From Space to Settlements: Earth Observations for support prioritization and monitoring

Steven Ndung'u[(1)], Adel Daoud[(1)(2)], Giulia Del Panta[(1)] and Andrea Pellandra[(1)]

(1) the United Nations High Commissioner for Refugees, Denmark; e-mail: ndungust@unhcr.org
(2) Linköping University, Sweden

Abstract

Timely socio-economic data on forcibly displaced persons (FDPs) remains critically scarce, constraining humanitarian prioritization and intervention monitoring. While household surveys provide detailed information, their resource-intensive nature creates substantial temporal and geographic gaps - particularly problematic given the 123.2 million FDPs worldwide by 2024. This research introduces a novel Earth Observation and Machine Learning (EO-ML) framework to address this gap through transfer learning. By adapting a foundation model pretrained on Demographic and Health Survey (DHS) data from 1.2 million households across 36 African countries to refugee and host community contexts in South Sudan, we successfully bridge the domain gap between general population surveys and displacement settings. Using UNHCR's Forced Displacement Survey (FDS) and Results Monitoring Survey (RMS) data, we constructed a harmonized socio-economic index and employed spatially stratified cross-validation across 28 administrative regions. Preliminary results show that the best-performing model achieves a mean absolute error of 5.37 on a 0–100 scale and R^2 of 0.49. These findings provide a proof of concept for leveraging EO data to fill critical information gaps between survey rounds, unlocking the potential of World Bank and UNHCR microdata libraries for humanitarian prioritization.

1 Introduction

The United Nations High Commissioner for Refugees (UNHCR) estimated 123.2 million forcibly displaced persons (FDPs) worldwide in 2024, a 6% increase compared to 2023 (approximately 7 million people) and representing a global doubling in displacement numbers over the last decade[1]. FDPs comprise refugees, asylum seekers, and internally displaced persons. They are characterized by acute poverty, limited access to healthcare, and heightened vulnerability, often having lost their financial assets, livelihoods, and social networks [1, 2]. Compounding these challenges, many reside in remote and overcrowded settlements, frequently subject to legal and institutional constraints that restrict mobility and access to formal labor markets and education [3]. Consequently, they depend heavily on humanitarian assistance, particularly in low- and middle-income host countries facing substantial fiscal and institutional pressures [4]. Therefore, there is a critical need for frequent monitoring and targeted resource prioritization to address the evolving situations of FDPs effectively.

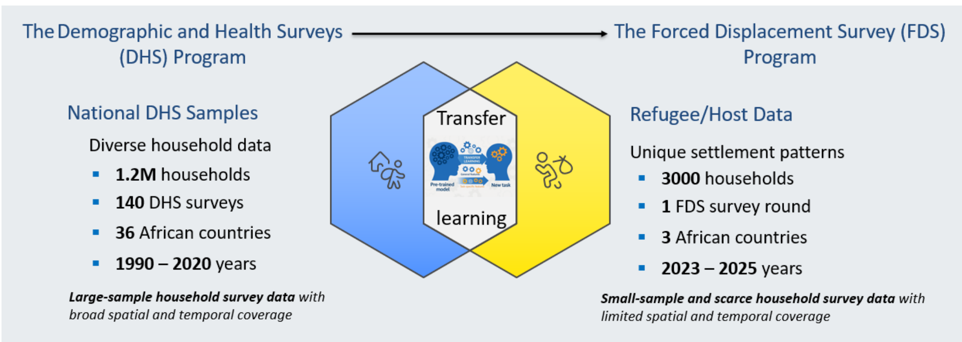


Figure 1: Domain adaptation strategy. The transfer learning model bridges the data gap by pretraining on the rich DHS data, spanning 1.2 million households across 36 countries (data collected in 30 years), before fine tuning on the data scarce FDS. This approach leverages broad spatial and temporal features (knowledge learned) from national samples to enhance predictive performance in localized, unique refugee settlement contexts.

The lack of timely, granular socio-economic data on FDPs and host populations constrains the design and implementation of effective support programs. While household surveys provide detailed information, their

[1] https://www.unhcr.org/global-trends

resource-intensive nature creates substantial temporal and geographic gaps in data coverage, particularly with regard to household indicators that can change quickly. These gaps in household indicators limit humanitarian and development actors' ability to understand evolving needs, target assistance effectively, and monitor the impact of interventions on living conditions. These gaps affect all refugee settlements, including camps, urban areas, and mixed settlements. UNHCR collects many survey data on FDPs, such as FDS data (FDS is a comprehensive survey providing household indicators for people forced to flee). While it is aimed to be more frequent (every 2–5 years) than a typical 10 year census, significant changes can occur between these survey rounds. New displacement events, such as the April 2023 Sudan crisis leading to additional refugee flows into South Sudan, create rapidly evolving situations that are not always captured by scheduled surveys. This creates an urgent need for methods to generate reliable socioeconomic estimates between survey rounds, particularly in contexts where sudden population movements or crises alter the humanitarian landscape.

This project adapts EO-ML models to estimate socio-economic conditions through transfer learning[2]. We used a foundation model trained on DHS Demographic and health data comprising approximately 1.2 million households in 36 African countries collected between 1990 and 2020; to transfer established wealth patterns to the data-scarce context of FDPs and host populations (see Figure 1 for illustration). The primary goal is to generate camp-level socio-economic proxies from satellite imagery to support prioritization and monitoring. This approach targets three critical challenges hindering the support of displaced populations: (1) limited geographic coverage in inaccessible areas; (2) temporal lags between survey rounds that obscure fast-changing wealth indicators; and (3) insufficient data on host communities. Bridging these gaps is essential for improving immediate aid delivery and designing evidence-based development strategies.

2 Data

The analysis primarily focuses on refugee, IDPs, and host community settlements in South Sudan (2023 data), with planned extensions to Cameroon (2024) and Zambia (2025). The data set is derived from two UNHCR household-level sources: the FDS[3], and the Results Monitoring Survey (RMS)[4]. The RMS tracks changes in rights and well-being (focusing on IDPs and refugee returnees in this case) to inform operational strategies.

As this study employs a transfer learning approach building upon the model established by Pettersson et al. [5] (see Section 3.4), we utilize a spatial grid resolution of 6.72 km $\times$ 6.72 km (45.16 km^2) to match the model's input requirements. To contextualize this spatial scale, Yida - the largest refugee camp analyzed in South Sudan, spans a surface area of 14.28 km^2, which is smaller than one grid cell. With this grid resolution, the household samples are distributed across a total of 211 grids, comprising 100 grids from the FDS data (3,000 households) and 111 grids from the RMS data (5,500 households). Each grid translates to a single observation during training.

2.1 Daytime multispectral optical imagery and nighttime light data

We employ daytime multispectral optical imagery and nighttime light data as core remote sensing inputs. These data sources were selected to capture complementary socioeconomic signals, utilizing optical features to characterize physical settlement morphology and dwelling quality while utilizing nighttime lights as a proxy for electrification and economic activity. For optical signals, we utilize Landsat Level-2 surface reflectance products, which are radiometrically corrected to ensure calibrated and comparable observations. This imagery is tiled into 6.72 km $\times$ 6.72 km grids (224px $\times$ 224px), matching the input specifications of the foundation model trained on Demographic and Health Survey (DHS) data across 36 African countries (as described in Section 3.4). With a spatial resolution of 10–30 m across visible, near-infrared, and shortwave-infrared bands, these inputs enable the model to capture vegetation, surface water, and built infrastructure. To complement daytime signals, nighttime light data from the VIIRS Day/Night Band are incorporated as a proxy for human presence and economic activity. All satellite sources are selected to align spatially and temporally with UNHCR socio-economic

[2] This technique enables the transfer of knowledge gained from solving one estimation task to address a related but slightly different estimation challenge.

[3] https://www.unhcr.org/what-we-do/reports-and-publications/data-and-statistics/forced-displacement-survey/forced-displacement-survey-south-sudan

[4] https://microdata.unhcr.org/index.php/catalog/1063

surveys. Finally, all imagery is accessed and processed at scale via Google Earth Engine, applying automated cloud masks and quality assessment filters to exclude anomalies.

2.2 Data augmentation

The constraints of a small sample size ($N \approx 200$) are mitigated through standard data augmentation, such as rotation and flipping. Furthermore, to leverage the high dwelling density characteristic of refugee and host community settlements, we employ a spatial sliding-window strategy to extract additional training samples. This involves systematically shifting the extraction frame around a central grid to generate spatially offset snapshots that retain significant overlap with the original view. This method amplifies the data set size while introducing partial spatial variations; finally, we apply a density filter to discard augmented grids containing fewer than five households.

3 Methodology

3.1 Data alignment and socio-economic index construction

FDS and RMS exhibit strong structural and semantic alignment, making them well suited for domain adaptation. In particular, both surveys include closely matched questions and value labels on living conditions, allowing the construction of transferable and comparable indicators between the data sets. We leverage this overlap to harmonize variables at the household-level variables such as housing quality, access to basic services, etc.

The socioeconomic index is constructed as a latent composite measure that encapsulates both physical and functional dimensions of welfare. Because holistic well being is not directly observable, we quantify it by using standardized proxies: visible features, such as dwelling quality and asset ownership, are combined with abstract indicators, such as sanitation access, to infer broader economic and health outcomes. This synthesis aggregates distinct observable characteristics into a comprehensive measure of household living standards. These variables are first standardized to ensure comparability across different measurement scales and survey instruments. We then apply Principal Component Analysis and interpret the first principal component as the underlying socio-economic dimension, as it captures the largest share of common variance across indicators. The resulting scores are linearly rescaled to a 0 to 100 index for interpretability.

3.2 Spatially stratified cross-validation

The data set encompasses households distributed across 28 level three administrative regions in South Sudan. Although the sampling frame covers both southern and northern territories, the distribution is geographically skewed toward the north. To mitigate biases arising from spatial heterogeneity and to prevent data leakage caused by spatial autocorrelation, we implemented a spatially stratified $k-$fold cross validation strategy. The data was partitioned into four spatially distinct folds (labeled A, B, C, and D). Unlike standard random splitting, we utilized a bespoke allocation heuristic to ensure strict spatial separation while maintaining demographic representativeness. This design guarantees that each fold reflects the diverse settlement structures of both refugees and host communities independent of population size, forcing the model to learn robust features applicable across varying geographical contexts.The experimental design follows a ”leave one group out” protocol. In each iteration, one specific geographic group is isolated as the Test Set, a second distinct group serves as the Validation Set, and the remaining two constitute the Training Set (e.g, A and B for training, C for validation, D for testing). This structure ensures that the model is trained, tuned, and tested on completely disjoint spatial distributions, preventing data leakage between neighboring administrative regions due to spatial autocorrelation.

3.3 Temporo-spatial vision transformer model

Vision transformers adapt the Transformer architecture to image analysis by representing each image as a sequence of patch tokens [6]. We utilise a multi-modal spatiotemporal transformer architecture designed to regress socioeconomic indicators from Satellite Image Time Series (SITS). The SITS framework leverages EO data that encapsulates spatial (space) and temporal (time) and spectral dimensions. Analogous to video processing, the model decomposes the SITS analysis into two distinct stages: frame-level spatial feature extraction and subsequent temporal aggregation. This hierarchical approach enables the effective detection of spatiotemporal patterns and dynamic evolution within the data. This allows the model to capture both local and global spatial relationships, including settlement structure, vegetation patterns, and built-environment characteristics.

3.4 Transfer learning and domain adaptation

Transfer learning reuses knowledge from models pre-trained on large data sets and adapts it to a new task through fine-tuning, which is particularly effective when the target data set is small. In our work, the pretrained model parameters encode general patterns of socioeconomic well-being across African countries, learned from DHS data using satellite imagery. This foundation model was trained on a comprehensive data set comprising 1.2 million households from 140 DHS surveys across 36 African countries, spanning a 30 year period (1990–2020) [5, 7]. During adaptation to the FDS setting, where labelled data for refugee populations is limited, we freeze the foundation model backbone weights so that the spatial encoder remains static and acts as a feature extractor and not a learner. Gradient updates are restricted to the Temporal Transformer encoder and the final regression head, allowing the model to recalibrate its spatiotemporal aggregation for the target distribution without altering the underlying visual semantics. This targeted fine-tuning enables the model to learn complementary features characteristic of refugee settlements, bridging the domain gap between general population surveys and the spatial signatures of displacement, and reducing DHS-FDS distribution mismatch.

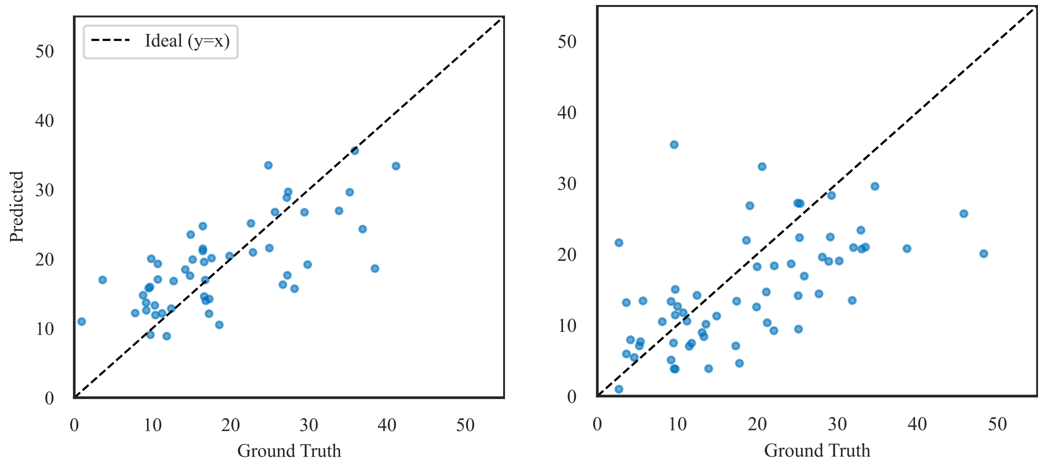
4 Preliminary results and discussion

We first established a baseline for transferability by quantifying the distribution shift between the source (DHS) and target (FDS) domains. To do this, we applied the model weights learned from the DHS context directly to the FDS data set without any retraining on FDS data (zero-shot inference). This step is critical for assessing whether the feature representations learned from general national surveys are immediately applicable to the unique environmental and structural characteristics of forced displacement settings.The results of this zero-shot evaluation reveal a substantial domain gap. While the model achieved a robust Coefficient of Determination ($R^2 \sim 0.7$) on the source DHS data [5, 7], direct transfer to the FDS target resulted in significant performance degradation, yielding a negative R^2 of -0.19 and a negative correlation of -0.13 (Table 1). This performance can belargely attributable to the distinct morphological signatures of refugee and host communities, such as high dwelling density, informal or temporary housing materials, and location in geophysically distinct regions, which fall outside the learned manifold of the encoder trained on DHS data. Notably, the DHS data set excludes areas inhabited by FDPs and host communities across the 36 countries. However, despite the poor goodness-of-fit metrics, the zero-shot model yielded a Mean Absolute Error (MAE) of 10.94 and a Root Mean Squared Error (RMSE) of 13.75. The comparable magnitude of these errors to the fine-tuned results suggests that while the zero-shot model fails to capture the specific variance and ordering of the target labels, the backbone effectively extracts plausible semantic representations, providing a robust initialization for subsequent fine-tuning.

Table 1: Summary of Results based on cross-validation. These results are based on validation data sets.

Fold/Model	MAE	MSE	RMSE	R^2	Correlation
A	6.70	78.12	8.84	0.33	0.59
B	**5.37**	**45.00**	**6.71**	**0.49**	**0.70**
C	8.10	107.92	10.39	0.33	0.58
D	10.47	182.88	13.52	0.02	0.17
zero-shot	10.94	171.24	13.75	-0.19	-0.13

Subsequent fine-tuning and evaluation using spatially stratified cross-validation further underscores the heterogeneity of the target domain. As detailed in Table 1, predictive performance varies substantially across the four geographical partitions, reflecting the localized nature of displacement settlements. Fold B demonstrated the most robust predictive capability, achieving the lowest error rates (MAE: 5.37, RMSE: 6.71) and the highest predictive power (R^2: 0.49, Correlation: 0.70), Figure 2. In contrast, Fold D exhibited significant struggles (R^2: 0.02), indicating that the morphological or socioeconomic patterns in this partition differ markedly from the training distributions in the other folds. This high variance validates the necessity of our spatially stratified



(a) Validation set (b) Test set

Figure 2: Predictive performance on validation and test datasets. Plots display ground truth versus predicted values; the dashed diagonal line ($y = x$) indicates perfect prediction accuracy.

approach; a standard random split would likely mask these regional disparities, leading to an overly optimistic assessment of model generalizability.

5 Conclusion

EO-ML techniques have been successfully applied to poverty estimation in low and middle income nations; however, their adaptation to the complex, data scarce environments of refugee and host community settlements remains unexplored. This study bridges that critical gap, presenting an innovative proof of concept that demonstrates the feasibility of transferring learned socioeconomic features to these unique contexts. Our preliminary results underscore the promise of this approach, offering a scalable pathway to operationalize the growing wealth of high quality, publicly available microdata libraries by the World Bank[5] and UNHCR[6]. Ultimately, this framework would provide a foundation for more frequent, granular monitoring to support humanitarian prioritization and evidence-based intervention.

References

[1] Utz Pape and Paolo Verme. “Measuring Poverty in Forced Displacement Contexts”. In: (2023).

[2] Mohammed Salah Alfahal et al. “Conflict and displacement in Sudan: Health challenges, socio-economic strain, trauma, and coping mechanisms among internally displaced persons in Port Sudan shelters”. In: *Medicine* 104.16 (2025), e42232.

[3] Isabel Ruiz and Carlos Vargas-Silva. “The labor market impacts of forced migration”. In: *American Economic Review* 105.5 (2015), pp. 581–586.

[4] Paolo Verme and Kirsten Schuettler. “The impact of forced displacement on host communities: A review of the empirical literature in economics”. In: *Journal of Development Economics* 150 (2021), p. 102606.

[5] Markus B Pettersson et al. “Time Series of Satellite Imagery Improve Deep Learning Estimates of Neighborhood-Level Poverty in Africa.” In: *IJCAI*. 2023, pp. 6165–6173.

[6] Alexey Dosovitskiy. “An image is worth 16x16 words: Transformers for image recognition at scale”. In: *arXiv preprint arXiv:2010.11929* (2020).

[7] Pettersson, Connor T Jerzak, and Adel Daoud. “Debiasing Machine Learning Predictions for Causal Inference Without Additional Ground Truth Data:” One Map, Many Trials” in Satellite-Driven Poverty Analysis”. In: *arXiv preprint arXiv:2508.01341* (2025).

[5] https://microdata.worldbank.org/index.php/catalog?page=1&sk=refugees&sort_by=rank&sort_order=desc&ps=15

[6] https://microdata.worldbank.org/catalog/unhcr?page=1&sk=refugees&ps=15&repo=unhcr