

UNIVERSIDAD POLITÉCNICA DE
TULANCINGO



MAESTRÍA EN COMPUTACIÓN ÓPTICA

Cuantificación de los niveles de urbanización en
ciudades capitales a partir del análisis de imágenes
multiespectrales del satélite WorldView-2

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Maestro en Computación Óptica

PRESENTA:

Aldo Alberto Aguilar Vallejo

ASESORES

Dra. Carina Toxqui Quitl

Dr. Raúl Castro Ortega

Dedicatoria

A mis padres,
por su amor incondicional, por enseñarme el valor del esfuerzo y por ser siempre mi mayor fuente de inspiración.

A mi familia,
por su comprensión, paciencia y apoyo constante durante este camino académico.

A quienes han creído en mí,
y me han alentado a continuar aprendiendo, aun en los momentos de mayor dificultad.

A la memoria de todos aquellos que sembraron en mí la curiosidad por entender el mundo a través de la ciencia.

Agradecimientos

Deseo expresar mi más sincero agradecimiento a todas las personas e instituciones que hicieron posible la realización de este trabajo.

A la Dra. Carina Toxqui Quitl, por su valiosa asesoría, acompañamiento y confianza durante todo el desarrollo de esta investigación. Su orientación académica y humana ha sido fundamental para el logro de este proyecto.

Al Dr. Raúl Castro Ortega, por sus aportes técnicos y su constante disposición para compartir su conocimiento y experiencia en el área de redes neuronales.

Al Dr. Alfonso Padilla Vivanco y al Dr. Enrique González Amador, por su apoyo en los procesos de investigación y gestión académica dentro del posgrado.

A la Universidad Politécnica de Tulancingo, por brindar las condiciones necesarias para el desarrollo de este trabajo, y al Consejo de Ciencia, Tecnología e Innovación de Hidalgo (CITNOVA) y al SECIHTI, por el apoyo económico otorgado a través de sus programas de becas.

A la PhD Anna Carbone, por su guía y colaboración durante mi estancia en el Politecnico di Torino, la cual representó una experiencia académica y personal invaluable.

Finalmente, agradezco a mis compañeros, profesores y amigos del posgrado por su apoyo, sus ideas y por compartir conmigo este camino de aprendizaje y crecimiento profesional.



Prefacio

El monitoreo de la extensión espacial y la densidad de construcción es fundamental para la planificación urbana sostenible, la gestión ambiental y el desarrollo de infraestructura. Las imágenes satelitales de alta resolución se han convertido en una herramienta valiosa para mapear superficies impermeables, permitiendo evaluaciones oportunas de los cambios en la cobertura del suelo en regiones con rápida urbanización.

Entre las diversas técnicas disponibles para la extracción de características urbanas, los métodos basados en índices espectrales siguen siendo ampliamente utilizados debido a su bajo costo computacional, facilidad de implementación y adaptabilidad a diferentes sensores. Estos métodos se basan en combinaciones algebraicas de bandas multiespectrales para resaltar el contraste espectral entre las estructuras artificiales y otros tipos de cobertura terrestre.

Por otro lado, los métodos supervisados han demostrado ser útiles en una amplia gama de aplicaciones de estudios urbanos, incluyendo el monitoreo del cambio urbano, la evaluación del bienestar socioeconómico, la dinámica urbana y la sostenibilidad urbana. Aunado a esto, en los últimos años, las redes neuronales convolucionales (CNN) han emergido como una opción para la segmentación de elementos en imágenes, ya que aprenden de manera automática patrones espaciales y espectrales complejos, diferenciando con alta precisión diversos tipos de cobertura terrestre, incluso en entornos con variaciones significativas de iluminación, textura o escala a partir de imágenes satelitales.

Esta tesis propone dos técnicas para la extracción y cuantificación de construcción a partir de imágenes satelitales: MBEI, un índice que aplica una máscara para aislar las áreas construidas eliminando el suelo desnudo, y Extract-Built-Net, un modelo de segmentación semántica basado en una arquitectura U-Net que extrae la expansión urbana utilizando imágenes multiespectrales de alta resolución del satélite WorldView-2. Las imágenes satelitales contienen diversas coberturas terrestres, incluyendo áreas construidas, suelo desnudo, agua, vegetación y nubes en una Región de Interés (RoI).

La estructura de este trabajo de tesis se compone del Capítulo 1 donde se abordan los antecedentes, introducción, objetivos y un estudio del marco teórico que hay en la literatura. El segundo Capítulo presenta los índices espectrales de construcción usados en el sentido

remoto, en el Capítulo 3 se presenta una propuesta de red neuronal para la identificación de construcción en imágenes satelitales. El Capítulo 4, está dedicado a la presentación de los resultados obtenidos. Las conclusiones, así como trabajo a futuro de esta tesis son mencionados en el Capítulo 5.

Resumen

Se propone un método de segmentación de áreas de cobertura urbana basado en índices de construcción. Lo anterior, a partir del análisis de imágenes multispectrales de alta resolución del satélite WorldView-2. Se analiza un conjunto de imágenes digitales de ciudades con diferentes niveles de urbanización. Cada índice se evalúa en diferentes coberturas del suelo, como vegetación, agua, suelo desnudo y edificación. Su rendimiento se evalúa mediante el Índice de Discriminación Espectral (SDI, spectral discrimination index). Los resultados indican que el índice de extracción de construcción normalizada (NBEI, Normalized Built-up Extraction Index) superó a los demás índices, al separar superficies impermeables de suelo desnudo, quienes son espectralmente similares. Además, se estima la densidad de construcción \mathcal{B} utilizando los mapas derivados de los índices de Extracción de construcción.

Por otro lado, se presenta una red neuronal convolucional (CNN, convolutional neural network) con una arquitectura U-Net modificada para la extracción de áreas de cobertura urbana. El modelo de red muestra resultados significativos en la identificación de superficies construidas, con un rendimiento consistente tanto en ciudades como en contextos espaciales. Nuestros resultados de segmentación se comparan con los proporcionados por el NBEI para la detección de superficies urbanas. Los píxeles de la imagen se clasifican como desarrollo urbano con una precisión superior al 85.11 %.

Finalmente, se propone un nuevo índice, el cual combina la robustez espectral de los índices tradicionales con un proceso de optimización de umbrales adaptativos y una máscara para eliminar las áreas de suelo desnudo, minimizando la confusión entre suelos desnudos y superficies impermeables. Este índice, denominado MBEI (Masked Built-up Extraction Index), mejora la delimitación de áreas urbanas al integrar información multibanda y ajustar dinámicamente los rangos de segmentación. Los experimentos realizados muestran que el MBEI alcanza una precisión superior al 88.32 %, superando el desempeño de los índices convencionales y ofreciendo una herramienta más confiable para el monitoreo del crecimiento urbano y la planificación territorial.

Abstract

A method for the segmentation of urban coverage areas based on built-up indices is proposed. This is achieved through the analysis of high-resolution multispectral images from the WorldView-2 satellite. A set of digital images from cities with different levels of urbanization is analyzed. Each index is evaluated on different land cover types, such as vegetation, water, bare soil, and built-up areas. Its performance is assessed using the Spectral Discrimination Index (SDI). The results indicate that the Normalized Built-up Extraction Index (NBEI) outperformed the other indices in separating impervious surfaces from bare soil, which are spectrally similar. In addition, the building density \mathcal{B} is estimated using maps derived from the built-up extraction indices.

On the other hand, a convolutional neural network (CNN) with a modified U-Net architecture is presented for the extraction of urban coverage areas. The network model shows significant results in identifying built-up surfaces, with consistent performance both in cities and in different spatial contexts. Our segmentation results are compared with those obtained using the NBEI for urban surface detection. The image pixels are classified as urban development with an accuracy higher than 85.11 %.

Finally, a new index is proposed, which combines the spectral robustness of traditional indices with an adaptive threshold optimization process and a masking step to remove bare soil areas, minimizing confusion between bare soil and impervious surfaces. This index, called the Masked Built-up Extraction Index (MBEI), improves the delineation of urban areas by integrating multiband information and dynamically adjusting the segmentation ranges. The experiments conducted show that the MBEI achieves an accuracy higher than 88.32 %, outperforming conventional indices and providing a more reliable tool for monitoring urban growth and territorial planning.

Índice general

1. Antecedentes	1
1.1. Introducción	1
1.2. Sistemas de sensado remoto	2
1.3. Detección de área construida	3
1.4. Planteamiento del Problema	5
1.5. Propuesta de solución	6
1.6. Estado del Arte	6
1.7. Justificación	10
1.8. Objetivos	11
1.9. Aportaciones	12
1.10. Conclusiones	13
Bibliografía	13
2. Segmentación de construcción basada en índices	17
2.1. Introducción	17
2.2. Imágenes multiespectrales de WorldView-2	19
2.3. Preprocesamiento de imágenes WorldView-2	20
2.4. Firma espectral de diferentes superficies	23
2.5. Índices para la Extracción de construcción	24
2.5.1. Built-up Area Index (BAI)	24
2.5.2. Built-up Spectral Index (BSI)	25
2.5.3. New Built-up Extraction Index (NBEI)	26
2.5.4. Red Edge - Green Index (RGI)	27
2.5.5. Masked Built-up Extraction Index (MBEI)	28
2.6. Cuantificación de la densidad de construcción	29
2.7. Conclusiones	31
Bibliografía	31

3. Aprendizaje profundo para segmentación de construcción en imágenes sa-	
telitales	35
3.1. Introducción	35
3.2. Extracción de construcción a partir de imágenes satelitales	36
3.2.1. Base de datos WV-2	36
3.2.2. Preprocesamiento de las imágenes	37
3.3. Arquitectura U-Net	38
3.4. Baseline	41
3.5. Conclusiones	42
Bibliografía	43
4. Resultados de segmentación de superficies urbanas	45
4.1. Introducción	45
4.2. Resultados basados en índices	46
4.3. Resultados basados en Aprendizaje Profundo	49
4.4. Densidad de construcción	50
4.5. Conclusiones	51
5. Conclusiones y Trabajo a Futuro	53
A. Trabajos derivados de la Tesis	55

Índice de figuras

1.1.	Proceso de adquisición de imágenes digitales de sensado remoto.	2
1.2.	Delimitación del área urbana usando el NBEI. Imagen tomada de [22].	8
1.3.	Mapa binarizado de construcción usando el Índice de segmentación de construcción. Imagen tomada de [23]	8
1.4.	Delimitación de carreteras usando el índice de construcción de área. Imagen tomada de [24].	9
1.5.	Segmentación de múltiples superficies usando el índice Red Edge/Green Index. Imagen tomada de [25].	10
2.1.	Imagen en Falso Color de la Ciudad de Helsinki, se pueden observar en tonos verdes áreas construidas, y en tonos rosas áreas de vegetación. Nótese que algunas zonas de suelo desnudo se confunden con construcción.	18
2.2.	Firma espectral de construcción, agua y vegetación	24
2.3.	Histograma acumulado y distribución espacial del BAI en las cuatro escenas urbanas. El histograma muestra que la mayoría de las superficies tienen valores por debajo de cero, separando de manera efectiva el agua, pero sin lograr distinguir completamente la vegetación de las áreas construidas y confundiendo áreas construidas con suelo desnudo y caminos.	25
2.4.	Histograma acumulado y distribución espacial del BSI en las cuatro escenas urbanas. Al igual que el BAI, el histograma muestra que la mayoría de las superficies tienen valores por debajo de cero, aislando de manera efectiva el agua. Ofrece una mejor separación entre vegetación y áreas construidas, pero aún presenta confusión entre áreas construidas, suelo desnudo y caminos.	26
2.5.	Histograma acumulado y distribución espacial del NBEI en las cuatro escenas urbanas. Los valores de áreas construidas se agrupan alrededor de cero, la vegetación aparece con valores más altos y el agua cerca de -1 . Al igual que otros índices, las áreas construidas suelen confundirse con el suelo desnudo.	27
2.6.	Histograma acumulado y distribución espacial del RGI en las cuatro escenas urbanas. Los valores de áreas construidas se agrupan entre -0.2 y 0.4 , la vegetación aparece entre 0.2 y 0.6 , y el agua por debajo de -0.4 . Al igual que otros índices, las áreas construidas suelen confundirse con el suelo desnudo.	28

2.7.	Resultados de extracción de construcción usando el método propuesto de enmascaramiento. Como se puede observar, la información de suelo desnudo fue eliminada. Se calcularon los valores de densidad de construcción por sectores suburbanos $\mathcal{B}_A = 0.13$, $\mathcal{B}_B = 0.11$, $\mathcal{B}_C = 0.63$, y $\mathcal{B}_D = 0.23$	30
3.1.	Ciudad de Berlín, Alemania.	37
3.2.	Arquitectura de la U-Net modificada utilizada para la segmentación urbana binaria.	39
3.3.	Resultados de segmentación utilizando la arquitectura U-Net propuesta.	40
3.4.	Distribución espacial de NBEI una escena urbana.	41
4.1.	Mapas binarios de áreas construidas obtenidos a partir de imágenes multiespectrales correspondientes a las cuatro zonas de estudio: Turín, Roma, Atenas y Bucarest. Los píxeles blancos indican zonas construidas; los píxeles negros indican fondo. De arriba hacia abajo: mapa BSI , mapa BAI , mapa NBEI y mapa RGI	47
4.2.	Comparación cualitativa de cinco muestras del conjunto de datos de Turín. Las columnas muestran: (a) imagen NIR1, (b) índice NBEI, (c) mapa de probabilidad de la red neuronal convolucional (CNN), (d) clasificación binaria de la CNN y (e) clasificación de la CNN superpuesta a la imagen NIR1 (verde: zonas urbanizadas, rojo: fondo).	49
4.3.	Resultados de extracción de construcción usando el método propuesto de enmascaramiento. Como se puede observar, la información de suelo desnudo fue eliminada. Se calcularon los valores de densidad de construcción por sectores suburbanos $\mathcal{B}_A = 0.13$, $\mathcal{B}_B = 0.11$, $\mathcal{B}_C = 0.63$, y $\mathcal{B}_D = 0.23$	50

Índice de tablas

2.1. Bandas espectrales del satélite WorldView-2.	19
2.2. Factores de calibración absoluta [12] para WorldView-2	21
2.3. Irradiancia exoatmosférica media (E_{sun}) según Thuillier 2013 [12]	22
4.1. Precisión general (%) e Índice de Discriminación Espectral (SDI) para cada índice espectral en todas las zonas de estudio.	48

Capítulo 1

Antecedentes

1.1. Introducción

El crecimiento y evolución espacial de las ciudades constituyen uno de los fenómenos centrales para la planificación territorial, la gestión ambiental, así como la provisión de los servicios públicos necesarios para la zona urbana. Llevar a cabo una cuantificación precisa de los niveles de urbanización de las ciudades requiere de información detallada y actualizada en cuanto a la presencia, extensión y configuración de la infraestructura construida (tales como construcciones, redes carreteras, superficies agrícolas, áreas de suelo desnudo, entre otras).

En este ámbito, el sensado remoto con sensores ópticos de muy alta resolución espacial y multiespectral, como el satélite WorldView-2 (WV-2), ofrece un valioso recurso. Los sistemas de sensado remoto combinan resolución espacial con ocho bandas espectrales que permiten diferenciar materiales de cubierta y detectar estructuras urbanas con un alto grado de detalle. Estas características han sido aprovechadas en estudios orientados a la clasificación y extracción de información de la construcción, así como de la delimitación de uso del suelo.

El sensado remoto, es la ciencia de obtener información de la superficie de la Tierra sin estar en contacto directo con ella. Esto se logra mediante la detección y el registro de la energía reflejada o emitida desde la superficie, así como el procesamiento, análisis y aplicación de dicha información.

En este trabajo de Tesis, se diseñan, implementan en MATLAB y evalúan algoritmos para la segmentación de coberturas urbanas usando imágenes multiespectrales del satélite World View-2 (WV-2). Se propone un índice enmascarado de extracción de construcción (MBEI) que detecta construcción en imágenes satelitales de ciudades con múltiples superficies como: vegetación, agua, construcción y suelo desnudo. La cobertura urbana incluye áreas de uso residencial, así como industrial y comercial que conforman el paisaje desarrollado o construido.

1.2. Sistemas de sensado remoto

De acuerdo al Centro de Sensado Remoto de Canada [1], el proceso de sensado remoto se ejemplifica en la Figura 1.1, este implica una interacción entre la radiación incidente y los objetos de interés.

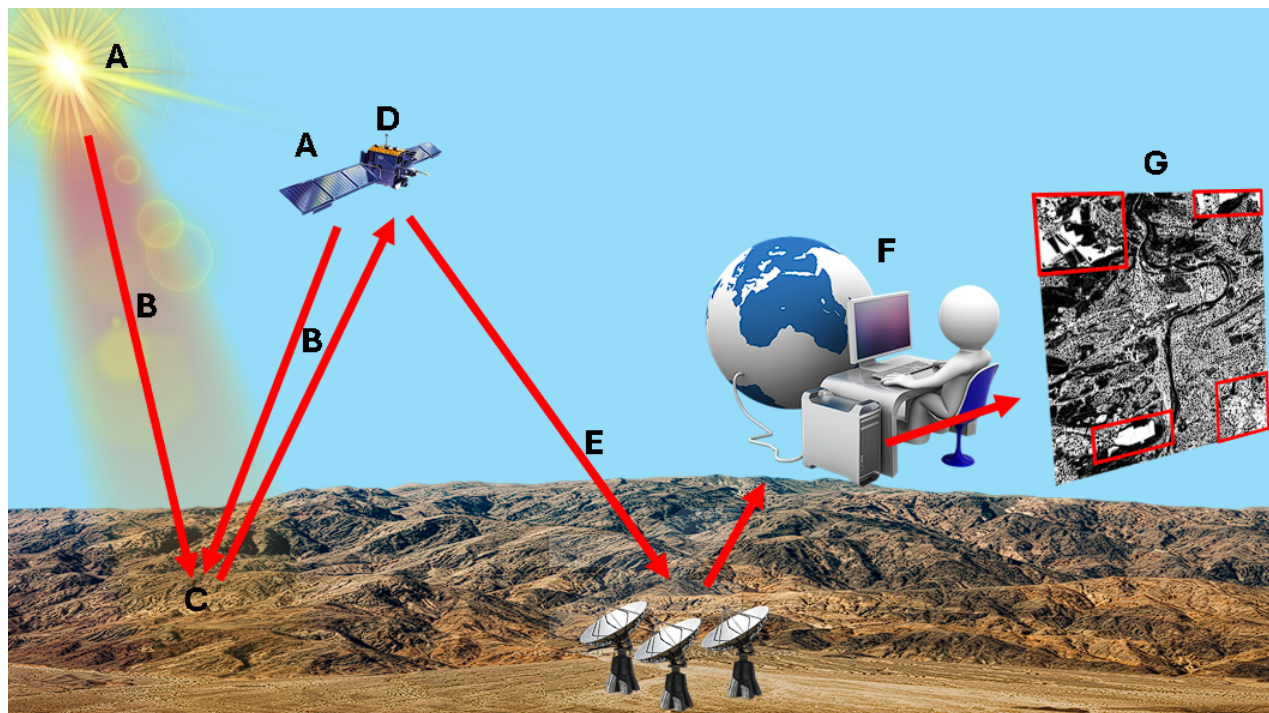


Figura 1.1: Proceso de adquisición de imágenes digitales de sensado remoto.

Esto permite el uso de sistemas de imágenes en los que intervienen los siguientes siete elementos:

1. **Fuente de energía o iluminación (A):** el sensado remoto requiere de una fuente de energía que emita o proporcione radiación electromagnética al objeto de estudio, permitiendo su detección y análisis.
2. **Radiación y la atmósfera (B):** a medida que la energía viaja desde su fuente hacia el objeto, entra en contacto e interactúa con la atmósfera que atraviesa. Esta interacción puede ocurrir nuevamente cuando la energía viaja desde el objeto hacia el sensor.
3. **Interacción con el objeto (C):** una vez que la energía alcanza al objeto a través de la atmósfera, la interacción con él depende de las propiedades y naturaleza tanto del objeto como de la radiación.

4. **Registro de la energía por el sensor (D):** una vez que la energía ha sido reflejada, emitida o dispersada por el objeto, es necesario un sensor remoto (sin contacto directo con la superficie) para captar y registrar la radiación electromagnética. El producto de esta captura son generalmente cubos multiespectrales, conformados por varias bandas espectrales λ_i con ubicación específica dentro del espectro electromagnético. Cada imagen digital se caracteriza por:
 - una resolución espacial, determinada por el tamaño del píxel,
 - una resolución espectral, definida por el número y la anchura de las bandas registradas, y
 - una resolución radiométrica, que corresponde al número de niveles digitales (bits) con los que el sensor discrimina la energía registrada en cada píxel.
5. **Transmisión, recepción y procesamiento (E):** la energía registrada por el sensor se transmite, generalmente en forma electrónica, hacia una estación receptora y de procesamiento. En este lugar, los datos son decodificados, corregidos y transformados en una imagen, la cual puede presentarse en formato digital y/o impreso
6. **Interpretación y análisis (F):** la imagen procesada resultante se interpreta, ya sea de manera visual y/o digital o electrónica, dependiendo de los objetivos y las herramientas disponibles, para extraer información sobre el objeto observado.
7. **Aplicación (G):** el elemento final del proceso de sensado remoto se logra cuando aplicamos la información obtenida a partir de los píxeles de la imagen sobre el objeto, con el fin de comprenderlo mejor, revelar nueva información o ayudar a resolver un problema específico.

1.3. Detección de área construida

Generalmente se considera “urbano” (o el área urbana) tanto el uso de suelo como la cobertura del suelo. Los píxeles “urbanos” forman la base de muchos análisis en el sensado remoto y consisten en píxeles de superficies desarrolladas e impermeables, que incluyen superficies construidas, techos, aceras, carreteras, pistas y estacionamientos, áreas cubiertas por materiales impenetrables como asfalto, concreto y piedra, estos materiales sellan eficazmente las superficies, repelen el agua y evitan que la precipitación y el deshielo se infiltren en los suelos [2].

En este caso, se trata de cobertura del suelo, que con frecuencia se etiqueta como área construida, cobertura construida o cobertura urbana. Sin embargo, considerando el uso de suelo, las “áreas urbanas” también pueden incluir píxeles no desarrollados y con vegetación,

como parques y bosques urbanos, y pueden excluir desarrollos que son componentes de otros usos de suelo [3].

Existen varios enfoques para determinar el límite de una ciudad o pueblo. La delimitación física de las áreas urbanas y aglomeraciones se ha caracterizado por dos enfoques. Por un lado, la delimitación basada en criterios físicos o morfológicos, donde el área continua construida, o la densidad de ámbitos contiguos, constituye el mecanismo básico para la delimitación. Por otro lado, los estudios basados en criterios funcionales o económicos, donde se enfatizan las relaciones y flujos existentes a lo largo del territorio urbanizado, siendo fundamental la relación entre el lugar de residencia y el lugar de trabajo [4]. Entre estos dos, el primero es el más utilizado.

Métodos basados en índices

El *band ratioing* [5] es una técnica ampliamente utilizada en el procesamiento de imágenes satelitales, cuyo principio básico consiste en calcular el cociente entre los valores de reflectancia registrados en dos bandas espectrales diferentes. Esta operación permite atenuar los efectos de variaciones en la iluminación solar, la topografía o las condiciones atmosféricas, ya que dichas influencias tienden a afectar de manera similar a ambas bandas. Al mismo tiempo, el cociente resalta los contrastes relativos en la respuesta espectral, incrementando la capacidad de discriminación entre diferentes coberturas del suelo.

Matemáticamente, si se denota como $R_{\lambda_i}(x, y)$ la reflectancia de un píxel en la posición (x, y) correspondiente a la longitud de onda λ_i , el cociente entre dos bandas λ_i y λ_j se define como:

$$BR(x, y) = \frac{R_{\lambda_i}(x, y)}{R_{\lambda_j}(x, y)} \quad (1.1)$$

donde:

- $R_{\lambda_i}(x, y)$ representa la reflectancia del píxel en la banda i ,
- $R_{\lambda_j}(x, y)$ representa la reflectancia del píxel en la banda j ,
- $BR(x, y)$ es el valor resultante del cociente de bandas en la posición (x, y) .

El resultado de este cociente puede interpretarse como una normalización espectral, en la que se reduce la influencia de factores externos y se favorece el realce de las diferencias intrínsecas en la firma espectral de los objetos. De este modo, el *band ratioing* ha demostrado ser una herramienta fundamental para distinguir entre clases con reflectancias similares en valores absolutos, como suelos desnudos y áreas urbanas.

De manera más general, esta técnica puede extenderse a múltiples combinaciones de bandas, generando un vector de razones espectrales:

$$\mathbf{BR}(x, y) = \left[\frac{R_{\lambda_1}(x, y)}{R_{\lambda_2}(x, y)}, \frac{R_{\lambda_1}(x, y)}{R_{\lambda_3}(x, y)}, \dots, \frac{R_{\lambda_m}(x, y)}{R_{\lambda_n}(x, y)} \right] \quad (1.2)$$

Este vector puede emplearse como insumo en algoritmos de clasificación supervisada y no supervisada, en la formulación de índices espectrales especializados, o como característica de entrada en métodos de segmentación basados en aprendizaje automático. En este sentido, el *band ratioing* no solo constituye una técnica de preprocesamiento, sino también una estrategia analítica robusta que ha dado origen a índices para la identificación de agua, vegetación, suelos desnudos y áreas urbanas.

Métodos basados en aprendizaje profundo

Para solventar estos problemas presentes en los métodos tradicionales, las redes neuronales convolucionales (CNN) y el aprendizaje profundo (DP) han transformado por completo la forma de analizar imágenes satelitales para propósitos de sensado remoto. Cada vez más hay nuevas arquitecturas e implementaciones diseñadas para aprender representaciones espaciales y que pueden “reconocer” patrones tales como texturas, bordes, formas, entre otros. Con esto, se pueden resolver tareas de segmentación semántica con una precisión superior a la de los métodos clásicos. Algunas configuraciones como las arquitecturas del tipo U-Net [6], variantes con codificadores profundos basados en ResNet y los modelos basados en transformadores (transformers) o en atención (attention based), han mejorado la extracción automática de superficies urbanas en imágenes de muy alta resolución.

1.4. Planteamiento del Problema

Los índices espectrales son ampliamente utilizados en sensado remoto para resaltar e identificar características específicas de la superficie terrestre, realizar detección de cambios y monitorear el medio ambiente. Estos índices espectrales son combinaciones matemáticas de valores de reflectancia de diferentes bandas de longitud de onda; al aumentar el contraste entre las respuestas espectrales, simplifican datos complejos multispectrales o hiperespectrales [7].

Los índices espectrales destinados a la detección de áreas construidas suelen enfrentar dificultades para distinguir entre superficies construidas del suelo desnudo o roca expuesta debido a la similitud de sus respuestas espectrales, lo que puede dar lugar a una sobreestimación del área urbanizada [8].

1.5. Propuesta de solución

Delaney et al. (2025) [9] realizan una revisión exhaustiva sobre la detección de suelo desnudo por sensado remoto y advierten que los errores en la identificación se deben en gran medida a la confusión espectral entre suelo y superficies similares, como áreas urbanas o residuos agrícolas. También subrayan la importancia de utilizar estrategias robustas de validación para mejorar la precisión. Valdiviezo-N et al [10]. señalan que este tipo de confusión es un problema común en la aplicación de índices de construcción. En ambientes áridos o semiáridos, el problema se acentúa. Rasul et al. (2018) [11] encontraron que el Índice de Área Construida Normalizado (NDBI) genera valores altos tanto para zonas construidas como para suelo seco, lo que limita su utilidad en dichos entornos climáticos.

Para superar las limitaciones de los índices tradicionales en la discriminación entre superficies construidas y suelo desnudo, se propone la optimización del algoritmo de extracción de áreas urbanas mediante el ajuste dinámico de los rangos de umbral.

La estimación de la densidad de construcción \mathcal{B} se realizó sobre el mapa obtenido $MBEI$ derivado de imágenes multiespectrales de alta resolución. Lo cual, permitió cuantificar de manera más confiable el grado de urbanización en diferentes escenarios. Este mapa contiene todos los píxeles clasificados como superficie urbana, alcanzando una precisión superior al 88.32% a diferencia de otros índices reportados en la literatura, el $MBEI$ minimiza la confusión espectral entre áreas urbanizadas y coberturas con alta reflectancia en bandas similares, como el suelo desnudo o la roca expuesta.

Finalmente, la propuesta metodológica abre la posibilidad de integrar el $MBEI$ con técnicas de aprendizaje automático y redes neuronales convolucionales (CNN), lo que potenciaría su desempeño en escenarios urbanos más complejos. Esta sinergia entre índices espectrales optimizados y modelos de segmentación basados en inteligencia artificial podría convertirse en una herramienta clave para abordar los desafíos contemporáneos en la teledetección urbana.

1.6. Estado del Arte

La delimitación del área urbana morfológica incluye algunos de los primeros trabajos que utilizaron sensores Landsat [12]. Desde entonces, ésto ha sido complementado con enfoques más sofisticados para la identificación de asentamientos. Estos van desde índices simples [13], hasta caracterizaciones regionales [14], y enfoques complejos de fusión de datos que utilizan modelado Bayesiano y campos aleatorios de Markov [15]. Las técnicas de mapeo urbano han progresado aún más hacia la extracción de medidas estadísticas de caracterización. Estas incluyen la obtención de medidas simples de morfología urbana a partir de imágenes satelitales [16], el uso de información de textura [17], y la geoestadística [18] o la estadística espacial [19] para

la caracterización de estructuras y patrones urbanos, y de esta forma la delimitación del área urbana.

Además de los sensores multiespectrales ópticos más utilizados, sensores satelitales alternativos también se usan cada vez más en la delimitación del área urbana morfológica. El poder de discriminación adicional sobre los sensores ópticos que ofrece el radar de imágenes ha llevado a su uso en el mapeo de áreas urbanas. Otro enfoque para mapear la extensión urbana es la interferometría radar satelital, mediante la cual se mide el grado de coherencia entre un par de imágenes de radar [20]. Las imágenes nocturnas también han encontrado un lugar en este campo. Las imágenes nocturnas que muestran luces de ciudades pueden emplearse para estimar el área de los asentamientos [21] ha mostrado un método interesante para determinar la extensión natural de un área urbana a partir de datos de percepción remota, el cual sugiere realizar una búsqueda vecinal para derivar el área urbana contigua a partir de una imagen clasificada binaria.

En el estado del arte, la mayoría de los algoritmos para la segmentación de construcción han sido diseñados para cubos de datos provenientes de los satélites LANDSAT y Sentinel. Nosotros listamos a continuación aquellos que hacen uso de las bandas multiespectrales de imágenes del satélite WV-2.

El Nuevo Índice de Extracción de Áreas Construidas (NBEI, por sus siglas en inglés) fue desarrollado para abordar las limitaciones de los índices espectrales tradicionales cuando se aplican a imágenes de alta resolución como las del WorldView-2. El NBEI mejora el contraste espectral entre las superficies impermeables y las coberturas del suelo circundantes al combinar las bandas más discriminantes identificadas mediante la selección de características ReliefF: NIR2, NIR1, Verde y Borde Rojo. Su formulación se basa en la diferencia normalizada entre la suma de las bandas del infrarrojo cercano (reflectantes) y las bandas del borde del visible (absorbentes), lo que incrementa la separabilidad entre materiales urbanos heterogéneos [22]. En la Figura 1.4 se muestra un ejemplo de segmentación de construcción de imágenes provenientes del satélite WorldView-2.

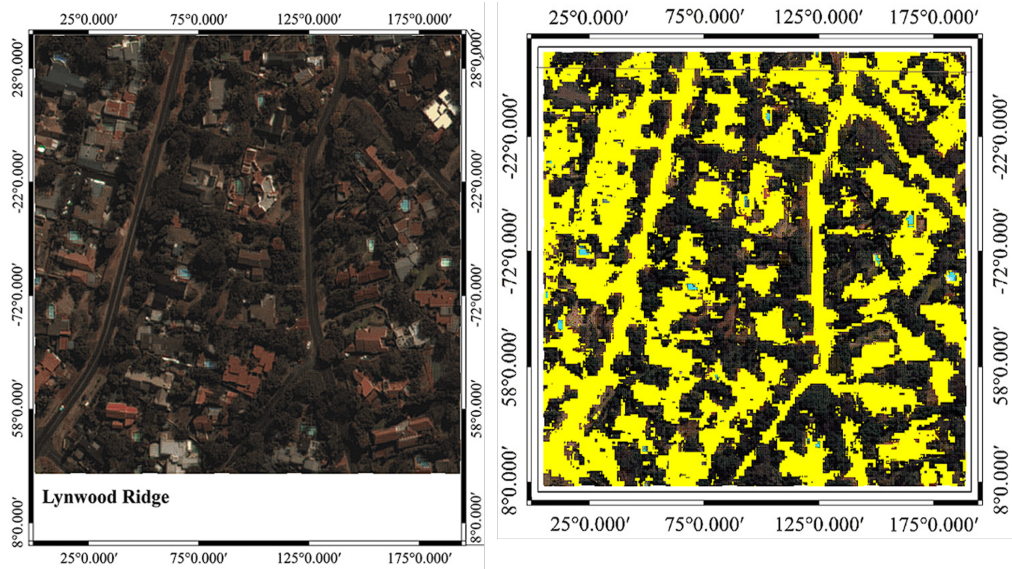


Figura 1.2: Delimitación del área urbana usando el NBEI. Imagen tomada de [22].

El Índice Espectral de Áreas Construidas (BSI, por sus siglas en inglés) fue diseñado para mejorar la extracción de superficies impermeables a partir de datos del satélite WorldView-2. El BSI fue desarrollado utilizando un algoritmo de optimización por enjambre de partículas multiobjetivo. Esta técnica identificó las bandas espectrales más relevantes y sus pesos óptimos para mejorar la separabilidad espectral entre superficies construidas y no construidas [23].

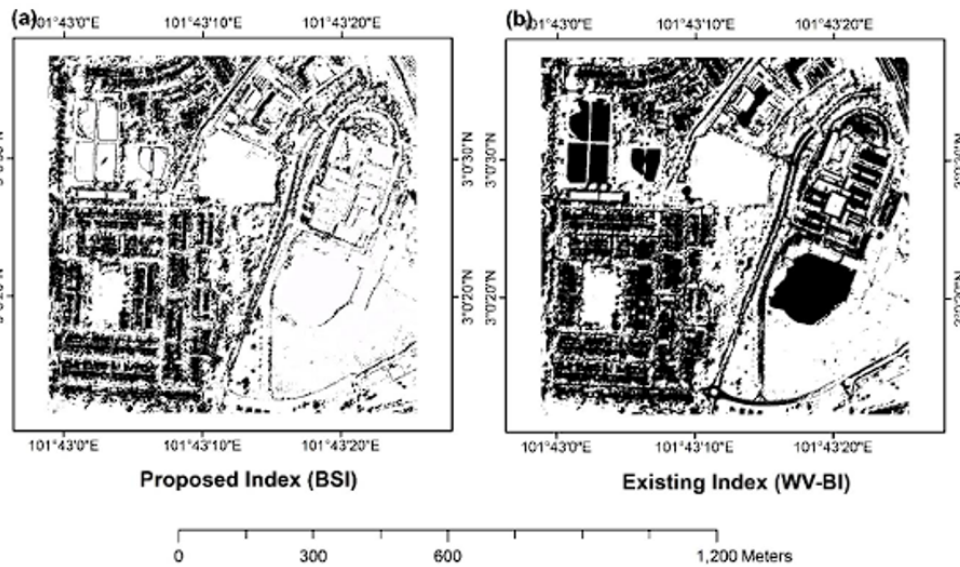


Figura 1.3: Mapa binarizado de construcción usando el Índice de segmentación de construcción. Imagen tomada de [23]

El Índice de Área Construida (BAI, por sus siglas en inglés) es un índice espectral simple, diseñado para mejorar la detección de superficies artificiales, particularmente asfalto y concreto, en imágenes de alta resolución. Hace uso de los valores de reflectancia entre las bandas del infrarrojo cercano y el azul, basado en la reflectancia característica de la mayoría de los materiales construidos. Para la delimitación de caminos y carreteras, el Índice BAI mostró buenos resultados [24]. En la Figura 1.2 se muestra el delineado de las carreteras usando este índice.



Figura 1.4: Delimitación de carreteras usando el índice de construcción de área. Imagen tomada de [24].

El Índice Red Edge-Verde (RGI, por sus siglas en inglés) es un índice espectral diseñado para mejorar la discriminación entre superficies impermeables y suelo desnudo en entornos urbanos y periurbanos. El índice se calcula como la diferencia normalizada entre las bandas del Borde Rojo y Verde [25]. En la Figura 1.3 se muestra un ejemplo de su eficiencia ante múltiples superficies como vegetación, agua y suelo desnudo.

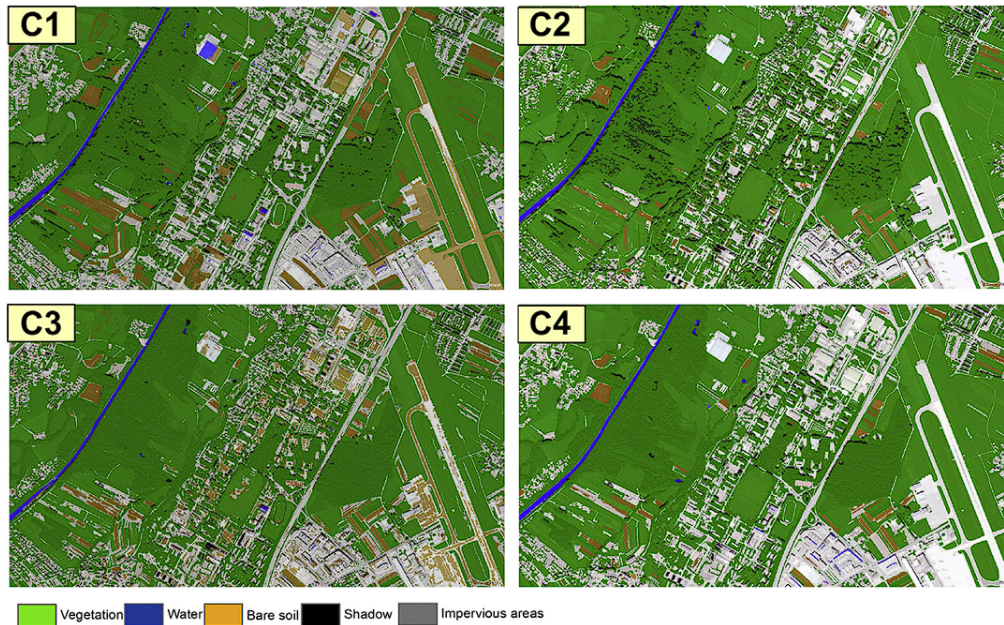


Figura 1.5: Segmentación de múltiples superficies usando el índice Red Edge/Green Index. Imagen tomada de [25].

1.7. Justificación

Ambientalmente, la expansión del suelo urbanizado altera procesos biológicos esenciales y degrada los entornos urbanos. Socialmente, la sobreexpansión puede generar una considerable tasa de desocupación de viviendas, mientras que la sobresaturación demográfica mina la sostenibilidad urbana. Por lo tanto, una comprensión integral de los patrones de expansión urbana y su equilibrio con la población es esencial para el desarrollo regional sostenible.

Las características espacio-temporales son cruciales en la investigación sobre urbanización, ya que la ubicación de los cambios en el suelo urbano refleja la transformación de la estructura espacial urbana, la cual está intrínsecamente relacionada con indicadores clave como la eficiencia y la compacidad del uso del suelo, los cuales afectan el consumo energético urbano con importantes implicaciones para la sostenibilidad. Además, la información espacio-temporal facilita el reconocimiento de los patrones de distribuciones urbana y permite una identificación

más robusta de parámetros urbanos como la densidad de construcción.

1.8. Objetivos

A continuación, se presentan el objetivo general y los objetivos específicos que guían el desarrollo de este trabajo de Tesis.

Objetivo general

Desarrollar un método para la extracción, cuantificación y análisis de los niveles de urbanización mediante el procesamiento de imágenes multiespectrales de alta resolución obtenidas por el satélite WorldView-2.

Objetivos específicos

- Construir una base de datos compuesta por imágenes multiespectrales de 26 ciudades con diferentes grados de urbanización, adquiridas por el satélite WorldView-2.
- Realizar un análisis espectral de superficies representativas (vegetación, agua, suelo desnudo y áreas construidas) a partir de las firmas espectrales derivadas de píxeles en el conjunto de bandas disponibles.
- Diseñar y evaluar índices espectrales orientados a la extracción de áreas construidas.
- Optimizar los parámetros de segmentación automática para mejorar la precisión en la detección de áreas urbanas.
- Implementar y entrenar una red neuronal convolucional basada en la arquitectura U-Net para la delimitación de la cobertura urbana.
- Aplicar técnicas de segmentación automática de áreas construidas en ciudades seleccionadas.
- Calcular métricas de densidad de construcción a partir de las imágenes segmentadas y analizar su relación con los niveles de urbanización.

1.9. Aportaciones

En este proyecto de Tesis se tienen las siguientes aportaciones:

Memorias de Congreso

1. C. Toxqui, A. Delgadillo, **A. Aguilar**, R. Castro Ortega, and A. Padilla. 2025 “*Deep learning in urban studies through multi-spectral satellite imagery*” SPIE Future Sensing Technologies, Yokohama, Japan. **Invited paper**
2. **A. Aguilar**, R. Castro Ortega , A. Delgadillo, C. Toxqui, A. Padilla and A. Carbone. 2025 “*Comparison of built-up indices for urban extraction.*” Proceedings of SPIE 13596, Current Developments in Lens Design and Optical Engineering XXVI. 1359632. (2025).

Participación en Congresos Nacionales

1. Armando Delgadillo Jiménez, Carina Toxqui-Quitl, Raúl Castro-Ortega, **Aldo Aguilar Vallejo**, Alfonso Padilla-Vivanco, Anna Carbone. 2025 “*Estimation of urban fractal dimension using a Convolutional Neural Network*”, XXIV Mexican International Conference on Artificial Intelligence - MICAI. Guanajuato, Mexico. **Ponencia**
2. **Aldo Aguilar Vallejo**, Armando Delgadillo Jiménez, Carina Toxqui-Quitl, Raúl Castro-Ortega, Alfonso Padilla-Vivanco, Anna Carbone. 2025 “*Índice espectral para la delimitación de suelo desnudo usando imágenes de alta resolución*”, LXVIII Congreso Nacional de Física. Centro de Convenciones EDOMEX, Ciudad de Toluca, Estado de México. **Ponencia**

Artículos indexados

1. Armando Delgadillo Jiménez, Carina Toxqui-Quitl, Raúl Castro-Ortega, **Aldo Aguilar Vallejo**, Alfonso Padilla-Vivanco, Anna Carbone. 2025 “*Estimation of urban fractal dimension using a Convolutional Neural Network*”, Lectures notes on Artificial Intelligence, Springer. **Aceptado para publicación**
2. Armando Delgadillo Jiménez, Carina Toxqui-Quitl, Raúl Castro-Ortega, **Aldo Aguilar Vallejo**, Alfonso Padilla-Vivanco, Anna Carbone. 2025 “*A Convolutional Neural Network approach to quantifying Urban Scaling Laws*”, Smart cities, Springer. **En evaluación**

1.10. Conclusiones

En este trabajo de Tesis, se diseñan, implementan en MATLAB y evalúan algoritmos para la segmentación de coberturas urbanas usando imágenes multispectrales del satélite WorldView-2 (WV-2). De acuerdo a la literatura encontrada, se analizan e implementan los índices de segmentación de construcción que han sido diseñados para imágenes provenientes del satélite WV-2. Se establecen las limitantes de cada uno de los índices propuestos en el área y se propone una solución que permite calcular la densidad de construcción de manera más exacta.

Los métodos basados en índices han demostrado ser fáciles de implementar por su simplicidad, son computacionalmente eficientes y se adaptan a diferentes bases de datos dados por los satélites LANDSAT, SENTINEL o WorldView-2. Se eligieron las imágenes provenientes del satélite WV-2 porque proporciona imágenes de mayor resolución respecto a las proporcionadas por otros satélites. A manera de comparativa, se implementa una arquitectura de redes neuronales para evaluar su capacidad de discriminar construcción sobre la misma base de datos.

De esta manera, se establece como objetivo principal el diseño de un índice de segmentación de construcción que sea capaz de distinguir entre construcción y suelo desnudo. Lo anterior usando la base de datos WV-2. Finalmente, se listan los trabajos de investigación derivados de la Tesis. Los cuales, muestran los resultados comparativos del Índice propuesto respecto de los existentes en la literatura.

Bibliografía

- [1] Natural Resources Canada Canada Centre for Remote Sensing. Fundamentals of remote sensing: A canada centre for remote sensing tutorial. Tutorial / PDF document, 2008.
- [2] Basudeb Bhatta. *Analysis of Urban Growth and Sprawl from Remote Sensing Data*. Advances in Geographic Information Science. Springer Berlin Heidelberg, 2010.
- [3] Sebastián Martinuzzi, William A. Gould, and Olga M. R. González. Land development, land use, and urban sprawl in puerto rico integrating remote sensing and population census data. *Landscape and Urban Planning*, 79(3-4):288–297, 2007.
- [4] Josep Roca, M. Carmen Burns, and Josep Maria Carreras. Monitoring urban sprawl around barcelona’s metropolitan area with the aid of satellite imagery. In *Proceedings of Geo-Imagery Bridging Continents, XXth ISPRS Congress*, Istanbul, Turkey, 2004. ISPRS.
- [5] Robert A. Schowengerdt. *Remote Sensing: Models and Methods for Image Processing*. Academic Press, Elsevier, Burlington, MA, USA, 3rd edition, 2007.

- [6] Libo Wang, Rui Li, Ce Zhang, Shenghui Fang, Chenxi Duan, Xiaoliang Meng, and Peter M. Atkinson. Unetformer: A unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190:196–214, 2022.
- [7] Mario Arreola-Esquivel, Carina Toxqui-Quitl, Maricela Delgadillo-Herrera, Alfonso Padilla-Vivanco, Gabriel Ortega-Mendoza, and Anna Carbone. Non-binary snow index for multi-component surfaces. *Remote Sensing*, 13(14):2777, 2021.
- [8] Moses A. Cho and Abel Ramoelo. Urban land-cover mapping using worldview-2 imagery: Evaluating spectral indices and classification algorithms. In *Proc. SPIE 9643, Earth Observing Missions and Sensors: Development, Implementation, and Characterization III*, volume 9643, page 96430K, 2015.
- [9] B. Delaney, K. Tansey, and M. Whelan. Satellite remote sensing techniques and limitations for identifying bare soil. *Remote Sensing*, 17(4):630, 2025.
- [10] Juan C. Valdiviezo-N, Alejandro Téllez-Quiñones, Adan Salazar-Garibay, and Alejandra A. López-Caloca. Built-up index methods and their applications for urban extraction from sentinel 2a satellite data: discussion. *Journal of the Optical Society of America A*, 35(1):35–44, 2018.
- [11] A. Rasul, H. Balzter, G. R. F. Ibrahim, H. M. Hameed, J. Wheeler, B. Adamu, S. Ibrahim, and P. M. Najmaddin. Applying built-up and bare-soil indices from landsat 8 to cities in dry climates. *Land*, 7(3):81, 2018.
- [12] B. C. Forster. Some urban measurements from landsat data. *Photogrammetric Engineering and Remote Sensing*, 49(8):1293–1307, 1983.
- [13] Y. Zha, J. Gao, and S. Ni. Use of normalized difference built-up index in automatically mapping urban areas from tm imagery. *International Journal of Remote Sensing*, 24(3):583–594, 2003.
- [14] D. L. Civco, J. D. Hurd, J. P. Wilson, and C. L. Arnold. Comparison of land use and land cover change detection approaches. Technical report, University of Connecticut, 2002.
- [15] L. Yu, J. Liu, and L. Wang. Classification of remote sensing imagery using a markov random field model. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5):2542–2549, 1999.
- [16] P.A. Longley and V. Mesev. Measuring urban morphology using remotely-sensed imagery. In J.P. Donnay, M.J. Barnsley, and P.A. Longley, editors, *Remote Sensing and Urban Analysis*, pages 163–183. Taylor Francis, 2001.

- [17] M. Pesaresi and A. Bianchin. Recognizing settlement structure using mathematical morphology and image texture. In J.P. Donnay, M.J. Barnsley, and P.A. Longley, editors, *Remote Sensing and Urban Analysis*, pages 55–67. Taylor Francis, 2001.
- [18] P.A. Brivio and E. Zilioli. Urban pattern characterization through geostatistical analysis of satellite images. In J.P. Donnay, M.J. Barnsley, and P.A. Longley, editors, *Remote Sensing and Urban Analysis*, pages 40–53. Taylor Francis, 2001.
- [19] V. Ramos and J. Silva. Estadística espacial para el análisis urbano: una revisión de técnicas con ejemplos. *Revista de Teledetección*, 22:5–19, 2003.
- [20] W.M.F. Grey and A.J. Luckman. Mapping urban extent using satellite radar interferometry. *Photogrammetric Engineering Remote Sensing*, 69(9):957–961, 2003.
- [21] B. Bhatta. *Analysis of Urban Growth and Sprawl from Remote Sensing Data*, volume 15 of *Remote Sensing Applications*. Springer, Heidelberg, 2009.
- [22] Adeniyi Adeyemi, Abel Ramoelo, Moses Cho, and Cecilia Masemola. Spectral index to improve the extraction of built-up area from worldview-2 imagery. *Journal of Applied Remote Sensing*, 15(2):024510, 2021.
- [23] Maher Ibrahim Sameen and Biswajeet Pradhan. A novel built-up spectral index developed by using multiobjective particle-swarm-optimization technique. In *Proc. of the International Conference on Geomatic and Geospatial Technology (GGT)*, 2016. Conference Paper.
- [24] Paidamwoyo Mhangara, John Odindi, Linda Kleyn, and Hardly Remas. Road extraction using object oriented classification. In *Proceedings of the 2015 ResearchGate Conference*, 2015.
- [25] Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using worldview-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:205–215, 2014.

Capítulo 2

Segmentación de construcción basada en índices

2.1. Introducción

Los métodos basados en índices son ampliamente adoptados en sensado remoto para extraer superficies impermeables a partir de imágenes multiespectrales. Estos métodos utilizan combinaciones computacionales de bandas espectrales para enfatizar tipos específicos de cobertura del suelo, particularmente las áreas construidas, aprovechando sus propiedades distintivas de reflectancia. Su simplicidad, eficiencia computacional y adaptabilidad a diferentes sensores y resoluciones los convierten en una opción común para análisis urbanos a gran escala y de múltiples periodos [1–3].

A pesar de su simplicidad, a menudo enfrentan limitaciones para separar con precisión las superficies construidas de clases espectralmente similares, como el suelo desnudo o la vegetación seca, especialmente en entornos urbanos heterogéneos [4]. Para abordar este desafío, se han propuesto varios índices que incorporan diferentes regiones espectrales de sensores de alta resolución como el WorldView-2. Estudios recientes han demostrado la utilidad de los índices construidos para mejorar el mapeo de superficies impermeables, particularmente cuando se combinan con una adecuada determinación de umbrales y marcos de validación estadística [1].

Este estudio evalúa cuatro métodos basados en índices: el Índice de Área Construida (BAI) [5], el Índice Espectral de Áreas Construidas (BSI) [6], el Nuevo Índice de Extracción de Áreas Construidas (NBEI) [7], el Índice Red Edge-Verde (RGI) [8] y el . Los cuatro índices se han seleccionado por su relevancia en la literatura reciente y su compatibilidad con las propiedades espectrales del WorldView-2. El rendimiento comparativo en la extracción de superficies impermeables se muestra utilizando un marco de evaluación consistente y cuantitativo en cuatro ciudades (Atenas, Bucarest, Roma y Turín). El análisis incluye métricas de

precisión e indicadores de discriminación espectral para evaluar la capacidad de cada índice para distinguir entre áreas construidas que incluyen carreteras asfaltadas, edificios, concreto y otras superficies impermeables artificiales, y superficies de fondo, particularmente suelo desnudo, agua y vegetación.

En la Figura 2.1 se muestra una imagen en falso color de la ciudad de Helsinki con múltiples tipos de superficie, como construcción, vegetación, agua, nubes, suelo desnudo, entre otros.

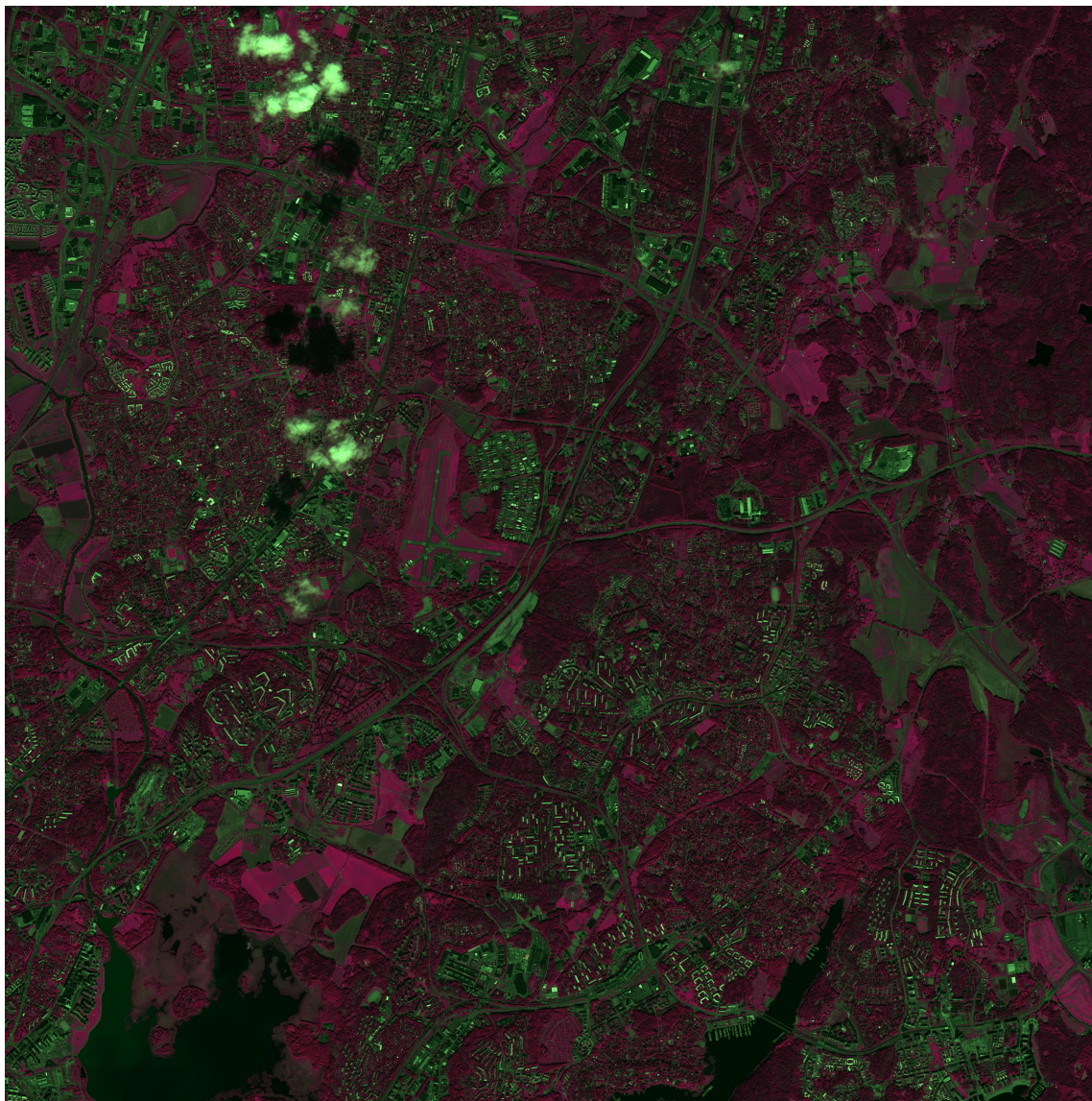


Figura 2.1: Imagen en Falso Color de la Ciudad de Helsinki, se pueden observar en tonos verdes áreas construidas, y en tonos rosas áreas de vegetación. Nótese que algunas zonas de suelo desnudo se confunden con construcción.

2.2. Imágenes multispectrales de WorldView-2

El conjunto de datos de prueba está compuesto por imágenes multispectrales de alta resolución adquiridas por el satélite WorldView-2 (WV-2), lanzado por DigitalGlobe en 2009 (Ahora MAXAR). WV-2 fue uno de los primeros satélites comerciales de alta resolución en proporcionar datos multispectrales de ocho bandas en el espectro visible e infrarrojo cercano (VNIR) con una resolución espacial de 1.6 metros, junto con una banda pancromática de 0.46 metros. Su configuración espectral única mejora la discriminación de elementos urbanos como superficies impermeables, vegetación y suelo desnudo, lo cual es fundamental para aplicaciones de mapeo urbano [9]. La Tabla 2.1 resume las bandas espectrales y sus rangos de longitud de onda.

Tabla 2.1: Bandas espectrales del satélite WorldView-2.

Banda	Nombre	Longitud de Onda (nm)
B1	Coastal Blue	400–450
B2	Blue	450–510
B3	Green	510–580
B4	Yellow	585–625
B5	Red	630–690
B6	Red Edge	705–745
B7	NIR1	770–895
B8	NIR2	860–1040

El conjunto de datos contiene escenas urbanas de veintiséis ciudades europeas. Estas ciudades fueron elegidas por sus estilos arquitectónicos distintivos, heterogeneidad en la cobertura del suelo y gradientes de densidad urbana, lo que proporciona un banco de pruebas integral para la evaluación de índices espectrales. Todas las imágenes fueron preprocesadas mediante ortorrectificación y corrección atmosférica para garantizar la consistencia geométrica y radiométrica. En este trabajo sólo se utilizaron las ocho bandas multispectrales. La banda pancromática fue excluida. Estos subconjuntos multispectrales son especialmente adecuados para la evaluación de índices de áreas construidas, como se ha demostrado en estudios previos de sensado remoto urbana [10].

2.3. Preprocesamiento de imágenes WorldView-2

El preprocesamiento de las imágenes multispectrales WorldView-2 nivel 2A se realizó siguiendo el procedimiento de corrección radiométrica absoluta de DigitalGlobe en 2017 [11] y actualizado por Maxar en 2019 [12]. Este proceso convierte los Números Digitales (DN) de la imagen satelital en reflectancia de la parte superior de la atmósfera (TOA) y posteriormente en reflectancia de superficie, aplicando el método de sustracción de objetos oscuros (DOS, por sus siglas en inglés).

Paso 1: Conversión de DN a radiancia TOA

Los productos satelitales de DigitalGlobe se entregan en formato digital como píxeles corregidos radiométricamente de manera relativa.

Los valores registrados por los sensores satelitales se almacenan en forma de números digitales (DN, *Digital Numbers*), que representan la respuesta cruda del detector ante la radiancia incidente. Sin embargo, estos valores carecen de significado físico directo, ya que están influenciados por las características del sensor, los ajustes instrumentales y los procesos de adquisición. Para que los datos puedan ser comparados en diferentes momentos, sensores o condiciones de observación, es necesario transformarlos a unidades físicas de radiancia espectral en la parte superior de la atmósfera. Este proceso de calibración radiométrica garantiza la consistencia entre imágenes y constituye un paso fundamental en aplicaciones de sensado remoto.

Los valores DN se convierten a radiancia espectral en la parte superior de la atmósfera (L) mediante la siguiente expresión:

$$L = \frac{GAIN \cdot DN \cdot AbsCalFactor}{EffectiveBandwidth} + OFFSET$$

donde:

- $GAIN$ corresponde al factor de ganancia aplicado en cada banda,
- $OFFSET$ representa el valor de compensación específico de cada banda,
- $AbsCalFactor$ es el factor de calibración radiométrica absoluta, proporcionado en el archivo de metadatos .IMD,
- $EffectiveBandwidth$ es el ancho de banda efectivo de cada canal espectral, también disponible en el archivo de metadatos .IMD.

La Tabla 2.2 presenta los valores de calibración utilizados en este estudio [12].

Tabla 2.2: Factores de calibración absoluta [12] para WorldView-2

Banda	GAIN	OFFSET
Coastal Blue	0.938	-13.099
Blue	0.946	-9.409
Green	0.958	-7.771
Yellow	0.979	-5.489
Red	0.969	-4.579
Red Edge	1.027	-5.552
NIR1	0.977	-6.508
NIR2	1.007	-3.699

Paso 2: Conversión de radiancia a reflectancia sobre la atmósfera

La reflectancia en la parte superior de la atmósfera (TOA, *Top-of-Atmosphere*) se obtiene a partir de la radiancia espectral registrada por el sensor, normalizando sus valores con respecto a la irradiancia solar incidente. Esta conversión es un paso fundamental, ya que permite expresar las mediciones en una escala adimensional que facilita la comparación entre diferentes fechas, sensores y condiciones de iluminación. Además, la transformación a reflectancia es un requisito común en numerosos análisis multispectrales, como índices de vegetación (NDVI), cocientes espectrales y transformaciones matriciales [11, 12].

En términos físicos, la radiancia registrada en un sensor satelital corresponde a la suma de varias contribuciones: la radiación reflejada directamente por la superficie, la radiación reflejada por el cielo (downwelling skylight) y la radiación de trayectoria ascendente (path radiance) generada en la atmósfera. Si bien un modelo completo debe considerar todos estos términos, en aplicaciones prácticas se asume un comportamiento lambertiano de la superficie y se ignoran los efectos atmosféricos, lo que lleva a una forma simplificada de la ecuación. De este modo, la reflectancia TOA se calcula como:

$$\rho_{\text{TOA}} = \frac{L \cdot d_{ES}^2 \cdot \pi}{E_{\text{sun}} \cdot \cos(\theta)}$$

donde:

- L es la radiancia espectral registrada por el sensor.

- d_{ES} es la distancia Tierra–Sol en unidades astronómicas.
- E_{sun} es la irradiancia solar exoatmosférica media para cada banda [?].
- θ es el ángulo cenital solar.

Es importante señalar que la reflectancia TOA no corrige efectos atmosféricos, topográficos ni de anisotropía bidireccional de la reflectancia (BRDF). Para mitigar parcialmente la influencia atmosférica, se recomienda aplicar técnicas como la *dark object subtraction (DOS)* o el uso de modelos radiativos más avanzados [13, 14].

Tabla 2.3: Irradiancia exoatmosférica media (E_{sun}) según Thuillier 2013 [12]

Banda	E_{sun} ($\text{W m}^{-2} \mu\text{m}^{-1}$)
Coastal Blue	1773.81
Blue	2007.27
Green	1829.62
Yellow	1701.85
Red	1538.85
Red Edge	1346.09
NIR1	1053.21
NIR2	856.599

Paso 3: Corrección atmosférica (DOS)

Tras la conversión de radiancia a reflectancia de cima de atmósfera (ρ_{TOA}), persisten los efectos de dispersión atmosférica, principalmente debidos a la radiancia de trayectoria ascendente. Para reducir este efecto, se empleó el método de *Dark Object Subtraction (DOS)*, el cual asume que ciertos píxeles de la escena —correspondientes a cuerpos de agua profunda, sombras o vegetación densa— deberían presentar valores de reflectancia cercanos a cero [14].

Sin embargo, debido a la dispersión atmosférica, estos píxeles muestran valores positivos en ρ_{TOA} . Este valor mínimo detectado se interpreta como una estimación de la reflectancia atribuida a la atmósfera:

$$\rho_{\lambda}^{\text{atm}} \approx \rho_{\lambda}^{\text{mín}}$$

donde $\rho_\lambda^{\text{mín}}$ corresponde al valor mínimo observado en la banda λ .

La reflectancia corregida en superficie se obtiene restando dicho valor a cada píxel:

$$\rho_\lambda^c = \rho_\lambda^{\text{TOA}} - \rho_\lambda^{\text{atm}}$$

con la restricción de que $\rho_\lambda^c \geq 0$ para evitar valores negativos.

Este procedimiento asegura que los objetos más oscuros de la imagen presenten valores de reflectancia próximos a cero, reduciendo la influencia de la radiancia de trayectoria atmosférica. No obstante, el método DOS no corrige efectos de absorción atmosférica ni la anisotropía angular de la superficie, por lo que se considera un método de primera aproximación.

2.4. Firma espectral de diferentes superficies

La firma espectral representa la variación de la reflectancia de un material en función de la longitud de onda. Cada tipo de superficie como agua, vegetación, suelo desnudo o áreas construidas interactúa con la radiación solar de manera distinta, generando patrones característicos en el espectro electromagnético [2].

Estas diferencias permiten discriminar superficies en imágenes satelitales, ya que, por ejemplo, la vegetación presenta alta reflectancia en el infrarrojo cercano, mientras que el agua absorbe fuertemente en casi todo el espectro visible e infrarrojo. Las superficies urbanas construidas y de suelo desnudo, en cambio, muestran firmas espectrales más similares, lo que representa un desafío para su separación precisa.

En esta sección se presentan ejemplos de firmas espectrales obtenidas a partir de imágenes multiespectrales, que ilustran el comportamiento característico de diferentes coberturas del terreno y su utilidad para la clasificación y segmentación en sensado remoto.

En la Figura 2.2 se pueden observar el comportamiento de la construcción, agua y vegetación.

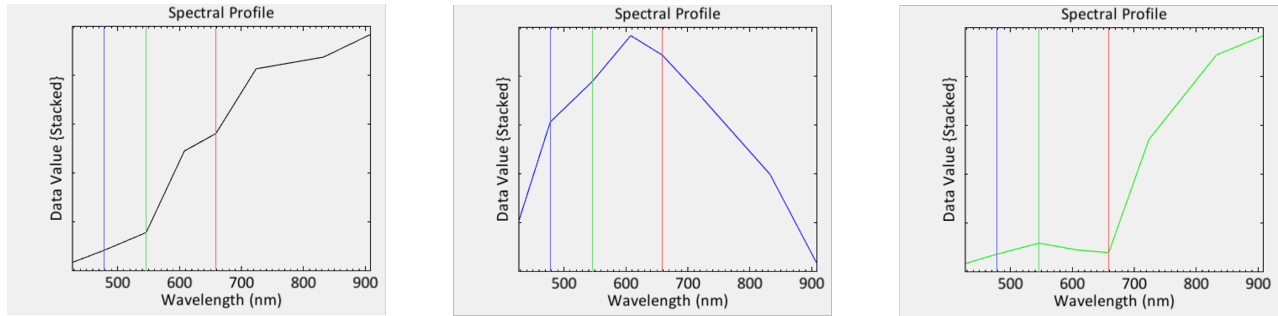


Figura 2.2: Firma espectral de construcción, agua y vegetación

2.5. Índices para la Extracción de construcción

Como se ha mencionado anteriormente, un problema común de los índices espectrales para el mapeo urbano es la superposición espectral entre áreas construidas y suelo desnudo. Ambos tipos de superficie a menudo presentan una reflectancia similar en los rangos del espectro visible e infrarrojo cercano, particularmente en condiciones secas o en entornos con vegetación escasa. [15].

2.5.1. Built-up Area Index (BAI)

El Índice de Área Construida (BAI, por sus siglas en inglés) es un índice espectral simple pero efectivo, diseñado para mejorar la detección de superficies artificiales, particularmente asfalto y concreto, en imágenes de alta resolución. Aprovecha el contraste en la reflectancia entre las bandas del infrarrojo cercano y el azul, basado en la reflectancia característica de la mayoría de los materiales construidos.

El BAI fue propuesto inicialmente para apoyar la clasificación orientada a objetos de caminos urbanos, pero desde entonces ha demostrado una aplicabilidad más amplia en la delimitación de superficies impermeables en entornos urbanos densos [5]. Su formulación matemática se expresa como:

$$\text{BAI} = \frac{\text{Blue} - \text{NIR1}}{\text{Blue} + \text{NIR1}} \quad (2.1)$$

La Figura 2.3 muestra el histograma acumulado para el BAI, así como los mapas de valores del BAI en las cuatro ciudades.

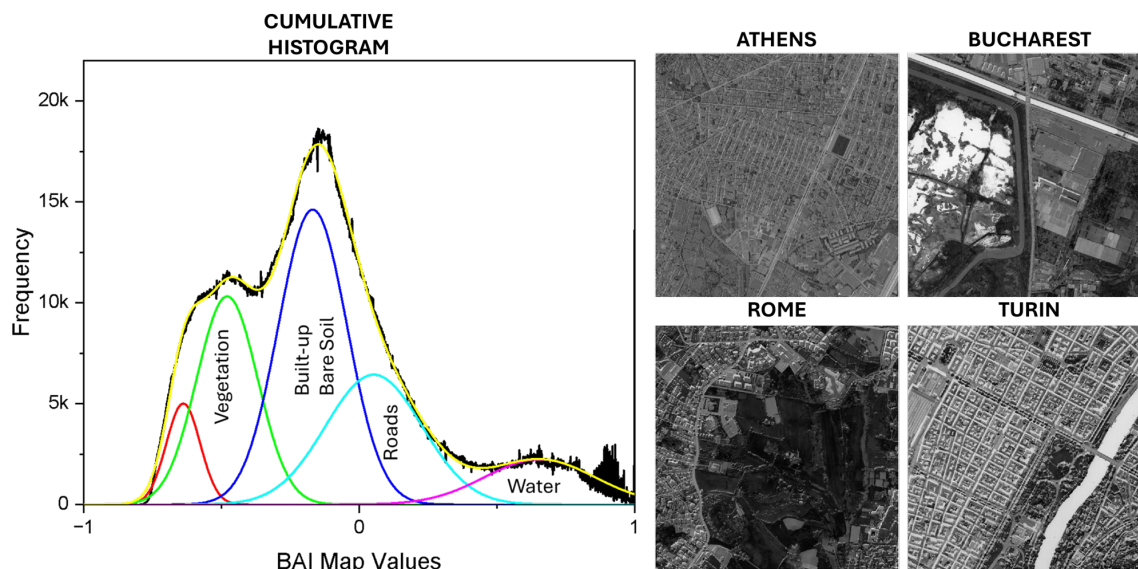


Figura 2.3: Histograma acumulado y distribución espacial del **BAI** en las cuatro escenas urbanas. El histograma muestra que la mayoría de las superficies tienen valores por debajo de cero, separando de manera efectiva el agua, pero sin lograr distinguir completamente la vegetación de las áreas construidas y confundiendo áreas construidas con suelo desnudo y caminos.

2.5.2. Built-up Spectral Index (BSI)

El Índice Espectral de Áreas Construidas (BSI, por sus siglas en inglés) fue diseñado para mejorar la extracción de superficies impermeables a partir de datos del satélite WorldView-2. A diferencia de los índices convencionales derivados de combinaciones heurísticas de bandas, el BSI fue desarrollado utilizando un algoritmo de optimización por enjambre de partículas multiobjetivo. Esta técnica identificó las bandas espectrales más relevantes y sus pesos óptimos para mejorar la separabilidad espectral entre superficies construidas y no construidas [6].

La formulación final del BSI se expresa como:

$$\text{BSI} = \frac{\text{Yellow} - 2 \times \text{NIR1}}{\text{Yellow} + 2 \times \text{NIR1}} \quad (2.2)$$

La Figura 2.4 muestra el histograma acumulado del BSI y los mapas de valores del BSI en las cuatro ciudades.

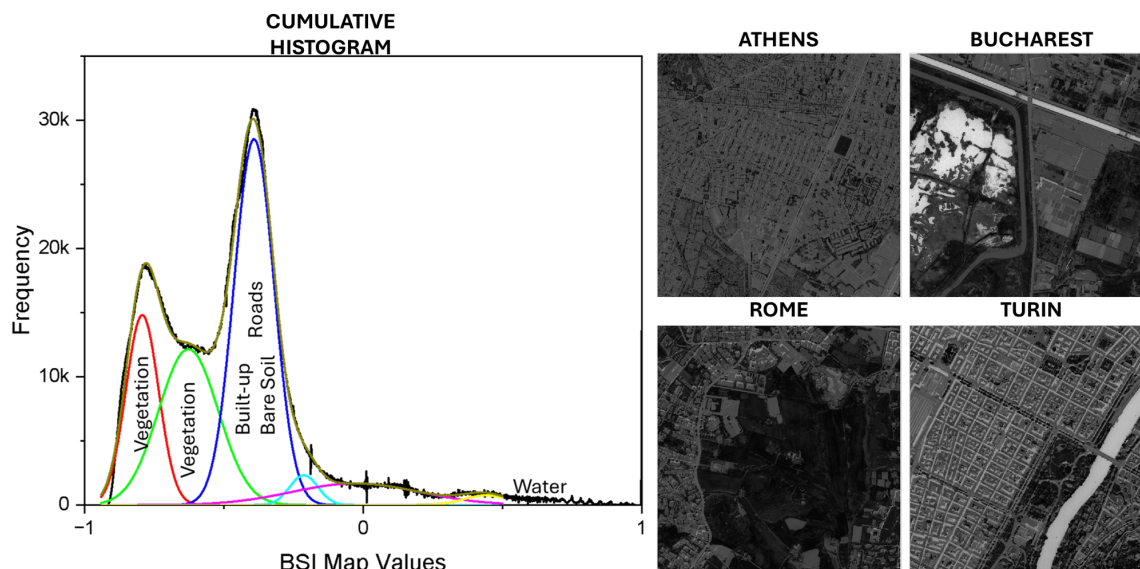


Figura 2.4: Histograma acumulado y distribución espacial del **BSI** en las cuatro escenas urbanas. Al igual que el BAI, el histograma muestra que la mayoría de las superficies tienen valores por debajo de cero, aislando de manera efectiva el agua. Ofrece una mejor separación entre vegetación y áreas construidas, pero aún presenta confusión entre áreas construidas, suelo desnudo y caminos.

2.5.3. New Built-up Extraction Index (NBEI)

El Nuevo Índice de Extracción de Áreas Construidas (NBEI, por sus siglas en inglés) fue desarrollado para abordar las limitaciones de los índices espectrales tradicionales cuando se aplican a imágenes de alta resolución como las del WorldView-2. El NBEI mejora el contraste espectral entre las superficies impermeables y las coberturas del suelo circundantes al combinar las bandas más discriminantes identificadas mediante la selección de características ReliefF: NIR2, NIR1, Verde y Borde Rojo. Su formulación se basa en la diferencia normalizada entre la suma de las bandas del infrarrojo cercano (reflectantes) y las bandas del borde del visible (absorbentes), lo que incrementa la separabilidad entre materiales urbanos heterogéneos [7]. El índice se expresa como:

$$\text{NBEI} = \frac{(\text{NIR2} + \text{NIR1}) - (\text{Green} + \text{RedEdge})}{(\text{NIR2} + \text{NIR1}) + (\text{Green} + \text{RedEdge})} \quad (2.3)$$

La Figura 3.4 muestra el histograma acumulado del NBEI y los mapas de valores del NBEI en las cuatro ciudades.

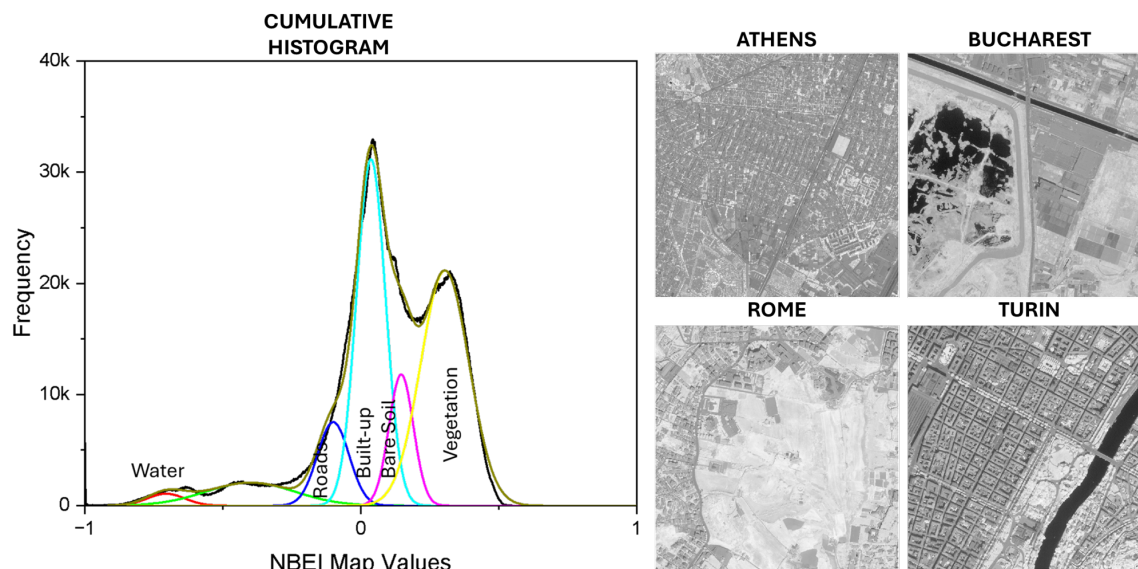


Figura 2.5: Histograma acumulado y distribución espacial del **NBEI** en las cuatro escenas urbanas. Los valores de áreas construidas se agrupan alrededor de cero, la vegetación aparece con valores más altos y el agua cerca de -1 . Al igual que otros índices, las áreas construidas suelen confundirse con el suelo desnudo.

2.5.4. Red Edge - Green Index (RGI)

El Índice Red Edge-Verde (RGI, por sus siglas en inglés) es un índice espectral diseñado para mejorar la discriminación entre superficies impermeables y suelo desnudo en entornos urbanos y periurbanos. Utiliza las bandas del Borde Rojo y Verde del sensor WorldView-2 para capturar diferencias sutiles en la reflectancia superficial [8]. El índice se calcula como la diferencia normalizada entre las bandas del Borde Rojo y Verde:

$$\text{RGI} = \frac{\text{RedEdge} - \text{Green}}{\text{RedEdge} + \text{Green}} \quad (2.4)$$

La Figura 2.6 muestra el histograma acumulado del RGI y los mapas de valores del RGI en las cuatro ciudades.

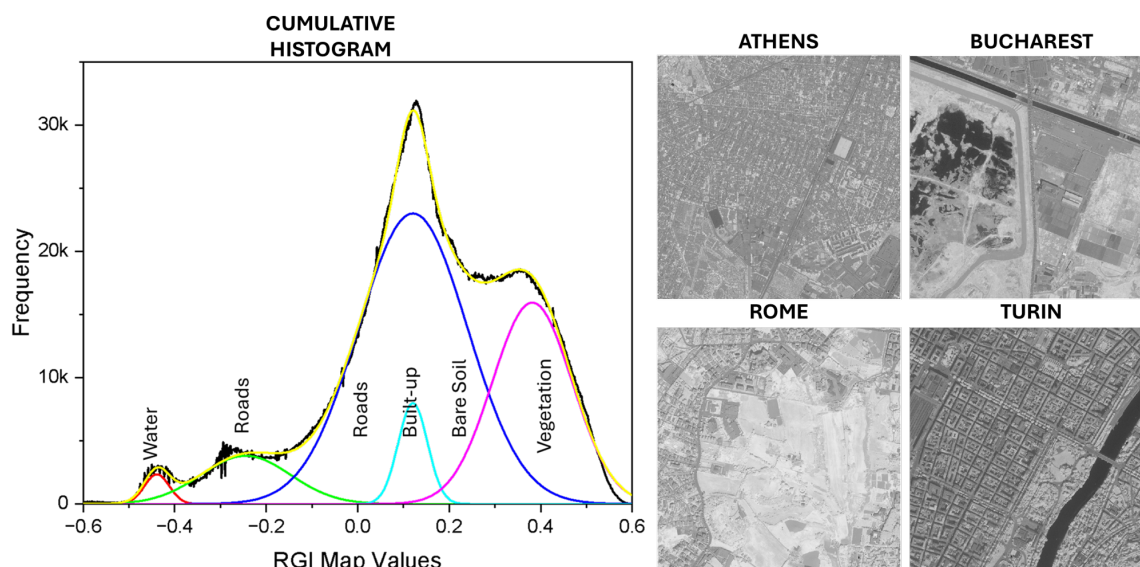


Figura 2.6: Histograma acumulado y distribución espacial del **RGI** en las cuatro escenas urbanas. Los valores de áreas construidas se agrupan entre -0.2 y 0.4 , la vegetación aparece entre 0.2 y 0.6 , y el agua por debajo de -0.4 . Al igual que otros índices, las áreas construidas suelen confundirse con el suelo desnudo.

2.5.5. Masked Built-up Extraction Index (MBEI)

Los índices de construcción suelen presentar limitaciones al diferenciar superficies edificadas de suelos desnudos o rocas expuestas, debido a la similitud espectral entre estas coberturas [16]. Esta superposición puede derivar en sobreestimaciones de las áreas urbanizadas [17].

Para superar esta limitación, se propone la incorporación de una máscara que aisle y elimine los píxeles correspondientes a suelo desnudo [18]. El *Índice para Suelo Desnudo* (ISD), diseñado originalmente para imágenes Landsat TM con las bandas SWIR2 y verde [19], caracteriza el comportamiento espectral del suelo desnudo, que suele mostrar alta reflectancia en el infrarrojo medio y baja en el verde. Dado que el sensor WorldView-2 no cuenta con la banda SWIR2, se adaptó el índice sustituyendo dicha banda por NIR2:

$$ISD = \frac{NIR2 - Green}{NIR2 + Green} \quad (2.5)$$

Usando un índice de construcción como el *NBEI*, y el índice de suelo *ISD* como máscara para una misma región de interés (*RoI*). Un píxel identificado como suelo desnudo en el mapa del ISD se elimina del mapa de construcción mediante una máscara \bar{B} definida como:

$$\bar{B}(i_1, i_2) = \begin{cases} 0 & \text{si } R_{min} < NBEI(i_1, i_2) < R_{max} \\ & \wedge T_{min} < ISD(i_1, i_2) < T_{max} \\ 1 & \text{en otro caso} \end{cases} \quad (2.6)$$

donde $[R_{min}, R_{max}]$ y $[T_{min}, T_{max}]$ son los rangos de umbral establecidos para suelo desnudo para los índices de construcción y suelo desnudo respectivamente.

Esta máscara se aplica al mapa binarizado del índice de construcción $NBEI$ el cuál se obtiene:

$$B_0(i_1, i_2) = \begin{cases} 1 & \text{si } U_{min} < NBEI(i_1, i_2) < U_{max} \\ 0 & \text{en otro caso} \end{cases} \quad (2.7)$$

donde $[U_{min}, U_{max}]$ corresponde al rango de umbral para construcción.

El *Masked Built-up Extraction Index* ($MBEI$) resulta de aplicar la máscara de construcción \bar{B} al mapa binarizado B_0 , según:

$$B = B_0 \circ \bar{B} \quad (2.8)$$

2.6. Cuantificación de la densidad de construcción

La densidad de construcción \mathcal{B} en un RoI se calcula como la proporción de píxeles clasificados como urbanos respecto al total de píxeles A_{tot} en la región [20]:

$$\mathcal{B} = \frac{1}{A_{tot}} \sum_{i_1=1}^M \sum_{i_2=1}^N MBEI(i_1, i_2) \quad (2.9)$$

donde M y N corresponden al tamaño del RoI .

En la Figura 4.3 se puede observar el resultado de extracción de construcción usando el índice propuesto $MBEI$ en la ciudad de Bruselas, Bélgica. A partir de este mapa, todos los píxeles clasificados como construcción son sumados para determinar la densidad de construcción por sector.

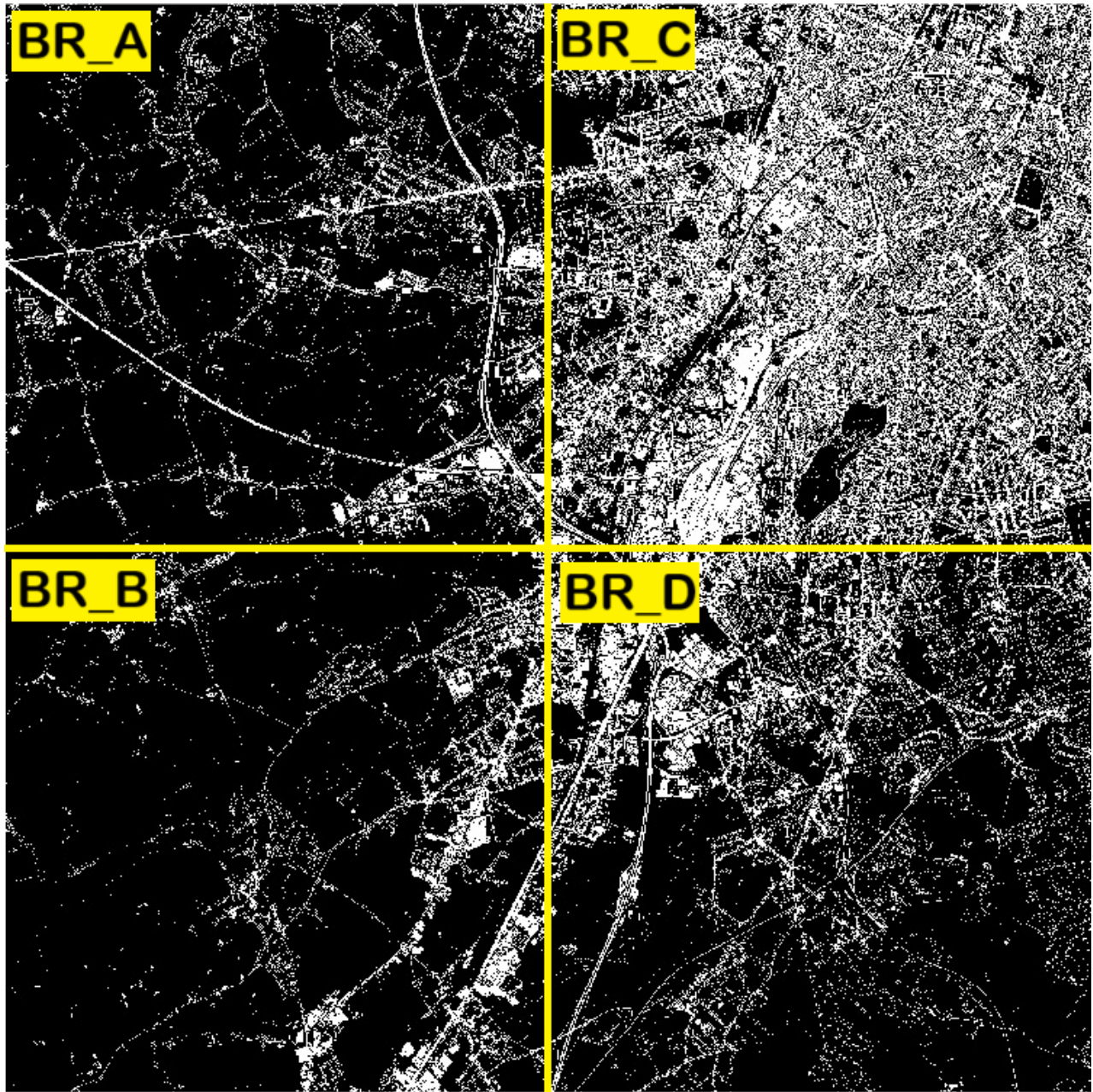


Figura 2.7: Resultados de extracción de construcción usando el método propuesto de enmascaramiento. Como se puede observar, la información de suelo desnudo fue eliminada. Se calcularon los valores de densidad de construcción por sectores suburbanos $\mathcal{B}_A = 0.13$, $\mathcal{B}_B = 0.11$, $\mathcal{B}_C = 0.63$, y $\mathcal{B}_D = 0.23$.

2.7. Conclusiones

En este capítulo se analizó la eficiencia de diversos índices espectrales aplicados a imágenes multiespectrales del satélite WorldView-2 para la extracción de áreas construidas. Los resultados muestran que, aunque índices como el BAI, BSI, NBEI y RGI son herramientas simples y computacionalmente eficientes, enfrentan limitaciones para separar con precisión las superficies edificadas de otras coberturas espectralmente similares, como el suelo desnudo.

Se define el nuevo índice propuesto *MBEI*. Y su relevancia radica en su capacidad para extraer zonas construidas sin incluir suelo desnudo, a diferencia de otros índices de construcción reportados en la literatura. La incorporación de una máscara basada en índices de suelo desnudo demostró ser una estrategia efectiva para reducir la confusión entre clases y mejorar la delimitación de superficies urbanas. Asimismo, se evidenció que la combinación de múltiples bandas, particularmente en las regiones del infrarrojo cercano y el borde rojo, incrementa la capacidad discriminante de los índices frente a escenarios urbanos heterogéneos.

La densidad de construcción \mathcal{B} se ha estimado en el mapa *MBEI* obtenido a partir de imágenes multiespectrales. Esta propuesta de mapa contiene todos los píxeles clasificados como desarrollo urbano con una precisión superior al 88.32%. El mapa binario *MBEI* resalta las superficies impermeables, incluidas edificaciones, viviendas, carreteras, asfalto y otros materiales de construcción.

En conjunto, estos resultados confirman que los métodos basados en índices son un punto de partida sólido para la segmentación urbana, pero requieren ajustes complementarios, como la integración de umbrales adaptativos y máscaras espectrales, para lograr una extracción más robusta y precisa de áreas construidas.

Bibliografía

- [1] Yassine Harrak, Rachid Ahmed, and Rahim Aguejdad. Evaluation of spectral indices and global thresholding methods for the automatic extraction of built-up areas: An application to a semi-arid climate using landsat 8 imagery. *Urban Science*, 9:1–26, 03 2025.
- [2] Mario Arreola-Esquivel, Carina Toxqui-Quitl, Maricela Delgadillo-Herrera, Alfonso Padilla-Vivanco, Gabriel Ortega-Mendoza, and Anna Carbone. Non-binary snow index for multi-component surfaces. *Remote Sensing*, 13(14):2777, 2021.
- [3] R. Kaur and P. Pandey. A review on spectral indices for built-up area extraction using remote sensing technology. *Arabian Journal of Geosciences*, 15(4):391, 2022.

- [4] Mohammad Karimi Firozjaei, Amir Sedighi, Majid Kiavarz, Salman Qureshi, Dagmar Haase, and Seyed Kazem Alavipanah. Automated built-up extraction index: A new technique for mapping surface built-up areas using landsat 8 oli imagery. *Remote Sensing*, 11(17), 2019.
- [5] Paidamwoyo Mhangara, John Odindi, Linda Kleyn, and Hardly Remas. Road extraction using object oriented classification. In *Proceedings of the 2015 ResearchGate Conference*, 2015.
- [6] Maher Ibrahim Sameen and Biswajeet Pradhan. A novel built-up spectral index developed by using multiobjective particle-swarm-optimization technique. In *Proc. of the International Conference on Geomatic and Geospatial Technology (GGT)*, 2016. Conference Paper.
- [7] Adeniyi Adeyemi, Abel Ramoelo, Moses Cho, and Cecilia Masemola. Spectral index to improve the extraction of built-up area from worldview-2 imagery. *Journal of Applied Remote Sensing*, 15(2):024510, 2021.
- [8] Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using worldview-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87:205–215, 2014.
- [9] C. Tarantino, Maria Adamo, Francesco Lovergine, Palma Blonda, and Valeria Tomaselli. *8-Band Image Data Processing of the Worldview-2 Satellite in a Wide Area of Applications*. 01 2012.
- [10] Shanshan Su, Jia Tian, Xinyu Dong, Qingjiu Tian, Ning Wang, and Yanbiao Xi. An impervious surface spectral index on multispectral imagery using visible and near-infrared bands. *Remote Sensing*, 14(14), 2022.
- [11] Michele A. Kuester. Absolute radiometric calibration: 2016v0. Technical report, DigitalGlobe, Westminster, CO, USA, jun 2017. GE01 and CAVIS update.
- [12] Michele Kuester and Tina Ochoa. Improvements in calibration, and validation of the absolute radiometric response of maxar earth-observing sensors. In *Joint Agency Commercial Imagery Evaluation (JACIE) Workshop-2019*, Reston, VA, USA, sep 2019.
- [13] Robert A. Schowengerdt. *Remote Sensing: Models and Methods for Image Processing*. Academic Press, Elsevier, Burlington, MA, USA, 3rd edition, 2007.
- [14] Todd Updike and Chris Comp. Radiometric use of worldview-2 imagery. Technical Note Revision 1.0, DigitalGlobe, nov 2010. Release Date: 01 November 2010.

- [15] Hong Fang, Yuchun Wei, and Qiuping Dai. A novel remote sensing index for extracting impervious surface distribution from landsat 8 oli imagery. *Applied Sciences*, 9:2631, 06 2019.
- [16] C.T. Nguyen, A. Chidthaisong, P. Kieu Diem, and L.-Z. Huo. A modified bare soil index to identify bare land features during agricultural fallow-period in southeast asia using landsat 8. *Land*, 10(3), 2021.
- [17] Nirupam Sarkar and Bidyut Chaudhuri. An efficient differential box counting approach to compute fractal dimension of image. *IEEE Transactions on Systems, Man, and Cybernetics*, 24:115–120, 1994.
- [18] G. Sun, X. Chen, X. Jia, Y. Yao, and Z. Wang. Combinational build-up index (cbi) for effective impervious surface mapping in urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), 2016.
- [19] K. Wentzel. Determination of the overall soil erosion potential in the nsikazi district (mpumalanga province, south africa) using remote sensing and gis. *Canadian Journal of Remote Sensing*, 28(2):322–327, 2002.
- [20] C. Boyko and R. Cooper. Clarifying and re-conceptualising density. *Progress in Planning*, 76:1–61, 2011.

Capítulo 3

Aprendizaje profundo para segmentación de construcción en imágenes satelitales

3.1. Introducción

Generalmente, el mapeo y posterior reconocimiento y clasificación de las zonas urbanas, se han realizado utilizando técnicas y métodos basados en la comparación pixel a pixel de la superficie terrestre, analizando áreas y zonas mediante Puntos de Verdad del Terreno (Ground Truth Points), así como el empleo de métodos estadísticos o aquellos que están basados en reglas. Si bien este tipo de acercamiento es útil, estos métodos pueden verse limitados y/o afectados cuando existe una heterogeneidad en cuanto a su información espectral, es decir algunos materiales pueden tener firmas espectrales similares y compartir información mutua, dificultando el reconocimiento y clasificación de zonas urbanas. Además de esto, la variabilidad geométrica y la probable confusión entre clases limitan el área de acción de estas técnicas.

Para solventar estos problemas presentes en los métodos tradicionales, las redes neuronales convolucionales (CNN) y el aprendizaje profundo (DP) han transformado por completo la forma de analizar imágenes satelitales para propósitos de sensado remoto. Cada vez más hay nuevas arquitecturas e implementaciones diseñadas para aprender representaciones espaciales y que pueden “reconocer” patrones tales como texturas, bordes, formas, entre otros. Con esto, se pueden resolver tareas de segmentación semántica con una precisión superior a la de los métodos clásicos. Algunas configuraciones como las arquitecturas del tipo U-Net, variantes con codificadores profundos basados en ResNet y los modelos basados en transformadores (transformers) o en atención (attention based), han mejorado la extracción automática de superficies urbanas en imágenes de muy alta resolución.

En este capítulo se presenta el diseño, entrenamiento y validación de una arquitectura de red neuronal convolucional (CNN) tipo U-Net, la cual está específicamente adaptada para la extracción de cobertura urbana a partir del análisis de imágenes satelitales WV-2. La elección de este tipo de arquitectura de red se encuentra justificada dada su capacidad comprobada para realizar segmentación semántica de precisión en imágenes de muy alta resolución, combinando un proceso de codificación y uno de decodificación. El primero de ellos, captura características jerárquicas y abstractas de las imágenes, mientras que el segundo, preserva la información espacial detallada.

3.2. Extracción de construcción a partir de imágenes satelitales

Diversos estudios, han confirmado y reportado la robustez y adaptabilidad que la U-Net tiene en el ámbito del área de sensado remoto. Investigaciones como las de Ma et al. [1], Zhang et al. [2] y Li et al. [3], demuestran su efectividad en tareas como la extracción de edificaciones y el mapeo de diversas coberturas del suelo, proporcionando un marco de referencia base para su implementación en este estudio. Estos antecedentes, sumados a las ventajas que ofrece la arquitectura de esta red previamente descrita, constituyeron la motivación principal para usar a la U-Net como modelo principal para segmentación en este trabajo.

3.2.1. Base de datos WV-2

El satélite WorldView-2 (WV-2), lanzado por DigitalGlobe en octubre de 2009, se consolidó como una de las primeras plataformas comerciales de muy alta resolución en ofrecer imágenes multiespectrales de 8 bandas, además de un canal pancromático con resolución espacial de 0.46 metros [4]. Su configuración espectral abarca desde el rango visible hasta el infrarrojo cercano (VNIR), con bandas multiespectrales adquiridas a una resolución espacial de 1.6 metros en NADIR. Esta diversidad espectral permite mejorar la discriminación entre superficies urbanas espectralmente similares, tales como construcción, suelo desnudo y vegetación, convirtiendo al WV-2 en una fuente de datos satelitales clave para estudios de clasificación, estimación y segmentación de zonas urbanas.

El conjunto de datos empleado en este trabajo de Tesis corresponde a una Región de Interés (RoI) ubicada en el área urbana de Berlín, Alemania. La imagen seleccionada posee una extensión espacial de $3,625 \times 3,625$ píxeles, lo que representa aproximadamente $5,800 \times 5,800$ metros en superficie terrestre. La escena incluye una amplia variedad de tipologías urbanas,

entre ellas zonas densamente construidas, áreas verdes y redes de transporte, lo que la convierte en un escenario adecuado para evaluar la capacidad de modelos de segmentación basados en aprendizaje profundo en contextos urbanos heterogéneos.



Figura 3.1: Ciudad de Berlín, Alemania.

3.2.2. Preprocesamiento de las imágenes

Para el entrenamiento de la red neuronal convolucional (CNN), la imagen completa se dividió en 784 sub-imágenes no superpuestas de 128×128 píxeles, correspondientes a áreas terrestres de aproximadamente 204.8×204.8 metros cada una. Se utilizó únicamente la banda NIR1 (B7) como imagen de entrada debido a su alta sensibilidad a las superficies urbanizadas y su robustez bajo condiciones variables de iluminación y atmósfera. Esta banda ha demostrado ser eficaz para la detección de edificaciones y superficies impermeables, especialmente en entornos urbanos complejos [5]. Todas las sub-imágenes fueron normalizadas en el rango $[0, 1]$ y almacenadas como arreglos en escala de grises de un solo canal con dimensiones $128 \times 128 \times 1$, de acuerdo a las prácticas estándar en la preparación de datos de entrada para redes neuronales convolucionales.

Además, en esta etapa se pueden implementar técnicas de aumento de datos, con el objetivo de incrementar la diversidad del conjunto de entrenamiento y con ello, reducir el riesgo de overfitting. La implementación de este aumento de datos, depende en gran medida de los datos de entrada y las condiciones de adquisición. Por ejemplo, si se cuenta con pocos datos, se pueden incluir versiones con rotaciones, con orientación horizontal y/o vertical, y cambios de brillo y contraste dentro de rangos controlados. Estas transformaciones permiten que la CNN aprenda representaciones más robustas frente a variaciones geométricas y de ilumina-

ción presentes en imágenes reales.

De manera similar, se consideraron estrategias de balanceo de clases, dado que algunas áreas urbanas densas pueden aparecer con frecuencia distinta respecto a espacios verdes (vegetación) o áreas despejadas (bare soil), asegurando así que la red no sesgue sus predicciones hacia las clases más predominantes. Los fragmentos procesados se organizaron en lotes (batches) para el entrenamiento, con técnicas de shuffle aleatorio y normalización dentro de cada lote, facilitando la convergencia del optimizador durante el proceso de aprendizaje.

3.3. Arquitectura U-Net

La arquitectura utilizada en este estudio corresponde a una versión modificada de *U-Net*, una Red Neuronal Convolutiva (CNN) introducida originalmente por Ronneberger et al. [6] la cual aborda problemas de segmentación de imágenes biomédicas. La arquitectura U-Net fue seleccionada en este trabajo debido a su eficacia comprobada en tareas y problemas de segmentación semántica, particularmente en escenarios en donde se dispone de un volumen reducido de datos etiquetados y es fundamental preservar la delimitación precisa de bordes y detalles de estructura.

Estas propiedades son completamente valiosas principalmente en aplicaciones de sensado remoto urbano, donde la elevada heterogeneidad espacial y la presencia de elementos de escala fina —tales como techos, pavimentos, zonas verdes o suelo desnudo— requieren una clasificación sumamente precisa nivel de pixel, es decir con muy alta resolución. Una arquitectura del tipo U-Net se encuentra fundamentada en un diseño simétrico de tipo *codificador-decodificador*, el cual es complementado por conexiones residuales (*skip connections*) entre sus capas correspondientes.

Esta arquitectura cuenta con dos divisiones o rutas a seguir. La primera es del lado del codificador, la resolución espacial de la imagen se reduce progresivamente mientras se capturan representaciones semánticas profundas a través de convoluciones y operaciones de agrupamiento máximo (*max-pooling*). Posteriormente, en la segunda ruta, el decodificador reconstruye esta resolución espacial mediante convoluciones transpuestas, y al tiempo reintegra la información de alta resolución proveniente del codificador a través de las conexiones residuales. Este mecanismo atenúa la pérdida de detalles durante el muestreo descendente y con ello, mejora la precisión de localización, convirtiendo a la arquitectura U-Net en una alternativa adecuada para la segmentación de áreas y zonas con construcción, así como de estructuras urbanas con un cierto grado de complejidad.

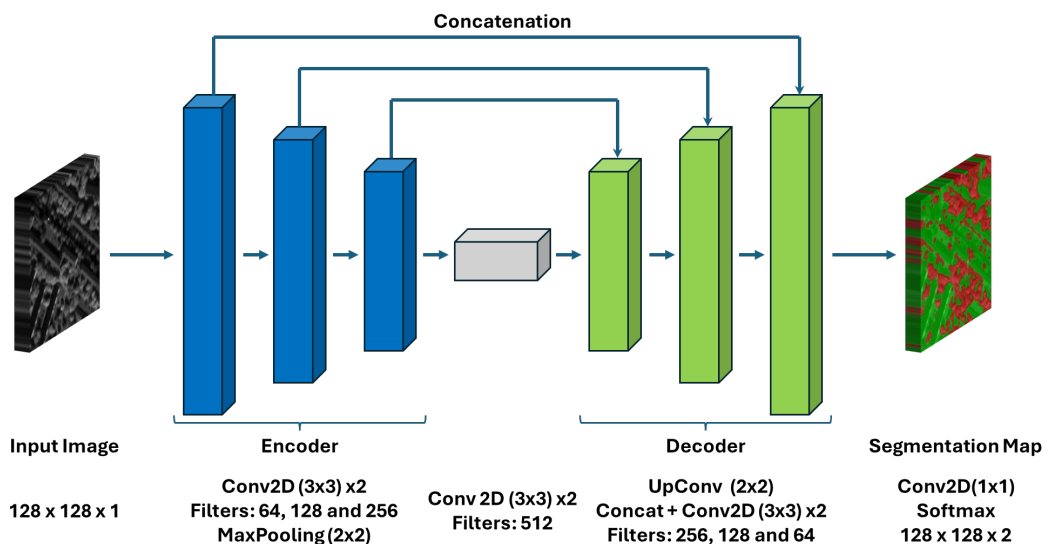


Figura 3.2: Arquitectura de la U-Net modificada utilizada para la segmentación urbana binaria.

Como puede verse en la Figura 3.2, el flujo del modelo inicia con el procesamiento y extracción de Regiones de Interés (RoI), la cual tiene una resolución espacial de 939×939 píxeles (equivalentes a 1502.4×1502.4 metros, considerando una resolución espacial de 1.6 metros), a partir de las cuales se generaron fragmentos de 128×128 píxeles en escala de grises. Cada fragmento fue emparejada con su máscara binaria correspondiente, donde cada píxel fue etiquetado como “construcción” o “no construcción”. El conjunto de datos se dividió en entrenamiento (70 %) y validación (30 %) siguiendo un muestreo estratificado para garantizar una representación equilibrada de ambas clases.

La red comienza con una capa de entrada, seguida de tres bloques convolucionales que conforman el codificador. Cada bloque incluye dos convoluciones con filtros de 3×3 , normalización por lotes (*batch normalization*), una función de activación ReLU y, finalmente, una operación de *max-pooling*, la cual reduce la resolución espacial a la vez que incrementa la profundidad de las características extraídas.

En el cuello de botella (*bottleneck*), la red incorpora dos capas convolucionales con 512 filtros, con ello, se refuerza la capacidad de representación semántica profunda, y una convolución transpuesta, marca el inicio del proceso de expansión espacial.

El decodificador consta de tres bloques que son simétricos al codificador. Cada bloque integra información espacial de alta resolución mediante la concatenación de las conexiones residuales con las características en expansión. Posteriormente, se aplican dos capas convolucionales con normalización por lotes y función de activación ReLU, seguidas por una convolución transpuesta que incrementa la resolución espacial de los mapas de características.

La capa final consiste en una convolución 1×1 , que reduce la dimensionalidad de la salida al número de clases (1 y 0), seguida de una capa *softmax*, encargada de calcular las probabilidades de pertenencia a cada clase y de asignar etiquetas a nivel de píxeles. El entrenamiento se realizó empleando un optimizador Adam con una tasa de aprendizaje de 10^{-4} , durante 30 épocas y con un tamaño de lote de 8. Para fomentar la generalización, los datos fueron reorganizados aleatoriamente en cada época. El conjunto de validación se utilizó como referencia para monitorear el desempeño de la red y prevenir sesgos y sobreajuste.

Una vez completado el entrenamiento, el modelo se evaluó con datos de validación no utilizados en el ajuste de parámetros, aplicando métricas de precisión por píxel y medidas específicas por clase. Los resultados mostraron que la red propuesta logra una segmentación robusta de áreas urbanas como se muestra en 3.3.

PREDICTED CLASSIFICATION

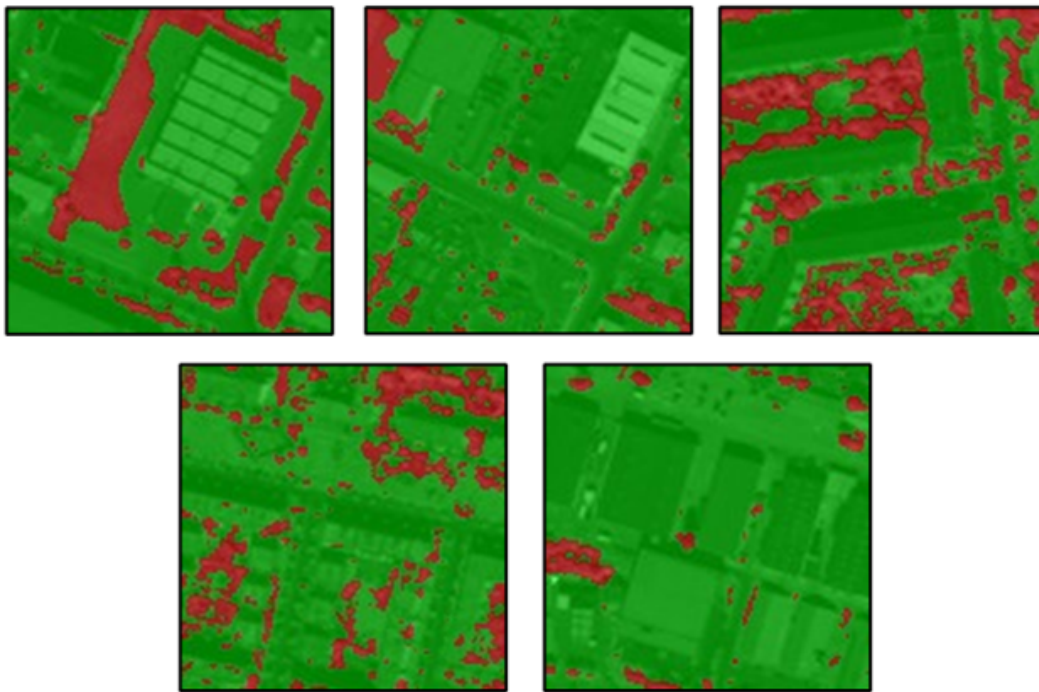


Figura 3.3: Resultados de segmentación utilizando la arquitectura U-Net propuesta.

3.4. Baseline

Para establecer una línea base de rendimiento, el estudio compara los resultados de la clasificación CNN con los obtenidos mediante un índice espectral tradicional: el *New Built-up Extraction Index (NBEI)*. Este índice mejora el contraste espectral entre superficies impermeables y no impermeables mediante la combinación de bandas visibles e infrarrojas cercanas. El NBEI se define de la siguiente manera [7]:

$$\text{NBEI} = \frac{(B8 + B7) - (B3 + B5)}{(B8 + B7) + (B3 + B5)} \quad (3.1)$$

donde $B3$ (Verde), $B5$ (Rojo), $B7$ (NIR1) y $B8$ (NIR2) se refieren a las bandas WorldView-2. Esta formulación se eligió por su capacidad para enfatizar las estructuras artificiales a la vez que suprime las señales de vegetación y suelo desnudo. La Figura 3.4 muestra el mapa de valores del NBEI en una ciudad.

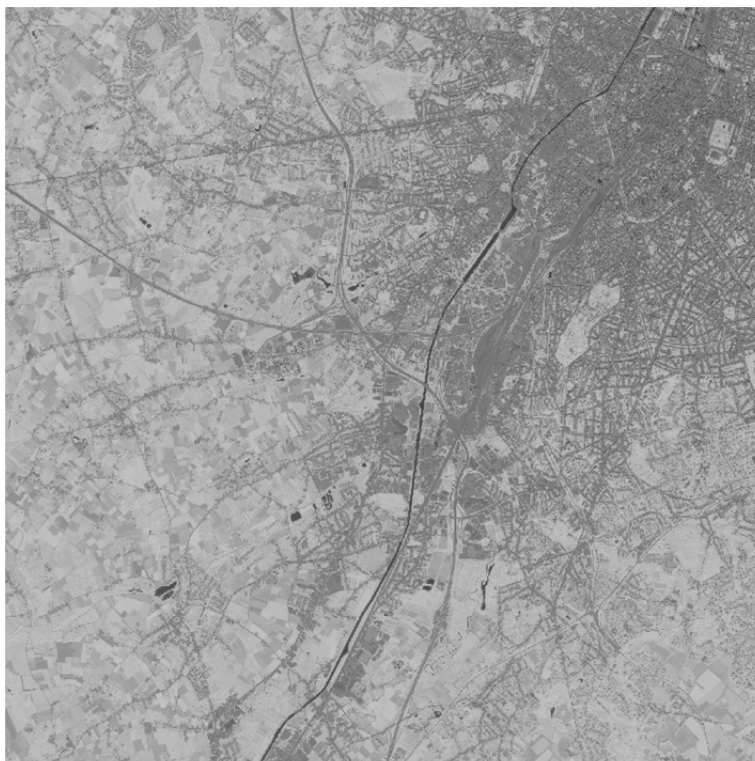


Figura 3.4: Distribución espacial de **NBEI** una escena urbana.

3.5. Conclusiones

En este Capítulo, se utilizó un modelo de segmentación semántica, el cual está basado en una versión adaptada de una arquitectura U-Net, focalizada a realizar clasificación binaria de áreas construidas y no construidas. El modelo de red fue entrenado con imágenes satelitales de alta resolución provenientes del sensor multiespectral WorldView-2.

El modelo propuesto, está centrado en la identificación y delimitación de las zonas urbanas, haciendo especial énfasis en zonas con construcción. Para la validación del desempeño, se utiliza como referencia el Normalized Built-up Extraction Index (NBEI), el cual, como se menciona en los capítulos anteriores es un índice que está diseñado para resaltar superficies con construcción frente a las zonas de vegetación, suelo desnudo, haciéndolo ampliamente utilizado en estudios de sensado remoto. En este ámbito, se utiliza el NBEI como una medida comparativa que permite evaluar la consistencia y la precisión de las predicciones generadas por la U-Net.

Con esto, se busca analizar la eficiencia computacional y la robustez del modelo en distintos escenarios de zonas urbanas. A partir de esto, se busca no solo demostrar las capacidades que tiene la U-Net para llevar a cabo la segmentación de la construcción en imágenes satelitales de WV-2, sino también establecer en cierta forma una metodología reproducible que integre técnicas de aprendizaje profundo junto con indicadores espectrales basados en índices.

Bibliografía

- [1] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B.A. Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152:166–177, 2019.
- [2] P. Liu, X. Liu, and M. Liu. Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network. *Remote Sensing*, 11(7):830, 2019.
- [3] M. Guo, H. Liu, Y. Xu, and Y. Huang. Building extraction based on u-net with an attention block and multiple losses. *Remote Sensing*, 12(9):1400, 2020.
- [4] B.L. Upchurch and M.A. Goforth. Worldview-2: A new tool for high-resolution, 8-band multispectral analysis. Technical paper, DigitalGlobe, 2010.
- [5] C. Tarantino, M. Adamo, F. Lovergine, P. Blonda, and V. Tomaselli. 8-band image data processing of the worldview-2 satellite in a wide area of applications. *Earth Observation*, 2012.
- [6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv:1505.04597*, 2015.
- [7] Adeniyi Adeyemi, Abel Ramoelo, Moses Cho, and Cecilia Masemola. Spectral index to improve the extraction of built-up area from worldview-2 imagery. *Journal of Applied Remote Sensing*, 15(2):024510, 2021.

Capítulo 4

Resultados de segmentación de superficies urbanas

4.1. Introducción

El presente Capítulo tiene como objetivo exponer el diseño experimental, la metodología aplicada y los resultados obtenidos en la evaluación de índices espectrales para la extracción de áreas construidas a partir de imágenes multiespectrales. Con este fin, se utilizaron datos satelitales del sensor WorldView-2, seleccionando distintas regiones de interés que representan escenarios urbanos con características heterogéneas en términos de densidad de construcción, presencia de suelo desnudo, cuerpos de agua y vegetación.

El análisis experimental se centra en la comparación de los índices BAI, BSI, NBEI y RGI, además de la propuesta metodológica basada en una máscara dinámica. Para cada caso, se construyen mapas binarios de áreas urbanas y se evalúa su precisión mediante métricas cuantitativas. Asimismo, se discuten las ventajas y limitaciones observadas en el desempeño de los índices frente a superficies espectralmente similares.

Finalmente, se presentan los resultados de densidad de construcción calculados a partir de los mapas generados, los cuales permiten cuantificar la distribución espacial del desarrollo urbano en las zonas de estudio. Estos hallazgos constituyen la base para la validación de la propuesta metodológica y su comparación con enfoques previamente reportados en la literatura.

4.2. Resultados basados en índices

Los cuatro índices espectrales —**BAI** (Ecuación2.1), **BSI** (Ecuación2.2), **NBEI** (Ecuación2.3) y **RGI** (Ecuación2.4)— fueron aplicados de forma independiente a cuatro escenas urbanas correspondientes a Regiones de Interés (RoIs) en las ciudades de **Atenas**, **Bucarest**, **Roma** y **Turín**. Cada índice fue calculado utilizando su formulación espectral respectiva, tal como se describió en las secciones anteriores. Los mapas resultantes fueron binarizados utilizando umbrales optimizados para extraer superficies construidas, minimizando al mismo tiempo la clasificación errónea del suelo desnudo.

Se realizó una evaluación cuantitativa mediante la selección manual de **puntos de referencia** para cuatro clases dominantes de cobertura superficial: vegetación, agua, suelo desnudo y zonas construidas. La alta resolución espacial de las imágenes de WorldView-2 permitió una identificación y etiquetado precisos de estos puntos de referencia mediante interpretación visual. Se recopilaron un total de 722 puntos para las clases de fondo (agua, vegetación y suelo desnudo), y 712 puntos para las superficies construidas. Esta distribución equilibrada garantiza una comparación justa entre las clases construidas y no construidas, reduciendo el sesgo y mejorando la fiabilidad de las métricas de precisión y discriminación espectral calculadas. La Tabla ?? muestra el número de muestras de referencia recolectadas por clase de superficie en cada ciudad.

Se generaron histogramas acumulativos para cada índice, basados en las muestras de referencia de las cuatro clases. Para cada índice, se seleccionó un rango óptimo de **umbrales de áreas construidas** con base en las regiones donde la frecuencia acumulada de las muestras construidas se maximiza y la del suelo desnudo se minimiza. La Figura 4.1 muestra los histogramas acumulativos y los umbrales seleccionados para cada índice. Los resultados de clasificación en las cuatro RoIs permiten una evaluación cualitativa de los métodos implementados basados en índices espectrales.

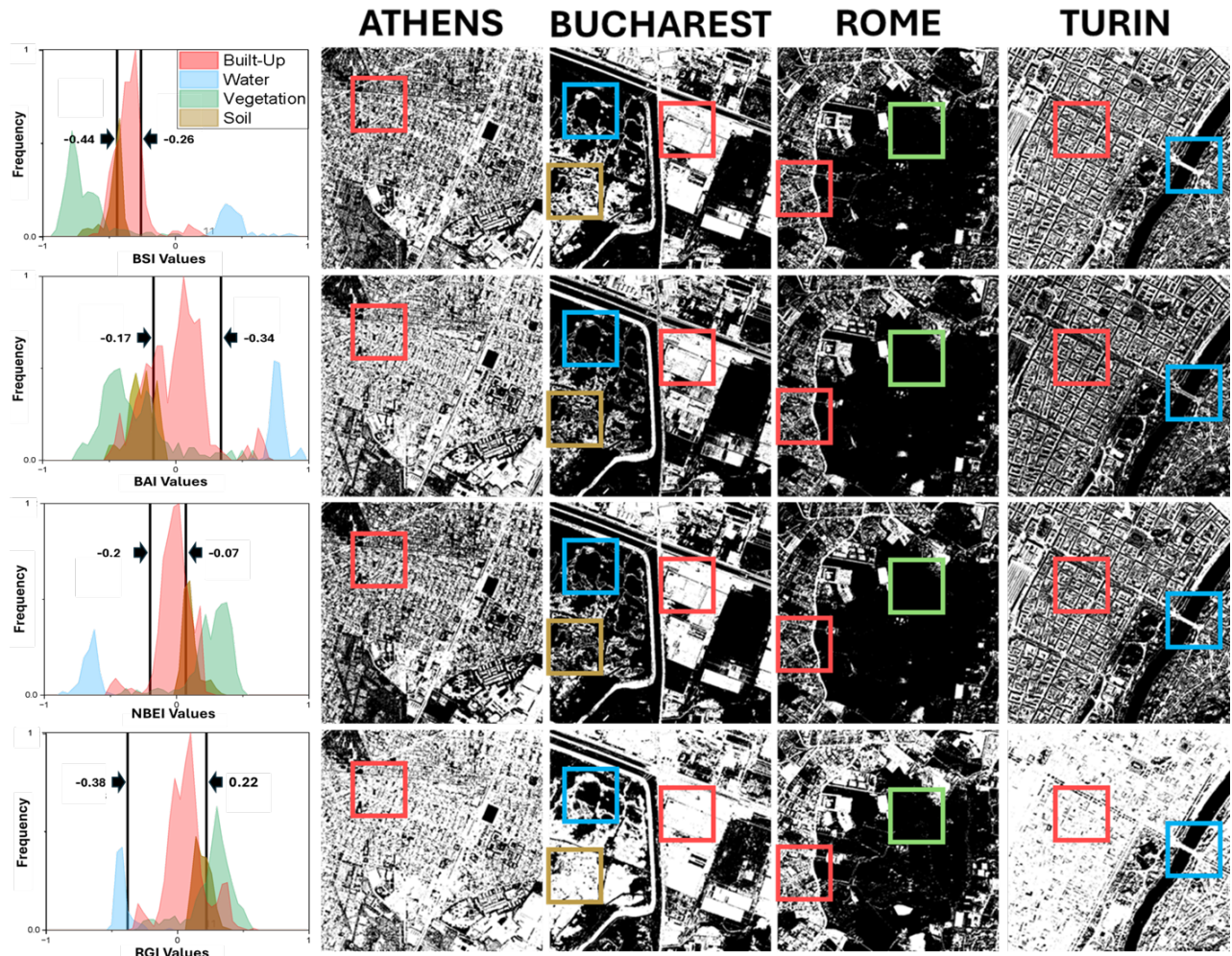


Figura 4.1: Mapas binarios de áreas construidas obtenidos a partir de imágenes multiespectrales correspondientes a las cuatro zonas de estudio: Turín, Roma, Atenas y Bucarest. Los píxeles blancos indican zonas construidas; los píxeles negros indican fondo. De arriba hacia abajo: mapa **BSI**, mapa **BAI**, mapa **NBEI** y mapa **RGI**.

Para evaluar el rendimiento de clasificación, se calculó la **precisión (AC)** para cada índice y RoI utilizando la siguiente expresión:

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

donde TP , TN , FP y FN representan el número de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, respectivamente, con base en los puntos de referencia.

Adicionalmente, para evaluar la separabilidad espectral entre las clases de **áreas construidas** y **suelo desnudo**, se calculó el **Índice de Discriminación Espectral (SDI)** mediante

la siguiente ecuación:

$$SDI = \frac{|\mu_1 - \mu_2|}{\sigma_1 + \sigma_2} \quad (4.2)$$

donde μ_1 , μ_2 y σ_1 , σ_2 son las medias y desviaciones estándar de los valores del índice para las clases de área construida y suelo desnudo, respectivamente. Un valor más alto de SDI indica una mayor separación espectral y, por tanto, un menor riesgo de clasificación errónea entre estas dos clases críticas.

La Tabla 4.1 resume el desempeño general de cada índice espectral en todas las zonas de estudio. Los resultados incluyen tanto la precisión binaria de clasificación como el SDI, que cuantifica la separabilidad espectral entre las áreas construidas y el suelo desnudo. Entre los índices evaluados, el NBEI alcanzó la mayor precisión (63.40 %) y la mejor discriminación espectral (SDI = 4.40), lo que refleja su robustez para distinguir superficies impermeables de materiales espectralmente similares.

El índice BAI mostró un rendimiento moderado en ambas métricas, mientras que el BSI arrojó un valor bajo de SDI (0.08), lo que indica una separación débil entre características construidas y suelo desnudo. El RGI presentó una precisión limitada, pero un SDI relativamente alto (2.19), lo que sugiere que, aunque su clasificación sea menos precisa, su respuesta espectral frente a materiales construidos es distintiva bajo ciertas condiciones. Estos hallazgos confirman los distintos grados de efectividad de cada índice, según su formulación espectral y sensibilidad ante la heterogeneidad de superficies urbanas.

Tabla 4.1: Precisión general (%) e Índice de Discriminación Espectral (SDI) para cada índice espectral en todas las zonas de estudio.

Índice	Precisión	SDI
BAI	61.43 %	0.97
BSI	60.91 %	0.08
NBEI	63.40 %	4.4
RGI	54.10 %	2.19
MBEI (propuesta)	85.11 %	XX

4.3. Resultados basados en Aprendizaje Profundo

La eficiencia en la clasificación de la red neuronal convolucional (CNN) propuesta se evaluó utilizando una región de interés (ROI) en Turín. La CNN se entrenó con imágenes satelitales WorldView-2 (banda NIR1) procesadas en bloques de 128×128 píxeles. Las predicciones de la red se compararon con los resultados del Índice de Extracción de Áreas Urbanizadas (NBEI) para evaluar su capacidad de distinguir entre áreas urbanizadas y no urbanizadas.

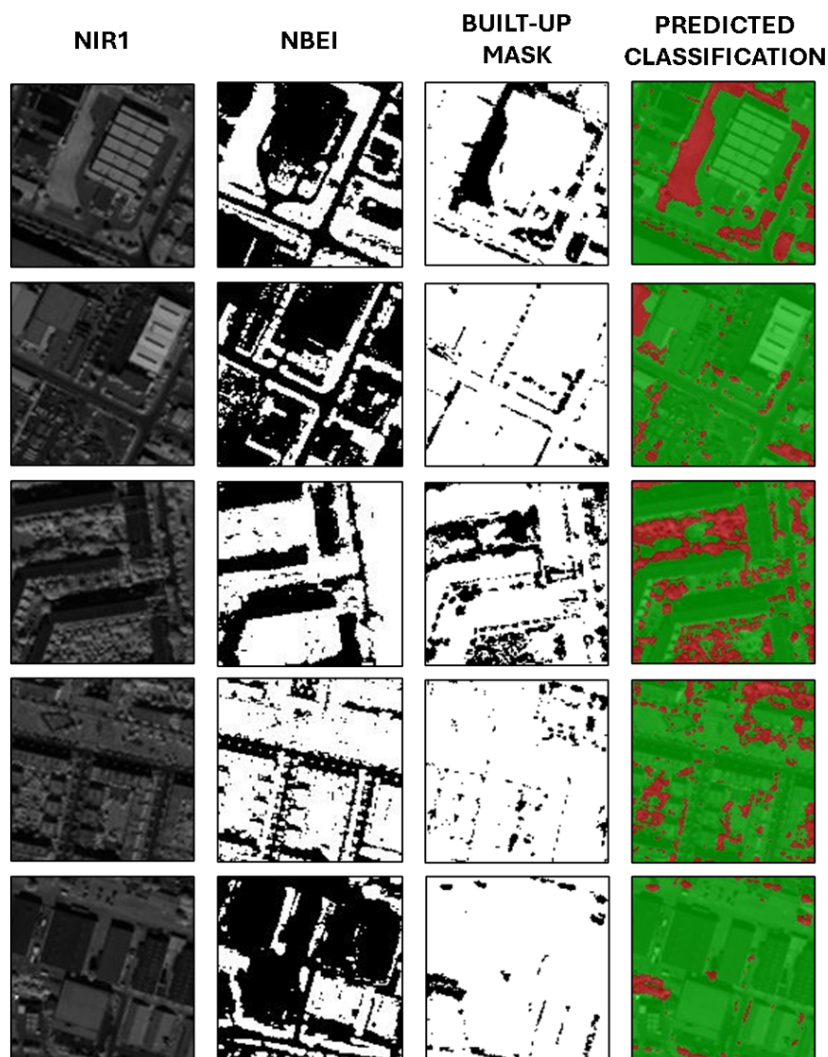


Figura 4.2: Comparación cualitativa de cinco muestras del conjunto de datos de Turín. Las columnas muestran: (a) imagen NIR1, (b) índice NBEI, (c) mapa de probabilidad de la red neuronal convolucional (CNN), (d) clasificación binaria de la CNN y (e) clasificación de la CNN superpuesta a la imagen NIR1 (verde: zonas urbanizadas, rojo: fondo).

La Figura 4.2 muestra los resultados cualitativos para cinco bloques de prueba representativos. Cada fila incluye: (a) el bloque de entrada en la banda NIR1, (b) el mapa NBEI obtenido a partir de los datos multiespectrales, (c) el mapa de probabilidad de área urbanizada de la CNN, (d) la máscara de clasificación binaria de la CNN y (e) la superposición visual de la clasificación predicha sobre la imagen NIR1, donde las áreas urbanizadas se muestran en verde y el resto en rojo. Estos ejemplos ilustran la capacidad de la CNN para reducir los falsos positivos, los cuales son comunes en los índices espectrales, además de preservar los detalles espaciales, especialmente en los límites de los edificios.

4.4. Densidad de construcción

En la Figura 4.3, se muestra el cálculo de la densidad de construcción por sector.

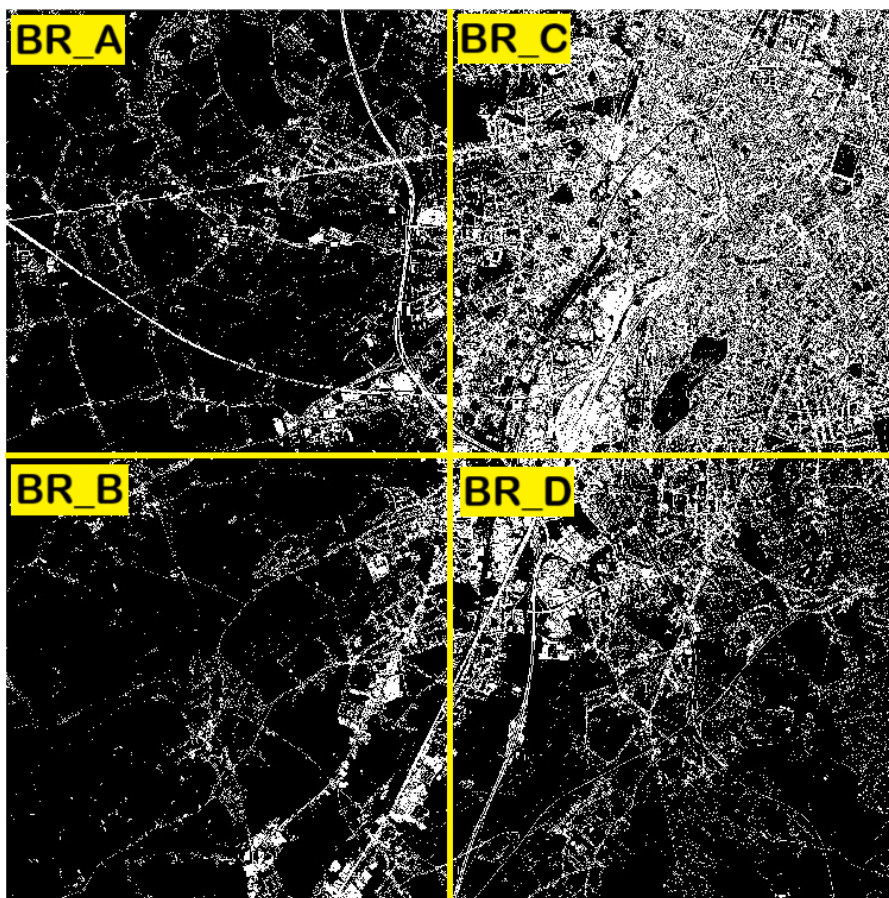


Figura 4.3: Resultados de extracción de construcción usando el método propuesto de enmascaramiento. Como se puede observar, la información de suelo desnudo fue eliminada. Se calcularon los valores de densidad de construcción por sectores suburbanos $\mathcal{B}_A = 0.13$, $\mathcal{B}_B = 0.11$, $\mathcal{B}_C = 0.63$, y $\mathcal{B}_D = 0.23$.

4.5. Conclusiones

En este Capítulo se evalúan dos enfoques para la clasificación de áreas urbanas a partir de imágenes provenientes del satélite WorldView-2:

- Segmentación basada en índices espectrales
- Segmentación semántica usando una arquitectura U-net

Se evaluaron cuatro índices de segmentación de áreas urbanas presentes en la literatura. Como se observa en la Tabla 4.1, el índice propuesto *MBEI* obtuvo un rendimiento superior. Aunado a esto, la implementación y los recursos computacionales son menores respecto a los requeridos usando un modelo neuronal.

Además, se implementó una arquitectura de red neuronal convolucional (CNN) para la segmentación de áreas urbanas. Cuantitativamente, se alcanzó una precisión global del 85,11 %, calculada mediante la comparación a nivel de píxel con datos de referencia etiquetados manualmente. A pesar de que la Red Neuronal basada en U-Net demostró una gran capacidad para capturar la distribución espacial de las características urbanas, la aplicación práctica de estas CNNs aún plantea retos concretos como: 1) La transferencia de los modelos entre ciudades con distintos materiales empleados en la construcción, 2) La presencia de sombras u objetos que causan oclusión en los edificios y 3) La necesidad de etiquetas suficientes para la etapa del entrenamiento supervisado.

Como trabajo a futuro, se evaluará la eficiencia del modelo para identificar áreas urbanas en imágenes que contienen múltiples superficies como agua, suelo desnudo, construcción y vegetación.

Capítulo 5

Conclusiones y Trabajo a Futuro

Este proyecto de Tesis propone un Índice Enmascarado de Extracción de Construcción (MBEI) que detecta construcción en imágenes satelitales de ciudades con múltiples superficies como: vegetación, agua, construcción y suelo desnudo. La cobertura urbana incluye áreas de uso residencial, así como industrial y comercial que conforman el paisaje desarrollado o construido.

Para mejorar el algoritmo de extracción de áreas construidas, los rangos de umbral de la máscara de construcción en la Ecuación (2.6) y del mapa binarizado de *NBEI* en la Ecuación (2.7) se optimizan para maximizar la diferencia entre píxeles urbanos y de suelo desnudo en el mapa propuesto *MBEI* de la Ecuación (2.8).

Una estrategia eficaz para mitigar esta confusión consiste en aplicar máscaras específicas para suelo desnudo, de modo que se excluyan los píxeles correspondientes a este tipo de superficie antes de aplicar los índices urbanos. Esta práctica permite reducir significativamente las sobreestimaciones de áreas construidas, al favorecer una segmentación más precisa basada en características espectrales y espaciales adicionales, como el patrón o textura del suelo.

Se compararon cuatro índices espectrales (BAI, BSI, NBEI, RGI) para la detección de áreas urbanizadas, utilizando imágenes multispectrales de alta resolución de WorldView-2. Estos índices se aplicaron a cuatro regiones urbanas con características espaciales y espectrales diferentes. Los resultados confirman que los métodos basados en índices siguen siendo una solución práctica y económica para el mapeo de superficies impermeables, especialmente cuando se requiere simplicidad computacional y facilidad de interpretación.

Entre los índices analizados, el NBEI demostró la mayor precisión general y la mejor capacidad para diferenciar entre áreas urbanizadas y suelo desnudo. El NBEI utiliza de forma equilibrada las bandas del infrarrojo cercano y el borde del visible, lo que garantiza un rendimiento constante en diversas estructuras y composiciones urbanas. Los índices BSI y BAI

ofrecen una precisión moderada, pero son menos adecuados para diferenciar el suelo desnudo, sobre todo en áreas con poca vegetación. El RGI, aunque menos preciso en la clasificación, muestra un buen potencial para captar respuestas espectrales específicas de diferentes tipos de construcciones. El análisis confirma que la similitud espectral entre las superficies urbanizadas y el suelo desnudo sigue siendo una limitación clave en los métodos de detección basados en índices. Sin embargo, al utilizar umbrales y métricas de discriminación adecuadas, como el SDI, estos métodos proporcionan resultados robustos y generalizables. En futuras investigaciones, se podría explorar la integración de características espaciales, textura o clasificadores de aprendizaje automático para mejorar aún más la precisión en la delimitación urbana.

A manera de comparativa, se implementó una red neuronal convolucional con una arquitectura U-Net modificada para la extracción de áreas urbanizadas a partir de imágenes WorldView-2 de alta resolución, en diversos entornos urbanos, con el objetivo de realizar la segmentación semántica de imágenes. El modelo demostró su capacidad para aprender patrones relevantes mediante un proceso de entrenamiento supervisado, alcanzando una precisión del 85,11 % en la segmentación de imágenes en escala de grises. La arquitectura propuesta mostró una buena capacidad de generalización, gracias a las conexiones directas que permiten recuperar detalles espaciales precisos. Además, la binarización previa de los datos facilitó el aprendizaje al simplificar las clases (áreas construidas y no construidas), lo que la hace idónea para el análisis y la monitorización urbana a gran escala en estudios que requieren una segmentación binaria precisa.

De manera general, los métodos basados en índices son capaces de discriminar áreas urbanas de áreas con vegetación, suelo desnudo y agua. Se logró un porcentaje de clasificación correcta de 88.32 % usando el método propuesto. Además de ser fáciles de implementar y requieren menos esfuerzo computacional.

Como trabajo futuro, proponemos incorporar índices de segmentación para facilitar la segmentación de los datos de entrada y mejorar la precisión, así como ampliar la base de datos de imágenes satelitales.

Apéndice A

Trabajos derivados de la Tesis

En este proyecto de Tesis se tienen las siguientes aportaciones:

1. **Comparison of built-up indices for urban features extraction.** A. Aguilar-Vallejo, R. Castro-Ortega, A. Delgadillo-Jiménez, C. Toxqui-Quitl, A. Padilla-Vivanco, A. Carbone. En: *Current Developments in Lens Design and Optical Engineering XXVI*, editado por Simon Thibault, Ching-Cherng Sun, Jessica DeGroot Nelson. **Proc. SPIE Vol. 13596, 135960T**. Publicado por SPIE, 2025. doi: [10.1117/12.3064061](https://doi.org/10.1117/12.3064061).
2. **Índice espectral para la delimitación de suelo desnudo usando imágenes de alta resolución.** A. Aguilar-Vallejo, A. Delgadillo-Jiménez, R. Castro-Ortega, A. Padilla-Vivanco, A. Carbone, C. Toxqui-Quitl. *Presentado en formato plática en la sesión de Óptica, LXVIII Congreso Nacional de Física*, Sociedad Mexicana de Física, 14 de octubre de 2025, Salón 2, Puebla, México. Clave única: LXVIII-010882.
3. **Monitoreo de la densidad de la construcción en ciudades usando imágenes multiespectrales.** A. Aguilar-Vallejo, R. Castro-Ortega, A. Delgadillo-Jiménez, C. Toxqui-Quitl, A. Padilla-Vivanco, A. Carbone. *Presentado en formato póster en el Mexican Optics and Photonics Meeting (MOPM 2025)*, Academia Mexicana de Óptica, 1–3 de octubre de 2025, Biblioteca Raúl Rangel Frías, Universidad Autónoma de Nuevo León, Monterrey, México. Referencia: AMO/MOPM2025-ACP-033.
4. **Estimation of urban fractal dimension using a Convolutional Neural Network.** A. Delgadillo, C. Toxqui-Quitl, R. Castro-Ortega, A. Aguilar, A. Padilla-Vivanco, A. Carbone (2025). *Lectures Notes in Artificial Intelligence, MICAI 2025, LNAI 16221*, pp. 1–12, L. Martínez-Villaseñor et al. (Eds.), Springer Nature Switzerland AG; <https://doi.org/10.1007/978-3-032-09037-9-31>

Comparison of built-up indices for urban features extraction

A. Aguilar-Vallejo, R. Castro-Ortega, A. Delgadillo-Jiménez, C. Toxqui-Quitl, A. Padilla-Vivanco^a and A. Carbone^b

^aComputer Vision Laboratory. Universidad Politécnica de Tulancingo, Ingenierías 100, 43629, Hidalgo, México.

^bDISAT, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Turin, 10129, Italy

ABSTRACT

A comparative evaluation of built-up indices applied to high-resolution multispectral imagery from the WorldView-2 satellite is presented. Four metropolitan areas are analyzed: Athens, Bucharest, Rome, and Turin. Each index is tested on different land covers, such as vegetation, water, bare soil, and built-up. A binarization strategy based on cumulative histograms of ground-truth samples is implemented. Performance was assessed using the Spectral Discrimination Index (SDI). The results indicate that NBEI outperformed the other indices in separating impervious surfaces from spectrally similar bare soil. The findings confirm the potential of index-based approaches for built-up extraction in high-resolution imagery, while also highlighting their limitations in heterogeneous urban environments.

Keywords: WorldView-2, Urban Segmentation, Spectral Indices, Urban Land Cover, Built-Up Index.

1. INTRODUCTION

Monitoring the spatial extent and growth of built-up areas is essential for sustainable urban planning, environmental management, and infrastructure development. High resolution satellite imagery has become a valuable tool for mapping impervious surfaces, allowing timely assessments of changes in land cover in rapidly urbanizing regions [1, 2]. Among the various techniques available for urban feature extraction, spectral index-based methods are still widely used because of their low computational cost, ease of implementation, and adaptability across sensors. These methods rely on algebraic combinations of multispectral bands to enhance the spectral contrast between artificial structures and other types of land cover. Despite their simplicity, they often face limitations in accurately separating built-up surfaces from spectrally similar classes such as bare soil or dry vegetation, particularly in heterogeneous urban environments [3]. To address this challenge, several indices have been proposed that incorporate different spectral regions of high-resolution sensors like WorldView-2. Recent studies have demonstrated the utility of built-up indices in improving the mapping of impervious surfaces, particularly when combined with careful thresholding and statistical validation frameworks [4].

This study evaluates four index-based methods: the Built-up Area Index (BAI) [5], Built-up Spectral Index (BSI) [6], New Built-up Extraction Index (NBEI) [7], and Red Edge-Green Index (RGI) [8]. The four indexes have been selected based on their relevance in the recent literature and their compatibility with WorldView-2 spectral properties. Comparative performance in extracting impervious surfaces is shown using a consistent and quantitative evaluation framework in four cities (Athens, Bucharest, Rome, and Turin). The analysis includes accuracy metrics and spectral discrimination indicators to assess the ability of each index to distinguish between built-up that includes asphalted roads, buildings, concrete and other manmade impervious surfaces, and background surfaces, particularly bare soil, water, and vegetation.

Further author information: (Send correspondence to R. Castro-Ortega.)
R. Castro-Ortega.: E-mail: raul.castro@upt.edu.mx, Telephone: +52 (776) 76-76-366

2. MATERIALS AND METHODS

2.1 Satellite Data

The test data set is made up of high-resolution multispectral imagery acquired by the WorldView-2 (WV-2) satellite launched by DigitalGlobe in 2009. WV-2 was among the first high-resolution commercial satellites to provide eight-band multispectral data in the visible and near-infrared (VNIR) spectrum at a spatial resolution of 1.6 meters alongside a 0.46-meter panchromatic band. Its unique spectral configuration enhances the discrimination of urban features such as impervious surfaces, vegetation, and bare soil, which is critical for urban mapping applications [9]. Table 1 summarizes the spectral bands and their wavelength ranges.

Table 1. Spectral bands from the WorldView-2 satellite.

Band	Name	Wavelength (nm)
B1	Coastal Blue	400–450
B2	Blue	450–510
B3	Green	510–580
B4	Yellow	585–625
B5	Red	630–690
B6	Red Edge	705–745
B7	NIR1	770–895
B8	NIR2	860–1040

The dataset contains urban scenes from four European cities: Athens, Bucharest, Rome, and Turin. For each city, a Region of Interest (RoI) of 939×939 pixels was selected - corresponding to approximately 1.5×1.5 km of ground area - to ensure spatial consistency between cities with diverse urban morphologies. These cities were chosen for their distinct architectural styles, land cover heterogeneity, and urban density gradients, which provide a comprehensive testbed for spectral index evaluation. All imagery was preprocessed through orthorectification and atmospheric correction to ensure geometric and radiometric consistency. Only the eight multispectral bands were used in this work. The panchromatic band was excluded. These multispectral subsets are particularly well-suited for evaluating built-up indices, as demonstrated in prior urban remote sensing studies [10].

2.2 Index-Based Methods

Index-based methods are widely adopted in remote sensing to extract impervious surfaces from multispectral imagery. These methods use computational combinations of spectral bands to emphasize specific land cover types, particularly built-up areas, by exploiting their distinct reflectance properties. Their simplicity, computational efficiency, and adaptability to different sensors and resolutions make them a common choice for large-scale and multi-period urban analysis [4, 11, 12]. However, a common problem of spectral indices for urban mapping is the spectral overlap between built-up areas and bare soil. Both surface types often exhibit similar reflectance in the visible and near-infrared ranges, particularly under dry conditions or in sparsely vegetated environments. This spectral overlap may result in overestimates of urban extents and reduced classification accuracy [13]. A comparative assessment of built-up index-based methods is conducted to evaluate their performance on cities with varying urban morphologies and geographical conditions.

2.2.1 Built-up Area Index (BAI)

The Built-up Area Index (BAI) is a simple yet effective spectral index designed to enhance the detection of artificial surfaces, particularly asphalt and concrete, in high-resolution imagery. It leverages the contrast in reflectance between the near-infrared and blue bands based on the characteristic reflectance of most built-up materials. The BAI was initially proposed to support the object-oriented classification of urban roads but has since demonstrated broader applicability in delineating impervious surfaces in dense urban fabrics [5]. Its mathematical formulation is given as:

$$\text{BAI} = \frac{\text{NIR} - \text{Blue}}{\text{NIR} + \text{Blue}} \quad (1)$$

Figure 1 shows the cumulative histogram for BAI, and the maps of BAI values across the four cities.

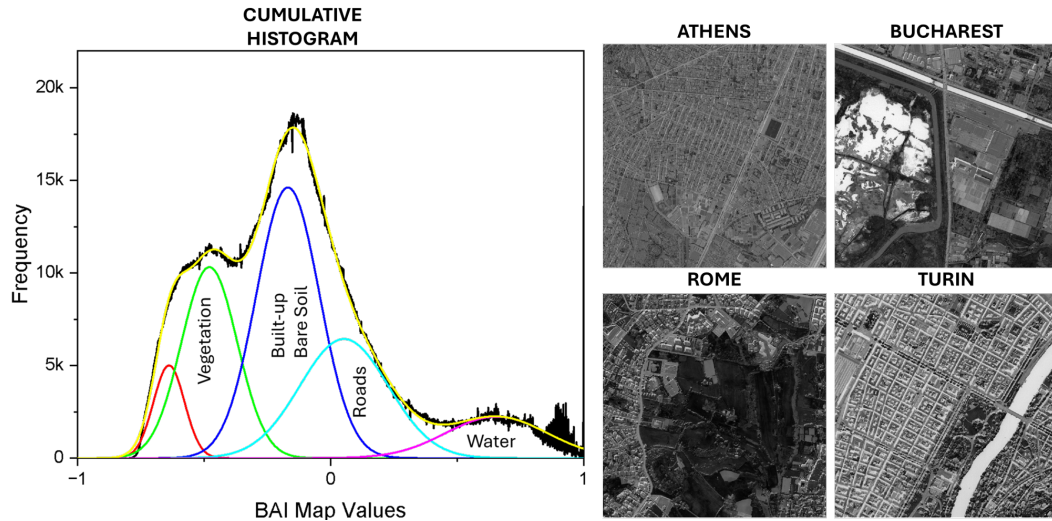


Figure 1. Cumulative histogram and spatial distribution of the **BAI** across the four urban scenes. The histogram shows most surfaces below values of zero, effectively separating water, but failing to distinguish completely vegetation from built-up areas and confusing built-up with bare soil and roads.

2.2.2 Built-up Spectral Index (BSI)

The Built-up Spectral Index (BSI) was designed to improve the extraction of impervious surfaces from WorldView-2 data. Unlike conventional indices derived from heuristic band combinations, BSI was developed using a multi-objective particle swarm optimization algorithm. This technique identified the most relevant spectral bands and their optimal weights to enhance spectral separability between built-up and non-built-up surfaces [6].

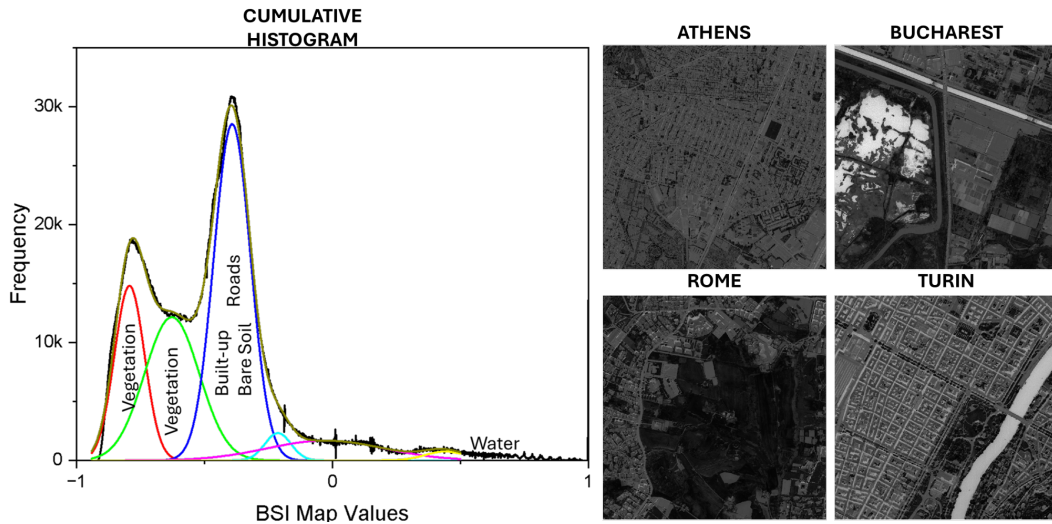


Figure 2. Cumulative histogram and spatial distribution of the **BSI** across the four urban scenes. Similar to BAI, the histogram displays most surfaces below zero, effectively isolating water. It offers improved separation between vegetation and built-up areas but still confuses built-up with bare soil and roads.

The final BSI formulation is expressed as:

$$BSI = \frac{Yellow - 2 \times NIR1}{Yellow + 2 \times NIR1} \quad (2)$$

Figure 2 shows the cumulative histogram for BSI, and the maps of BSI values across the four cities.

2.2.3 New Built-up Extraction Index (NBEI)

The New Built-up Extraction Index (NBEI) was developed to address the limitations of traditional spectral indices when applied to high-resolution imagery such as WorldView-2. NBEI enhances the spectral contrast between impervious surfaces and surrounding land covers by combining the most discriminative bands identified through ReliefF feature selection: NIR2, NIR1, Green, and Red Edge. Its formulation is based on the normalized difference between the sum of the near-infrared (reflective) and visible-edge (absorptive) bands, which increases separability among heterogeneous urban materials [7]. The index is expressed as:

$$NBEI = \frac{(NIR2 + NIR1) - (Green + RedEdge)}{(NIR2 + NIR1) + (Green + RedEdge)} \quad (3)$$

Figure 3 shows the cumulative histogram for NBEI, and the maps of NBEI values across the four cities.

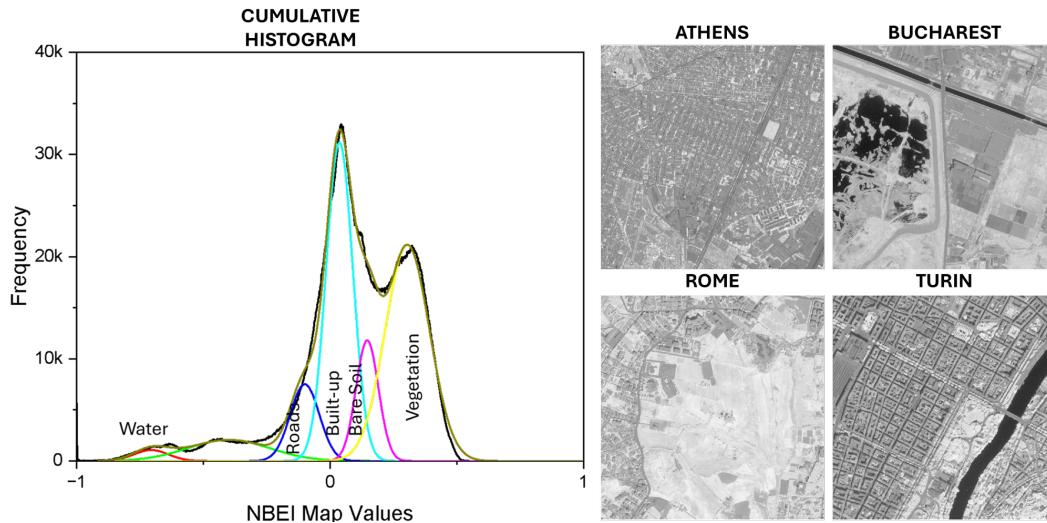


Figure 3. Cumulative histogram and spatial distribution of the NBEI across the four urban scenes. Built-up values cluster around zero, vegetation appears at higher values, and water near -1. Like other indices, built-up areas are often confused with bare soil.

2.2.4 Red Edge - Green Index (RGI)

The Red Edge-Green Index (RGI) is a spectral index designed to enhance the discrimination between impervious surfaces and bare soil in urban and peri-urban environments. It uses the Red Edge and Green bands of the WorldView-2 sensor to capture subtle differences in surface reflectance [8]. The index is computed as the normalized difference between the Red Edge and Green bands:

$$RGI = \frac{RedEdge - Green}{RedEdge + Green} \quad (4)$$

Figure 4 shows the cumulative histogram of RGI and the maps of RGI values in the four cities.

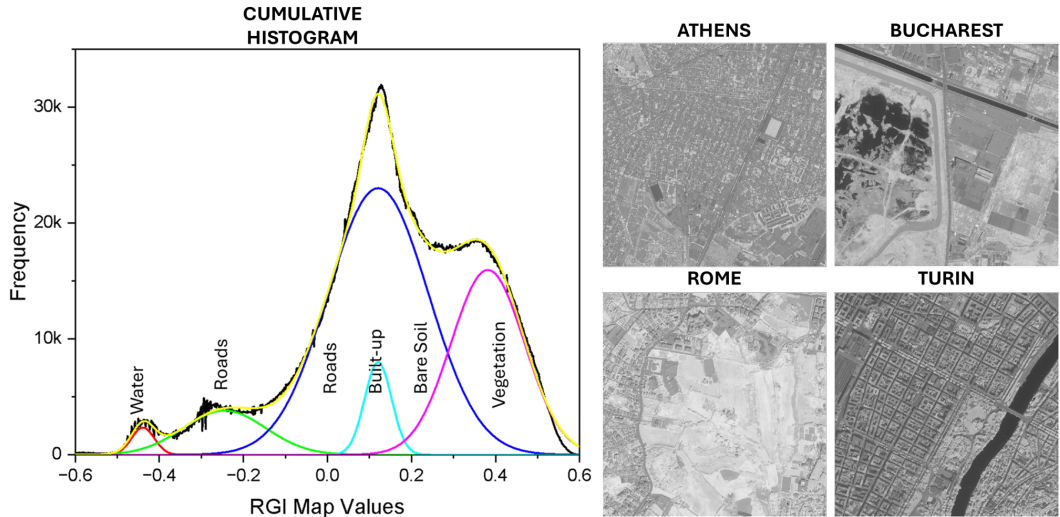


Figure 4. Cumulative histogram and spatial distribution of the **RGI** across the four urban scenes. Built-up values clusters between -0.2 and 0.4, vegetation appears between 0.2 and 0.6, and water below -0.4. Like other indices, built-up areas are often confused with bare soil.

3. BUILT-UP EXTRACTION RESULTS

The four spectral indices—**BAI** (Eq. 1), **BSI** (Eq. 2), **NBEI** (Eq. 3), and **RGI** (Eq. 4)— have been applied independently to four urban scenes corresponding to Regions of Interest (RoIs) in **Athens**, and **Bucharest**, **Rome**, **Turin**. Each index was computed using its respective spectral formulation, as described in the previous sections. The resulting index maps were binarized by using thresholds, optimized to extract built-up surfaces while minimizing the misclassification of bare soil.

A quantitative evaluation was conducted using manually selected **ground truth points** for four dominant surface classes: vegetation, water, bare soil, and built-up areas. The high spatial resolution of the WorldView-2 imagery facilitated accurate identification and labeling of these reference points through visual interpretation. A total of 722 points were collected for background classes (water, vegetation, and bare soil), and 712 points were collected for built-up surfaces. This balanced distribution ensures a fair comparison between built-up and non-built classes, reducing bias and enhancing reliability, computed accuracy and spectral discrimination metrics. Table 2 lists the number of ground-truth samples collected per surface class in each city.

Table 2. Number of ground truth points per surface type in each study area.

City	Vegetation	Water	Bare Soil	Built-up
Athens	115	0	56	210
Bucharest	22	52	54	127
Rome	75	0	79	141
Turin	179	69	21	234
Total	391	121	210	712

Cumulative histograms were generated for each index, based on the ground truth samples of the four classes. For each index, an optimal **Built-up threshold range** was selected based on regions where the cumulative frequency of built-up samples was maximized, and bare soil was minimized. Figure 5 shows the cumulative histograms and built-up thresholds for each index. Classification results across the four RoIs allow a qualitative evaluation of the implemented built-up index-based methods.

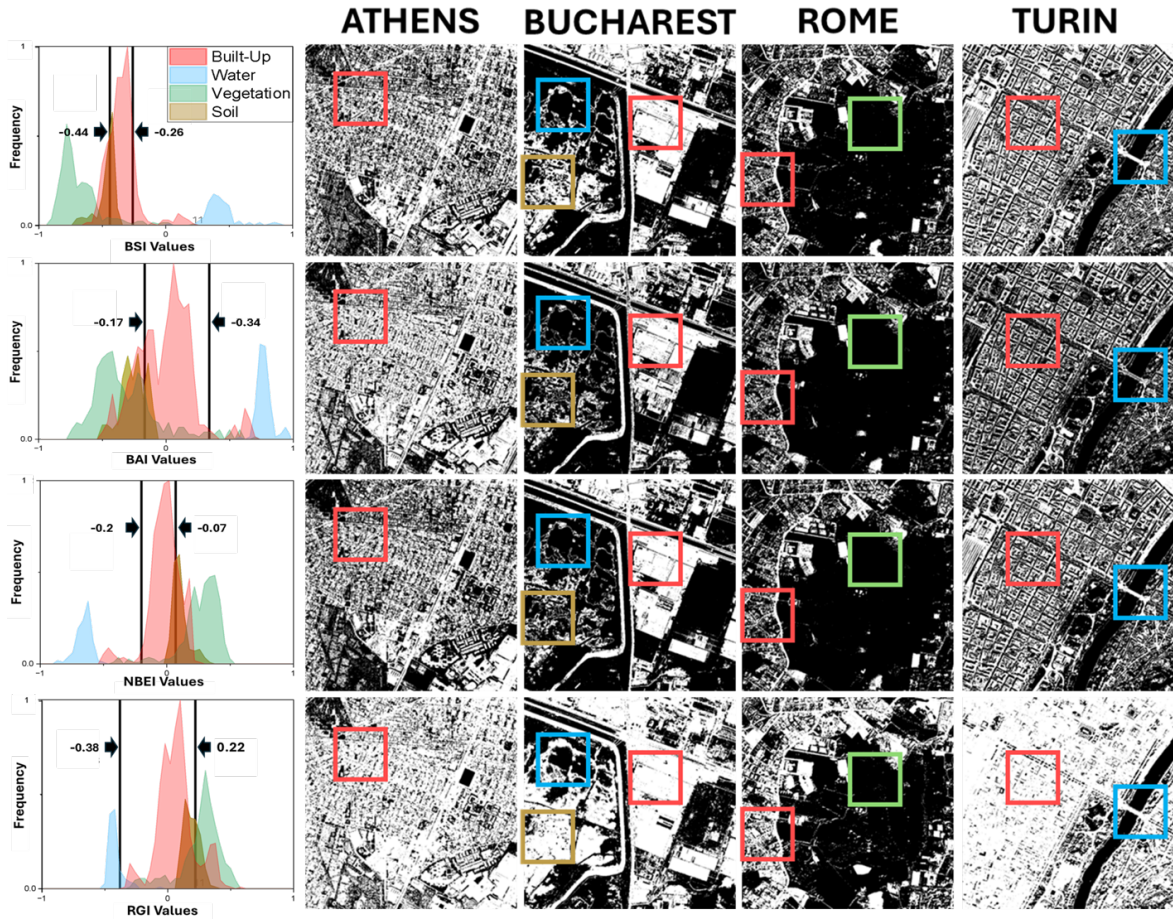


Figure 5. Binary built-up maps obtained from multispectral imagery corresponding to the four study areas: Turin, Rome, Athens, and Bucharest. White pixels indicate a built-up area; black pixels indicate a background. From top to down **BSI** map, **BAI** map, **NBEI** map, and **RGI** map.

To evaluate classification performance, the **accuracy (AC)** was computed for each index and RoI using the following expression:

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where TP , TN , FP , and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively, based on the validation ground truth.

Additionally, to assess the spectral separability between **Built-up** and **bare soil**, the **Spectral Discrimination Index (SDI)** was calculated as:

$$SDI = \frac{|\mu_1 - \mu_2|}{\sigma_1 + \sigma_2} \quad (6)$$

where μ_1 , μ_2 and σ_1 , σ_2 are the means and standard deviations of the index values for Built-up and bare soil classes, respectively. A higher SDI indicates greater spectral separation, and therefore, a lower risk of

misclassification between these two critical classes.

Table 3 summarizes the overall classification performance of each spectral index across all study areas. The results include both the binary classification accuracy and the Spectral Discrimination Index (SDI), which quantifies the spectral separability between Built-up and bare soil classes. Among the indices evaluated, the NBEI achieved the highest accuracy (63.40 %) and the best spectral discrimination (SDI = 4.40), reflecting its robustness in distinguishing impervious surfaces from spectrally similar materials. The BAI showed moderate performance in both metrics, while the BSI yielded a low SDI value (0.08), indicating a weak separation between built-up and bare soil features. The RGI demonstrated limited accuracy but a relatively high SDI (2.19), suggesting that although its classification output may lack precision, its spectral response to Built-up materials is distinct under certain conditions. These findings confirm the varying degrees of effectiveness of each index depending on their spectral formulation and sensitivity to urban surface heterogeneity.

Table 3. Overall Accuracy (%) and Spectral Discrimination Index (SDI) for each spectral index across all study areas.

Index	Accuracy (%)	SDI
BAI	61.43	0.97
BSI	60.91	0.08
NBEI	63.40	4.40
RGI	54.10	2.19

4. CONCLUSIONS

Four spectral indices—BAI, BSI, NBEI, RGI—for the extraction of built-up areas have been compared by using the WorldView-2 high-resolution multispectral imagery. The indexes have been applied to four urban regions with different spatial and spectral characteristics. The results confirm that index-based methods remain a practical and low-cost solution for impervious surface mapping, particularly when computational simplicity and interpretability are required. Among the investigated indices, the NBEI demonstrates highest overall accuracy and spectral separation between built-up and bare soil classes. The NBEI makes a balanced use of near-infrared and visible-edge bands thus providing consistent performance across diverse urban structure and composition. The BSI and BAI offer moderate accuracy but are less suitable to separate bare soil, especially in areas with low vegetation cover. The RGI, although less accurate in classification, shows good potential to capture distinct spectral responses from specific built-up structures. The analysis confirms that the spectral similarity between built-up and bare soil surfaces remains a key limitation in index-based extraction approaches. However, when appropriate thresholds and discrimination metrics such as SDI are adopted, these methods yield robust and transferable results. Future work may explore the integration of spatial features, texture, or machine learning classifiers to further improve urban delineation accuracy.

ACKNOWLEDGMENTS

A. Aguilar-Vallejo (CVU No. 541295) and A. Delgadillo-Jimenez (CVU No. 1323784) express gratitude to the Secretaría de Ciencias, Humanidades, Tecnología e Innovación (SECIHTI) for the Master's scholarship. The authors thank the European Space Agency (ESA) for granting access to the WorldView satellite imagery for the Politecnico di Turin. Financial support from the TED4LAT project under the WIDERA initiative of the Horizon Europe Programme (Grant Agreement No. 101079206) is also gratefully acknowledged.

References

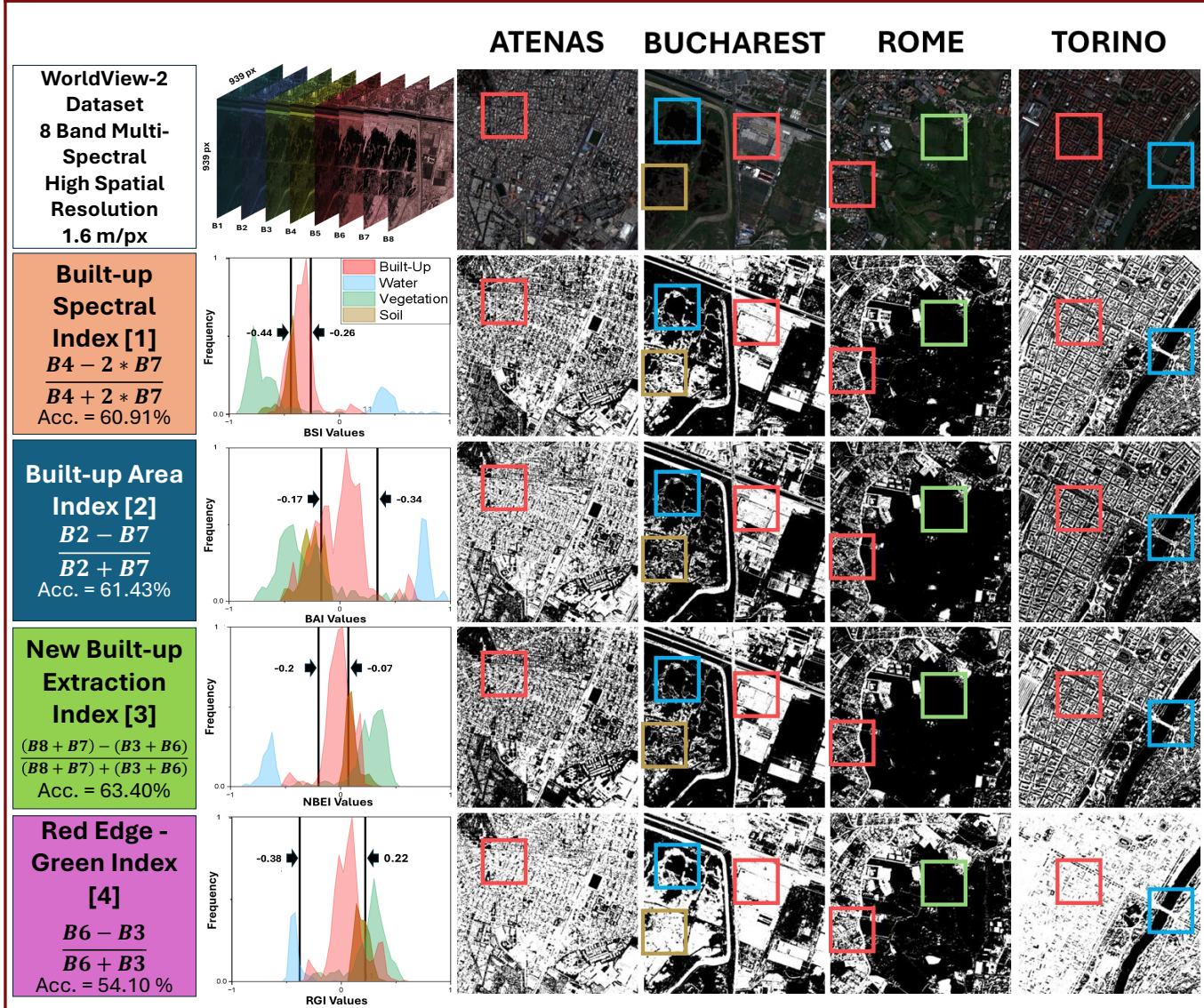
- [1] Isabel Georg, Thomas Blaschke, and Hannes Taubenböck. "Spatial delineation of urban corridors in North America: An approach incorporating fuzziness based on multi-source geospatial data". In: *Cities* 133 (2023), p. 104129. ISSN: 0264-2751. DOI: <https://doi.org/10.1016/j.cities.2022.104129>. URL: <https://www.sciencedirect.com/science/article/pii/S0264275122005686>.
- [2] Thomas Esch et al. "Urban footprint processor—Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission". In: *IEEE Geoscience and Remote Sensing Letters* 10.6 (2013), pp. 1617–1621. DOI: [10.1109/LGRS.2013.2272953](https://doi.org/10.1109/LGRS.2013.2272953).

- [3] Mohammad Karimi Firozjaei et al. “Automated Built-Up Extraction Index: A New Technique for Mapping Surface Built-Up Areas Using LANDSAT 8 OLI Imagery”. In: *Remote Sensing* 11.17 (2019). ISSN: 2072-4292. DOI: [10.3390/rs11171966](https://doi.org/10.3390/rs11171966). URL: <https://www.mdpi.com/2072-4292/11/17/1966>.
- [4] Yassine Harrak, Rachid Ahmed, and Rahim Aguejdad. “Evaluation of Spectral Indices and Global Thresholding Methods for the Automatic Extraction of Built-Up Areas: An Application to a Semi-Arid Climate Using Landsat 8 Imagery”. In: *Urban Science* 9 (Mar. 2025), pp. 1–26. DOI: [10.3390/urbansci9030078](https://doi.org/10.3390/urbansci9030078).
- [5] Paidamwoyo Mhangara et al. “Road extraction using object oriented classification”. In: *Proceedings of the 2015 ResearchGate Conference*. Available at: <https://www.researchgate.net/publication/267856733>. 2015.
- [6] Maher Ibrahim Sameen and Biswajeet Pradhan. “A novel built-up spectral index developed by using multiobjective particle-swarm-optimization technique”. In: *Proc. of the International Conference on Geomatic and Geospatial Technology (GGT)*. Conference Paper. 2016. URL: <https://www.researchgate.net/publication/301348545>.
- [7] Adeniyi Adeyemi et al. “Spectral index to improve the extraction of built-up area from WorldView-2 imagery”. In: *Journal of Applied Remote Sensing* 15.2 (2021), p. 024510. DOI: [10.1117/1.JRS.15.024510](https://doi.org/10.1117/1.JRS.15.024510).
- [8] Mariana Belgiu, Lucian Drăguț, and Josef Strobl. “Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using WorldView-2 imagery”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 87 (2014), pp. 205–215. ISSN: 0924-2716. DOI: <https://doi.org/10.1016/j.isprsjprs.2013.11.007>. URL: <https://www.sciencedirect.com/science/article/pii/S0924271613002670>.
- [9] C. Tarantino et al. “8-Band Image Data Processing of the Worldview-2 Satellite in a Wide Area of Applications”. In: Jan. 2012. ISBN: 978-953-307-973-8. DOI: [10.5772/27499](https://doi.org/10.5772/27499).
- [10] Shanshan Su et al. “An Impervious Surface Spectral Index on Multispectral Imagery Using Visible and Near-Infrared Bands”. In: *Remote Sensing* 14.14 (2022). ISSN: 2072-4292. DOI: [10.3390/rs14143391](https://doi.org/10.3390/rs14143391). URL: <https://www.mdpi.com/2072-4292/14/14/3391>.
- [11] Mario Arreola-Esquivel et al. “Non-binary snow index for multi-component surfaces”. In: *Remote Sensing* 13.14 (2021), p. 2777. DOI: [10.3390/rs13142777](https://doi.org/10.3390/rs13142777).
- [12] R. Kaur and P. Pandey. “A review on spectral indices for built-up area extraction using remote sensing technology”. In: *Arabian Journal of Geosciences* 15.4 (2022), p. 391. DOI: [10.1007/s12517-022-09688-x](https://doi.org/10.1007/s12517-022-09688-x).
- [13] Hong Fang, Yuchun Wei, and Qiuping Dai. “A Novel Remote Sensing Index for Extracting Impervious Surface Distribution from Landsat 8 OLI Imagery”. In: *Applied Sciences* 9 (June 2019), p. 2631. DOI: [10.3390/app9132631](https://doi.org/10.3390/app9132631).

ABSTRACT

Accurate segmentation of built-up areas in satellite imagery is crucial for urban planning, environmental monitoring, and land-use management applications. This study evaluates the performance of multiple spectral indices designed for high-resolution WorldView-2 (WV2) satellite imagery. The high-resolution images enable the detection of fine surface details, enhancing the accuracy of built-up area segmentation.

BUILT-UP MAPS OF WV-2 SATELLITE IMAGERY



CONCLUSIONS

In this study, built-up indices were evaluated across various cities to assess their effectiveness in delineating urban areas. Spectral indices were tested among several land covers, including water, bare soil, vegetation, and built-up areas. These results were obtained using a thresholding approach specifically designed to reduce false positives from bare soil areas, which often spectrally overlap with built-up surfaces.

REFERENCES

- [1] M. I. Sameen., et al (2016). *A novel built-up spectral index developed by using multiobjective particle-swarm-optimization technique*, IOP C. Ser. Earth Environ. Sci. 37(1).
- [2] Mhangara, P., et al. (2011). *Road extraction using object-oriented classification*. *Visual Technology*, 20, 45–50.
- [3] Adeyemi, A., et al. (2021). *Spectral index to improve the extraction of built-up area from WorldView-2 imagery*. *J. of Applied Remote Sensing*, 15(2).
- [4] A. Wolf. (2010). *Using WorldView 2 Vis-NIR MSI Imagery to Support Land Mapping and Feature Extraction Using Normalized Difference Index Ratios*, Longmont, Colorado.

A quien corresponda.

Tenemos el gusto de informarle que el trabajo con clave única: LXVIII-010882 y cuyo título es: Índice espectral para la delimitación de suelo desnudo usando imágenes de alta resolución, de los autores:

1. Aldo Alberto Aguilar Vallejo (Presentador)
Universidad Politécnica de Tulancingo
2. Armando Aníbal Delgadillo Jiménez
Universidad Politécnica de Tulancingo
3. Raul Castro Ortega
Universidad Politécnica de Tulancingo
4. Alfonso Padilla Vivanco
Universidad Politécnica de Tulancingo
5. Anna Filomena Carbone
Politécnico de Turín
6. Carina Toxqui Quitl
Universidad Politécnica de Tulancingo

Fue aceptado para ser presentado en formato tipo plática en la sesión de Óptica el día 14 de octubre de 2025 en Salón 2 dentro del programa del LXVIII Congreso Nacional de Física.

Se extiende la presente constancia el día 23 de septiembre de 2025.

ATENTAMENTE,

Comité Científico
Sociedad Mexicana de Física





MEXICAN OPTICS AND
PHOTONICS MEETING

CONEXIÓN CON LA
COMUNIDAD
MEXICANA DE ÓPTICA

PHOTONICS MATERIALS
AND TECHNIQUES FOR
SERS AND
PHOTOCATALYTIC
APPLICATIONS

Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas
Academia Mexicana de Óptica (AMO)

otorgan a

Aldo Alberto Aguilar Vallejo

la presente

CONSTANCIA

Por su asistencia al Mexican Optics and Photonics MOPM 2025, realizado en la sala de usos múltiples de la Biblioteca Universitaria Raúl Rangel Frías, de la Universidad Autónoma de Nuevo León, del 1 al 3 de octubre del 2025.

“Alere Fammam Veritatis”
Ciudad Universitaria, a 3 de octubre de 2025.

Dr. Arturo Alberto Castillo Guzmán
Presidente de la AMO





MEXICAN OPTICS AND
PHOTONICS MEETING

CONEXIÓN CON LA
COMUNIDAD
MEXICANA DE ÓPTICA

PHOTONICS MATERIALS
AND TECHNIQUES FOR
SERS AND
PHOTOCATALYTIC
APPLICATIONS

Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico Matemáticas
Academia Mexicana de Óptica (AMO)

otorga a

**Aldo Alberto Aguilar Vallejo, Raúl
Castro Ortega, Armando Aníbal
Delgadillo Jiménez, Carina Toxqui Quitl,
Alfonso Padilla Vivanco, y Anna
Filomena Carbone**

el presente

RECONOCIMIENTO

Por su valiosa contribución como coautores del póster "**Monitoreo de la densidad de la construcción en ciudades usando imágenes multiespectrales**" en el Encuentro Mexicano de Óptica A.C., que se llevó a cabo en la sala de usos múltiples de la Biblioteca Universitaria Raúl Rangel Frías de la Universidad Autónoma de Nuevo León.

"Alere Fammam Veritatis"
Ciudad Universitaria a 03 de octubre de 2025.

Dr. Arturo Alberto Castillo Guzmán
Presidente de la AMO





Estimation of Urban Fractal Dimension Using a Convolutional Neural Network

Armando Delgadillo Jimenez¹ (✉), Carina Toxqui Quitl¹, Aldo Aguilar Vallejo¹, Raul Castro Ortega¹, Alfonso Padilla Vivanco¹, and Anna Carbone²

¹ Universidad Politécnica de Tulancingo, Hidalgo, Mexico
armando.delgadillo2331024@upt.edu.mx

² Politecnico Di Torino, Turin, Italy

Abstract. A deep learning approach to estimate the urban fractal dimension D_f using high-resolution WorldView-2 (WV-2) imagery is proposed. The networks are trained on fractal Brownian Motion (FBM) surfaces generated through computational models to simulate natural textures with varying degrees of roughness. Each surface is characterized by the Hurst exponent (H) related to the fractal dimension as $D_f = 2 - H$. For the classification task, the images are divided into nine distinct classes, each corresponding to a defined range of H values. The regression task uses the same dataset for training and predicts the value of the fractal dimension. The CNNs learn to detect spatial patterns that reflect differences in the fractal geometry of the surfaces. Trained on a large dataset of synthetic images, the models can accurately estimate the urban fractal dimension D_f from unseen satellite data. Our CNN-based predictions are compared against well-established methods for estimating the fractal dimension using real-world satellite data.

[AQ1](#)

[AQ2](#)

Keywords: Fractal Brownian Motion · Convolutional Neural Network · Detrended Moving Average · fractal dimension · urban systems

1 Introduction

Fractals are geometric entities exhibiting unique characteristics such as infinite length, fractal dimension, scale invariance, and self-similarity [1]. The fractal dimension (D_f) can be considered a measure of complexity and indicates the degree to which a random set fills the Euclidean space in which it is embedded. Another parameter commonly used in fractal analysis is the Hurst exponent, which measures the level of persistence and roughness in time series and random surfaces. The fractality has been shown to be related to changes in the environmental conditions of the geographical areas studied [2,3]. Hurst exponent values tend to decrease over time as human exploitation of rural area increases. Similarly the Hurst exponent estimated in cities decreases as the urbanization degree of the area increases. Overall, these studies provide indication of interesting relationships between the temporal and spatial changes in the fractality of the various regions of the Earth surface. It could be inferred that the rural or urban ecosystems are losing complexity over time and space.

The analysis of ecosystems and natural resources is crucial for a better understanding of the environment. It enables the sustainable management and planning of future decisions to mitigate adverse impacts caused by over-exploitation, global warming, and urban growth, among other potential issues. Technological advances in geographic information systems and improvements in computational power help supporting this task. A relevant tool is satellite systems, which allow large-scale monitoring and quantifying changes in the Earth surface over time. Once geospatial data are downloaded from various platforms, they are analyzed using classical digital image processing techniques [4, 5].

The morphology and function of cities are prominent examples of fractals with Hausdorff dimension D_f providing a measure of urban concentration over multiple scales [1, 2]. Fractal analysis applied to satellite imagery can measure and characterize the overall spatial complexity of remotely sensed images. The use of fractal geometry allows the identification of complex patterns that are not captured by traditional parameters. In remote sensing, the use of fractal parameters is particularly attractive because they are expected to be independent of the sensor properties [6, 7]. The variogram, which has been used for spatial analysis in geostatistics, can also be used to estimate the fractal dimension of natural surfaces [8]. Accurate estimates of fractal dimension have been obtained by implementing the Detrended Moving Average (*DMA*) algorithm on WorldView-2 (WV-2) satellite images of urban areas [9]. Results show higher fractal dimension values for urban rather than suburban and peripheral sectors, with D_f ranging from 1.65 to 1.90, respectively.

On the other hand, deep learning networks provide a powerful approach for parameter identification by automatically extracting features from high-dimensional and nonlinear data, thereby uncovering hidden patterns. Recent advances have shown promising results in applying deep learning to identify parameters of dynamical systems driven by fractional Brownian motion (FBM). Granik et al. [10] employed a deep learning model to estimate the Hurst exponent of FBM, demonstrating its effectiveness with both simulated and experimental data. Similarly, Wentao et al. [11] use deep learning for parameter identification of nonlinear dynamical system driven by fractional Brownian motion. David et al. [12] propose an approach based on deep learning techniques to measure the fractal dimension.

This paper proposes a deep learning approach to estimate and classify the urban fractal dimension using high-resolution WorldView-2 satellite imagery. Two Convolutional Neural Networks (CNNs) are designed for: *i* regression (RegFractNet) to predict continuous Hurst exponent values, and *ii* classification (ClassFractNet) to assign images to discrete H -value classes. Training and validation are first performed on synthetic Fractional Brownian Motion (FBM) surfaces with controlled Hurst exponents. The networks are then applied to real satellite images, classifying urban areas according to their morphological complexity. Results show high accuracy on synthetic data (regression error ≈ 0.04 , classification accuracy 95.56%) and a reasonable correspondence with literature values for real urban areas. To verify the consistency of the Hurst exponent values on satellite images, a comparison is done using the Detrended Moving Average (*DMA*) and variogram

algorithms. Our results demonstrate that both models can effectively characterize and classify different urbanization areas.

2 Description of the Fractal Surface Artificial Datasets

This section describes the datasets employed for training and validating the network models. The datasets are composed of synthetic *FBM* surfaces to verify the consistency of the Hurst exponent $H = 2 - D_f$ values.

Synthetic 2D images are generated with controlled Hurst exponent values ranging from 0.1 to 0.9, with increments of 0.1, with resolution of 128×128 pixels in grey level. The datasets were generated by the Cholesky-Levinson factorization (CLF) using the FRACLAB tool [13]. As for the image set, the CNN has used 1890 images corresponding to FBM, 210 generated images for each H value. Figure 1 shows nine *FBM* surface samples.

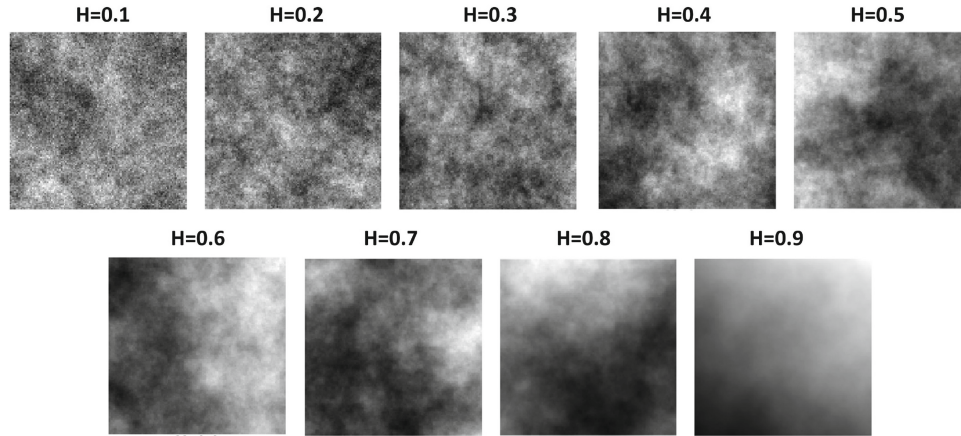


Fig. 1. Samples of Fractional Brownian motion surfaces with size 128×128 and controlled values of Hurst exponent H ranging from 0.1 to 0.9, steps of 0.1

Fractional Brownian fields, i.e. continuous functions of two variables $f_H(x_1, x_2)$ with the Hurst exponent H as a parameter, can be conveniently adopted as model of surfaces with random textures. The generic property of invariance under iteration describes the concept of fractal geometry, which has allowed us to understand structures and features observable in the real world that traditional geometry cannot accurately describe. Scaling relations of the form:

$$f(\lambda) \propto \lambda^{D_f}, \quad (1)$$

are shown by self-similar textures where λ is a characteristic scale, a measuring unit size, and D_f the fractal (Hausdorff) dimension given by:

$$D_f = d - H, \quad (2)$$

with d the Euclidean dimension and H the Hurst exponent that characterizes the degree of irregularity. $H > 0.5$ indicates a persistent correlation $H < 0.5$ anti-persistent correlation, and $H = 0.5$ corresponds to the uncorrelated ordinary Brownian function.

3 Architecture of the CNN

In this research, two convolutional Neural Network (CNN) architectures were designed: one for classification and another for regression. The ClassFractNet assigns images to discrete H values classes, and the RegFracNet predicts continuous Hurst exponent values from FBM synthetic surfaces and real-world urban areas. For training and validation, both networks used images that have a size of 128×128 grayscale pixels, corresponding to simulated Brownian surfaces. The CNN classification and regression models use a database of 1890 images for training and validation with 95% and 5%, and 90% and 10%, respectively. In this work, the CNN models have been implemented by the Deep Learning Toolbox of the MATLAB program. The experiments are conducted using a Lenovo laptop equipped with 8 GB DDR3L RAM and an AMD Ryzen 7 5700U processor with 2 GB Radeon Graphics.

3.1 ClassFractNet Model

A CNN architecture is designed to classify according to Hurst exponent values from images of Brownian surfaces with a Classification layer (Classification-Layer), named **ClassFractNet**. A subsampling layer follows each convolutional layer as a *max-pooling* operation of 2×2 , this helps to reduce the risk of overfitting, because the classification is performed in nine discrete categories. A fully connected layer passes as the final *softmax* output layer. Batch normalization is applied just before the activation function [14]. The architecture is trained using an Adam optimizer, ReLU activation function, a learning rate of 0.01, and a validation frequency of 12 iterations. It is trained for 50 epochs, with a random reorganization of the data performed at the beginning of each epoch to enhance generalization. Figure 2 shows the **ClassFractNet** architecture used in this case. The input images pass through five convolutional layers with kernel sizes of 3×3 and 5×5 and an output with nine classifications.

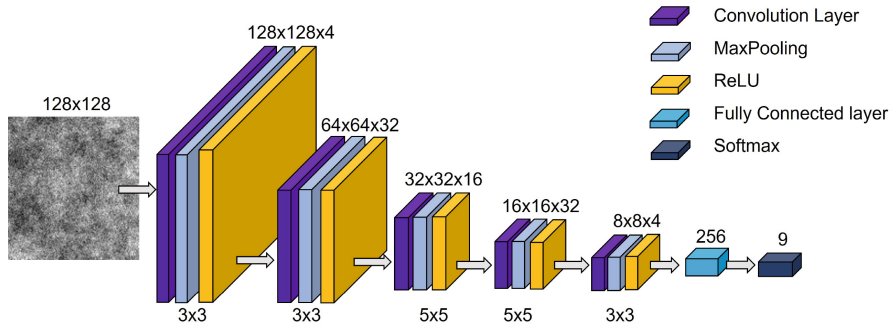


Fig. 2. **ClassFractNet** architecture. The model is trained entirely on a large dataset of synthetically generated FBM surfaces with known Hurst exponent values. The trained model assigns images to discrete H -value classes.

3.2 RegFractNet Model

Second, the network for regression named **RegFractNet** begins with an input layer, `imageInputLayer` ($128 \times 128 \times 1$), which defines the size and number of channels of the processed images. Based on this input, the architecture comprises six convolutional blocks with filter sizes of 5×5 and 3×3 , enabling it to capture complex spatial patterns related to the fractal structure of the images. Each block consists of a convolutional layer followed by a batch normalization layer and a ReLU activation function. A maxPooling subsampling layer is then used to reduce spatial dimensionality and retain the most relevant features [15].

The hierarchical progression of these blocks allows the network to learn deep representations of the texture and geometric complexity of the image. After the convolutional blocks, two fully connected layers are added. The first, with 256 neurons, condenses the learned information; the second, with a single neuron, gives the final prediction of the fractal dimension value. The architecture ends with a regression layer, suitable for tasks with continuous output. The training process is performed using the Adam optimizer, with an initial learning rate of 0.01. It is trained for 50 epochs with a mini-batch size of 32 samples, performing a random reorganization of the data at the beginning of each epoch to improve generalization. This architecture enables the capture of complex, nonlinear relationships between the texture patterns of images and their fractal dimension, providing an automated and efficient alternative to traditional estimation methods, such as box-counting, frequently used in fractal analysis [16]. Figure 3 shows the **RegFractNet** architecture used in this approach.

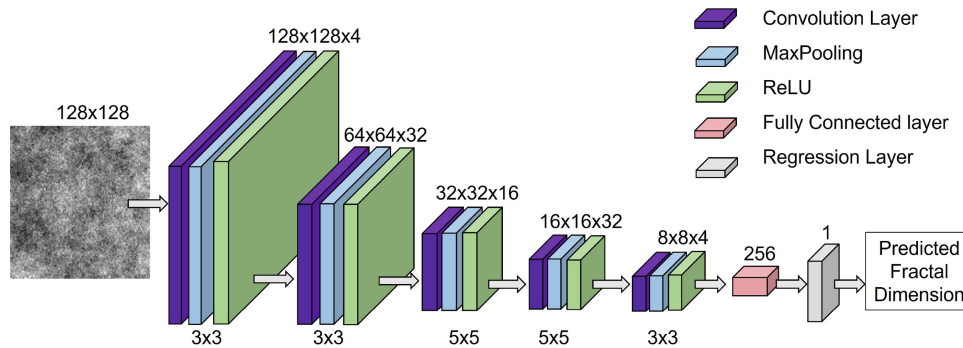


Fig. 3. RegFractNet architecture. The model is trained entirely on a large dataset of synthetically generated *FBM* surfaces with known Hurst exponent values. The trained model predicts continuous *H*-values.

4 Results and Discussion

The **ClassFractNet** and **RegFractNet** models are trained and validated entirely on controlled datasets. This section presents the testing results on synthetically generated *FBM* surfaces. Afterward, the trained models are applied

to real satellite images, classifying urban areas according to their morphological complexity. The CNN-based predictions are compared against the ones given by the *DMA* and variogram algorithms. Fractal metrics are evaluated for their ability to capture texture complexity, periodicity, self-similarity, and spatial heterogeneity across different image types.

4.1 Estimation of the Fractal Dimension of *FBM* Surfaces

Figures 4 and 5 shown the testing results given by the **ClassFractNet** and **RegFractNet** on artificial images that are not included in the training or validation datasets.

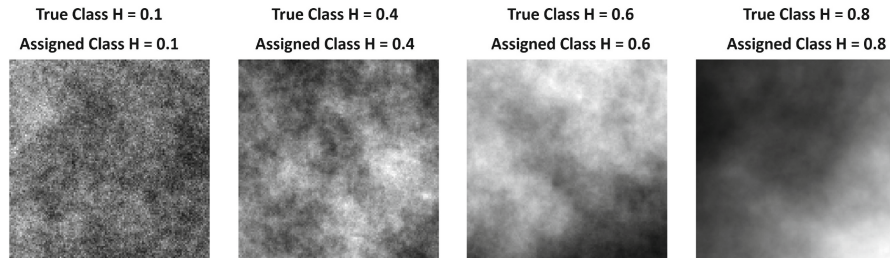


Fig. 4. Assigned class given by **ClassFractNet** to external artificial surfaces of FBM. The assigned H -value in each image corresponds to its true H value.

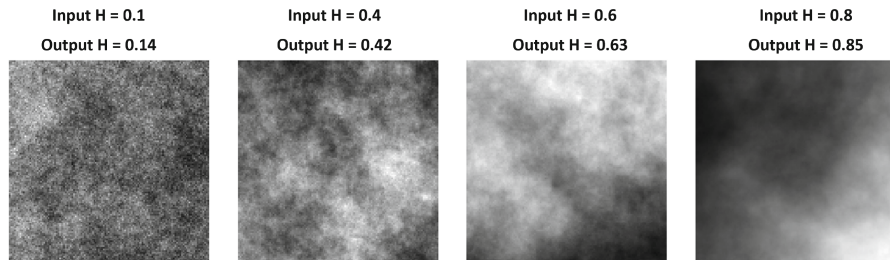


Fig. 5. Predicted H value given by **RegFractNet** to external artificial surfaces of FBM. The output H value in each image is close to its input H value, with an average error of 0.035. As can be seen, the **RegFractNet** model is more sensitive to the different fractional Brownian surfaces.

The **ClassFractNet** exhibits an accuracy of 95.56% on synthetically generated *FBM* surfaces. Figure 6 shows the confusion matrix generated after the training model. As can be seen, the Assigned class agrees with the True class in most cases.

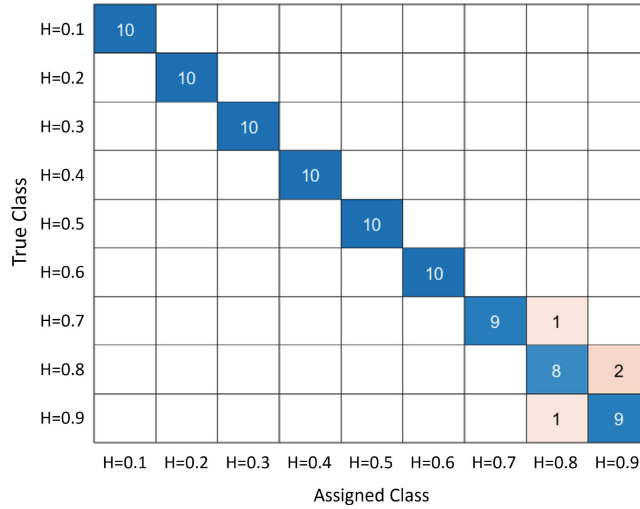


Fig. 6. **ClassFractNet**: Confusion matrix for validation data.

The **RegFractNet** exhibits a regression error of $\simeq 0.04$ on the *FBM* surfaces used as validation. Figure 7 shows a linear regression of the comparison between the H values of the labeling and the H values that the network predicts.

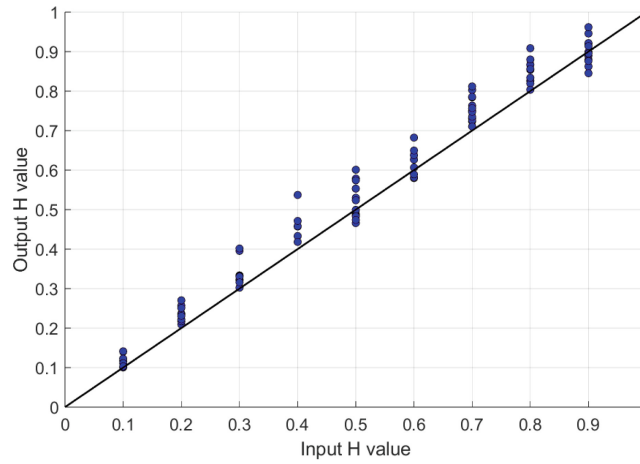


Fig. 7. **RegFractNet**: Regression results for validation data corresponding to 10% of full dataset.

4.2 Estimation of Fractal Dimension of WV-2 Imagery

High-resolution satellite images with varying levels of urbanization are used as testing data. A given Region of Interest (*RoI*) is selected from Image N45-024 (Torino) downloaded from the Urban Atlas collection of the largest European cities of WorldView-2 satellite images [17] with a spatial resolution of 10.73 m. The *RoI* that covers approximately 67,908 km² for the selected city is divided into 36 fixed-size non-overlapping patches of 128 × 128 pixels, matching the input size supported by the network model. In Fig. 8, the *RoI* image of Torino city N45-024 is shown. Since WV-2 satellite images are provided in RGB format, the red band was selected for testing.

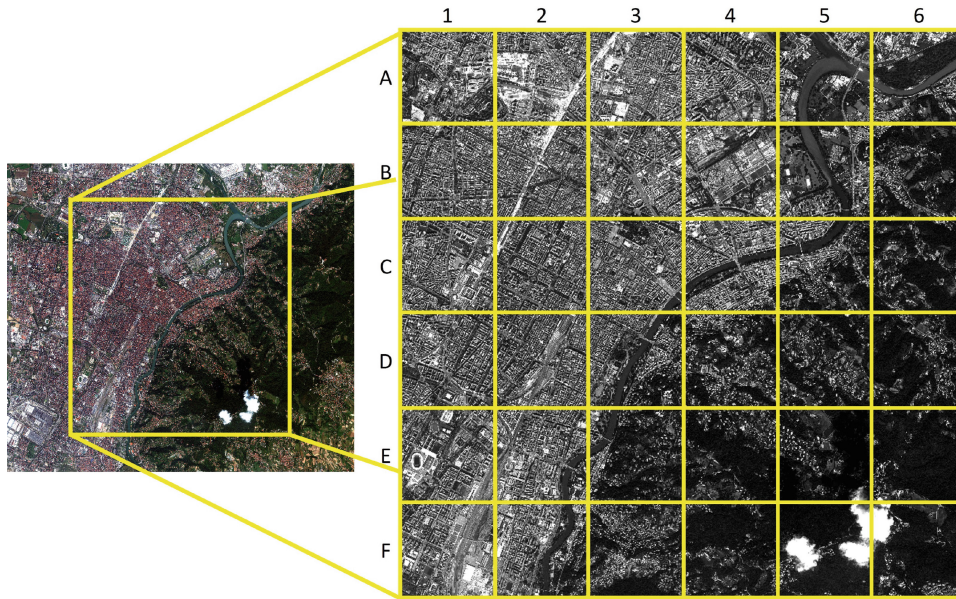


Fig. 8. A total of 36 fixed-size non-overlapping patches, each measuring 128 × 128 pixels, are extracted from image N45-024 (Torino). The red band is selected for testing and arranged into a matrix outlined in yellow, labeled with letters (A–F) for rows and numbers (1–6) for columns, in order to identify the different sections of the *RoI*. (Color figure online)

After the Region of Interest (*RoI*) is segmented into patches aligned with the specifications of the trained models, an experimental validation is conducted using randomly selected sections from the testing dataset. This procedure aims to evaluate the capacity of the models to generalize from synthetically generated data to real-world imagery. As depicted in Fig. 9, a subset of randomly sampled patches is employed to test the performance of the **RegFractNet** architecture, estimate the H parameter.

For regions with high urban density, the model predicts Hurst exponent values around $H = 0.12$. As the level of urbanization decreases, a noticeable increase in H values occurs. Some very noisy samples exhibit values as high as $H = 0.57$ due to cloudy occlusion in the satellite image. In this initial approach –keeping in mind that H is related to the fractal dimension, (see Eq. 2)– the range of D_f values varies from 1.43 to 1.89, consistently with previous studies reported in the literature [9, 18, 19].

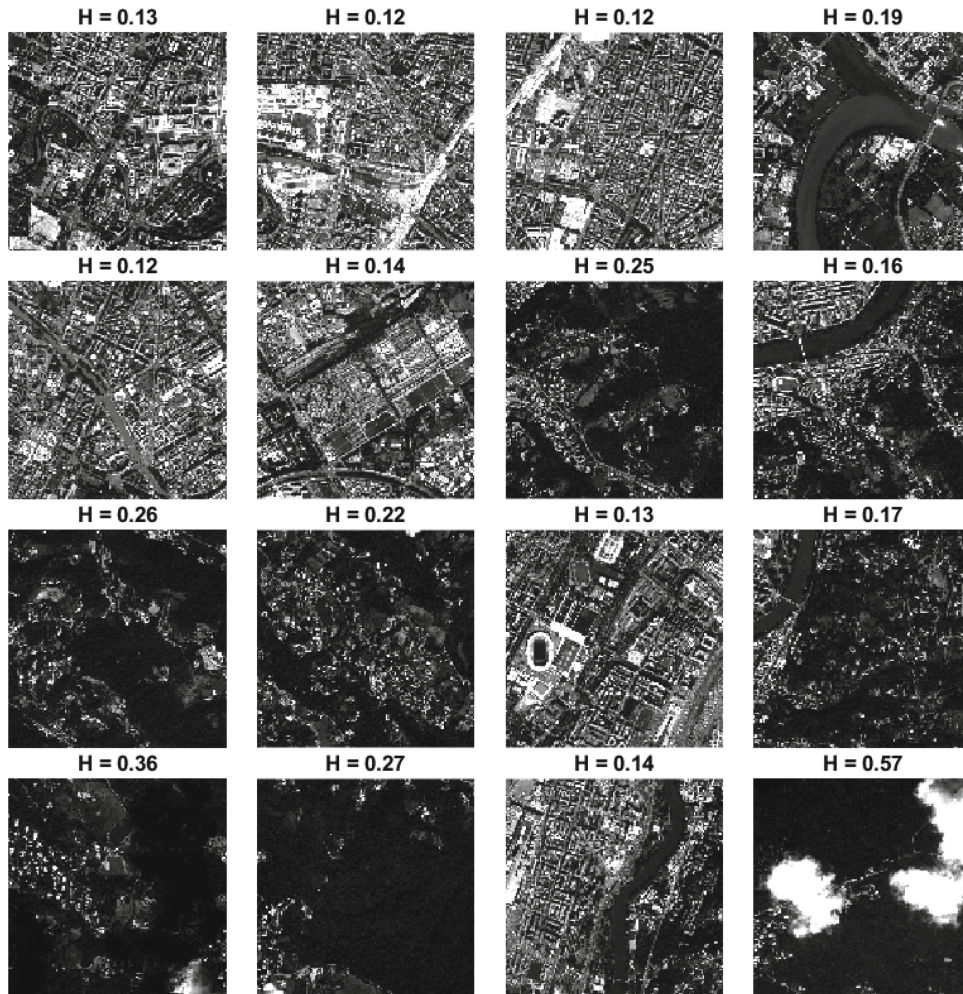


Fig. 9. Predicted Hurst exponent values given by the **RegFractNet** model for different urban areas. Random Patches of size 128×128 obtained from the WV-2 satellite image N45-024 (Torino) result in mean values of $D_f = 1.88, 1.81,$ and 1.75 for highly urbanized areas, urbanized areas, and rural areas, respectively.

Table 1. Fractal dimension D_f for the WV-2 images of Torino city. The WV-2 are divided in six sectors labelled by capital letters in the first column. The urban fractal dimension D_f is obtained by using: [1] **ClassFractNet**, [2] **RegFractNet**, [3] DMA, and [4] Variogram algorithms.

Areas	D_f [1]	D_f [2]	D_f [3]	D_f [4]
A1	1.90	1.87	1.90	1.90
A2	1.90	1.88	1.89	1.86
A3	1.90	1.88	1.90	1.88
A4	1.90	1.86	1.88	1.87
A5	1.90	1.81	1.87	1.83
A6	1.90	1.81	1.87	1.85
B1	1.90	1.86	1.90	1.90
B2	1.90	1.86	1.90	1.87
B3	1.90	1.88	1.90	1.88
B4	1.90	1.86	1.89	1.85
B5	1.90	1.82	1.86	1.85
B6	1.90	1.75	1.89	1.86
C1	1.90	1.86	1.90	1.87
C2	1.90	1.87	1.90	1.89
C3	1.90	1.84	1.89	1.88
C4	1.90	1.82	1.89	1.87
C5	1.90	1.84	1.89	1.84
C6	1.80	1.74	1.89	1.87
D1	1.90	1.87	1.90	1.89
D2	1.90	1.86	1.90	1.89
D3	1.90	1.82	1.89	1.86
D4	1.90	1.87	1.90	1.85
D5	1.90	1.78	1.90	1.84
D6	1.90	1.79	1.89	1.85
E1	1.90	1.87	1.89	1.85
E2	1.90	1.87	1.90	1.87
E3	1.90	1.83	1.89	1.85
E4	1.90	1.85	1.89	1.86
E5	1.80	1.64	1.86	1.83
E6	1.70	1.73	1.88	1.85
F1	1.90	1.87	1.89	1.85
F2	1.90	1.86	1.89	1.85
F3	1.90	1.83	1.90	1.84
F4	1.80	1.75	1.89	1.85
F5	1.40	1.43	1.77	1.54
F6	1.80	1.73	1.85	1.80

To verify the consistency of the CNN-based predictions, we implement two classical fractal dimension estimation methods for urban areas: the detrended moving average [9] and the variogram [8] algorithms. Table 1 presents the urban fractal dimension D_f values obtained on the 36 fixed-size, non-overlapping patches shown in Fig. 8.

5 Conclusions and Future Work

The fractal dimension is a key metric for urban morphology, complexity, and spatial planning. This approach is useful for large-scale, and automated analysis. The application to real satellite imagery demonstrates the potential of the proposed methodology for remote sensing and urban studies.

This paper proposes a deep learning approach to estimate and classify the urban fractal dimension using high-resolution WorldView-2 satellite imagery. CNN models are trained and validated on synthetic Fractional Brownian Motion (FBM) surfaces with controlled Hurst exponents. The networks are then applied to real satellite images, classifying urban areas according to their morphological complexity. The results demonstrated that the network was not only able to classify urban regions based on their morphological complexity correctly but also to assign Hurst exponent values that correlate accurately with the level of urbanization present in the images. This behavior reflects how the network can capture self-similarity and roughness patterns characteristic of urban surfaces.

Using CNN-based predictions and classical fractal dimension estimation methods, we found a clear trend: highly urbanized areas show low values of the Hurst exponent H . In contrast, less urbanized areas exhibit higher values. This pattern appears in techniques such as the variogram and DMA, and also in neural networks like **ClassFractNet** and **RegFractNet**. Among them, **RegFractNet** stands out for its greater ability to distinguish between different levels of urbanization due to its sensitivity. Overall, both traditional algorithms and neural networks provide different ranges of fractal dimensions. However, they all reveal the same classification trend: the degree of urbanization is closely reflected in fractal behavior.

Future work aims to train neural networks with larger Brownian surface datasets (512×512 and 1024×1024 pixels), which will enable the processing of higher-resolution images and increase the accuracy in classifying and regressing the Hurst exponent for larger urban areas with greater spatial complexity. This approach has the potential to improve the quantitative analysis of urban morphology and contribute to the fractal characterization of urban environments in satellite imagery.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Mandelbrot, B.B.: *Fractals: Form, Freeman, Chance and Dimension* (1977)
2. Batty, M., Longley, P.A.: Fractal-based description of urban form. *Environ. Plann. B. Plann. Des.* **14**(2), 123–134 (1987)
3. Frankhauser, P.: The fractal approach: a new tool for the spatial analysis of urban agglomerations. *Population* **10**(1), 205–240 (1998)
4. Arreola Esquivel, M., Toxqui Quitl, C., Delgadillo Herrera, M., Padilla Vivanco, A., Ortega Mendoza, G., Carbone, A.: Non-binary snow index for multi-component surfaces. *Remote Sens.* **13**(14), 2777 (2021)
5. Aguilar, A., Castro Ortega, R., Delgadillo, A., Toxqui, C., Padilla, A., Carbone, A.: Comparison of built-up indices for urban extraction. *Proceedings of SPIE 13596, Current Developments in Lens Design and Optical Engineering XXVI*, pp. 1359632 (2025)
6. Goodchild, M.F., Mark, D.M.: The fractal nature of geographic phenomena. *Ann. Assoc. Am. Geogra.* **77**(2), 265–278 (1987)
7. Di Martino, G., Iodice, A., Riccio, D., Ruello, G., Zinno, I.: The role of resolution in the estimation of fractal dimension maps from SAR data. *Remote Sens.* (2018). <https://doi.org/10.3390/rs10010009>
8. Develi, K., Babadagli, T.: Quantification of natural fracture surfaces using fractal geometry. *Math. Geology* **30**, 971–998 (1998)
9. Carbone, A., Murialdo, P., Pieroni, A., Toxqui Quitl, C.: Atlas of urban scaling laws. *Journal of Physics: Complexity* (2022). <https://doi.org/10.1088/2632-072X/ac718e>
10. Granik, N., et al.: TSingle-particle diffusion characterization by deep learning. *Biophys. J.* **117**(2), 185–192 (2019)
11. Hou, W., Ma, S.: Deep learning for parameter identification of nonlinear dynamical system driven by fractional Brownian motion. *Res. Square* (2024)
12. Valle, D., Wagemakers, A., Daza, A., Sanjuán, M.A.F.: Characterization of fractal basins using deep convolutional neural networks. *Int. J. Bifurcat. Chaos* **32** (2022)
13. FRACLAB, last visit on 05/2025. We use the CLF algorithm included in the package FRACLAB downloadable at <https://project.inria.fr/fraclab/>
14. Grandini, M., Bagli, E., Visani, G.: Metrics for Multi-Class Classification: an Overview ArXiv abs/2008.05756 (2020)
15. Alzubaidi, L., et al.: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **8**(1), 1–74 (2021). <https://doi.org/10.1186/s40537-021-00444-8>
16. Sun, Z.: Pattern Recognition in Convolutional Neural Network (CNN) ICMMIA 2022: Application of Intelligent Systems in Multi-modal Information Analytics, pp. 295–302
17. Urban Atlas. last visit on 03/2025. The ESA third party mission collection of the largest European urban areas recorded by the WorldView-2 satellite. <https://tpm-ds.eo.esa.int/oads/access/collection/WorldView-2>
18. Thomas, I., Frankhauser, P., Keersmaecker, M.L.: Fractal dimension versus density of the built-up surfaces in the periphery of Brussels. *Pap Reg. Sci.* **86**(2), 287–307 (2007)
19. Lagarias, A., Prastacos, P.: Estimation of fractal dimensions of Mediterranean cities using data for different built-up densities. *Conference: ERSAGR. In: Cities and regions in a changing Europe: challenges and prospects*, pp. 5–7 (2017)