

Summary

This paper presents a standardized methodology that combines machine learning, satellite imagery, and household survey data to estimate slum populations at the grid-cell level (~6.72 km) across 129 Global South countries. The authors construct a slum indicator using DHS survey data, extract features from Landsat and nighttime light imagery with a ResNet-34 network, and then apply XGBoost for prediction, ultimately producing a spatial distribution map of slum populations for the year 2018. The model achieves a median R^2 of 0.89 for within-country predictions and 0.85 for country-level holdout validation. To the best of the authors' knowledge, this is the first large-scale, spatially continuous inventory of slum populations covering the entire Global South.

General Comments

This study involves a large volume of data and a complex methodology, and it carries significant social and policy implications. However, as a reviewer not specialised in slum research, I encountered considerable difficulties in understanding the manuscript. The paper is written under the assumption that readers possess deep background knowledge in remote sensing, machine learning, and slum studies, which is not friendly to interdisciplinary readers or policy makers. The comments below are intended to help the authors improve the readability and transparency of the paper, so that a wider audience (including researchers from other fields and students) can understand and use this valuable data product.

Major Comments

1. Add in-text explanations of key technical terms at first appearance for non-specialist readers.

Many terms (e.g., “ResNet-34”, “XGBoost”, “DHS cluster”, “GHS-POP”, “SDG 11.1 framework”) are introduced without any intuitive explanation. Readers from other will struggle to follow the logic. I suggest that for each important term, the authors add a short plain-language sentence the first time it appears – for example: “ResNet-34, a widely used deep learning model for image recognition” or “GHS-POP, a global population dataset that distributes census counts into 1-km grid cells.”

2. Address the coarseness of the 6.72 km resolution and clarify where this resolution comes from.

The mapping resolution (6.72 km at the equator) is derived from Landsat 30 m pixels $\times$ 224 pixels. However, a 6.72 km grid cell ($\approx 45 \text{ km}^2$ at the equator) is quite coarse for studying slums, which are often small clusters of a few hundred meters. The authors mention in Section 5.2 that this resolution is “not intended for site-specific operational applications”, but I suggest they strengthen this limitation earlier in the paper (e.g., in the abstract or results) and explicitly state which input data force this resolution. Also, please clarify whether any alternative higher-resolution data (e.g., Sentinel-2, Planet) were considered and why they were not used. Readers will question whether such coarse cells can meaningfully capture slum populations, especially in heterogeneous urban areas.

Minor Comments

1. **Line 20 (“Slums are a visible manifestation...”)** – The term “Global South” is used throughout but never defined in the abstract or main text. Please add a brief definition (e.g., “developing regions as defined by UNCTAD”) on first use.
2. **Line 42–46** – The opening paragraph cites several human-rights references. For a non-legal reader, a one-sentence explanation of why “adequate housing” is a human right would be helpful. Consider a short footnote.
3. **Line 74–76 (“Know Your City Campaigns”)** – Most readers will not know what this is. Add a short explanation (e.g., “a participatory slum mapping initiative by Slum Dwellers International”).
4. **Line 101–102** – “Grid-cell level (6.72 km × 6.72 km at the equator)” – Add a parenthetical example of how many grid cells a typical city (e.g., Nairobi or Mumbai) would contain. This gives a concrete sense of scale.
5. **Line 135–139** – “A cluster refers to a DHS enumeration unit, approximately the size of a neighbourhood or village.” – Please add approximate population or household range (e.g., “typically 50–300 households”) so readers can understand the granularity of the training data.
6. **Line 150–152** – Mention of random displacement (2 km urban, 5 km rural) is critical. Please add a short sentence in Section 3.2 or 3.3 explaining how your model handles this uncertainty – for example, “the 6.72 km grid cell size was chosen to absorb this displacement.”
7. **Line 177–181** – The description of the NPP-VIIRS-like nighttime light dataset is technical. Add a plain-English statement: “Brighter nighttime lights generally indicate more developed areas, so this helps the model distinguish slums from better-served neighbourhoods.”
8. **Line 228–232 (SDG 11.1 and “access to electricity”)** – Explain why electricity was added while “security of tenure” was excluded, in one simple sentence. Currently it reads as an arbitrary choice.
9. **Line 273–276** – “ResNet-34” and “transfer learning” will be opaque to many. Add a short intuitive explanation: “Like reusing a general-purpose

image-recognition system (trained on everyday photos) and adapting it to recognise slum-like patterns in satellite images.”

- 10.**Line 385–388** – “Excluded grid-cells with population below 1,000” – Explain the reason (e.g., “to avoid unreliable predictions in very sparsely populated areas”). This reassures readers that you did not arbitrarily cut off data.
- 11.**Line 402–409** – The assumption that slum and non-slum households have the same average size is crucial. The discussion already does a good sensitivity analysis. Please move a short version of that sensitivity (e.g., “assuming $\pm 10\%$ difference changes national estimates by $\sim 8\%$ ”) into the main results section (Section 4.2 or 4.4), not only in the Discussion. This makes it more visible to readers who skim.
- 12.**Figure 5a and 6a (maps)** – The maps use color gradients. For colour-blind or greyscale printing, add a note in the caption that patterns (e.g., hatching) are not used, but the digital version relies on colour. Also, define “dark blue” as “highest absolute slum population” explicitly in the caption.
- 13.**Line 532–541** – The text mentions “very low ($<10\%$) ... very high ($>60\%$)” concentration levels. Please add these exact thresholds to the legend of Figure 6a, not only in the text.