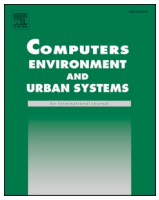
Computers, Environment and Urban Systems 121 (2025) 102306

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus



On the power of CNNs to detect slums in Brazil

João P. da Silva [a,*], José F. Rodrigues-Jr [a], João P. de Albuquerque [b,c]

[a] Institute of Mathematics and Computer Sciences, University of São Paulo, Avenida Trabalhador São-carlense, 400, São Carlos, 13566-590 São Paulo, Brazil
[b] School of Social and Political Sciences, University of Glasgow, 7 Lilibank Gardens, Glasgow G12-8RZ, UK
[c] School of Business Administration of São Paulo, Fundação Getulio Vargas, FGV-EAESP, São Paulo, Brazil.

ARTICLE INFO

Keywords:
Deep learning
Satellite imagery
VHR image
Slums
Informal settlements

ABSTRACT

The rapid expansion of slums poses a critical challenge for urban planning in Low- and Middle-Income Countries (LMICs), where traditional data collection methods like censuses are often outdated and insufficient. This study examines the transferability and generalization capabilities of deep learning models, specifically Convolutional Neural Networks (CNNs), for automated slum detection across six Brazilian cities with varying urban morphologies: São Paulo, Rio de Janeiro, Belo Horizonte, Brasília, Salvador, and Porto Alegre. Utilizing Very High Resolution (VHR) and High Resolution (HR) satellite imagery, we trained and evaluated models based on the EfficientNetV2L architecture. Our experimental results show that CNN models trained on data from a single city achieved high accuracy within that city (F1 scores exceeding 0.90 with VHR imagery), but their performance significantly decreased when applied to other cities (F1 scores dropping below 0.80), highlighting the impact of regional variations in urban morphology. Conversely, a generalized model trained on combined data from all six cities maintained robust performance across all cities, achieving F1 scores above 0.80 with VHR imagery. These findings indicate that while CNNs are effective for automated slum mapping, regional diversity necessitates training on diverse datasets to ensure generalization. We provide a comprehensive methodology over an openly shared dataset, and code to facilitate future research and applications in urban geoscience. The aim is to enhance the scalability and generalization of remote sensing and deep learning methods for slum identification across diverse urban environments.

1. Introduction

Inadequate housing poses a significant challenge, particularly in Low- and Middle-Income Countries (LMICs). The proliferation of slums is a clear symptom of this issue, serving as one of the most visible outcomes of unregulated urbanization. The term “slum” can vary in definition depending on the organization or country. In 2002, UN-Habitat convened international experts to establish objective criteria for defining slums. These criteria include: (i) inadequate access to safe water; (ii) inadequate sanitation and infrastructure; (iii) poor structural housing quality; (iv) overcrowding; and (v) insecure residential status UN-Habitat (2003).

In Brazil, slums—commonly referred to as favelas—align with UN-Habitat’s criteria but also encompass aspects such as illegal land occupation, noncompliance with urban regulations, and lack of essential public services IBGE (2024). The global significance of slums is underscored by the United Nations’ Sustainable Development Goals (SDGs), particularly Goal 11, which aims to make cities inclusive, safe, resilient, and sustainable UN-Stats (2023). A key target of this goal is to ensure universal access to adequate, safe, and affordable housing and basic services by 2030. In LMICs, where urban populations are rapidly expanding, the growth of slums poses a significant challenge to achieving this goal.

Globally, over 1 billion people are estimated to live in slums, with projections suggesting this number could rise if current urbanization trends persist UN-Habitat (2023). In Brazil, a country with one of the highest urbanization rates, slums house a significant share of the population. According to the 2010 census, more than 11 million people lived in 6329 slums across the country IBGE (2010). Preliminary data from the 2022 census indicate the existence of 11,421 subnormal agglomerations IBGE (2019).

Slum mapping is crucial for providing accurate, up-to-date information, enabling policymakers to target interventions and allocate resources effectively. By identifying areas in need, slum mapping can significantly improve living conditions in these communities Abascal, Rothwell, et al. (2022). In Brazil, the characteristics of slums vary widely

* Corresponding author.
E-mail addresses: jp.silva@usp.br (J.P. da Silva), junio@icmc.usp.br (J.F. Rodrigues-Jr), joao.porto@glasgow.ac.uk (J.P. de Albuquerque).

https://doi.org/10.1016/j.compenvurbsys.2025.102306
Received 5 December 2024; Received in revised form 2 March 2025; Accepted 4 May 2025
Available online 31 May 2025

depending on factors such as topography, environmental risks, urban layout, and population density. These variations present challenges to the mapping and monitoring process. Traditional methods like censuses collect valuable demographic and socioeconomic data but are often costly and logistically demanding, particularly in a vast country like Brazil, with over 200 million inhabitants Congresso Nacional (2021); IBGE (2023). The decennial national census, last conducted in 2010, faced significant delays in 2020 due to the COVID-19 pandemic IBGE (2021). Such gaps can hinder timely responses to the needs of vulnerable populations.

Recent advancements in Earth Observation (EO), artificial intelligence (AI), and Geographic Information Systems (GIS) have revolutionized slum mapping and monitoring. The declining cost of High-Resolution (HR) and Very High-Resolution (VHR) satellite imagery, combined with the availability of free platforms like Google Earth, has democratized access to large-scale data Mahabir et al. (2018). Integrating HR/VHR imagery with advanced AI algorithms enables the detection of informal settlements and the monitoring of dynamic changes within these areas Kuffer et al. (2016); Mahabir et al. (2018); Raj et al. (2024); Neupane et al. (2024).

This study aims to assess the transferability and generalization capabilities of deep learning models, specifically Convolutional Neural Networks (CNNs), for slum mapping across different Brazilian cities. Additionally, it compares the effectiveness of Very High Resolution (VHR) and High Resolution (HR) imagery in this context. Fig. 1 illustrates the main constituents our methodology. The research addresses the following questions:

- Can deep learning models accurately identify slums in Brazilian cities?
- How well do these models generalize across the vast variability of urban agglomerates?
- What are the best practices for obtaining and utilizing VHR and HR imagery for slum identification?

2. Related work

Over the past few decades, numerous studies have highlighted the potential of Remote Sensing (RS) for urban analyses. Early approaches often relied on single metrics, such as income, to map deprived areas. For instance, nighttime satellite imagery was employed to correlate poverty rates, economic activity, population density, and electricity consumption in urban environments Elvidge et al. (1997, 2007). However, this approach proved limited in deprived areas, as these regions typically exhibit low and uniform levels of nighttime light, making it difficult to differentiate between varying economic activities Jean et al. (2016).

The increasing availability of high-resolution (HR) satellite imagery and large datasets has since transformed the field. Satellite imagery combined with artificial intelligence (AI) has become a cornerstone for predicting urban indicators of Union of Concerned Scientists (2022). Recent studies have demonstrated the potential of combining satellite and street-level imagery with deep learning to understand socioeconomic conditions across diverse contexts. Yeh et al. (2020) used publicly available satellite imagery and deep learning to map economic well-being in Africa, showcasing how remote sensing can bridge data gaps in resource-constrained settings. Similarly, Suel et al. (2023) explored the visual characteristics of poverty and wealth in 12 cities across five high-income countries using street images, highlighting the importance of contextual and visual cues in assessing socioeconomic disparities. A common methodology involves using deep learning models to extract features from satellite images, which are then fed into regression algorithms to predict poverty indicators. This approach has been applied successfully in African countries Xie et al. (2016); Jean et al. (2016) and Brazilian cities Silva and Rodrigues (2024). However, its focus on isolated indicators limits its applicability in large urban areas, where the variability of characteristics within deprived regions brings about new obstacles. Moreover, this method is computationally intensive, requiring vast amounts of imagery to comprehensively capture the reality of these areas Silva and Rodrigues (2024); Owusu et al. (2024).

Newer approaches have shifted focus to directly identifying deprived areas rather than individual metrics. Some studies emphasize morphological characteristics, which refer to the physical and structural features of regions, such as shape, size, distribution, layout, and orientation. These approaches extract spatial (contextual) features and apply them in machine learning models for binary classification (deprived, non-deprived) Chao et al. (2021); Owusu et al. (2024); Vanhuysse et al. (2021); Kuffer et al. (2023); Owusu et al. (2023) or multi-class classification to categorize deprived areas Georganos et al. (2021); Trento Oliveira et al. (2023).

Deep learning models, particularly Convolutional Neural Networks (CNNs), have also been used to learn spatial features and automatically detect informal settlements. CNNs have been employed for binary classification El Moudden and Amnai (2023); Raj et al. (2024); Mboga et al. (2017), multi-class classification Verma et al. (2019), and regression tasks to compare local citizens' perception of deprivation with AI-

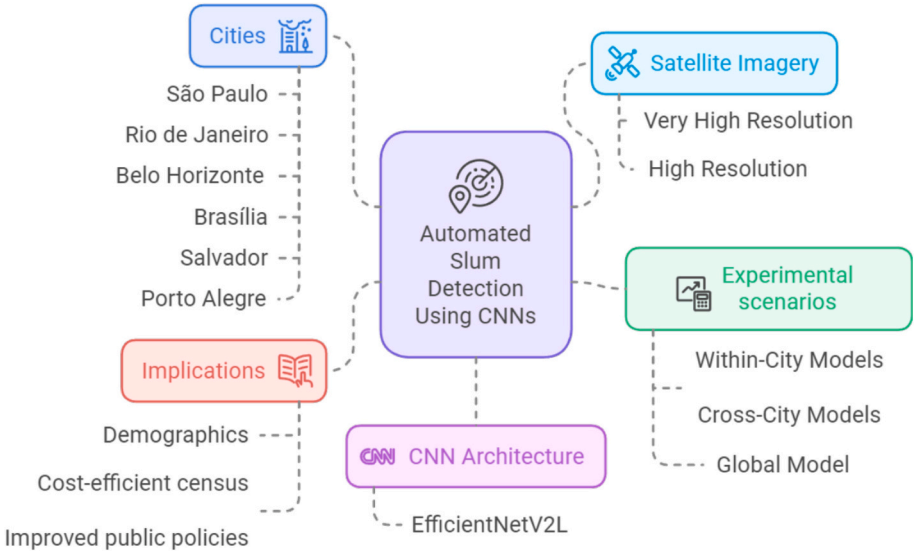


Fig. 1. Summary of this work's methodology.

generated assessments Abascal et al. (2024). Some studies propose hybrid approaches, combining machine learning and deep learning. For example, U-Net architectures have been used for semantic segmentation to extract building footprints, followed by clustering algorithms (e.g., K-Means) to identify deprivation levels or other characteristics Abascal, Rodríguez-Carreño, et al. (2022); Wang et al. (2023).

Despite these advancements, most studies fail to address the generalization capability of their models. Experiments are often limited to specific urban contexts, and the proposed methods are rarely validated across different locations. Consequently, the generalization and transferability of these models remain uncertain, with significant performance discrepancies observed when models trained in one city are applied to others Owusu et al. (2024); Vanhuysse et al. (2021); Georganos et al. (2021). The generalization capabilities of deep learning models for automatic slum mapping, in particular, have received little attention.

Therefore, further exploration is needed to assess the transferability of deep learning models, especially in a country as diverse as Brazil, where slum characteristics vary widely. Additionally, it is relevant to investigate whether the use of Very High Resolution (VHR) imagery impacts model accuracy, as the complexity of object-level details in these images could potentially reduce performance Owusu et al. (2024). In Section 5, we compare our methodology and results with those of previous studies to address these gaps.

3. Materials and methods

This study's methodology is structured into four stages: data acquisition, data processing, model training, and analysis. The process begins with the collection of relevant geographic and satellite data, followed by processing steps to prepare the data for modeling. At the core of the methodology is the training of a Convolutional Neural Network (CNN) to classify urban areas into slum and non-slum categories. The model is then evaluated in three scenarios (With-city, Cross-city, and Global) to assess its generalization capability. The overall workflow, detailing each step from data acquisition to analysis, is illustrated in Fig. 2.

3.1. Cities of reference

To ensure robust evaluation, models were trained and tested in multiple Brazilian cities, selected for their diverse urban environments and varying prevalence of informal settlements. These cities are:

- São Paulo, São Paulo (SP): the largest and most densely populated city in Brazil, São Paulo features a complex urban fabric interwoven with significant slum areas. Although slums occupy only 4 % of the city's surface, they house over 15 % of its population, resulting in extremely high population densities IBGE (2022).
- Salvador, Bahia (BA): known for its historical and cultural significance, Salvador hosts numerous informal settlements, often located along the coastline and characterized by dense, irregular housing. Slums make up 9 % of the city's area, housing 42 % of the population IBGE (2022).
- Belo Horizonte, Minas Gerais (MG): as the capital of the state of Minas Gerais, Belo Horizonte presents a mix of planned urban areas and spontaneous settlements. Slums frequently occupy hilly terrain. Slums occupy 5 % of the city's area and house 13 % of its population IBGE (2022).
- Brasília, Distrito Federal (DF): despite its status as a planned city, Brasília has developed informal settlements on its periphery, including Sol Nascente, the largest slum in Brazil. This makes it a compelling case for testing the model in less conventional environments. Slums represent less than 1 % of the city's area and house 7 % of the population IBGE (2022).
- Rio de Janeiro, Rio de Janeiro (RJ): renowned for its dramatic topography, Rio de Janeiro's slums vary widely in their characteristics, with some situated in flat urban areas and others on iconic hillsides, such as Rocinha, one of the largest slums in the country. Slums occupy almost 5 % of the city's area, housing 21 % of its population IBGE (2022).
- Porto Alegre, Rio Grande do Sul (RS): located in southern Brazil, Porto Alegre's slums are primarily found in flatter, urbanized areas, with fewer settlements on hillsides. Slums occupy 4 % of the city's area and house 13 % of the population IBGE (2022).

The selection of these cities provides a comprehensive evaluation of the CNN model's scalability and generalization across diverse urban landscapes, each with unique physical and socioeconomic characteristics. Figs. 4–9 showcase examples of informal settlements in each city, while Fig. 3 highlights the city boundaries considered in this study.

3.2. Data acquisition

Data for this study was obtained from a combination of governmental datasets and satellite imagery, specifically selected for robust experimentation.

3.2.1. Dataset AGSN

The primary dataset used for identifying and classifying slum areas is the Aglomerados Subnormais (AGSN), translated as Subnormal Agglomerations IBGE (2010). According to the Brazilian Institute of Geography and Statistics (IBGE), AGSNs refer to irregular land occupations primarily used for housing, characterized by disorganized urban layouts, limited access to essential public services, and locations often subject to land-use restrictions. In 2019, IBGE conducted a preliminary mapping of AGSNs as a preparatory step for the 2020 census. This initiative not only supported census operations but also raised public awareness about vulnerable populations during the COVID-19 pandemic. The preliminary mapping identified over 13,000 informal settlements, comprising more than 5 million households IBGE (2020).

Dataset AGSN is ideally suited for supervised machine learning and is among the most comprehensive open-source datasets on slums worldwide. It includes vector data that precisely identifies city regions as either slum or non-slum, providing a reliable data source. Fig. 10 illustrates the dataset provided for the city of São Paulo. In the following, we use São Paulo for illustration purposes; but, in the experiments, we use all the reference cities (Section 3.1) over the same training-testing protocols.

3.2.1.1. Vector data. For this study, vector data was obtained from IBGE platforms, providing information on administrative boundaries[1] and AGSN mappings.[2] These data was used to create the grid system used in data processing and to overlay AGSN data with satellite imagery.

3.2.2. Satellite imagery

Two types of satellite imagery were employed: Very High Resolution (VHR) and High Resolution (HR), as illustrated in Fig. 11. Additionally, Fig. 12 presents the color histograms for "slum" and "non-slum" images in the city of São Paulo, highlighting differences in their color distributions. From the plot, it is possible to notice subtle differences in the three color channels; we hypothesize that color is a feature that guided the CNNs in our problem domain.

[1] https://portaldemapas.ibge.gov.br/portal.php

[2] https://www.ibge.gov.br/geociencias/organizacao-do-territorio/tipologias-do-territorio/15788-favelas-e-comunidades-urbanas.html

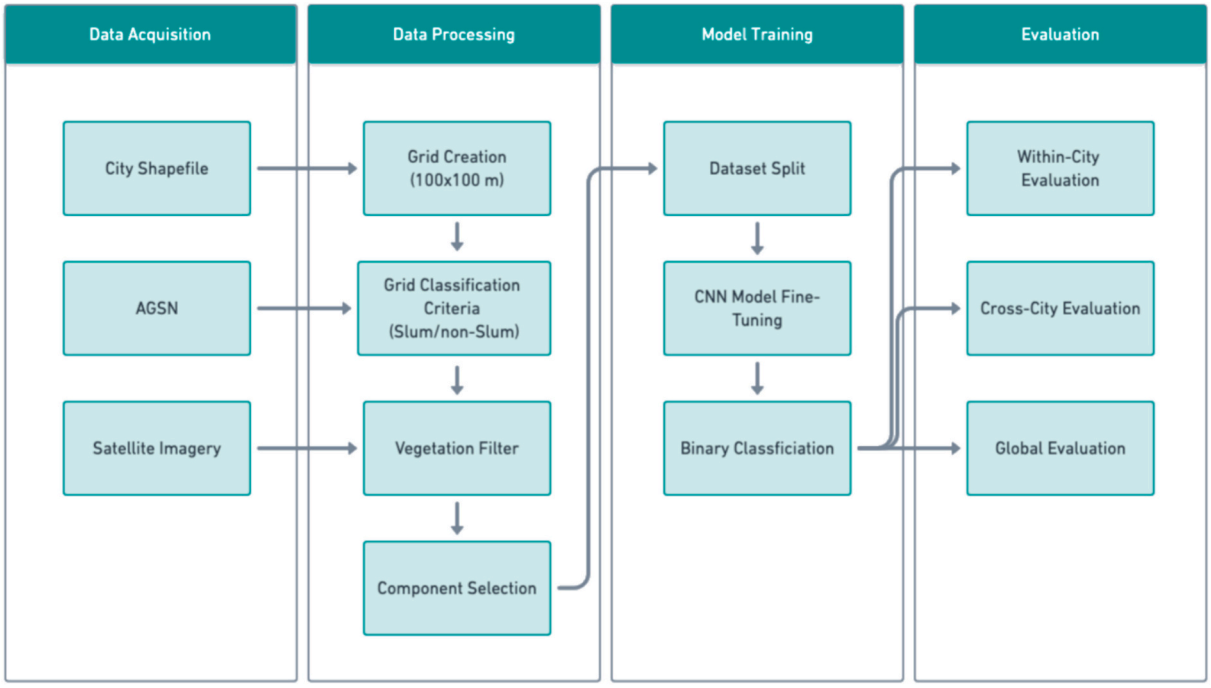


Fig. 2. Overview of the methodology, including data acquisition, processing, model training, and evaluation. We repeat this process for every city in our dataset.

- VHR images, with a spatial resolution of less than 1 m per pixel, were sourced from the Google Maps Static, freely accessible by the time of this work API.[3] These RGB images have a resolution of 400 × 400 pixels at a zoom level of 19, covering an area of approximately 100 × 100 m (10,000 m^2)—the reference size for grid components (see Section 3.3.1).
- HR images were sourced from the Sentinel-2 satellite, which offers free imagery with a spatial resolution of 10 m per pixel. Each image covers the same area as the VHR images (100 × 100 m) but at a lower resolution. To ensure cloud-free data, Google Earth Engine was used alongside the precomputed s2cloudless product. The s2cloudless algorithm, developed by the EO Research team at Sinergise,[4] employs gradient boosting to detect clouds across any resolution. The algorithm uses ten Sentinel-2 bands as input to generate a cloud probability map, which is converted into a cloud mask using a defined threshold Skakun et al. (2022). The image selection process involved applying an initial cloud cover filter to exclude images with high cloud presence. Subsequently, the s2cloudless mask is used for refinement.

3.3. Data processing

The data processing pipeline was implemented using QGIS, an open-source GIS software that offers robust tools for spatial data manipulation. The following steps outline the process used to prepare the data for model training.

3.3.1. Grid creation

Vector data for each city were divided into a uniform grid system, with each grid component covering an area of 100 × 100 m (10,000 m^2). This grid structure provided a systematic framework for classifying and analyzing different parts of the city.

3.3.2. Grid classification criteria

The classification of each grid component was based on the AGSN dataset (Section 3.2.1), which provides detailed information on slum coverage. For each grid component, the proportion of the area covered by slums was calculated. Grid components where more than 50 % of the area overlapped with slums were labeled as "slum"; the remaining grid components, with less than 50 % of the area overlapped with slums, were classified as "non-slum", as detailed in Eq. 1.

$$P = \frac{A_{\text{slum}}}{A_{\text{grid}}} \tag{1}$$

where P is the proportion of the grid area covered by slums; A_{slum} is the area of the grid component overlapping with slum regions; and A_{grid} is the total area of the grid component. If $P > 0.5$, a grid component is labeled as "slum".

3.3.3. Vegetation filter

Many cities in the dataset contain extensive vegetation, rivers, and coastal areas. Including such regions in the model would not enhance performance, as they lack urban characteristics. Since the analysis utilizes 3-band (RGB) images, the widely used Normalized Difference Vegetation Index (NDVI), which requires the near-infrared band, could not be applied. Instead, we employed the VIGreen vegetation index Gitelson et al. (2002), which uses only the red and green bands to effectively identify regions with dense vegetation, rivers, and oceans. This allowed us to focus the analysis on urban areas.

3.3.4. Component selection

After filtering grid components based on urban characteristics, a balanced and randomized selection process was used to build the training dataset. The dataset consisted of 50 % "slum" and 50 % "non-slum" grid components. For each selected grid component, a corresponding satellite image was extracted, covering the same 100 × 100 m area. This process is illustrated in Fig. 13.

A total of 3000 grid components/images were selected from each of the cities—São Paulo, Salvador, Brasília, Rio de Janeiro, Belo Horizonte, and Porto Alegre—with an equal distribution between the "slum" and "non-slum" categories. Table 1 provides a detailed breakdown of the

[3] https://developers.google.com/maps/documentation/maps-static/overview

[4] https://www.sinergise.com/en/news/eo-browser-goes-public

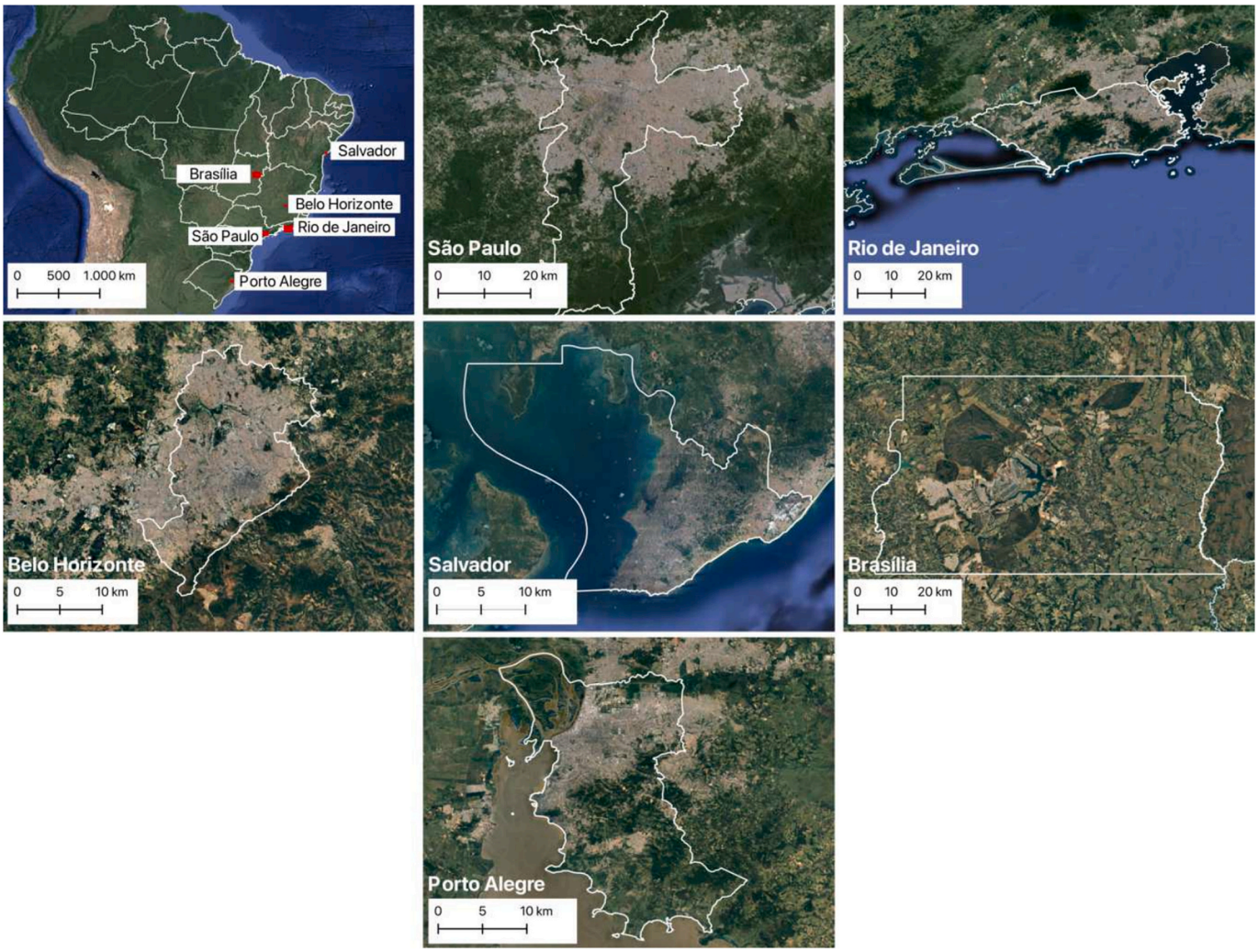


Fig. 3. Maps and photos illustrating the cities included in this study.

Fig. 4. Paraisópolis - São Paulo.

grid components selected for each city.

3.4. Model training

A deep learning model was trained to identify slum areas from satellite imagery. The experiments were conducted on a Ryzen 9 CPU with 32GB of RAM and an NVIDIA RTX 3090 GPU with 24GB of VRAM. The models were implemented using Keras with TensorFlow as the backend, running on Linux Pop!_OS 22.04 LTS. The following steps outline the training process:

3.4.1. Dataset Split

The dataset was randomly stratified and divided into three subsets: training (70 %), validation (15 %), and testing (15 %), following the work of Roshan Joseph (2022). The training set was used to optimize the model's parameters, the validation set for hyperparameter tuning and overfitting prevention, and the test set to evaluate final model performance. This split ensures robust generalization to unseen data.

Fig. 5. Beiru – Salvador.

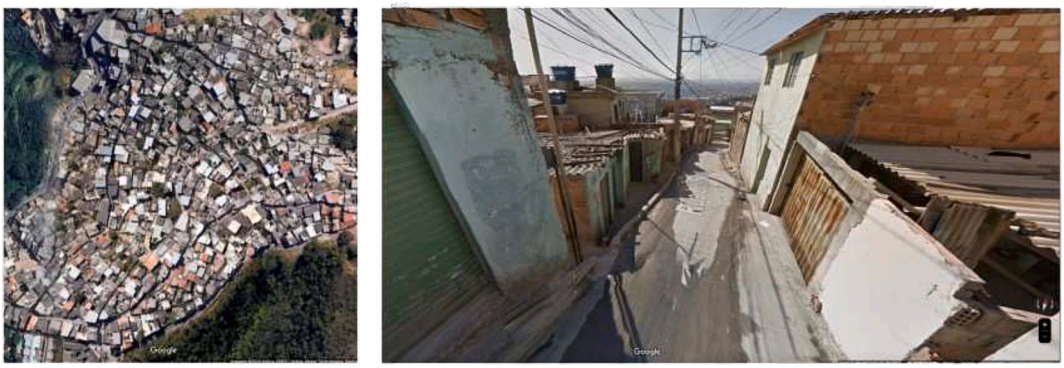


Fig. 6. Aglomerado da Serra - Belo Horizonte.

Fig. 7. Sol Nascente – Brasília.

Fig. 8. Rocinha - Rio de Janeiro.

Fig. 9. Vila Cruzeiro do Sul - Porto Alegre.

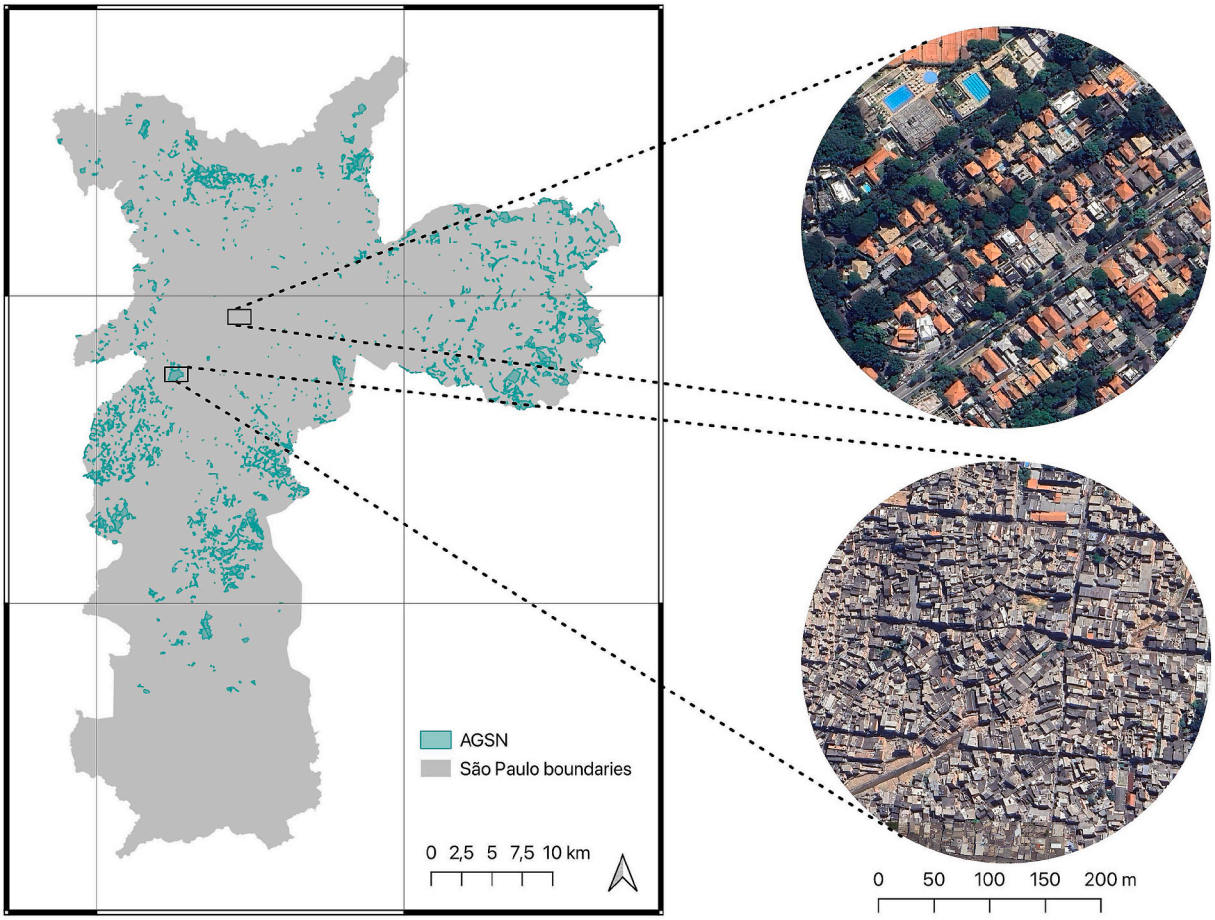


Fig. 10. Example of AGSN data for São Paulo, highlighting slum (spot below) and non-slum (spot above) regions.

3.4.2. CNN model fine-tuning

For this task, the EfficientNetV2L architecture Tan and Le (2021) was selected due to its superior balance between accuracy and computational efficiency. EfficientNetV2L improves upon EfficientNetV1 by incorporating enhanced "Compound Scaling" and introducing "Fused-MBConv" layers, which optimize the trade-offs between depth, width, and resolution.

The model's weights were initialized with those from a pre-trained EfficientNetV2L model on the ImageNet dataset Russakovsky et al. (2015). The convolutional layers remained unfrozen, allowing the model to be fine-tuned specifically for satellite imagery. This fine-tuning process helped the model to adapt to the unique characteristics of the satellite images, improving its performance in identifying slum areas.

3.4.3. Binary classification

The model was trained for binary classification, distinguishing between "slum" and "non-slum" grid components. Its output layer consists of a single neuron with a sigmoid activation function, mapping the output to a probability between 0 and 1. The sigmoid function, $\sigma(x) = \frac{1}{1+e^{-x}}$, compresses the output into a narrow range, making it sensitive to variations near 0 and 1. This allows the model to produce confidence levels, which can be adjusted to prioritize metrics like recall in specific applications.

The training process employed the Adam optimizer, see Eq. 2:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{m_t}{\sqrt{v_t} + \epsilon} \tag{2}$$

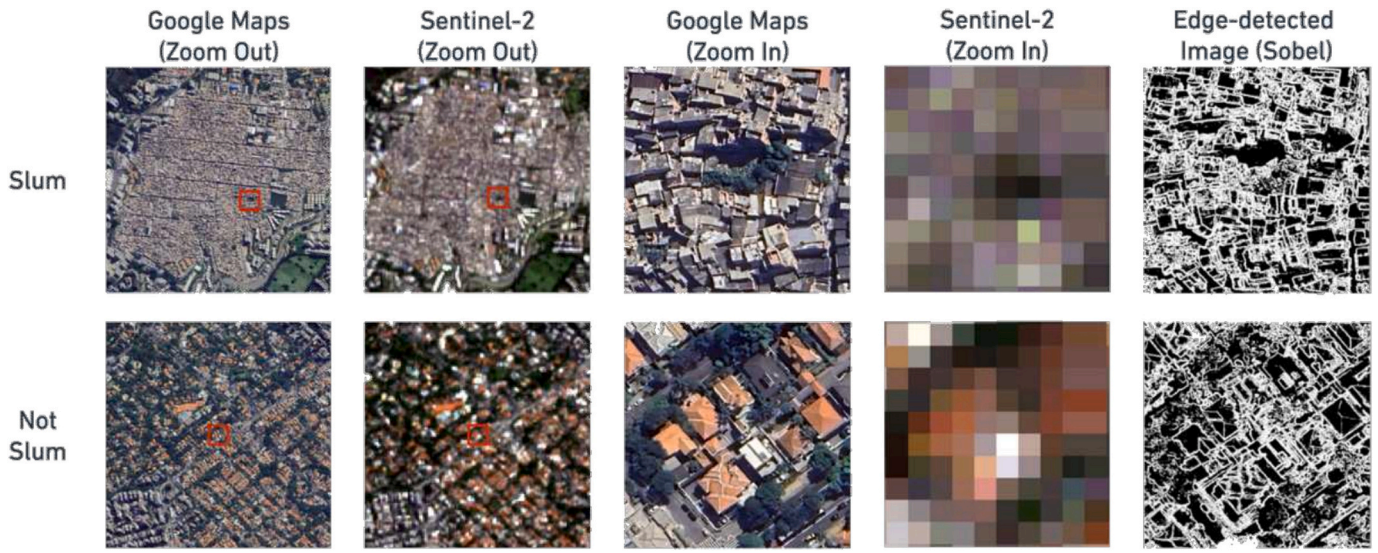


Fig. 11. Comparison between satellite imagery with Very High Resolution (VHR) – sourced from Google Maps Static API, and with High Resolution (HR) – sourced from satellite Sentinel-2, demonstrating the level of detail captured by each. At the last column, we present the result of the Sobel edge detection algorithm Gonzalez and Woods (2008) – we hypothesize that the density of edges is another feature, together with color, guiding the CNNs of our methodology.

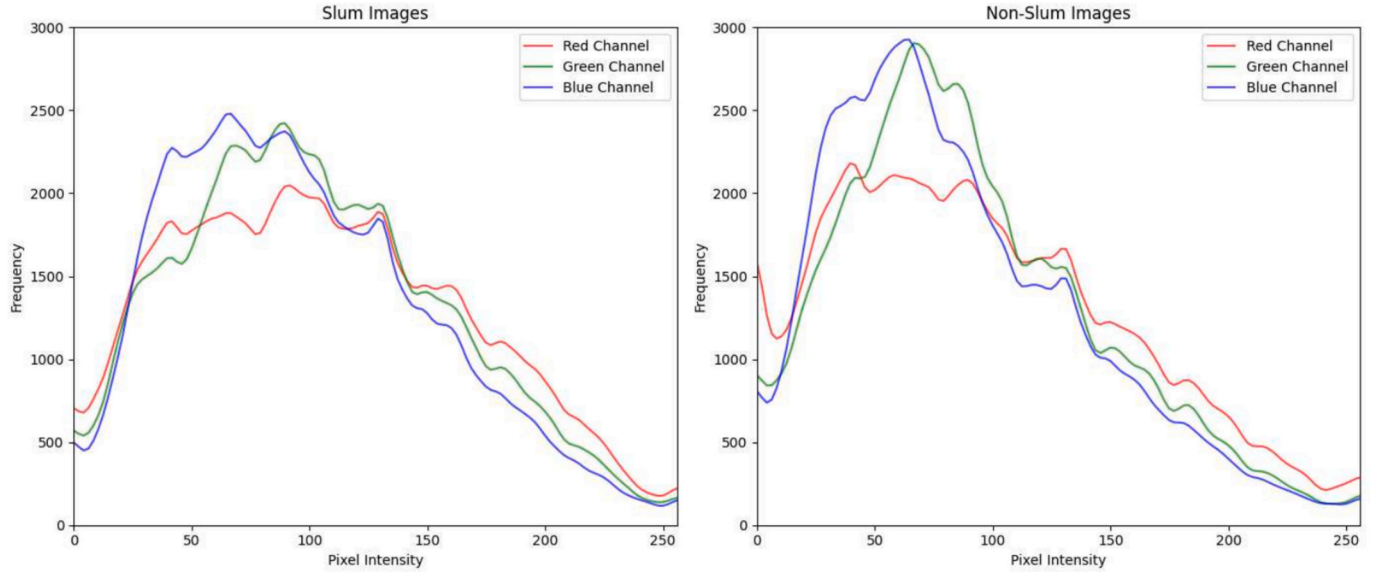


Fig. 12. Average color histograms (RGB) summarizing 3000 satellite images from São Paulo. The left column displays the histograms for "slum" images, while the right column shows the histograms for "non-slum" images.

where θ_t are the model parameters at iteration t; η is the learning rate (set to 1×10^{-4}); m_t and v_t are the estimates of the first and second moments of the gradients; and ϵ is a small constant to prevent division by zero. Eq. 2 adjusts the model's parameters during training by combining momentum and adaptive learning rates. Momentum (m_t) smooths updates by considering past gradients, while adaptive scaling (v_t) ensures stable updates by adjusting step sizes based on gradient magnitudes. This balance allows Adam to converge efficiently and reliably, even in noisy or complex optimization landscapes.

On average, the models required 20 epochs to converge, with early stopping to prevent overfitting – we employed the binary cross-entropy loss function, as detailed in Eq. 3. Each training session took approximately 30 min to complete.

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log(p_i) + (1-y_i)\cdot\log(1-p_i)\right] \quad (3)$$

where L is the loss; N is the number of samples; y_i is the true label (0 for non-slum, 1 for slum) of sample i; and p_i is the predicted probability that sample i is a slum area.

The choice of binary classification over segmentation was driven by the study's focus on capturing broader patterns of slum presence within predefined grid areas rather than achieving pixel-level precision. In satellite images, segmentation may introduce ambiguity due to the blending of slum and non-slum features within individual pixels. Classification, by leveraging grid-level information, enables robust discrimination of mixed areas. This approach aligns better with the study's objectives and the resolution of the available imagery, balancing computational efficiency with actionable insights.

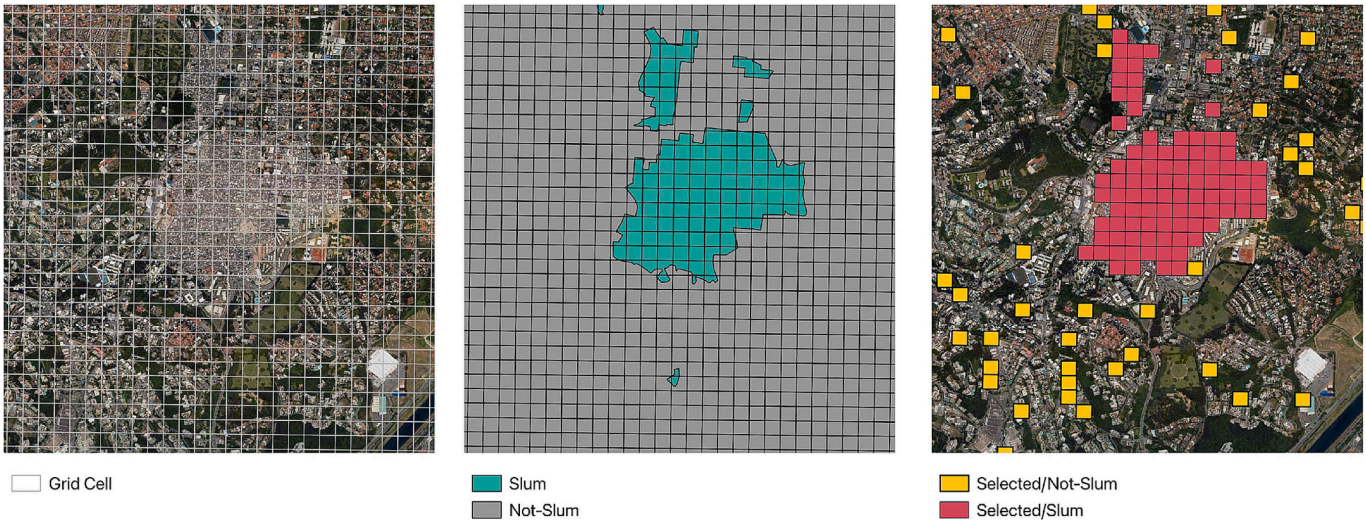


Fig. 13. Illustration of the grid creation process based on imagery, grid component discrimination based on AGSN data, and random balanced selection of components.

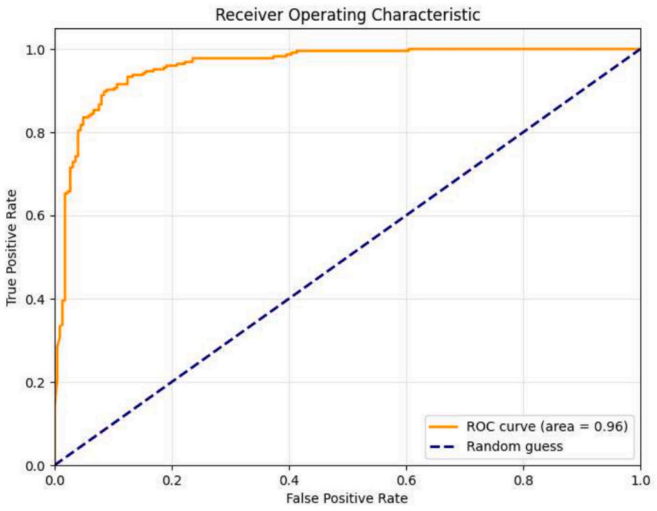


Fig. 14. ROC curve of the model trained on Very High Resolution (VHR) images for the city of São Paulo.

4. Results

4.1. Within-city evaluation

The initial evaluation considered training a separate model over the training data of each selected city and testing it over the testing data of the same city—within-city evaluation. As outlined in Section 3.3.4, a total of 3000 grid components (100 × 100 m each) and their corresponding satellite images were selected for each city, with equal representation of "slum" and "non-slum" classifications.

Two models were trained per city: one utilizing Very High Resolution (VHR) images and the other High Resolution (HR) images. To ensure consistency and compatibility, all images were resized to 224 × 224 pixels prior to training. HR images, originally at 10 × 10 pixels, underwent an upscaling using the nearest neighbor resampling method proposed by Patil (2018) – see Eq. 4, aligning their spatial resolution to the input size. Similarly, VHR images, initially at 400 × 400 pixels, were downsampled to 224 × 224 pixels using nearest neighbor interpolation. The models were based on the EfficientNetV2L architecture, as described in Section 3.4.2 and underwent fine-tuning with satellite images to adapt their weights for slum detection.

$$I_{\text{resized}}(x,y) = I\left(\text{round}\left(\frac{x}{s}\right), \text{round}\left(\frac{y}{s}\right)\right) \quad (4)$$

where $I_{\text{resized}}(x,y)$ is the pixel value at position (x,y) in the resized image; I is the original image; s is the scaling factor; and function round() rounds to the nearest integer.

Models trained on VHR images achieved superior performance across most of the cities, with F1 scores exceeding 0.90 in São Paulo, Rio de Janeiro, Belo Horizonte, and Brasília. In contrast, Salvador and Porto Alegre reported F1 scores of 0.85 and 0.81, respectively. When using HR images, the models achieved F1 scores above 0.80 in São Paulo, Belo Horizonte, and Brasília, while Rio de Janeiro and Porto Alegre scored only above 0.70. Salvador obtained an F1 score of 0.67. These results underscore the impact of image resolution on model performance, particularly in cities with more complex or heterogeneous urban layouts. Table 2 summarizes the evaluation metrics for each city and Fig. 14 illustrates the ROC curve of the model trained on São Paulo using VHR images – the plot demonstrates the high performance for a within-city experiment over the largest Brazilian city.

4.2. Cross-city evaluation

To assess the model's ability to generalize across different Brazilian cities, we conducted a cross-city evaluation by testing models trained with data from one city (source) on the test data from each of the other cities (targets). Importantly, the models were not retrained or fine-tuned with data from the test cities; instead, they were directly evaluated using the pre-trained weights. This experiment was performed for both VHR and HR models.

Fig. 15 presents the F1 scores for the cross-city evaluation using VHR images. The y-axis corresponds to the source cities where the models were trained, and the x-axis corresponds to the target cities where the models were tested. Diagonal values indicate the F1 scores for models tested on the same city they were trained on, providing a baseline for within-city performance. Off-diagonal values show the model's performance when applied to unseen cities, highlighting its generalization capabilities. Similarly, Fig. 16 displays the results for the HR models.

Detailed results of these evaluations are provided in Tables 3 and 4.

4.3. Global evaluation

To further evaluate the generalization of our CNN methodology, we

Fig. 15. F1 scores for the cross-city evaluation using VHR images. The y-axis represents the source city (training), and the x-axis represents the target city (testing).

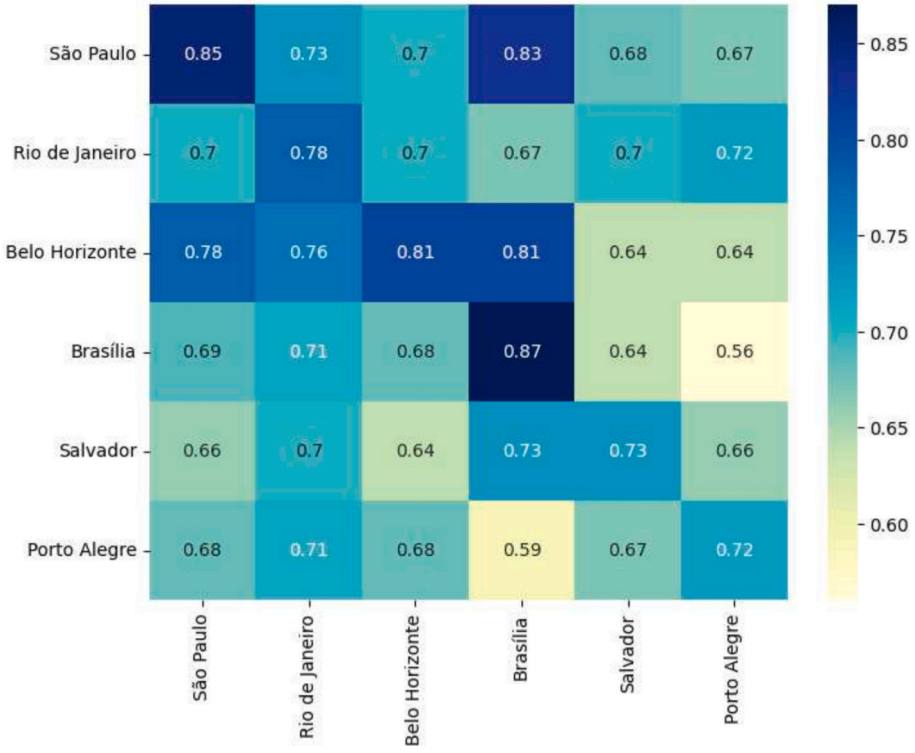


Fig. 16. F1 scores for the cross-city evaluation using HR images. The y-axis represents the source city (training), and the x-axis represents the target city (testing).

trained a global model using a dataset comprising images from all the selected cities: São Paulo, Rio de Janeiro, Belo Horizonte, Brasília, Salvador, and Porto Alegre. The aim was to assess the performance of a model that learned from a diverse range of urban environments.

The results are summarized in Table 5. When using VHR images, the model achieved F1 scores of 0.91 in São Paulo and Belo Horizonte, 0.89 in Rio de Janeiro and Brasília, 0.85 in Salvador, and 0.78 in Porto Alegre. Conversely, the model trained with HR images exhibited lower performance, with F1 scores ranging from 0.55 to 0.70 across all cities.

The generalized model consistently outperformed most models when

Table 1
Grid components selected for model training, with total grid components and class distribution per city.

City	Total Components	Slum	Non-slum	Total Selected
São Paulo	156,623	3781	152,842	3000
Salvador	70,202	6256	63,946	3000
Brasília	579,153	4235	574,918	3000
Rio de Janeiro	122,424	5579	116,845	3000
Belo Horizonte	33,875	1673	32,202	3000
Porto Alegre	50,586	2961	47,625	3000

compared to the cross-city evaluation, achieving F1 scores above 0.85 in São Paulo, Belo Horizonte, Rio de Janeiro, Brasília and Salvador for VHR images. Additionally, it maintained acceptable performance in Porto Alegre, with F1 scores exceeding 0.78.

4.4. Global evaluation - leave-one-out

To assess the robustness of the global model, we conducted yet another experiment – we used technique leave-one-out with 6 cities. We trained the model with five cities and tested it over the left-out city; we trained and tested for each of the 6 cities according to the leave-one-out protocol. This experiment allowed us to evaluate the potential of the methodology over unseen urban contexts. The quality of global predictions is significantly influenced by the spatial distribution and representativeness of the training data, particularly when models are applied to regions with conditions that differ from those in the training set Meyer and Pebesma (2022).

Table 2
Classification results for the within-city evaluation using VHR and HR images.

City	Data source	Accuracy	Precision	Recall	AUC	PRC	F1
São Paulo	**VHR-Google Maps**	**0.94**	**0.94**	**0.94**	**0.96**	**0.96**	**0.94**
	HR-Sentinel-2	0.85	0.85	0.85	0.91	0.91	0.85
Belo Horizonte	**VHR-Google Maps**	**0.93**	**0.93**	**0.92**	**0.98**	**0.97**	**0.93**
	HR-Sentinel-2	0.81	0.81	0.81	0.87	0.85	0.81
Rio de Janeiro	**VHR-Google Maps**	**0.90**	**0.90**	**0.91**	**0.96**	**0.95**	**0.91**
	HR-Sentinel-2	0.79	0.79	0.78	0.86	0.85	0.78
Brasília	**VHR-Google Maps**	**0.90**	**0.90**	**0.90**	**0.94**	**0.93**	**0.90**
	HR-Sentinel-2	0.86	0.87	0.86	0.95	0.95	0.87
Salvador	**VHR-Google Maps**	**0.85**	**0.85**	**0.85**	**0.90**	**0.89**	**0.85**
	HR-Sentinel-2	0.67	0.67	0.67	0.76	0.76	0.67
Porto Alegre	**VHR-Google Maps**	**0.81**	**0.81**	**0.81**	**0.86**	**0.84**	**0.81**
	HR-Sentinel-2	0.72	0.72	0.72	0.78	0.76	0.72

Table 3
Classification results for the cross-city evaluation using VHR images. The training occurred over data of the source city, and the test occurred over the data of the target city.

Source	Target	Accuracy	Precision	Recall	AUC	PRC	F1
São Paulo	São Paulo	0.93	0.93	0.93	0.97	0.97	0.93
	Rio de Janeiro	0.87	0.87	0.87	0.93	0.91	0.87
	Belo Horizonte	0.90	0.90	0.90	0.94	0.93	0.90
	Brasília	0.81	0.81	0.81	0.83	0.81	0.81
	Salvador	0.78	0.78	0.78	0.81	0.79	0.78
	Porto Alegre	0.75	0.75	0.75	0.77	0.72	0.75
Belo Horizonte	Belo Horizonte	0.93	0.93	0.92	0.98	0.97	0.93
	Rio de Janeiro	0.85	0.85	0.85	0.90	0.89	0.85
	São Paulo	0.85	0.85	0.85	0.91	0.90	0.86
	Brasília	0.80	0.80	0.80	0.84	0.82	0.80
	Salvador	0.74	0.74	0.74	0.80	0.78	0.74
	Porto Alegre	0.76	0.76	0.77	0.80	0.73	0.76
Rio de Janeiro	Rio de Janeiro	0.90	0.90	0.91	0.96	0.95	0.91
	Belo Horizonte	0.86	0.87	0.86	0.93	0.93	0.87
	São Paulo	0.86	0.86	0.85	0.92	0.92	0.86
	Brasília	0.76	0.76	0.76	0.78	0.77	0.76
	Salvador	0.80	0.80	0.80	0.85	0.83	0.80
	Porto Alegre	0.76	0.76	0.75	0.76	0.76	0.76
Brasília	Brasília	0.90	0.90	0.90	0.94	0.93	0.90
	Belo Horizonte	0.77	0.77	0.77	0.85	0.85	0.77
	São Paulo	0.77	0.77	0.77	0.84	0.82	0.78
	Rio de Janeiro	0.77	0.77	0.77	0.83	0.80	0.78
	Salvador	0.73	0.73	0.73	0.79	0.76	0.72
	Porto Alegre	0.69	0.69	0.69	0.70	0.64	0.69
Salvador	Salvador	0.85	0.85	0.85	0.90	0.89	0.85
	Belo Horizonte	0.82	0.82	0.82	0.89	0.88	0.83
	São Paulo	0.82	0.82	0.82	0.90	0.89	0.82
	Rio de Janeiro	0.82	0.82	0.82	0.90	0.89	0.83
	Brasília	0.83	0.83	0.83	0.89	0.88	0.83
	Porto Alegre	0.74	0.74	0.74	0.76	0.72	0.74
Porto Alegre	Porto Alegre	0.81	0.81	0.81	0.86	0.84	0.81
	Belo Horizonte	0.79	0.79	0.79	0.84	0.82	0.79
	São Paulo	0.77	0.77	0.77	0.81	0.79	0.77
	Rio de Janeiro	0.83	0.83	0.83	0.88	0.87	0.83
	Brasília	0.71	0.71	0.71	0.75	0.69	0.71
	Salvador	0.76	0.76	0.76	0.81	0.79	0.76

Table 4
Classification results for the cross-city evaluation using HR images. The training occurred over data of the source city, and the test occurred over the data of the target city.

Source	Target	Accuracy	Precision	Recall	AUC	PRC	F1
São Paulo	São Paulo	0.85	0.85	0.85	0.91	0.91	0.85
	Rio de Janeiro	0.73	0.73	0.72	0.78	0.75	0.73
	Belo Horizonte	0.70	0.70	0.69	0.73	0.68	0.70
	Brasília	0.83	0.83	0.84	0.87	0.85	0.83
	Salvador	0.68	0.68	0.68	0.70	0.65	0.68
	Porto Alegre	0.67	0.67	0.67	0.69	0.64	0.67
Belo Horizonte	Belo Horizonte	0.81	0.81	0.81	0.87	0.85	0.81
	Rio de Janeiro	0.76	0.76	0.75	0.78	0.76	0.76
	São Paulo	0.78	0.78	0.78	0.85	0.83	0.78
	Brasília	0.81	0.81	0.81	0.88	0.85	0.81
	Salvador	0.64	0.65	0.64	0.66	0.63	0.64
	Porto Alegre	0.64	0.64	0.64	0.65	0.63	0.64
Rio de Janeiro	Rio de Janeiro	0.79	0.79	0.78	0.86	0.85	0.78
	Belo Horizonte	0.70	0.70	0.69	0.72	0.73	0.70
	São Paulo	0.70	0.70	0.70	0.73	0.73	0.70
	Brasília	0.67	0.67	0.67	0.70	0.69	0.67
	Salvador	0.70	0.70	0.70	0.72	0.72	0.70
	Porto Alegre	0.72	0.72	0.72	0.76	0.74	0.72
Brasília	Brasília	0.86	0.87	0.86	0.95	0.95	0.87
	Belo Horizonte	0.68	0.67	0.68	0.71	0.71	0.68
	São Paulo	0.69	0.69	0.69	0.71	0.70	0.69
	Rio de Janeiro	0.71	0.71	0.71	0.75	0.75	0.71
	Salvador	0.64	0.64	0.64	0.61	0.60	0.64
	Porto Alegre	0.56	0.56	0.56	0.60	0.61	0.56
Salvador	Salvador	0.73	0.73	0.73	0.78	0.78	0.73
	Belo Horizonte	0.64	0.64	0.64	0.66	0.65	0.64
	São Paulo	0.66	0.66	0.66	0.70	0.69	0.66
	Rio de Janeiro	0.72	0.72	0.73	0.79	0.79	0.73
	Brasília	0.73	0.73	0.73	0.74	0.75	0.73
	Porto Alegre	0.66	0.66	0.66	0.68	0.62	0.66
Porto Alegre	Porto Alegre	0.72	0.72	0.72	0.78	0.76	0.72
	Belo Horizonte	0.68	0.68	0.68	0.70	0.67	0.68
	São Paulo	0.68	0.68	0.68	0.70	0.63	0.68
	Rio de Janeiro	0.71	0.71	0.71	0.78	0.76	0.71
	Brasília	0.59	0.59	0.58	0.62	0.58	0.59
	Salvador	0.67	0.67	0.67	0.73	0.71	0.67

Table 5
Performance of the global model on individual cities using VHR and HR images.

City	Data source	Accuracy	Precision	Recall	AUC	PRC	F1
São Paulo	**VHR-Google Maps**	**0.91**	**0.91**	**0.91**	**0.96**	**0.95**	**0.91**
	HR-Sentinel-2	0.86	0.86	0.86	0.93	0.92	0.86
Belo Horizonte	**VHR-Google Maps**	**0.91**	**0.91**	**0.91**	**0.97**	**0.97**	**0.91**
	HR-Sentinel-2	0.75	0.75	0.74	0.85	0.85	0.75
Rio de Janeiro	**VHR-Google Maps**	**0.89**	**0.89**	**0.89**	**0.94**	**0.93**	**0.89**
	HR-Sentinel-2	0.79	0.79	0.78	0.85	0.83	0.79
Brasília	**VHR-Google Maps**	**0.89**	**0.89**	**0.89**	**0.96**	**0.95**	**0.89**
	HR-Sentinel-2	0.86	0.86	0.86	0.93	0.93	0.86
Salvador	**VHR-Google Maps**	**0.85**	**0.86**	**0.85**	**0.91**	**0.89**	**0.85**
	HR-Sentinel-2	0.68	0.68	0.67	0.74	0.73	0.68
Porto Alegre	**VHR-Google Maps**	**0.78**	**0.79**	**0.78**	**0.86**	**0.85**	**0.78**
	HR-Sentinel-2	0.72	0.72	0.72	0.78	0.75	0.72

Table 6
Performance of the global model on individual cities using VHR and HR images - Leave-one-out technique.

City	Data source	Accuracy	Precision	Recall	AUC	PRC	F1
São Paulo	**VHR-Google Maps**	**0.86**	**0.86**	**0.86**	**0.91**	**0.90**	**0.86**
	HR-Sentinel-2	0.74	0.74	0.74	0.81	0.79	0.74
Belo Horizonte	**VHR-Google Maps**	**0.90**	**0.90**	**0.90**	**0.95**	**0.94**	**0.90**
	HR-Sentinel-2	0.74	0.73	0.75	0.81	0.78	0.74
Rio de Janeiro	**VHR-Google Maps**	**0.87**	**0.87**	**0.87**	**0.91**	**0.89**	**0.87**
	HR-Sentinel-2	0.77	0.77	0.77	0.83	0.80	0.77
Brasília	**VHR-Google Maps**	**0.76**	**0.76**	**0.75**	**0.81**	**0.78**	**0.76**
	HR-Sentinel-2	0.75	0.75	0.75	0.81	0.78	0.75
Salvador	**VHR-Google Maps**	**0.81**	**0.80**	**0.81**	**0.86**	**0.84**	**0.81**
	HR-Sentinel-2	0.63	0.63	0.63	0.67	0.64	0.63
Porto Alegre	**VHR-Google Maps**	**0.73**	**0.73**	**0.73**	**0.75**	**0.71**	**0.73**
	HR-Sentinel-2	0.66	0.66	0.66	0.67	0.62	0.66

The results in Table 6 show that when using VHR images, the model maintained satisfactory performance in most cities, with F1 scores of 0.90 in Belo Horizonte, 0.87 in Rio de Janeiro, 0.86 in São Paulo, 0.81 in Salvador, 0.76 in Brasília and 0.73 in Porto Alegre. In comparison, models using HR images exhibited lower performance, with F1 scores ranging from 0.63 to 0.77.

5. Discussion

This research provides insights into the application of deep learning for slum detection using satellite imagery in three settings: within-city, cross-city, and global; and considering two image resolutions: High Resolution (HR) and Very High Resolution (VHR).

5.1. Cross-City evaluation

The cross-city evaluation, where models trained on data from one city were tested on data from other cities without retraining or fine-tuning, yielded the lowest performance compared to within-city and global evaluations. As shown in Figs. 15 and 16, F1 scores for cross-city evaluations were generally lower, with significant variability depending on the source-target city pair. This highlights the difficulty of transferring models across regions with distinct urban and slum characteristics.

5.2. Within-City vs. global models

The results of the global evaluation (Tables 5 and 6) in comparison to the results of the within-city evaluation (Table 2) demonstrate that exposing the model to a broader range of slum characteristics and urban patterns does not necessarily lead to superior performance compared to models trained on data from individual cities. This suggests that the generalization benefits of a mixed dataset come at the cost of reduced specificity, as the model must learn features that are broadly applicable across diverse urban landscapes, rather than optimizing for the unique characteristics of a single city.

5.3. HR vs. VHR imagery

Across all evaluation settings—within-city, cross-city, and global—models utilizing Very High Resolution (VHR) images consistently outperformed those using High Resolution (HR) images. As detailed in Tables 2 and 5, VHR models achieved higher F1 scores, with improvements ranging from 5 to nearly 20 percentage points depending on the city and evaluation setting. Despite this pronounced difference in performance, the High Resolution 10 × 10 images still demonstrated discriminatory potential in the proposed task; this is evidence that such low-cost images can still perform reasonable if images with higher resolutions are not available.

The comparison of performance metrics between HR and VHR images requires careful consideration, as their varying spatial resolutions introduce differences in the scale and granularity of the input data. VHR images, with finer spatial details, provide a more precise representation of urban morphology, which can enhance the model's ability to detect nuanced patterns associated with slums. In contrast, HR images, with coarser resolution, may aggregate features from mixed areas, such as regions containing both slum and non-slum characteristics, potentially reducing model accuracy. This difference in spatial scale has implications for model interpretability and generalization. To ensure fair comparisons, future studies could normalize the input data by aggregating VHR pixels to match HR resolution or assess model performance on metrics that account for scale differences, such as grid-level precision. In this study, we considered the trade-offs between computational efficiency, data availability, and model accuracy, which become relevant for data with varying resolutions.

5.4. Comparison to previous works

In Table 7, we summarize the performance of our methodology in direct comparison to other works. In the table, one can see that we achieved results similar or superior to every other methodology. It is important to note that these comparisons do not rely on the same dataset (Location); accordingly, these numbers provide a relative estimative of performance, instead of an absolute perspective.

In the first stage of our experiments, where models were trained and tested within each city, F1 scores ranged from 0.81 to 0.94 using VHR imagery. These results align closely with those reported in previous studies. For instance, Mboga et al. Mboga et al. (2017) achieved an accuracy of 0.91 in Dar es Salaam, Tanzania, using a CNN model with VHR Quickbird imagery. Similarly, Verma et al. Verma et al. (2019) reported an accuracy of 0.94 in Mumbai, India, using a CNN trained on VHR Pleiades imagery, while performance dropped to 0.90 when using HR Sentinel-2B images. Using contextual features with machine learning, Owusu et al. Owusu et al. (2023) reported F1 scores of 0.77, 0.86, and 0.77 in Accra, Lagos, and Nairobi, respectively, with Sentinel-2 imagery. In a follow-up study, Owusu et al. Owusu et al. (2024) achieved F1 scores of 0.93, 0.58, and 0.92 in the same cities.

Regarding cross-city results, F1 scores varied significantly depending on the source-target city pairs, ranging from 0.69 to 0.93 across 36

Table 7
Non-absolute comparison of slum detection studies using deep learning and satellite imagery - each author used a different dataset.

Study	Location	Imagery Type	Method	Performance
Mboga et al. (2017)-Within-City	Dar es Salaam, TZ	VHR QuickBird	CNN	Accuracy: 0.91
Verma et al. (2019)-Within-City	Mumbai, IN	VHR Pleiades	CNN	Accuracy: 0.94
		HR Sentinel-2B	CNN	Accuracy: 0.90
Owusu et al. (2023)-Within-City	African Cities	HR Sentinel-2	ML + Features	F1 Score: 0.77–0.76
Owusu et al. (2023)-Cross-City	African Cities	HR Sentinel-2	ML + Features	F1 Score: 0.57–0.69
Owusu et al. (2023)-Global	African Cities	HR Sentinel-2	ML + Features	F1 Score: 0.63–0.84
Owusu et al. (2024)-Within-City	African Cities	HR Sentinel-2	ML + Features	F1 Score: 0.58–0.93
Owusu et al. (2024)-Cross-City	African Cities	HR Sentinel-2	ML + Features	F1 Score: 0.13–0.81
Owusu et al. (2024)-Global	African Cities	HR Sentinel-2	ML + Features	F1 Score: 0.68–0.86
This Work-Within-City	Brazilian Cities	VHR GMaps	EfficientNetV2L	F1 Score: 0.81–0.94 Accuracy: 0.81–0.94
This Work-Cross-City	Brazilian Cities	VHR GMaps	EfficientNetV2L	F1 Score: 0.69–0.93 Accuracy: 0.69–0.93
This Work-Global	Brazilian Cities	VHR GMaps	EfficientNetV2L	F1 Score: 0.78–0.91 Accuracy: 0.79–0.91
This Work-Global (Leave-one-out)	Brazilian Cities	VHR GMaps	EfficientNetV2L	F1 Score: 0.73–0.90 Accuracy: 0.73–0.90

combinations. Similar variability was observed by Owusu et al. Owusu et al. (2023), who reported F1 scores between 0.57 and 0.69 in their cross-city evaluations. In their subsequent work, Owusu et al. Owusu et al. (2024) observed even wider variability, with F1 scores ranging from 0.13 to 0.81 across African cities, highlighting the challenges of generalizing models across diverse urban landscapes.

For the global results, a generalized model trained on data from all six Brazilian cities, we achieved F1 scores ranging from 0.78 to 0.91. When applying the Leave-one-out technique, where the model was trained on five cities and tested on the remaining one, F1 scores ranged from 0.73 to 0.90. These results are comparable to those of Owusu et al. Owusu et al. (2023), who reported scores between 0.63 and 0.84, and Owusu et al. Owusu et al. (2024), who achieved F1 scores from 0.68 to 0.86 in three African cities. These findings reinforce the effectiveness of generalized models in capturing diverse urban patterns, albeit with some performance trade-offs compared to city-specific models.

These results demonstrate the success of our methodology in slum detection across diverse urban contexts using deep learning. Notably, our use of the EfficientNetV2L architecture, combined with fine-tuning on satellite imagery, consistently delivered competitive performance, particularly in the within-city evaluation where VHR models achieved F1 scores exceeding 0.90 in several cities. Additionally, even in the challenging cross-city setting, our models demonstrated transferability comparable to or exceeding previous works, with F1 scores as high as 0.90. The generalized model further underscores the robustness of our approach, achieving strong performance across all cities, with F1 scores between 0.78 and 0.91. These outcomes validate the effectiveness of our specific training pipeline, including the careful preprocessing of both VHR and HR images, the strategic use of diverse datasets, and the emphasis on scalability for practical applications in urban analysis.

5.4.1. *Morphological indicators*

In another approach, morphological indicators have gained traction in slum detection tasks Wang et al. (2023) because they provide valuable contextual and structural insights into urban landscapes, such as building density, size, layout, and spatial organization, which are often key characteristics of slum areas. These indicators complement satellite imagery by capturing physical patterns that may not always be evident in pixel-based data alone, enabling a more nuanced understanding of urban morphology. In future works, incorporating such indicators can improve model performance and generalizability, particularly when combined with high-resolution imagery.

5.5. *Challenges of ground truth variability*

One of the critical challenges in slum detection is the variability in ground truth definitions, which differ significantly across countries and cities due to variations in socioeconomic, cultural, and legal contexts. While this study benefits from the Brazilian AGSN dataset, which provides detailed and well-defined labels tailored to the local context, these definitions may not align with those used in other regions. Such disparities could hinder the generalizability of models trained on localized datasets, particularly in applications requiring global consistency, such as monitoring progress toward Sustainable Development Goals (SDGs). This challenge underscores the need for deeper discussions on how local definitions influence the generalizability of slum detection models. Future research could address this issue by incorporating domain adaptation techniques or additional contextual features to bridge regional differences. Furthermore, developing standardized global definitions for slums, while challenging, would significantly enhance the scalability and impact of deep learning models in this field. By situating this study within the broader context of slum detection challenges, we aim to highlight the importance of addressing these disparities to advance the field and support global urban monitoring efforts.

The study's use of a consistent definition of slums from the AGSN dataset ensures uniformity in labeling and robust model training, enhancing reproducibility. While tailored to the Brazilian context, the methodology could be adapted to other countries by incorporating local definitions and employing techniques like transfer learning or domain adaptation, enabling broader applicability.

6. Conclusion

The findings of this study demonstrate the significant potential of deep learning models, particularly CNNs, in leveraging satellite imagery for automated slum detection in urban environments. The successful application of these models across multiple Brazilian cities highlights the effectiveness of using VHR images to capture the complex characteristics of slums. However, the variation in model performance across different cities underscores the importance of accounting for regional differences in slum morphology and urban patterns.

Future work should focus on developing models that can better generalize across diverse urban contexts, possibly through the use of more advanced architectures like encoder-decoder networks (EDNs) that can handle imbalanced data and preserve spatial information. Additionally, expanding the classification beyond binary categories to include different types of slum environments could enhance model adaptability and usefulness for urban planning and policy-making.

By refining these approaches, we can improve the ability to monitor and address the challenges of urban inequality, contributing to more effective interventions in rapidly developing regions.

Availability of data and materials

The dataset used in this study is publicly available at https://data.mendeley.com/datasets/xg7p7rfrrp/1 (DOI: 10.17632/xg7p7rfrrp.1). The source code for data processing, model training, and evaluation is available at https://github.com/dev-jotape/cnn-slum-prediction.

CRediT authorship contribution statement

João P. da Silva: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **José F. Rodrigues-Jr:** Writing – review & editing, Validation, Supervision. **João P. de Albuquerque:** Writing – review & editing, Supervision.

Acknowledgements

This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation INV-045252. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

References

Abascal, A., Rodríguez-Carreño, I., Vanhuysse, S., Georganos, S., Sliuzas, R., Wolff, E., & Kuffer, M. (2022). Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. *Computers, Environment and Urban Systems, 95*, Article 101820.

Abascal, A., Rothwell, N., Shonowo, A., Thomson, D. R., Elias, P., Elsey, H., … Kuffer, M. (2022). "domains of deprivation framework" for mapping slums, informal settlements, and other deprived areas in lmics to improve urban planning and policy: A scoping review. *Computers, Environment and Urban Systems, 93*, Article 101770.

Abascal, A., Vanhuysse, S., Grippa, T., Rodriguez-Carreño, I., Georganos, S., Wang, J., … Wolff, E. (2024). Ai perceives like a local: predicting citizen deprivation perception using satellite imagery. *npj Urban Sustainability, 4*(1), 20.

Chao, S., Engstrom, R., Mann, M., & Bedada, A. (2021). Evaluating the ability to use contextual features derived from multi-scale satellite imagery to map spatial patterns of urban attributes and population distributions. *Remote Sensing, 13*(19), 3962.

Congresso Nacional. (2021). Draft budgetary plan – PLOA 2022 (PL n° 19/2021-CN). URL https://www2.camara.leg.br/orcamento-da-uniao/estudos/2021/subsidios-a-apreciacao-do-projeto-de-lei-orcamentaria-ploa-para-2022.

El Moudden, T., & Amnai, M. (2023). Building an efficient convolution neural network from scratch: A case study on detecting and localizing slums. *Scientific African, 20*, Article e01612.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., & Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing, 18*(6), 1373–1379.

Elvidge, C. D., Safran, J., Tuttle, B., Sutton, P., Cinzano, P., Pettit, D., … Small, C. (2007). Potential for global mapping of development via a nightsat mission. *GeoJournal, 69*, 45–53.

Georganos, S., Abascal, A., Kuffer, M., Wang, J., Owusu, M., Wolff, E., & Vanhuysse, S. (2021). Is it all the same? mapping and characterizing deprived urban areas using worldview-3 superspectral imagery. a case study in nairobi, kenya. *Remote Sensing, 13*(24), 4986.

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment, 80*(1), 76–87.

Gonzalez, R. C., & Woods, R. E. (2008). *Digital Image Processing* (3rd ed.). Prentice Hall.

IBGE. (2010). Subnormal agglomerates 2010. URL https://www.ibge.gov.br/en/geosciences/territorial-organization/territorial-typologies/17553-subnormal-agglomerates.html?edicao=17587.

IBGE. (2019). Subnormal agglomerates 2019 - preliminary results. URL https://www.ibge.gov.br/en/geosciences/territorial-organization/territorial-typologies/17553-subnormal-agglomerates.html?edicao=27744.

IBGE. (2020). Aglomerados subnormais 2019: Classificação preliminar e informações de saúde para o enfrentamento à covid-19. URL https://biblioteca.ibge.gov.br/visualizacao/livros/liv101717_notas_tecnicas.pdf.

IBGE. (2021). Postponement of the demographic census. URL https://www.ibge.gov.br/novo-portal-destaques/30569-adiamento-do-censo-demografico.html.

IBGE. (2022). Panorama do censo 2022. URL https://censo2022.ibge.gov.br/panorama/.

IBGE. (2023). Official note on the 2022 population census. URL https://www.ibge.gov.br/en/highlights/36135-official-note-on-the-2022-population-census.html?lang=en-GB.

IBGE. (2024). *Favelas e Comunidades Urbanas [Slums and Urban Communities]*. Rio de Janeiro: Brazilian Institute of Geography and Statistics. URL https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2102062.

Jean, N., Burke, M., Michael Xie, W., Davis, M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*(6301), 790–794.

Kuffer, M., Ali, I. M. M., Gummah, A., Da Silva, A., Mano, W. S., Kushieb, I., , … Abdallah, M., et al. (2023). Ideamapsudan: Geo-spatial modelling of urban poverty. In *2023 Joint Urban Remote Sensing Event (JURSE)* (pp. 1–4). IEEE.

Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from space—15 years of slum mapping using remote sensing. *Remote Sensing, 8*(6), 455.

Mahabir, R., Croitoru, A., Crooks, A. T., Agouris, P., & Stefanidis, A. (2018). A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities. *Urban Science, 2*(1), 8.

Mboga, N., Persello, C., Bergado, J. R., & Stein, A. (2017). Detection of informal settlements from vhr images using convolutional neural networks. *Remote Sensing, 9*(11), 1106.

Meyer, H., & Pebesma, E. (2022). Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications, 13*(1), 2208.

Neupane, B., Aryal, J., & Rajabifard, A. (2024). Cnns for remote extraction of urban features: A survey-driven benchmarking. *Expert Systems with Applications, 255*, Article 124751.

Oliveira, L. T., Kuffer, M., Schwarz, N., & Pedrassoli, J. C. (2023). Capturing deprived areas using unsupervised machine learning and open data: a case study in são paulo, brazil. *European Journal of Remote Sensing, 56*(1), 2214690.

Owusu, M., Engstrom, R., Thomson, D., Kuffer, M., & Mann, M. L. (2023). Mapping deprived urban areas using open geospatial data and machine learning in africa. *Urban Science, 7*(4), 116.

Owusu, M., Nair, A., Jafari, A., Thomson, D., Kuffer, M., & Engstrom, R. (2024). Towards a scalable and transferable approach to map deprived areas using sentinel-2 images and machine learning. *Computers, Environment and Urban Systems, 109*, Article 102075.

Patil, M. S. M. M. (2018). Interpolation techniques in image resampling. *Int. J. Eng. Technol, 7*(3.34), 567–570.

Raj, A., Mitra, A., & Sinha, M. (2024). Deep learning for slum mapping in remote sensing images: A meta-analysis and review. *arXiv preprint arXiv*, 2406.08031.

Roshan Joseph, V. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: An ASA Data Science Journal, 15*(4), 531–538. https://doi.org/10.1002/sam.11583

Russakovsky, O., Deng, J., Hao, S., Krause, J., Satheesh, S., Ma, S., , … Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision, 115*, 211–252.

Silva, J. P., & Rodrigues, J. F. (2024). Utilizing satellite imagery to predict socioeconomic indicators: a study conducted in brazil. *International Journal of Semantic Computing, 18*(3), 329–345.

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., , … Hagolle, O., et al. (2022). Cloud mask intercomparison exercise (cmix): An evaluation of cloud masking algorithms for landsat 8 and sentinel-2. *Remote Sensing of Environment, 274*, Article 112990.

Suel, E., Muller, E., Bennett, J. E., Blakely, T., Doyle, Y., Lynch, J., , … Nathvani, R., et al. (2023). Do poverty and wealth look the same the world over? a comparative study of 12 cities from five high-income countries using street images. *EPJ Data Science, 12*(1), 19.

Tan, M., & Le, Q. (2021). Efficientnetv2: Smaller models and faster training. In *International conference on machine learning* (pp. 10096–10106). PMLR.

UN-Habitat. (2003). *The challenge of slums: global report on human settlements, 2003*. London and Sterling, VA: Earthscan Publications Ltd. URL https://unhabitat.org/the-challenge-of-slums-global-report-on-human-settlements-2003.

UN-Habitat. (2023). Annual report 2022. URL https://unhabitat.org/annual-report-2022.

Union of Concerned Scientists. (2022). UCS satellite database. URL https://www.ucsusa.org/resources/satellite-database.

UN-Stats. (2023). The sustainable development goals report 2023: Goal 11. URL https://unstats.un.org/sdgs/report/2023/goal-11/.

Vanhuysse, S., Georganos, S., Kuffer, M., Grippa, T., Lennert, M., & Wolff, E. (2021). Gridded urban deprivation probability from open optical imagery and dual-pol sar data. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS* (pp. 2110–2113). IEEE.

Verma, D., Jana, A., & Ramamritham, K. (2019). Transfer learning approach to map urban slums using high and medium resolution satellite imagery. *Habitat International, 88*, Article 101981.

Wang, J., Fleischmann, M., Venerandi, A., Romice, O., Kuffer, M., & Porta, S. (2023). Eo + morphometrics: Understanding cities through urban morphology at large scale. *Landscape and Urban Planning, 233*, Article 104691.

Xie, M., Jean, N., Burke, M., Lobell, D., & Ermon, S. (2016). Transfer learning from deep features for remote sensing and poverty mapping. In *, 30. Proceedings of the AAAI conference on artificial intelligence*.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., … Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature Communications, 11*(1), 2583.