

# LEVERAGING SATELLITE IMAGERY FOR CHILDHOOD POVERTY ESTIMATION

**Fan Yang\***, **Makkunda Sharma\***, **Duy-Nhat Vo**, **Matthew Sutcliffe**

Department of Computer Science, University of Oxford  
Oxford, UK

{luke.yang, makkunda.sharma, duy.vo, matthew.sutcliffe}@cs.ox.ac.uk

**Mengyan Zhang**, **William Rudgard**, **Seth Flaxman**<sup>†</sup>

University of Oxford  
Oxford, UK

{mengyan.zhang, william.rudgard, seth.flaxman}@cs.ox.ac.uk

**Jack Gidney**, **H Juliette T Unwin**

University of Bristol  
Bristol, UK

{db21148, juliette.unwin}@bristol.ac.uk

**Esra Suel**

CASA, University College London  
London, UK

e.suel@ucl.ac.uk

**Swapnil Mishra**

Saw Swee Hock School of Public Health, National University of Singapore  
Singapore

swapnilmishra.anu@gmail.com

**Samir Bhatt**

Section of Epidemiology, University of Copenhagen  
Copenhagen, Denmark

samir.bhatt@sund.ku.dk

**Oliver Fiala**

Save the Children

O.Fiala@savethechildren.org.uk

## ABSTRACT

Methods using satellite imagery have been increasingly explored for analyzing demographic, health, and development indicators. This paper introduces a new dataset that pairs satellite imagery with high-quality survey data to benchmark state-of-the-art computer vision methods targeting childhood poverty estimation. The dataset includes 33,608 images from 16 countries in Eastern and Southern Africa between 1997 and 2022, integrating six childhood poverty indicators (UNICEF, 2021) derived from Demographic and Health Surveys (DHS). Baseline approaches based on band statistics and spectral indices are compared against deep learning foundation models (e.g., DINOv2 (Oquab et al., 2023) and SatMAE (Cong et al., 2022)), demonstrating that classical methods remain strong baselines while deep learning vision models with high-resolution input further improve accuracy. Open-source code is provided to reproduce and extend the pipeline, including dataset construction and model comparisons.

## 1 INTRODUCTION

Satellites like Landsat and Sentinel regularly capture and host updated, publicly available, high-resolution imagery every 1–2 weeks, providing valuable data for applications in agriculture, health, development, and disaster response. Recent work demonstrates how satellite imagery can be used

---

\*Both authors contributed equally to this research.

<sup>†</sup>Corresponding author: seth.flaxman@cs.ox.ac.uk

to estimate the causal impact of electricity access on livelihoods (Ratledge et al., 2021), measure urban income and overcrowding (Suel et al., 2021), and predict population in the absence of census data (Wardrop et al., 2018). Of the 8 billion people globally, over 2 billion are children (aged <18 as defined in the UN Convention on the Rights of the Child (UN General Assembly, 1989)), whose poverty differs from that of adults due to distinct growth, health, and education needs (Brooks-Gunn & Duncan, 1997). These specific needs are inferred beyond household-level wealth or resources, as households may be unequal internally or fail to address the requirements of children. Consequently, UNICEF (2021) emphasize the need to measure child poverty at the level of children themselves, with a focus on the “constitutive rights of poverty” that facilitate cross-country comparisons.

In this work, we introduce **KidSat**, a new dataset designed for advanced machine learning methods targeting multidimensional child poverty estimation. KidSat pairs geo-coded survey data on child poverty indicators from 16 countries in Eastern and Southern Africa with the associated multispectral satellite imagery, offering a standardized framework to benchmark and refine approaches in computer vision for childhood poverty monitoring. Our contributions are as follows:

- We aggregate geo-coded DHS data on multidimensional child poverty alongside corresponding satellite imagery. The evaluation benchmarks are provided on a univariate measure (the percentage of children experiencing deprivation, from 0% to 100%) and its sub-indicators, making model outputs directly interpretable by policymakers.
- We evaluate a range of methods, from baselines leveraging spatial correlation to satellite-based foundational vision models. Notably, our experiments underscore the importance of high-resolution imagery and the capability of foundational vision models to effectively process such detailed inputs.

## 2 RELATED WORK

Publicly available high-resolution satellite imagery from Landsat and Sentinel has facilitated the creation of numerous datasets for machine learning tasks. Well-known examples include functional map of the world (fMoW) (Christie et al., 2018), EuroSAT (Helber et al., 2019), XView (Lam et al., 2018), Spacenet (Van Etten et al., 2018), and Floodnet (Rahnemoonfar et al., 2021), all of which focus primarily on object detection, segmentation, or similar computer vision-centric objectives. In the field of social science, Yeh et al. (2021) introduced SustainBench to systematically link remote-sensing imagery with broader development metrics such as health, wealth, and clean water indices; it remains the only standardized dataset designed for socio-economic prediction tasks to date.

Beyond these generic datasets, the application of machine learning to satellite images for demographic, health, and development indicators has gathered increasing attention. Jean et al. (2016) pioneered the use of convolutional neural networks (CNNs) to estimate human development levels at finer spatial and temporal resolutions. Subsequent research has leveraged satellite imagery to infer both directly observable indicators like deforestation and agricultural yields (Ball et al., 2022; Estes et al., 2022; Xu et al., 2024) and more abstract metrics such as poverty rates (Ayush et al., 2021), health conditions (Daoud et al., 2023), or composite measures of development (Sherman et al., 2023), especially in regions with sparse on-ground data.

Progress in self-supervised learning (He et al., 2022; Caron et al., 2021) has driven the emergence of large-scale foundation models, which are pre-trained on vast, unlabeled datasets and later fine-tuned for specific tasks. In computer vision, this paradigm is demonstrated by SimCLR (Chen et al., 2020), CLIP (Radford et al., 2021), and DINO (Caron et al., 2021). More recently, satellite-specific foundation models have appeared, including SatMAE (Cong et al., 2022) (based on masked autoencoders), SatCLIP (Klemmer et al., 2023) (adapting CLIP to satellite imagery), and DiffusionSat (Khanna et al., 2024) (a diffusion-based generator (Rombach et al., 2022)). Whether these specialized models consistently outperform general-purpose foundation models on smaller or more specialized datasets remains an open question.

## 3 KIDSAT DATASET

KidSat is a novel dataset that pairs high-resolution satellite imagery with summarized demographic and health data from the DHS Program in Eastern and Southern Africa. By leveraging the DHS’s

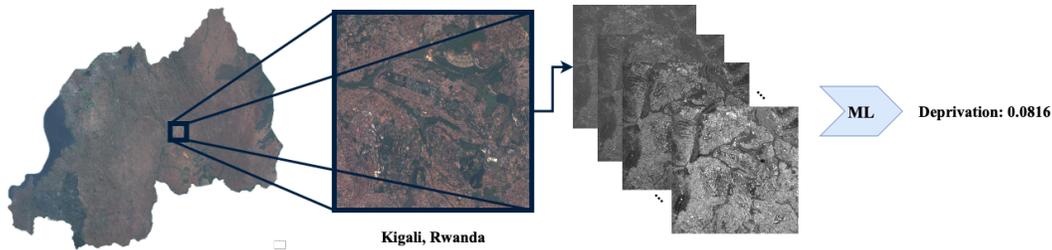


Figure 1: Pipeline for predicting child poverty from satellite images. Each cluster’s location is used to extract a  $10\text{km}\times 10\text{km}$  patch, which is processed by baseline or transformer-based models to predict the `severe_deprivation` measure.

rigorous sampling methods and geo-located cluster information, KidSat offers a unique opportunity to apply advanced machine learning approaches for estimating socio-economic and health indicators in under-surveyed areas. Basic dataset statistics are provided in Appendix A.1.

The DHS Program (UN General Assembly, 1989) has conducted over 400 nationally representative household surveys in 90 countries since 1984 using a stratified two-stage cluster design in which enumeration areas (EAs) serve as the primary sampling unit. Anonymized cluster coordinates (with random jitter) facilitate the mapping of survey data to precise locations. Child poverty indicators in our dataset draw on UNICEF’s (UNICEF, 2021) multidimensional framework, which encompasses six dimensions of deprivation (housing, water, sanitation, nutrition, health, and education). A child is considered severely deprived if they are severely deprived in at least one dimension. We aggregate these outcomes at the cluster level, forming our primary target variable, `severe_deprivation`. Detailed variable definitions of childhood poverty indicators appear in Appendix A.1.2.

To build the satellite imagery component, we utilize Sentinel-2 and Landsat (5, 7, and 8) data for their public availability, large-scale coverage, and suitability for vision-based analyses. Each EA is paired with a  $10\text{km}\times 10\text{km}$  image patch obtained via Google Earth Engine, filtered by survey year and low cloud coverage ( $\leq 20\%$ ). Median compositing is applied to minimize cloud artifacts. Sentinel-2 images yield a  $1000\times 1000$  pixel patch, while Landsat results in  $333\times 333$  pixels for the same spatial coverage, encompassing both RGB and extended spectral bands (Appendix A.1.1).

## 4 BENCHMARK AND RESULTS

### 4.1 EXPERIMENTAL DESIGN

We evaluate child poverty estimation under two distinct settings: a *spatial* and a *temporal* benchmark. For the spatial task, we use 5-fold **spatial cross-validation** at the cluster level, training on 80% of the clusters and holding out the remaining 20%. This setup tests the model’s ability to generalize to unsurveyed locations. For the temporal task, we use **historical data** from before 2019 (inclusive) to predict child poverty for the years 2020–2022, thus assessing the model’s capability to forecast future conditions based on prior satellite imagery.

### 4.2 BENCHMARK METHODS

We compare baseline methods and advanced computer vision techniques, each representing a different strategy for processing satellite imagery:

- **Baseline:** *Mean Regression* always predicts the mean training-set value. *Gaussian Process Regression* (GPR) uses only geographic coordinates, exploiting spatial proximity in the target. Additionally, we combine spectral band statistics (mean and standard deviation) and common remote-sensing indices (e.g. normalized difference vegetation index, normalized difference water index) in a *Random Forest Regression* (RFR) model.
- **MOSAICS** (Rolf et al., 2021): A framework that derives global feature embeddings from satellite imagery. We feed these embeddings into a regression model for poverty prediction.

Table 1: Comparison of various baseline methods, random forest regressions, and Transformer-based models (DINOv2 and SatMAE) for predicting *severe\_deprivation*. Performance is reported as MAE (with standard errors) on two benchmarks: a *spatial* split (where random clusters are held out) and a *temporal* split (where the years 2020–2022 are held out). Model inputs range from precomputed feature sets (e.g., band statistics, MOSAIKS features) to RGB/multispectral Landsat and Sentinel imagery at varying dimensions.

Type	Model	Model Input	Spatial	Temporal
Baseline	Mean Regression	-	$0.2930 \pm 0.0018$	0.3183
	GPR	Coordinates	$0.2436 \pm 0.0002$	0.5656
RGB	RFR	Spectral Stats	$0.2184 \pm 0.0011$	0.2671
	MOSAIKS	MOSAIKS Features	$0.2356 \pm 0.0114$	0.2588
	DINOv2 (Raw)	Landsat Imagery	$0.2260 \pm 0.0005$	0.2704
	DINOv2 (Raw)	Sentinel-2 Imagery	$0.2013 \pm 0.0019$	0.2597
	DINOv2 (Fine-tuned)	Landsat Imagery	$0.2042 \pm 0.0015$	0.2574
	DINOv2 (Fine-tuned)	Sentinel-2 Imagery	$0.1663 \pm 0.0023$	0.2858
	SatMAE (Raw)	Landsat Imagery	$0.2341 \pm 0.0017$	0.3453
	SatMAE (Raw)	Sentinel-2 Imagery	$0.2347 \pm 0.0027$	0.3067
	SatMAE (Fine-tuned)	Landsat Imagery	$0.2125 \pm 0.0019$	0.3376
	SatMAE (Fine-tuned)	Sentinel-2 Imagery	$0.2093 \pm 0.0039$	0.3139
Multispectral	RFR	Spectral Stats + Indices	$0.1924 \pm 0.0012$	0.2685
	SatMAE (Raw)	Landsat Imagery	$0.2325 \pm 0.0014$	0.3783
	SatMAE (Raw)	Sentinel-2 Imagery	$0.2322 \pm 0.0024$	0.4230
	SatMAE (Fine-tuned)	Landsat Imagery	$0.2001 \pm 0.0017$	0.3827
	SatMAE (Fine-tuned)	Sentinel-2 Imagery	$0.1993 \pm 0.0015$	0.3931

- **DINOv2** (Oquab et al., 2023): A self-supervised vision transformer (ViT) model pre-trained on RGB images. We evaluate it *as-is* (Raw) and also *fine-tune* it on 17 DHS variables; a separate ridge regression layer then maps extracted features to childhood poverty.
- **SatMAE** (Cong et al., 2022): A masked autoencoder tailored for satellite imagery, evaluated in its *Raw* form and with *fine-tuning* on DHS variables. We test RGB-only and multispectral variants with both Landsat and Sentinel images.

### 4.3 RESULTS AND DISCUSSION

Table 1 summarizes mean absolute error (MAE) on both spatial and temporal benchmarks. The spatial benchmark indicates clear benefits from incorporating satellite imagery, with Random Forest (RGB) ( $0.2184 \pm 0.0011$ ) and baseline models outperforming simple geographic GPR. Adding multispectral statistics (Random Forest Multispectral) further reduces error ( $0.1924 \pm 0.0012$ ). Transformer-based methods, particularly DINOv2 with high-resolution Sentinel-2 imagery, achieve the best spatial performance when fine-tuned ( $0.1663 \pm 0.0023$ ). This highlights the value of detailed spatial information and the strong feature extraction capacity of large ViT models.

Temporal predictions prove more challenging, as models must infer time-related changes solely from static images. While raw transformer-based features (e.g., DINOv2 Raw) can outperform baselines, the fine-tuned versions sometimes degrade in temporal performance, possibly overfitting historical data. Notably, GPR performs poorly (0.5656), reflecting its inability to capture temporal shifts without explicit time signals. Overall, the results underscore the difficulty of forecasting with limited temporal cues but confirm the promise of deep vision architectures for spatial generalization.

### 4.4 INTERPRETATION AND CONCLUSION

An MAE of 20 indicates a deviation of 20 percentage points in estimating the share of children experiencing severe deprivation, which would require further efforts and improvement for practical usage. In the *spatial* task, combining even simple spectral statistics with a random forest regressor can yield clear improvements over naive baselines and coordinate-based GPR. Further gains are

observed with large-scale vision transformers, when leveraging high-resolution satellite imagery and fine-tuning on relevant DHS variables (e.g., DINOv2-ViT on Sentinel).

*Temporal* prediction, where models attempt to forecast poverty in later years, is inherently more difficult. Here, performance often degrades compared to spatial prediction, presumably because historical data alone do not capture evolving local conditions. Fine-tuning on time-stamped data can risk overfitting, particularly if year-over-year changes are not well-represented in the training set. Overall, this benchmark underscores the complexity of forecasting real-world socio-economic trends from satellite imagery without direct temporal signals.

DINOv2 emerges as the strongest performer in the spatial benchmark, likely due to its ability to exploit higher-resolution RGB data. SatMAE, although pre-trained on satellite imagery, is constrained by a lower input resolution ( $224 \times 224$ ), limiting its performance relative to DINOv2’s larger patches (e.g.,  $1000 \times 1000$  in Sentinel-2). In the temporal task, models frequently struggle to capture changing conditions, suggesting that fine-tuning on historical data alone may induce overfitting. While multi- or hyperspectral data can strengthen performance, these additional bands are underutilized in the current best-performing DINOv2 model, pointing to directions for multispectral foundational models capable of processing high-resolution inputs.

## REFERENCES

- Kumar Ayush, Burak Uz kent, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. Efficient poverty mapping from high resolution remote sensing images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 12–20, 2021.
- James GC Ball, Katerina Petrova, David A Coomes, and Seth Flaxman. Using deep convolutional neural networks to forecast spatial patterns of amazonian deforestation. *Methods in Ecology and Evolution*, 13(11):2622–2634, 2022.
- Jeanne Brooks-Gunn and Greg J Duncan. The effects of poverty on children. *The future of children*, pp. 55–71, 1997.
- Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.
- Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.
- Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6172–6180, 2018.
- Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David Lobell, and Stefano Ermon. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35:197–211, 2022.
- Adel Daoud, Felipe Jordán, Makkunda Sharma, Fredrik Johansson, Devdatt Dubhashi, Sourabh Paul, and Subhashis Banerjee. Using satellite images and deep learning to measure health and living standards in india. *Social Indicators Research*, 167(1):475–505, 2023.
- Lyndon D Estes, Su Ye, Lei Song, Boka Luo, J Ronald Eastman, Zhenhua Meng, Qi Zhang, Dennis McRitchie, Stephanie R Debats, Justus Muhando, et al. High resolution, annual maps of field boundaries for smallholder-dominated croplands at national scales. *Frontiers in artificial intelligence*, 4:744863, 2022.
- Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

- Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019. doi: 10.1109/JSTARS.2019.2918242.
- 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.
- Samar Khanna, Patrick Liu, Linqi Zhou, Chenlin Meng, Robin Rombach, Marshall Burke, David B. Lobell, and Stefano Ermon. Diffusionsat: A generative foundation model for satellite imagery. In *The Twelfth International Conference on Learning Representations*, 2024.
- Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. Satclip: Global, general-purpose location embeddings with satellite imagery. *arXiv preprint arXiv:2311.17179*, 2023.
- Darius Lam, Richard Kuzma, Kevin McGee, Samuel Dooley, Michael Laielli, Matthew Klaric, Yaroslav Bulatov, and Brendan McCord. xview: Objects in context in overhead imagery. *arXiv preprint arXiv:1802.07856*, 2018.
- Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR, 2021.
- Maryam Rahnemoonfar, Tashnim Chowdhury, Argho Sarkar, Debvrat Varshney, Masoud Yari, and Robin Roberson Murphy. Floodnet: A high resolution aerial imagery dataset for post flood scene understanding. *IEEE Access*, 9:89644–89654, 2021.
- Nathan Ratledge, Gabriel Cadamuro, Brandon De la Cuesta, Matthieu Stigler, and Marshall Burke. Using satellite imagery and machine learning to estimate the livelihood impact of electricity access. Technical report, National Bureau of Economic Research, 2021.
- 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):4392, 2021.
- Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.
- Luke Sherman, Jonathan Proctor, Hannah Druckenmiller, Heriberto Tapia, and Solomon M Hsiang. Global high-resolution estimates of the united nations human development index using satellite imagery and machine-learning. Technical report, National Bureau of Economic Research, 2023.
- Esra Suel, Samir Bhatt, Michael Brauer, Seth Flaxman, and Majid Ezzati. Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. *Remote Sensing of Environment*, 257:112339, 2021.
- UN General Assembly. Convention on the rights of the child. *United Nations, Treaty Series*, 1577(3):1–23, 1989.
- UNICEF. Child poverty profiles: Understanding internationally comparable estimates, 2021. URL <https://data.unicef.org/resources/child-poverty-profiles-understanding-internationally-comparable-estimates/>.
- Adam Van Etten, Dave Lindenbaum, and Todd M Bacastow. Spacenet: A remote sensing dataset and challenge series. *arXiv preprint arXiv:1807.01232*, 2018.

NA Wardrop, WC Jochem, TJ Bird, HR Chamberlain, Donna Clarke, David Kerr, Linus Bengtsson, Sabrina Juran, Vincent Seaman, and AJ Tatem. Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences*, 115(14):3529–3537, 2018.

Jonathan Xu, Amna Elmustafa, Liya Weldegebriel, Emnet Negash, Richard Lee, Chenlin Meng, Stefano Ermon, and David Lobell. Harvestnet: A dataset for detecting smallholder farming activity using harvest piles and remote sensing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 22438–22446, 2024.

Christopher Yeh, Chenlin Meng, Sherrie Wang, Anne Driscoll, Erik Rozi, Patrick Liu, Jihyeon Lee, Marshall Burke, David B Lobell, and Stefano Ermon. Sustainbench: Benchmarks for monitoring the sustainable development goals with machine learning. *arXiv preprint arXiv:2111.04724*, 2021.

## A APPENDIX

### A.1 DATASET DESCRIPTION

The KidSat dataset we present in this work includes both cluster-wise child poverty derived from the DHS data and the satellite imagery corresponding to each cluster. Due to the confidentiality of the survey data, DHS requires registration prior to accessing the data. We include detailed procedures for acquiring the satellite imagery and DHS data in our (anonymous) GitHub repository.

Our dataset consists of 33,608 images, each covering a  $10 \text{ km} \times 10 \text{ km}$  region from 16 countries in Eastern and Southern Africa between 1997 and 2022. Each image is paired with a summarized DHS data entry, linked via the geographic coordinates from the DHS record. Figure 2 shows the spatial distribution of the DHS cluster locations included in the KidSat dataset.

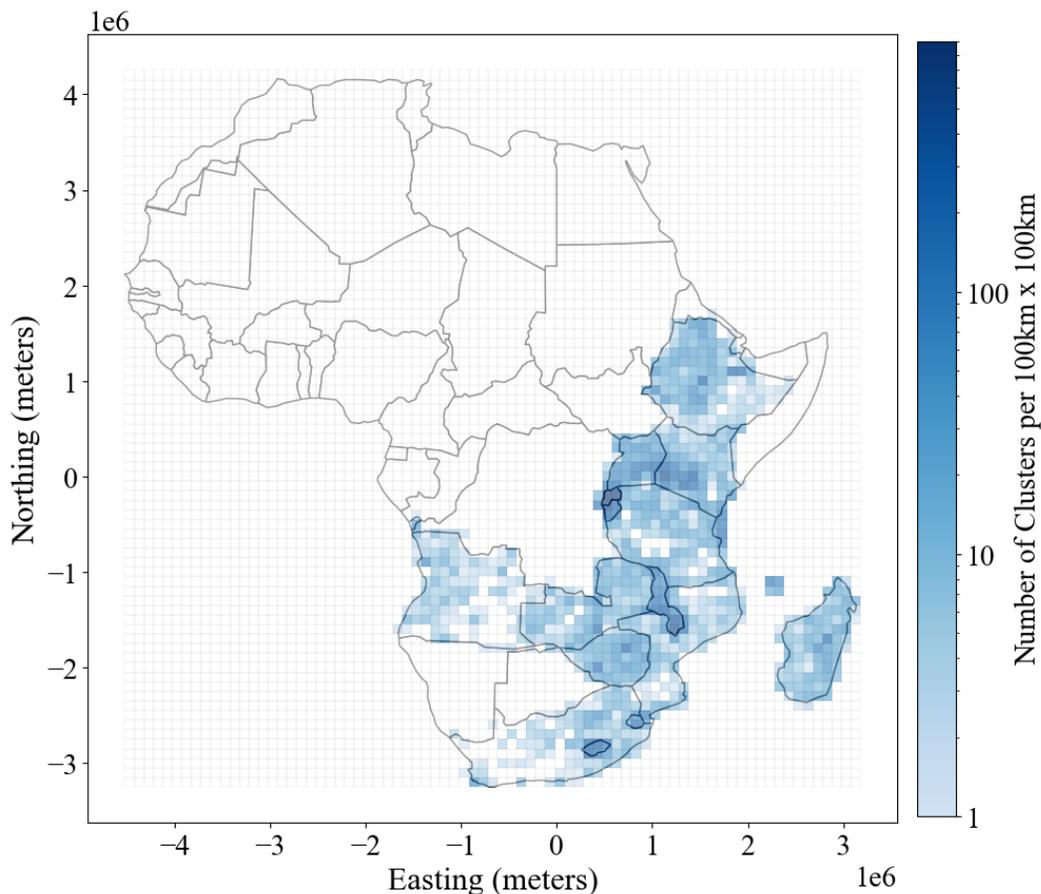


Figure 2: Heatmap illustrating the spatial density of DHS cluster locations in Eastern and Southern Africa. The color represents the count of cluster locations per  $100 \text{ km} \times 100 \text{ km}$  grid cell.

#### A.1.1 IMAGERY STATISTICS

**Sentinel-2** is a mission from the European Space Agency (ESA), part of the Copernicus Programme. It consists of two satellites (Sentinel-2A and Sentinel-2B) and provides imagery in 13 spectral bands shown in Table 2.

#### Key Statistics for Sentinel-2:

- **Spatial Resolution:** 10 m, 20 m, and 60 m depending on the band.
- **Temporal Resolution:** 5 days revisit time at the equator (with two satellites).

Band Name	Band Number	Central Wavelength (nm)	Resolution (m)
Coastal Aerosol	1	443	60
Blue	2	494	10
Green	3	560	10
Red	4	665	10
Red Edge 1	5	703	20
Red Edge 2	6	740	20
Red Edge 3	7	782	20
NIR (Near Infrared)	8	835	10
NIR Narrow	8A	864	20
Water Vapour	9	945	60
SWIR 1	11	1610	20
SWIR 2	12	2190	20
Cirrus	10	1375	60

Table 2: Sentinel-2 Band Information

- **Spectral Range:** 13 spectral bands, ranging from visible light (Blue, Green, Red) to infrared (NIR and SWIR).
- **Coverage:** Global, with a swath width of 290 km.
- **Radiometric Resolution:** 12-bit data (values range from 0 to 4096).

**Landsat 5, 7, and 8** are parts of a long-running Earth observation mission managed by NASA and the U.S. Geological Survey (USGS). Here we provide band information in Table 3.

Band Name	Landsat 5	Landsat 7	Landsat 8	Wavelength (nm)
Coastal Aerosol	-	-	1	433–453
Blue	1	1	2	450–520
Green	2	2	3	520–600
Red	3	3	4	630–670
NIR (Near Infrared)	4	4	5	850–880
SWIR 1	5	5	6	1550–1750
SWIR 2	7	7	7	2080–2350
Panchromatic	-	8	8	500–680
Cirrus	-	-	9	1360–1380
Thermal Infrared 1	6	6	10	10400–12500
Thermal Infrared 2	-	-	11	10400–12500

Table 3: Landsat 5, 7, and 8 Band Information

#### Key Statistics for Landsat (Landsat 5, 7 & 8):

- **Spatial Resolution:** 30 m for multispectral bands, 15 m for panchromatic, 100 m for thermal bands (resampled to 30 m).
- **Temporal Resolution:** 16 days revisit time.
- **Spectral Range:** 7 bands for Landsat 5, 8 bands for Landsat 7, and 11 bands for Landsat 8, spanning visible, infrared, and thermal wavelengths.
- **Coverage:** Global, with a swath width of 185 km.
- **Radiometric Resolution:** 16-bit data (values range from 0 to 65536).

We follow the conventional approach used in the Google Earth Engine for imagery normalization. To preprocess the Sentinel-2 imagery, we normalize the pixel values by scaling the original range of 0 to 3000 to a range of 0 to 255. Values outside this range are clipped. This method preserves the relative intensity of the pixel values while adapting the data for image rendering. For Landsat imagery, the pixel values are normalized from 0 to 30000 to a range of 0 to 255. Similarly, values

outside this range are clipped to ensure that they conform to the appropriate visualization range. We used the raw imagery data from both Sentinel-2 and Landsat datasets before the normalization and pre-processing.

### A.1.2 CHILDHOOD POVERTY

The `severe_deprivation` variable is used in this work to represent the percentage of children experiencing severe poverty for individual responses within the cluster. It is calculated by aggregating several indicators of severe deprivation across multiple dimensions such as housing, water, sanitation, nutrition, health, and education. The detailed definition of severe deprivation can be found in Table 4. Note that a child is classified in severe deprivation if they experience severe deprivation in any of the dimensions.

In addition, deprivation in each subcategory, as well as `moderate_deprivation`, is also included in the dataset. Further definitions can be found in the work by UNICEF (2021).

We present the histograms of the variable `severe_deprivation`, faceted by country, in Figure 3. The distributions of `severe_deprivation` vary significantly across countries. Most countries exhibit right-skewed distributions, with exceptions such as Malawi and Zimbabwe, which show left-skewed distributions. Additionally, some countries display Gaussian-like distributions (e.g., Rwanda), while others show U-shaped patterns (e.g., Tanzania). Given the variation in distribution across countries, spatial modeling for all of Eastern and Southern Africa poses a considerable challenge. For both optimization and policy-making purposes, country-specific modeling could improve the applicability and effectiveness of this approach.

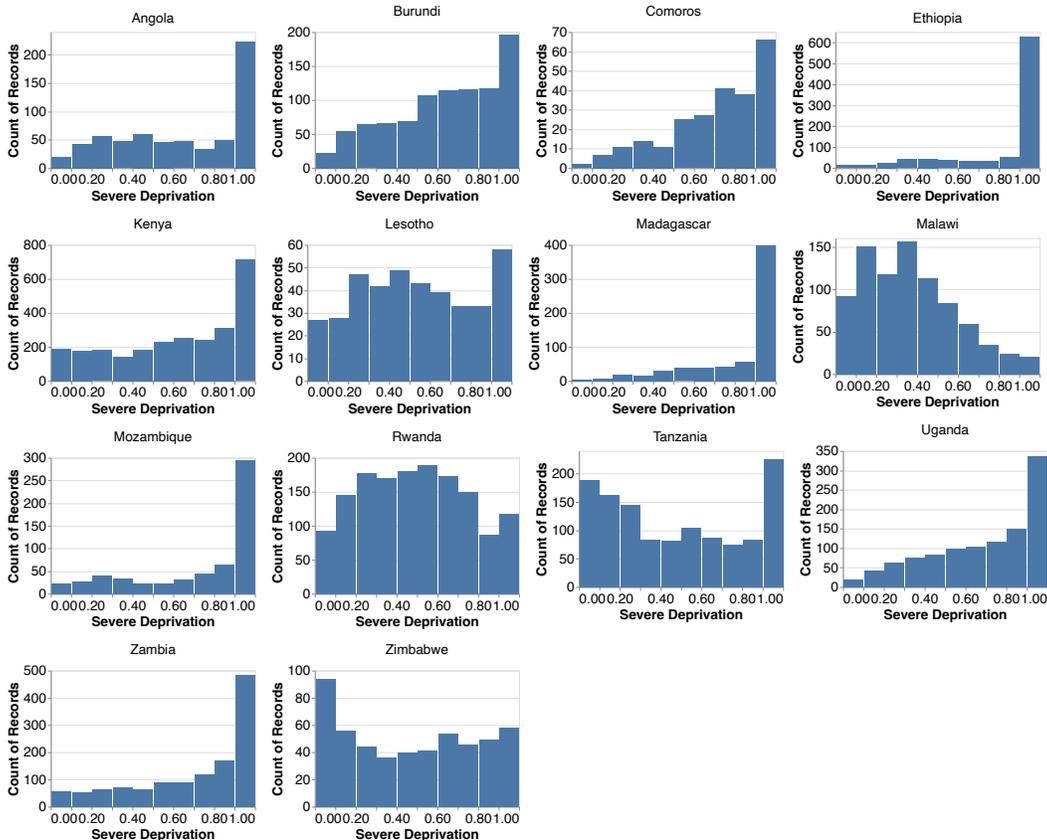


Figure 3: Histogram of the target severe deprivation variable, faceted by country.

Among all DHS variables used in child poverty calculation, we selected 17 variables, as presented in Table 5, as the prediction targets during model fine-tuning. Continuous variables were scaled to the range  $[0, 1]$ , and categorical variables were expanded using one-hot encoding, where each category

<b>Dimension</b>	<b>Unit of Analysis</b>	<b>Severe Deprivation Definition</b>	<b>Moderate Deprivation Definition</b>
<b>Housing</b>	Children under 17 years of age	Children living in a dwelling with five or more persons per sleeping room.	Children living in a dwelling with three or more persons per sleeping room.
<b>Sanitation</b>	Children under 17 years of age	Children with no access to a toilet facility of any kind (i.e., open defecation, pit latrines without slabs, hanging latrines, bucket latrines, etc.).	Children using improved facilities but shared with other households.
<b>Water</b>	Children under 17 years of age	Children with no access to water facilities of any kind (i.e., using surface water or unimproved supplies).	Children using improved water sources but located more than 15 minutes away (30 minutes roundtrip).
<b>Nutrition</b>	Children under 5 years of age	Stunting (3 standard deviations below the international reference population).	Stunting (2 standard deviations below the international reference population).
<b>Education</b>	Children between 5–14 years of age	Children who have never been to school.	Children who are not currently attending school.
	Children between 15–17 years of age	Children who have not completed primary school.	Children who are not currently attending secondary school (or did not complete secondary school).
<b>Health</b>	Children 12–35 months old	Children who did not receive immunization against measles nor any dose of DPT.	Children who received fewer than four vaccines (measles plus three rounds of DPT).
	Children 36–59 months old	Children with severe cough and fever who received no treatment of any kind.	Children with severe cough and fever who did not receive professional medical treatment.
	Children 15–17 years old	Unmet contraception needs.	Unmet contraception needs.

Table 4: Severe and moderate deprivation definitions by dimension and unit of analysis. Table adapted from UNICEF (2021).

was represented by a binary indicator. This resulted in a 99-dimensional vector representing each cluster, based on the 17 selected DHS variables. We then used this vector to map satellite imagery for prediction and update as part of the model fine-tuning process.

Table 5: This table categorizes various Demographic and Health Survey (DHS) variables by their respective child deprivation categories. The categories include Water, Sanitation, Nutrition, Health, Education, and Housing. Each category lists specific variables and their descriptions relevant to assessing child deprivation.

Deprivation Category	Description	Variable
Water	Main drinking water source	hv201
	Time to water source	hv204
Sanitation	Type of toilet facility	hv205
	Toilet sharing status	hv225
Nutrition	Height-for-age z-score	hc70
Health	Child received any vaccination	h10
	DPT 1 vaccination	h3
	DPT 2 vaccination	h5
	DPT 3 vaccination	h7
	Measles 1 vaccination	h9
	Child had cough recently	h31
Education	Current contraceptive method	v312
	Highest education level in household	hv106
	Educational attainment recoded	hv109
Housing	School attendance current year	hv121
	Sleeping rooms in household	hv216
	Wealth index score	hv271