



COLEÇÃO
ESTUDOS
CARIOCAS

Detección de favelas de Río de Janeiro en ortoimágenes utilizando deep learning con la arquitectura U-Net

Detection of Rio de Janeiro favelas in orthoimages using deep learning with the U-Net architecture

Detección de favelas de Río de Janeiro en ortoimágenes utilizando deep learning con la arquitectura U-Net

Jedielso Sales de Souza¹, Andrea Diniz da Silva² y Ian Monteiro Nunes³

¹ Instituto Brasileiro de Geografia e Estatística, Superintendência Estadual do Rio de Janeiro, Av. Beira Mar, 436 - Centro, Rio de Janeiro/RJ, CEP 20021-060, ORCID: 0009-0009-6053-7065, jedielso.souza@ibge.gov.br

² Instituto Brasileiro de Geografia e Estatística, Escola Nacional de Ciências Estatísticas, Rua André Cavalcanti, 106 - Santa Teresa, Rio de Janeiro/RJ, CEP 20231-050, ORCID: 0000-0001-9116-0162, andrea.diniz@ibge.gov.br

³ Instituto Brasileiro de Geografia e Estatística, Diretoria de Pesquisas, Rua Pacheco Leão, 1235 - Jardim Botânico, Rio de Janeiro/RJ, CEP 22460-905, ORCID: 0000-0003-3445-4169, ian.nunes@ibge.gov.br

Resumen

Este estudio evaluó el desempeño de la arquitectura U-Net en la identificación de asentamientos informales en ortoimágenes de alta resolución, utilizando máscaras anotadas manualmente como referencia. Se compararon modelos entrenados con y sin data augmentation. El modelo con aumento de datos presentó un mejor desempeño en términos de IoU, F1-Score y Precisión, mientras que el modelo sin *augmentation* obtuvo una mayor Revocación, evidenciando el trade-off entre sensibilidad y control de falsos positivos. A pesar de las dificultades asociadas a áreas pequeñas y de bajo contraste visual, los resultados confirman el potencial de la U-Net para el mapeo de asentamientos informales.

Palabras clave: favela, ortoimagen, deep learning, planificación urbana, geotecnologías

Abstract

This study evaluated the performance of the U-Net architecture for identifying informal settlements in high-resolution orthoimagery, using manually annotated masks as reference. Models trained with and without data augmentation were compared. The model with data augmentation achieved better performance in terms of IoU, F1-Score, and Precision, while the model without augmentation obtained higher Recall, highlighting the trade-off between sensitivity and false-positive control. Despite challenges related to small areas and low visual contrast, the results confirm the potential of U-Net for mapping informal settlements.

Keywords: favela, orthoimage, deep learning, urban planning, geotechnologies

Resumo

Este estudo avaliou o desempenho da arquitetura U-Net na identificação de favelas em ortoimagens de alta resolução, utilizando máscaras manuais como referência. Foram comparados modelos treinados com e sem data augmentation. O modelo com data augmentation apresentou melhor desempenho em IoU, F1-Score e Precisão, enquanto o modelo sem *augmentation* obteve maior Revocação, evidenciando o trade-off entre sensibilidade e controle de falsos positivos. Apesar das dificuldades em áreas pequenas e de baixo contraste visual, os resultados confirmam o potencial da U-Net para o mapeamento de assentamentos precários.

Palavras-chave: favela, ortoimagem, deep learning, planejamento urbano, geotecnologias

Volumen
13

Número
4

*Autor(a) correspondiente
jedielso.souza@ibge.gov.br

Envío 15 nov 2025

Aceptación 13 ene 2026

Publicación 05 feb 2026

¿Cómo citar?

SOUZA, J. S.; SILVA, A. D.;
NUNES, I. M. Detección de
favelas de Río de Janeiro en
ortoimágenes utilizando deep
learning con la arquitectura
U-Net. *Coleção Estudos
Cariocas*, v. 13, n. 4, 2026.
DOI 10.71256/19847203.13.4.195.2025..

El artículo fue
originalmente enviado en
PORTUGUÉS. Las
traducciones a otros
idiomas fueron revisadas
y validadas por los
autores y el equipo
editorial. Sin embargo,
para una representación
más precisa del tema
tratado, se recomienda
que los lectores
consulten el artículo en
su idioma original.



1 Introducción

El crecimiento urbano acelerado y muchas veces desordenado impone desafíos crecientes a la gestión de las ciudades, comprometiendo el desarrollo sostenible. Según el Informe Mundial de las Ciudades (ONU BRASIL, 2022), la población urbana mundial debe pasar del 56% en 2021 al 68% en 2050.

En Brasil, la urbanización estuvo históricamente asociada a la expansión de asentamientos precarios, resultado de la ausencia de políticas públicas eficaces y de la falta de acceso a vivienda e infraestructura adecuadas. En Río de Janeiro, el proceso de remoción de casas de vecindad a finales del siglo XIX llevó a familias a ocupar colinas y áreas periféricas, originando las primeras favelas (Ling, 2018; Marins, 1998).

Ante la escala global de la urbanización, la Agenda 2030 de la ONU estableció los 17 Objetivos de Desarrollo Sostenible (ODS), entre los cuales el ODS 11 – Ciudades y Comunidades Sostenibles tiene como objetivo “lograr que las ciudades sean más inclusivas, seguras, resilientes y sostenibles”, previendo la mejora y urbanización de las favelas hasta 2030 (ONU BRASIL, 2023). Uno de los principales desafíos para alcanzar esta meta es la identificación y el monitoreo continuo de los asentamientos precarios, esenciales para la planificación urbana y la formulación de políticas públicas.

Tradicionalmente, esta información se obtiene mediante los censos demográficos, realizados cada diez años, lo que limita el seguimiento de las transformaciones en el período intercensal. En este contexto, el uso de imágenes de satélite y técnicas de inteligencia artificial surge como una alternativa prometedora, permitiendo identificar, mapear y monitorear favelas de forma más ágil y precisa.

La teledetección ha sido ampliamente empleada para la identificación y el monitoreo de asentamientos informales, aprovechando imágenes de satélite de diferentes resoluciones espaciales y temporales en contextos urbanos diversos. Estudios demuestran que las características espectrales, de textura, geométricas y morfológicas extraídas de estas imágenes permiten distinguir áreas formales e informales, especialmente cuando se combinan con técnicas de procesamiento digital de imágenes y aprendizaje automático (Kemper *et al.*, 2015; Alrasheedi *et al.*, 2021; Cinnamon; Noth, 2023).

Los enfoques tradicionales, como el análisis basado en objetos y los algoritmos de aprendizaje automático clásico, presentan buenos resultados, aunque demandan extracción y selección manual de características y ajustes específicos para cada contexto urbano (Ghaffarian; Emtehani, 2021; Oliveira *et al.*, 2023). En los últimos años, los métodos basados en aprendizaje profundo, especialmente las redes neuronales convolucionales y los modelos de segmentación semántica, se han consolidado como el estado del arte en el análisis de imágenes de satélite, demostrando una mayor capacidad para capturar la complejidad morfológica y la heterogeneidad espacial características de las favelas (Wurm *et al.*, 2019; Maiya; Babu, 2018; Lu *et al.*, 2021; Abascal *et al.*, 2022). Estos enfoques amplían el potencial de mapeo y monitoreo continuo de los asentamientos informales, proporcionando subsidios más precisos para la planificación urbana y la formulación de políticas públicas.

El presente trabajo busca contribuir a este esfuerzo, aplicando modelos de aprendizaje profundo en imágenes de satélite para la identificación y el monitoreo de la dinámica de las favelas en la ciudad de Río de Janeiro.

2 Materiales y Métodos

2.1 Identificación y Delimitación de las Áreas de Interés

En el estudio se utilizaron ortofotografías RGBI con resolución espacial de 15 cm (Ground Sample Distance - GSD), adquiridas durante un vuelo fotogramétrico realizado en abril de 2024 por la cámara aerofotogramétrica UltraCam Osprey 4.1. Los datos fueron cedidos por el Instituto Municipal de Urbanismo Pereira Passos (IPP), órgano de la Prefectura de la Ciudad de Río de Janeiro. Las ortofotografías están compuestas por las bandas espectrales del rojo (Red), verde (Green), azul (Blue) e infrarrojo cercano (Near Infrared - NIR).

Las imágenes se encuentran en el sistema de referencia de coordenadas EPSG:31983, que corresponde al sistema geodésico SIRGAS 2000 asociado a la proyección cartográfica Universal Transversa de Mercator (UTM), huso 23 Sur (meridiano central 45°W). Este sistema de referencia es adecuado para el área de estudio, ya que el municipio de Río de Janeiro está ubicado cerca del meridiano de secancia del huso, lo que contribuye a la minimización de las distorsiones lineales y de área en los análisis espaciales.

Para este estudio, se utilizaron exclusivamente las bandas del espectro visible (RGB). Considerando el elevado costo computacional asociado al procesamiento de imágenes de gran extensión territorial, se elaboró una cuadrícula espacial compuesta por cuadrados de 512 × 512 metros, delimitada por el área del municipio de Río de Janeiro.

Con el objetivo de reducir la resolución espacial de las imágenes, originalmente de 15 cm, se realizó la remezcla de píxeles por el método bilineal, resultando en una resolución de 50 cm. Esta nueva resolución preserva una adecuada capacidad de identificación de los componentes urbanos, al mismo tiempo que reduce significativamente el costo computacional del procesamiento.

Los cuadrados localizados en islas deshabitadas fueron descartados, mientras que aquellos situados en la frontera municipal se mantuvieron únicamente cuando al menos un tercio de su área se encontraba dentro de los límites del municipio.

Al final de este proceso, la cuadrícula resultante pasó de 5.002 a 4.647 cuadrados. Estos cuadrados fueron, entonces, utilizados para recortar la imagen del municipio de Río de Janeiro (con resolución espacial de 1 m), generando múltiples imágenes menores y, por lo tanto, más viables para el análisis computacional.

La ciudad de Río de Janeiro, conforme definido en su Plan Director de Desarrollo Urbano Sostenible (PDUS), está subdividida en diferentes unidades territoriales destinadas a la planificación y control del desarrollo urbano. Entre estas unidades se encuentran las Áreas de Planificación (APs), cinco áreas definidas a partir de criterios de compartimentación ambiental, características históricas y patrones de uso y ocupación del suelo (Río de Janeiro, 2025):

- Área de Planificación 1 (AP1) – corresponde a la región central de la ciudad;
- Área de Planificación 2 (AP2) – abarca la Zona Sur y la Gran Tijuca;
- Área de Planificación 3 (AP3) – comprende la Zona Norte, con excepción de la Gran Tijuca y de la Región Administrativa VII (San Cristóbal);
- Área de Planificación 4 (AP4) – corresponde a la Región Suroeste, creada más recientemente;
- Área de Planificación 5 (AP5) – engloba los barrios de la Zona Oeste.

Con el objetivo de asegurar una representatividad territorial amplia, se optó por utilizar las Áreas de Planificación definidas oficialmente por la Prefectura. Esta decisión considera que las diferencias físicas y socioeconómicas de la población ejercen influencia directa sobre la configuración y las características del espacio urbano.

Así, cada cuadrado de la cuadrícula fue asociado a una Área de Planificación (AP) del municipio (Figura 1), conforme a la división oficial disponible en la plataforma DATA.RIO (2023). En los casos en que un cuadrado se superpusiera a dos o más APs, se establecieron criterios específicos de atribución, asegurando la consistencia de la clasificación territorial:

1. Los cuadrados que contenían regiones clasificadas como favelas por el IBGE o por la Prefectura de Río fueron atribuidos a la AP que englobara esa área clasificada;
2. Los cuadrados sin áreas de favela, o que contenían áreas de favela en más de una AP, fueron designados a la AP que presentara la mayor extensión territorial dentro del cuadrado.

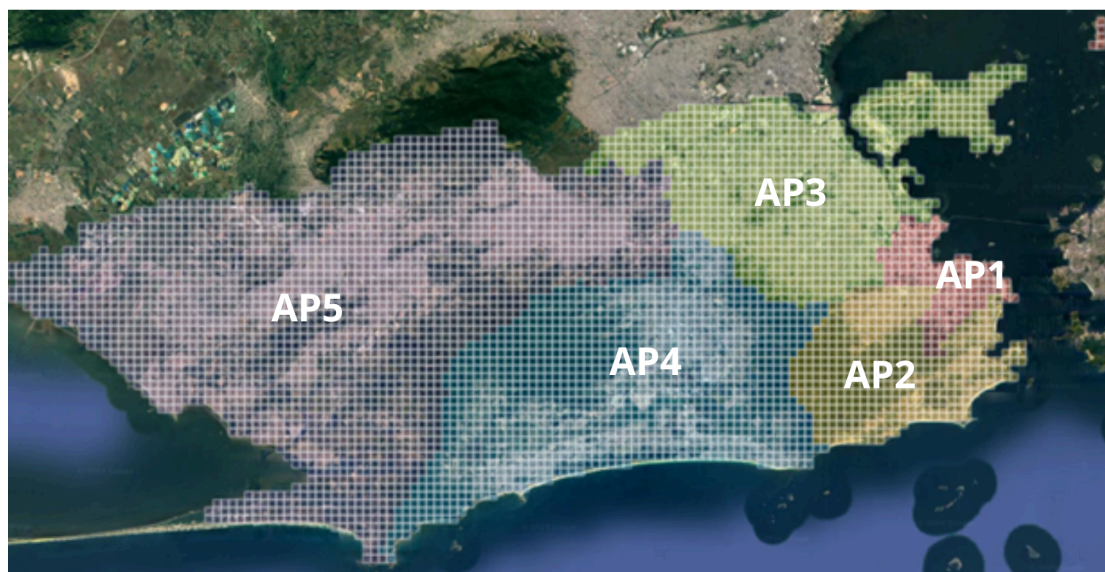


Figura 1: Cuadrícula de cuadrados de 512 m² aplicada sobre la imagen de la ciudad de Río de Janeiro. La malla fue utilizada para segmentar el área urbana en recortes menores y estandarizados, facilitando el procesamiento de las imágenes de satélite. Distribuida por las Áreas de Planificación de la Prefectura de Río de Janeiro.

Fuente: Elaboración propia

La identificación de los cuadrados de la cuadrícula que contenían las áreas de interés se realizó con base en datos espaciales vectoriales en formato shapefile. Estos datos corresponden a la malla de sectores censales del Censo 2022, disponible por el IBGE, inicialmente clasificados como Aglomerados Subnormales (AGSN), denominación posteriormente actualizada a Favelas y Comunidades Urbanas (FCU).

El cambio de nomenclatura refleja un proceso de revisión conceptual conducido por el IBGE, en diálogo con movimientos sociales, la comunidad académica y otros órganos gubernamentales. Esta actualización busca adoptar una terminología más adecuada y respetuosa, sustituyendo la expresión “Aglomerados Subnormales”, que venía siendo considerada inadecuada por reforzar estigmas sociales. Sin embargo, es importante destacar que la alteración fue apenas terminológica, no habiendo cambios en los criterios técnicos utilizados para la identificación y el mapeo de estas áreas (IBGE, 2024).

Complementariamente, se incorporaron datos del Instituto Municipal de Urbanismo Pereira Passos (IPP). Estos datos, disponibles en el portal DATA.RIO (2019), delimitan las áreas oficialmente reconocidas como favelas.

Para el IBGE, la caracterización de un aglomerado subnormal ocurre cuando hay ocupación irregular de la tierra, asociada a al menos una de las siguientes condiciones (IBGE, 2020):

- Precariedad de servicios públicos esenciales, como abastecimiento de agua, suministro de energía eléctrica, recolección de basura o saneamiento;
- Patrón urbanístico irregular, reflejado en la presencia de vías de circulación estrechas, alineamiento irregular, lotes de tamaños y formas desiguales, ausencia de aceras o construcciones no regularizadas por órganos públicos;
- Restricción a la ocupación del suelo, cuando los domicilios se encuentran en áreas protegidas por legislación ambiental, franjas de dominio de carreteras o ferrovías, áreas contaminadas, entre otras situaciones de uso inadecuado del suelo urbano.

Además, después de la identificación y delimitación de las áreas, el IBGE asocia los aglomerados subnormales a las unidades operacionales del Censo Demográfico, denominadas sectores censales. Cada sector debe comprender áreas contiguas con, al menos, 51 domicilios, respetando límites político-administrativos y garantizando la coherencia territorial y operacional de la recolección censal (IBGE, 2020).

Por su parte, el IPP — órgano vinculado a la Prefectura de Río de Janeiro — utiliza el término favela con un enfoque socioespacial y administrativo, buscando reconocer y mapear territorios consolidados en el tejido urbano. La definición adoptada por el IPP toma en cuenta un conjunto más amplio de características, que incluye tanto aspectos físicos como sociales y catastrales. Según el Instituto, un área es considerada favela cuando presenta (Río de Janeiro, 2012):

1. Ocupación irregular de la tierra;
2. Ausencia de títulos de propiedad formales (lo que no implica ilegalidad de la ocupación);
3. Tejido urbano dispuesto de forma irregular;
4. Lotes pequeños e indefinidos;
5. Vías estrechas;
6. Infraestructura de saneamiento precaria;
7. Equipamientos sociales inexistentes o insuficientes;
8. Viviendas precarias y en desacuerdo con las normas urbanísticas;
9. Inexistencia de normas urbanísticas especiales aplicables al área;
10. No inserción de los inmuebles en los catastros inmobiliarios municipales;
11. Predominio de población de bajos ingresos.

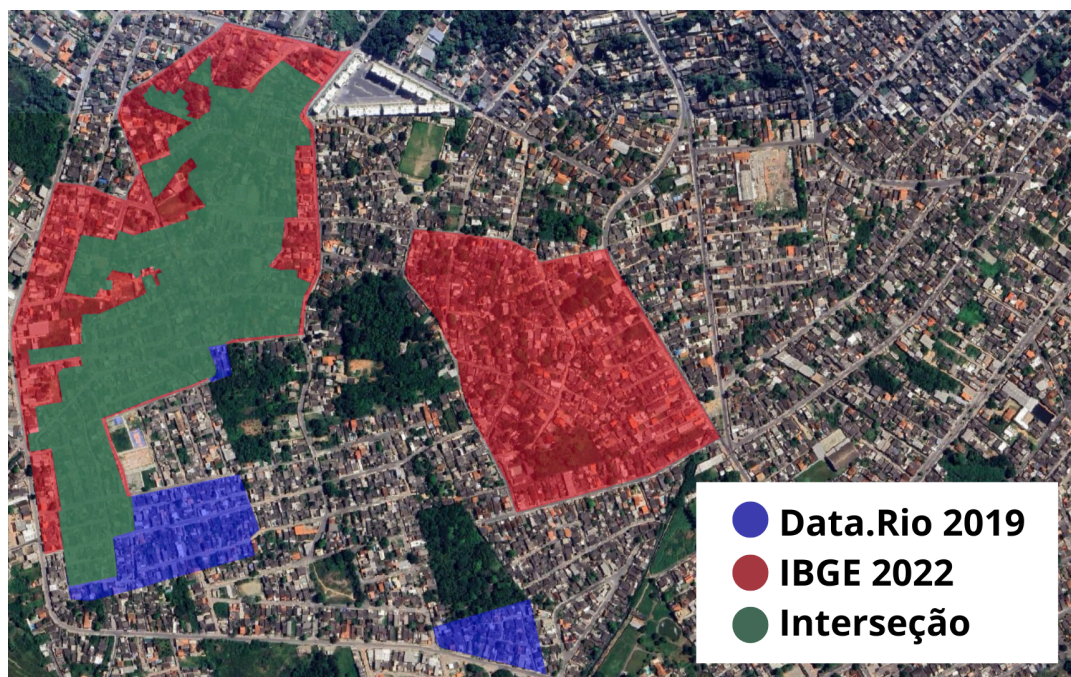


Figura 2: Áreas de favelas clasificadas por el IBGE y por el Data.Rio. La superposición de las dos bases permite visualizar áreas de concordancia y divergencia entre las clasificaciones, evidenciando tramos reconocidos por ambas fuentes y regiones identificadas por solo una de ellas.

Fuente: Elaboración propia

La superposición de los polígonos de las dos bases permitió identificar discrepancias de clasificación, revelando áreas reconocidas como Favelas y Comunidades Urbanas (FCU) por el IBGE, pero no registradas como favelas en el DATA.RIO, y viceversa (Figura 2). La consolidación de estas bases permitió ampliar el conjunto de áreas preclasificadas, de modo de extraer de la cuadrícula de cuadrados únicamente los recortes de imagen que presentaban alguna identificación previa de favela en al menos una de las fuentes.

Como resultado, se observó una reducción significativa en el número de imágenes a ser procesadas. Al final, fueron identificados 1.512 cuadrados potencialmente asociados a la presencia de áreas de favela (Figura 3).

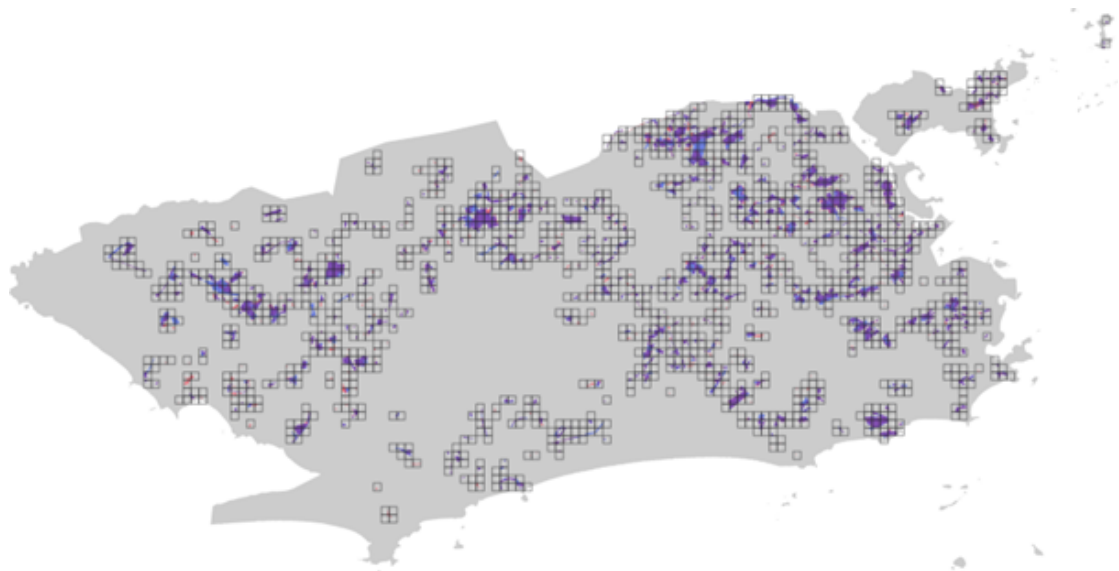


Figura 3: Superposición de la cuadrícula de cuadrados con las áreas de favelas clasificadas por el IBGE y por el Data.Rio. La combinación de estas capas permite identificar qué recortes de la cuadrícula presentan indicios de contener favelas dentro de la escena, con base en las clasificaciones previamente realizadas por los órganos.

Fuente: Elaboración propia

2.2 Muestra de Entrenamiento

2.2.1 Identificación de las áreas de favelas

Las imágenes de la cuadrícula espacial resultante de la superposición de las áreas del IBGE y del Data.Rio (Figura 3) fueron consideradas como la población objetivo, de la cual fue extraída la muestra de entrenamiento. Fue seleccionada una muestra estratificada con muestreo inverso, correspondiente al 30% de los cuadrados pertenecientes a las Áreas de Planificación 1 a 4.

El Área de Planificación 5 fue excluida en esta etapa debido al elevado esfuerzo necesario para la creación de las máscaras de referencia. En razón de esta limitación operacional, la AP5 no fue incluida en la muestra actual, estando prevista su incorporación en estudios futuros.

En el muestreo estratificado, la población (U) se divide en estratos (H) — grupos distintos y exhaustivos, homogéneos respecto a la variable de interés. La selección ocurre de forma independiente en cada estrato, y la muestra final resulta de la unión de las unidades seleccionadas (Silva; Bianchini; Dias, 2023).

En este estudio, los estratos corresponden a las Áreas de Planificación, una vez que esta división se basa en criterios previamente mencionados que influyen en el patrón urbanístico. La estratificación espacial tiende a aumentar significativamente la eficiencia del muestreo, especialmente en regiones con características heterogéneas (Dong *et al.*, 2022).

Para la selección dentro de los estratos, se utilizó muestreo inverso simple, un procedimiento secuencial en el que, en lugar de seleccionar directamente un número fijo n de unidades, se evalúan sucesivamente las unidades de la población hasta que n de ellas satisfagan la condición de interés (Silva; Bianchini; Dias, 2023). En el presente estudio, esto correspondió a evaluar m imágenes hasta que n de ellas contuvieran áreas de favelas y comunidades urbanas.

En total, fueron evaluadas 358 imágenes, correspondientes a los cuadrados seleccionados en las APs 1 a 4. De estas, 286 imágenes (30% de las elegibles) compusieron la muestra final, mientras que 72 fueron descartadas (Tabla 1).

Tabla 1: Cantidad de imágenes de la muestra en las Áreas de Planificación

AP (<i>H</i>)	Población (<i>U</i>)	Evalradas (<i>m</i>)	Descartadas (<i>r</i>)	Muestra (<i>n</i>)
1	70	23	2	21
2	117	53	18	35
3	444	157	24	133
4	322	125	25	97

Fuente: Elaboración propia

La identificación de las áreas de favela se realizó por medio del análisis de la cuadrícula de cuadrados de la muestra en Google Earth Pro. En las regiones con cobertura de Street View, se procedió a la inspección visual directa de las edificaciones. En las áreas sin esta cobertura, la evaluación se basó en el relieve del terreno y en las construcciones en 3D, examinadas desde múltiples ángulos y puntos de referencia dentro de las imágenes, lo que garantizó una identificación visual consistente de las áreas de interés (Figura 4).



Figura 4: Visualización, en Google Earth Pro, de la superposición de la cuadrícula de cuadrados con las áreas de favelas clasificadas por el IBGE y por el Data.Rio. El análisis incluye la inspección de las vías internas por medio del Street View y, en los lugares donde esta herramienta no está disponible, la observación del relieve y de las construcciones en 3D. La fijación de elementos de referencia (círculo amarillo) auxilia en la localización geoespacial de cada imagen durante el proceso de análisis visual.

Fuente: Elaboración propia

Simultáneamente a la identificación visual de las favelas en Google Earth Pro, las imágenes de la muestra fueron segmentadas manualmente en la aplicación de dibujo Procreate, en una tableta, para la producción de las máscaras de referencia utilizadas en el entrenamiento y prueba del modelo. En las máscaras, las áreas de favelas fueron rellenadas con el color blanco, mientras que las demás áreas urbanas recibieron el color negro. Durante el proceso, se buscó contornear elementos que, aunque ubicados dentro de las favelas, no constituyen viviendas, como árboles, calles y campos de fútbol, con el fin de refinar el delineamiento de las áreas de interés (Figura 5).

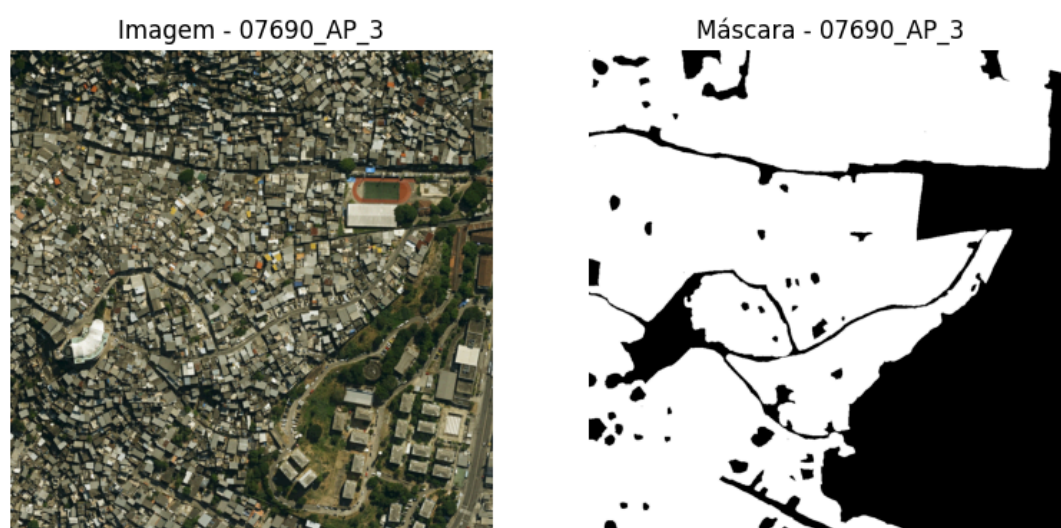


Figura 5: Imagen de satélite 7690, correspondiente a parte del Complexo do Alemão (Área de Planificación 3). En la esquina inferior derecha, se observa el Colégio Estadual Jornalista Tim Lopes y, un poco más arriba, la Vila Olímpica Carlos Castilho. La figura presenta también la máscara de referencia producida en el estudio: las áreas en blanco corresponden a las regiones clasificadas como favela después del proceso de análisis y segmentación manual de la imagen, mientras que las áreas en negro representan las demás regiones urbanas no clasificadas como favela.

Fuente: Elaboración propia

En total, fue analizada un área aproximada de 75 km², cuyas imágenes fueron integralmente segmentadas de forma manual. Del total del área segmentada, el 12,10% corresponde a la variable de interés, es decir, a las áreas de favelas. Este porcentaje es superior al de las áreas oficialmente demarcadas como favelas por el IBGE y por el IPP cuando se considera la totalidad de la extensión territorial del municipio de Río de Janeiro, en las cuales esas áreas representan menos del 5%.

La muestra fue dividida en dos grupos distintos: entrenamiento y prueba. Del total, el 80% de las imágenes (229) fueron destinadas al entrenamiento del modelo, mientras que el 20% (57) compusieron el conjunto de prueba (Tabla 2). Las imágenes fueron distribuidas entre las cuatro Áreas de Planificación (Figura 6).

Tabla 2: Cantidad de imágenes de entrenamiento y prueba

AP	Entrenamiento	Prueba
1	17	4
2	28	7
3	106	27
4	78	19

Fuente: Elaboración propia

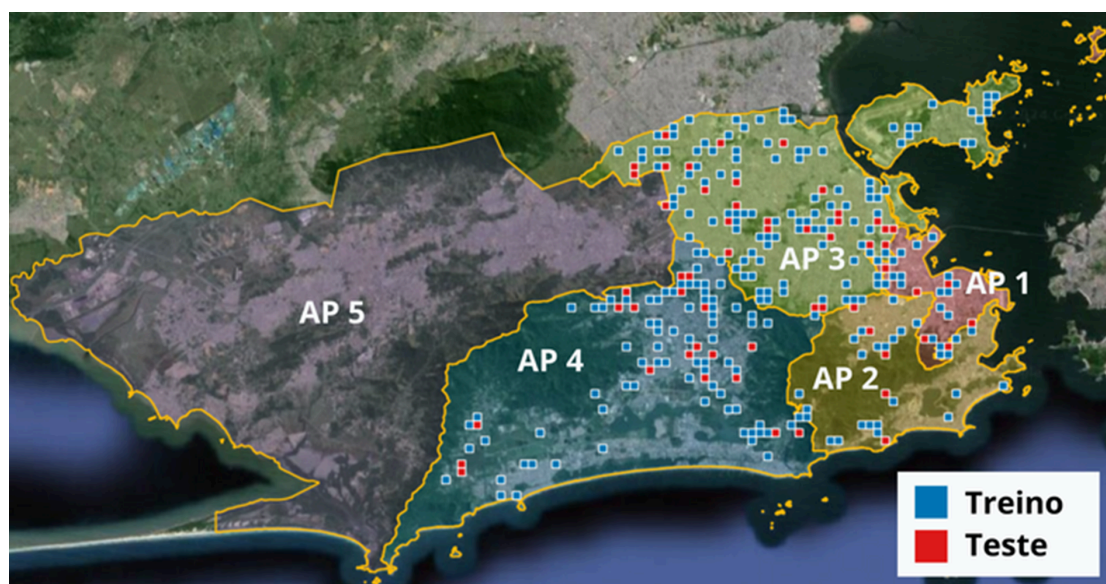


Figura 6: Distribución espacial de los conjuntos de entrenamiento y prueba por las Áreas de Planificación 1 a 4, correspondientes a los estratos definidos en el proceso de muestreo.
Fuente: Elaboración propia

2.2.2 Criterios para selección de áreas de favelas

El descarte de parte de las imágenes durante el proceso de muestreo inverso fue realizado por medio de inspección visual individual, con el objetivo de garantizar que solo aquellas que presentaran píxeles correspondientes a construcciones características de favelas fueran incluidas en la muestra.

La exclusión de imágenes ocurrió por la ausencia de construcciones típicas de favelas dentro de la escena analizada. Este hecho se debe a que las delimitaciones espaciales utilizadas — tanto del IBGE como del Data.Rio — frecuentemente presentan extensión territorial superior al área efectiva de la favela. De esta forma, en el proceso de superposición de estas áreas con la cuadrícula de cuadrados utilizada para definir la población, fueron incluidas imágenes que no contenían favelas (Figura 7).

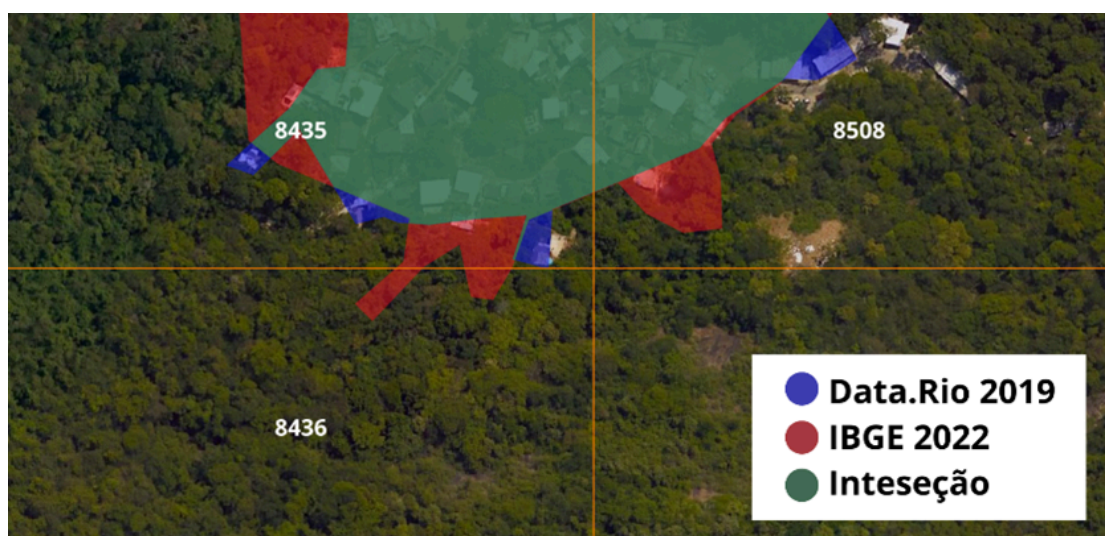


Figura 7: La imagen 8436, ubicada en las proximidades del Complexo do Turano, en el barrio de Río Comprido, fue descartada. Aunque la base del IBGE indica una pequeña área de favela dentro del recorte, no hay construcciones con características compatibles en esa porción de la escena. A diferencia, las imágenes 8435 y 8508 presentan edificaciones que justifican su inclusión en el análisis.

Fuente: Elaboración propia

Por tratarse de un análisis basado en imágenes, la identificación de las áreas de favela se fundamentó en un conjunto de criterios visuales (Figura 8):

1. **Patrón irregular de las construcciones:** disposición desordenada de las edificaciones, sin indicios de planificación urbana.
2. **Materiales constructivos precarios:** uso de materiales improvisados o de baja durabilidad.
3. **Localización en laderas:** presencia de construcciones en terrenos inclinados, muchas veces sujetos a riesgos de deslizamientos.
4. **Proximidad a carreteras y ferrovías:** ocupaciones en franjas restrictas junto a infraestructuras de transporte.
5. **Proximidad a cunetas/alcantarillas:** ocurrencia de edificaciones junto a alcantarillas a cielo abierto o cursos de agua canalizados de forma precaria.
6. **Calles estrechas:** vías de ancho reducido, generalmente incompatibles con la circulación de vehículos de emergencia o transporte público.



Figura 8: Criterios visuales para identificación de favelas. Las áreas destacadas presentan: (1) patrón irregular de construcciones sin planificación urbana; (2) materiales constructivos precarios; (3) ocupación en laderas; (4) proximidad a carreteras y ferrovías en franjas restrictas; (5) edificaciones junto a cunetas y alcantarillas a cielo abierto; (6) calles estrechas incompatibles con circulación de vehículos de emergencia.

Fuente: Elaboración propia

2.3 Aumento de Imágenes para Ampliación del Conjunto de Entrenamiento

Con el objetivo de ampliar el conjunto de entrenamiento sin la necesidad de producir nuevas máscaras manualmente, se utilizó la biblioteca *Albumentations* (Buslaev, 2018), que permite aplicar transformaciones rápidas, eficientes y flexibles en las imágenes, como reflejos, rotaciones y distorsiones geométricas.

Fueron aplicadas al conjunto original (OR) de imágenes (Figura 9) las siguientes transformaciones:

- **HorizontalFlip (HF)** – reflejo horizontal de la imagen;
- **VerticalFlip (VF)** – reflejo vertical de la imagen;
- **Transpose (TP)** – transposición de la imagen, invirtiendo filas y columnas.

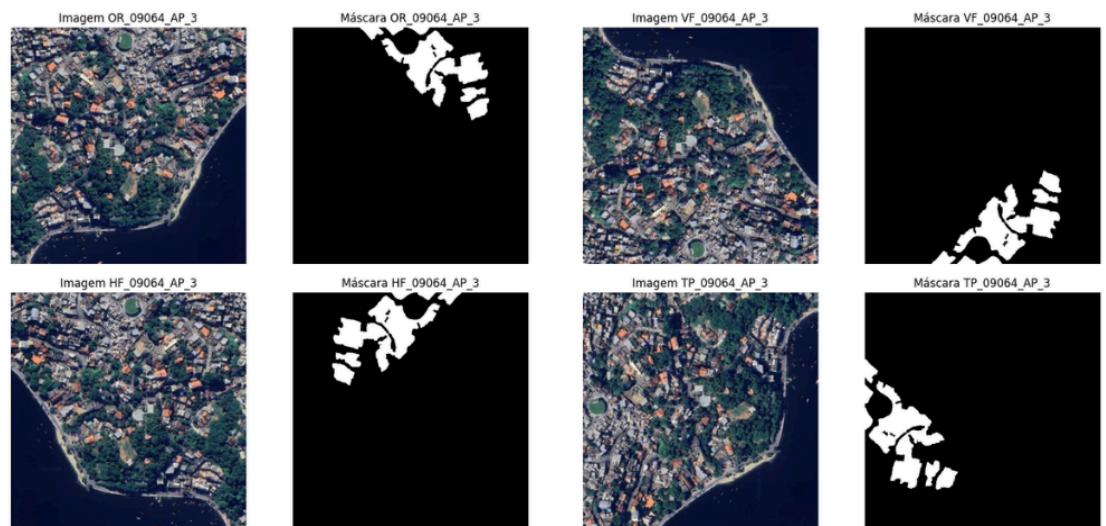


Figura 9: Imagen 09064, ubicada en el barrio de Cocotá, en la Ilha do Governador, en las proximidades de la Praia de Cocotá (Área de Planificación 3). La imagen original fue sometida, por medio de la biblioteca Albumentations, a las transformaciones de reflejo horizontal, reflejo vertical y transposición de filas y columnas. Las mismas transformaciones fueron aplicadas a la máscara de referencia correspondiente, asegurando la consistencia del par imagen-máscara después del aumento de datos.

Fuente: Elaboración propia

A partir de estas transformaciones, fueron definidos dos conjuntos de datos para el entrenamiento del modelo:

1. **Modelo entrenado sin *data augmentation*;**
2. **Modelo entrenado con *data augmentation*.**

En el modelo entrenado con *data augmentation*, después del aumento de imágenes, el número de imágenes de entrenamiento se volvió cuatro veces superior al del conjunto original (sin *data augmentation*), mientras que el conjunto de prueba permaneció inalterado.

2.4 Identificación Automática de las Áreas de Favelas

La identificación automática de las áreas de favelas en las imágenes de satélite fue realizada a partir del conjunto de imágenes de entrenamiento, procesado por medio de un modelo basado en la arquitectura U-Net (Ronneberger; Fischer; Brox, 2015). Esta red neuronal convolucional (CNN) está diseñada específicamente para tareas de segmentación semántica, en las cuales cada píxel de la imagen es clasificado en una categoría específica.

La aplicación de redes neuronales profundas, en particular la arquitectura U-Net, para el mapeo de asentamientos precarios a partir de imágenes de satélite ha sido explorada en estudios recientes, demostrando resultados prometedores.

Lu *et al.* (2021) propusieron GASlum-Net, una arquitectura innovadora que combina los principios de la U-Net y la ConvNeXt en una estructura de dos corrientes para integrar imágenes RGB (del satélite Jilin-1, 5 m) y características geoespaciales derivadas del Sentinel-2. Su modelo, entrenado en 2.892 imágenes de 64x64 píxeles, superó significativamente los modelos de referencia (U-Net, ConvNeXt-UNet y FuseNet), alcanzando una mejora de hasta el 10,97% en el IoU. El estudio destacó la eficacia del modelo en la detección de favelas de mediano y gran tamaño (5 a >25 ha) y analizó el impacto de diferentes características y ajustes de la arquitectura en el desempeño.

Complementariamente, Abascal *et al.* (2022) aplicaron la U-Net para mapear patrones de privación urbana en Nairobi, utilizando imágenes de altísima resolución (WorldView-3, 30 cm). El modelo alcanzó una precisión de 0,92 y un IoU de 0,73 en el conjunto de prueba, validando su eficacia para extracción de características urbanas. El estudio también reveló limitaciones en la distinción de edificaciones individuales en áreas de alta densidad, donde el tamaño del píxel superaba los espacios entre construcciones. Para superar esto, los autores recurrieron a un análisis morfológico agregado, clasificando áreas en niveles de privación (Alta, Media, Baja) con una precisión general de 0,71. La investigación concluye que la extracción semiautomática de características es viable y que la combinación de deep learning con análisis espacial proporciona información valiosa para la planificación urbana.

La principal característica de esta arquitectura es su estructura simétrica en forma de “U”, compuesta por dos partes complementarias, el encoder y el decoder (Figura 10).

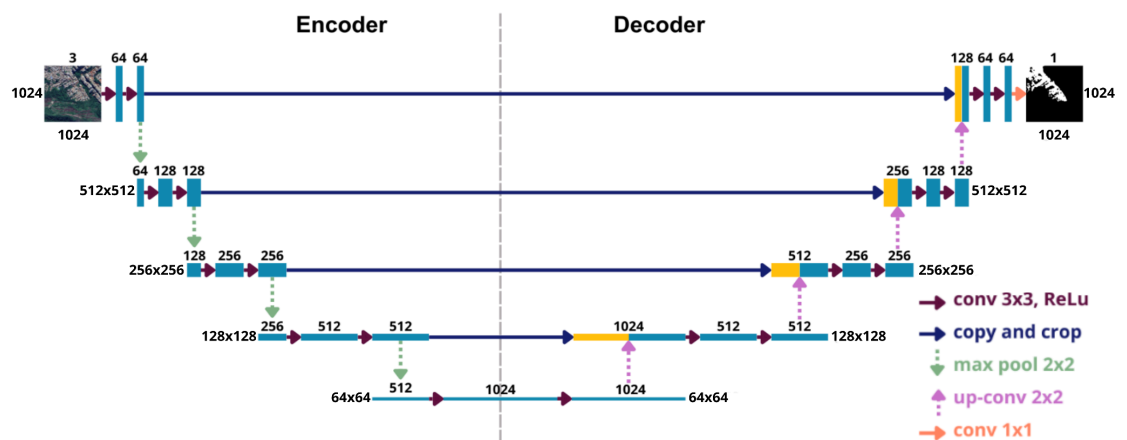


Figura 10: Ilustración de la arquitectura U-Net aplicada a imágenes de 1024 × 1024 píxeles con tres canales, presentada en su formato simétrico característico, compuesto por los bloques de *encoder* (contracción) y *decoder* (expansión).

Fuente: Elaboración propia

Las cajas azules en la Figura 10 representan los mapas de características extraídas de las imágenes, mientras que las cajas amarillas muestran las copias de estas características utilizadas posteriormente. Las flechas indican las diferentes operaciones realizadas a lo largo de la red.

En el lado izquierdo está el encoder, o camino de contracción, responsable de extraer características de la imagen de entrada. A medida que la imagen pasa por las capas convolucionales (flecha roja) seguidas de la aplicación de la función de activación ReLU (*Rectified Linear Unit*), la red identifica patrones en diferentes niveles, aumentando el número de canales. En seguida, las capas de *max pooling* 2×2 (flecha verde) reducen la resolución espacial, permitiendo que la red capture información más general.

En el lado derecho está el decoder, o camino de expansión, responsable de recuperar la resolución de las características y reconstruir el mapa de segmentación. Este proceso es realizado por medio de convoluciones transpuestas (flecha rosa), que aumentan la resolución espacial, seguido de la aplicación de la función de activación ReLU y de nuevas capas convolucionales (flecha roja), que refinan los detalles recuperados.

Entre los dos caminos están las conexiones de salto (flecha azul), que conectan directamente las capas de la contracción a las capas correspondientes de la expansión. Estas conexiones concatenan los mapas de características de los dos

lados, permitiendo combinar detalles finos — preservados en el *encoder* — con la información más abstracta del *decoder*, generando segmentaciones más precisas.

Por último, la capa de salida (flecha naranja) es una capa convolucional responsable de producir la segmentación final, con la misma resolución de la imagen original.

Durante el entrenamiento del modelo, fue utilizada como función de pérdida la *Dice Loss*, definida como $1 - \text{Coeficiente de Dice}$ (Figura 11). Esta elección fue motivada por el desbalance entre las clases, una vez que las áreas de favelas ocupan una proporción significativamente menor de la imagen en comparación a las demás clases urbanas. La Dice Loss es especialmente adecuada para tareas de segmentación en escenarios desbalanceados, pues penaliza de forma más severa errores en la predicción de la clase minoritaria y favorece la maximización de la superposición entre las máscaras predichas y de referencia.

Dice Loss = 1 - Coef. Dice

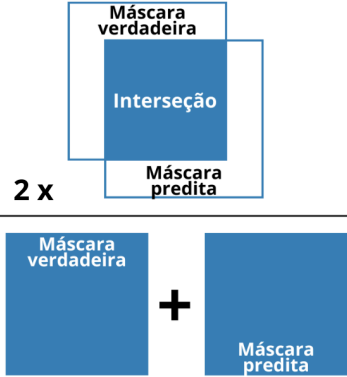
$$\text{Coeficiente de Dice} = \frac{2 \times \text{Área da interseção}}{\text{Área total}} \Leftrightarrow 2 \times \frac{\text{Interseção}}{\text{Máscara verdadeira} + \text{Máscara predita}}$$


Figura 11: Ilustración del concepto del Coeficiente de Dice y *Dice Loss*, mostrando la relación entre el doble del área de intersección de las regiones de la máscara verdadera y de la máscara predicha y la suma de las áreas de ambas máscaras.

Fuente: Elaboración propia

Para ambos conjuntos de datos, el entrenamiento fue realizado con un número máximo de 400 épocas, utilizando *early stopping* con paciencia de 50 épocas, técnica que interrumpe el entrenamiento cuando no hay mejora en el desempeño en validación por un determinado número de épocas, evitando sobreajuste. En ambos casos, se adoptó un batch size de 8 y una tasa de aprendizaje de 1×10^{-4} . En el Conjunto de Datos 2, el uso de data augmentation aumentó el número efectivo de muestras, permitiendo mantener esta configuración y garantizando una comparación justa entre los modelos.

La implementación del modelo, así como todo el pipeline de entrenamiento y evaluación, fue desarrollada en el lenguaje Python, utilizando bibliotecas especializadas en aprendizaje profundo, con destaque para TensorFlow y Keras en la construcción de la arquitectura neuronal. El preprocesamiento de las imágenes fue realizado con OpenCV (cv2) y NumPy, mientras que la evaluación cuantitativa empleó métricas de la biblioteca scikit-learn.

2.5 Métricas para Evaluación del Desempeño de los Modelos

Las métricas de evaluación utilizadas en este estudio fueron el IoU (*Intersection over Union*), *F1-Score* (o Estadística F), Sensibilidad (*Recall*) y Precisión (*Precision*). Ambas varían entre 0 y 1, siendo que valores más cercanos a 1 indican mayor precisión del modelo en la tarea de segmentación.

El IoU, también conocido como índice de Jaccard, mide el grado de superposición entre la máscara predicha y la máscara verdadera, siendo calculado por la razón entre el área de intersección y el área de unión de las dos máscaras (Lakshmanan;

Görner; Gillard, 2021). Cuanto mayor el valor del IoU, más precisa es la segmentación, lo que lo torna especialmente relevante en tareas que demandan alta exactitud espacial (Figura 12).

$$\text{IoU} = \frac{\text{Área da interseção}}{\text{Área da união}}$$



Figura 12: Ilustración del concepto de IoU (Intersection over Union), mostrando el área de intersección entre la máscara verdadera y la máscara predicha en relación al área de unión entre ambas.

Fuente: Elaboración propia

Para una comprensión más profunda de las demás métricas, es fundamental conocer los conceptos asociados a la matriz de confusión (Figura 13): Verdadero Positivo (VP), Verdadero Negativo (VN), Falso Positivo (FP) y Falso Negativo (FN) (Provost; Fawcett, 2016).

	Valor predito: Verdadeiro	Valor predito: Falso
Valor real: Verdadeiro	Verdadeiro Positivo (VP) <small>Predição correta</small>	Falso Negativo (FN) <small>Erro do Tipo II</small>
Valor real: Falso	Falso Positivo (FP) <small>Erro do Tipo I</small>	Verdadeiro Negativo (VN) <small>Predição correta</small>

Figura 13: Matriz de confusión, que compara los valores predichos del modelo con los valores reales, organizando los resultados en cuatro categorías: Verdaderos Positivos (VP) y Verdaderos Negativos (VN) representan aciertos; Falsos Positivos (FP) y Falsos Negativos (FN) representan los errores.

Fuente: Elaboración propia

- **Verdadero Positivo (VP):** Caso en que el modelo predice la clase positiva y esta predicción corresponde a la clase positiva real. Es decir, el modelo identifica correctamente la presencia de la clase de interés.
- **Verdadero Negativo (VN):** Situación en la cual tanto la predicción del modelo como la realidad pertenecen a la clase negativa. De esta forma, el modelo reconoce adecuadamente la ausencia de la clase de interés.
- **Falso Positivo (FP):** Ocurre cuando el modelo predice la clase positiva, pero la clase real es negativa. En otras palabras, el modelo señala erróneamente la presencia de la clase de interés.
- **Falso Negativo (FN):** Caso en que el modelo predice la clase negativa, pero la clase real es positiva. Esto significa que el modelo falla al no identificar la presencia de la clase de interés.

Precisión y Sensibilidad son métricas complementarias ampliamente utilizadas en la evaluación de modelos de clasificación, especialmente en contextos de desbalance de clases (Figura 14). Mientras que la Precisión está asociada a la confiabilidad de las predicciones positivas realizadas por el modelo, la Sensibilidad refleja su capacidad de identificar correctamente los casos positivos existentes (Provost; Fawcett, 2016):

- **Precisión:** evalúa la proporción de predicciones positivas que son efectivamente correctas, siendo calculada por la razón entre los verdaderos positivos (VP) y la suma de los verdaderos positivos y de los falsos positivos (VP + FP).
- **Sensibilidad:** mide la capacidad del modelo de identificar todos los casos positivos reales, correspondiendo a la razón entre los verdaderos positivos (VP) y la suma de los verdaderos positivos y de los falsos negativos (VP + FN).

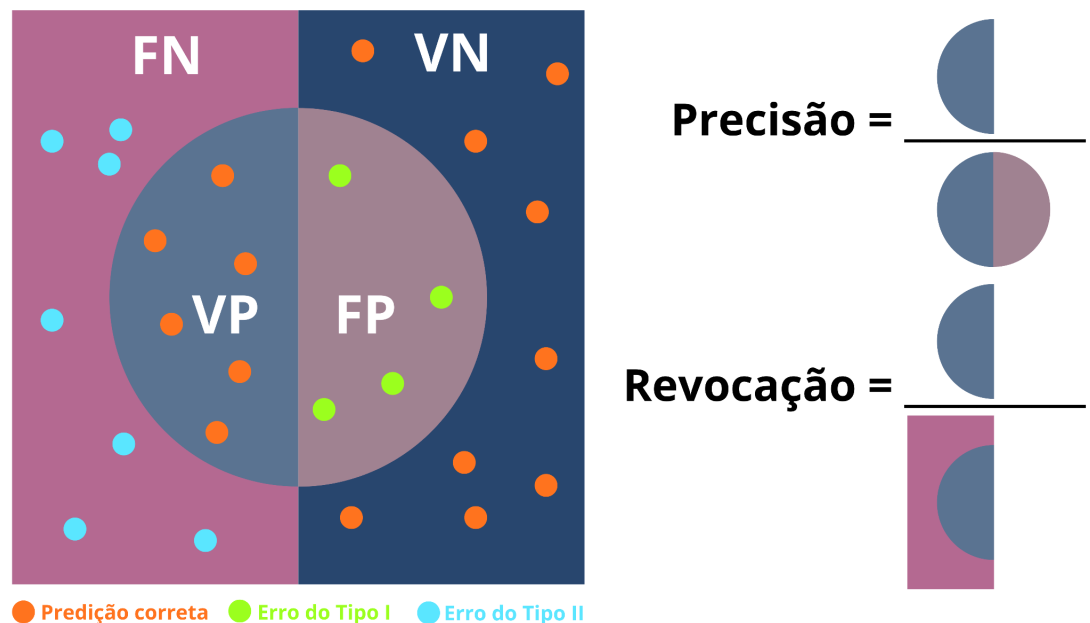


Figura 14: Ilustración del concepto de Precisión y Sensibilidad. La figura ilustra cómo los falsos positivos reducen la Precisión y los falsos negativos reducen la Sensibilidad, enfatizando que el énfasis en una métrica impacta la otra.

Fuente: Elaboración propia

El *F1-Score*, a su vez, integra las métricas de Precisión y Sensibilidad en una sola medida, permitiendo evaluar simultáneamente la confiabilidad de las predicciones positivas y la capacidad del modelo de identificar correctamente todos los casos positivos (Provost; Fawcett, 2016).

Esta métrica es calculada por medio de la media armónica entre Precisión y Sensibilidad, lo que penaliza valores extremos y resulta en una evaluación más equilibrada y representativa del desempeño del modelo, especialmente en escenarios con desbalance de clases.

$$F1 = 2 \times \frac{\text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} \quad (1)$$

3 Resultados

Tras el entrenamiento de los modelos con y sin data augmentation, los menores valores de la función de pérdida fueron alcanzados en la época 52 para el modelo sin aumento de datos y en la época 71 para el modelo con aumento de datos. En ambos casos, no hubo mejora en las 50 épocas siguientes, lo que activó el criterio de *early stopping*.

El análisis global de los resultados, considerando el promedio de las cuatro áreas de planificación, indica que el modelo entrenado con *data augmentation* presenta un desempeño superior en tres de las cuatro métricas evaluadas. Obtuvo mayores valores promedio de IoU, Precisión y *F1-Score*, evidenciando una mejora en la calidad general de la segmentación y en el equilibrio entre errores de falso positivo y falso negativo.

Por otro lado, el modelo sin *data augmentation* presentó un mayor valor promedio de Sensibilidad, indicando una mayor capacidad de identificar correctamente las regiones de interés, aunque a costa de una mayor incidencia de falsos positivos. Este comportamiento refuerza la existencia de un *trade-off* entre Precisión y Sensibilidad, en el cual el aumento de una métrica tiende a impactar negativamente a la otra.

De forma general, los resultados sugieren que la utilización de data augmentation contribuye a hacer el modelo más equilibrado y robusto, con ganancias consistentes en métricas agregadas como IoU y *F1-Score*, que son particularmente relevantes para la evaluación global del desempeño en tareas de segmentación (Tabla 3).

Tabla 3. Desempeño promedio de las métricas de evaluación de la U-Net entrenada sin y con data augmentation. Los valores superiores entre los dos modelos por Área de Planificación y General están destacados en azul.

AP	Modelo sin <i>data augmentation</i>				Modelo con <i>data augmentation</i>			
	IoU	Prec	Sens	F1	IoU	Prec	Sens	F1
1	0,7037	0,5101	0,7850	0,5887	0,7053	0,5328	0,7031	0,5828
2	0,7491	0,6953	0,7962	0,7287	0,7576	0,7419	0,7659	0,7344
3	0,6334	0,5188	0,7785	0,5628	0,6847	0,6034	0,7548	0,6256
4	0,7283	0,5188	0,7785	0,6791	0,7441	0,6902	0,7282	0,6955
General	0,6842	0,5189	0,7751	0,6238	0,7149	0,6444	0,7437	0,6592

Fuente: Elaboración propia

El análisis desagregado por Área de Planificación (AP) evidencia que los efectos del *data augmentation* varían conforme al contexto espacial, aunque se observa un patrón general de mejora en métricas agregadas. En las APs 2, 3 y 4, el modelo

con *data augmentation* presentó un desempeño superior o equivalente al modelo sin aumento de datos en las métricas de IoU, Precisión y *F1-Score*, indicando ganancias consistentes en la calidad de la segmentación y en el equilibrio entre errores de clasificación.

En la AP 3, en particular, la ventaja del modelo con *data augmentation* es más pronunciada, con aumentos expresivos en IoU, Precisión y *F1-Score*, además de una Sensibilidad elevada. Resultados similares, aunque menos acentuados, se observan en las APs 2 y 4, donde el modelo con *data augmentation* mantiene un desempeño globalmente superior.

La AP 1 presenta un comportamiento distinto. Aunque el modelo con *data augmentation* obtiene un IoU ligeramente superior (0,7053 contra 0,7037) y mayor Precisión (0,5328 contra 0,5101), el modelo sin *augmentation* presenta una Sensibilidad considerablemente más elevada (0,7850 contra 0,7031), resultando en un *F1-Score* marginalmente superior. Este resultado indica un *trade-off* más acentuado entre Precisión y Sensibilidad en esta área específica, posiblemente asociado a las características particulares de la región y a la limitada cantidad de muestras disponibles.

En síntesis, el análisis por AP confirma que el *data augmentation* tiende a producir modelos más equilibrados y con mejor desempeño global, especialmente en términos de IoU y *F1-Score*, aunque su impacto pueda variar localmente en función de la distribución y de la representatividad de los datos de entrenamiento.

En el diagrama de cajas que presenta la distribución de las métricas para las 57 imágenes del conjunto de prueba, estratificadas por Área de Planificación, se observan patrones importantes (Figura 15). El modelo entrenado con *data augmentation* demuestra menor dispersión en los valores de IoU y *F1-Score*, indicando un desempeño más consistente y predecible en estas métricas. En contrapartida, el modelo sin *data augmentation* presenta valores medianos superiores en la métrica de Sensibilidad. En cuanto a la Precisión, el modelo con *augmentation* muestra una distribución levemente desplazada hacia arriba, con primer, tercer cuartiles y mediana más elevados, sugiriendo un desempeño superior.

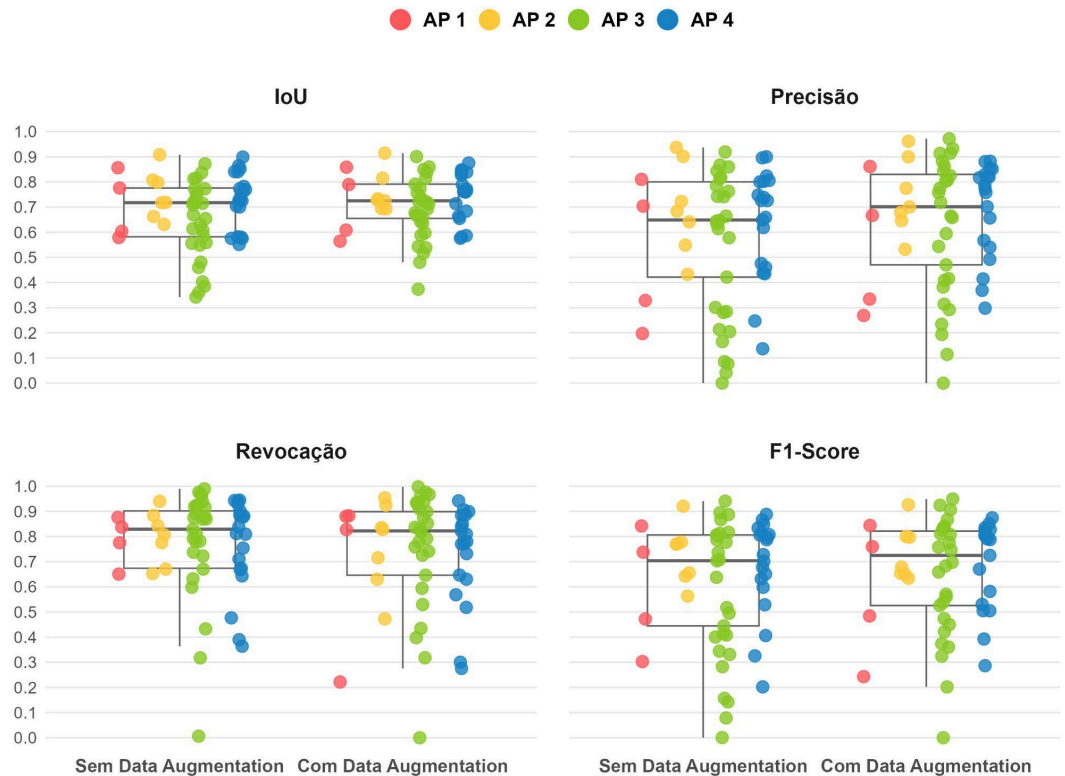


Figura 15: Distribución de las métricas de evaluación (IoU, Precisión, Sensibilidad y F1-Score) de la U-Net entrenada sin y con data augmentation, estratificada por Área de Planificación (AP). El análisis revela que el modelo con aumento de datos presenta menor dispersión en las métricas de IoU, Precisión y F1-Score, mientras el modelo sin augmentation muestra cuartiles superiores en Sensibilidad.

Fuente: Elaboración propia

Para evaluar objetivamente si la ganancia en desempeño observada era estadísticamente significativa, se condujeron pruebas de hipótesis comparando los valores de IoU y *F1-Score* obtenidos por el modelo con data augmentation contra los del modelo sin data augmentation. Como los datos de las 57 imágenes del conjunto de prueba representan mediciones pareadas (cada imagen evaluada por ambos modelos), se seleccionaron pruebas estadísticas adecuadas para muestras pareadas.

Inicialmente, la prueba de *Shapiro–Wilk* aplicada a las diferencias individuales de los valores de IoU para cada imagen rechazó la hipótesis de normalidad (valor $p = 0,0019$), violando así un presupuesto fundamental de la prueba *t* pareada. Ante esto, se optó por la prueba de *Wilcoxon* para muestras pareadas, un método no paramétrico robusto a desviaciones de la normalidad, que evalúa si las diferencias entre los pares presentan un desplazamiento sistemático en relación a cero.

La prueba fue configurada con hipótesis alternativa (H_1) unilateral de que los valores de IoU del modelo con data augmentation (M_2) tienden a ser mayores que los del modelo sin *data augmentation* (M_1), conforme se define a continuación:

- H_0 : Las distribuciones de los valores de IoU de M_1 y M_2 son iguales. Es decir, no hay evidencia de que los valores de IoU de M_2 sean sistemáticamente mayores que los de M_1 .
- H_1 : Los valores de IoU de M_2 son sistemáticamente mayores que los de M_1 .

Considerando un nivel de significancia (α) de 0,05, el valor p de 0,0006 permitió rechazar la hipótesis nula (H_0), proporcionando evidencia estadística de que los valores de IoU del modelo entrenado con *data augmentation* son superiores. La mediana de las diferencias fue positiva (+0,0183), corroborando esta conclusión. Adicionalmente, en el 71,9% de las imágenes (41 de 57), el modelo con data

augmentation superó al modelo sin augmentation, reforzando la consistencia práctica de la ventaja observada.

Para el *F1-Score*, se adoptó un procedimiento similar. La prueba de *Shapiro–Wilk* también rechazó la normalidad de las diferencias (valor $p = 0,0002$), justificando nuevamente el uso de la prueba no paramétrica de *Wilcoxon*. Se aplicó el mismo esquema de hipótesis:

- H_0 : Las distribuciones de los valores de *F1-Score* de M_1 y M_2 son iguales.
- H_1 : Los valores de *F1-Score* de M_2 son sistemáticamente mayores que los de M_1 .

El resultado, considerando también un nivel de significancia (α) de 0,05, con valor $p = 0,0060$, permitió rechazar la hipótesis nula. La superioridad del modelo con *data augmentation* también se manifestó en la diferencia de las medianas del *F1-Score* (+0,0122 puntos). El modelo M_2 superó a M_1 en el 66,67% de las imágenes (38 de 57), presentando un desempeño inferior en solo el 33,33% de los casos (19 de 57).

Con el objetivo de visualizar y comprender el comportamiento de las predicciones en la identificación de favelas en imágenes de satélite, se presentan a continuación resultados visuales referentes a dos muestras de cada una de las cuatro Áreas de Planificación (AP): un ejemplo de predicción exitosa y otro en que el desempeño fue inferior. Cada imagen es acompañada de sus respectivas métricas de IoU y *F1-Score*. Para facilitar el análisis comparativo, las áreas predichas como favela fueron sobrepuestas a las imágenes originales con aplicación de transparencia, permitiendo visualizar claramente las regiones urbanas clasificadas por el modelo U-Net y, al mismo tiempo, preservar la forma original para comparación directa con las máscaras de referencia.

En la imagen 08135, correspondiente al Área de Planificación 1, se observa en la esquina superior derecha un tramo de la favela Parque Horácio Cardoso Franco, ubicada en Benfica, en la Calle Couto de Magalhães (Figura 16). Ambos modelos presentaron un desempeño satisfactorio en la segmentación del área de interés; sin embargo, el modelo entrenado con *data augmentation* obtuvo resultados superiores, reflejados en mayores valores de IoU y *F1-Score*. Se observa que el modelo sin *data augmentation* clasificó erróneamente una pequeña región en la esquina superior derecha de la imagen, mientras que los errores del modelo con aumento de datos se mostraron más localizados y restringidos a las proximidades de la favela, indicando una segmentación más precisa y espacialmente coherente.

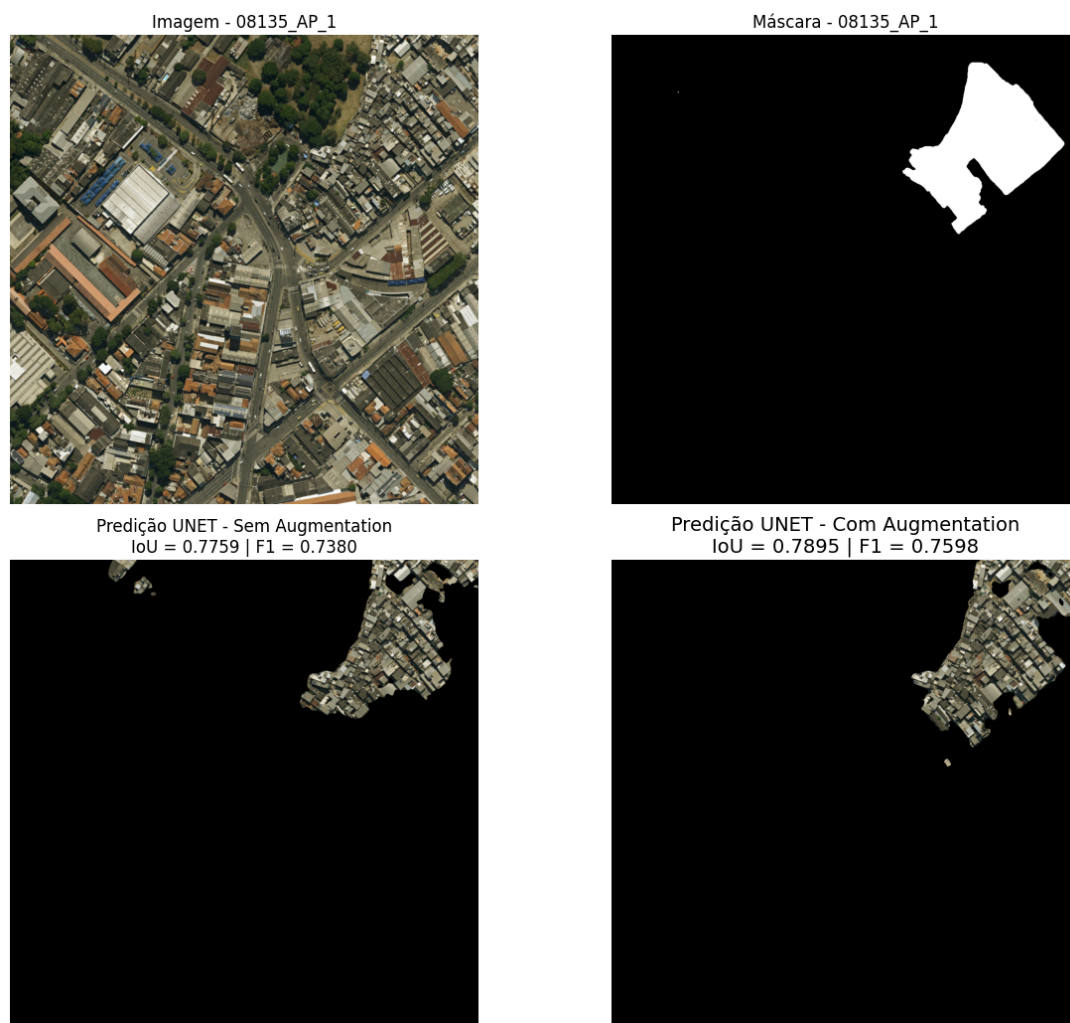


Figura 16: Comparación visual de las predicciones para la imagen 08135 (Área de Planificación 1). De izquierda a derecha: (a) imagen de satélite original; (b) máscara de referencia; (c) predicción del modelo sin data augmentation; (d) predicción del modelo con data augmentation. La escena incluye, en la esquina superior derecha, un segmento de la favela Parque Horácio Cardoso Franco, en Benfica (Calle Couto de Magalhães).

Fuente: Elaboración propia

La imagen 08430, correspondiente al Área de Planificación 1, contiene, en su porción inferior, una pequeña favela registrada exclusivamente en la base de datos del Instituto Pereira Passos, ubicada en la esquina de la Calle José Eugênio con la Calle Francisco Eugênio, en el barrio de San Cristóbal (Figura 17). En función de la reducida dimensión del área de interés, ambos modelos presentaron una baja superposición con la región de favela, aunque con valores de IoU superiores a 0,5. El modelo sin *data augmentation* obtuvo un desempeño superior tanto en IoU como en *F1-Score*; sin embargo, presentó un área significativamente mayor clasificada erróneamente como favela en comparación al modelo entrenado con aumento de datos, evidenciando un *trade-off* entre superposición global y control de falsos positivos.

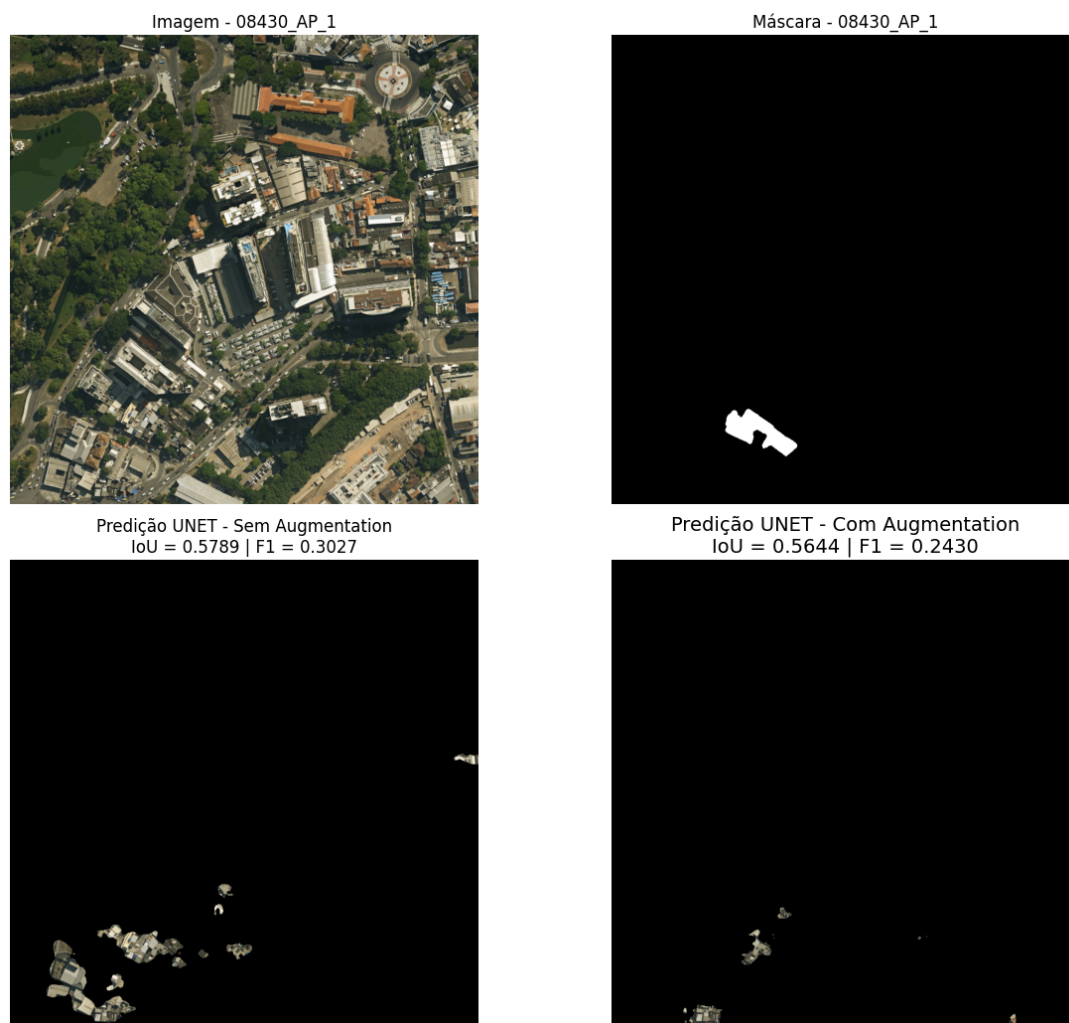


Figura 17: Comparación visual de las predicciones para la imagen 08430 (Área de Planificación 1). De izquierda a derecha: (a) imagen de satélite original; (b) máscara de referencia; (c) predicción del modelo sin data augmentation; (d) predicción del modelo con data augmentation. La escena contiene, en la porción inferior, una pequeña favela catalogada solo por el Instituto Pereira Passos, situada en la esquina de la Calle José Eugênio con la Calle Francisco Eugênio, en San Cristóbal.
Fuente: Elaboración propia

En la imagen 07848, referente al Área de Planificación 2, se observan tramos del Parque Vila Isabel, separados en la escena por el Parque Recanto del Trovador y por la Vila Olímpica Artur da Távola, en el barrio de Vila Isabel, en las proximidades de la Calle Visconde de Santa Isabel con la Calle Barão do Bom Retiro (Figura 18). Ambos modelos presentaron un desempeño elevado, con valores de IoU y *F1-Score* superiores a 0,9. Aun así, el modelo entrenado con *data augmentation* demostró una ligera superioridad, al delimitar con mayor precisión los contornos de la favela y discriminar adecuadamente elementos internos que no corresponden a domicilios precarios. Se nota que pequeños tramos en la porción inferior de las predicciones de ambos modelos fueron clasificados erróneamente como favela; tales regiones corresponden, en realidad, a áreas de terraza en edificaciones residenciales de pequeño porte.

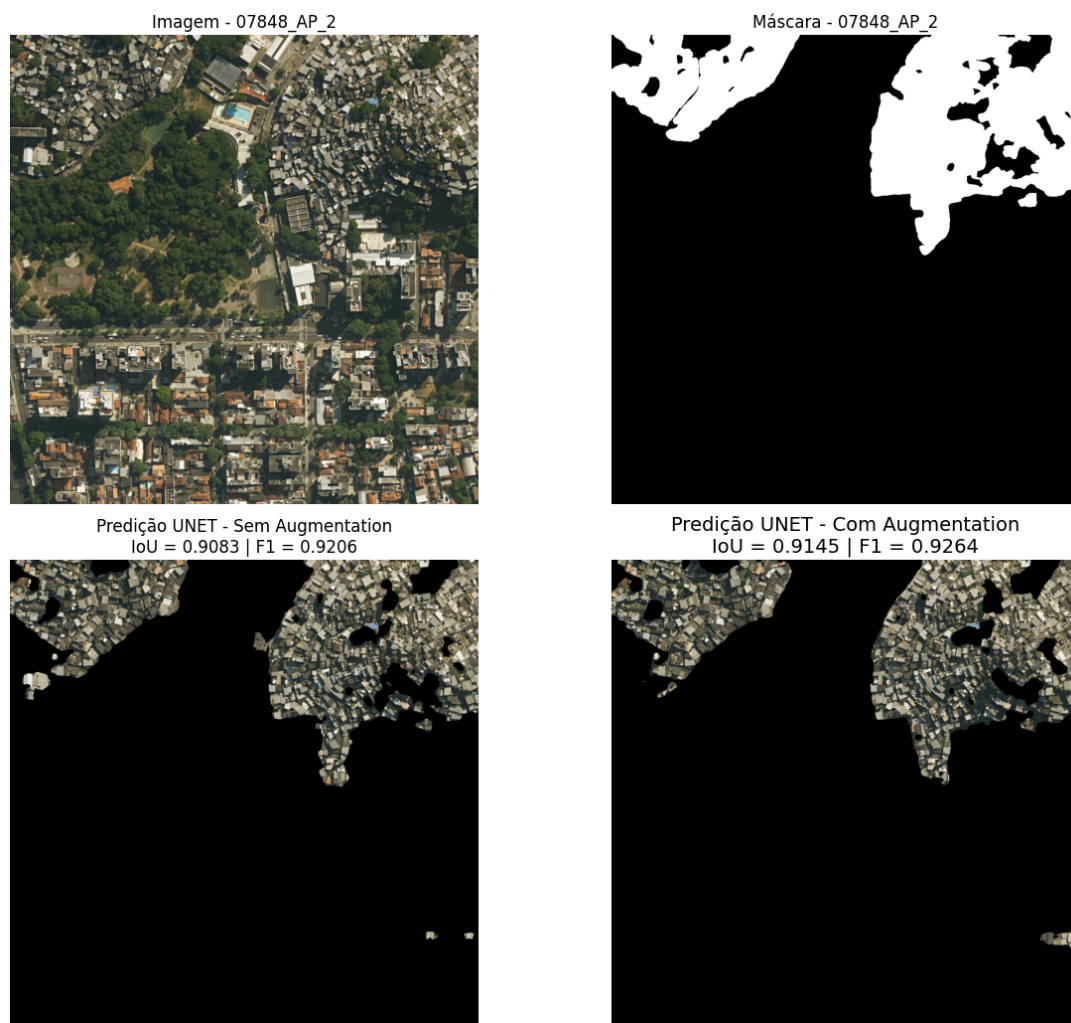


Figura 18: Comparación visual de las predicciones para la imagen 07848 (Área de Planificación 2). De izquierda a derecha: (a) imagen de satélite original; (b) máscara de referencia; (c) predicción del modelo sin data augmentation; (d) predicción del modelo con data augmentation. La escena exhibe tramos del Parque Vila Isabel, separados por el Parque Recanto del Trovador y la Vila Olímpica Artur da Távola, en el barrio de Vila Isabel, próximo al cruce de las calles Visconde de Santa Isabel y Barão do Bom Retiro.
Fuente: Elaboración propia

En la imagen 08157, referente al Área de Planificación 2, se observa la favela Vidigal, ubicada en el barrio homónimo, próximo a la Avenida Niemeyer (Figura 19). Se nota que ambos modelos clasificaron erróneamente una extensa área como favela, la cual no corresponde a la máscara de referencia ni está registrada como tal en las bases del IPP o del IBGE. Esta región presenta un patrón mixto de ocupación: aunque contiene construcciones que claramente no se enmarcan en la clasificación de favela, otras poseen características visuales ambiguas que podrían suscitar cuestionamientos. Adicionalmente, la morfología accidentada del terreno, con fuerte inclinación, contribuye a una cierta irregularidad en las formas de los lotes y de las edificaciones, lo que puede haber influenciado los errores de clasificación.

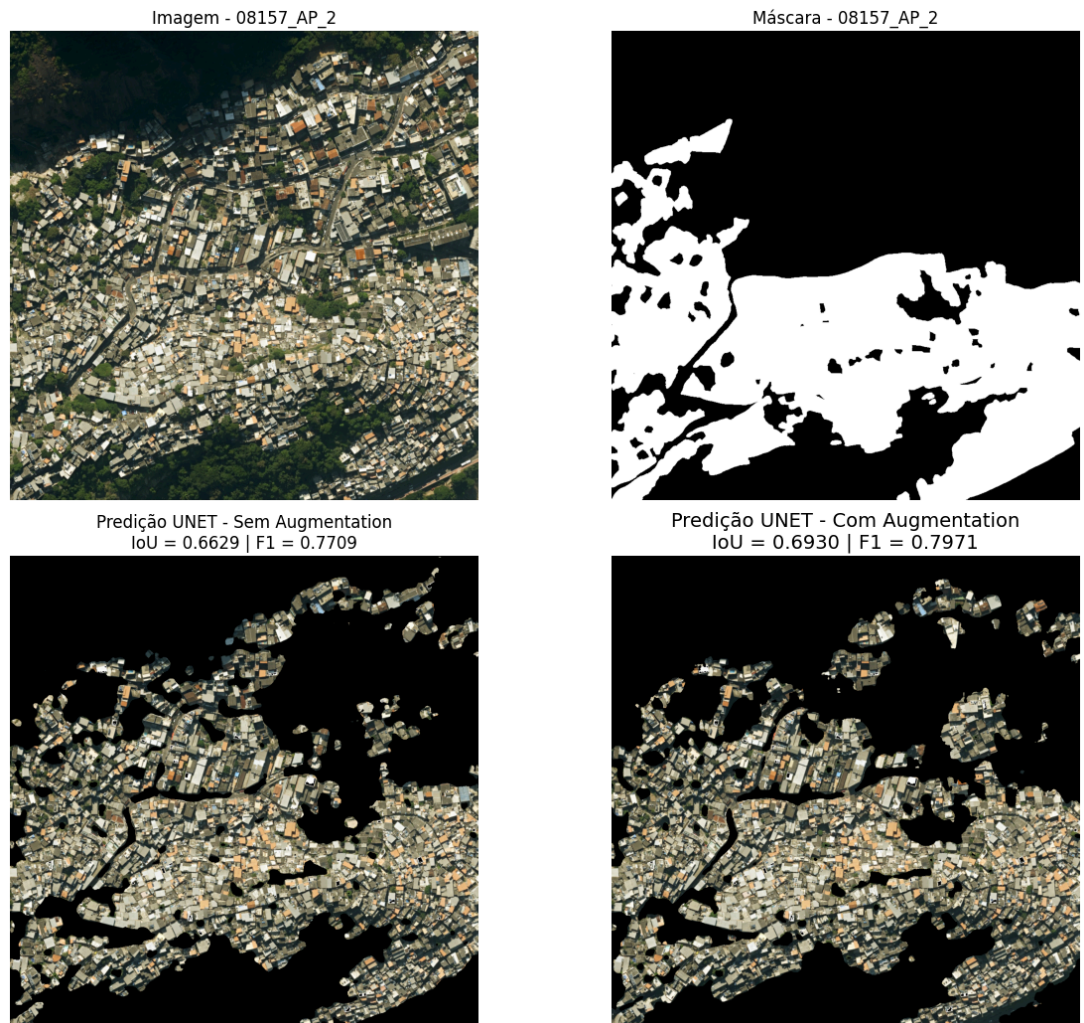


Figura 19: Comparación visual de las predicciones para la imagen 08157 (Área de Planificación 2). De izquierda a derecha: (a) imagen de satélite original; (b) máscara de referencia; (c) predicción del modelo sin data augmentation; (d) predicción del modelo con data augmentation. La escena muestra la favela Vidigal, en el barrio homónimo, en las proximidades de la Avenida Niemeyer.

Fuente: Elaboración propia

La imagen 07691 muestra la favela Nova Brasília, ubicada en el Complexo do Alemão en el Área de Planificación 3, en las proximidades del cruce de la Estrada de Itararé con la Avenida Itaóca (Figura 20). La predicción generada por el modelo con *data augmentation* presentó un desempeño superior, caracterizado por una menor ocurrencia de falsos positivos (áreas no favela erróneamente clasificadas) en comparación al modelo sin aumento de datos. Las áreas incorrectamente segmentadas como favela por ambos modelos corresponden, en su mayoría, a condominios residenciales.

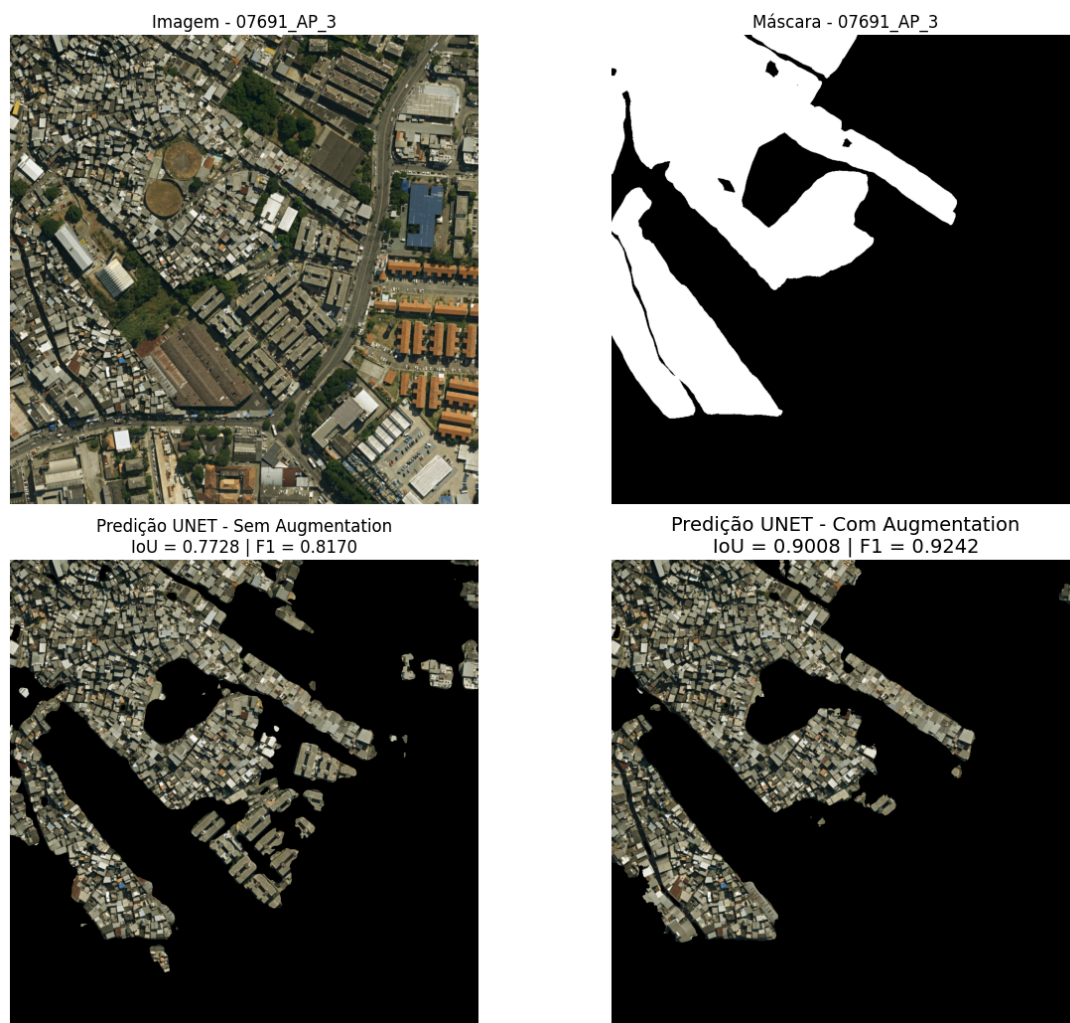


Figura 20: Comparación visual de las predicciones para la imagen 07691 (Área de Planificación 3). De izquierda a derecha: (a) imagen de satélite original; (b) máscara de referencia; (c) predicción del modelo sin data augmentation; (d) predicción del modelo con data augmentation. La escena exhibe la favela Nova Brasília, en el Complexo do Alemão, próxima al encuentro de la Estrada de Itararé con la Avenida Itaóca.

Fuente: Elaboración propia

La imagen 07170, del Área de Planificación 3, exhibe la pequeña favela Castelo de Lucas, registrada solo en la base de datos del IPP y ubicada en el barrio Parada de Lucas, en el encuentro de la Calle Cordovil con la Calle Amadeu Amaral (Figura 21). En un área reducida como esta, ambos modelos tuvieron una baja superposición con IoU inferior a 0,5. El modelo sin *data augmentation* aún consiguió una superposición con el área de favela evidenciada por un *F1-Score* de 0,0006, sin embargo clasificó erróneamente como favela diversas áreas. Ya el modelo con data augmentation clasificó erróneamente un área menor como favela, pero tuvo un *F1-Score* de cero.

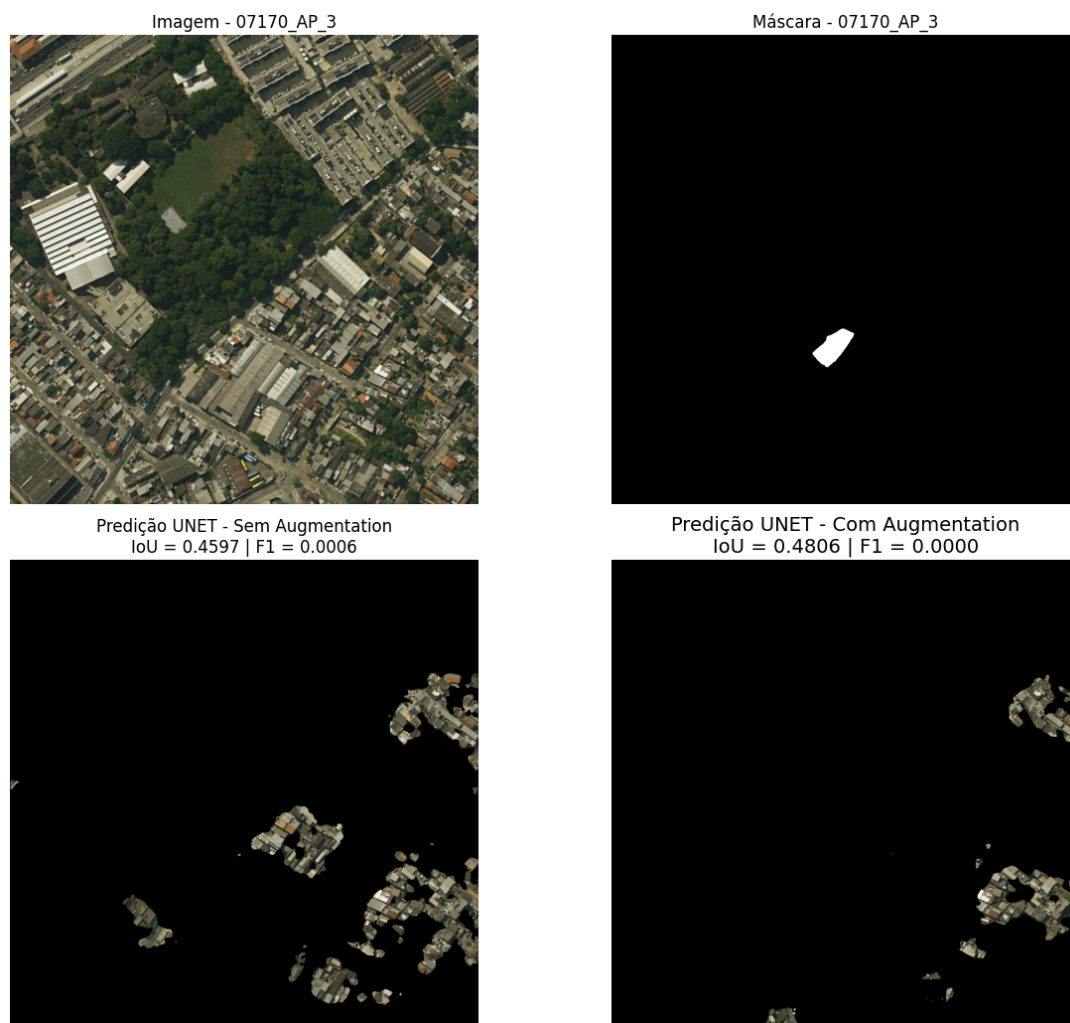


Figura 21: Comparación visual de las predicciones para la imagen 07170 (Área de Planificación 3). De izquierda a derecha: (a) imagen de satélite original; (b) máscara de referencia; (c) predicción del modelo sin data augmentation; (d) predicción del modelo con data augmentation. La escena muestra la pequeña favela Castelo de Lucas, registrada exclusivamente por el IPP, situada en el barrio Parada de Lucas, en la confluencia de las calles Cordovil y Amadeu Amaral.

Fuente: Elaboración propia

La imagen 05804, del Área de Planificación 4, muestra la favela Comunidad Corumau, ubicada en el barrio de la Taquara, frente a la Estrada Curumau (Figura 22). Ambos modelos alcanzaron valores elevados de IoU y *F1-Score*; sin embargo, el modelo entrenado con *data augmentation* presentó un desempeño superior. Desde el punto de vista cualitativo, se observa que este modelo delimitó con mayor fidelidad los contornos de la favela y generó una cantidad menor de falsos positivos — es decir, áreas no pertenecientes a la favela clasificadas incorrectamente — cuando se comparó con el modelo sin aumento de datos. Las principales regiones erróneamente segmentadas corresponden a terrazas de casas adosadas y de edificaciones residenciales de pequeño porte, cuyas características espectrales y de textura pueden asemejarse a las de áreas de favela en imágenes de satélite, dificultando la distinción por ambos modelos.

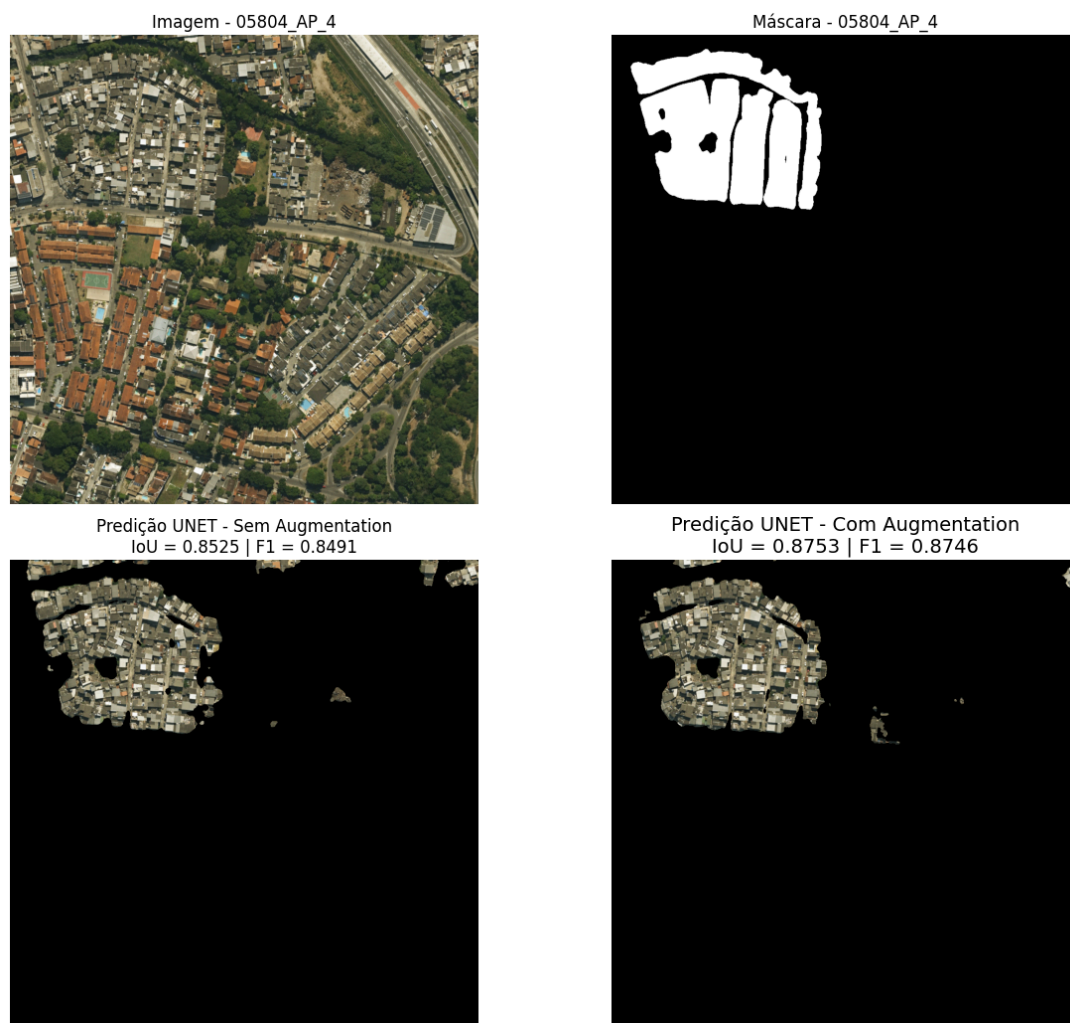


Figura 22: Comparación visual de las predicciones para la imagen 05804 (Área de Planificación 4). De izquierda a derecha: (a) imagen de satélite original; (b) máscara de referencia; (c) predicción del modelo sin data augmentation; (d) predicción del modelo con data augmentation. La escena exhibe la favela Comunidad Corumau, en la Taquara, frente a la Estrada Curumau.

Fuente: Elaboración propia

La imagen 05729, del Área de Planificación 4, muestra un pequeño aglomerado residencial clasificado como favela solo en la base del IPP, situado frente a la Estrada da Boiúna, en el barrio de la Taquara, en la esquina inferior izquierda de la escena (Figura 23). Ambos modelos consiguieron identificar parcialmente este conjunto de edificaciones, atribuyéndole correctamente la clase de favela. Sin embargo, también se observaron falsos positivos en otras regiones que no constan como favela ni en la máscara de referencia ni en las bases oficiales. El área segmentada en la esquina superior izquierda de la imagen, en las proximidades de la Calle Pereira, presenta características morfológicas y espectrales ambiguas, lo que puede justificar la divergencia de clasificación y suscitar cuestionamientos en cuanto a su real categorización. Por otro lado, las edificaciones identificadas en la esquina inferior derecha solo por el modelo sin data augmentation corresponden, en realidad, a un pequeño condominio situado en la Estrada Curumau, próximo al Túnel Senador Nelson Carneiro, configurando un error inequívoco de clasificación.

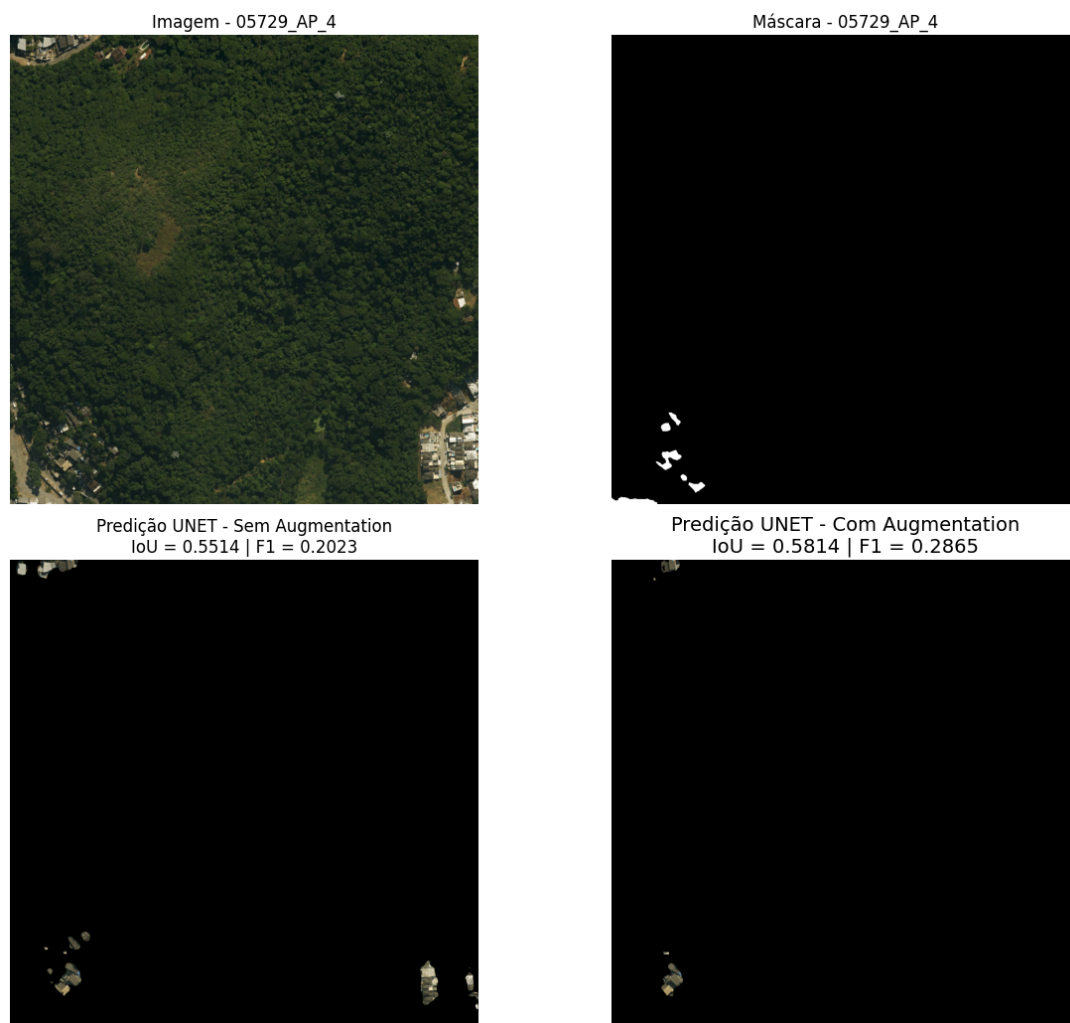


Figura 23: Comparación visual de las predicciones para la imagen 05729 (Área de Planificación 4). De izquierda a derecha: (a) imagen de satélite original; (b) máscara de referencia; (c) predicción del modelo sin data augmentation; (d) predicción del modelo con data augmentation. La escena exhibe, en la esquina inferior izquierda, un pequeño aglomerado residencial catalogado como favela solo por el IPP, frente a la Estrada da Boiúna, en la Taquara.

Fuente: Elaboración propia

4 Consideraciones Finales

Los resultados obtenidos confirman el desempeño prometedor de la arquitectura U-Net para la identificación de favelas en imágenes de satélite. De modo general, el modelo entrenado con *data augmentation* presentó un desempeño superior en las métricas de IoU, *F1-Score* y Precisión, indicando mayor calidad en la superposición de las áreas segmentadas, mejor equilibrio entre errores de clasificación y mayor control sobre la ocurrencia de falsos positivos. En contrapartida, este modelo registró valores ligeramente inferiores de Sensibilidad en comparación al modelo entrenado sin aumento de datos, evidenciando un comportamiento más conservador en la identificación de las áreas de interés.

Este resultado refleja el *trade-off* clásico entre Precisión y Sensibilidad, en el cual el *data augmentation* contribuye a hacer el modelo más selectivo, reduciendo clasificaciones incorrectas (falsos positivos), aunque a costa de la pérdida de algunas detecciones verdaderas. Tal característica puede interpretarse como deseable en aplicaciones en las cuales la reducción de falsos positivos es prioritaria, como en análisis urbanos y planificación territorial. Además, los análisis cuantitativos y cualitativos indican que el uso de *data augmentation* favorece una segmentación más estable y espacialmente coherente, especialmente en áreas con mayor heterogeneidad morfológica, reforzando su relevancia como estrategia para

aumentar la capacidad de generalización del modelo.

A pesar de los avances observados, persisten desafíos relevantes. El modelo entrenado con *data augmentation*, aunque presenta mayor Precisión, demuestra una tendencia más conservadora en la identificación de áreas de favela, reduciendo la ocurrencia de falsos positivos, pero, en algunos casos, dejando de detectar áreas legítimas, especialmente aquellas de pequeña extensión o insertadas en contextos visuales complejos. Por otro lado, el modelo entrenado sin aumento de datos, al presentar mayor Sensibilidad, se mostró más sensible a la detección de estas áreas, aunque con mayor propensión a clasificar erróneamente regiones que no corresponden a este tipo de ocupación.

En ambos casos, los errores recurrentes estuvieron asociados a áreas que presentan características visuales semejantes a las de favelas, como terrazas residenciales, condominios de pequeño porte y zonas urbanas degradadas. Además, la identificación de favelas de pequeña extensión, fragmentadas por la cuadrícula de recorte o insertadas en ambientes de bajo contraste visual, se mostró particularmente desafiante, evidenciando limitaciones en la distinción de estas áreas en función de la ambigüedad de sus características espectrales y espaciales.

Para superar estas limitaciones y perfeccionar el modelo, se sugieren las siguientes direcciones para trabajos futuros:

1. Incorporación de bandas espectrales adicionales, como Infrarrojo Cercano (NIR) e Infrarrojo de Onda Corta (SWIR), que pueden capturar información sobre vigor vegetativo, humedad y materiales constructivos, auxiliando en la distinción entre favelas y otras tipologías urbanas.
2. Refinamiento del esquema de clases, con la subdivisión de la amplia categoría "no favela" en clases semánticamente distintas. La creación de categorías específicas para tipologías urbanas recurrentemente confundidas — como conjuntos residenciales formales, edificaciones comerciales, galpones industriales y cubiertas de terraza — permitiría al modelo aprender representaciones más discriminatorias. Esta mayor granularidad en la definición del objetivo de aprendizaje tiene potencial para reducir significativamente los falsos positivos, proporcionando una comprensión más estratificada y contextual del tejido urbano.
3. Evaluación de arquitecturas de red neuronal más avanzadas, como HRNet, transformers para visión computacional (e.g., SegFormer, Mask2Former) y métodos de Aprendizaje Autosupervisado (SSL), que pueden mejorar la capacidad de representación de características espaciales y de textura, especialmente en escenarios con datos anotados limitados.
4. Exploración de estrategias de posprocesamiento para refinamiento de máscaras, como técnicas de suavizado de bordes, fusión de segmentaciones y métodos basados en geometría, con el objetivo de corregir incoherencias espaciales y reducir ruidos en las predicciones.
5. Integración de datos contextuales y socioeconómicos a nivel de sector censal, que pueden ofrecer información adicional para la diferenciación de áreas urbanas de bajos ingresos con morfologías semejantes.
6. Revisión colaborativa del proceso de anotación, involucrando a múltiples especialistas en la generación de las máscaras de referencia. Como la interpretación de características visuales que definen una favela no siempre es consensual, la consolidación de las anotaciones por medio de operaciones de unión o intersección puede producir máscaras de referencia más consistentes y robustas. Este enfoque permite capturar tanto una definición más abarcadora (unión) como una más conservadora (intersección) del fenómeno, reduciendo sesgos subjetivos individuales y creando bases de entrenamiento que mejoran la capacidad de generalización del modelo para diferentes contextos interpretativos.

7. Ampliación y diversificación de las técnicas de data augmentation, utilizando bibliotecas como Albumentations para introducir transformaciones más complejas y estocásticas, aumentando la variabilidad del conjunto de entrenamiento y la robustez del modelo a diferentes condiciones de imagen.

La implementación de estas propuestas tiene potencial para elevar significativamente la precisión, la confiabilidad y la aplicabilidad de modelos de segmentación semántica en el mapeo de asentamientos precarios, contribuyendo a políticas urbanas más asertivas y basadas en evidencia.

Referencias

- ABASCAL, A. *et al.* Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. **Journal of Environmental Management**, v. 95, 2022. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0198971522000643>. Acceso en: 12 dic. 2023.
- ALRASHEEDI, K.; DEWAN, A.; EL-MOWAFY, A. Using local knowledge and remote sensing in the identification of informal settlements in riyadh city, Saudi Arabia. **Remote Sensing**, v. 15, n. 15, p. 3895, 2021. Disponible en: <https://www.mdpi.com/2072-4292/15/15/3895>. Acceso en: 25 sep. 2023.
- BUSLAEV, A. *et al.* Albumentations: fast and flexible image augmentations. **arXiv preprint arXiv: 1809.06839**. 2018. Disponible en: <https://arxiv.org/pdf/1809.06839>. Acceso en: 16 ene. 2024.
- CINNAMON, J.; NOTH, T. Spatiotemporal development of informal settlements in cape town, 2000 to 2020: An open data approach. **Habitat International**, v. 112, 2023. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0197397523000139>. Acceso en: 11 abr. 2024.
- DATA.RIO. Limite Favelas 2019. 2019. Disponible en: <https://www.data.rio/datasets/limite-favelas-2019/explore>. Acceso en: 20 dic. 2023.
- DATA.RIO. Limite Áreas de Planejamento (AP). 2023. Disponible en: https://www.data.rio/datasets/b9e30861acfe4bea947e6278a6b30ce3_1/explore. Acceso en: 20 dic. 2023.
- DONG, Shiwei. Spatial Stratification Method for the Sampling Design of LULC Classification Accuracy Assessment: A Case Study in Beijing, China. **Remote Sensing**, v. 14, n. 4, 2022. Disponible en: <https://www.mdpi.com/2072-4292/14/4/865>. Acceso en: 03 feb. 2024.
- GHAFFARIAN, S.; EMTEHANI, S. Monitoring urban deprived areas with remote sensing and machine learning in case of disaster recovery. **Urban Science**, v. 9, n. 4, 2021. Disponible en: <https://www.mdpi.com/2225-1154/9/4/58>. Acceso en: 12 dic. 2023.
- IBGE. Aglomerados Subnormais 2019: Classificação preliminar e informações de saúde para o enfrentamento à COVID-19. 2020. Disponible en: https://biblioteca.ibge.gov.br/visualizacao/livros/liv101717_notas_tecnicas.pdf. Acceso en: 06 dic. 2023.
- IBGE. Favelas e Comunidades Urbanas: IBGE muda denominação dos aglomerados subnormais. 2024. Disponible en: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/38962-favelas-e-comunidades-urbanas-ibge-muda-denominacao-dos-aglomerados-subnormais>. Acceso en: 05 nov. 2025.
- KEMPER, T. *et al.* Towards an automated monitoring of human settlements in south africa using high resolution spot satellite imagery. **International Society for Photogrammetry and Remote Sensing**. v. XL-7/W3, [S.n.], 2015. Disponible en: <https://core.ac.uk/download/pdf/38631179.pdf>. Acceso en: 4 mar. 2024

LAKSHMANAN, Valliappa; GÖRNER, Martin; GILLARD, Ryan. **Practical machine learning for computer vision**. " O'Reilly Media, Inc.", 2021. O'Reilly Media.2021

LING, Antony. Cortiços eram melhores que as favelas. **Caos Planejado**, 2018. Disponible en: <https://caosplanejado.com/corticoseeram-melhores-que-favelas/>. Acceso en: 23 dic. 2023.

LU, W. et al. A Geoscience-Aware Network (GASlumNet) Combining UNet and ConvNeXt for Slum Mapping. **Remote Sensing**, v. 16, n. 2, p. 260, 2021. Disponible en: <https://www.mdpi.com/2072-4292/16/2/260>. Acceso en: 14 may. 2024.

MAIYA, S. R.; BABU, S. C. Slum segmentation and change detection: A deep learning approach. **arXiv preprint arXiv**: 1811.07896. 2018. Disponible en: <https://arxiv.org/pdf/1811.07896>. Acceso en: 13 mar. 2024.

MARINS, Paulo. **História da Vida Privada no Brasil**: Habitação e Vizinhaça: Limites da Privacidade no Surgimento das Metrôpoles Brasileiras. 3. ed. São Paulo: Companhia Das Letras, 1998. P. 86.

OLIVEIRA, L. T. et al. Capturing deprived areas using unsupervised machine learning and open data: a case study in São Paulo, Brazil. **European Journal of Remote Sensing**, 2023. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/22797254.2023.2214690>. Acceso en: 5 mar. 2024.

ONU BRASIL. **ONU-Habitat**: população mundial será 68% urbana até 2050. NAÇÕES UNIDAS BRASIL. 2022. Disponible en: <https://brasil.un.org/pt-br/188520-onu-habitat-popula%C3%A7%C3%A3o-mundial-ser%C3%A1-68-urbana-at%C3%A9-2050>. Acceso en: 17 sep. 2023.

ONU BRASIL. **Os objetivos de desenvolvimento sustentável no Brasil**. NAÇÕES UNIDAS BRASIL. 2023. Disponible en: <https://brasil.un.org/pt-br/sdgs/11>. Acceso en: 17 sep. 2023.

PREFEITURA DA CIDADE DO RIO DE JANEIRO. **Conselho Estratégico de Informações da Cidade**: Atas de Reuniões. 2012. Disponible en: <https://www.rio.rj.gov.br/documents/91329/1f8a19d9-91d6-430d-81f4-52081055114e>. Acceso en: 13 nov. 2025.

PREFEITURA DA CIDADE DO RIO DE JANEIRO. **Painel Rio**: Conheça mais o rio de hoje para construir o rio de amanhã. 2025. Disponible en: <https://pds-pcrj.hub.arcgis.com/pages/unidades>. Acceso en: 12 nov. 2025.

PROVOST, Foster; FAWCETT, Tom. **Data Science para Negócios**: O que você precisa saber sobre mineração de dados e pensamento analítico de dados. Rio de Janeiro: Alta Books, 2016.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation. **arXiv preprint arXiv**: 1505.04597, 2015. Disponible en: <https://arxiv.org/pdf/1505.04597>. Acceso en: 16 ene. 2024.

SILVA, Pedro; BIANCHINI, Zélia; DIAS, Antonio. **Amostragem: Teoria e Prática Usando R**. Rio de Janeiro: [S.n.], 2023. Disponible en: <https://amostragemcomr.github.io/livro/index.html>. Acceso en: 2 feb. 2026.

USGS (United States Geological Survey). **The Universal Transverse Mercator (UTM) Grid**. [S.l.: s.n.], 2001. P. 2. Documento online. Disponible en: <https://pubs.usgs.gov/fs/2001/0077/report.pdf>. Acceso en: 10 ene. 2024.

WURM, M. et al. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 150, 2019. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0924271619300383>. Acceso en: 11 dic. 2023

Sobre los Autores

Jedielso Sales de Souza es graduado en Estadística por la Escuela Nacional de Ciencias Estadísticas (ENCE). Actúa en el Instituto Brasileño de Geografía y Estadística (IBGE) como Jefe de la Agencia Centro de la Superintendencia Estatal de Río de Janeiro, siendo responsable de la gestión y capacitación de equipos involucrados en la recolección de diversas investigaciones estadísticas. En el Censo Demográfico de 2022, ejerció la función de coordinador de área en la región central de la ciudad de Río de Janeiro, donde adquirió experiencia en el mapeo de favelas y comunidades urbanas. Su actuación integra intereses en estadística, análisis de datos espaciales y estudios sobre favelas.

Andrea Diniz da Silva es profesora e investigadora en los programas de grado y de posgrado de la Escuela Nacional de Ciencias Estadísticas - ENCE/IBGE; líder del grupo de investigaciones "Big Data y Estadísticas Públicas"; investigadora de los grupos de investigaciones "DHPJS - Derechos Humanos, Poder Judicial y Sociedad" y "CAST - Computational Agriculture Statistics Laboratory". Bachiller en estadística por la Universidad del Estado de Río de Janeiro - UERJ, magíster en Estudios de Población e Investigaciones Sociales y doctora en Población, Territorio y Estadísticas Públicas, ambos por la ENCE. Realizó un doctorado sandwich en la University of Wollongong, en Australia. Trabaja con métodos para investigaciones y uso de big data para la producción de estadísticas públicas. Ocupa el cargo de jefa de relaciones internacionales y de coordinadora de GT sobre política de innovación, en el IBGE.

Ian Monteiro Nunes es Ingeniero de Computación con maestría y doctorado en Ciencia de Datos por el Departamento de Informática de la PUC-Río. Actualmente, es el Jefe de la Gerencia de Inteligencia en Datos e Innovación del IBGE, área pionera volcada a I+D y modernización del Censo Agropecuario, y profesor de Ciencia de Datos en la Escuela Nacional de Ciencias Estadísticas (ENCE). Su actuación incluye la coordinación del proyecto de I+D "Representation Learning for Crop Segmentation in Brazil" en el LNCC y la participación en el grupo de investigación Big Data y Estadísticas Oficiales. Con sólida experiencia en la frontera entre la academia y la gestión pública, Ian lidera iniciativas de transformación digital utilizando Visión Computacional, Teledetección y Aprendizaje Automático. Sus áreas de investigación abarcan segmentación de imágenes, deep learning, adaptación de dominio y open-set recognition.

Contribuciones de los Autores

Conceptualización, [J.S.S.]; metodología, [J.S.S., A.D.S., I.M.N.]; software, [J.S.S.]; validación, [J.S.S.]; análisis formal, [J.S.S.]; investigación, [J.S.S.]; recursos, [J.S.S.]; curaduría de datos, [J.S.S.]; redacción—preparación del borrador original, [J.S.S.]; redacción—revisión y edición, [J.S.S., A.D.S., I.M.N.]; visualización, [J.S.S.]; supervisión, [A.D.S., I.M.N.]. Todos los autores leyeron y aprobaron la versión publicada del manuscrito.

Disponibilidad de Datos

Los datos de esta investigación están disponibles en: <https://drive.google.com/drive/folders/10j4KrWzsuuQuK8yODZRRhCscWvNig17J?usp=sharing>

Conflictos de Interés

Los autores declaran no tener conflictos de interés.

Sobre la *Coleção Estudos Cariocas*

La *Coleção Estudos Cariocas* (ISSN 1984-7203) es una publicación dedicada a estudios e investigaciones sobre el Municipio de Río de Janeiro, vinculada al Instituto Pereira Passos (IPP) de la Secretaría Municipal de la Casa Civil de la Alcaldía de Río de Janeiro.

Su objetivo es divulgar la producción técnico-científica sobre temas relacionados con la ciudad de Río de Janeiro, incluyendo sus conexiones metropolitanas y su inserción en contextos regionales, nacionales e internacionales. La publicación está abierta a todos los investigadores (sean empleados municipales o no), abarcando áreas diversas — siempre que aborden, parcial o totalmente, el enfoque espacial de la ciudad de Río de Janeiro.

Los artículos también deben alinearse con los objetivos del Instituto, a saber:

1. promover y coordinar la intervención pública en el espacio urbano del Municipio;
2. proveer e integrar las actividades del sistema de información geográfica, cartográfica, monográfica y de datos estadísticos de la Ciudad;
3. apoyar el establecimiento de las directrices básicas para el desarrollo socioeconómico del Municipio.

Se dará especial énfasis a la articulación de los artículos con la propuesta de desarrollo económico de la ciudad. De este modo, se espera que los artículos multidisciplinarios enviados a la revista respondan a las necesidades de desarrollo urbano de Río de Janeiro.