

This manuscript presents a technically sophisticated effort to create a first-of-its-kind, spatially explicit inventory of slum populations for the Global South. The authors' framework, which integrates diverse data sources using a hybrid deep learning approach, is novel and addresses a topic of significant importance for sustainable development and urban studies.

However, the manuscript's claims of robustness currently rest on a methodological foundation that requires significant clarification and justification. Several critical design choices, from the definition of the ground-truth variable to the modeling architecture and data processing steps, are not sufficiently defended. Therefore, it requires a major revision to address these fundamental issues before it can be considered for publication.

Major Comments

- 1. Line 241-252:** The formula appears to calculate a weighted average of three within-dimension proportions. Why is this structure superior to more standard aggregation methods?
- 2. Lines 289-315 & 638-642:** The manuscript employs a complex, two-stage model (ResNet-34 for feature extraction + XGBoost for regression), but its necessity and true performance are not adequately proven. The high R^2 values reported (82-96%) raise concerns about potential overfitting, especially given the presence of spatial autocorrelation in the data. The authors mention constructing regionally stratified models" in the Discussion (Lines 638-642), which is a crucial methodological detail. Was a formal **spatial cross-validation** procedure implemented (e.g., holding out entire geographic regions or countries for testing)? If not, the reported performance metrics are likely inflated and do not reflect the model's true generalizability.
- 3. Lines 200-201:** The manuscript fails to state *how* the 1km GHS-POP data was aggregated to the 6.72km analysis grid. For population **counts**, **summation** is the only method that preserves demographic totals.

Minor Comments

- Lines 638-642:** Key details about the modeling strategy (e.g., the use of regionally stratified models) are currently in the Discussion section. All descriptions of what was done should be consolidated within the Methods section to ensure a logical flow.
- Section 2.1 vs 2.2:** The Methods section inconsistently switches between present tense (This study uses...) and past tense (We obtained...). The established convention is to use the **past tense** to describe the work performed. This should be corrected for professionalism.
- Lines 46 & 835:** The data is repeatedly cited as being in the Zalando repository. The correct name is **Zenodo**.
- Lines 767-831:** The Implications and future research section is good but could be more

impactful by proposing **1-2 concrete, actionable research questions** that this new dataset uniquely enables (e.g., How does the spatial distribution of slum populations correlate with exposure to specific climate hazards at a regional scale?).