Estimating Household-Level Economic Characteristics from High-Resolution Satellite Imagery

Satej Soman [a] Susana Constenla-Villoslada [a] Emily Aiken [a] Joshua E. Blumenstock [a,b]

{satej, susana_constenla, emilyaiken, jblumenstock}@berkeley.edu

a: School of Information, UC Berkeley b: Goldman School of Public Policy, UC Berkeley

---

PRELIMINARY AND INCOMPLETE: PLEASE DO NOT CITE WITHOUT PERMISSION

Submitted to the World Bank Land Conference

13-17 May, 2024

Washington, D.C., USA

keywords

remote sensing
poverty mapping
targeting
machine learning
big data

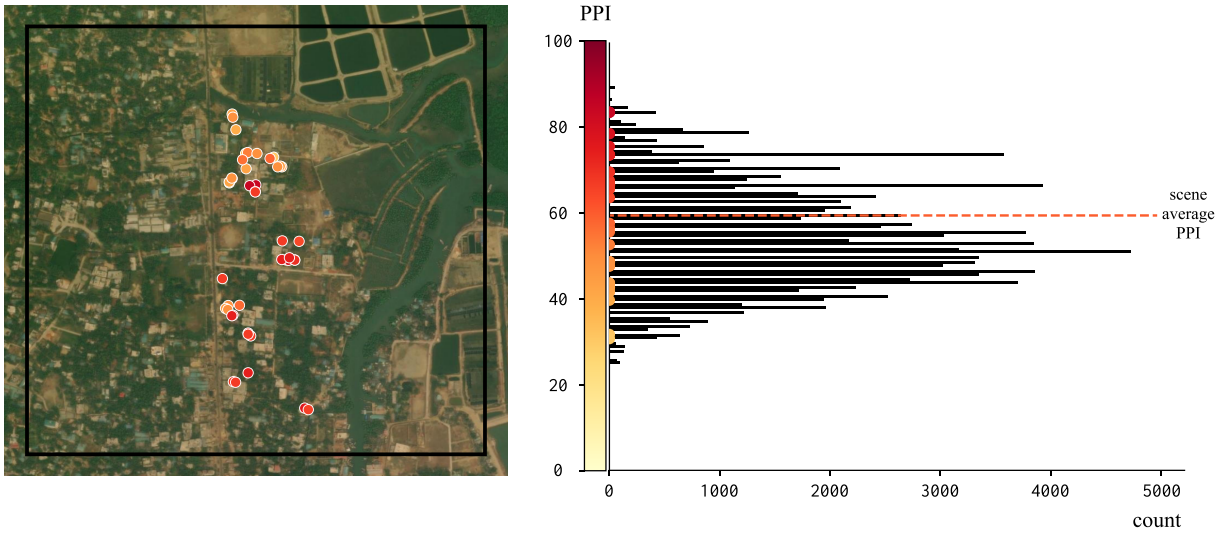
1 Introduction

Understanding economic development, land rights and management, and structural transformation requires accurate and granular measurements of poverty and growth. Fine-grained estimates of living standards are also critical to effectively target policies and evaluate development interventions (Smythe and Blumenstock, 2022; Elbers et al., 2007). However, most low-income countries lack the resources and administrative capacity to regularly collect household-level socioeconomic information (cf. Jerven, 2013).

In the past decade, there has been considerable innovation in methods for constructing estimates of living standards from non-traditional sources of data. Satellite-based poverty estimates have been produced at the village level (Jean et al., 2016; Yeh et al., 2020; Engstrom et al., 2017), the neighborhood level (Smythe and Blumenstock, 2022) and the satellite tile level (e.g., tiles that are 1-2 square kilometres in area (Chi et al., 2022; Rolf et al., 2021).) In general, these studies find that machine learning and satellite data can explain a substantial amount of the in survey-based ground-truth measures of wealth and poverty (e.g., 70% explained variation in an asset index at the village level in Sub-Saharan Africa (Yeh et al., 2020) and 60% explained variation in poverty rates at the village level in Sri Lanka (Engstrom et al., 2017)).

While these methods have been successful at estimating the wealth and poverty of relatively large geographic units that contain over large numbers of households, a great deal of variation in living standards is lost when households are aggregated to larger geographic units. As an example, Figure 1a highlights a 1-square kilometer region in Bangladesh, and indicates with dots the locations of 41 households that were surveyed. We observe that, even within this one small geographic area, there is considerable variation in living standards (as measured by the *Progress Out of Poverty*, or PPI, score); indeed the distribution of PPI scores within this $1km^2$ tile has roughly the same variance as the distribution for the entire region covered by the survey (Figure 1b).

In this study, we explore the extent to which machine learning and high-resolution satellite imagery can be used to accurately estimate the living standards of *individual households*. Using ground truth data from a large survey in Bangladesh, we compare methods that rely on "black-box" deep learning algorithms, to more interpretable algorithms that extract specific characteristics from satellite imagery as a precursor to supervised learning.



(a) geographic locations of survey responses within a given satellite tile, or "scene", colored by PPI score

(b) distribution of PPI scores for the entire survey, with colored dots from the left panel overlaid on the vertical axis and scene-average PPI marked by a dashed line

Figure 1: Average tile size versus distribution of PPI scores in Bangladesh. The left panel maps survey response coordinates, colored by *progress out of poverty* (PPI) score, for a single village and the size of the corresponding 1 square kilometer tile (black box). The right panel shows the distribution of the entire dataset's PPI scores, with colored values from the left panel overlaid on the vertical axis. While state-of-the-art in poverty mapping can accurately predict the scene-average PPI at the 1-2 square kilometer imagery scale, household-level predictions require predicting outcome variables that span almost the entire dataset's range, a more challenging task.

1.1 Related Work

Two previous studies have estimated household-level poverty from satellite imagery, demonstrating the feasibility of the overall approach. However, these studies' generalizability is limited by a number of factors. Both studies suffer from small sample sizes ($N = 231$ households in Kenya (Watmough et al., 2019) and $N = 238$ households in China (Han et al., 2021)), focus on very small geographies (within a single village or city), do not explore deep learning approaches to poverty prediction (which have proven most accurate in past work), and only evaluate predictive accuracy for a single measure of poverty (asset-based wealth, rather than a suite of measures). They also do not evaluate the real-world utility of household-level satellite-based poverty estimates, such as humanitarian aid targeting or tax administration (Knebelmann et al., 2023).

Most work in the literature on satellite-based poverty prediction, independent of the scale at which poverty estimates are produced, takes one of two approaches: either an *explicit featurization* approach or a *deep learning*-based approach. In the explicit featurization approach, researchers define a set of meaningful features from satellite images at the outset, (such as: the number of houses in a community, local road network density, forest cover extent, and average rooftop size). These features are paired with ground-truth village-level socioeconomic data (typically from national surveys such as the Demographic and Health Surveys or Living Standards Measurement Surveys, which conduct households surveys in a small subset of communities in a country). This matched dataset comprise the labels and predictors used to train an ML model to predict socioeconomic status from the satellite-based features (Hersh et al., 2021; Engstrom et al., 2017). In the deep learning approach to satellite-based poverty prediction, introduced by Jean et al. (2016), artificial neural networks initially trained for generic image recognition tasks are "fine-tuned", or further trained, to predict poverty from satellite bands without human

intervention in explicitly defining salient image characteristics at the outset. Such models are typically more accurate than the explicit featurization approach (Jean et al., 2016), but they lack the ability to identify which image attributes are most salient to poverty prediction. Recent papers have generalized the deep learning poverty prediction pipeline, producing and evaluating satellite-based poverty estimates in all of Sub-Saharan Africa (Yeh et al., 2020), all low and middle-income countries (Chi et al., 2022), and globally (CIESIN, 2022). In both approaches, the trained model is then used to predict poverty in all areas of a country, including the large number of areas not surveyed. We test both interpretable explicit feature extraction from images and the deep learning approach to predict poverty from satellite images.

2 Data & Methods

2.1 Data Sources

The relevant datasets to enable household-level poverty prediction from satellite imagery are: a) ground truth, geo-located poverty labels from survey data; b) polygon definitions of building footprints; and c) high-resolution satellite imagery of the relevant prediction area.

For this project's ground truth poverty measure, we use the results of a large geolocated field survey ($N \approx 100{,}000$) conducted in the Cox's Bazar District of southeastern Bangladesh by GiveDirectly. The survey asked respondents demographic questions (number of children in the household; size of household; age, gender identity, and years of education of the household head), asset ownership questions (whether the household owned a bicyle, refrigerator, fan, and mobile phones), and dwelling quality questions (number of walls and access to sanitation). From these survey responses, a *Progress Out of Poverty Index* (PPI)[1] score was calculated, and this score is used as the target variable in this prediction exercise.

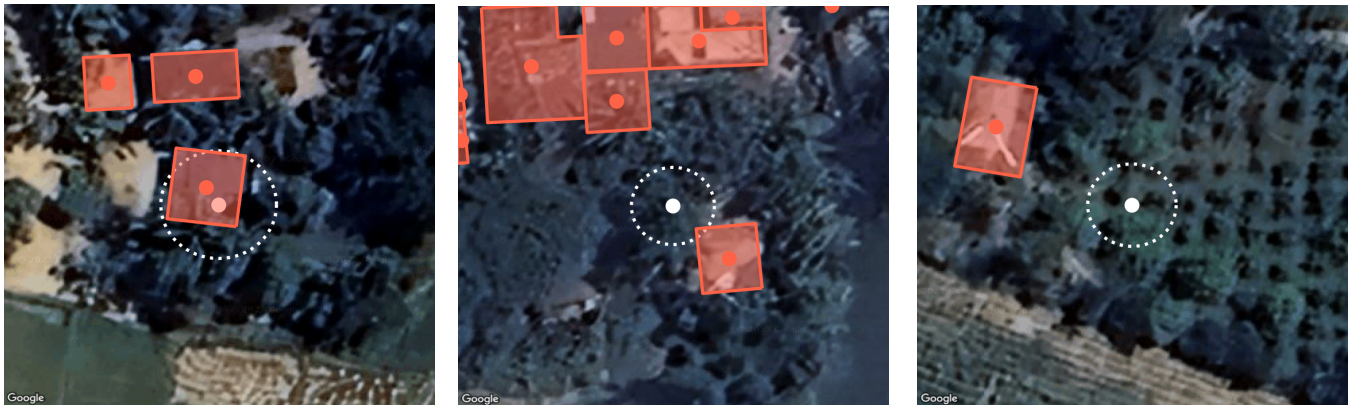
To link the survey responses to buildings, we match GPS locations in the survey with building polygons from the Open-Buildings dataset (Sirko et al., 2021). The GPS coordinates for survey responses were considered to match to a building polygon if the reported coordinate fell within a building polygon (point-in-polygon), if a building footprint fell within a radius of the reported GPS coordinate corresponding to the GPS device's reported accuracy, or if a single GPS coordinate fell within the bounds of the highest-resolution satellite imagery tile for a single building (see Figure 2 for examples of each match type). We were able to match 11,243 households to buildings using the point-in-polygon approach, 11,504 households using the GPS buffer approach, and 4,817 households using the tile bounds-based approach, for a total dataset size of 27,564 household-building matches.

For satellite imagery, we used the Google Static Maps application programming interface (API) to download satellite imagery centered at each matched building polygon's centroid. We procured satellite imagery at the three most granular zoom levels offered by the service: $z = 19$ (roughly the scale of a neighborhood), $z = 20$ (capturing the area around a house), $z = 21$ (corresponding to a given house), as seen in Figure 3.

2.2 Machine Learning Methods

The resulting dataset (PPI score matched to building footprint polygons and satellite imagery tiles) was then featurized in two ways to create predictive variables for a machine learning predictor: with a set of explicit features known from the literature to be predictive of socioeconomic outcomes, and with a set of artificial neural networks that extracted high-dimensional vector representations of images suitable for classification. Table 1 lists the manually-curated features with a description of

[1] Unlike the similarly-named *probability of poverty index*, the progress out of poverty index is positively correlated with wealth.



(a) Point-in-polygon match: reported GPS coordinate falls within a building footprint.

(b) GPS buffer match: buffering the reported GPS component by the recorded GPS uncertainty creates a polygon which uniquely intersects a building footprint.

(c) Scene match: the reported GPS coordinate and building footprint are the only geometries in the bounding box of the highest-resolution satellite tile.

Figure 2: Types of match conditions to assign survey responses to physical structures. The white dots represent the survey response's reported GPS coordinate, and the white dotted line is the boundary of a circle centered at the GPS coordinate with radius equal to the GPS recording's uncertainty. The red polygons are building footprints as reported by OpenBuildings, with building centroids represented by red dots.

(a) $z = 19$

(b) $z = 20$

(c) $z = 21$

Figure 3: Example images from the Google Static Maps API at the three zoom levels used, each centered at the same coordinates of a house in the sample.

their implementation. The featurization architectures investigated in this work are the ResNet (He et al., 2016) and ConvNeXt (Liu et al., 2022) networks.

Proxy measure	feature name	description
Crowdedness	`building_area`	building footprint size (MB)
	`building_area_wavg`	building footprint size (SWA)
	`num_buildings_in_scene`	total count of buildings in scene
	`min_NN_dist`	minimum distance to nearest neighboring building (MB)
	`min_NN_dist_wavg`	minimum distance to nearest neighboring building (SWA)
Dwelling characteristics	`aver_4NN_dist`	average distance to nearest four neighboring buildings (MB)
	`aver_4NN_dist_wavg`	average distance to nearest four neighboring buildings (SWA)
	`R_band`	spectral band: Red (MB)
	`G_band`	spectral band: Green (MB)
	`B_band`	spectral band: Blue (MB)
	`RGB`	RGB 16-bit composite (MB)
	`R_band_wavg`	spectral band: Red (SWA)
	`G_band_wavg`	spectral band: Green (SWA)
	`B_band_wavg`	spectral band: Blue (SWA)
	`RGB_wavg`	RGB 16-bit composite (SWA)
Urban planning	`angle_to_road`	footprint orientation with respect to nearest road (MB)
	`angle_to_road_wavg`	footprint orientation with respect to nearest road (SWA)
Road network	`distance_to_road`	proximity to nearest road (MB)
	`distance_to_road_wavg`	proximity to nearest road (SWA)
Other controls	`distance_to_point`	distance of scene building centroids to matched survey GPS coordinate
	`inv_d_weight`	inverse distance weight (centroid to matched point)

MB: Matched building; SWA: Scene weighted average. It refers to if the corresponding feature refers to the matched building in the scene, or is an inverse distance weighted average of the values of all the buildings in the scene. The weighting distance corresponds to the one from each building centroid to the GPS coordinate from the survey

Table 1: List of domain expert-curated features used in the explicit featurization approach.

After feature extraction, each set of features is run through principal component analysis to identify the most salient 5,000 feature dimensions, and then passed into a gradient boosting machine to predict household PPI scores. The total dataset was split into a training set (70% of dataset; N_{train} = 19,295), validation set (15% of the dataset; $N_{\text{validation}}$ = 4,134), and holdout set (the remaining 15% of the dataset; N_{test} = 4,134). The training set was used to identify patterns, while predictor hyperparameters were tuned by evaluation on the validation set. Final performance was reported by evaluation of mean-squared-error and R^2 on the test set. The hyperparameters for the gradient boosting machine searched over were the number of ensemble estimators, the learning rate, and the maximum tree depth of the underlying decision trees.

3 Preliminary Results

We present the prediction performance of gradient boosted machines on features from both the explicit featurization approach and deep learning approaches in Table 2. With the caveat that this work is still in its early stages, we find comparable explanatory performance of the two approaches, especially comparing prediction performance on the manual features to that of ConvNeXt-featurized images. However, we expect the smaller explicitly-curated feature space to have limited generalizability in other settings, with less homogenous label distributions, as we explain in our discussion section.

feature set	performance		optimal hyperparameters		
	R^2	MSE	# estimators	learning rate	maximum tree depth
explicit featurization	0.1199	121.09	100	0.1	3
ConvNeXt-featurized	0.1182	121.34	5000	0.01	none
ResNet18-featurized	0.0188	135.01	1000	0.01	1

Table 2: Performance of gradient boosting machine predictions on different feature sets, with corresponding hyperparameter choices.

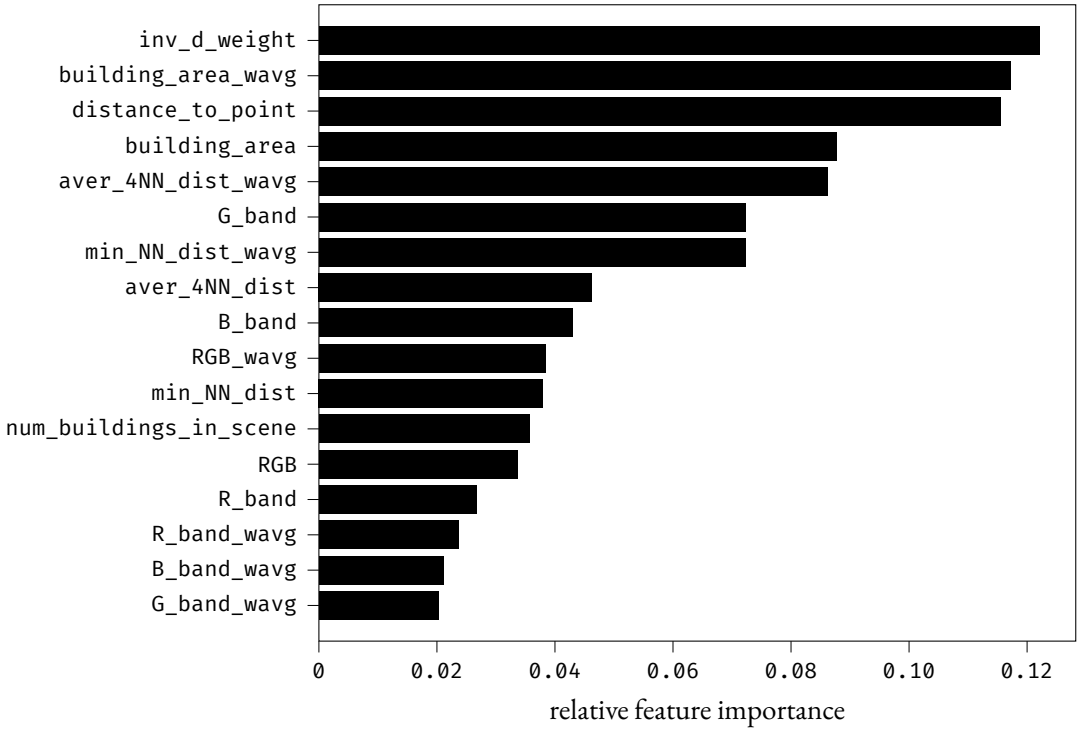


Figure 4: Relative importance in prediction of each manually-curated feature. Importance is measured by normalizing the information gain of each feature column by the total information gain across all estimators. Higher relative importance indicates more explanatory power.

For the explicit featurization, we also report the relative importance in explanatory power of each feature in Figure 4. Perhaps surprisingly, the inverse distance weighting, initially included as an intermediate calculation for other features, is the most predictive feature. We interpret the utility of this feature as reflecting a local building clustering density, thereby picking up the spatial distribution of the household's surrounding buildings. Other potential "neighborhood effects" are picked up by the weighted average of the area of the buildings in the vicinity of the household. The distance from the building centroid to the GPS survey response coordinate is also an important predictive feature, indicating the quality of GPS connection may be correlated with socioeconomic outcomes.

The top 5 most predictive features are, interestingly, derived entirely from vector geometry calculations as opposed to data derived from satellite imagery. The most predictive satellite imagery feature that was manually extracted is the concentration of green pixels in the image, which likely corresponds to the amount of agricultural land seen in a given satellite image tile.

4 Discussion & Future Work

This work represents the first systematic evaluation of whether household-level socioeconomic status for entire regions can be estimated from high-resolution satellite imagery. Leveraging recent, large-scale ground-truth survey data from Bangladesh, we evaluate how accurately measures of vulnerability can be estimated from satellite imagery at the household level.

Such granular estimates of poverty and vulnerability from satellite imagery facilitate development research and programming in contexts with low administrative capacity or limited budgets for survey-based data collection. These estimates may be useful in a number of downstream measurement tasks, such as estimating interim subnational poverty statistics, household-level targeting of humanitarian aid, determination of land value and tenurial rights, and the monitoring and evaluation of aid program efficacy.

There are three main directions in which we plan to expand this work in the future: (i) improving and refining the analysis pipeline; (ii) expanding our analysis to additional geographies; and (iii) policy simulations to explore potential use cases of household-level poverty estimates. To the first point, there are many margins on which we are actively working to improve the core results. These include the construction of additional manually curated features (e.g., distance to agricultural land, proximity to roads, shadow-based proxies for building height, and the number and areal density of buildings); the use of other machine learning architectures, such as vision transformers (Dosovitskiy et al., 2020) to automatically extract features; and additional fine-tuning of the deep learning models. To expand geographically, we have immediate plans to replicate our analysis from Bangladesh using nationally representative household survey data from Togo. In addition to geographic and cultural distinctions, the Togo survey data contain consumption and expenditures, and are representative of the entire nation (as opposed to a single district in Bangladesh), including urban and peri-urban regions. Finally, and more speculatively, we are interested in benchmarking one or two specific policy applications – such as the targeting of social assistance – to better understand the extent to which household-level poverty estimates can improve program effectiveness and efficiency.

More broadly, we hope this project can contribute to the literature on satellite-based poverty measurement in two key ways:

1. Create and evaluate novel methods for estimating household-level standards of living from satellite imagery at national scale. These methods rely on advances in the resolution of commercially-available satellite imagery to below 1 square meter per pixel, as well as novel deep learning architectures (such as transformers (Vaswani et al., 2017)) for identifying spatial characteristics of human settlements from imagery - leading to better building footprint identification and improved ability to make inferences about characteristics of the built environment (e.g. land use, road network completeness, nighttime luminosity).

2. Provide guidance to policymakers about the strengths and weaknesses of household-level poverty estimation techniques. This guidance covers which metrics (e.g. wealth vs. vulnerability vs. food security) can be reliably measured from satellite imagery, and at which scales (household, block, neighborhood) estimation is feasible, given data availability. We also provide guidance on expert-curated feature extraction versus fine-tuned deep learning models in poverty prediction. Critically, we expect the latter approach to provide more accurate estimates, but stakeholders often require more interpretable models compatible with the former approach. Our systematic comparison outlines the cost-benefit tradeoffs that users of these approaches face. We simulate anti-poverty program targeting from these household-level poverty estimates in Togo and Bangladesh, and compare these estimates to traditional poverty targeting approaches (e.g. proxy-means testing, community-based targeting).

references

Chi, G., H. Fang, S. Chatterjee, and J. E. Blumenstock (2022). Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences 119*(3), e2113658119.

CIESIN (2022). Global gridded relative deprivation index (grdi), version 1. =https://sedac.ciesin.columbia.edu/data/set/povmap-grdi-v1.

Dosovitskiy, A., L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Elbers, C., T. Fujii, P. Lanjouw, B. Özler, and W. Yin (2007). Poverty alleviation through geographic targeting: How much does disaggregation help? *Journal of Development Economics 83*(1), 198–213.

Engstrom, R., J. S. Hersh, and D. L. Newhouse (2017). Poverty from space: using high-resolution satellite imagery for estimating economic well-being. *World Bank Policy Research Working Paper 8284*.

Han, P., Q. Zhang, Y. Zhao, and F. Y. Li (2021). High-resolution remote sensing data can predict household poverty in pastoral areas, inner mongolia, china. *Geography and Sustainability 2*(4), 254–263.

He, K., X. Zhang, S. Ren, and J. Sun (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Hersh, J., R. Engstrom, and M. Mann (2021). Open data for algorithms: mapping poverty in belize using open satellite derived features and machine learning. *Information Technology for Development 27* (2), 263–292.

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

Jerven, M. (2013). *Poor numbers: how we are misled by African development statistics and what to do about it*. Cornell University Press.

Knebelmann, J., V. Pouliquen, and B. Sarr (2023). Discretion versus algorithms: Bureaucrats and tax equity in senegal. Technical report, Working Paper.[6].

Liu, Z., H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986.

Rolf, E., J. Proctor, T. Carleton, I. Bolliger, V. Shankar, M. Ishihara, B. Recht, and S. Hsiang (2021). A generalizable and accessible approach to machine learning with global satellite imagery. *Nature communications 12*(1), 4392.

Sirko, W., S. Kashubin, M. Ritter, A. Annkah, Y. S. E. Bouchareb, Y. Dauphin, D. Keysers, M. Neumann, M. Cisse, and J. Quinn (2021). Continental-scale building detection from high resolution satellite imagery. *arXiv preprint arXiv:2107.12283*.

Smythe, I. S. and J. E. Blumenstock (2022). Geographic microtargeting of social assistance with high-resolution poverty maps. *Proceedings of the National Academy of Sciences 119*(32), e2120025119.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin (2017). Attention is all you need. *Advances in neural information processing systems 30*.

Watmough, G. R., C. L. Marcinko, C. Sullivan, K. Tschirhart, P. K. Mutuo, C. A. Palm, and J.-C. Svenning (2019). Socioecologically informed use of remote sensing data to predict rural household poverty. *Proceedings of the National Academy of Sciences 116*(4), 1213–1218.

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