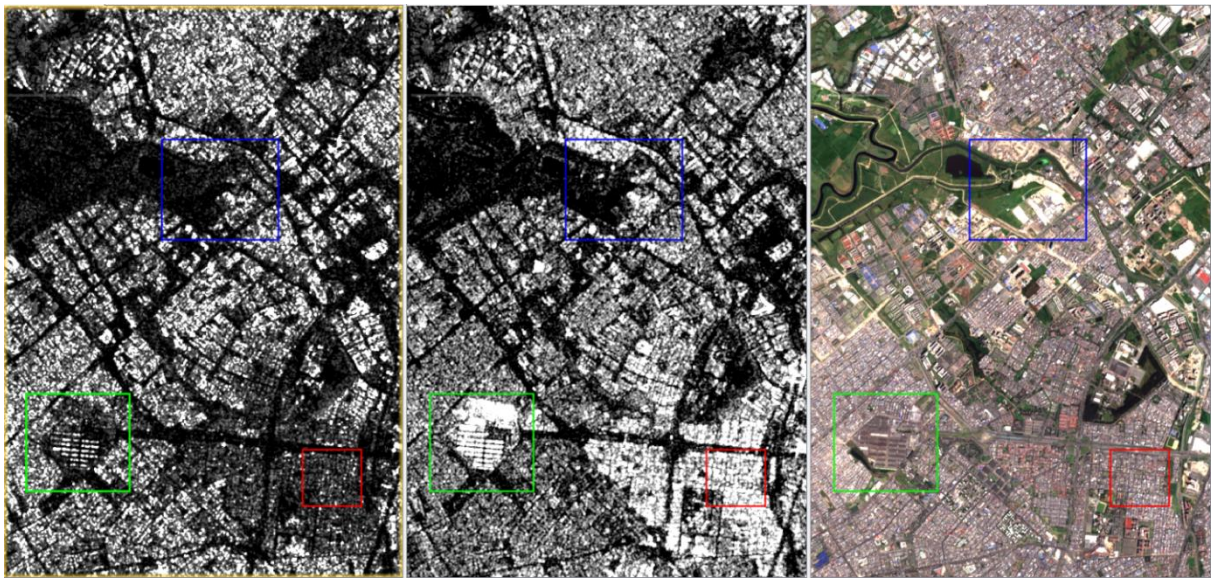
Por otra parte, la misión Sentinel 1, está conformada por dos satélites (A y B), sin embargo, el satélite Sentinel 1B dejó de operar correctamente desde el mes diciembre de 2021, por lo que la frecuencia de revisita de 6 días que se lograba con los dos satélites no estará disponible hasta que el satélite Sentinel 1C sea puesto en órbita, lo que significa que la frecuencia de revisita mínima actualmente es de 12 días. Esto quiere decir que cada 12 días se podrá dar las mejores condiciones en cuanto a base temporal y perpendicular para la creación de un mapa de coherencia, minimizando las interferencias creadas por variaciones atmosféricas, climáticas y topográficas.

Adicionalmente, la oferta de imágenes en dirección ascendente es muy baja en comparación con la dirección descendente, al menos para Suramérica. Esto significa que no es posible utilizar imágenes en modo ascendente y descendente para Suramérica, lo que ayudaría a minimizar los efectos y distorsiones introducidas por la topografía presente en el área Andina (zonas de alta nubosidad).

Por otra parte, la polarización ofrecida en dirección descendente en modo IW es VV y VH, y según algunos autores, la polarización VH es la más apropiada o genera mejores resultados en la clasificación de tejido urbano, aunque se puede complementar con la polarización VV, la cual muestra buena respuesta a patrones geométricos como los característicos de las zonas urbanas. En la Imagen 6, se presenta de izquierda a derecha, imagen Sentinel 1A polarización VH, imagen Sentinel 1A polarización VV e imagen Sentinel 2 (Misión de satélites ópticos

multiespectrales) en RGB como referencia, en las cuales se aprecia en recuadro de color azul que la polarización VH diferencia mejor lo urbano con lo no urbano, mientras que la polarización VV tiene problemas para dicha distinción en cuanto a intensidad de la señal recibida, en recuadro de color rojo se aprecian diferencias en las zonas que cuentan con construcciones techadas con teja de barro y techos a dos aguas, en donde la intensidad es mayor para este tipo de techos para la polarización VV, mientras que en la polarización VH se obtiene una baja intensidad. Por último, en el cuadro color verde se aprecia como la polarización VH puede diferenciar mejor aun siendo dentro del área urbana los edificios de los que no, aunque sacrificando intensidad, a diferencia de la polarización VV que muestra excesiva intensidad, dificultando la identificación de calles al interno del centro de abastecimiento de alimentos localizado en dicha área.

Imagen 6. Comparación polarizaciones VV y VH de Sentinel 1.



Nota. De izquierda a derecha, imagen Sentinel 1 en polarización VH, polarización VV e imagen Sentinel 2 RGB. *Adaptado de EO Browser app*, Sinergise Ltd., (s.f.), https://apps.sentinel-hub.com/eo-browser/. (Contains modified Copernicus Service information [2017] for Copernicus Service Information).

Por lo tanto, teniendo en cuenta que las dos polarizaciones ofrecen información complementaria en la identificación de áreas con cobertura urbana, conviene considerar trabajar con ambas para una correcta clasificación del tejido urbano. Además, los productos

derivados de dos imágenes en el tiempo, como el caso de los mapas de coherencia, ofrecen una descripción de las características de la superficie de los objetos en tierra (Bolter, 2000).

Por otra parte, las imágenes de RADAR cuentan generalmente con un efecto de sal y pimienta, o moteado que introduce ruido y dificulta la interpretación, tal y como se aprecia en la Imagen 5 en el recuadro verde. Para mitigar este efecto se han desarrollado técnicas que permiten filtrar la imagen reduciendo este efecto a costo de un poco de textura, ya que en la etapa de filtrado se busca homogenizar con base en los píxeles vecinos y a su vez puede presentarse perdida de información. Otra técnica de filtrado es por medio de la creación del mapa de coherencia, la cual solo es posible con imágenes IW y EW ya que estas contienen la información de fase necesaria para el cálculo interferométrico. Además, el mapa de coherencia facilita la identificación de áreas urbanas dada su poca variabilidad de respuesta entre imágenes, por lo tanto, el filtro consolida las áreas urbanas y minimiza el efecto sal y pimienta destacando los píxeles que en verdad tienen alta coherencia.

Finalmente, dadas a las características y limitaciones en la oferta de imágenes RADAR SAR Sentinel 1, el modo IW es el modo más apropiado para el estudio de áreas con cobertura urbana en el tiempo, ya que provee la revisita más cercana de 12 días, doble polarización, y aporta la información de intensidad y fase necesaria para la formación del mapa de coherencia.

1.6. Uso de inteligencia artificial en la clasificación de imágenes satelitales

La clasificación de imágenes satelitales ha sido utilizada desde hace varias décadas en diferentes sectores como meteorología, ambiental, agricultura, defensa, ecología, forestal, manejo de desastres, planeamiento urbano y muchas más. Este tipo de imágenes ofrecen la posibilidad de analizar el territorio con una perspectiva general para diagnosticar, planear y monitorear acciones. Sin embargo, las tareas de clasificación tradicionalmente han sido realizadas por personal altamente calificado y demandando bastante recursos computacionales. En la última década se han incorporado más compañías en el mercado de imágenes satelitales, lo que ha impulsado su uso y la exploración en métodos para simplificar los procesos de clasificación que permiten extraer información de las imágenes. Esta búsqueda por nuevas alternativas de procesamiento ha encontrado en la inteligencia artificial un aliado importante, capaz de realizar tareas complejas y optimizando el uso de recursos computacionales. El

estudio de la inteligencia artificial ha llevado a descubrir formas de procesamiento de imágenes y visión por computadora, capaces de realizar tareas de clasificación y segmentación de imágenes, esto incluyendo las imágenes satelitales ya que estas comparten características básicas con cualquier imagen digital, solo que agregan información de referencia espacial que obliga a darles un tratamiento diferencial, pero en esencia, las técnicas de clasificación automática supervisada o no supervisada convencionales pueden ser aplicadas a las imágenes satelitales.

1.7. Aprendizaje profundo (Deep Learning)

La inteligencia artificial (IA) puede ser definida como la ciencia e ingeniería de fabricación de máquinas inteligentes, especialmente programas informáticos inteligentes. Si bien la inteligencia puede descomponerse en razonamiento, aprendizaje, percepción, resolución de problemas e inteligencia lingüística, existen diferentes subcampos de IA que se encargan de estudiar cada uno de estos componentes. Dentro de estos subcampos se puede encontrar el aprendizaje automático (machine learning ML) y el aprendizaje profundo (deep learning DL).

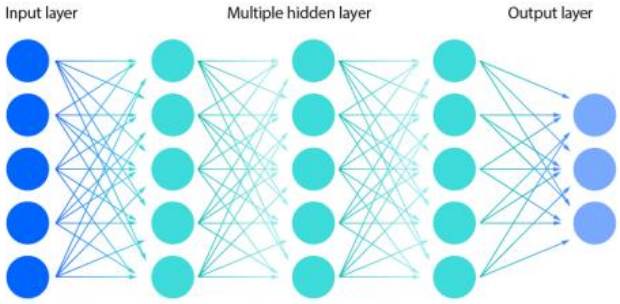
Por una parte, ML requiere más intervención humana, más cantidad de datos estructurados y la jerarquía es más dependiente de la experiencia humana para comprender la diferencia en los datos, mientras que el DL trabaja más en la automatización de extracción de características, minimizando la intervención humana y permitiendo trabajar con cantidades más grandes de datos y no estructurados.

Para poder entender a qué hace referencia DL, se trata de la recreación de la estructura de la red neuronal humana, en la que varias neuronas se interconectan para intercambiar información, de la misma manera se define un modelo matemático e implementado computacionalmente, el cual crea una red de neuronas sintéticas organizadas para trabajar en paralelo, imitando así el comportamiento de ciertos procesos del cerebro humano.

En la Imagen 7, se presenta una ilustración de una red neuronal profunda, en la cual se puede apreciar la conectividad entre las diferentes capas que la componen, donde cada círculo representa una neurona y las flechas indican la conexión con las demás neuronas. En la imagen, también se identifican las neuronas en color verde como parte de las capas ocultas de la red, las azules representan los datos de entrada que son conocidos, y en color azul más claro al

costado derecho, se encuentran los datos de salida, que son los datos resultantes del modelo. Las capas ocultas generan gran cantidad de información útil para el aprendizaje del algoritmo, pero no de fácil entendimiento para el ser humano, por lo que también se identifican como modelos no interpretables o comúnmente denominados “caja negra”.

Imagen 7. Estructura general de red neuronal profunda.



Nota. Tomado de *Qué es una red neuronal?* [Figura], por IBM, s.f., IBM, https://www.ibm.com/mx-es/topics/neural-networks.

Existen varios tipos de redes neuronales profundas, diferenciadas por la arquitectura que usan y los problemas que intentan resolver, tales como las redes neuronales convolucionales (CNN), redes neuronales recurrentes (RNN), redes neuronales de base radial (RBF), redes neuronales de crecimiento (GNN) y redes neuronales antagónicas (GAN). Sin embargo, las redes neuronales profundas más comúnmente usadas son las CNN y RNN, las cuales son empleadas en casos de uso más habituales (IBM, sf), mientras que las demás son usadas en problemas muy particulares como en aplicaciones de grafos y modelos generativos.

Las CNN pueden ser asociadas al procesamiento de datos bidimensionales, tales como imágenes o video, con capacidad de aprender relaciones complejas, patrones geométricos o espaciales, superando las ofrecidas por el ML, aunque su entendimiento y explicación resulta más complejo. Estas capacidades hacen que las CNN sean reconocidas por trabajar muy bien con datos no estructurados, en aplicaciones de detección y clasificación de imágenes satelitales, ya que al trabajar con datos bidimensionlaes pueden extraer patrones geométricos como los generados en contextos urbanos, preservando así la información estructural local (Adams et al., 2021).

Por otra parte, las redes neuronales recurrentes (RNN), a diferencia de las convolucionales ofrecen mejores resultados en otros procesos cognitivos como el procesamiento de texto. La principal diferencia entre estos dos tipos de redes neuronales radica en que las CNN son capaces de identificar patrones espaciales en 2 dimensiones, ideal para procesamiento de imágenes, mientras que las RNN son capaces de identificar patrones de secuencia ideales en el procesamiento de texto, audio o datos temporales (Vosta & Yow, 2022).

Estos tipos de redes buscan simular el procesamiento de datos que realizan las neuronas de una persona, mediante procesos iterativos que dejan aprendizaje en cada iteración. Esto, permite que una computadora pueda, luego de un proceso de entrenamiento, realizar tareas complejas como una clasificación con base en el contenido de una imagen, por ejemplo, una imagen satelital.

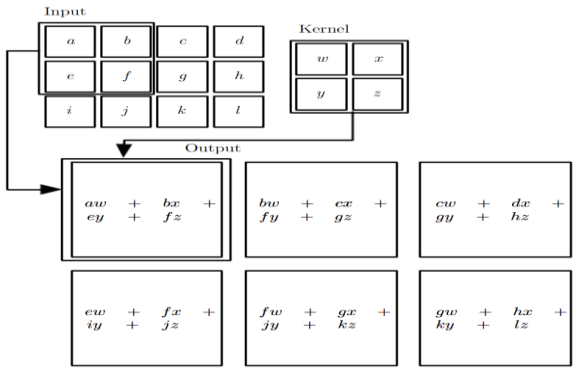
1.8. Redes neuronales convolucionales

Las redes neuronales convolucionales CNN, son redes neuronales artificiales que simulan más el comportamiento de neuronas biológicas encargadas del campo visual, las cuales permiten el procesamiento de imágenes para la extracción de características, para luego discriminar dentro de la imagen los elementos reconocibles. Debido a la forma en que se extraen las características, utilizando información bidimensional, las CNN son capaces de detectar características geométricas, por ejemplo, destacar los bordes de los objetos presentes en la imagen. Esta característica le permite mayor eficiencia en el trabajo con imágenes ya que agrega información extraída de la imagen como un todo y no como píxeles totalmente individuales.

En las etapas de convolución de una CNN, una serie de filtros matriciales se aplican sobre la imagen, ya que una imagen no es más que una matriz en la cual el valor de cada entrada es representa por un color, lo que permite visualmente diferenciar objetos en la imagen. Por lo tanto, para los sistemas computacionales, las imágenes son solo valores almacenados y organizados en matrices, lo cual le permite procesar estas imágenes como operaciones matriciales utilizando operaciones básicas, extrayendo así características de la imagen. Así como se presenta en la Imagen 8, la imagen “Matriz input” es operada por una matriz más

pequeña “kernel” la cual recorre toda la imagen desplazándose a través de esta para generar una imagen de salida “output” de menores dimensiones.

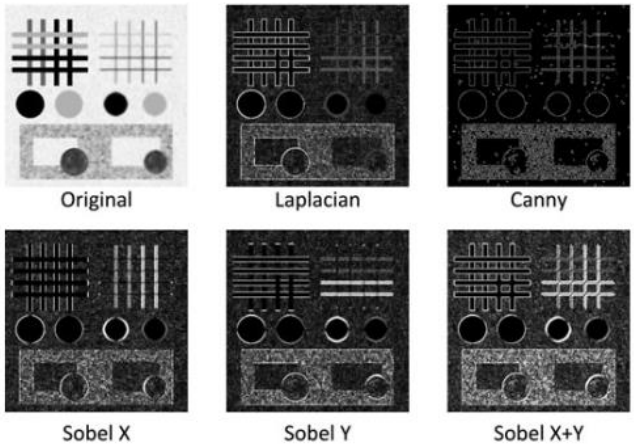
Imagen 8. Operación kernel.



Nota. Tomado de *Deep Learning* (p. 330), por Goodfelow et al., 2016, The MIT press.

El kernel, también llamado filtro, se diseña de forma que al operar la imagen original destaque por ejemplo elementos en forma vertical, horizontal, o en ambos sentidos como presenta la Imagen 9. En las redes convolucionales, diferentes filtros son aplicados esperando extraer la mayor cantidad de información de la imagen original de manera abstracta.

Imagen 9. Filtros en imágenes.

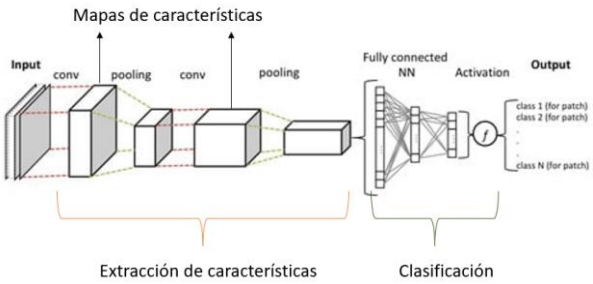


Nota. Tomado de *Image procesing and recognition for biological images*. (p. 531), por Uchida, 2013, Development, Growth & Differentiation.

Una red neuronal convolucional se estructura en capas, la primera capa corresponde a la imagen de entrenamiento o imagen de entada, posteriormente nuevas imágenes llamadas mapas de características son generadas aplicando una serie de filtros a la imagen de entrada (Xie et al., 2019). En estos mapas se almacena la información de manera más abstracta y tras aplicarles una función de activación sirven como entrada para una nueva capa, así sucesivamente por la cantidad de capas que se defina en la estructura. Las funciones de activación permiten validar la información encontrada por el kernel en la creación del mapa de características, evitando la linealidad que se encontraría si todas las características continúan de capa en capa. Algunas de las funciones de activación más comunes en el uso de redes neuronales convolucionales son Sigmoid, Softmax, Tanh, y ReLU, con esta última mayormente utilizada en las capas de extracción de características gracias a que permite un aprendizaje más rápido con costo computacional menor.

Dependiendo de las dimensiones del mapa de características, puede convenir realizar una reducción de dimensión utilizando el proceso llamado pooling para disminuir la cantidad de cómputo (Xie et al., 2019), el cual utiliza un kernel pero esta vez para tomar el valor indicado (máximo o promedio), así minimiza el error de traslaciones locales en la imagen, por ejemplo, si un objeto se mueve levemente entre mapas de características el pooling permite retener la información. Una vez la información es extraída se ingresa a una capa de neuronas densa, es decir, donde cada una conecta con todas las demás para poder realizar el cálculo de pesos que debe darle a cada una de las características encontradas en las imágenes. Finalmente, se define una última capa en la cual la cantidad de neuronas son el número de clases que se desea identificar.

Imagen 10. Estructura red neuronal convolucional.



Nota. Adaptado de *Deep learning for use and land cover classification base don hyperspectral and multispectral earth observation data: A review*. (Figura 6-a), por Vali et al., 2020, Remote sensing.

En la Imagen 10 se aprecia la estructura de una CNN como dos etapas, la primera hace referencia al proceso realizado para la extracción de características y la segunda hace referencia a la etapa de clasificación. La primera etapa tiene como entrada la imagen de entrenamiento y la segunda tiene como entrada las características extraídas. Dado que es un proceso supervisado, en la fase de entrenamiento se le suministran adicionalmente las etiquetas de la clase a la que pertenece cada imagen de entrenamiento, así al entrar en la red neuronal se puede asociar las características encontradas con la categoría indicada, para posteriormente poder clasificar nuevas imágenes en alguna de las clases. De esta manera se puede entender de una manera abstracta el funcionamiento de una CNN, aunque hay que tener en cuenta que es mucho más complejo y que existen diferentes enfoques, combinaciones y variaciones de acuerdo a como el problema lo demande.

1.9. Métricas de evaluación de redes neuronales convolucionales

Como en muchos modelos de aprendizaje automático, las métricas son un paso importante en el proceso, ayudan a entender la efectividad y eficiencia del algoritmo. Algunas de las más comunes en problemas de clasificación son precisión, sensibilidad (recall), especificidad, F1 score, AUROC y accuracy (Maxwell et al., 2021), sin embargo, la elección y uso depende de cada problema en particular. En un modelo de clasificación binaria, como el de identificar si una imagen contiene tejido urbano o no, una herramienta útil para entender qué métrica conviene usar es la matriz de confusión, la cual aporta información acerca de las clasificaciones correctas y erróneas, discriminadas por categoría, de manera que se tiene:

Tabla 1. Matriz de confusión.

Predicción / Clase	Positiva	Negativa
Positiva	Verdadero positivo (VP)	Falso negativo (FN)
Negativa	Falso positivo (FP)	Verdadero negativo (VN)

Con base en la Tabla 1, se puede construir las siguientes métricas:

Precisión: Toma en consideración todos los valores positivos predichos correctamente de una categoría y su relación con la cantidad de predicciones positivas en general:

$$\text{Precisión} = \frac{VP}{VP+FP}$$

Sensibilidad (Recall): Toma en consideración todos los valores positivos predichos correctamente de una categoría y su relación con la cantidad de casos positivos reales. Mide la capacidad del modelo de identificar los casos positivos de una categoría:

$$\text{Recall} = \frac{VP}{VP+FN}$$

Especificidad: Toma en consideración todos los valores negativos predichos correctamente de una categoría y su relación con la cantidad de casos negativos reales:

$$\text{Especificidad} = \frac{VN}{VN+FP}$$

Exactitud (Accuracy): Toma en consideración la cantidad total de predicciones correctas sobre la cantidad de casos. Presenta un panorama general acerca de la calidad del modelo para ambas categorías. Es susceptible al desbalanceo de datos, es decir, que si los datos están desbalanceados puede dar valores buenos que no precisamente reflejan la calidad del modelo para las dos categorías. Se calcula como:

$$\text{Accuracy} = \frac{VP+VN}{VP+VN+FP+FN}$$

A partir de las anteriores métricas, se puede obtener la gráfica de la curva ROC, la cual resume la información de la matriz de confusión y ayuda a mitigar los efectos que las clases desbalanceadas ejercen sobre las métricas. El gráfico se construye con un eje vertical que presenta la tasa de verdaderos positivos o sensibilidad, mientras que el eje horizontal presenta la tasa de falsos positivos o la proporción de clase negativa clasificada como positiva. Este gráfico presenta una forma sencilla de entender el rendimiento, ya que de acuerdo con la forma que genere se infiere si el rendimiento es bueno o no. Sin embargo, la curva ROC es un elemento gráfico el cual hay que analizar, mientras que el área bajo la curva ROC (AUROC) se emplea como métrica de rendimiento ya que resume la información del gráfico, con valor máximo de 1 indica una clasificación perfecta, mientras que un valor de 0.5 indica que el modelo no tiene capacidad de discriminar entre las clases, y valores menores a 0.5 indican un modelo que predice incluso peor que el azar.

Si bien la elección de la métrica puede ser un conjunto de estas, depende del tipo de pregunta y acción a desencadenar a partir de su respuesta. Por lo anterior, si lo que se quiere es abarcar la mayor cantidad de casos positivos predichos correctamente de una categoría tolerando la clasificación de positivos erróneamente, entonces la métrica a tener en cuenta es el Recall, la cual es más alta a mayor cantidad de casos clasificados correctamente de la categoría de interés sin importar cuantos se han clasificado erróneamente de las demás. Por otra parte, si lo que se quiere es evitar la clasificación errónea, por ejemplo, por el alto costo que implica, la métrica de Precisión debería ser alta, ya que esta indica si lo que se clasificó como verdadero en realidad lo era, es decir que se quiere minimizar los casos de falsos positivos. Por otra parte, se encuentra el Accuracy, que integra todas las predicciones correctas con relación a la cantidad de datos, en este caso se pierde información acerca de en qué casos es mejor el predictor por lo que conviene usarla cuando la Precisión y Recall muestran valores similares. Finalmente, se encuentra el F1 Score, el cual es una combinación de Precisión y Recall como una media armónica de estas, utilizando la fórmula:

$$\text{F1 Score} = 2\frac{(Precisión)(Recall)}{(Precisión)+(Recall)}$$

En la anterior fórmula, un valor más alto obtenido indica que el modelo es más fiable y balanceado en las 2 métricas, aunque hay que tener en cuenta que es una métrica susceptible a los datos no balanceados, es decir cuando se tienen mucha mayor cantidad de datos de una clase que de otra, por lo que su adopción podría verse limitada en cuyo caso.

1.10. Clasificación de imágenes satelitales

Existen varios antecedentes del uso de inteligencia artificial en la clasificación de imágenes, modelos de Deep Learning que han mostrado resultados favorables en comparación con otros métodos de clasificación como árboles de decisión, K-vecinos más cercanos y support vector machine (Elngar et al., 2021).

Algunas de las propuestas de arquitectura de CNN entrenadas y probadas que han logrado excelentes resultados en tareas de clasificación de imágenes, están disponibles para su uso libremente. Dentro de estas, se pueden destacar de la última década la ResNet50, VGG16, GoogleNet y LeNet, las cuales han sido entrenadas con conjuntos muy grandes de imágenes,

procesos que requieren de capacidades computacionales extremadamente exigentes. Estos modelos se les conoce como pre-entrenados, y la ventaja de su uso radica en la posibilidad de omitir la etapa de entrenamiento o extracción de características y pasar directamente a la etapa de clasificación. También, es posible que algunos casos requieren de un refinamiento de entrenamiento, pero no es comparable con el entrenamiento con el cual fueron desarrollados, traduciendo esto en una optimización de recursos. Sin embargo, en algunos casos estos modelos refinados superan incluso las capacidades computacionales locales dificultando su uso, así como en algunos casos no reflejan una gran diferencia de rendimiento con modelos más pequeños o diseñados para problemas específicos.

En el caso de las imágenes satelitales, hay varios trabajos acerca del uso de CNN para la extracción de información, como es el caso de ResNet50 para la extracción automática de construcciones en imágenes satelitales (Alsabhan & Alotaib, 2022), o para la clasificación de escenas utilizando un refinamiento de los pesos con base en imágenes satelitales (Shabbir et al., 2021). Así mismo, otra red que se ha estudiado en tareas de clasificación de imágenes satelitales es la VGG16 para estudiar patrones en ambientes urbanos a gran escala (Albert et al., 2017), o la privación ambiental en áreas urbanas.

Como previamente en este documento se mencionó, las imágenes satelitales de RADAR SAR Sentinel 1 pueden abarcar hasta 400 kilómetros de terreno, lo que significa que ciudades enteras pueden ser cubiertas con solo una imagen. Sin embargo, cuando se utilizan para tareas de monitoreo, es necesario utilizar más de una imagen, las cuales deben ser clasificadas y comparadas a lo largo del tiempo, y dado el gran tamaño que tienen las imágenes satelitales, la clasificación automática es un recurso valioso para las tareas de análisis espacial y diagnósticos del territorio, permitiendo extraer información más rápida y oportuna. En el caso de imágenes de RADAR SAR, por su característico efecto de sal y pimienta, las clasificaciones automáticas basadas en el valor de píxel pueden llevar a una incorrecta clasificación, por tal motivo, los métodos de clasificación automática que permiten también analizar el contexto espacial, como las CNN, resultan mucho más adecuadas.

Capítulo 2. De la selección de productos Sentinel-1 a la clasificación de áreas urbanas

Como se mencionó en el capítulo anterior, las imágenes satelitales y en especial las provenientes de sistemas activos como el RADAR SAR, son útiles en la observación y monitoreo de objetos persistentes en el tiempo en la superficie terrestre, tales como desarrollos urbanos, ya que una vez construido puede tomar décadas en ser modificado, generando la misma respuesta para el sistema de RADAR SAR a lo largo del tiempo. No obstante, como cualquier otro dato, por sí mismo puede que no aporte cantidad significativa de información, por lo que en muchos casos es necesario refinar, mejorar, destacar y extraer la información presente en las imágenes, a través de procesamientos enfocados en minimizar distorsiones y ruido, maximizar el contraste, y nivelar los valores obtenidos de las diferentes imágenes.

Por lo tanto, en el presente capítulo se exploran las fuentes de adquisición de las imágenes satelitales Sentinel 1, entendiendo que, aunque son propiedad del programa Copernicus de la Agencia espacial europea ESA, existen varios repositorios que ofrecen acceso a las imágenes, incluso con procesamientos adicionales.

También, se presentan los pasos de procesamiento de las imágenes y su importancia para destacar las características urbanas, para con estas poder extraer las muestras de áreas con y sin tejido urbano, necesarias para el entrenamiento de modelos de inteligencia artificial supervisados.

Por último, se definió una arquitectura de red neuronal convolucional la cual es parte del conjunto de técnicas de aprendizaje profundo, y que haciendo uso de las muestras es capaz de identificar y clasificar una imagen en alguna de las dos categorías. Adicionalmente, incluye una alternativa para refinar los resultados de la clasificación con el fin de mejorar la cuantificación.

2.1. Adquisición de imágenes Sentinel 1

Las imágenes Sentinel 1 pueden ser adquiridas a través del sitio web de Dataspace como fuente oficial del programa Copernicus (https://browser.dataspace.copernicus.eu/), donde se tiene acceso a todo el repositorio de imágenes desde su inicio de operaciones en 2014, y en diferentes

niveles de preprocesamiento, tales como sin procesamiento (raw), imagen compleja (SLC) e imagen ajustada a resolución de terreno (GRD). Otras alternativas incluyen los servicios de Google Earth Engine, o Sentinel Hub, los cuales son servicios que permiten la consulta y procesamiento de imágenes satelitales en la nube, aunque son de tipo comercial y requieren de una suscripción paga, con algunas funcionalidades sin costo, pero limitadas.

Otra alternativa para la descarga de estas imágenes es a través de la Alaska Satellite Facility (ASF), la cual es una entidad especializada en el tratamiento y estudio de datos de imágenes de RADAR SAR, y ofrece en su sitio web (https://search.asf.alaska.edu/) la opción de acceder a una colección de imágenes Sentinel 1 y aplicar algunos procesamientos a estas, tales como la corrección radiométrica de terreno (RTC) e interferometría diferencial SAR (INSAR), procesos que demandan recursos computacionales considerables y se generan por solicitud del usuario, ya sea a través de su sitio web utilizando la interfaz gráfica o por medio de su software development kit (SDK) para Python.

La zona elegida de estudio es conocida por su alto porcentaje de nubosidad y por su gran extensión urbana, lo que la hace más práctica para la recolección de muestras. Se seleccionó un recuadro que incluye el área metropolitana de Bogotá, y se almacenaron las coordenadas de la extensión de este (latitud máxima: 5.0968, latitud mínima: 4.3744, longitud máxima: -73.7865, longitud mínima: -74.4455) con el fin de mantener la misma selección en la etapa de procesamiento. Posteriormente, se realizó una exploración del catálogo de imágenes Sentinel 1 y, aunque existen imágenes desde el año 2014, para el área metropolitana de Bogotá no hay suficientes imágenes que cumplan con la base temporal mínima de 12 días sino hasta finales del mes de septiembre de 2017, siendo dicha base temporal la condición ideal para lograr buenos resultados en el mapa de coherencia. Con base en lo anterior, se definió la imagen adquirida el 30 de septiembre de 2017 como la primera imagen del estudio, ya que es la primera que presentó tener continuidad de toma cada 12 días y, por consiguiente, las siguientes 3 imágenes que ofrecen menor desplazamiento en la posición del satélite al momento de la toma (base perpendicular). Con este conjunto de 4 imágenes, es posible crear 3 mapas de coherencia consecutivos, es decir pares de imágenes con base temporal de 12 días proporcionan un mapa de coherencia; Imágenes con fecha de toma 2017/09/30-2017/10/12 forman un mapa de coherencia, al igual que las parejas 2017/10/12-2017/10/24, y 2017/10/24-2017/11/05.

Asimismo, a modo de contraste temporal y al momento de esta investigación, se definió las 4 imágenes más recientes que cumplieran con los mismos requisitos de base temporal y perpendicular, con fechas 2023/01/08, 2023/01/20, 2023/02/01 y 2023/02/13.

Una vez definidas las imágenes, se procedió a la descarga por medio de los sitios web de Dataspace y ASF. El primer sitio permitió la obtención de las imágenes 8 imágenes con un preprocesamiento GRD, suficiente para la inclusión inmediata en software de análisis de imágenes, y necesario para los procesamientos que permiten la comparación y operación con otras imágenes. Por otra parte, utilizando la interfaz gráfica web del segundo sitio, se solicitó el procesamiento de los mapas de coherencia de los pares de imágenes mencionados previamente. Los mapas de coherencia descargados no requieren de ningún procesamiento adicional más que la integración con las imágenes GRD. La tabla 2 presenta un resumen de las 14 imágenes descargadas, 8 imágenes preprocesadas en nivel GRD y 6 mapas de coherencia.

Tabla 2. Imágenes adquiridas.

Imagen GRD	Fecha de adqusición	Fuente de datos	Mapa de coherencia	Fecha de adqusición	Fuente de datos
1	2017-09-30	Dataspace	1	2017-09-30 2017-10-12	ASF
2	2017-10-12	Dataspace	2	2017-10-12 2017-10-24	ASF
3	2027-10-24	Dataspace	3	2017-10-24 2017-11-05	ASF
4	2017-11-05	Dataspace	4	2023-01-08 2023-01-20	ASF
5	2023-01-08	Dataspace	5	2023-01-20 2023-02-01	ASF
6	2023-01-20	Dataspace	6	2023-02-01 2023-02-13	ASF
7	2023-02-01	Dataspace			
8	2023-02-13	Dataspace			

2.2. Preparación de imágenes RADAR SAR Sentinel 1

Las imágenes en formato GRD requirieron de un procesamiento para ser comparables o interoperables con otras. Era necesario garantizar la correcta sobreposición de píxeles para hacerlos comparables a través del tiempo. Por consiguiente, se debió realizar las correcciones de posición en órbita y distorsión por relieve, además de estandarizar los valores de los píxeles para que los rangos de estos sean comparables. La preparación de datos se realizó localmente por medio del software open source especializado para imágenes Sentinel SNAP, el cual es de uso libre y puede ser descargado gratuitamente del sitio web de la Agencia Espacial Europea (https://step.esa.int/main/download/snap-download/). Una vez instalado fue posible aplicar los siguientes procesamientos a las imágenes:

Orbit correction: El trazado de la órbita seguida por el satélite al momento de la toma con respecto a la órbita planificada puede diferir, por lo tanto, algunos días después de la adquisición el cálculo de la órbita real realizada por el satélite es habilitada para la corrección, logrando que las coordenadas de la imagen sean más precisas y confiables.

Calibration: La calibración es imprescindible para la comparación de imágenes, ya que permite relacionar directamente el valor del píxel con la retrodispersión eliminando sesgos radiométricos del proceso de adquisición.

Terrain correction: La corrección por terreno permite disminuir distorsiones generadas por la geomorfología (shadow, layover, foreshortening).

DEM-Assistend coregistration: Permite co-registrar varias imágenes de RADAR teniendo en cuenta el relieve por medio de un DEM, logrando mejores resultados en la creación de una imagen a partir de otras. La cantidad de bandas resultante es la suma de la cantidad de bandas de cada imagen, en este caso se realizó un proceso para las imágenes GRD obteniendo una imagen de 8 bandas por periodo, ya que cada imagen pose una banda para polarización VH y otra para VV.

Por otra parte, los mapas de coherencia no requirieron el procesamiento mencionado, ya que para poder generar el mapa de coherencia el proceso requiere de co-registrar a la más alta precisión las imágenes, de manera que el producto obtenido de ASF ya poseen dichas correcciones. Sin embargo, para poder operar las imágenes GRD y los mapas de coherencia se utilizaron las siguientes herramientas del mismo software:

Collocation: Permite crear una imagen a partir de otras imágenes. Dado que los mapas de coherencia ya han sido co-registrados y corregidos geométricamente, se puede usar una herramienta más sencilla para crear la imagen que integre los 3 mapas de coherencia, de lo cual se obtuvo una imagen de 3 bandas, una por cada mapa de coherencia.

Band Math: Permite operaciones entre bandas de una imagen, las cuales se utilizaron para el cálculo del valor medio píxel a píxel, creando dos nuevas bandas en la imagen GRD, una para cada polarización (VH y VV), de esta manera homogenizando la imagen y reduciendo ruido de las imágenes originales, mientras que la finalidad en el valor medio de los mapas de coherencia es la de ratificar los elementos con alta coherencia en el tiempo, como son las áreas urbanas.

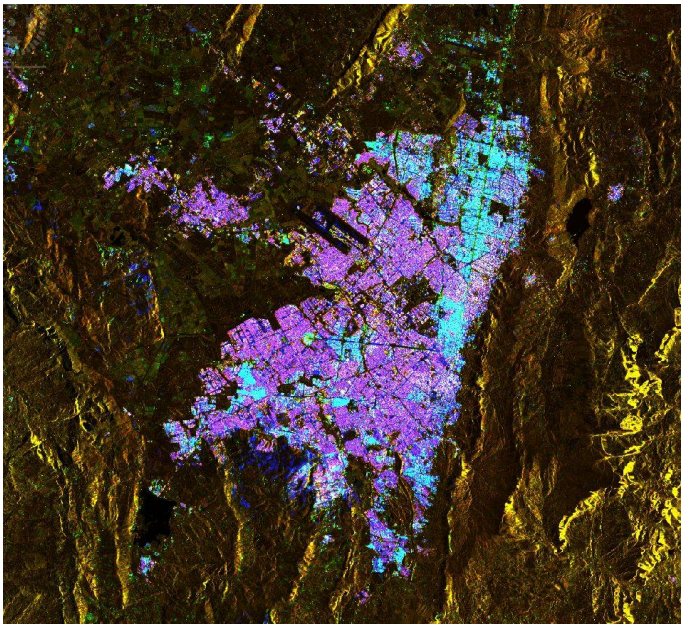
Subset: El proceso de subset permite extraer solo la información deseada de una imagen, tanto en selección geográfica o recorte como en selección de bandas. El proceso se aplicó para las nuevas imágenes creadas a partir del cálculo de valor medio, obteniendo una imagen mucho más pequeña la cual consume menos recursos computacionales.

Reproject: Dado que es necesario crear una sola imagen compuesta por la imagen GRD y mapa de coherencia, esta última se re proyectó al sistema de referencia WGS 1984 coincidiendo así con el sistema de referencia de la imagen GRD.

Collocation: Finalmente, la imagen GRD y el mapa de coherencia promedio se integran, logrando sobreponer las dos imágenes con la misma referencia espacial, asignando canales o bandas para cada una de las imágenes del proceso, obteniendo una imagen con 3 bandas, correspondientes al promedio de amplitud en polarización VV y VH, además del promedio del mapa de coherencia.

La Imagen 11 presenta el resultado de la integración de las imágenes GRD y mapas de coherencia en una visualización a color de 3 bandas (RGB), asignando color rojo a los valores correspondientes a la polarización VH, verde a la polarización VV y azul al mapa de coherencia.

Imagen 11. Imagen Sentinel 1 procesada de 3 bandas.



Nota. Imagen Sentinel 1 procesada de 3 bandas (Rojo: VH media, Verde: VV media, Azul: media mapa de coherencia). (Contains modified Copernicus Service information [2017] for Copernicus Service Information).

2.3. Imágenes de muestra para el entrenamiento

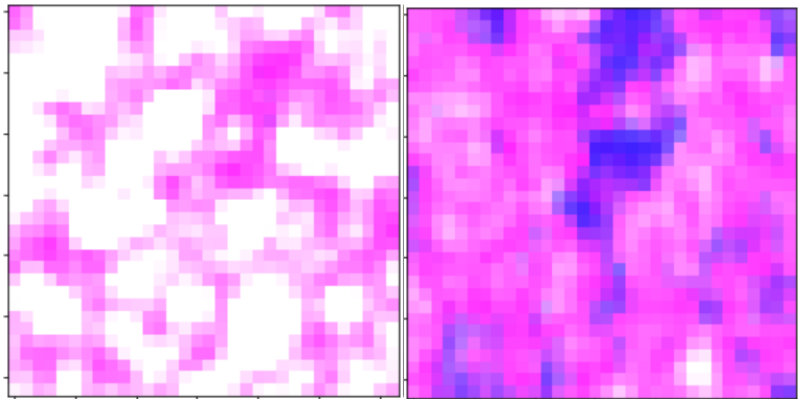
Los modelos de inteligencia artificial requieren de datos de los cuales aprender, por lo cual se utilizan conjuntos de datos seleccionados como muestras de las cuales los sistemas de aprendizaje automático extraen las características que les dan criterios para realizar una clasificación, estimación o predicción. En el caso de las imágenes, es posible clasificar una imagen en alguna categoría, como una imagen de un animal que pueda ser clasificado como una imagen de un perro, gato, etc. Sin embargo, en el caso de las imágenes satelitales, estas deben ser clasificadas no como un todo, sino discriminando el contenido en estas, por lo que se usan porciones de la imagen, las cuales se asignan a una clase si tiene un alto porcentaje de esta. De esta manera una imagen satelital no es clasificada como un todo, sino que se clasifican pequeñas porciones de imagen que forman la imagen original.

Con base en las características urbanas observadas en las imágenes de RADAR SAR presentadas en el capítulo anterior, así como en los patrones visuales de geometría y brillo que

identifican las zonas urbanizadas (Zhao et al., 2020), se decidió tomar las muestras de áreas que presenten tales características.

Para tal fin, fue necesario tomar porciones de la imagen procesada de 3 bandas, teniendo en cuenta que cada píxel de la imagen representaba un área de 10 metros cuadrados, se definió tomar recortes de 32 píxeles por 32 píxeles en dos categorías (Imagen 12). Las categorías se definieron con base en la presencia de tejido urbano, una categoría representa áreas total o parcialmente urbanizadas, y la otra con áreas sin o muy baja urbanización, y se obtuvo 32477 imágenes de muestras distribuidas en 17906 y 14571 respectivamente.

Imagen 12. Muestra de área urbanizada.



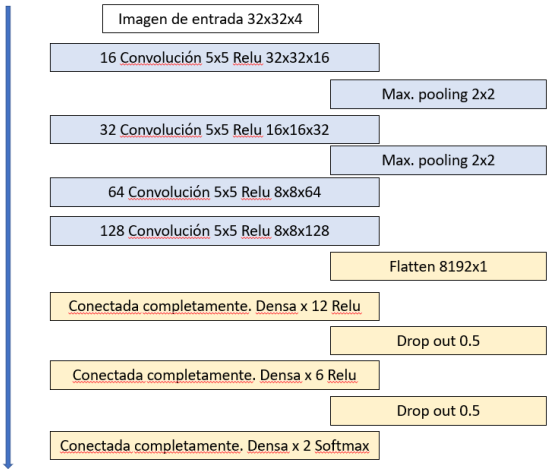
Nota. De izquierda a derecha muestra de área urbanizada y muestra de área no urbanizada. (Contains modified Copernicus Service information [2017] for Copernicus Service Information).

Por otra parte, para llevar a cabo el entrenamiento de modelos de aprendizaje automático de manera supervisada, como es el caso de los modelos en los que se selecciona las muestras de las que se quiere que aprenda, es común utilizar parte del conjunto de muestras para hacer una prueba de su eficiencia, es decir que se obtienen dos conjuntos, uno de entrenamiento y otro de prueba. Se definió entonces tomar el 80% de las muestras para el conjunto de entrenamiento y 20% para la etapa de prueba. Así mismo, se asignaron etiquetas a las muestras, 1 para aquellas pertenecientes a áreas urbanas y 0 a áreas no urbanas.

2.4. Arquitectura de red neuronal convolucional

Una vez definidas las imágenes de muestras de las cuales se pretendía extraer las características de las áreas urbanas y no urbanas, se construyó una red neuronal de 7 capas enfocada en la clasificación de imágenes en estas 2 categorías, 4 capas correspondientes a la extracción de características y 3 dedicadas a la clasificación como se ilustra en la Imagen 13.

Imagen 13. Arquitectura de red neuronal entrenada.



En la imagen anterior se puede visualizar en color celeste el bloque de extracción de características, el cual se compone de 4 capas de 16, 32, 64 y 128 filtros con dimensión de kernel 5 x 5, con la finalidad de crear los mapas de características. Se utilizó una reducción de dimensionalidad de los mapas de características utilizando un pooling del valor máximo (max pooling) solo en las dos primeras capas, lo que permitió sintetizar el mapa de características en uno de menor dimensión, 16x16 y 8x8 respectivamente. Sin embargo, en las siguientes capas no se aplicó max pooling porque los mapas de características ya eran suficientemente pequeños, y una menor dimensión implicaba una pérdida de información y por consiguiente una mala clasificación. Una vez finalizado el bloque de extracción, se realizó la transformación del tensor multidimensional a un vector unidimensional por medio del proceso de aplanamiento (flatten) llevandolo a la forma (q,p) donde q es la cantidad de mapas de características, p la multiplicación de las dimensiones de ancho y alto de los mapas, dimensiones requeridas para el ingreso al bloque de clasificación o a la red de neuronas conectadas completamente, es decir,

que cada entrada del vector se conectó con cada una de las 12 neuronas de la capa siguiente. Posteriormente, se utilizó la instrucción dropout, la cual desconecta aleatoriamente un porcentaje de las neuronas con el fin de minimizar el sobreajuste por la memorización de los datos de entrenamiento, permitiendo una solución más general. Se aplicó un dropout del 50%, valor que experimentalmente ha demostrado mejorar los resultados (Hinton et al., 2012), para conectar las salidas activas a la siguiente capa de 6 neuronas. Nuevamente, se aplicó un dropout del 50% para finalmente conectar con la capa de salida o de clasificación, la cual contenía solo 2 neuronas, una por cada clase (urbano, no urbano). En cuanto a la función de activación, con base en el estado del arte (Tabla 3) se definió utilizar en las capas de extracción de características ReLU, mientras que en el bloque de clasificación se definió la función sigmoid ya que en este caso por ser un problema de clasificación binaria esta función proporciona valores entre 0 y 1, facilitando la interpretabilidad como una probabilidad de pertenecer a una de las dos clases. Sin embargo, su desempeño en la clasificación con las imágenes de prueba presentó bastantes errores en clasificación, por lo que se decidió probar con la función de activación softmax, la cual devuelve una distribución probabilística de cada clase y maximiza las diferencias de probabilidades entre estas, obteniendo mejores resultados en la etapa de validación.

Tabla 3. Funciones de activación más usadas en CNN.

Architecture	Hidden Layers	Output Layer	Cross Reference
AlexNet	ReLU	Softmax	Krizhevsky et al., 2012
NiN	No activation	Softmax	Lin et al., 2013
ZFNet	ReLU	Softmax	Zeiler & Fergus, 2013
VGGNet	ReLU	Softmax	Simonyan & Zisserman, 2015
SegNet	ReLU	Softmax	Badrinarayanan et al., 2015
GoogleNet	ReLU	Softmax	Szegedy et al., 2015
SqueezeNet	ReLU	Softmax	Golkov et al., 2016
ResNet	ReLU	Softmax	He et al., 2016
ResNeXt	ReLU	Softmax	Xie et al., 2017
MobileNets	ReLU	Softmax	Howard et al., 2017
SeNet	ReLU	Sigmoid	Fu et al., 2017

Nota. Tomado de Nwankpa et al., (2018, p. 15)

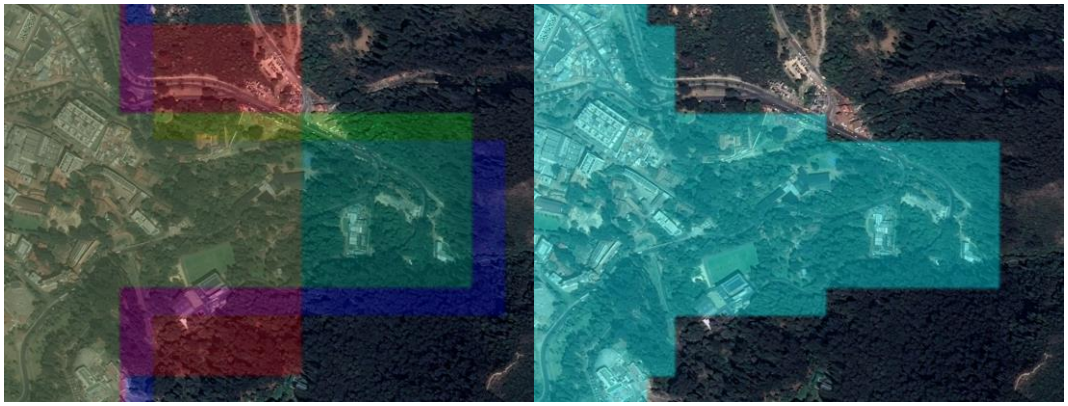
De igual manera, con base en estudios previos acerca de la efectividad de reconocidas arquitecturas de redes neuronales convolucionales, se definió utilizar las arquitecturas

ResNet50 y VGG16, las cuales destacan por su alta capacidad en tareas de clasificación (Shah et al., 2023), con el ánimo de tener un punto de referencia y así comparar el desempeño de la arquitectura propuesta.

2.5. Construcción de imagen satelital clasificada

El último paso fue la construcción de la imagen satelital clasificada, esto debido a que el clasificador solo clasificaba imágenes de 32 pixeles de ancho y alto, tal como fue entrenado, mientras que una imagen satelital puede contener miles de pixeles de ancho y alto. Por lo tanto, la imagen satelital fue subdividida en pequeños recortes de 32 pixeles de ancho y alto los cuales fueron la entrada del clasificador, para después reconstruir con estos la imagen total clasificada. Cómo ya se mencionó, los recortes a clasificar eran de 32 pixeles de ancho y alto, esto significa que cada recorte a clasificar representaba una superficie en terreno de 320 metros cuadrados, en la cual se podía encontrar tejido urbano y no urbano, dificultando la discriminación entre clases, sobre todo en las áreas limítrofes entre urbano y no urbano ya que se pierde la homogeneidad de cada una (Dell'Acqua, 2006). Para contrarrestar esta situación, se estableció un sistema de clasificación píxel a píxel, utilizando varias predicciones para lograr redundancia en las predicciones y obtener algo similar a un sistema de ensamble o votación (Mohammed & Kora, 2023) basado en 10 predicciones, de las cuales se tomó el valor de la moda de cada píxel, dando más consistencia a la clase predicha. Sin embargo, para que el sistema de votación funcionara, las imágenes satelitales a clasificar tuvieron que ser recortadas 10 veces, agregando un ligero desplazamiento con relación a la anterior, con el fin de que en el proceso de predicción de los recortes no se produjera siempre el mismo resultado. Por ejemplo, en la Imagen 14 se presentan en azul, verde y rojo, 3 de las 10 clasificaciones de áreas urbanas obtenidas, mientras que en el costado derecho una imagen con el resultado de la votación en color turquesa.

Imagen 14. Predicciones sobrepuestas.



Nota. De izquierda a derecha, Resultado de las diferentes predicciones sobrepuestas, resultado de votación de clasificación. (Contains modified Copernicus Service information [2017] for Copernicus Service Information).

Tomando el resultado de votación se mejoró la resolución espacial de la predicción, así como también la cuantificación de áreas de tejido urbano ya que se pudo discriminar a más detalle las áreas que no poseen características urbanas.

Capítulo 3. Hallazgos de procesamiento y resultados de clasificación.

Una vez definido el proceso metodológico, pasando por obtener los datos del repositorio de imágenes hasta la clasificación de una escena completa o zona de interés, restó entonces medir y poner a prueba la eficiencia del procedimiento, por lo que este capítulo se enfoca en presentar las métricas obtenidas del modelo de aprendizaje automático utilizado, así como una comparativa con los modelos ResNet50 y VGG16. Esto permitió entender que tan efectivo fue el método de clasificación visto en el capítulo anterior, además de proporcionar un soporte cuantificable acerca de su eficiencia. Por otra parte, se exploró de manera visual los resultados, complementando y corroborando lo que cuantitativamente se obtuvo de las métricas, entendiendo los casos en los que el método de clasificación tiene ventajas y desventajas.

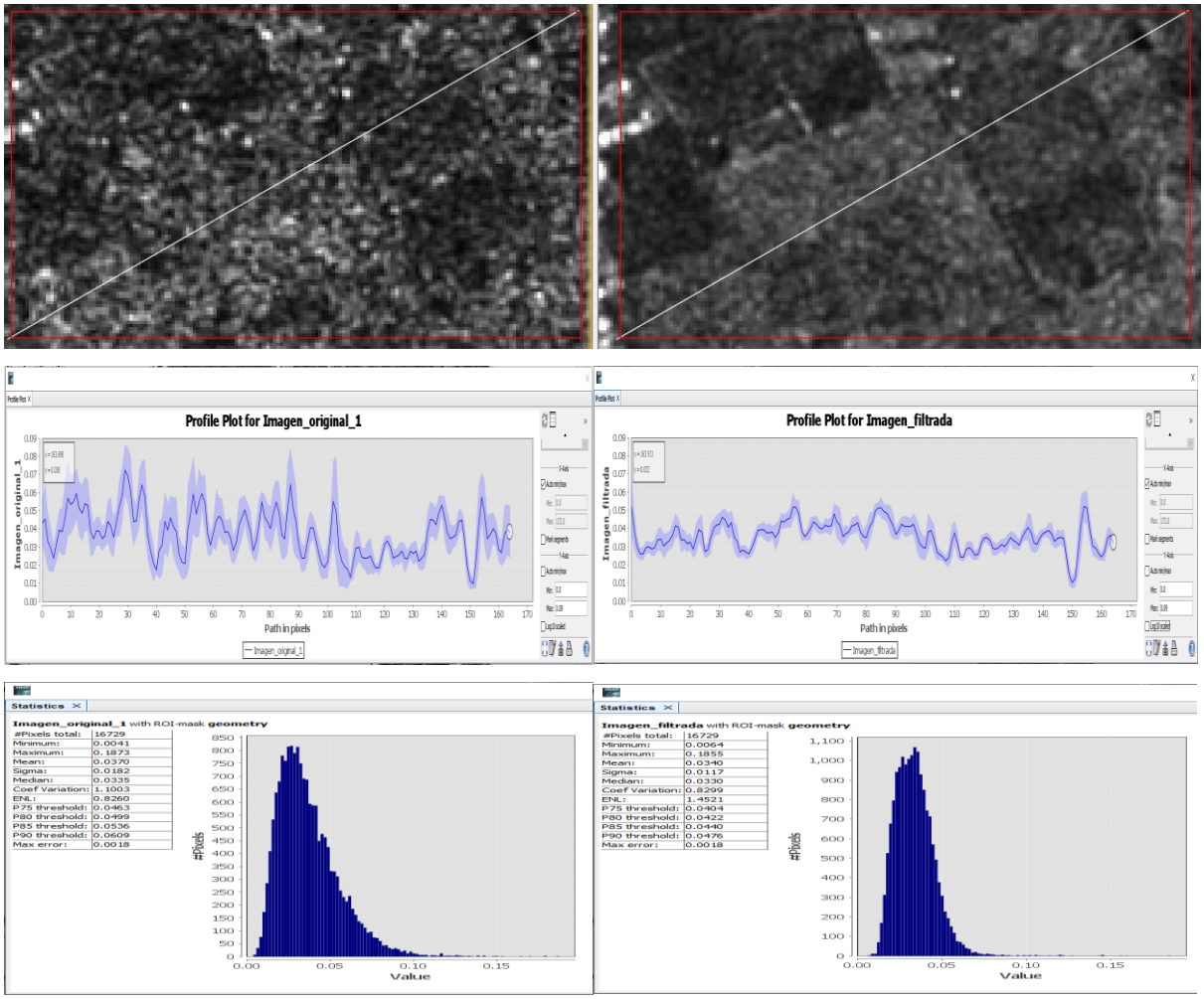
3.1. Medidas de tendencia central en la eliminación de ruido de las imágenes

Como se mencionó en el capítulo 1, las imágenes de RADAR SAR tienen la característica de contar con un efecto sal y pimienta, el cual se genera al momento de formar la imagen, introduciendo ruido aleatorio en esta y dificultando la extracción de información. Se evidenció que el efecto de sal y pimienta produce una respuesta radiométrica similar a la que puede generar un área urbana, esto quiere decir, valores altos de retro dispersión que desfavorece el proceso de clasificación automática haciendo necesarios procesos adicionales para mitigar tal desventaja. Por otra parte, en el capítulo 2 se aplicó un tratamiento para reducir dicho efecto, utilizado una medida de tendencia central como la media, lo cual permitió homogenizar los píxeles con sus vecinos, logrando altos valores de retro dispersión para las zonas que en verdad aportan esta característica y valores mucho más bajos para las zonas no urbanizadas.

En la Imagen 15, se presenta una comparativa del antes y después de la aplicación del filtro de valor medio, evidenciando la mejora en las geometrías con la disminución de ruido, así como dejando ver más claramente los puntos brillantes correspondientes a construcciones. Por otra parte, se puede apreciar en un perfil trazado en zona rural, como disminuye la desviación estándar y se minimiza los cambios abruptos entre píxeles vecinos, mientras que cuantitativamente se logró para el área del perfil, una disminución de aproximadamente 9% en el valor medio de píxel y de 35% en la desviación estándar, lo que significa para la identificación una condición más favorable en los procesos de extracción de características y posteriormente la clasificación realizada por la red neuronal. Asimismo, como lo resalta la Imagen 16, en los rectángulos amarillos, rojo y azul, en el antes y después de la aplicación del filtro, las áreas de cultivo (en color rojo y amarillo) no se aprecian bien definidas antes de la aplicación, así como sus bordes que tienen apariencia difusa, mientras que, en el cuadro azul se presenta un área con algunas construcciones, mostrando que antes de la aplicación del filtro presentaba bastante ruido, el cual hace parecer que hay más construcciones de las que en realidad existen, ya que después de la aplicación del filtro las construcciones se aprecian más consolidadas y con mayor claridad.

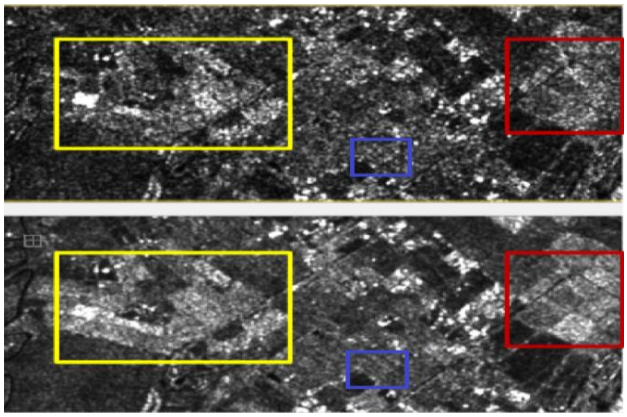
Luego de aplicar el proceso de filtrado, las áreas con cultivos se definen claramente en su extensión y se discrimina cuales cultivos son más altos que otros, mientras que donde parecía haber bastantes construcciones en realidad solo había unas cuantas.

Imagen 15. Comparación filtro valor medio de píxel.



Nota. En la parte izquierda, arriba se encuentra la imagen original sin filtrado con un perfil trazado diagonalmente, en el medio se visualiza los valores radiométricos del perfil y en la parte inferior el histograma de la imagen. En la parte derecha, arriba se aprecia la imagen filtrada con su respectivo perfil, en el medio los valores radiométricos del este, y en la parte inferior el histograma de la imagen resultante. (Contains modified Copernicus Service information [2017] for Copernicus Service Information).

Imagen 16. Mejora en distinción de elementos tras la aplicación del filtro.



Nota. En la parte superior imagen sin filtro, en la parte inferior imagen filtrada. Cultivos en cuadros de color rojo y amarillo, y en azul zonas urbanizadas. (Contains modified Copernicus Service information [2017] for Copernicus Service Information).

3.2. Métricas de la clasificación automática

Las métricas proveen un entendimiento del alcance de los modelos de clasificación automática, así como permiten evidenciar las debilidades y comparar con otras alternativas. Para los problemas de clasificación, es común usar las métricas Precisión, Sensibilidad (recall) y puntaje F1 (F1-score), ya que aportan una perspectiva de la eficiencia del modelo. Por lo tanto, estas métricas fueron calculadas para el modelo de aprendizaje automático propuesto en este documento, así como para los modelos ResNet50 y VGG16, con el fin de tener un punto de comparación con otras arquitecturas comúnmente usadas en problemas de clasificación de imágenes.

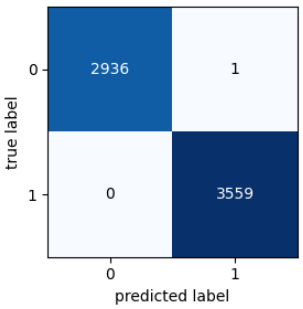
En cuanto a los resultados, los 3 modelos mostraron buen desempeño en la tarea de clasificación, mostrando valores muy altos en todas las métricas. Sin embargo, destaca la arquitectura propuesta y entrenada por completo (CNN entrenada), como se puede apreciar en la Tabla 4, la cual logró clasificar casi perfectamente las muestras de prueba obteniendo el valor más alto en Precisión, Recall y F1 Score, seguido del modelo VGG16 y por último ResNet50.

Tabla 4. Métricas de evaluación de modelos.

CNN entrenada	ResNet50	VGG16

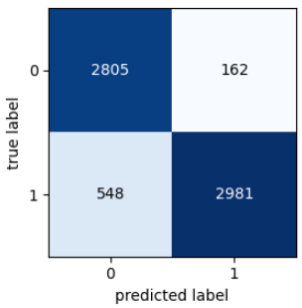
CNN entrenada

	precision	recall	f1-score	support
0	1.0000	0.9997	0.9998	2937
1	0.9997	1.0000	0.9999	3559
accuracy			0.9998	6496
macro avg	0.9999	0.9998	0.9998	6496
weighted avg	0.9998	0.9998	0.9998	6496



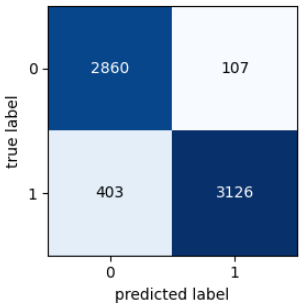
ResNet50

	precision	recall	f1-score	support
0	0.8366	0.9454	0.8877	2967
1	0.9485	0.8447	0.8936	3529
accuracy			0.8907	6496
macro avg	0.8925	0.8951	0.8906	6496
weighted avg	0.8974	0.8907	0.8909	6496



VGG16

	precision	recall	f1-score	support
0	0.8765	0.9639	0.9181	2967
1	0.9669	0.8858	0.9246	3529
accuracy			0.9215	6496
macro avg	0.9217	0.9249	0.9214	6496
weighted avg	0.9256	0.9215	0.9216	6496



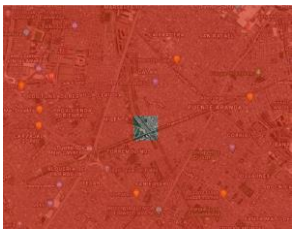
De acuerdo con los resultados presentados en la Tabla 4, la arquitectura más adecuada para clasificar las imágenes de RADAR SAR Sentinel 1 en el contexto binario de características urbanas y no urbanas, es la arquitectura propuesta o entrenada desde cero, puede ser debido a la alta variabilidad entre las clases que sirvieron de entrenamiento (Castelluccio et al., 2015). Sin embargo, aunque los resultados en las etapas de entrenamiento y validación mostraron excelentes resultados, era conveniente producir una predicción de la clasificación del área completa de interés e inspeccionar visualmente los resultados, con el fin de verificar que los valores predichos corresponden a la realidad.

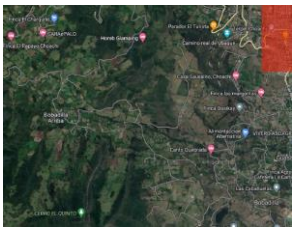
3.3. Comparación visual de resultados

A través de una visualización de la clasificación resultante, se pueden apreciar varias imprecisiones generadas por los 3 modelos. La Tabla 5 presenta una inspección visual y comparativa de los 3 modelos propuestos en las tareas de clasificación, encontrando lo siguiente:

1. En áreas de alta densidad VGG16 muestra mejores resultados, mucho más contundentes al clasificar como urbano. Sin embargo, la CNN muestra más sensibilidad a las áreas verdes o no construidas dentro de estas áreas altamente densificadas, mientras que ResNet50 muestra errores de clasificación del tipo falso negativo.
2. En áreas urbanizadas de baja densidad es más complicado lograr una correcta clasificación, los resultados de las 3 arquitecturas muestran imprecisiones, en común se tiene que las coberturas de tipo tejados traslucidos e inclinados dificulta clasificarlos como urbanos. Sin embargo, destaca VGG16 ya que clasifica la mayor parte de áreas con construcciones como urbano, aun incluyendo áreas con pocas construcciones, implicando un recall más efectivo y errores de clasificación urbana por exceso. Por otro lado, ResNet50 y CNN presentan más falsos negativos, es decir clasifican como no urbano algo que si lo es.
3. En las zonas rurales ResNet50 mostró los peores resultados, clasificando como urbano áreas completamente rurales, mientras que CNN entrenada clasificó solo las áreas con alta densidad urbana como urbano y VGG16 además agregó algunas construcciones en zonas rural.
4. En las zonas mixtas las 3 arquitecturas muestran un comportamiento similar, aunque ResNet50 muestra más áreas clasificadas como urbano cuando la mayor parte del área a clasificar no está construida.

Tabla 5. Comparación gráfica de resultados de los diferentes modelos.

	CNN entrenada	ResNet50	VGG16
Urbano denso			

	CNN entrenada	ResNet50	VGG16
Urbano baja densidad			
Zona rural			
Zona mixta			

Nota. Adaptado de *Google Earth*. (Google, s.f.).

3.4. Refinamiento en la clasificación

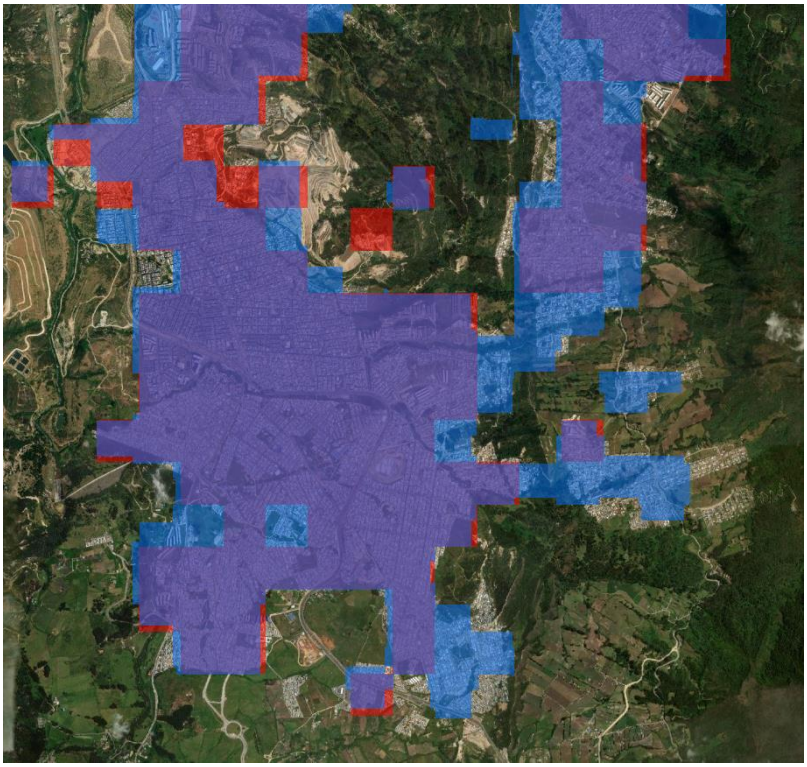
Tras la inspección visual de los resultados se evidenció la dificultad que tienen los modelos para clasificar correctamente las zonas con tejido mixto, debido a las dimensiones de las imágenes de entrenamiento de 32 píxeles cuadrados que equivalen en terreno a 320 metros cuadrados. El proceso de predicción funciona igual que en la etapa de entrenamiento, es decir tomando recortes de 32 píxeles de alto y ancho del área de interés, clasificando entonces 320 metros cuadrados de terreno en cada paso, la cual es un área suficiente para albergar ambas clases.

Una solución encontrada para mejorar el resultado fue la de mover ligeramente el recorte a predecir en varias ocasiones y posteriormente a nivel de píxel realizar un sistema de votación, clasificando el píxel como la mayoría de las predicciones lo defina. De esta manera la clasificación mínima ya no es de 320 metros cuadrados en todos los casos y varía de acuerdo

con las diferentes predicciones. Este enfoque de redundancia de datos es particularmente útil en las zonas limítrofes de áreas urbanas y rurales, ya que minimiza el tamaño de mínimo clasificable favoreciendo la heterogeneidad. Además, esta redundancia de datos ayuda a minimizar los errores de clasificar un recorte como urbano en zona no urbana.

La Imagen 17 presenta en dos capas un ejemplo del resultado de la aplicación del sistema de votación, mostrando en color rojo una sola predicción y en color azul el resultado del refinamiento.

Imagen 17. Refinamiento de predicción por votación.



Nota. En color rojo predicción inicial, en color azul predicción refinada por votación. (Contains modified Copernicus Service information [2017] for Copernicus Service Information).

De la imagen anterior, se puede apreciar como áreas en azul no fueron clasificadas como urbanas en la predicción inicial, lo cual implica una pérdida de información. En contraparte, las áreas en color rojo muestran una mala clasificación de algunas áreas circundantes o muy cercanas a áreas urbanas, lo cual se corrigió con la votación. Así mismo, las áreas en color purpura, son áreas en las cuales coinciden la predicción inicial y el proceso de votación. Por

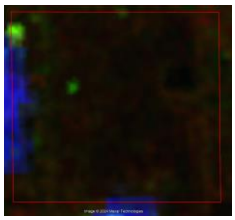
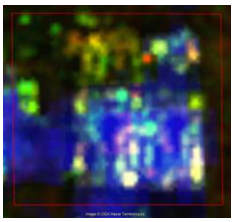
último, aunque el resultado de votación también muestra algunas imprecisiones, se aprecia que es mucho más acertado que la predicción inicial y por lo tanto conviene hacer uso de este para el resultado final.

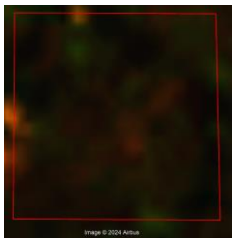
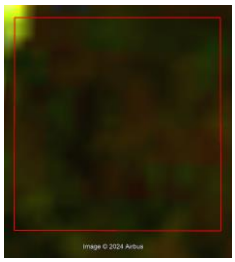
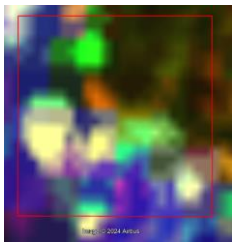
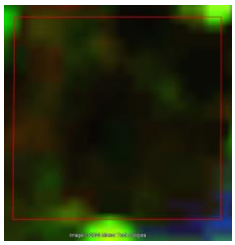
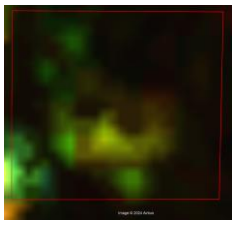
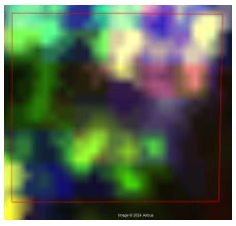
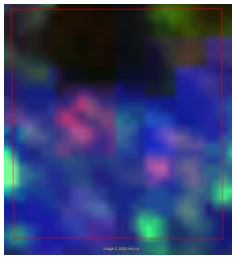
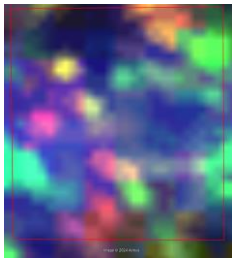
3.5. Comparación temporal de la clasificación de CNN entrenada

Según los resultados analizados, se decidió utilizar el modelo CNN entrenado, ya que para clasificar a urbano y no urbano usando imágenes de RADAR, las métricas y la inspección visual mostraron mejores resultados.

El paso siguiente fue la inspección temporal de la clasificación del total del área de interés (área metropolitana de Bogotá) de las imágenes 2017 y 2023. Para tal fin, se utilizaron las imágenes de Google Earth, el cual ofrece la posibilidad de ver imágenes de diferentes años a muy alta resolución, lo que permitió constatar casos de densificación y construcciones nuevas (Tabla 6), y en algunos casos entender la clasificación obtenida debido a la perdida de cobertura arbórea en zonas urbanas, lo que permite una mejor interacción de la señal de RADAR con las construcciones.

Tabla 6. Ejemplos de cambio de clasificación entre años 2017 y 2023

	Clasificado No urbano imagen 2017	Imagen RADAR 2017	Clasificado Urbano imagen 2023	Imagen RADAR 2023
1				

	Clasificado No urbano imagen 2017	Imagen RADAR 2017	Clasificado Urbano imagen 2023	Imagen RADAR 2023
2				
3				
4				
5				
6				

Nota. Comparación temporal de la aplicación de CNN entrenada del área metropolitana de Bogotá entre los años 2017 y 2023. Adaptado de *Google Earth*. (Google, s.f.).

En la anterior tabla se aprecian dos características que permiten clasificar como urbano en 2023 que en 2017 no. La primera hace referencia a la aparición de construcciones de gran tamaño que modifican la apariencia rural como en los ejemplos del 1 al 4, las cuales muestran nuevos edificios y galpones para 2023. La segunda característica corresponde a los procesos de densificación, tales como se aprecian en los ejemplos del 5 y 6, mostrando mayor cantidad de construcciones para 2023 que en 2017, lo que favorece a una respuesta más fuerte en las imágenes de RADAR y por tanto mayor similitud con las características que el clasificador considera urbano.

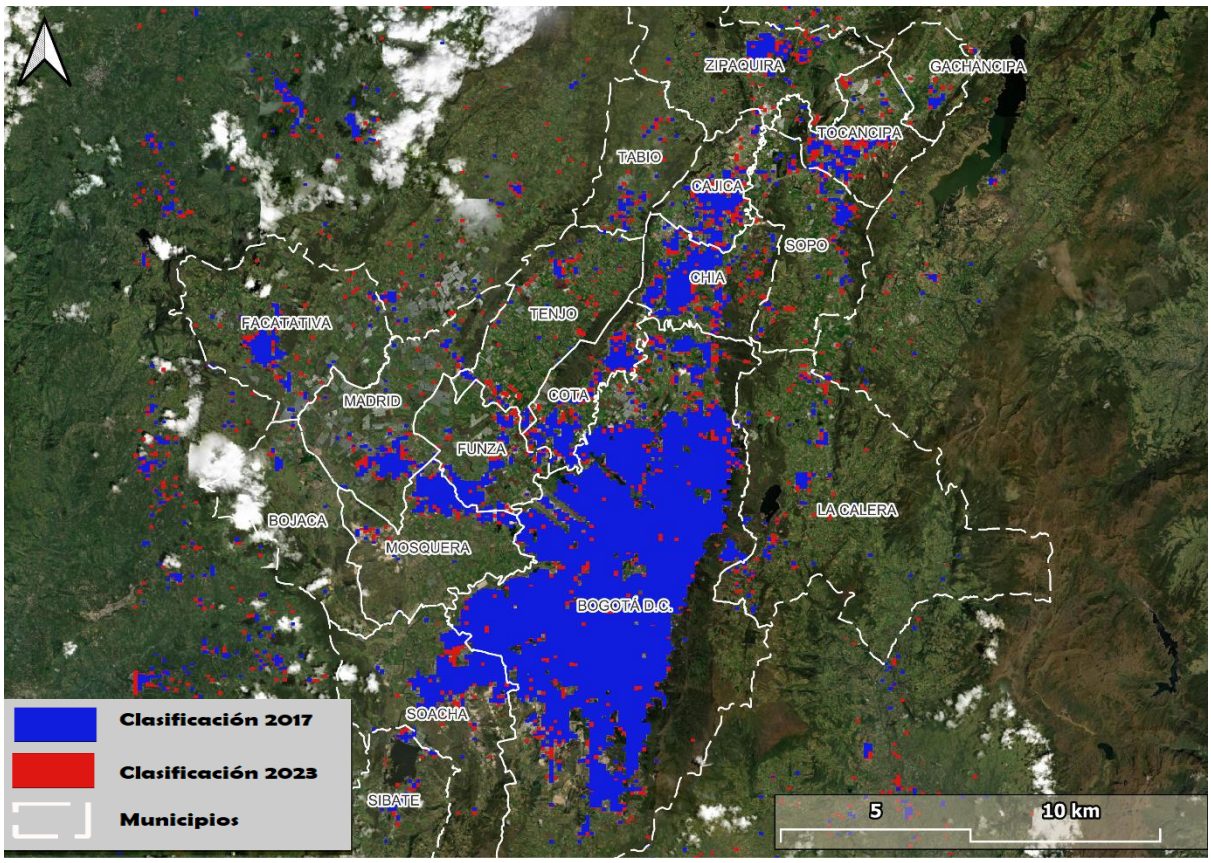
3.6. Cuantificación de áreas Bogotá

Por último, se cuantificó la cantidad de área clasificada con tejido urbano para los años 2017 y 2023, para el área metropolitana de Bogotá, la cual cuenta con 17 municipalidades, encontrando un incremento en el tejido urbano en toda el área, lo cual era esperado dado las altas cifras de desplazamientos rurales hacía la capital, ya sea por falta de oportunidades en otras regiones o por desplazamientos forzados producto del conflicto armado del país, lo que ha llevado a un aumento de población para el área metropolitana y por consiguiente un aumento en la demanda de vivienda. Sin embargo, no todas las municipalidades muestran la misma tasa de crecimiento. Se encontró que la ciudad de Bogotá D.C., es una de las de menor crecimiento, algo que no sorprende dado que las áreas rurales existentes son en la mayoría de las zonas de reserva o conservación natural, lo que desplaza las urbanizaciones a los municipios aledaños que cuentan con una vía de conexión a Bogotá D.C., de mayor capacidad y de tránsito pesado, como es el caso de las salidas sur, norte y dos occidentales. Sin embargo, las municipalidades que más presentaron aumento son aquellas que por tradición han sido rurales, las cuales cuentan con un pequeño centro histórico y actividad económica fuerte en agricultura y ganadería, por lo que los procesos de urbanización y densificación son más evidentes y fáciles de cuantificar en estas municipalidades.

Para realizar una comparación visual e identificar geográficamente donde se encontraron los cambios, se utilizó el software especializado en manejo de datos geográficos QGIS, el cual es Open Source y se puede obtener mediante descarga directa de su sitio web oficial (https://qgis.org/download/). La comparación visual se presenta en la Imagen 18, donde se

visualiza la clasificación obtenida para el 2017 en color azul sobrepuesta a la del 2023 en color rojo, de esta manera se destacando las áreas clasificadas como urbano para el 2023 que en 2017 no fueron clasificadas como tal. Adicionalmente, se agregó la información de contornos de los municipios que conforman el área metropolitana de Bogotá, datos que son de acceso público a través del portal de datos abiertos del Instituto Geográfico Agustín Codazzi (http://geoportal.igac.gov.co), y una imagen base del servicio de imágenes Tiles provisto gratuitamente por Bing imágenes para mapas Web (http://ecn.t3.tiles.virtualearth.net/tiles/a{q}.jpeg?g=0&dir=dir_n').

Imagen 18. Comparación áreas clasificadas con tejido urbano área metropolitana de Bogotá, años 2017 y 2023.



Nota. Adaptado de *(Sources: Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community)*.

Tabla 7. Diferencia porcentual en área con tejido urbano, años 2017 y 2023.

Municipio	2017	2023	% Diferencia	Municipio	2017	2023	% Diferencia
TENJO	4	7,68	91,94%	MADRID	9,3	11,86	27,42%
TOCANCIPÁ	8,27	13,37	61,75%	CHÍA	20,71	25,97	25,43%
LA CALERA	6,07	9,49	56,38%	COTA	14,19	17,48	23,19%
GACHANCIPÁ	1,33	2,06	55,46%	FUNZA	10,25	12,31	20,10%
SOPÓ	5,91	8,84	49,60%	MOSQUERA	12,99	14,56	12,03%
TABIO	3,08	4,58	48,69%	SOACHA	24,12	26,71	10,76%
SIBATE	2,73	4,01	47,22%	BOGOTÁ D.C.	332,79	345,28	3,75%
ZIPAQUIRÁ	7,51	10,87	44,78%	BOJACÁ	0,95	0,98	3,08%
CAJICÁ	12,12	16,79	38,56%	Total	484,62	544,11	12,28%

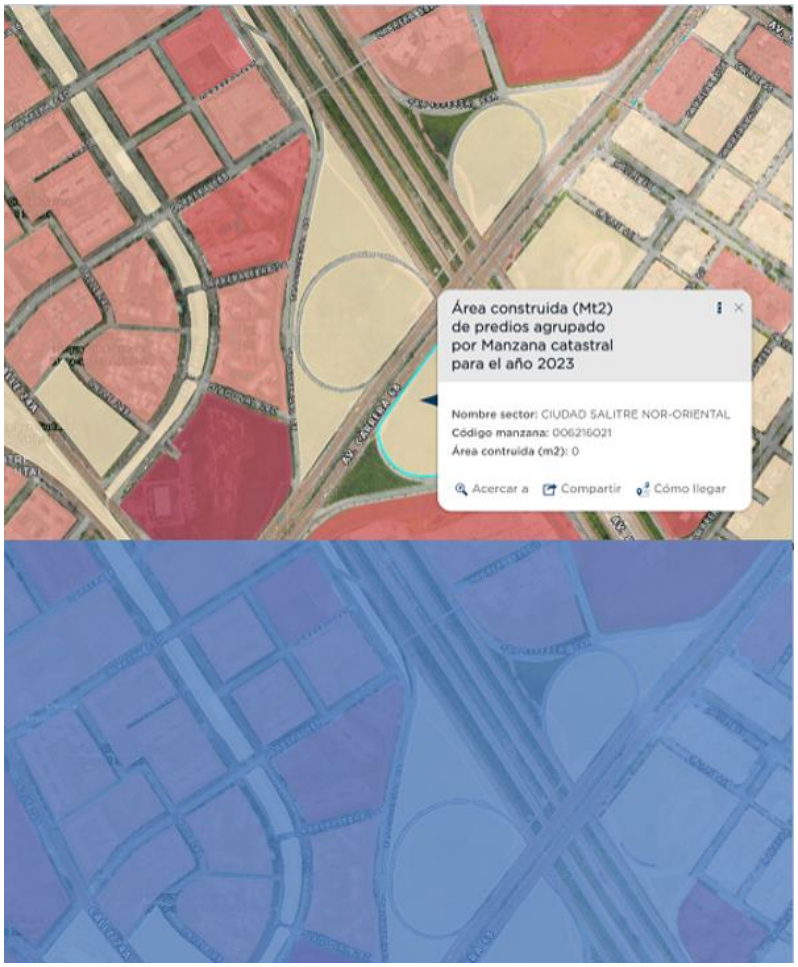
Nota. Valores expresados en km².

Los datos obtenidos de 2023 fueron comparados con los reportados por catastro Bogotá, los cuales indican un área clasificada como urbana de 299 km², es decir un 15% menos que lo que el clasificador determinó como urbano (345.28 km²). Una explicación para tal diferencia puede deberse a la forma en que se clasifica el territorio, ya que catastro Bogotá tiene en cuenta solo las áreas de predios urbanizados, es decir a nivel predio a predio, no incluyendo áreas comunes como parques, vías, plazas, puentes, entre otros, que a diferencia de la clasificación generada por el modelo CNN entrenado, esta sí las incluye, ya que hacen parte de un entorno urbanizado siendo esta la capacidad del clasificador y no la de desagregar predio a predio.

En la parte superior de la Imagen 19 se presenta la información suministrada por catastro Bogotá, mostrando un polígono del área entre vías y puentes vehiculares, siendo esta un área sin ningún tipo construcción. Por otro lado, la imagen inferior es el resultado de la clasificación como urbano para el 2023, en la que se incluyen vías, puentes, y áreas como área urbana, lo que produce inconsistencia en la cantidad de área clasificada como construida entre las dos

fuentes. Sin embargo, aunque el propósito de ambas cuantificaciones es diferente, es un buen punto de referencia para entender los resultados del clasificador.

Imagen 19. Comparación de resultados con catastro Bogotá.



Nota. En la parte superior áreas en zona urbana sin construcción, en la parte inferior áreas identificadas como urbanas por el modelo CNN entrenado. Adaptado de *Mapas Bogotá* [Mapa], por IDECA, s.f., IDECA, https://mapas.bogota.gov.co/

4. Conclusiones

El procedimiento realizado en este trabajo representa una potencial herramienta en la identificación y cuantificación de áreas con características urbanas en regiones que cuenten con alta nubosidad, sirviéndose de imágenes satelitales RADAR SAR Sentinel 1 y redes neuronales

convolucionales. Representa también una alternativa económica para el monitoreo de zonas de protección o de riesgo, así como para ayudar al control de procesos de urbanización, lo cual puede significar un accionar rápido y planificado en materia de prevención y disminución de los costos asociados a proveer los servicios básicos en zonas no habilitadas o no planificadas, eventuales procesos de reubicación y reparación de la población. Por lo anterior, en este apartado se presenta las ventajas y desventajas del procedimiento, la coherencia de su aplicación y algunas recomendaciones para trabajos posteriores.

4.1. Ventajas y desventajas del clasificador

Teniendo en cuenta que las imágenes satelitales Sentinel 1 son de acceso libre, y que gracias a sus características de adquisición se puede acceder a una nueva imagen de la misma escena cada 12 días, se posibilita obtener una clasificación recurrente habilitando el monitoreo en el tiempo, brindando a las organizaciones públicas o privadas interesadas en administrar o gestionar el territorio una herramienta de bajo costo e intuitiva para el diagnóstico y monitoreo. Sin embargo, debido a su resolución espacial, las imágenes obtenidas para el entrenamiento tuvieron que ser recortes de 32 píxeles para lograr el mejor desempeño, esto quiere decir que el área más pequeña clasificable es de 320 metros cuadrados, ya que una dimensión de píxel mayor abarcaría áreas muy extensas y generalizaría en zonas limítrofes urbano-rurales. En contrapeso una dimensión de píxel más pequeña se queda con muy pocos píxeles para la extracción de características geométricas, perdiendo así la capacidad de extraer características en contexto urbano. Teniendo en cuenta que el área mínima clasificable fue de 320 metros cuadrados, se encontró un reto adicional ya que, los pequeños extractos de imagen a clasificar pueden diferir en su distribución interna, en algunos casos mayor porcentaje de áreas sin urbanizar y en otros lo opuesto, por lo que, al clasificar dos extractos de imagen ligeramente trasladados entre sí, puede significar una clasificación diferente en áreas donde la cobertura sea mixta. Por lo tanto, las áreas con más baja densidad urbana no gozan igualmente de la precisión que áreas completamente urbanizadas o no urbanas. Para mitigar este efecto, se optó por realizar varias clasificaciones del mismo periodo, pero ingresando dimensiones diferentes de imagen de entrada, de manera de tener un sistema de votación en la clasificación. Con este sistema de votación, se encontró una disminución de clasificación de áreas mayormente no urbanas que fueron clasificadas con urbanas, mejorando la precisión en la cuantificación.

Por otra parte, se apreció que el clasificador es más sensible a los procesos de densificación urbana, los cuales se evidenciaron en las periferias de los centros históricos de las municipalidades, en los cuales ya existían algunas edificaciones, pero con el paso del tiempo se incrementó la densidad urbana. Además, el clasificador presentó un buen comportamiento en la distinción de respuestas urbanas y orografía, pero perdiendo precisión para clasificar tejido urbano en zonas con mucho relieve, esto debido a que las correcciones geométricas de las imágenes generan una pérdida de información y unas pequeñas porciones de imagen sin información útil para el clasificador.

De esta manera se evidenció que el uso de redes neuronales convolucionales en combinación con imágenes de RADAR SAR Sentinel 1, muestran un buen desempeño para tareas de clasificación urbana, aunque cabe destacar el trabajo previo requerido en tareas de procesamiento de las imágenes, parte indispensable para la comparación entre estas, permitiendo interoperarlas y crear nuevas características que sirven de entrada al clasificador.

4.2. Coherencia en su aplicación

La aplicación de la metodología en el área metropolitana de Bogotá mostró para el área urbana de la ciudad de Bogotá un aumento de 3.75% en 2023 con respecto al año 2017, dejando ver como los corredores viales de conexión a la ciudad de Bogotá D.C., mantienen una dinámica más activa en los procesos de urbanización, en los cuales se ubican industrias, zonas de almacenamiento de mercadería y centros de distribución. Aunque, por otra parte, se evidenció la densificación en las zonas de periferia de áreas urbanas o centro histórico de los diferentes municipios del área metropolitana, zonas usualmente definidas como zonas de expansión urbana que cuentan con los servicios básicos, donde se transforman las áreas de rurales en áreas urbanizadas y perdiendo las capacidades agropecuarias. Así, los municipios aledaños a la ciudad de Bogotá mostraron incrementos porcentuales más altos en áreas clasificadas como urbanas para el periodo de estudio, sin embargo, nominalmente las cantidades son mucho menores ya que son municipios pequeños en comparación. Además, se constató con la información ofrecida por catastro oficial de la ciudad de Bogotá, que la cantidad de área urbana encontrada por el clasificador es coherente con esta, con diferencias debidas en mayor medida por la diferencia entre el objeto de cada fuente de datos, ya que catastro no reporta las áreas

comunes que hacen parte del entorno urbano como plazas y parques, al igual que las obras de infraestructura necesarias para el funcionamiento de la ciudad como vías y puentes, mientras que el clasificador sí engloba estos elementos que pertenecen al entorno urbano y los clasifica como tal. Por lo cual, el procedimiento de clasificación desarrollado ofrece una perspectiva acerca de la cuantificación y distribución de las áreas ocupadas por el tejido urbano, pudiendo servir en la identificación de los procesos de urbanización no autorizados o en zonas de restricción urbanística.

4.3. Recomendaciones para estudios posteriores

Una de las limitaciones del modelo radica en la resolución espacial de la imagen de RADAR SAR utilizada, ya que cada píxel representa 10 metros cuadrados en terreno, las áreas a clasificar pueden incluir características urbanas y no urbanas, y el clasificador debe asignar algunas de las dos categorías. Por lo que al usarse una imagen de RADAR SAR con resolución espacial menor, habría menor lugar para la ambigüedad que enfrenta el clasificador, mejorando el resultado final. Sin embargo, la elección de imágenes Sentinel 1 para este trabajo se debe a que son de acceso libre y gratuito, a diferencia de las imágenes con mayor resolución que son de pago, lo que significa contar con un presupuesto para ello.

Otra alternativa que podría ayudar a resolver la ambigüedad es la de combinar la clasificación con otros métodos basados en píxeles, de manera de obtener un consenso entre los dos métodos, aprovechando los patrones geométricos característicos de las zonas urbanas identificados por las redes neuronales y utilizando la salida de la arquitectura propuesta para limitar las áreas que realmente contienen información relacionado a lo urbano, tratando de obtener un área de clasificación más pequeña que facilite cuantificar a más detalle las áreas urbanizadas.

5. Referencias bibliográficas

Adams, J. Qiu, Y. Posadas, L. Eskridge, K. & Graef, G. (2021). Phenotypic trait extraction of soybean plants using deep convolutional neural networks with transfer learning. Big Data and Information Analytics, Volumen 6, 26-40. https://doi.org/10.3934/bdia.2021003

Albert, A., Kaur, J., & Gonzalez, M. (2017). Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. Massachusetts Institute of Technology, y Philips Lighting Research. https://doi.org/10.48550/arXiv.1704.02965

Alsabhan, W., & Alotaiby, T. (2022). Automatic Building Extraction on Satellite Images Using Unet and ResNet50. Computational Intelligence and Neuroscience, Volumen 2022. Artículo 5008854. https://doi.org/10.1155/2022/5008854

Asobancaria. (2022). Banca y Economía, edición 1317, Información en la vivienda y propuestas de solución. https://www.asobancaria.com/wp-content/uploads/2022/02/1317_BE.pdf

Boldrini, P., & Malizia, M. (2017). Urbanizaciones informales en ciudades intermedias. Los casos del Gran San Salvador de Jujuy y Gran San Miguel de Tucumán (Noroeste argentino). Volumen 22, Número 22, 085-106. ISNN-1666-6186

Bolter, R. (2000). Detection and reconstruction of buildings from multiple view interferometric SAR Data. IGARSS 2000, 749-751. https://doi.org/10.1109/IGARSS.2000.861691

Castelluccio, M., Poggi, G., Sansone, C., & Verdoliva, L. (2015). Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. https://doi.org/10.48550/arXiv.1508.00092

Chan, Y.K., & Koo, V.C. (2008). An introduction to synthetic aperture RADAR (SAR). Progress In Electromagnetics Research B, Volumen 2, 27–60. https://doi.org/10.1109/MSPEC.1969.5213674

Dell'Acqua, F. (2006). Discriminating urban environments using multi-scale texture and multiple SAR images. International Journal of Remote Sensing, Volumen 27, Número 18, 3797-3812. https://doi.org/10.1080/01431160600557572

Elngar, A., Arafa, M., Fathy, A., Moustafa, B., Mahmoud, O., Shaban, M., Fawzy, N. (2021). Image Classification Based On CNN: A Survey. Journal of Cybersecurity and Information Management, Volumen 6, 18-50. 10.5281/zenodo.4897990

ESA. (2007). InSAR Principles: Guidelines for SAR Interferometry Processing and Interpretation (ESA TM-19). ESA Publications.

European Commision (JRC) & Institute for the Protection and Security of the Citizen (IPSC). (2009). RADAR Earth Observation Imgaery for Urban Area Characterisation. https://publications.jrc.ec.europa.eu/repository/bitstream/JRC50451/st_report_sar_urban_final.pdf

Freire, M. (2006). Urban planning: Challenges in developing countries. International congress on Human Development Madrid.

Goodfellow, I. Bengio, Y. & Courville, A. (2016). Deep Learning (Adaptative computation and Machine Learning series). The MIT press, ISBN 9780262035613

Google. (s.f.). Imagen de Google Earth Bogotá año 2017 y 2023.

Satélite Simón Bolívar. (2020). Salto Ángel, Venezuela. [Imagen satelital]. Google Earth. https://www.googleearth.com/AquíVaElEnlaceDirectoALaImagen

Hinton, E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. Improving neural networks by preventing co-adaptation of feature detectors. https://arxiv.org/abs/1207.0580v1

IBM. (s.f.). ¿Qué son las redes neuronales?. https://www.ibm.com/mx-es/topics/neural-networks

IDECA, (s.f.). Área construida (Mt2) de predios agrupado por Manzana catastral para el año 2023 (bogota.gov.co). https://mapas.bogota.gov.co/#

Lincoln Institute of Land Policy. (2011). Regularización de asentamientos informales en América Latina. Lincoln Institute of Land Policy. https://www.lincolninst.edu/sites/default/files/pubfiles/regularizacion-asentamientos-informales-full_0.pdf

Maxwell, A., Warner, T., & Guillén, L. (2021). Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 2: Recommendations and Best Practices. Remote sensing, Volumen 13, Artículo 2591. https://doi.org/10.3390/rs13132591

Md Tazmul, I., & Qingmin, M. (2022). An exploratory study of Sentinel-1 SAR for rapid urban flood mapping on Google Earth Engine. International Journal of Applied Earth Observation and Geoinformation, Volumen 113, Artículo 103002. https://doi.org/10.1016/j.jag.2022.103002

Mohammed, A., & Kora, R. (2023). A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University - Computer and Information Sciences, Volumen 35-2, 757–774. https://doi.org/10.1016/j.jksuci.2023.01.014

Nwankpa, C., Ijomah, W., Gachagan. A., & Marshall, S. (2018). Activation Functions: Comparison of Trends in Practice and Research for Deep Learning. https://doi.org/10.48550/arXiv.1811.03378

Pandya, B., & Gajjar, D.N. (2014). Parallelization of Synthetic Aperture RADAR (SAR) Imaging Algorithms on GPU. IJCSC, Volumen 5, Número 1, 143-146. ISSN-0973-7391

Parag, M., Lottering, R., Peerbhay, K., Agjee, N., & Poona, N. (2024). The use of synthetic aperture RADAR technology for crop biomass monitoring: A systematic review, Remote Sensing Applications: Society and Environment, Volumen 33, Artículo 101107. https://doi.org/10.1016/j.rsase.2023.101107

Provost, F. & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data. Volumen 1-1, 51-59. http://doi.org/10.1089/big.2013.1508

Sandoval, V., & Sarmiento, J. (2019). A neglected issue: Informal settlements, urban development, and disaster risk reduction in Latin America and the Caribbean. Disaster Prevention and Management, Volumen 29-5, 731-745. https://doi.org/10.1108/DPM-04-2020-0115

SATNow. (2022). What are Polar Orbits?. SatNow. https://www.satnow.com/community/what-are-polar-orbits

Shabbir, A., Ali, N., Ahmed, J., Zafar, B., Rasheed, A., Sajid, M., Ahmed, A., & Hanif Dar, S. (2021). Satellite and Scene Image Classification Based on Transfer Learning and Fine Tuning of ResNet50. Mathematical Problems in Engineering, Volumen 2021, 1-18. https://doi.org/10.1155/2021/5843816

Shah, S., Qadri, S., Bibi, H., Waqas Shah, S., Sharif, M., & Marinello, F. (2023). Comparing Inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A Case Study on Early Detection of a

Rice Disease. Agronomy, Volumen 13-6, Artículo 1633. https://doi.org/10.3390/agronomy13061633

Thaens, R., & Coman, C. (2018). A Deep Learning SAR Target Classification Experiment on MSTAR Dataset. 19th International RADAR Symposium (IRS), 1-6. https://doi.org/10.23919/IRS.2018.8448048

Ushida, S. (2012). Image procesing and recognition for biological images. Development, Growth & Differentiation, Volumen 55, 523-549, DOI: 10.1111/dgd.12054

Vali, A., Comail, S., & Matteucci, M. (2021). Deep learning for use and land cover classification based on hyperspectral and multispectral earth observation data: A review. Remote sensing, Volumen 12, Número 15, Artículo 2495. https://doi.org/10.3390/rs12152495

Vosta, S., & Yow, K-C. (2022). A CNN-RNN Combined structure for real-word violence detection in survillance cameras. Applied sciences, Volumen 12, Número 3, Artículo 1021. https://doi.org/10.3390/app12031021

Wellmann, T., Lausch, A., Andersson, E., Knapp, S., Cortinovis, C., Jache, J., Scheuer, S., Kremer, P., Mascarenhas, A., Kraemer, R., Haase, A., Schug, F., & Haase, D. (2020). Remote sensing in urban planning: Contributions towards ecologically sound policies?. Landscape and Urban Planning, Volumen 204, Artículo 103921. https://doi.org/10.1016/j.landurbplan.2020.103921

Xie, Y., Dai, W., Hu, Z., Liu, Y., Li, C., & Pu, X. (2019). A Novel Convolutional Neural Network Architecture for SAR Target Recognition. Advanced Sensor Technologies in Geospatial Sciences and Engineering, Volumen 2019, Artículo 1246548. https://doi.org/10.1155/2019/1246548

Zhao, J., Zhang, Z., Yao, W., Datcu, M., Xiong, H., & Yu, W. (2020). OpenSARUrban: A Sentinel-1 SAR Image Dataset for Urban Interpretation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volumen 13, 187-203. https://doi.org/10.1109/JSTARS.2019.2954850