

EVALUACIÓN DE LA ARQUITECTURA U-NET PARA LA SEGMENTACIÓN SEMÁNTICA DE COBERTURAS NATURALES EN IMÁGENES SATELITALES RGB

EVALUATION OF THE U-NET ARCHITECTURE FOR SEMANTIC SEGMENTATION OF NATURAL LAND COVERS IN RGB SATELLITE IMAGES

Jonathan Ruiz Apablaza^{1*}, Fredy Andrés Cristancho² y Víctor Andrés Martínez²

RESUMEN

Este estudio evalúa la viabilidad de aplicar redes neuronales convolucionales, específicamente la arquitectura U-Net, para la segmentación semántica de coberturas naturales en imágenes satelitales RGB del conjunto de datos DeepGlobe. La investigación se enmarca en el proyecto binacional COMIXTA entre el Instituto Geográfico Militar de Chile (IGM) y el Instituto Geográfico Agustín Codazzi de Colombia (IGAC), orientado al fortalecimiento de metodologías cartográficas basadas en inteligencia artificial. Se entrenaron dos versiones del modelo: una sin un conjunto de validación explícito y otra empleando una estrategia de validación simple con una división 80/20 de los datos y un mecanismo de parada anticipada (early stopping). Los resultados muestran que el modelo sin validación incurrió en sobreajuste, alcanzando métricas artificialmente elevadas ($\text{IoU} > 0.83$), mientras que el modelo con validación obtuvo predicciones más conservadoras pero generalizables ($\text{IoU} \approx 0.42$). La evaluación cualitativa reveló errores sistemáticos en la clase “agua” debido al desbalance en el conjunto de datos. Se utilizaron técnicas como entrenamiento en precisión mixta, normalización robusta y activación GELU para mejorar la eficiencia y estabilidad del aprendizaje. La implementación fue realizada en un entorno computacional accesible (GPU NVIDIA T1000), demostrando que estas metodologías pueden ser replicadas en instituciones públicas con recursos limitados. Este trabajo establece una base técnica sólida para futuras extensiones hacia modelos multiclase, integración de imágenes multispectrales y producción cartográfica automatizada a gran escala.

Palabras clave: segmentación semántica, U-Net, cartografía automatizada, imágenes satelitales, inteligencia artificial geoespacial.

ABSTRACT

This study evaluates the feasibility of applying convolutional neural networks, specifically the U-Net architecture, for the semantic segmentation of natural land covers in RGB satellite images from the DeepGlobe dataset. The research is part of the binational COMIXTA project between the Military Geographic Institute of Chile (IGM) and the Agustín Codazzi Geographic Institute of Colombia (IGAC), aimed at strengthening cartographic methodologies based on artificial intelligence. Two versions of the model were trained: one without an explicit validation set and another using a simple validation strategy with an 80/20 data split and an early stopping mechanism. The results show that the model without validation suffered from overfitting, reaching artificially high metrics ($\text{IoU} > 0.83$), while the model with validation produced more conservative but generalizable predictions ($\text{IoU} \approx 0.42$). Qualitative evaluation revealed systematic errors in the “water” class due to data imbalance. Techniques such as mixed precision training, robust normalization, and GELU activation were used to improve training efficiency and stability. The implementation was carried out in an accessible computing environment (NVIDIA T1000 GPU), demonstrating that these methodologies can be replicated in public institutions with limited resources. This work establishes a solid technical foundation for future extensions toward multiclass models, integration of multispectral imagery, and large-scale automated cartographic production.

Keywords: semantic segmentation, U-Net, automated cartography, satellite imagery, geospatial artificial intelligence.

¹Instituto Geográfico Militar, Santiago, Chile.

²Instituto Geográfico Agustín Codazzi, Bogotá, Colombia.

*Autor de correspondencia: jruiz@igmm.cl

1. INTRODUCCIÓN

La generación automatizada de cartografía temática a partir de imágenes satelitales ha experimentado avances significativos durante la última década, impulsada por el desarrollo de técnicas de aprendizaje profundo que permiten abordar tareas complejas como la clasificación del uso y cobertura del suelo con una resolución sin precedentes (Zhu et al., 2017; Tuia et al., 2011). En este contexto, la segmentación semántica asigna una etiqueta de clase a cada píxel de una imagen, emergiendo como una herramienta prometedora para actualizar bases cartográficas, monitorear el ambiente y apoyar la planificación territorial.

El uso creciente de satélites ópticos de muy alta resolución espacial (VHR), como WorldView-2/3 o Pleiades, ha permitido capturar imágenes multispectrales con resoluciones inferiores a un metro, facilitando la detección precisa de coberturas como cuerpos de agua, vegetación, estructuras urbanas y áreas agrícolas (Gómez et al., 2016). Sin embargo, este incremento en la calidad espacial de los datos ha hecho inviable el procesamiento manual o semi-manual a gran escala, exigiendo nuevas herramientas que automaticen tareas sin comprometer la precisión topológica ni semántica.

En América Latina, donde la heterogeneidad territorial, la presión sobre los recursos naturales y los riesgos ambientales requieren de estrategias robustas de monitoreo y gobernanza espacial, la incorporación de tecnologías basadas en Inteligencia Artificial (IA) se ha vuelto prioritaria. Instituciones públicas como el Instituto Geográfico Agustín Codazzi (IGAC) de Colombia, han comenzado a integrar modelos de aprendizaje automático en sus procesos cartográficos. No obstante, la adaptación de estas herramientas requiere superar limitaciones institucionales comunes en la región: escasez de GPU especializadas, datos de entrenamiento etiquetados limitados y capacidades técnicas dispares.

Entre las soluciones técnicas más destacadas para la segmentación semántica de imágenes se encuentra la red neuronal convolucional U-Net, una arquitectura profunda que ha sido validada tanto en dominios médicos como geospaciales (Ronneberger et al., 2015; Volpi & Tuia, 2017). Diseñada originalmente para segmentar células en imágenes microscópicas, U-Net se caracteriza por una estructura tipo encoder-decoder que permite reducir la resolución espacial para extraer características globales y luego reconstruirla utilizando skip connections que preservan detalles locales.

Este estudio se enmarca dentro del proyecto de cooperación binacional COMIXTA entre el Instituto Geográfico Militar (IGM) de Chile y el IGAC de Colombia, cuyo fin es fortalecer las capacidades técnicas para la producción cartográfica mediante IA. La presente investigación constituye una primera aproximación orientada a evaluar la factibilidad de aplicar la arquitectura U-Net para la segmentación semántica de coberturas de bosque y agua en imágenes RGB del conjunto de datos de referencia DeepGlobe. Para evaluar el impacto del sobreajuste, un riesgo inherente en modelos de aprendizaje profundo, se compara el desempeño de un modelo entrenado sin validación explícita frente a otro entrenado con una división 80/20 y un mecanismo de parada anticipada.

2. MATERIALES Y MÉTODOS

Este apartado describe los datos utilizados, la configuración computacional empleada, la arquitectura del modelo y las técnicas de preprocesamiento adoptadas. La metodología ha sido diseñada bajo principios de transparencia, reproducibilidad y adaptabilidad institucional.

2.1 Datos

El presente estudio utiliza como fuente primaria el conjunto de datos del DeepGlobe Land Cover Classification Challenge (Demir et al., 2018), una referencia internacional ampliamente empleada en competencias de segmentación semántica aplicadas a imágenes satelitales de muy alta resolución (VHR). Este dataset fue originalmente curado para fomentar el desarrollo de algoritmos robustos de clasificación de cobertura del suelo a partir de imágenes RGB capturadas por el satélite WorldView-3, el cual ofrece una resolución espacial nativa de 50 cm por píxel. El conjunto completo utilizado contiene 803 imágenes RGB anotadas con máscaras de segmentación que cubren siete clases principales de uso del suelo: bosque, agua, tierras agrícolas, áreas

urbanas, pastizales, tierras áridas y otros. No obstante, para efectos de esta primera implementación y con el fin de simplificar la tarea desde el punto de vista computacional y analítico, se restringió la segmentación a dos clases de interés geográfico: bosque y agua.

Para reducir la carga computacional y adaptar las entradas al modelo, las imágenes fueron redimensionadas de su resolución original de 1448x1448 a 256x256 píxeles mediante interpolación bilineal. Las máscaras multiclase originales fueron procesadas para generar una máscara de salida de dos canales, donde cada canal corresponde a una de las clases de interés. Para ello, se aplicó una codificación estricta basada en umbralización de color: el verde puro (0, 255, 0) se asignó a la clase "bosque" y el azul puro (0, 0, 255) a la clase "agua".

El enfoque binario adoptado también responde a la necesidad de desarrollar una línea base robusta y escalable. Trabajar inicialmente con dos clases permite aislar los efectos de la arquitectura del modelo y del preprocesamiento, sin la interferencia de problemas de desbalance extremo entre clases o ambigüedades semánticas entre coberturas similares. La imagen utilizada para la inferencia de los modelos corresponde al área que cubre la carta G-102 NE escala 1:25.000, ubicada en la región de La Araucanía, provincia de Cautín, comuna de Cunco, Chile.

2.2 Hardware y entorno de ejecución

El desarrollo completo de los experimentos fue realizado utilizando Python 3.10 como lenguaje principal, implementado sobre el entorno interactivo Jupyter Notebook, lo que permitió una rápida iteración y visualización intermedia de resultados. Este entorno es reconocido por su utilidad en tareas de ciencia de datos, aprendizaje profundo, y a su vez, ha sido recomendado por plataformas como EarthAI, Radiant MLHub y otras iniciativas de inteligencia geoespacial reproducible (Kedron et al., 2021).

Desde el punto de vista del hardware, la ejecución se llevó a cabo en un workstation con sistema operativo Windows 10, procesador Intel Xeon Silver 4208, y una GPU NVIDIA T1000 de 8 GB, compatible con la arquitectura CUDA 11.8. Esta configuración representa una infraestructura básica-intermedia, disponible en muchos centros de investigación e instituciones técnicas de América Latina, lo que la convierte en un entorno representativo para validar la viabilidad de soluciones de IA aplicada en contextos con limitaciones presupuestarias.

Se utilizó la biblioteca PyTorch 2.0 como marco principal para la definición y entrenamiento del modelo, dada su flexibilidad y compatibilidad con entrenamiento en GPU. Se integraron además bibliotecas complementarias como torchvision, numpy, scikit-learn, matplotlib, tqdm, Pillow y rasterio. Este entorno demostró ser suficiente para entrenar modelos U-Net con lotes de hasta 8 imágenes por iteración sin saturar la memoria de la GPU.

2.3 Arquitectura del modelo

Se implementó la arquitectura U-Net, originalmente propuesta por Ronneberger et al. (2015) para la segmentación de estructuras celulares en imágenes biomédicas. Esta red convolucional profunda ha demostrado resultados excepcionales en tareas de segmentación semántica densa, incluyendo aplicaciones remotas y cartográficas (Maggiori et al., 2017; Volpi & Tuia, 2017). Su diseño tipo encoder-decoder permite capturar información contextual de largo alcance al tiempo que preserva detalles espaciales locales mediante skip connections simétricas.

En este estudio, la U-Net fue adaptada para trabajar con 3 canales de entrada (RGB) y 2 canales de salida, correspondientes a las clases "bosque" y "agua". El encoder consta de cinco bloques secuenciales, cada uno compuesto por capas Conv2d con kernel 3x3, normalización por lotes (BatchNorm2d), y activación GELU (Gaussian Error Linear Unit).

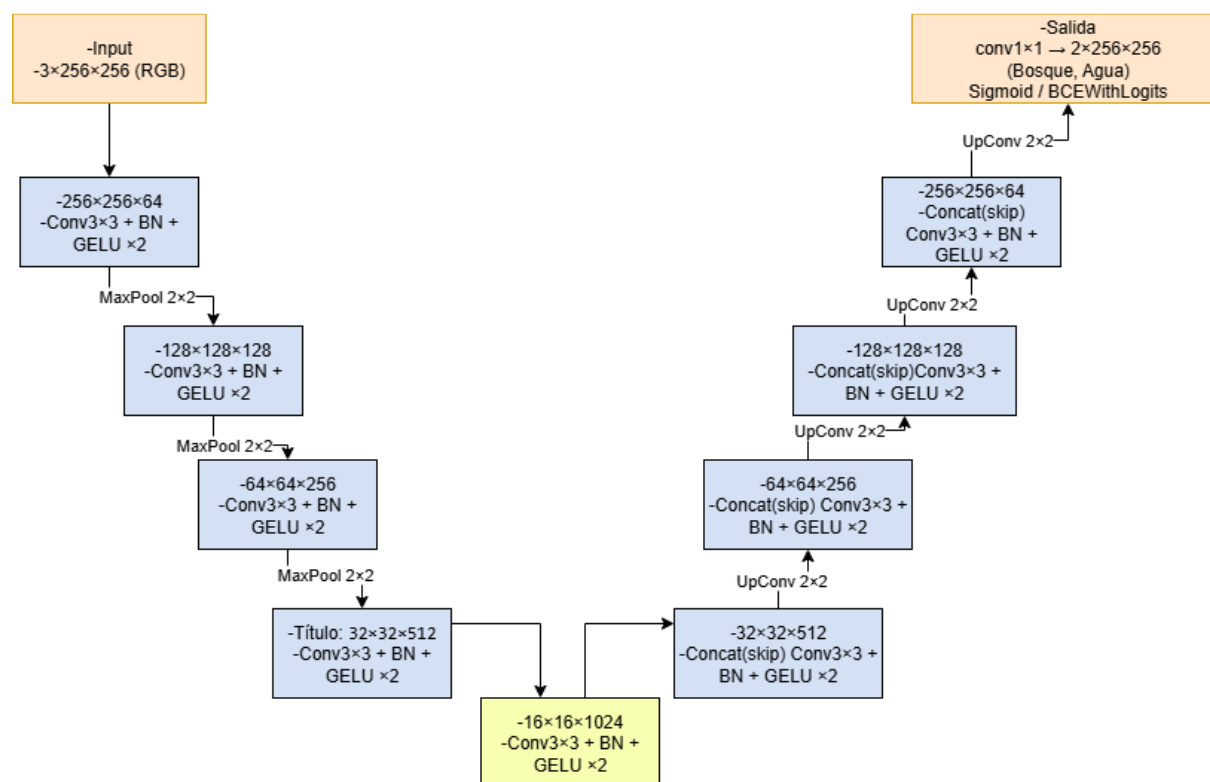


Figura 1. Estructura modelo U-Net. Fuente: Elaboración propia, 2025.

La elección de GELU en lugar de la activación ReLU estándar se basó en la hipótesis de que su naturaleza estocástica y su transición suave podrían favorecer una mejor propagación de gradientes en el contexto de coberturas naturales con bordes difusos (Hendrycks & Gimpel, 2016). El decoder replica la estructura del encoder pero en orden inverso, utilizando ConvTranspose2d para realizar upsampling progresivo. En cada nivel, se concatenan las características del encoder mediante skip connections, lo que permite recuperar detalles de borde perdidos durante la contracción.

2.4 Normalización y preprocesamiento

El preprocesamiento de los datos es una etapa crítica en la eficacia del entrenamiento de redes neuronales profundas, especialmente cuando se trabaja con datos visuales provenientes de sensores remotos (Zhu et al., 2017). En este trabajo, se adoptó una estrategia de normalización robusta, basada en la mediana y el rango intercuartílico de los datos de entrenamiento, utilizando el algoritmo RobustScaler de la biblioteca scikit-learn.

Esta técnica se diferencia de las normalizaciones clásicas al no verse afectada por valores extremos o distribuciones sesgadas, lo que resulta particularmente útil cuando las imágenes contienen zonas saturadas, sombras profundas o artefactos por nubes. Tal como lo demuestran Ioffe y Szegedy (2015), la normalización adecuada de las entradas puede acelerar la convergencia del modelo y reducir la dependencia de la inicialización de pesos.

El pipeline de preprocesamiento siguió las siguientes etapas: redimensionamiento a 256x256 píxeles mediante interpolación bilineal, conversión a tensor mediante ToTensor() de torchvision, y normalización robusta ajustada sobre una muestra del conjunto de entrenamiento. Para las máscaras, se aplicó un procedimiento paralelo de redimensionamiento seguido por umbralización vectorizada para generar tensores binarios por clase.

2.5 Optimización y entrenamiento

El proceso de entrenamiento del modelo U-Net fue diseñado considerando principios de eficiencia computacional y estabilidad numérica. Se implementó un esquema de entrenamiento en precisión mixta (mixed precision training), activado mediante el módulo torch.amp de PyTorch. Esta técnica permite realizar cálculos intermedios en coma flotante de 16 bits (FP16) mientras se mantienen acumuladores en 32 bits (FP32), reduciendo el uso de memoria y acelerando los tiempos de cómputo sin pérdida significativa de precisión (Micikevicius et al., 2018).

Se utilizó como función de pérdida la BCEWithLogitsLoss, especialmente diseñada para tareas de segmentación binaria multicanal. El optimizador adoptado fue AdamW, una variante de Adam con desacoplamiento explícito del término de regularización weight_decay, lo que mejora la generalización al prevenir la acumulación de pesos excesivos en redes profundas (Loshchilov & Hutter, 2019). Se fijó un learning_rate inicial de $3e-4$ y un weight_decay de $1e-4$ como configuración base.

La estrategia experimental se dividió en dos fases para evaluar el impacto del sobreajuste. Modelo 1 (Sin Validación): se entrenó utilizando el 100% de los datos durante 100 épocas, sin un conjunto de validación para monitoreo. Modelo 2 (Con Validación y Parada Anticipada): los datos se dividieron en conjuntos de entrenamiento (80%) y validación (20%) mediante una partición estratificada y reproducible. Este modelo se entrenó con un mecanismo de parada anticipada que detenía el proceso si el

2.6 Estrategia de validación y parada anticipada

La validación simple fue implementada como una estrategia fundamental para evaluar la capacidad del modelo de generalizar a datos no vistos. Se utilizó una división simple (hold-out) del 80/20 mediante train_test_split, asegurando una distribución aleatoria y balanceada de clases. Esta decisión metodológica se fundamenta en la alta dependencia espacial inherente a los datos satelitales (Tuia et al., 2011).

En cada época, el modelo fue evaluado contra el conjunto de validación y se calcularon métricas promedio por canal. Como criterio de interrupción temprana, se definió que el entrenamiento debía finalizar si no se observaba mejora en el IoU de validación durante 15 épocas consecutivas. Esta técnica de early stopping ha demostrado ser efectiva para prevenir el sobreajuste en tareas con conjuntos de datos de tamaño medio o reducido (Bengio, 2012). La implementación práctica permitió detener el entrenamiento en la época 68, representando una reducción de 60% en el tiempo de cómputo respecto al límite superior originalmente previsto.

2.7 Métricas utilizadas

La evaluación del rendimiento del modelo fue realizada mediante tres métricas clásicas en tareas de segmentación semántica binaria, calculadas de forma independiente para cada canal de salida y luego promediadas para generar un indicador global:

- IoU (Intersection over Union): También conocido como índice de Jaccard, mide la intersección entre la predicción y la verdad terreno, dividida por su unión. Es especialmente útil en segmentación semántica, debido a que penaliza tanto los falsos positivos como los falsos negativos (Everingham et al., 2010).
- Precisión: Define la proporción de verdaderos positivos sobre todos los positivos predichos. Una alta precisión indica que las áreas detectadas como pertenecientes a una clase realmente lo son, lo que reduce el riesgo de sobresegmentación (Sokolova & Lapalme, 2009).
- Recall: Mide la proporción de verdaderos positivos sobre el total de instancias reales. Es clave en contextos donde las omisiones pueden tener consecuencias significativas, como la subdetección de cuerpos de agua o masas forestales dispersas.

2.8 Evaluación visual y validación cualitativa

La evaluación cuantitativa fue complementada mediante una validación cualitativa, consistente en la visualización de resultados sobre imágenes satelitales no vistas durante el entrenamiento. Esta evaluación visual tuvo como objetivo verificar la coherencia espacial de las predicciones y su correspondencia con la morfología de los objetos geográficos segmentados. Se prestó especial atención a la continuidad de los bordes, la detección de fragmentos pequeños y la supresión de falsas detecciones.

3. RESULTADOS

Esta sección presenta los resultados del estudio en cuatro análisis complementarios: comparación de modelos entrenados con y sin validación, evaluación cualitativa visual de las predicciones, síntesis de métricas finales, y análisis del desempeño desagregado por clase (bosque vs. agua).

3.1 Comparación entre modelos con y sin validación

La investigación evaluó el desempeño de dos configuraciones de la arquitectura U-Net para segmentación semántica de coberturas naturales, centrándose en las clases "bosque" y "agua". El modelo 1 fue entrenado durante 100 épocas utilizando la totalidad del conjunto de datos, sin segmentar un subconjunto específico para validación. Por su parte, el modelo 2 fue entrenado bajo una estrategia metodológicamente más robusta: se reservó un 20% de los datos para validación, y se utilizó una técnica de early stopping, que detuvo el entrenamiento tras 15 épocas consecutivas sin mejora en el IoU de validación.

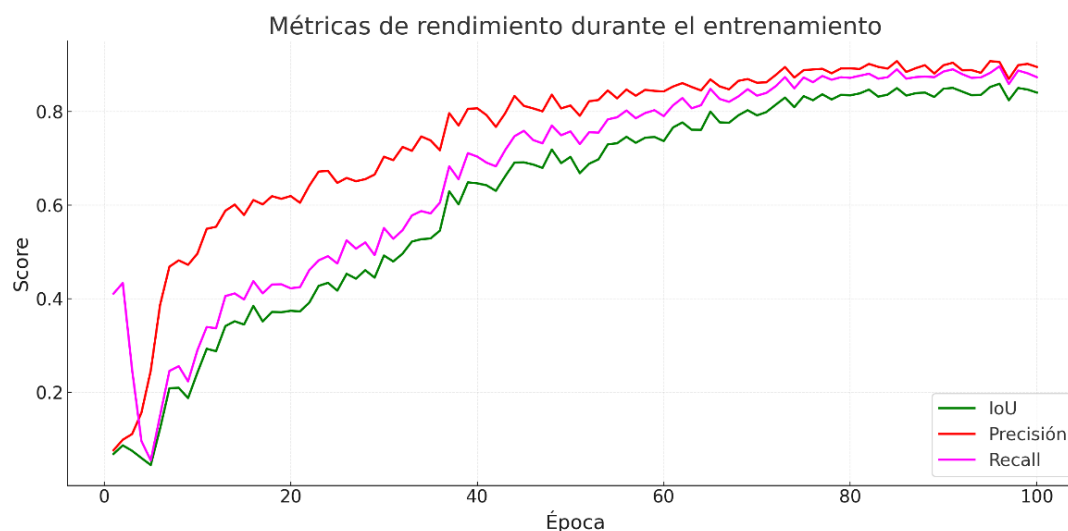


Figura 2: Gráfico comparativo de métricas de evaluación de modelo 1. Fuente: Elaboración propia, 2025.

En el caso del modelo 1 (Figura 2), se observa una disminución continua de la función de pérdida hasta valores cercanos a 0.01 al finalizar el entrenamiento, junto con un aumento progresivo del IoU que alcanza valores superiores a 0.83. A primera vista, esto podría interpretarse como una señal de éxito, no obstante, al no existir un conjunto de validación independiente, estos resultados deben ser considerados con cautela, debido a que reflejan el ajuste del modelo al conjunto de entrenamiento y no necesariamente su capacidad para predecir nuevos datos.

En la Figura 3, correspondiente al modelo 2, se observa una dinámica distinta. Durante las primeras 30-35 épocas, tanto la pérdida de entrenamiento como la de validación disminuyen significativamente. Sin embargo,

a partir de ese punto, la curva de validación se estabiliza en torno a 0.08, mientras que la de entrenamiento sigue descendiendo. Esta divergencia indica el inicio del sobreajuste, pero gracias al mecanismo de parada anticipada, el entrenamiento se interrumpe antes de que el modelo se sobreentrene gravemente. El IoU de validación se estabiliza en torno a 0.42.

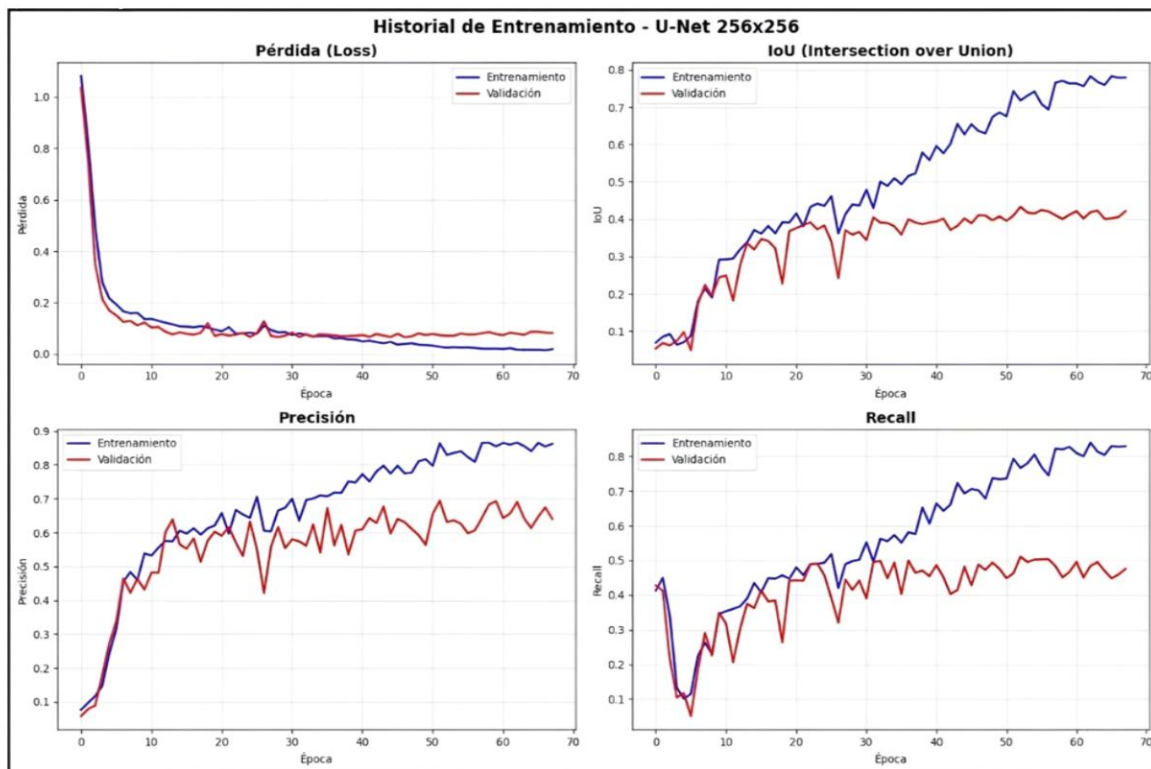


Figura 3. Gráfico comparativo de métricas de evaluación de modelo 2. Fuente: Elaboración propia, 2025.

Otro aspecto relevante es el comportamiento de las métricas de precisión y recall. El modelo 1 exhibe valores superiores al 89% en ambas, pero solo sobre entrenamiento. El modelo 2 alcanza un máximo de 64.1% de precisión y 47.58% de recall sobre validación. Esta asimetría revela que, aunque el modelo validado tiene un rendimiento cuantitativo inferior, es mucho menos propenso a cometer errores de generalización. En aplicaciones cartográficas, donde la fiabilidad y coherencia morfológica son claves, estos resultados sugieren que el modelo con validación ofrece una base más segura para futuras implementaciones.

3.2 Resultados cualitativos

La evaluación cualitativa de los modelos es crucial para complementar el análisis numérico y captar dimensiones morfológicas, espaciales y operativas que no siempre son visibles a través de métricas tradicionales. Se examinan los resultados visuales generados por ambos modelos sobre escenas no vistas del conjunto de datos DeepGlobe, con especial énfasis en la detección de las clases "bosque" y "agua".

El modelo 1, entrenado sin validación, produjo segmentaciones extensas y con alta densidad temática para la clase "bosque". En la Figura 4, el modelo clasifica como vegetación zonas periféricas, caminos, áreas agrícolas y otras superficies que, si bien pueden compartir tonalidades verdes, no corresponden a bosque real. Esta tendencia a la sobresegmentación puede atribuirse a una sobreoptimización de los pesos del modelo hacia características superficiales frecuentes en el conjunto de entrenamiento. Además, se evidencian errores de clasificación en cuerpos de agua, como falsas detecciones en zonas de sombra o sobre áreas húmedas no permanentes.

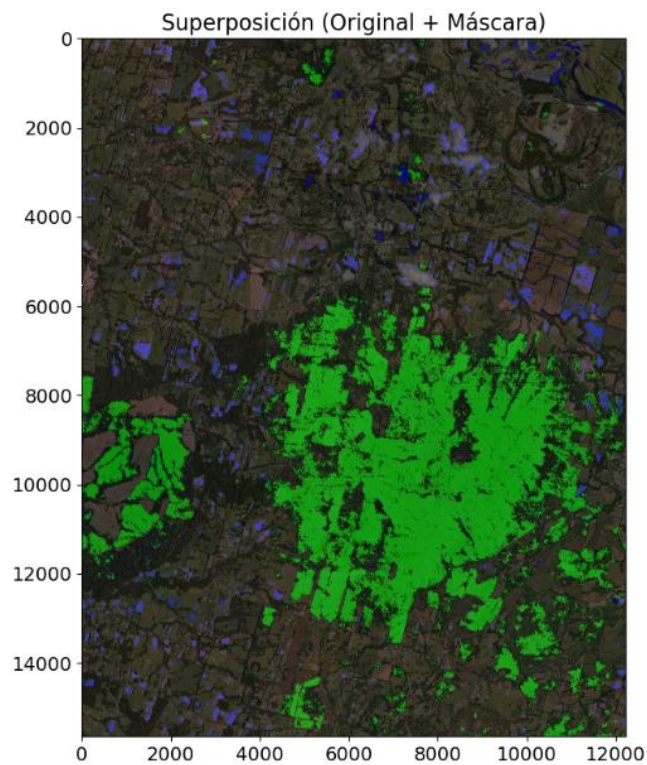


Figura 4. Resultado de inferencia de modelo 1. Fuente: Elaboración propia, 2025.

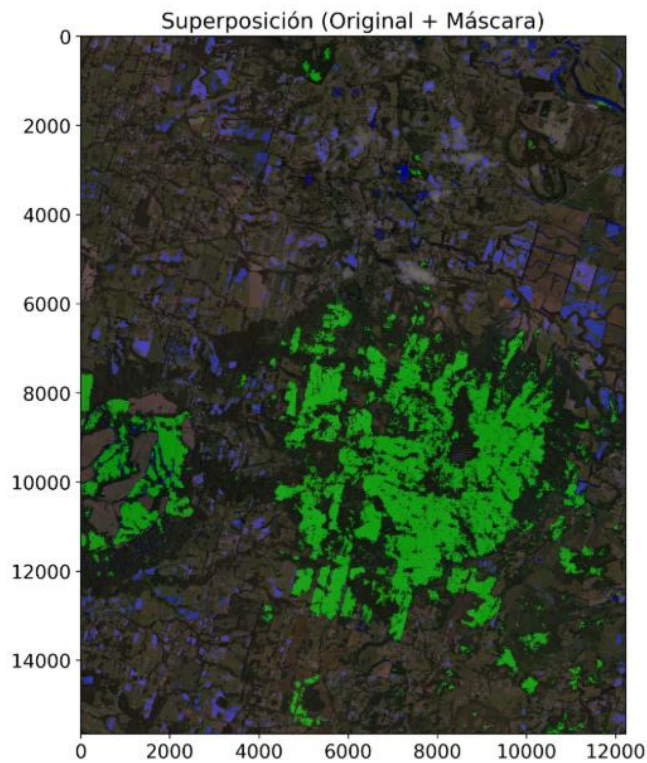


Figura 5. Resultado de inferencia de modelo 2. Fuente: Elaboración propia, 2025.

Por el contrario, el modelo 2, entrenado con validación y parada anticipada, muestra un comportamiento más prudente y ajustado. En la Figura 5 se observa que las áreas clasificadas como bosque son menos extensas, pero coinciden con mayor precisión con los objetos reales visibles en la imagen RGB.

La cobertura de agua, aunque reducida, se localiza preferentemente en zonas correctas, y los límites entre clases presentan mayor nitidez y congruencia morfológica. Este modelo logra preservar la geometría original de las coberturas con mayor fidelidad, especialmente en regiones de transición entre bosque y suelo desnudo.

El contraste entre ambos modelos no solo revela diferencias en la cantidad de área clasificada, sino también en la calidad espacial de las predicciones. El modelo 1 tiende a generar máscaras densas, continuas y con bordes irregulares, mientras que el modelo 2 opta por segmentaciones más parciales pero coherentes, con geometría interna más ajustada a los objetos reales. Esto se traduce en un menor número de falsos positivos en el modelo 2, incluso a costa de omitir algunos fragmentos pequeños de cobertura verdadera (falsos negativos), tal como lo representa la figura 6.

Desde un punto de vista operativo, las segmentaciones del modelo 2 resultarían más útiles para tareas de monitoreo ambiental o generación de mapas temáticos, debido a que reducen el ruido cartográfico y mejoran la confianza en la interpretación visual. La fidelidad morfológica de los objetos detectados, la supresión de errores sistemáticos y la coherencia espacial en zonas heterogéneas refuerzan su aplicabilidad institucional.

La evaluación cualitativa corrobora que el modelo con validación no solo generaliza mejor desde el punto de vista numérico, sino que también produce salidas más realistas, cartográficamente aceptables y útiles para análisis geoespacial. Esta dimensión visual, frecuentemente descuidada en la literatura técnica, debe ser integrada sistemáticamente en la validación de modelos de segmentación, especialmente en instituciones públicas donde la calidad del producto final tiene implicancias legales, operativas y sociales.

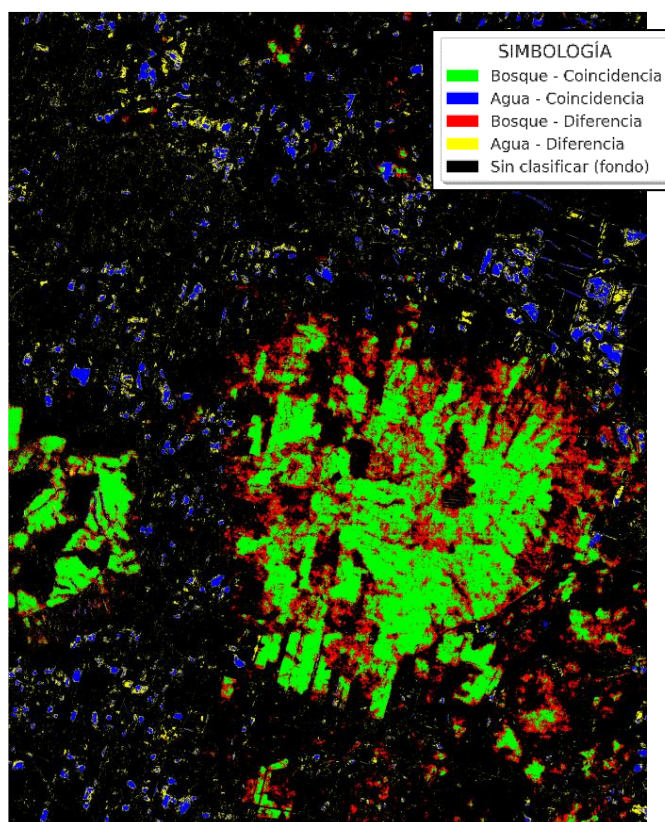


Figura 6: Diferencia de segmentación semántica entre el modelo 1 y 2. Fuente: Elaboración propia, 2025.

3.3 Comparación de métricas finales

La comparación de métricas finales entre los modelos entrenados con y sin validación permite cuantificar de manera precisa el impacto del sobreajuste, la estabilidad del entrenamiento y la generalización sobre datos no vistos.

El modelo 1 completó 100 épocas y alcanzó un IoU final de 0.8409, con una precisión de 0.8957 y un recall de 0.8738. La pérdida final fue de apenas 0.0096. Si bien estos valores pueden interpretarse como un éxito en términos de ajuste al conjunto de datos, fueron calculados exclusivamente sobre el mismo conjunto con el cual el modelo fue entrenado, lo que impide evaluar su capacidad de generalización.

En contraste, el modelo 2 se detuvo en la época 68, con un IoU de validación de 0.4213, precisión de 0.6409 y recall de 0.4758. A diferencia del modelo 1, los valores del modelo 2 fueron obtenidos sobre un conjunto de validación independiente, lo que les confiere una mayor validez para medir el desempeño real del modelo ante datos no vistos.

La siguiente tabla resume y compara ambas configuraciones:

Tabla 1: Tabla comparativa de métricas de evaluación entre el modelo 1 y 2. Fuente: Elaboración propia, 2025.

Métrica	Modelo 1 (Sin Validación)	Modelo 2 (Con Validación)
Pérdida final	0.0096 (train)	0.0817 (val) / 0.0187 (train)
IoU	0.8409 (train)	0.4213 (val) / 0.7798 (train)
Precisión	0.8957 (train)	0.6409 (val) / 0.8629 (train)
Recall	0.8738 (train)	0.4758 (val) / 0.8297 (train)
Época	100	68

Como se puede observar en la tabla 1, las métricas del modelo 2 son sistemáticamente inferiores en validación en comparación con el entrenamiento del modelo 1, pero presentan una evaluación más certera y fidedigna de la capacidad del modelo para generalizar. Es importante destacar que el modelo 2 también logró una reducción del 32% en tiempo de entrenamiento gracias a la parada anticipada.

Desde una perspectiva operativa, el modelo validado se muestra como una opción más robusta y confiable, pese a que sus cifras puedan parecer más modestas. Este resultado refuerza la idea de que el desempeño en entrenamiento no es un indicador suficiente para validar modelos destinados a aplicaciones geoespaciales o cartográficas. La validación externa, la estabilidad de las curvas de métricas y el monitoreo continuo del sobreajuste son condiciones necesarias para garantizar resultados interpretables y reproducibles.

3.4 Evaluación de clase: bosque vs. Agua

La evaluación desagregada por clase permitió detectar diferencias significativas en el comportamiento del modelo frente a las coberturas "bosque" y "agua", revelando una segmentación notablemente más eficaz en la primera clase. Esta asimetría se explica por varios factores: la clase bosque estuvo presente en la mayoría de las imágenes, otorgándole mayor representación estadística durante el entrenamiento. En cambio, la clase agua apareció en menor proporción, afectando negativamente su aprendizaje. Además, las zonas boscosas presentan patrones de textura y color más consistentes en las bandas RGB, mientras que el agua puede variar considerablemente en tonalidad según profundidad, turbidez o reflejos.

4. DISCUSIÓN

Los resultados obtenidos permiten establecer una base crítica para la evaluación del desempeño de arquitecturas convolucionales en tareas de segmentación semántica sobre imágenes satelitales RGB. Los hallazgos revelan implicancias metodológicas relevantes para futuras aplicaciones cartográficas institucionales

y contribuyen al desarrollo de mejores prácticas en contextos de datos limitados y recursos computacionales moderados.

4.1 Sobreajuste y validación simple

Los resultados ilustran una tensión fundamental en aprendizaje profundo aplicado a geoinformación: el equilibrio entre precisión técnica y utilidad operativa. La literatura en inteligencia artificial geoespacial (Tuia et al., 2011; Zhu et al., 2017) documenta consistentemente este dilema, donde modelos aparentemente exitosos fallan al enfrentar variabilidad espacial real.

Entrenar modelos sin validación externa puede generar una falsa sensación de éxito y conducir a la adopción de modelos que fallan catastróficamente al ser aplicados en nuevos contextos. Esta situación es particularmente crítica en instituciones cartográficas o ambientales, donde las decisiones basadas en estos modelos tienen implicancias operativas, económicas y legales. Por ello, la validación simple debe considerarse un mínimo obligatorio, incluso en contextos con datos escasos.

La eficiencia computacional observada (32% reducción en tiempo de entrenamiento) tiene implicaciones estratégicas para instituciones públicas latinoamericanas. Esta optimización temporal resulta crucial cuando los recursos computacionales son limitados y la sostenibilidad operativa depende de ciclos de entrenamiento eficientes.

4.2 Balance de clases y segmentación desigual

El desbalance observado refleja un problema estructural en datasets de teledetección: la sobrerrepresentación de coberturas extensas versus elementos puntuales o lineales. Esta asimetría, documentada por Buda et al. (2018), requiere estrategias correctivas específicas como funciones de pérdida ponderadas o técnicas de remuestreo adaptativo.

Además del desbalance numérico, se identificaron desafíos relacionados con la ambigüedad espectral de la clase agua, la cual fue confundida con sombras, suelos oscuros o áreas húmedas temporales. Frente a este escenario, se sugiere que futuras investigaciones consideren el uso de estrategias adicionales para mitigar el efecto del desbalance, tales como el uso de funciones de pérdida focalizadas (Lin et al., 2020), muestreo estratificado por clase o aumento sintético de muestras subrepresentadas.

4.3 Activación GELU y normalización robusta

La elección de la función de activación GELU en lugar de ReLU representó un cambio estratégico que aportó beneficios en la estabilidad del entrenamiento. GELU, introducida por Hendrycks y Gimpel (2016), permite transiciones suaves entre regiones lineales y no lineales, mejorando la propagación de gradientes y reduciendo la probabilidad de neuronas inactivas. Esta propiedad fue especialmente útil en el contexto de segmentación sobre datos naturales con alta variabilidad local.

Además, se implementó una normalización robusta de los datos de entrada mediante el escalador RobustScaler, basado en la mediana y el rango intercuartílico. A diferencia de la normalización estándar (media y desviación típica), este enfoque es menos sensible a valores atípicos, condición común en imágenes ópticas con nubosidad parcial, sombras topográficas o reflexiones especulares. Este tipo de normalización ha sido sugerido por Ioffe y Szegedy (2015) y resulta particularmente efectiva en escenarios donde las condiciones de adquisición varían significativamente entre escenas.

En conjunto, estas decisiones técnicas contribuyeron a un entrenamiento más estable, permitiendo converger sin requerir arquitecturas más profundas o mecanismos de regularización adicionales.

4.4 Comparación con estudios previos

Al comparar los resultados obtenidos con la literatura existente, se observa una correspondencia razonable en términos de desempeño general. Li et al. (2020), en su trabajo sobre detección de deforestación con imágenes Sentinel-2, lograron resultados robustos utilizando una base arquitectónica similar, empleando la arquitectura U-Net e incorporando mecanismos de atención espacial.

El valor de IoU ≈ 0.42 alcanzado en validación se considera un punto de partida coherente con la literatura, especialmente considerando que se obtuvo sin aplicar técnicas adicionales como data augmentation, sin mecanismos de atención en la arquitectura base, utilizando únicamente insumo de imágenes RGB sin información multiespectral, y entrenando en un entorno de recursos computacionales limitados.

Esto sugiere que el pipeline desarrollado constituye una línea base sólida sobre la cual se pueden incorporar mejoras incrementales, como arquitecturas más complejas (U-Net++ o Attention U-Net), funciones de pérdida adaptativas, segmentación multiclase o integración de datos multitemporales. Es fundamental reconocer que el dataset DeepGlobe, si bien sirve como estándar para validación inicial, presenta limitaciones en su representatividad para los ecosistemas específicos de Chile y Colombia. Un paso futuro ineludible será la curación de datasets locales para validar el modelo en los territorios de interés del proyecto COMIXTA, permitiendo evaluar su desempeño en condiciones geográficas y ambientales propias de la región.

5. CONCLUSIONES

Este estudio exploratorio establece una promisorio línea base metodológica para la implementación de un pipeline de segmentación semántica basado en U-Net sobre imágenes RGB del conjunto de datos DeepGlobe, incluso bajo restricciones computacionales típicas de instituciones cartográficas públicas en América Latina. Los resultados obtenidos permiten derivar un conjunto de conclusiones metodológicas, técnicas e institucionales:

1. Importancia de la validación simple como mecanismo de control del sobreajuste: La comparación experimental demostró de manera concluyente que entrenar redes profundas sin un mecanismo de evaluación intermedio conduce a un sobreajuste severo y a métricas de rendimiento engañosas. La implementación de una estrategia de validación con parada anticipada fue fundamental para obtener un modelo con predicciones más conservadoras y morfológicamente coherentes, tal como recomiendan Zhu et al. (2017) y Bengio (2012).
2. Eficacia y adaptabilidad de la arquitectura U-Net: La arquitectura U-Net, a pesar de haber sido originalmente diseñada para segmentación biomédica, demostró ser eficaz en tareas geoespaciales de clasificación binaria. Su diseño basado en encoder-decoder y skip connections permite preservar tanto el contexto global como los detalles espaciales finos, lo que resulta fundamental en la generación de mapas temáticos de alta resolución.
3. Limitaciones asociadas al conjunto de datos y desbalance de clases: La reducción del problema a una clasificación binaria facilitó el entrenamiento, pero también expuso las limitaciones inherentes al desequilibrio entre clases. La clase "agua" fue sistemáticamente subsegmentada por el modelo, reflejando la necesidad de aplicar técnicas correctivas en futuros desarrollos. Es fundamental reconocer dos limitaciones metodológicas clave. Primero, si bien el dataset DeepGlobe sirve como un estándar para la validación inicial, su representatividad para los ecosistemas específicos de Chile y Colombia es limitada. Segundo, la decisión de reducir la resolución de las imágenes de 1448x1448 a 256x256 píxeles constituye una limitación técnica significativa que afecta la delineación precisa de objetos pequeños.
4. Viabilidad técnica en entornos institucionales con recursos limitados: Este estudio demuestra que es factible prototipar y validar metodologías de segmentación semántica sin recurrir a infraestructura de alto rendimiento. Mediante el uso de software de código abierto y una GPU de gama media, se logró un flujo de trabajo eficiente, replicable y modular.
5. Relevancia estratégica en el marco de cooperación binacional COMIXTA: Este estudio se enmarca en un esfuerzo conjunto entre el IGM de Chile y el IGAC de Colombia, como parte del proyecto binacional COMIXTA. La colaboración entre ambos institutos ha permitido avanzar hacia la estandarización de

procesos de validación cartográfica asistida por algoritmos, proyectando la posibilidad de una futura red regional de producción cartográfica automatizada.

A partir de esta base, se proyectan líneas futuras de investigación orientadas a consolidar un marco metodológico institucional robusto, incluyendo la extensión del modelo a tareas multiclase, la integración de información adicional como bandas multiespectrales o datos multitemporales, y la incorporación de flujos de trabajo de georreferenciación y post-procesamiento vectorial para la producción cartográfica oficial. A partir de esta base, se proyectan líneas futuras de investigación orientadas a consolidar un marco metodológico institucional robusto, estructuradas jerárquicamente según su prioridad de implementación: Fase 1 – Optimización del modelo base: aplicar técnicas de data augmentation para balancear clases y aumentar la robustez del modelo, implementar funciones de pérdida ponderadas (Focal Loss, Dice Loss) para mitigar el desbalance de clases, y desarrollar una estrategia de inferencia por mosaico (tiling) para procesar imágenes en resolución nativa; Fase 2 – Expansión de capacidades: extender el modelo a tareas multiclase incluyendo áreas urbanas, suelo desnudo y cultivos, integrar información espectral adicional como bandas multiespectrales (NDVI, SWIR), e incorporar datos multitemporales para la detección de cambios en coberturas; Fase 3 – Escalamiento operativo: implementar flujos de trabajo de georreferenciación y post-procesamiento vectorial, validar el modelo sobre datasets locales de Chile y Colombia, y desarrollar protocolos de producción cartográfica oficial automatizada. Estos avances permitirán responder con mayor eficacia a los desafíos actuales en planificación territorial, monitoreo ambiental y generación de información geoespacial de calidad.

6. AGRADECIMIENTOS

Se agradece especialmente a la Agencia Chilena de Cooperación Internacional para el Desarrollo (AGCID) y a la Agencia Presidencial de Cooperación Internacional de Colombia (APC-Colombia), por su apoyo estratégico y financiamiento, que permitió consolidar esta alianza técnica entre Chile y Colombia en el ámbito de la IA aplicada a la cartografía.

Se extiende el debido reconocimiento al Instituto Geográfico Militar de Chile (IGM) y al Instituto Geográfico Agustín Codazzi de Colombia (IGAC) por proporcionar el marco institucional, los recursos técnicos y el respaldo necesario para el desarrollo de esta investigación. Su compromiso con la innovación cartográfica y el fortalecimiento de las capacidades geoespaciales regionales ha sido fundamental para el éxito de esta colaboración binacional.

De manera particular se agradece a todo el equipo técnico que participó en las diferentes etapas de esta investigación: desde la conceptualización metodológica hasta la implementación computacional y el análisis de resultados. Su dedicación, competencia y colaboración interdisciplinaria fueron esenciales para alcanzar los objetivos propuestos y establecer las bases para futuras investigaciones en el área de inteligencia artificial geoespacial.

Finalmente, se reconoce el apoyo de las comunidades de software libre y código abierto, particularmente los desarrolladores de PyTorch, scikit-learn y las demás bibliotecas utilizadas en este estudio, cuyas contribuciones facilitan la democratización del acceso a herramientas avanzadas de aprendizaje profundo en instituciones públicas de la región.

Esta investigación representa un paso significativo en el fortalecimiento de la cooperación técnica latinoamericana y en el desarrollo de capacidades autónomas para la producción cartográfica asistida por inteligencia artificial.

7. REFERENCIAS

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. En G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Lecture notes in computer science: Vol. 7700. Neural networks: Tricks of the trade* (2ª ed., pp. 437–478). Springer. https://doi.org/10.1007/978-3-642-35289-8_26

- Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259. <https://doi.org/10.1016/j.neunet.2018.07.011>
- Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., & Raskar, R. (2018). DeepGlobe 2018: A challenge to parse the Earth through satellite images. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 172–181). IEEE. <https://doi.org/10.1109/CVPRW.2018.00031>
- Everingham, M., van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2), 303–338.
- Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 55–72. <https://doi.org/10.1016/j.isprsjprs.2016.03.008>
- Hendrycks, D., & Gimpel, K. (2016). Gaussian Error Linear Units (GELUs). *ArXiv*. <https://arxiv.org/abs/1606.08415>
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. En *Proceedings of the 32nd International Conference on Machine Learning (ICML)* (Vol. 37, pp. 448–456). PMLR.
- Kedron, P., Frazier, A. E., Goodchild, M. F., & Li, W. (2021). Reproducibility and replicability: A new hope for quantitative geography. *Annals of the American Association of Geographers*, 111(5), 1271–1274. <https://doi.org/10.1080/24694452.2020.1863548>
- Li, X., He, Y., & Chen, Z. (2020). U-Net based deep learning for deforestation detection in Amazon rainforest using Sentinel-2 imagery. *Remote Sensing Letters*, 11(12), 1085–1094.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2020). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 318–327. <https://doi.org/10.1109/TPAMI.2018.2858826>
- Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. En *International Conference on Learning Representations (ICLR)*.
- Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), 645–657. <https://doi.org/10.1109/TGRS.2016.2612821>
- Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., & Wu, H. (2018). Mixed precision training. En *International Conference on Learning Representations (ICLR)*.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. En N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Lecture Notes in Computer Science: Vol. 9351. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>

- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958.
- Tuia, D., Volpi, M., Copa, L., & Kanevski, M. (2011). A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing*, 10(8), 1325-1337.
- Volpi, M., & Tuia, D. (2017). Fully convolutional networks for semantic segmentation of aerial images. En 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 3452-3455). IEEE.
- Zhang, F., Liu, C., & Wang, L. (2019). Water body segmentation in urban areas from high-resolution images using a U-Net deep network. *Remote Sensing*, 11(21), 2530.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36. <https://doi.org/10.1109/MGRS.2017.2762307>