https://doi.org/10.5194/egusphere-2025-5987
Preprint. Discussion started: 22 January 2026

Semi-Supervised Segmentation for Mapping Urban Expansion and Hazard Exposure in Lima, Peru

Javier Jaimes[1], Luis Moya[1], Marta Vilela[2], Maria Santa-Maria[2], Sandra Santa-Cruz[1], and
Carlos Gonzales[3]

[1]GERDIS Research Group, Department of Engineering, Pontificia Universidad Católica del Perú, Peru
[2]CIAC, Department of Architecture, Pontificia Universidad Católica del Perú, Peru
[3]Centro Peruano Japonés de Investigaciones Sísmicas y Mitigación de Desastres, Facultad de Ingeniería Civil, Universidad Nacional de Ingeniería, Peru

Correspondence: Javier Jaimes (javier.jaimes@pucp.edu.pe)

Abstract. Urban expansion in rapidly growing cities increases exposure to natural hazards but remains difficult to monitor in regions with limited data. This challenge is amplified in places such as Metropolitan Lima, where global datasets of urban areas lack precision along complex and rapidly changing city boundaries. As a result, recent growth in informal and peripheral zones is not well defined. This study introduces a practical application of a semi-supervised mapping approach that combines satellite imagery with partially labeled information and targeted manual refinement to identify new built-up areas in Metropolitan Lima from 2016 to 2025. The method improves the detection of small and fragmented structures, including emerging informal settlements that global datasets frequently miss. Results show that Metropolitan Lima expanded by approximately 76 km^2 during the study period. A portion of this growth occurred in coastal zones exposed to tsunamis, in areas with medium to high landslide susceptibility, and on soil types where strong ground shaking is amplified during large earthquakes. These findings highlight the continued concentration of people and infrastructure in hazard-prone terrain.

1 Introduction

Urbanization is one of the main drivers of territorial transformation in the 21st century. Since 1985, urban areas worldwide have expanded rapidly, often without the spatial, social, or environmental planning required for sustainable growth (Huang and Xu, 2022). This process has altered land-use patterns, increased resource consumption, and heightened population exposure to natural and climate-related hazards (UN-Habitat, 2024). Today, more than one billion people live in informal settlements—nearly one-quarter of the global urban population—where access to basic services, secure land tenure, resilient infrastructure, and political representation remains limited. Global analyses indicate a strong correlation between the proportion of informal urban development and climate vulnerability ($\rho = 0.82$), and a negative correlation with climate preparedness ($\rho = -0.64$) (UN-Habitat, 2024). These findings highlight that the most excluded populations are simultaneously the most exposed and the least equipped to adapt or recover from environmental disasters.

On a global scale, urban growth between 1985 and 2015 has been largely inefficient in terms of land use. Recent analyses based on high-resolution satellite imagery and population datasets show that newly developed urban areas accommodate fewer

https://doi.org/10.5194/egusphere-2025-5987
Preprint. Discussion started: 22 January 2026

people per hectare than pre-existing urban zones (Taubenböck et al., 2025). This trend toward low-density, sprawling urbanization increases per capita infrastructure demand and contributes to higher emissions. Maintaining historical levels of land-use efficiency could have accommodated an additional 1.466 billion people without consuming more land. Using deep learning methods, researchers have classified global urban forms into 138 morphological patterns, revealing a dominance of dispersed and fragmented configurations associated with peripheral or suburban development, while only a small fraction of cities exhibit compact and mixed morphologies (Debray et al., 2025). As suitable land becomes scarcer, informal expansion increasingly takes place in environmentally fragile and high-risk areas—such as riverbanks, steep slopes, and floodplains—thereby amplifying the exposure of vulnerable populations to natural hazards. Consequently, balancing urban development with disaster-risk reduction has emerged as a complex and urgent planning challenge (Espinoza and Fort, 2020).

Remote sensing and machine learning have become indispensable tools for urban monitoring, enabling large-scale analysis of satellite and aerial imagery (Gram-Hansen et al., 2019; Tingzon et al., 2020; Assarkhaniki et al., 2021; Fallatah et al., 2022; Marconcini et al., 2021). In rapidly expanding cities, these technologies support detailed mapping of informal settlements and risk-prone zones (Li and Liu, 2020; Moya et al., 2022; Das et al., 2022; Mansour et al., 2023; Moya et al., 2024; Kesavan et al., 2024). Among these, deep learning (DL) has shown particular promise due to its strong capabilities in feature extraction and spatial pattern recognition (Yue et al., 2022; Bruno Adriano et al.; Wei et al., 2022; Geiß et al., 2020; Ding et al., 2022; Yokoya et al., 2024; Qiao et al., 2025). Despite these advances, most existing approaches rely on fully supervised learning, which demands extensive and highly accurate training data. However, in many regions—such as Peru—reliable datasets are limited or incomplete, constraining the effectiveness of machine learning-based urban footprint mapping (Kuffer et al., 2016). To overcome these limitations, recent studies have explored semi-supervised learning approaches (Wang et al., 2022; Guo et al.; Shi et al., 2022; Saha et al., 2021). Nevertheless, these methods have been analyzed mainly in developed countries, where urban patterns related to informal construction are largely absent. In parallel, considerable global efforts have been made to compile large-scale training datasets for land-use mapping. The World Settlement Footprint (WSF) products provide 10-m resolution built-up extents using optical and SAR data (German Aerospace Center, 2023), while the Global Human Settlement Layer offers built-up surfaces derived from Sentinel-2 imagery (Pesaresi et al., 2024). Furthermore, Microsoft's Global Building Footprints deliver building-level maps at continental scales. Such datasets are essential for understanding the global impact of urbanization on natural resources, exposure to hazards, and sustainable planning, enabling governments and organizations to make informed decisions. However, although these global-scale products are valuable for continental or national analyses, their accuracy decreases at finer spatial scales, such as the district level. For instance, Figure 1 illustrates an urbanized area on the outskirts of Lima, Peru, where a significant portion of informal settlements is missing from global-scale urban maps. This limitation underscores the need for localized, adaptable approaches capable of capturing fine-scale urban growth dynamics in data-scarce contexts.

In Lima, most informal urban growth occurs along the city's periphery, where state control and spatial planning are weak. These areas are frequently affected by land trafficking, with plots subdivided and sold without formal property rights (Dammert-Guardia et al., 2025). Housing is typically self-built and non-engineered, increasing vulnerability to earthquakes. Many settlements expand on steep slopes using *pircas* (dry-stone walls) to create flat surfaces, which remain unstable during seismic

https://doi.org/10.5194/egusphere-2025-5987
Preprint. Discussion started: 22 January 2026

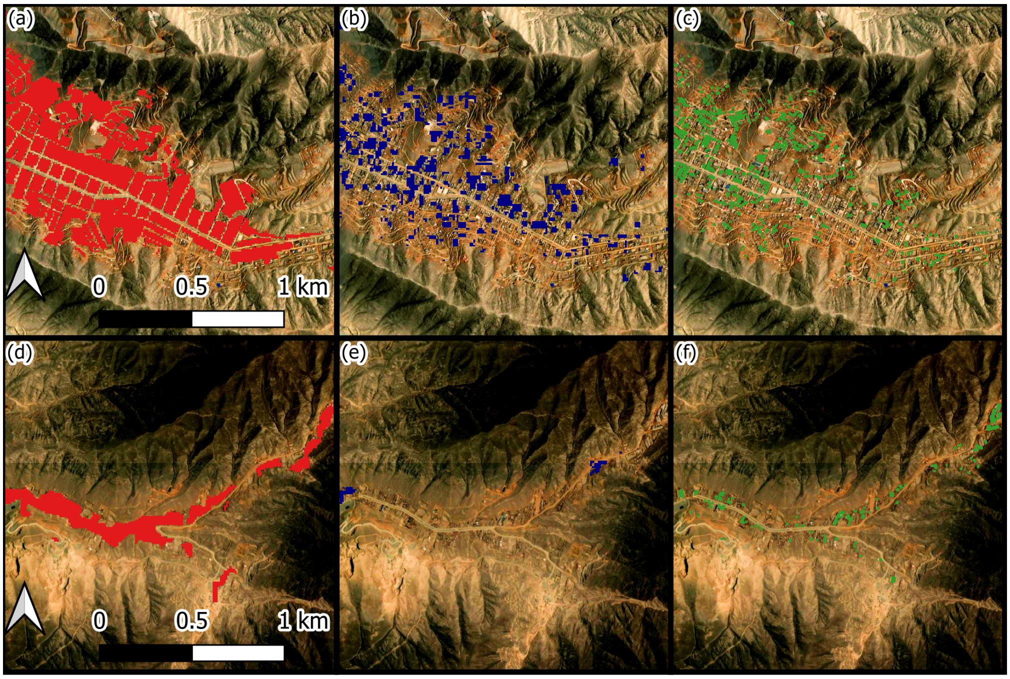

Figure 1. Built-up areas in the outskirts of Metropolitan Lima. Panels (a) and (d) show built-up areas identified by visual inspection (red). Panels (b) and (e) display built-up areas from the World Settlement Footprint (blue), and panels (c) and (f) show those identified by Microsoft's Global Building Footprint (green). Basemaps: Google, ©2024 MAXAR Technologies.

events (Santa-Cruz et al., 2021; Ita et al., 2023, 2024). Basic services such as water, sewage, and electricity are often absent in the early years, compounding exposure to health and environmental risks. In several peripheral districts, delinquency and land ownership disputes also emerge. Due to their fragmented growth and limited spatial data, these boundary zones are poorly represented in existing urban maps, yet they constitute critical areas for monitoring. Accurately delineating these zones is technically challenging because of their heterogeneous morphology and rapid change, but doing so is essential for effective urban planning and disaster-risk management. This priority aligns with one of Peru's four national development objectives for 2050—the promotion of a sustainable territory—as outlined in the *National Strategic Development Plan* (CEPLAN, 2023).

In this study, we present a practical methodology based on semi-supervised segmentation to improve urban footprint mapping along the boundaries of Metropolitan Lima, Peru. The approach integrates open-source datasets—including Sentinel-2 imagery, OpenStreetMap (OSM), and publicly available local government data—providing a scalable and transferable framework for monitoring urban expansion in data-scarce regions. The remainder of this paper is organized as follows. Section 2 describes the proposed approach, Section 3 presents the experimental results for Metropolitan Lima, Section 4 analyzes the growth of exposure against natural hazards, and Section 5 offers a critical interpretation of urban growth patterns, and the scope and limitations of the proposed approach, followed by the main conclusions in Section 6.

https://doi.org/10.5194/egusphere-2025-5987
Preprint. Discussion started: 22 January 2026

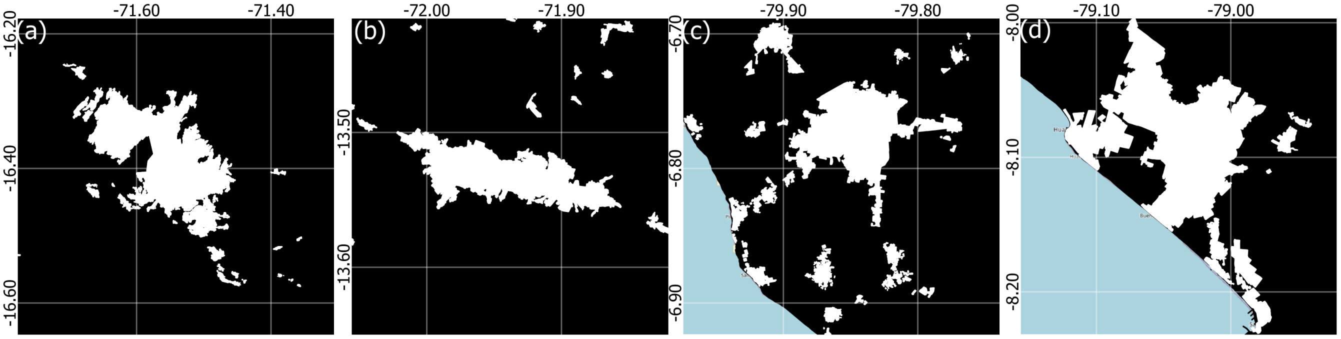

Figure 2. Examples of urban conglomeration maps for the cities of (a) Arequipa, (b) Cusco, (c) Lambayeque, and (d) La Libertad.

2 Proposed built-up mapping approach

Consider an image $X \in \mathbb{R}^{w \times h \times c}$, where w and h denote the spatial dimensions (width and height), and c indicates the number of spectral bands. Each pixel is defined by its position (i,j), and its spectral values are represented by the vector $X_{ij} = x \in \mathbb{R}^c$. The main objective is to determine whether X_{ij} belongs to the urban class ($Y_{ij} = 1$) or the non-urban class ($Y_{ij} = -1$). In this formulation, the category of a subset $I \subset \{(i,j)\}$ of pixels is known in advance. The proposed methodology to classify the whole image involves three main steps: (i) using a closed boundary to increase the labeled subset I, (ii) applying a semi-supervised scheme to calibrate a discriminant function, and (iii) classifying the entire set of pixels $\{(i,j)\}$.

2.1 Closed boundary

Let us define the set $I_p = \{(i,j) \mid Y_{ij} = 1, (i,j) \in I\}$ as the subset of pixels labeled as urban areas. Similarly, $I_n = \{(i,j) \mid Y_{ij} = -1, (i,j) \in I\}$ represents the subset of pixels labeled as non-urban areas. The initial labeled set I is typically derived from outdated built-up maps provided by local governments or open-access sources such as OpenStreetMap (QuickOSM, 2025). Consequently, most of its elements belong to I_p. The non-urban samples, on the other hand, generally correspond to vegetation or open-space areas within the city. Because these data sources do not fully capture recent urban expansion, the closed boundary step is introduced to roughly approximate the most recent urban extent prior to model calibration, and increase the size of I.

Given the scarcity of fine-grained urban maps, traditional urban studies in developing countries often rely on urban agglomeration maps (Vilela and Moschella, 2017; Peirano et al., 2023), which consist of closed polygons that encompass the built-up areas (Fig. 2). These polygons can be used to automatically increase the number of non-urban samples. Let $\Omega = \bigcup_i P_i$ denote the union of a set of planar polygon areas. If $(i,j) \notin \Omega$, then $Y_{ij} = -1$; that is, all pixels in the study area that are located outside the polygons will be labeled as non-urban to increase the training set. In the specific context of this study, Ω is constructed from the urban agglomeration maps. Users may directly employ these maps or adapt them—for instance, by applying buffers or manually editing the polygons—to better approximate the actual urban extent. This strategy enables the generation of additional reliable non-urban samples without requiring pixel-level manual labeling, thereby strengthening the subsequent semi-supervised calibration process.

2.2 Discriminant function and its calibration

Consider a classifier function $g : \mathbb{R}^{w\times h\times c} \to \mathbb{B}^{w\times h}$, where $\mathbb{B} = \{1,-1\}$, used to generate a binary image $\mathbf{B} \in \mathbb{B}^{w\times h}$. The objective is to calibrate g so that $\mathbf{B}$ accurately represents the spatial distribution of built-up areas. For this purpose, the method proposed by Kiryo et al. (2017) is adopted. To quantify the similarity between the predicted value B_{ij} from the discriminant function and the true label Y_{ij} at pixel (i,j), we introduce the following loss function $l : \mathbb{R}^2 \to \mathbb{R}$:

$$l(B_{ij}, Y_{ij}) = \frac{1}{1+e^{5Y_{ij}B_{ij}}} \tag{1}$$

If only training samples of built-up areas are used, the Non-Negative Positive-Unlabeled (NNPU) risk estimator proposed by Kiryo et al. (2017) can be applied:

$$\tilde{R}_{pu}^{\text{NNPU}}(g) = \pi_p \mathbb{E}_p[l(B_{ij},+1)] + \max\left\{0, \mathbb{E}_u[l(B_{ij},-1) - \pi_p \mathbb{E}_p[l(B_{ij},-1)]\right\}, \tag{2}$$

where π_p is the prior probability of the urban samples, and $\mathbb{E}_p[l(B_{ij},+1)]$ represents the expected loss when built-up areas are considered as class $+1$, and is approximated as described below.

$$\mathbb{E}_p[l(B_{ij},+1)] = \frac{1}{n_p} \sum_{(i,j)\in I_p} l(t_{ij},+1) \tag{3}$$

Likewise, $\mathbb{E}_p[l(B_{ij},-1)]$ denotes the expected loss when built-up areas are treated as class -1, and $\mathbb{E}_u[l(B_{ij},-1)]$ represents the expected loss when the unlabeled samples are treated as class -1:

$$\mathbb{E}_p[l(B_{ij},-1)] = \frac{1}{n_p} \sum_{(i,j)\in I_p} l(t_{ij},-1) \tag{4}$$

$$E_u[l(B_{ij},-1) = \frac{1}{n_{p^c}} \sum_{(i,j)\in I_p^c} l(t_{ij},-1) \tag{5}$$

where I_p^c denotes the complement of I_p. These expectations jointly define the empirical components of the NNPU risk function, allowing the model to balance the contributions of labeled and unlabeled samples during optimization. Note also that Equation (2) uses both labeled and unlabeled pixels. In the specific context of this study, training samples of non-built-up areas are also available (e.g., vegetation or pixels outside the closed boundary). Therefore, a second loss term for Non-Negative Negative-Unlabeled (NNNU) learning is introduced to take advantage of these additional negative samples.

$$\tilde{R}^{\mathrm{NNNU}}_{pu}(g) = \pi_n \mathbb{E}_n[l(B_{ij}, -1)] + \max\left\{0, \mathbb{E}_u[l(B_{ij}, +1) - \pi_n \mathbb{E}_n[l(B_{ij}, +1)]\right\}, \tag{6}$$

where $\pi_m = 1 - \pi_p$ is the prior probability of non-urban samples. This complementary formulation enhances model robustness by explicitly incorporating reliable non-urban information into the semi-supervised training process. A combined risk function is then proposed:

$$\tilde{R}(g) = \max\left\{\tilde{R}^{\mathrm{NNPU}}_{pu}(g), \tilde{R}^{\mathrm{NNNU}}_{nu}(g)\right\}. \tag{7}$$

This formulation allows the model to adaptively focus on the most informative supervision signal, balancing the contribution of both positive and negative classes. Finally, the optimal discriminant function is obtained by solving

$$\hat{g} = \underset{g \in \mathcal{G}}{\arg\min}\ \tilde{R}(g), \tag{8}$$

where $\mathcal{G}$ denotes the set of candidate discriminant functions. The optimal function $\hat{g}$ minimizes the combined risk defined in Equation (7), yielding a calibrated model capable of distinguishing built-up and non-built-up pixels under limited supervision.

3 Urban growth in Metropolitan Lima

We evaluate the performance of the proposed approach in quantifying urban growth in Metropolitan Lima. The city has undergone rapid and mostly unregulated expansion, altering its physical and socio-economic structure over the past century (Espinoza and Fort, 2020; Sobrino et al., 2023). Lima is a suitable case study due to its heterogeneous urban form, informal expansion, and limited spatial data.

3.1 Dataset

Sentinel-2 satellite imagery was used as input for the model. One image per year from 2016 to 2025 was selected (Table 1). For each year, the image corresponds to the period between January and March, when cloud coverage is minimal. All scenes belong to the 18LTM Sentinel-2 tile code, which covers most of Metropolitan Lima (Figure 3a). The blue, green, red, and near-infrared bands from the L2A Surface Reflectance product were used. Random image patches of size $512 \times 512 \times 4$ (width, height, bands) were extracted within the study area to built I.

Regarding the samples of built-up areas (I_p), the main source was the urban block dataset from the 2017 National Census (National Institute of Statistics and Informatics (INEI), 2017). This information was complemented with building footprints from OpenStreetMap (QuickOSM, 2025) and urban growth clusters identified by Moya et al. (2024). That study used radar satellite imagery to detect 25 recently urbanized areas showing significant backscatter change. Its results were constrained by topography and speckle noise, but the dataset remains valuable because it documents newly formed informal clusters across

Table 1. Sentinel-2 acquisition dates for Metropolitan Lima

Acquisition date	Acquisition date	Acquisition date
2016-03-27	2019-04-11	2022-04-30
2017-02-20	2020-04-20	2023-01-25
2018-04-11	2021-04-20	2024-02-14
		2025-01-14

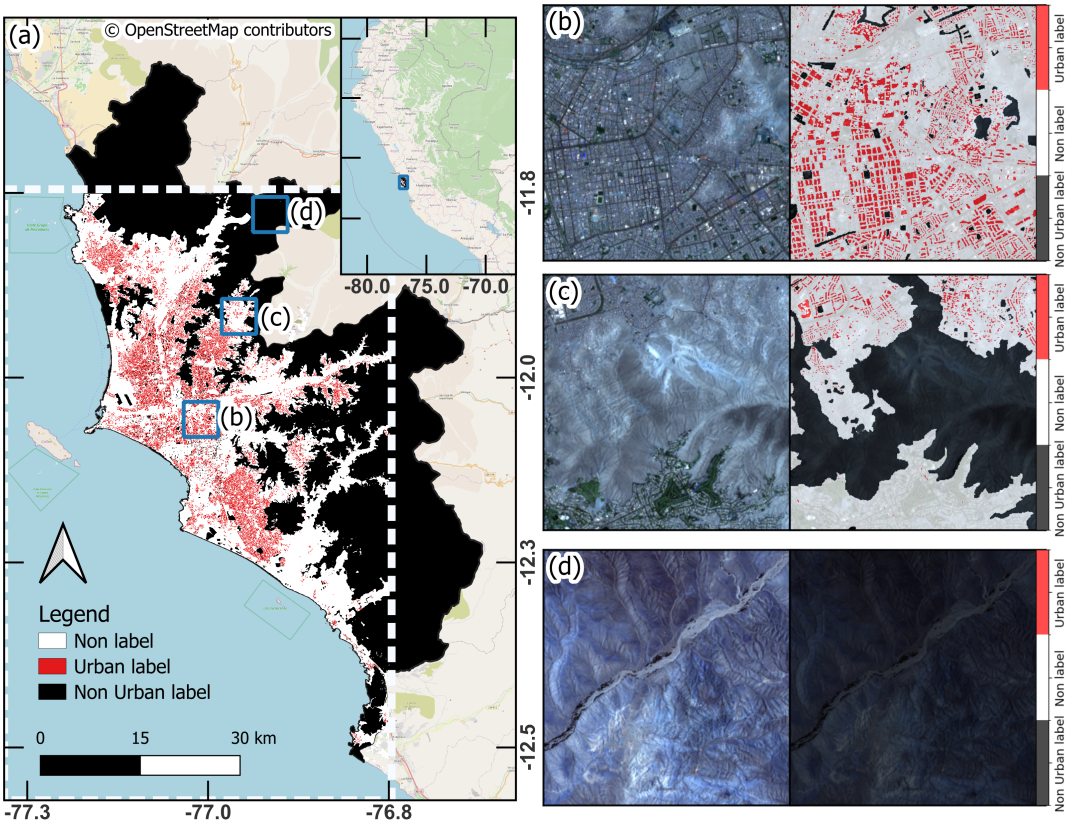

Figure 3. (a) Metropolitan Lima. The inset shows the location of the study area within Peru. Red tones indicate built-up areas, white tones represent unlabeled pixels within the closed boundary polygons, and black pixels correspond to non–built-up areas within Lima Metropolitan. The white dashed rectangle delineates the extent of the Sentinel-2 imagery used in the study. Closer views of a consolidated urban area, a peripheral area, and a non-urban area are shown in panels (b), (c), and (d), respectively. Basemaps (b,c & d): Copernicus Sentinel data [2024]

Lima, increasing the diversity of I_p samples. The set I_n was derived from the urban conglomerate of Metropolitan Lima, which was manually delineated following standard procedures for urban planning studies, and serves here as the closed boundary Ω. In addition, green areas from OpenStreetMap were included to define I_n. Figures 3b–3d illustrate three sample locations: within the city center, along the urban boundary, and outside the closed boundary.

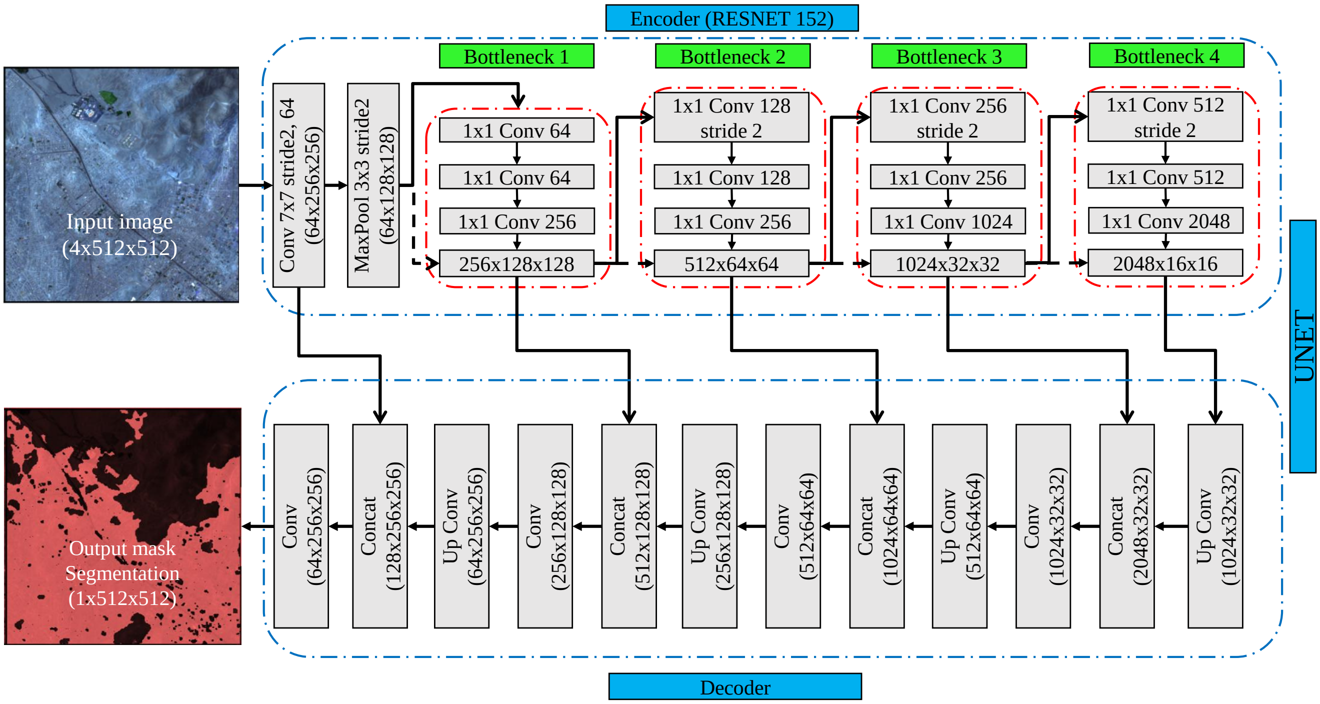

Figure 4. U-Net architecture with a ResNet-152 encoder. The input image is progressively encoded into high-level semantic features through residual blocks and then decoded to generate a detailed segmentation mask. Dashed arrows represent skip connections that merge encoder and decoder feature maps to recover spatial detail. Red dashed polygons highlight the bottleneck region, and blue dashed polygons delineate the encoder and decoder components.

3.2 Classifier model

We employed a U-Net segmentation model (Ronneberger et al., 2015) with a ResNet-152 backbone (He et al., 2016) to generate pixel-level classification maps. U-Net is a convolutional neural network designed for semantic segmentation. In our configuration, ResNet-152 acts as the encoder, extracting hierarchical image features through 152 layers organized into Bottleneck blocks. These residual blocks compress and expand feature dimensions to maintain computational efficiency while preserving spatial information. Residual connections link the blocks, preventing accuracy loss in deep networks. Figure 4 illustrates the encoder–decoder process used to produce the segmentation mask.

For model calibration, only images recorded after 2021 were used to include the built-up areas detected by Moya et al. (2024). Thus, the images from 2021 and 2024 were selected to extract 12,000 random samples (Figure 3b–3d). Using two acquisition dates introduces temporal variability in the samples, which helps the model generalize better to different urban conditions. The Adam optimization algorithm was used to update the model weights, with a learning rate of 5×10^{-5}, batch size of 32, and 200 epochs. From the study area and available datasets, the cardinalities of I_p and I_n are approximately 2.4 million and 25.4

https://doi.org/10.5194/egusphere-2025-5987
Preprint. Discussion started: 22 January 2026

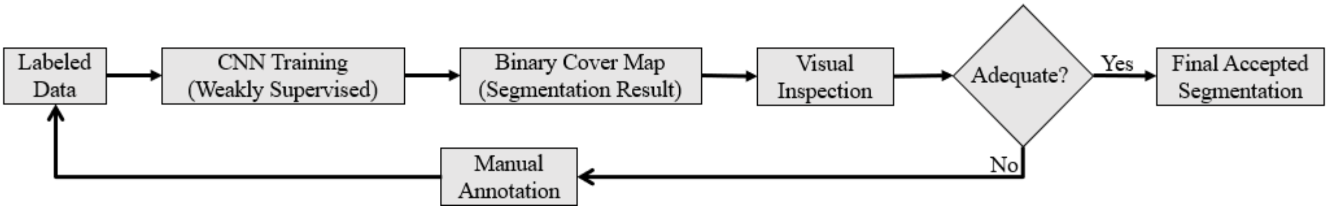

Figure 5. Iterative human-in-the-loop workflow for weakly supervised segmentation, adopted from Yokoya et al. (2024). Labeled data and manual annotations are progressively refined through CNN training, mapping, and targeted review.

million pixels, respectively. The entire study area contains about 40 million pixels. Therefore, the prior probability π_p used in the loss functions (Eqs. 2 and 6) must lie within the following range:

$$\begin{aligned} \frac{2.4}{40} &\leq \pi_p \leq \frac{40-25.4}{40} \\ 0.06 &\leq \pi_p \leq 0.37 \end{aligned} \tag{9}$$

These bounds reflect two limiting interpretations of the unlabeled pixels: if all are considered non-urban, urban pixels account for only 6% of the whole study area; conversely, if all are considered urban, the proportion increases to 37%. Accordingly, π_p was adjusted empirically within the range defined in Equation (9). To improve segmentation accuracy, additional labeled samples were progressively incorporated following the workflow proposed by Yokoya et al. (2024) (Figure 5). After each training cycle, segmentation outputs were evaluated both quantitatively and through visual inspection. Areas showing inconsistent or unsatisfactory results were manually corrected. When required, new polygons were delineated with adequate spatial precision to enhance local consistency. The revised samples were integrated into the training dataset, and the model was retrained iteratively. Approximately 600 polygons were annotated through this human-in-the-loop process, adding about 490,000 labeled pixels (0.5% of the total training data). After the iterative calibration, 70% of the samples were used for training and 30% for validation. Figure 6 shows the training and validation losses, both exhibiting stable convergence toward the final epochs.

To assess the performance of our model, we compared our results with the World Settlement Footprint (WSF) 2019 (German Aerospace Center, 2023). The WSF is a global 10 m resolution binary mask delineating human settlements. It was generated by the German Aerospace Center (DLR) using multitemporal Sentinel-1 and Sentinel-2 imagery from 2019 through a binary Random Forest classification. From the Lima region, 20 validation samples were randomly selected: (i) five from the city center, where WSF shows high accuracy (*consolidated* samples); (ii) ten around the closed urban boundary with high urban coverage (*peripheral* samples); (iii) three around the boundary with limited urban presence (*remote* samples); and (iv) two from the ocean. Peripheral and remote samples correspond to areas where informal urban expansion is most active. For these samples, the built-up areas were delineated manually by visual inspection of high-resolution satellite imagery (examples in Figure 7). It is important to note that WSF underwent post-processing to remove roads (German Aerospace Center, 2023),

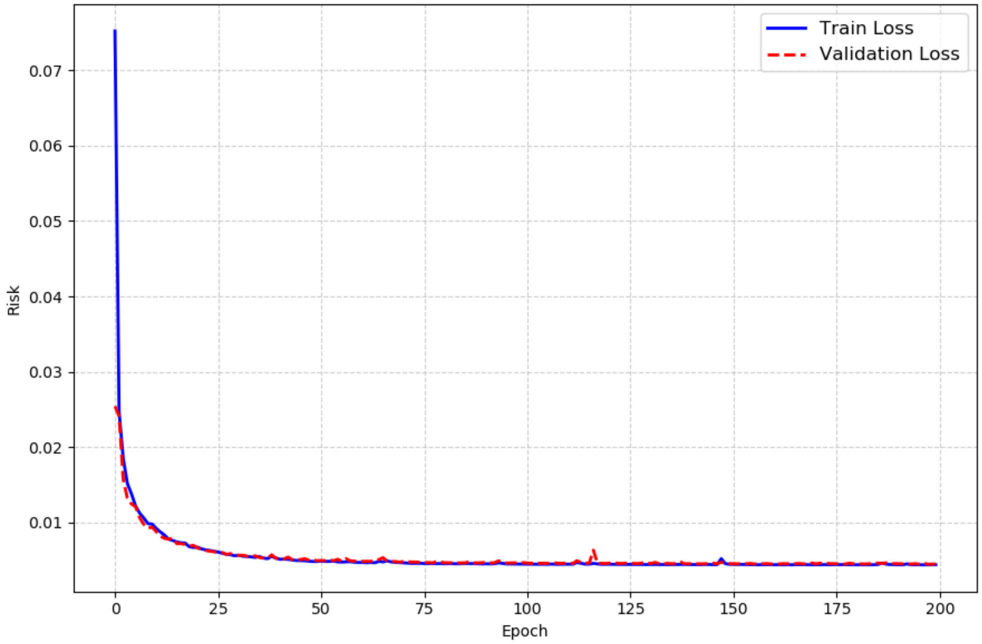

Figure 6. Variation of the risk function (Equation 7) across epochs during the calibration process, using a learning rate of 5×10^{-5} and $\pi_p = 0.35$. The blue line denotes the risk computed from the training set, and the red line represents the risk computed from the validation set.

whereas our model considers roads as built-up areas. For a fair comparison, road pixels were excluded from the validation process. Model performance and WSF accuracy were evaluated using *precision*, *recall*, and the *F1* score metrics.

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{10}$$

185 $$\text{Recall} = \frac{TP}{TP + FN}, \tag{11}$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{12}$$

where TP, FP, and FN denote the numbers of *true positives*, *false positives*, and *false negatives*, respectively. For a given class i (urban or non-urban), *precision* is the fraction of pixels classified as class i that are correctly identified. Conversely,
190 *recall* represents the fraction of pixels belonging to class i that are correctly detected.

The accuracy scores of our model and the WSF urban map are summarized in Table 2. In consolidated urban areas, when considering average scores, the WSF outperforms our model in recall by 2% and in F1 by 1%, but underperforms in precision

https://doi.org/10.5194/egusphere-2025-5987
Preprint. Discussion started: 22 January 2026

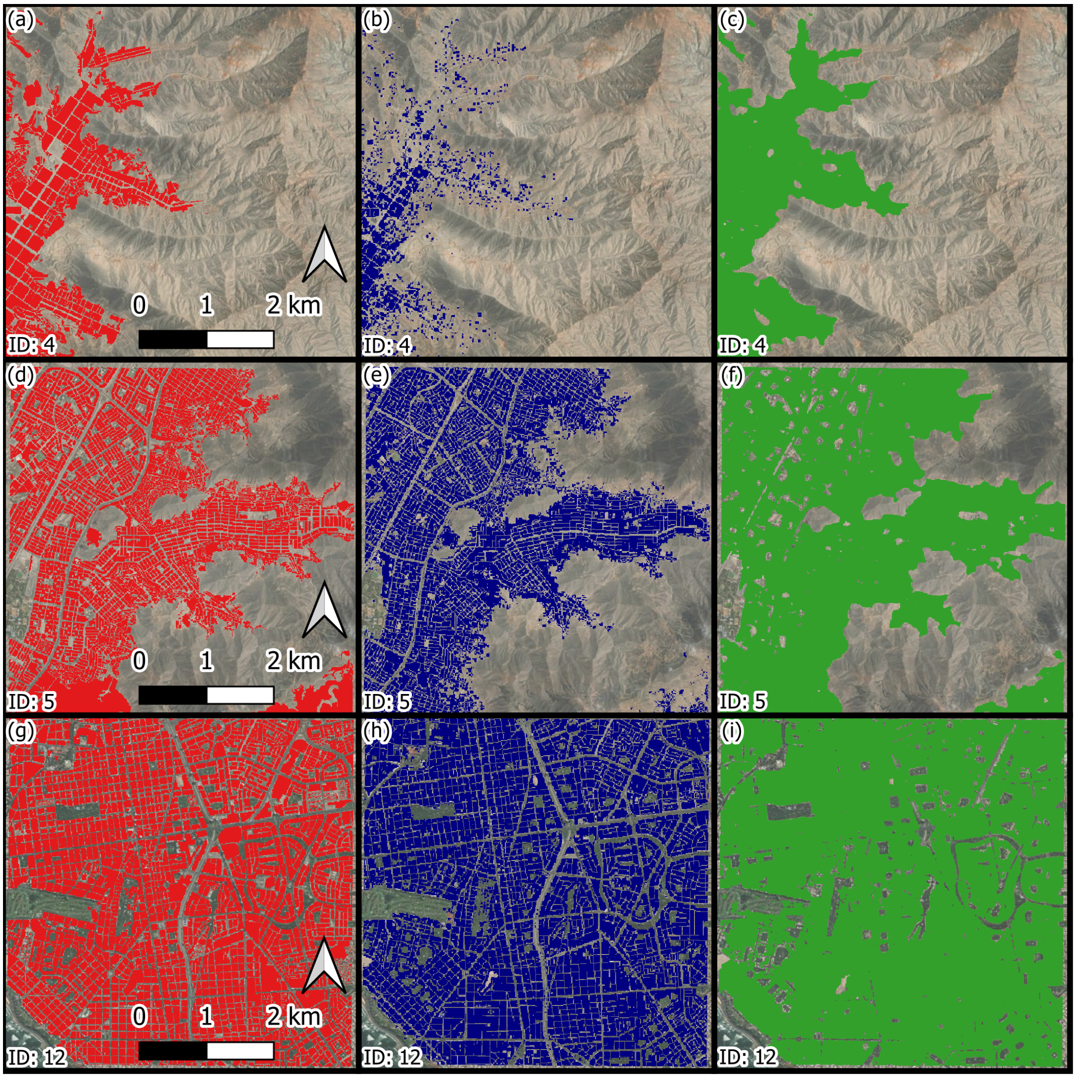

Figure 7. Comparison of built-up areas obtained from the proposed methodology, the WSF product, and the reference built-up map. Built-up areas detected by our method are shown in green (right), those from the WSF are shown in blue (middle), and the reference built-up areas are shown in red (left). Basemaps: Basemaps: Google, ©2024 MAXAR Technologies.

by 2%. This result is expected since WSF effectively represents consolidated urban zones worldwide. In peripheral areas, our model outperforms WSF in recall by 12% and in F1 by 2%, while underperforming in precision by 4%. In remote areas, the 195 improvement is more significant: our model exceeds WSF in recall by 37% and in F1 by 12%, with a 4% decrease in precision. These results indicate that the proposed model is particularly effective in detecting emerging urban growth zones, which aligns with the main objective of this study. It is also noteworthy that the largest performance gap favoring WSF (17%) occurs in

Table 2. Validation results comparing WSF and CNN-based classification across different areas in Lima.

Type Area	Model	Precision			Recall			F1-Score		
		U	NU	Av.	U	NU	Av.	U	NU	Av.
Consolidated	WSF	**0.8483**	0.7833	0.8158	0.8174	**0.8187**	**0.8180**	0.8326	**0.8006**	**0.8166**
	Ours	0.7680	**0.9066**	**0.8373**	**0.9464**	0.6454	0.7959	**0.8479**	0.7540	0.8010
Peripheral	WSF	**0.7295**	0.9180	**0.8238**	0.6061	**0.9515**	0.7788	0.6621	**0.9345**	0.7983
	Ours	0.5810	**0.9881**	0.7846	**0.9526**	0.8518	**0.9022**	**0.7218**	0.9149	**0.8183**
Remote	WSF	**0.3989**	0.9928	**0.6958**	0.1605	**0.9979**	0.5792	0.2289	**0.9953**	0.6121
	Ours	0.3132	**0.9994**	0.6563	**0.9268**	0.9824	**0.9546**	**0.4682**	0.9908	**0.7295**
Ocean	WSF	-	1	-	-	1	-	-	1	-
	Ours	-	1	-	-	1	-	-	1	-

consolidated areas for non-urban recall, whereas the largest improvement from our model (77%) appears in remote areas for urban recall.

3.3 Results

Once the model was calibrated, built-up areas were detected for each of the selected Sentinel-2 images (Section 3.1). Figure 8 shows the built-up extent by 2025, where blue tones represent areas existing in 2016 and red tones denote growth between 2016 and 2025. The insets provide detailed views of the urban boundaries in 2021 and 2024, when high-resolution imagery is available for this study, showing that the model successfully delineates built-up limits and their temporal evolution. Overall, Lima's urban footprint expanded by approximately 76.45 km^2 during this period (Figure 9). Annual expansion peaked between 2016 and 2017, reaching 18.67 km^2, followed by a gradual slowdown until 2021. A renewed acceleration was observed from 2022 onward, indicating the continued dynamism of urban expansion. The spatial distribution of growth shows a clear outward pattern concentrated in peripheral districts, consistent with previous reports of Lima's horizontal expansion. This trend highlights the limited containment capacity of existing urban boundaries and the persistent demand for new residential areas at the city's margins.

At the district level, the spatial distribution of urban expansion reveals pronounced heterogeneity across Lima's metropolitan area (Figure 10). During the study period, Carabayllo recorded the highest cumulative increase in urbanized area (approximately 11.4 km^2), followed by Lurín (7.5 km^2), Pachacamac (7.4 km^2), Punta Hermosa (5.0 km^2), Cieneguilla (5.0 km^2), Villa María del Triunfo (4.8 km^2), Ventanilla (4.4 km^2), Lurigancho (4.0 km^2), San Juan de Lurigancho (4.0 km^2), and Puente Piedra (3.1 km^2). These results highlight that the most substantial cumulative growth between 2016 and 2025 occurred in peripheral districts.

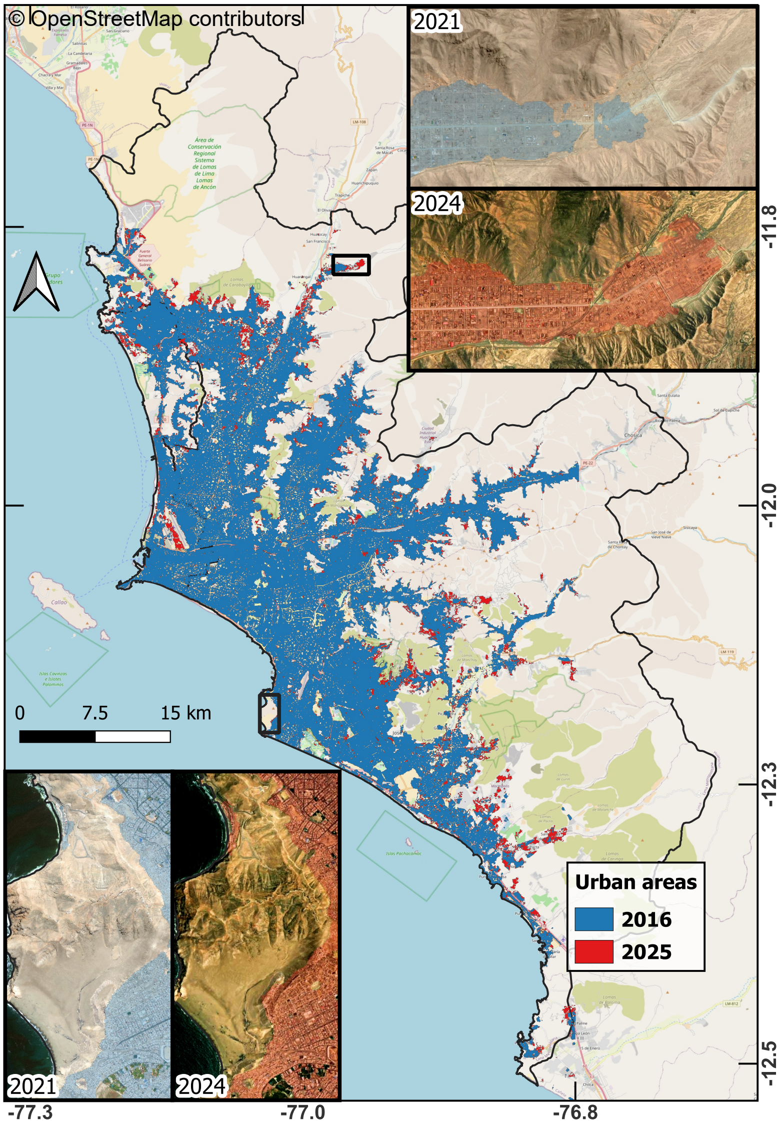

Figure 8. Spatial distribution of urban growth in Metropolitan Lima from 2016 to 2025. Blue tones indicate built-up areas present in 2016, and red tones depict the growth that occurred between 2016 and 2025. Insets show closer views of the predictions for 2021 and 2024. Basemaps of insets: Basemaps: Google, ©2024 MAXAR Technologies.

4 Hazard exposure and socioeconomic growth type

To quantify the growth in exposure to natural hazards, urban expansion was analyzed in relation to flood susceptibility, landslide susceptibility, seismic microzonation, and tsunami inundation zones. The tsunami inundation map represents the expected flooding extent from a potential Mw 9.0 earthquake off the coast of Lima, prepared by the Directorate of Hydrography and

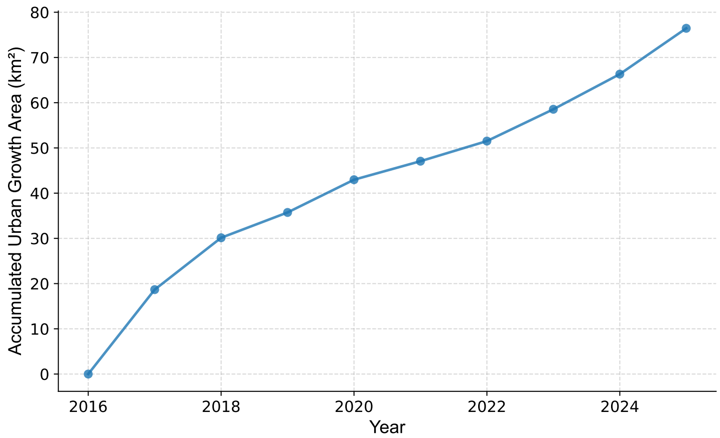

Figure 9. Cumulative increase in urbanized area between 2016 and 2025 in Metropolitan Lima.

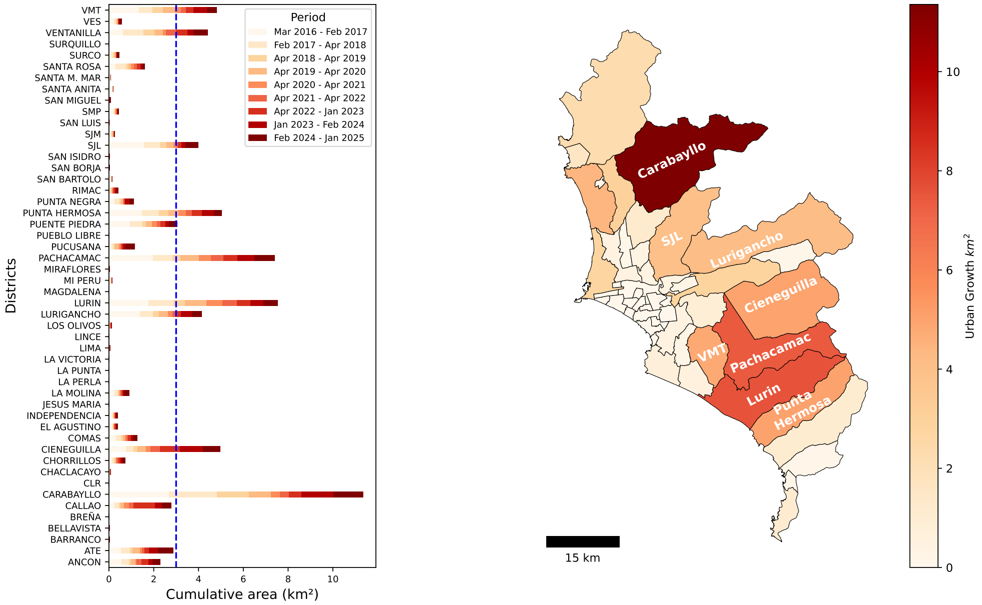

Figure 10. (a) Increase in urbanized area between 2016 and 2025 by district. Lighter shades indicate earlier phases of expansion, while darker tones represent more recent growth. (b) Spatial distribution of districts in Metropolitan Lima. Shade intensity denotes the total urban growth between 2016 and 2025. SJL: San Juan de Lurigancho; SJM: San Juan de Miraflores; SMP: San Martin de Porres; VES: Villa El Salvador; VMT: Villa Maria del Triunfo; CML: Carmen de la Legua-Reynoso.

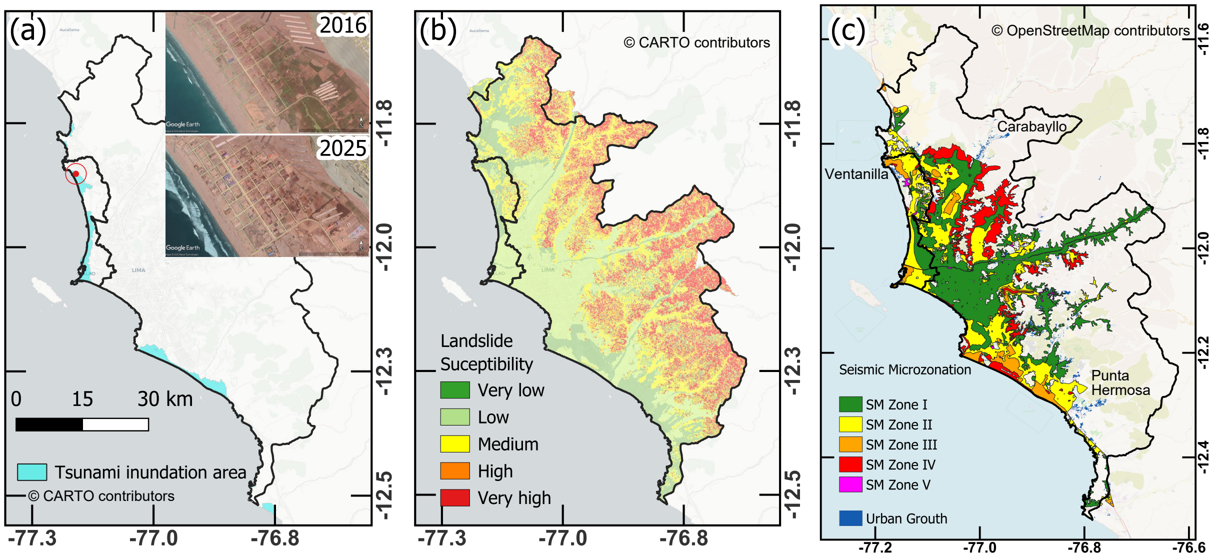

Figure 11. (a) Tsunami inundation map for Metropolitan Lima (Directorate of Hydrography and Navigation – Peruvian Navy, 2025). The inset shows a closer view of urban growth in the bay area of the District of Ventanilla, where red tones indicate zones subject to tsunami inundation. Basemaps of insets: Google, ©2025 MAXAR Technologies.(b) Landslide susceptibility classification for Metropolitan Lima (Villacorta Chambi et al., 2015). (c) Seismic microzonation map (Gonzales et al., 2023).

Navigation of Peru (DHN, 2014) (Figure 11a). The landslide susceptibility map proposed by Villacorta Chambi et al. (2015) classifies the territory into five categories—very low, low, medium, high, and very high—according to the probability of occurrence (Figure 11b). The map indicates that the eastern sectors of the city are highly vulnerable to mass movements such as rockfalls and mudslides (*huaycos*). Approximately 35% of Lima's territory is classified as having high or very high susceptibility to landslides.

The seismic microzonation (SM) map characterizes the soil deposits of Metropolitan Lima based on their mechanical and dynamic properties (Calderon et al., 2014; Gonzales et al., 2023; Calderon et al., 2025) (Figure 11c). It includes five zones: SM Zone I corresponds to rock formations with different degrees of fracturing (excluding steep slopes), dense to very dense alluvial gravels, and sedimentary deposits with gravelly layers at shallow depths; SM Zone II comprises medium-dense to dense sands and medium-to-stiff silty–clayey soils; SM Zone III includes coastal deposits of deep loose to medium-dense sands and soft to medium silts and clays; SM Zone IV is characterized by thick aeolian sands and loose marine deposits prone to liquefaction, as well as steep rocky slopes due to their high geological risk; and SM Zone V includes landfills and waste-material deposits resulting from anthropogenic activities throughout Metropolitan Lima. This classification also incorporates vibrational parameters obtained from geophysical surveys, such as the fundamental site period and the average shear-wave velocity in the upper 30 m (Vs30). Accordingly, SM Zone I is characterized by short fundamental periods (shorter than 0.2 s) or flat H/V spectral ratios and the highest Vs30 values (on the order of 550–650 m/s), whereas SM Zone IV exhibits the longest fundamental periods recorded in Metropolitan Lima (slightly exceeding 1 s) and the lowest Vs30 values in the city (on the order of 250 m/s). These contrasts clearly demonstrate the strong spatial variability of seismic amplification patterns driven by the diverse geological conditions across Metropolitan Lima.

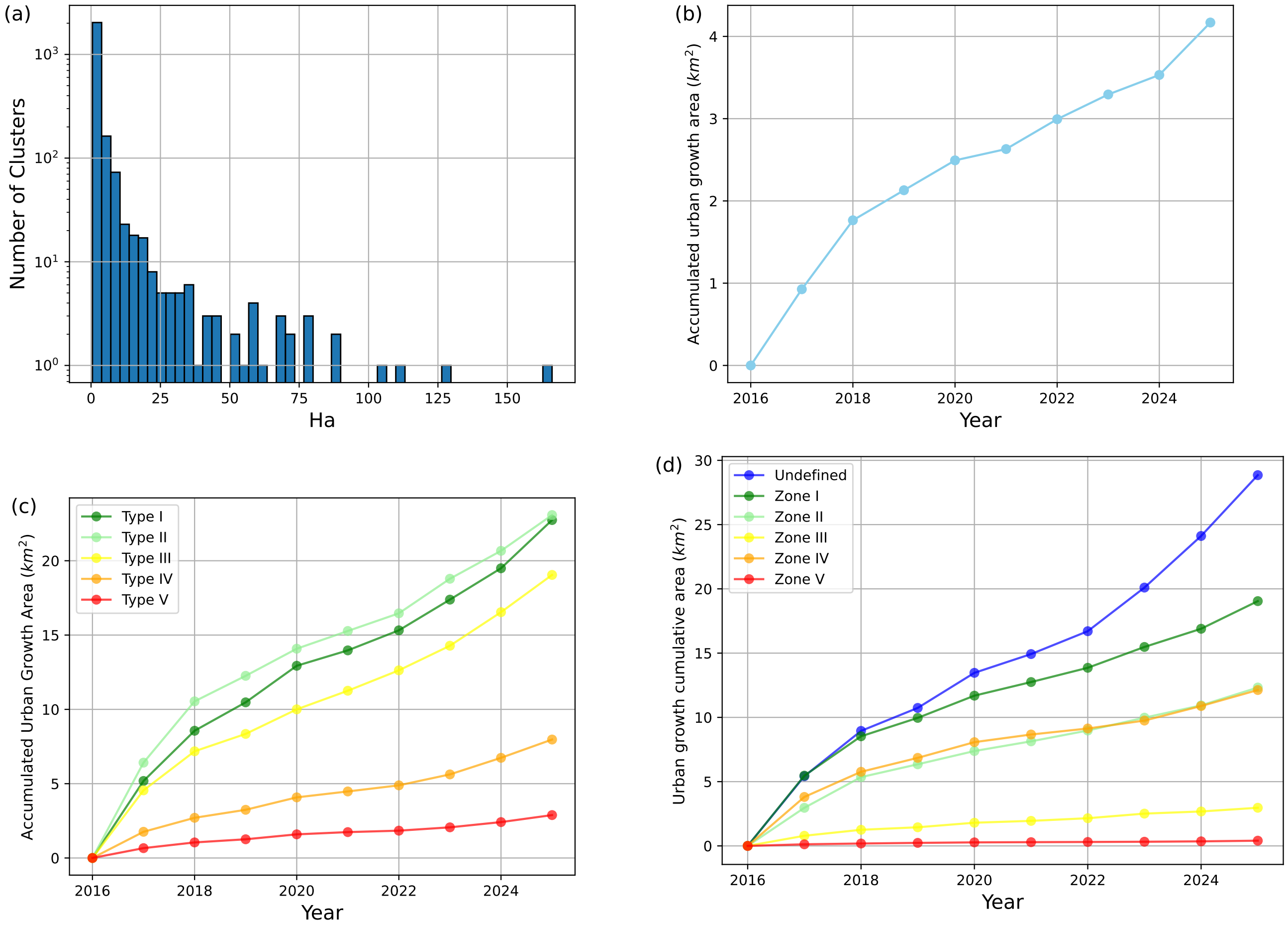

Figure 12. (a) Distribution of urban growth between 2016 and 2025 by clusters. (b) Cumulative urban growth by year in tsunami-prone coastal areas. (c) Cumulative urban growth by year across different landslide susceptibility levels. (d) Cumulative urban growth by year across different seismic microzonation classes.

Regarding tsunami exposure, approximately 4 km^2 of new urban area was developed within coastal zones exposed to tsunamis between 2016 and 2025 (Figure 12b). The District of Ventanilla experienced the largest expansion within tsunami-prone areas (inset in Figure 11a), particularly in its northern sector, where urban growth has intensified in recent years. The second largest expansion occurred in the District of Callao, where new land has been gained from the sea through natural sediment deposition and anthropogenic reclamation. These reclaimed areas are now occupied by residential and industrial developments. Urban expansion within landslide-susceptible zones was even greater: nearly 30 km^2 of new development overlaps areas with medium (Type III) to very high (Type V) susceptibility (Figure 12c), while zones classified as low or very low exceed 40 km^2. Notably, more than 36.5% of recent urban growth lacks detailed seismic microzonation coverage, underscoring a major gap in the current framework for integrated urban planning and seismic risk assessment (Figure 12d). In addition, approximately 12.5 km^2 of new construction occurred on soils classified as SM Zone IV and SM Zone V, where strong ground-motion amplification, liquefaction potential, and/or geotechnical instability are expected.

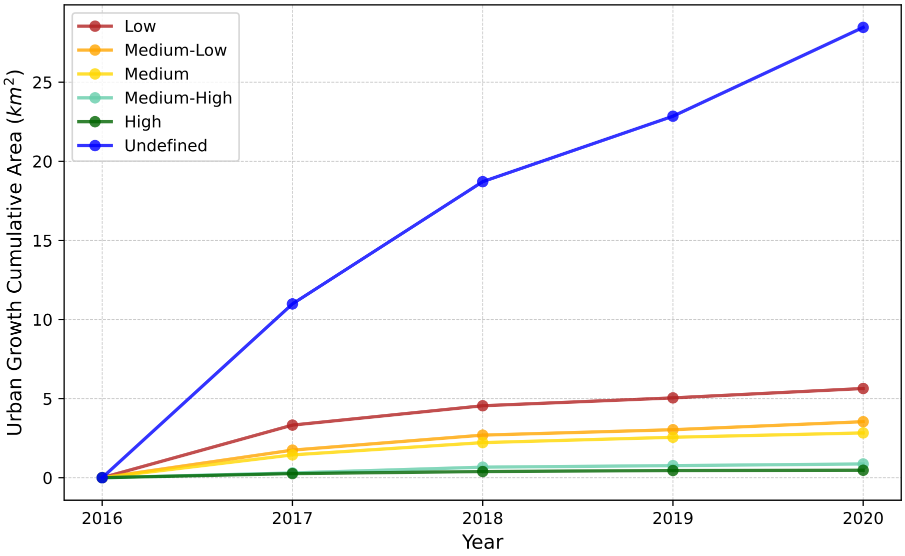

Figure 13. Cumulative urban growth by socioeconomic type between 2016 and 2025. A substantial portion remains unclassified due to limited population data in informal settlements.

In 2020, the National Institute of Statistics and Informatics (INEI) released a fine-scale socioeconomic stratification of Lima and Callao at the block level. Household per capita income was estimated from the 2017 Population and Housing Census, complemented by household surveys and auxiliary data sources (National Institute of Statistics and Informatics (INEI), 2020). Additional variables—such as distance to schools, health centers, and district capitals—were included to strengthen the spatial estimations. The study classifies urban households into five income groups: high, medium-high, medium, medium-low, and low. Our results show that about 75% of newly urbanized areas lack a defined socioeconomic classification, reflecting limited demographic coverage in informal and peripheral settlements (Figure 13). Among classified areas, most of the expansion occurred in low (5.6 km^2) and medium-low (3.5 km^2) income zones, corresponding to the lowest socioeconomic levels. This pattern highlights an increasing concentration of urban expansion in areas associated with higher levels of social vulnerability.

https://doi.org/10.5194/egusphere-2025-5987
Preprint. Discussion started: 22 January 2026

5 Discussion

This study shows that semi-supervised learning offers a practical and scalable solution for mapping the urban footprint in data-scarce environments. The NNPU-based model accurately delineated built-up areas from incomplete geocoded data, helping to close the gap between global-scale products and the demand for locally detailed urban information. By combining limited reference data with Sentinel-2 imagery, the model captured Lima's heterogeneous urban fabric, including informal settlements often omitted from global datasets such as the World Settlement Footprint (WSF).

This study builds upon the work presented in Moya et al. (2024), where large clusters of recent settlements were identified through changes in the backscatter of SAR satellite imagery. Because SAR systems are side-looking, their images are affected by geometric distortions and limited visibility in areas of complex topography—such as the steep hills and narrow valleys that dominate Lima's periphery. Consequently, not all settlements were captured in those SAR-based analyses. Here, we map the entire spectrum of new urban areas to produce more comprehensive estimates of the exposed population, including small clusters that could not be detected previously due to speckle noise in SAR images. Spatially, Lima's urban expansion between 2016 and 2025 ($\approx$76 km^2) reflects persistent outward growth concentrated in peripheral districts such as Carabayllo, Lurín, and Pachacamac. These areas illustrate the common pattern of peri-urban sprawl in Latin American cities, where weak land regulation and speculative land trafficking drive rapid settlement beyond formal planning control. Our total estimate of 8.3 km^2 per year contrasts with earlier studies. Espinoza and Fort (2020) reported an increase of about 200 km^2 between 2000 and 2018 ($\approx$11 km^2/year), based on manually delineated urban agglomeration polygons that encompass built-up areas. Similarly, the National Center for Strategic Planning (CEPLAN) (2023) estimated a growth of 100 km^2 between 2014 and 2018 ($\approx$25 km^2/year), notably higher than the previous estimate. A closer comparison reveals substantial differences in the delineation of urban agglomerations, highlighting a key source of uncertainty when different criteria are applied to define urban boundaries. Remote sensing–based approaches, by contrast, are less affected by such subjective interpretations.

Our findings have important implications for more efficient and sustainable land planning and management. Official datasets are often outdated or incomplete, partly due to limited road access, steep slopes, and insufficient safety conditions. In contrast, our results provide accurate, up-to-date information on built-up areas, enabling a clearer understanding of socio-spatial dynamics within the urban fabric. They can also support efforts to mitigate land trafficking. Field surveys indicate that early land occupations frequently consist of temporary and uninhabited dwellings (Moya et al., 2024). Over time, as these plots are informally commercialized, new occupants settle and later request formal land tenure. The proposed model can help identify such early-stage occupations, allowing authorities to prevent illegal land trafficking and reduce the associated violence and criminal activity (Vera, 2018; Pimentel-Sanchez, 2021; Dammert-Guardia et al., 2025). Additionally, implementing this methodology would enable more timely detection of informal expansion and facilitate early warnings for communities exposed to natural hazards.

The integration of hazard information revealed that nearly 30 km^2 of new urban development occurred within areas of medium to high landslide susceptibility. In contrast, the largest share—about 45 km^2—is located in zones classified as very low or low susceptibility. At first glance, these results appear to contradict those of Moya et al. (2024), who reported that informal urban growth is mainly concentrated in high and very high susceptibility areas. This difference arises from the analytical scale: our study is pixel based, whereas Moya et al. (2024) conducted a cluster-based analysis. In their approach, if any portion of a cluster overlapped a high or very high susceptibility zone, the entire cluster was classified accordingly. Moreover, Moya et al. (2024) could not detect small clusters, while our method captures these smaller settlements, which are numerous across the study area (Figure 12a).

Urban exposure to tsunami hazards increased by approximately 4 km^2 during the study period. These coastal areas require special consideration in urban development and land-use planning. Tsunami evacuation routes must be clearly communicated to residents, and the urban fabric should ensure that all neighborhoods have direct access to safe evacuation zones. Communities located near the shoreline should also be aware that, following a major earthquake, evacuation should begin immediately without waiting for an official warning. Past earthquakes have shown that reclaimed land is highly susceptible to liquefaction (Kramer, 1996; Konagai et al., 2013); furthermore, previous reports indicate that the District of Callao experienced liquefaction during past seismic events (Alva-Hurtado and Ortiz-Salas, 2020). Therefore, detailed geotechnical assessments are needed for the recently developed areas in the District of Callao, as these conditions will directly influence the foundation design of new constructions.

It is important to emphasize that the seismic microzonation map considered in this study (Calderon et al., 2025) builds upon the original soil seismic classification developed in 2004 and subsequently updated between 2011 and 2021, at an average rate of approximately five districts per year. Consequently, several of the districts evaluated at that time did not yet include the urban expansion areas documented in this study. The largest currently urbanized zones that still lack seismic classification are located in the districts of Ventanilla, Carabayllo, and Punta Hermosa. In the case of Ventanilla, an urban expansion area of approximately 1.09 km^2 is observed in the eastern sector of the district, near the coastline, which formerly constituted part of a protected wetland zone. At present, this area is being used for residential and industrial purposes, severely compromising the conservation of local flora and fauna. The uncontrolled anthropogenic fill used for urbanization, combined with the presence of a shallow groundwater table, may significantly increase the susceptibility to soil liquefaction and seismic amplification. In the northern sector of Carabayllo, an expansion area of approximately 3.5 km^2 reveals multiple geotechnical concerns. The largest portion of this growth is located on former agricultural land near the Chillón River, where the soil profile, although predominantly of alluvial origin, is associated with a high groundwater level. Additionally, urbanization has expanded into zones adjacent to hillsides that were previously exploited as construction-material quarries. After quarry closure, anthropogenic filling, land subdivision, and land sales for residential use were carried out. This practice is particularly hazardous, as buildings in these areas are founded on poorly compacted and heterogeneous fill materials with unknown dynamic behavior under seismic loading. A particularly critical case is the urbanization of the Río Seco ravine, which becomes active during intense summer rainfall associated with El Niño phenomena, generating debris flows from highlands. The most recent activation of this ravine occurred in 2023 during Cyclone Yaku. Finally, an urban expansion area of approximately 4.56 km^2 has been identified in the district of Punta Hermosa, located in southern Metropolitan Lima. This expansion is concentrated near a ravine that, similar to the Río Seco case, becomes active during periods of intense rainfall. The ongoing urbanization process has altered the natural transport routes of sediments and debris, increasing susceptibility to flooding and mass movement processes. The cases

https://doi.org/10.5194/egusphere-2025-5987
Preprint. Discussion started: 22 January 2026

described above clearly demonstrate the urgent need for governmental policies that regulate urban expansion in areas unsuitable for real estate development. Furthermore, it is crucial to promote the systematic updating of microzonation and multi-hazard risk studies, particularly for seismic and meteorological phenomena, in newly urbanized areas. This will provide the diagnostic tools required to support disaster risk management policies and to minimize future impacts on both human life and property.

The socioeconomic segmentation results further highlight that approximately 75% of the newly urbanized area remains undefined in terms of socioeconomic type due to gaps in demographic and census coverage in peripheral and informal settlements. Among classified zones, expansion is dominated by low (5 km^2) and medium-low (4 km^2) income strata, highlighting a pattern of socioeconomic segregation that extends Lima's inequality toward its periphery. This concentration of expansion in lower-income zones indicates that recent urban growth is reinforcing patterns of social inequality and spatial marginalization at the city's edges.

Methodologically, the improved recall in peripheral and remote areas—where the WSF underperforms—demonstrates the model's ability to detect small and fragmented structures typical of informal settlements. The iterative human-in-the-loop refinement effectively enhanced local consistency without requiring exhaustive manual labeling. Despite these advantages, the approach has several limitations. The 10 m resolution of Sentinel-2 imagery constrains the detection of fine-scale features, and residual cloud contamination may cause localized classification errors. The reliance on open-source data, while beneficial for reproducibility, can introduce positional inaccuracies and inconsistent labels. Finally, although the model performs well for Lima, its transferability to other cities depends on the similarity of their urban morphologies and spectral characteristics. Retraining with region-specific samples remains necessary to ensure generalization.

Overall, this study helps bridge the gap between urban remote sensing and disaster risk management in the Global South. The proposed framework enables urban planners and local authorities to quantify and monitor unregulated urbanization using open data and modest computational resources. By identifying where and how Lima continues to expand into hazard-prone zones, the approach supports the formulation of proactive, risk-informed planning measures aligned with Peru's 2050 National Development Plan, which aims to promote a sustainable and resilient territory.

6 Conclusions

This study presented a semi-supervised segmentation framework for mapping urban expansion in Metropolitan Lima using Sentinel-2 imagery and weakly labeled data. By integrating incomplete reference information with a NNPU-based learning scheme and targeted human-in-the-loop refinement, the approach produced detailed and internally consistent built-up maps that overcome key limitations of global products such as the World Settlement Footprint. The method proved particularly effective in peripheral and rapidly evolving areas, where informal settlements are common and conventional datasets tend to underrepresent recent construction. Between 2016 and 2025, Metropolitan Lima expanded by approximately 76 km^2, with growth concentrated in Carabayllo, Lurín, Pachacámac, and other districts undergoing active land subdivision. A portion of this growth occurred within zones exposed to natural hazards: about 4 km^2 in tsunami-prone coastal areas and nearly 30 km^2 in medium-to-high landslide susceptibility zones. Additional development covering approximately 15.5 km^2 took place on soils classified as Type IV and Type V in seismic microzonation studies, where strong ground amplification and potential instability are expected. These findings underscore the need for risk-informed land-use planning and the systematic incorporation of hazard information in urban management decisions.

The results also highlight the importance of this framework for detecting early-stage land occupations, which in Lima often begin as temporary, uninhabited structures and later evolve into consolidated informal settlements. Early identification of such occupations may support interventions aimed at reducing land trafficking, improving territorial governance, and mitigating the social and environmental impacts associated with unregulated urban expansion. Overall, this study demonstrates that semi-supervised learning provides a viable and scalable pathway for improving urban footprint mapping in data-scarce environments. The resulting products can support local governments in monitoring urban growth, evaluating exposure to natural hazards, and designing strategies consistent with Peru's 2050 National Development Plan.

Code availability. The code for the semi-supervised CNN is available on request to the corresponding author.

Data availability. The Sentinel-2 data can be downloaded from https://scihub.copernicus.eu/dhus/#/home (Copernicus), and the OSM plugin for QGIS, QuickOSM, is available at https://plugins.qgis.org/plugins/QuickOSM/ (QuickOSM, 2025).

Author contributions. JJ and LM conceptualized and designed the study. CG, SS, and JJ analyzed urban growth and exposure to hazardous areas. MV and MS prepared the closed polygons and contributed to the interpretation of urban expansion along city boundaries. JJ and MV analyzed urban growth across different socioeconomic categories. LM and JJ drafted the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This study was supported by CONCYTEC-PROCIENCIA under the program Undergraduate and Graduate Theses in Science, Technology, and Technological Innovation 2024 [Contract No. PE501091200-2024].

https://doi.org/10.5194/egusphere-2025-5987
Preprint. Discussion started: 22 January 2026

References

Alva-Hurtado, J. and Ortiz-Salas, C.: Actualización de la Ocurrencia del Fenómeno de Licuación de suelos en el Perú, TECNIA, 30, 6–17, https://doi.org/10.21754/tecnia.v30i2.756, 2020.

Assarkhaniki, Z., Sabri, S., and Rajabifard, A.: Using open data to detect the structure and pattern of informal settlements: an outset to support inclusive SDGs' achievement, Big Earth Data, 5, 2021.

Bruno Adriano, Luis Moya, Erick Mas, Hiroyuki Miura, Masashi Matsuoka, and Shunichi Koshimura: Urban Vulnerability Analysis in the Tributary Basin of the Rimac River, Peru Using High-Resolution Remote Sensing Imagery, in: IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium.

Calderon, D., Aguilar, Z., Lazares, F., Alarcon, S., and Quispe, S.: Development of a seismic microzoning map for Lima City and Callao, Peru, Journal of Disaster Research, 9, 939–945, 2014.

Calderon, D., Lazares, F., Gonzales, C., Carrasco, G., Aguilar, Z., and Alarcon, S.: Update of the Seismic Microzonation Map of Lima and Callao, Peru, Journal of Disaster Research, 20, 1010–1022, 2025.

CEPLAN: Tomo II: Indicadores, Plan Estratégico de Desarrollo Nacional al 2050, Tech. Rep. Decreto Supremo N.º 103-2023-PCM, Presidencia del Consejo de Ministros (PCM), Centro Nacional de Planeamiento Estratégico, Lima, Perú, documento oficial del gobierno peruano; contiene indicadores y metas del PEDN hasta 2050., 2023.

Copernicus: Copernicus Open Access, copernicus [data set], available at: https://scihub.copernicus.eu/dhus/#/home (last access: November 2024).

Dammert-Guardia, M., Aiquipa, A., Obregón, D. T., et al.: Tráfico de tierra, crecimiento urbano especulativo y gobernanza de las periferias de Lima: caso Pasamayito–San Juan de Lurigancho, Apuntes. Revista de ciencias sociales, 52, 2025.

Das, T., Shahfahad, Naikoo, M. W., Talukdar, S., Parvez, A., Rahman, A., Pal, S., Asgher, M. S., Islam, A. R. M. T., and Mosavi, A.: Analysing Process and Probability of Built-Up Expansion Using Machine Learning and Fuzzy Logic in English Bazar, West Bengal, Remote Sensing, 14, 2022.

Debray, H., Gassilloud, M., Lemoine-Rodríguez, R., Wurm, M., Zhu, X., and Taubenböck, H.: Universal patterns of intra-urban morphology: Defining a global typology of the urban fabric using unsupervised clustering, International Journal of Applied Earth Observation and Geoinformation, 141, 2025.

DHN: Carta de inundación en caso de tsunami: La Punta, Callao, estudio de la Provincia Constitucional del Callao que cubre los distritos de La Punta, Callao y La Perla, 2014.

Ding, L., Zhang, H., and Li, D.: Monitoring and Analysis of Urban Sprawl Based on Road Network Data and High-Resolution Remote Sensing Imagery: A Case Study of China's Provincial Capitals, Photogrammetric Engineering & Remote Sensing, 88, 479–486, 2022.

Directorate of Hydrography and Navigation – Peruvian Navy: Tsunami Inundation Map, https://www.dhn.mil.pe/cnat/cartas-inundacion, accessed: 2025-11-01, 2025.

Espinoza, A. and Fort, R.: Mapeo y tipología de la expansión urbana en el Perú, Tech. rep., Grupo de Análisis para el Desarrollo (GRADE), https://grade.org.pe/publicaciones/mapeo-y-tipologia-de-la-expansion-urbana-en-el-peru/, 2020.

Fallatah, A., Jones, S., Wallace, L., and Mitchell, D.: Combining Object-Based Machine Learning with Long-Term Time-Series Analysis for Informal Settlement Identification, Remote Sensing, 14, 1226, 2022.

Geiß, C., Zhu, Y., Qiu, C., Mou, L., Zhu, X. X., and Taubenböck, H.: Deep relearning in the geospatial domain for semantic remote sensing image segmentation, IEEE Geoscience and Remote Sensing Letters, 19, 1–5, 2020.

German Aerospace Center: World Settlement Footprint (WSF) Evolution - Landsat-5/-7 - Global, https://download.geoservice.dlr.de/WSF_EVO/, accessed on February, 2025, 2023.

Gonzales, C., Garay, R., Moya, L., Lazares, F., Aguilar, Z., Calderon, D., Diaz, M., Matsuoka, M., and Zavala, C.: Preliminary System for the Estimation of Peak Ground Acceleration Distribution in Metropolitan Lima and Callao: Application in Recent Seismic Events, Journal of Disaster Research, 18, 319–328, 2023.

Gram-Hansen, B., Helber, P., Varatharajan, I., Azam, F., Coca-Castro, A., Kopackova, V., and Bilinski, P.: Mapping Informal Settlements in Developing Countries using Machine Learning and Low Resolution Multi-spectral Data, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, pp. 361–368, 2019.

Guo, Y., Wang, F., Xiang, Y., and You, H.: Semisupervised Semantic Segmentation With Certainty-Aware Consistency Training for Remote Sensing Imagery, in: IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, IEEE, 2016.

Huang, C. and Xu, N.: Quantifying Urban Expansion from 1985 to 2018 in Large Cities Worldwide, Geocarto International, 37, 1–17, 2022.

Ita, P., Santa-Cruz, S., Daudon, D., Tarque, N., Párraga, A., and Ramos, V.: Out-of-plane analysis of dry-stone walls using a pseudo-static experimental and numerical approach in natural-scale specimens, Engineering Structures, 288, 116 153, 2023.

Ita, P., Santa-Cruz, S., Daudon, D., and Tarque, N.: Numerical and experimental pseudo-static study on block arrangement effects in traditional dry-stone walls for out-of-plane mechanical behavior evaluation, Engineering Failure Analysis, 166, 108 900, 2024.

Kesavan, H., Choudhary, S., Khan, S., R, S., and Sharma, A.: AI Based Urban Resilience Planning: Opportunities and Challenges, Journal of Environmental & Earth Sciences, 6, 200–214, 2024.

Kiryo, R., Niu, G., Plessis, M. C. d., and Sugiyama, M.: Positive-Unlabeled Learning with Non-Negative Risk Estimator, 2017.

Konagai, K., Kiyota, T., Suyama, S., Asakura, T., Shibuya, K., and Eto, C.: Maps of soil subsidence for Tokyo bay shore areas liquefied in the March 11th, 2011 off the Pacific Coast of Tohoku Earthquake, Soil Dynamics and Earthquake Engineering, 53, 240–253, 2013.

Kramer, S. L.: Geotechnical Earthquake Engineering, chap. 9, Prentice Hall, Upper Saddle River, New Jersey, ISBN 0133749436, chapter 9: Liquefaction, 1996.

Kuffer, M., Pfeffer, K., and Sliuzas, R.: Slums from Space—15 Years of Slum Mapping Using Remote Sensing, Remote Sensing, 8, 455, 2016.

Li, B. S. and Liu, C. N.: Research on Intelligent Recognition of Violation Based on Big Data of Urban Construction, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-3-W10, 721–724, 2020.

Mansour, S., Ghoneim, E., El-Kersh, A., Said, S., and Abdelnaby, S.: Spatiotemporal Monitoring of Urban Sprawl in a Coastal City Using GIS-Based Markov Chain and Artificial Neural Network (ANN), Remote Sensing, 15, 601, 2023.

Marconcini, M., Metz-Marconcini, A., Esch, T., and Gorelick, N.: Understanding current trends in global urbanisation-the world settlement footprint suite, GI_Forum, 9, 33–38, 2021.

Moya, L., Garcia, F., Gonzales, C., Diaz, M., Zavala, C., Estrada, M., Yamazaki, F., Koshimura, S., Mas, E., and Adriano, B.: Brief communication: Radar images for monitoring informal urban settlements in vulnerable zones in Lima, Peru, Natural Hazards and Earth System Sciences, 22, 65–70, 2022.

Moya, L., Vilela, M., Jaimes, J., Espinoza, B., Pajuelo, J., Tarque, N., Santa-Cruz, S., Vega-Centeno, P., and Yamazaki, F.: Vulnerabilities and exposure of recent informal urban areas in Lima, Peru, Progress in Disaster Science, 23, 100 345, 2024.

National Center for Strategic Planning (CEPLAN): Análisis del crecimiento y expansión urbana a nivel nacional y el impacto a nivel regional, Tech. rep., Observatorio de Políticas Públicas, CEPLAN, https://observatorio.ceplan.gob.pe/publicacion/detalle/274, accedido el 03 noviembre 2025, 2023.

National Institute of Statistics and Informatics (INEI): XII Population Census, VII Housing Census and III Indigenous Communities Census 2017, https://www.inei.gob.pe/estadisticas/censos/, 2017.

National Institute of Statistics and Informatics (INEI): Planos Estratificados de Lima Metropolitana a Nivel de Manzana, Tech. rep., Instituto Nacional de Estadística e Informática, https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1744/libro.pdf, documento basado en los Censos Nacionales de Población y Vivienda 2017, ENAHO 2017-2018 y registros administrativos, 2020.

Peirano, G., de la Flor Sáenz, L. E., and Astucuri, J. V.: Análisis del crecimiento y expansión urbana a nivel nacional y el impacto a nivel regional, Centro Nacional de Planeamiento Estratégico, 2023.

Pesaresi, M., Schiavina, M., Politis, P., Freire, S., Krasnodębska, K., Uhl, J. H., Carioli, A., Corbane, C., Dijkstra, L., Florio, P., et al.: Advances on the Global Human Settlement Layer by joint assessment of Earth Observation and population survey data, International Journal of Digital Earth, 17, 2390 454, 2024.

Pimentel-Sanchez, N.: Tráfico de Tierras en Áreas Periurbanas de Lima, Perú: Tipos e Implicancias Político-Institucionales y Socioculturales, Global Journal of Human-Social Science: C Sociology Culture, 21, 2021.

Qiao, Y., Zhong, B., Du, B., Cai, H., Jiang, J., and Liu, Q.: SAM Enhanced Semantic Segmentation for Remote Sensing Imagery without Additional Training, IEEE Trans. Geosci. Remote Sens., 63, 5610 816, 2025.

QuickOSM: QuickOSM Plugin for QGIS, available at: https://plugins.qgis.org/plugins/QuickOSM/ (last access: February 2025), 2025.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), vol. 9351 of *Lecture Notes in Computer Science*, pp. 234–241, Springer, 2015.

Saha, S., Mou, L., Zhu, X. X., Bovolo, F., and Bruzzone, L.: Semisupervised Change Detection Using Graph Convolutional Network, IEEE Geosci. Remote Sens. Lett., 18, 607–611, 2021.

Santa-Cruz, S., Daudon, D., Tarque, N., Zanelli, C., and Alcántara, J.: Out-of-plane analysis of dry-stone walls using a pseudo-static experimental and numerical approach in scaled-down specimens, Engineering Structures, 245, 112 875, 2021.

Shi, J., Wu, T., Qin, A. K., Lei, Y., and Jeon, G.: Semisupervised Adaptive Ladder Network for Remote Sensing Image Change Detection, IEEE Trans. Geosci. Remote Sens., 60, 2022.

Sobrino, E., Patroni, F., and Francisco, T.: Análisis del crecimiento y expansión urbana a nivel nacional y el impacto a nivel regional, Tech. rep., CEPLAN, 2023.

Taubenböck, H., Mast, J., Lemoine Rodríguez, R., Debray, H., Wurm, M., and Geiß, C.: Was global urbanization from 1985 to 2015 efficient in terms of land consumption?, Habitat International, 160, 103 397, 2025.

Tingzon, I., Dejito, N., Flores, R. A., Guzman, R. D., Carvajal, L., Erazo, K. Z., Cala, I. E. C., Villaveces, J., Rubio, D., and Ghani, R.: Mapping New Informal Settlements using Machine Learning and Time Series Satellite Images: An Application in the Venezuelan Migration Crisis, in: 2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G), pp. 198–203, 2020.

UN-Habitat: World Cities Report 2024: Cities and Climate Action, UN-Habitat, Nairobi, ISBN 978-92-1-133395-4, 2024.

Vera, E.: Tráfico de terrenos: invasión y crimen a gran escala, https://elcomercio.pe/peru/tierras-nadie-invasion-crimen-gran-escala-informe-noticia-49616
accedido el 03 noviembre 2025, 2018.

Vilela, M. and Moschella, P.: Paisaje y expansión urbana sobre espacios naturales en ciudades intermedias. El caso de Purrumpampa en Huamachuco, La Libertad, Perú, Bulletin de l'Institut français d'études andines, pp. 529–550, 2017.

Villacorta Chambi, S. P., Núñez Juárez, S., Vásquez Acuña, J. M., Pari Pinto, W., Ochoa Zubiate, M. B., Benavente Escobar, C. L., Tatard, L., Luque Poma, G., Rosado Seminario, M., Fídel Smoll, L., et al.: Peligros geológicos en el área de Lima Metropolitana y la región Callao-[Boletín C 59], 2015.

Wang, J.-X., Li, T., Chen, S.-B., Tang, J., Luo, B., and Wilson, R.: Reliable Contrastive Learning for Semi-Supervised Change Detection in Remote Sensing Images, IEEE Trans. Geosci. Remote Sens., 60, 2022.

Wei, S., Zhang, T., and Ji, S.: A Concentric Loop Convolutional Neural Network for Manual Delineation-Level Building Boundary Segmentation From Remote-Sensing Images, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–11, 2022.

Yokoya, N., Xia, J., and Broni-Bediako, C.: Submeter-level land cover mapping of Japan, International Journal of Applied Earth Observation and Geoinformation, 127, 2024.

Yue, J., Fang, L., Ghamisi, P., Xie, W., Li, J., Chanussot, J., and Plaza, A. J.: Optical Remote Sensing Image Understanding with Weak Supervision: Concepts, Methods, and Perspectives, IEEE Geoscience and Remote Sensing Magazine, 10, 250–269, 2022.