URBAN-RURAL DISPARITIES IN SATELLITE-BASED POVERTY PREDICTION

Emily Aiken *
UC Berkeley

Esther Rolf *
Harvard University

Joshua Blumenstock
UC Berkeley

ABSTRACT

Poverty maps derived from satellite imagery are increasingly used to inform high-stakes policy decisions, such as the targeting of humanitarian aid and the allocation of government resources. These maps are typically constructed by training machine learning algorithms with country- or continent-scale data, but many real-world applications are focused on specific urban or rural areas. This paper shows that satellite-based poverty predictions are less accurate at distinguishing levels of wealth within urban and rural areas than they are at distinguishing wealth differences between urban and rural areas, investigates why this may be the case, and documents the implications of these disparities for downstream policy decisions.

1 INTRODUCTION

Satellite-based poverty maps are increasingly being used to inform critical policy decisions, from estimating interim sub-national statistics (Hofer et al., 2020) to the targeting of humanitarian aid (Aiken et al., 2022; Smythe & Blumenstock, 2022) and social services (Gentilini et al., 2022). Such poverty maps are typically constructed by training machine learning algorithms to recognize poverty from high resolution imagery on a relatively modest amount of 'ground truth' data, and using the outputs of such algorithms to fill in spatial gaps in regions where ground truth data are not available. The deployment of poverty maps for real-world policies requires that the maps provide accurate and fair estimates of living conditions on the ground, particularly in settings where policymakers may consider such maps technocratic and therefore 'objective' measures of poverty (Kondmann & Zhu, 2021).

However, results in past work have hinted that machine learning models based on satellite imagery may rely heavily on indicators of urbanization observable in imagery as a predictor of wealth, and therefore have limited predictive power across *just* urban or rural regions. For example, Yeh et al. (2020) document that a satellite-based machine learning model for the African continent explains much less of the variance within urban (R^2=0.40) or rural (R^2=0.32) areas than in all areas together (R^2=0.70). Zhang et al. (2022) document urban-rural accuracy disparities for land-cover mapping in China, and Engstrom et al. (2017) add complementary evidence showing that nearly all of the most important satellite-based features used in satellite-based poverty prediction models in Sri Lanka are related to urban build-up. These disparities are particularly concerning in the context of policy deployments of satellite-based poverty maps that are restricted to just urban or just rural areas — a fairly common approach in anti-poverty programming (Lindert et al., 2020).

In this paper we build on the existing literature to more thoroughly explore representational disparities in satellite-based poverty maps between urban and rural areas and their implications for fairness. Using survey data and satellite imagery from seven countries (Table 1), our analysis produces three main results: (1) we document substantial disparities in the degree to which poverty can be estimated from satellite imagery between urban and rural areas, (2) we show how such disparities propagate into downstream policy decisions by simulating a hypothetical social protection program using satellite-based poverty maps, and (3) we explore more generally the role of indicators of urbanization in predicting poverty from satellite imagery, finding that proxies for urbanization explain most of the variation in wealth observed in satellite-based poverty maps.

*These authors contributed equally to this work

	DHS Year	Number of Clusters	% Rural Clusters	Tiles Per Cluseter		
				Minimum	*Mean*	*Maximum*
Colombia	2010	4,868	30.1%	16	38	88
Honduras	2011	1,128	56.2%	16	54	86
Indonesia	2017	1,319	57.8%	16	55	87
Kenya	2014	1,585	61.2%	16	57	86
Nigeria	2018	1,359	58.8%	16	56	86
Peru	2012	1,131	38.8%	16	43	85
Philippines	2017	1,213	64.0%	16	59	88

Table 1: Summary of datasets.

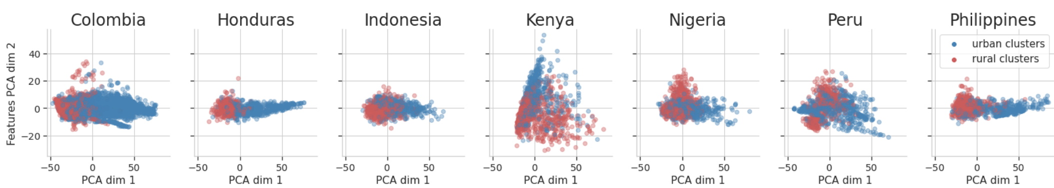

Figure 1: **Distribution analysis of imagery features across urban and rural clusters**. First and second dimensions of per-country PCA projections of the MOSAIKS features. Across countries, these two dimensions explain between 93.5% and 98.5% of the variation in the 4000 features.

2 Data and Methods

Survey Data. We use ground truth wealth measures from the most recent demographic and health (DHS) surveys from seven countries: Colombia, Honduras, Indonesia, Kenya, Nigeria, Peru, and the Philippines. Each of these surveys was conducted in 2010 or later and interviewed between 20,000 and 60,000 households in 1,000-5,00 clusters. Clusters are small groups of geographically adjacent households, sampled at random or stratified random in each country. Clusters are roughly equivalent to a neighborhood in urban areas (for which the provided cluster center location is jittered with a 2km radius) or a village in rural areas (for which the provided cluster center location is jittered with a 5km radius). Our ground truth measure of poverty for each cluster is the average asset-based wealth index — as recorded in each DHS survey — for households in that cluster. Each DHS survey uses a country-specific rule to define which areas are rural and which are urban; rural ratios range from 30% to 64% (Table 1).

Satellite Features. We obtain a set of tabularized features summarizing satellite tiles in each country we study from MOSAIKS (Rolf et al., 2021), accessed via `siml.berkeley.edu` (Carleton et al., 2022). The underlying RGB satellite images are from Planet Labs in 2019. Features are generated through an unsupervised machine learning approach based on random convolutional features (RCFs), which are shown to carry skill across a variety of prediction tasks (Rolf et al., 2021).

RCF embedding functions are essentially a wide and shallow feed-forward convolutional neural network with random but fixed (non-optimized) weights. We use RCFs as convenient way to obtain images features with a single, fixed featurization method across countries. Figure 1 visualizes the feature distributions across urban and rural instances. These embeddings of satellite imagery – which are created completely independently of the poverty prediction task, or any downstream task for that matter – already encode a high amount of signal as to whether a region is urban or rural.

In urban areas, where DHS clusters centroids have a jitter of up to 2km, each cluster is represented by around 16 MOSAIKS tiles; in rural areas, where DHS cluster centroids have up to a 5km jitter, each cluster is represented by around 87 MOSAIKS tiles (Table 1) — thus each rural cluster is represented by approximately six times as many tiles as urban clusters. For each cluster we calculate the mean of each MOSAIKS feature across all tiles that fall within the jitter radius of the cluster centroid, weighted for degree of overlap with the cluster extent.

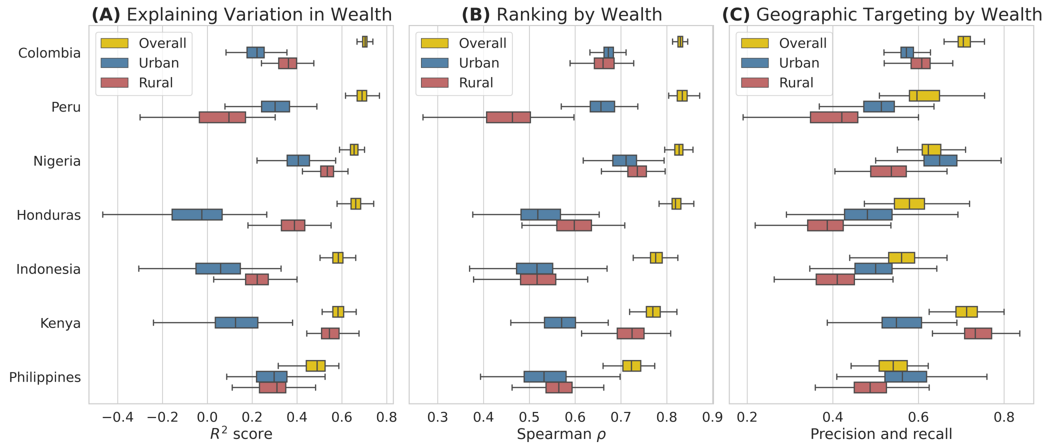

Figure 2: **Performance measured overall and within groups**. *Panel A:* Fraction in variation of true wealth explained by predictions (R^2 score) is higher overall (yellow) than within urban areas (blue) or rural areas (red) *Panel B:* Rank correlation (Spearman's ρ) between predicted and ground-truth wealth are higher in each country as a whole (yellow) than within urban (blue) and rural (red) regions in each country. *Panel C:* As a result, a simulated aid program that targets the poorest 20% of regions in urban (blue) or rural (red) parts of a country generally has lower accuracy than a program that targets within the entire country (yellow).

Methods. Our machine learning simulations begin by assigning 75% of regions in each country to a training set and 25% to a test set, uniformly at random. Following Rolf et al. (2021), in each country we train a ridge regression model to predict average household wealth in training set regions from the associated satellite-derived MOSAIKS features, tuning the ℓ_2 penalty via three-fold cross-validation. We then use the trained model to produce wealth estimates for every region in the test set. We report the mean and uncertainty across 100 simulations in all our results.

Our analysis focuses on understanding differences in how well satellite-based poverty prediction models perform between urban and rural areas (both predictive performance and applicability to a downstream decision-making task). First, we document patterns of representational disparities within and between urban and rural areas, by measuring predictive accuracy (measured with R^2 and Spearman's ρ) in the test set overall, in just urban regions, and in just rural regions.

Second, to understand how representational disparities manifest in downstream policy applications, we simulate hypothetical social protection programs using satellite-based predictions. We compare the precision and recall (equal by definition in this application (Brown et al., 2018)) of a hypothetical geographically targeted program that (1) is run in an entire country, targeting the poorest 20% of clusters in each country as a whole, (2) is run in only urban areas, targeting the poorest 20% of urban clusters, and (3) is run in only rural areas, targeting the poorest 20% of rural clusters.

Third, to explore the importance of urban build-up as an indicator of poverty in satellite-based poverty prediction models more generally, we compare the accuracy of two sets of predictions for identifying poverty in each country as a whole: (1) regression predictions from a model trained to estimate poverty (as in the first two analyses), and (2) probabilistic predictions of each region being urban, from a classification model trained to predict whether regions are urban or rural.

3 Results

Consistent with past work (Jean et al., 2016; Engstrom et al., 2017; Yeh et al., 2020; Chi et al., 2022), we find that satellite-based wealth predictions explain a significant portion of the variance in ground-truth wealth within each of the seven countries we study (mean R^2 = 0.47-0.70), and there is a high rank correlation between wealth predictions and ground truth (mean Spearman's ρ = 0.72-0.83). However, in most countries performance is substantially lower when predictions are evaluated

	(A) $R^2(w, \hat{w})$	**(B)** $\rho(w, \hat{w})$	**(C)** AUC$(u, \hat{u})$	**(D)** $\rho(w, u)$	**(E)** $\rho(w, \hat{u})$
Colombia	0.70 (0.01)	0.83 (0.01)	0.94 (0.01)	0.72 (0.01)	0.70 (0.02)
Honduras	0.66 (0.04)	0.82 (0.02)	0.95 (0.01)	0.75 (0.02)	0.75 (0.02)
Indonesia	0.58 (0.03)	0.78 (0.02)	0.93 (0.02)	0.72 (0.02)	0.71 (0.03)
Nigeria	0.65 (0.02)	0.83 (0.01)	0.87 (0.02)	0.58 (0.03)	0.72 (0.03)
Kenya	0.58 (0.03)	0.77 (0.02)	0.84 (0.02)	0.60 (0.03)	0.55 (0.04)
Philippines	0.47 (0.09)	0.72 (0.03)	0.90 (0.02)	0.53 (0.03)	0.63 (0.03)
Peru	0.69 (0.03)	0.83 (0.02)	0.96 (0.01)	0.77 (0.02)	0.74 (0.02)

Table 2: **Relationship between urban build-up and predicting wealth from satellite imagery.** **(A)** and **(B)** Predictive accuracy of satellite-based wealth predictions measured with R^2 and rank correlation. **(C)** Predictive accuracy of satellite-based urban/rural classifications measured with AUC. **(D)** Rank correlation between wealth and an indicator variable for being urban. **(E)** Rank correlation between wealth and a satellite-based prediction of being urban. w represents ground-truth wealth; $\hat{w}$ predicted wealth; u ground-truth urban (a binary indicator), and $\hat{u}$ predicted urban (a probabilistic prediction between 0 and 1). Standard deviations across bootstrapped runs in parentheses.

within urban (mean R^2 =-0.05-0.41 ; mean ρ = 0.51-0.71) and rural (mean R^2 = 0.06-0.55; mean ρ = 0.45-0.74) areas (Figure 2 Panels A, B). There is heterogeneity across countries in terms of which areas are hardest to predict: in two countries (Colombia, Peru) ranking accuracy (ρ) is higher among urban areas than among rural areas. In the remaining five countries (Nigeria, Honduras, Indonesia, Kenya, and the Philippines) ranking accuracy is higher among rural areas. In all countries, at least one of urban or rural areas has substantially lower predictive accuracy than the country as a whole (mean difference in $R^2 > 0.2$; mean difference in $\rho > 0.1$, Figure 2 Panels A, B).

These disparities in predictive power between urban and rural areas propagate into reduced accuracy in downstream policy decisions using satellite-based poverty maps. A simulated social protection program aiming to select the poorest 20% of regions in a country using satellite-based poverty maps tends to have relatively high recall and precision (54-71%), whereas programs identifying the poorest 20% of regions within urban or rural areas have lower recall and precision (38-73% in rural areas and 48-65% in urban areas, Figure 2 Panel C).

More generally, there is a strong relationship between wealth and urbanization across countries (ρ = 0.53-0.77), and urban classifications from satellite imagery have strong performance (AUC = 0.84-0.96, Table 2). It is thus possible that distinguishing between levels of urbanization is central to predicting wealth at a country scale. In fact, the difference between the predictive accuracy of satellite-based wealth predictions and satellite-based predictions of a region being *urban* for identifying wealth is relatively small (difference in Spearman's ρ = 0.07-0.22, Table 2), suggesting that representations of poverty in this featurization of RGB satellite imagery beyond urban build-up are present but limited.

4 DISCUSSION

Our main finding is that representational disparities in satellite imagery between rural and urban areas lead to differences in predictive accuracy for estimating poverty from imagery, consistently across all countries studied. In particular, it is much easier to differentiate poverty between urban and rural areas than within urban or rural regions. A key implication of this result for real-world deployments is that while satellite-based poverty programming at a country scale may be accurate, satellite-based poverty maps may not be as useful in guiding programs that require accurate differentiation across just urban areas or across just rural areas.

The main implication of our results for researchers in machine learning and development economics is that a focus on building predictive models that can distinguish wealth levels within urban and rural areas will be essential for making satellite-based poverty maps a useful and fair measurement tool. Other digital data sources, such as mobile phone data (Blumenstock et al., 2015; Steele et al., 2017), social media data (Fatehkia et al., 2020), or information from crowd-sourced maps (Tingzon et al., 2019) may be helpful alongside novel modeling techniques for improving within-urban and within-rural differentiation.

ACKNOWLEDGMENTS

Aiken acknowledges support from a Microsoft Research PhD Fellowship. Rolf acknowledges support from the Harvard Data Science Initiative (HDSI) and the Center for Research on Computation and Society (CRCS). Blumenstock acknowledges support from the National Science Foundation under CAREER Grant IIS-1942702. We thank Tamma Gabriel Cadamuro, Tamma Carleton, and Guanghua Chi for helpful feedback on earlier versions of this paper.

REFERENCES

Emily Aiken, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E Blumenstock. Machine learning and phone data can improve targeting of humanitarian aid. *Nature*, 603(7903):864–870, 2022.

Joshua Blumenstock, Gabriel Cadamuro, and Robert On. Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076, 2015.

Caitlin Brown, Martin Ravallion, and Dominique Van de Walle. A poor means test? Econometric targeting in Africa. *Journal of Development Economics*, 134:109–124, 2018.

Carleton, Chong, Druckenmiller, Noda, Proctor, Rolf, and Hsiang. Multitask observation using satellite imagery and kitchen sinks (MOSAIKS) API. `https://siml.berkeley.edu`, 2022.

Guanghua Chi, Han Fang, Sourav Chatterjee, and Joshua E Blumenstock. Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences*, 119 (3):e2113658119, 2022.

Ryan Engstrom, Jonathan Samuel Hersh, and David Locke Newhouse. Poverty from space: Using high-resolution satellite imagery for estimating economic well-being. *World Bank Policy Research Working Paper*, (8284), 2017.

Masoomali Fatehkia, Isabelle Tingzon, Ardie Orden, Stephanie Sy, Vedran Sekara, Manuel Garcia-Herranz, and Ingmar Weber. Mapping socioeconomic indicators using social media advertising data. *EPJ Data Science*, 9(1):22, 2020.

Ugo Gentilini, Mohamed Bubaker Alsafi Almenfi, TMM Iyengar, Yuko Okamura, John Austin Downes, Pamela Dale, Michael Weber, David Locke Newhouse, Claudia P Rodriguez Alas, Mareeha Kamran, et al. Social protection and jobs responses to COVID-19. 2022.

Martin Hofer, Tomas Sako, Arturo Martinez Jr, Mildred Addawe, Joseph Bulan, Ron Lester Durante, and Marymell Martillan. Applying artificial intelligence on satellite imagery to compile granular poverty statistics. *Asian Development Bank Economics Working Paper Series*, (629), 2020.

Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

Lukas Kondmann and Xiao Xiang Zhu. Under the radar–auditing fairness in ml for humanitarian mapping. *arXiv preprint arXiv:2108.02137*, 2021.

Kathy Lindert, Tina George Karippacheril, Inés Rodríguez Caillava, and Kenichi Chávez. *Sourcebook on the foundations of social protection delivery systems*. World Bank Publications, 2020.

Esther Rolf, Jonathan Proctor, Tamma Carleton, Ian Bolliger, Vaishaal Shankar, Miyabi Ishihara, Benjamin Recht, and Solomon Hsiang. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature communications*, 12(1):1–11, 2021.

Isabella Smythe and Joshua E Blumenstock. Geographic micro-targeting of social assistance with high-resolution poverty maps. *Proceedings of the National Academy of Sciences*, forthcoming, 2022.

Jessica E Steele, Pål Roe Sundsøy, Carla Pezzulo, Victor A Alegana, Tomas J Bird, Joshua Blumenstock, Johannes Bjelland, Kenth Engø-Monsen, Yves-Alexandre De Montjoye, Asif M Iqbal, et al. Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface*, 14(127):20160690, 2017.

Isabelle Tingzon, Ardie Orden, KT Go, S Sy, V Sekara, Ingmar Weber, M Fatehkia, M García-Herranz, and D Kim. Mapping poverty in the Philippines using machine learning, satellite imagery, and crowd-sourced geospatial information. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2019.

Christopher Yeh, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature communications*, 11(1):1–11, 2020.

Miao Zhang, Harvineet Singh, Lazarus Chok, and Rumi Chunara. Segmenting across places: The need for fair transfer learning with satellite imagery. *arXiv preprint arXiv:2204.04358*, 2022.