

S Supplementary Material

S.1 Tanzania 2015 DHS dataset

S.1.1 Indicator variables and the wealth index

The DHS Wealth Index is an asset-based index designed to compare relative, rather than absolute, economic wellbeing of households independently from education and health. The index is based on households' ownership of assets (such as radio, televisions, telephones, refrigerators, computers, bicycles, motorcycles, cars, bank accounts, house and lands); access to services (such as electricity, toilet facilities, and water sources, and sources of energy for cooking); and building materials (primarily flooring, walls and roofing materials). Responses to questions on assets, services, and enumerators' observations of building construction materials in the DHS survey are considered more reliable than self-reported income- and expenditure-based estimation of welfare and economic wellbeing [5]. The combined wealth index of clusters, the dependent variable, has been grouped into quintiles: *poorer*, *poor*, *average*, *wealthy* and *wealthier*. While others, such as [4] and [2], who have used this dataset have regrouped these quintiles into poor and non-poor, we opted to maintain the original categorizations to bring further nuance to the relative poverty ratings. At each cluster we calculate the mean of the combined wealth index and see which of the quintiles it belongs to. Consequently, since the quintiles are calculated for households and not clusters the number of clusters that fall in different quintiles is not necessarily the same.

S.1.2 Geolocation of clusters

While it is now standard practice for DHS datasets to come with location data of clusters, the DHS programme anonymises it by adding noise to the location variable for ethical reasons. In the current dataset, urban clusters are displaced by a distance of up to two kilometres and up to five kilometres for rural clusters, with a further randomly selected one per cent of the rural clusters displaced by up to ten kilometres [1]. This introduces uncertainties in the predictions as we matched coordinates with clusters. We overcome this by “correcting” the GPS data to the most likely location in an iterative manner on Google Maps. We achieve this by ensuring that we follow the guidelines for adding noise to the locations. That is, for rural locations, we assume that the nearest settlement within 5 km is the most likely “actual” cluster and expand the search area to 10 km in the few cases where we could not find any possible settlement in the 5km radius. We then use the “corrected coordinates” of clusters to extract cluster images.

S.2 Satellite images and online survey

The Google Maps Platform hosts a set of APIs through which developers can retrieve data from the platform. It is notoriously difficult to find metadata about images provided in the service. Generally, images are from different sensors and combined into a mosaic of images taken over multiple periods, with seams between different images being possible to identify sometimes. You can see some examples of the used images



Fig. S1: Samples of images from four different locations in Tanzania. Note that images are reduced by 40% from original size.

in Fig. S1. The corrected coordinates were fed into an R-script accessing the Google Maps platform and downloading corresponding ultra-high-resolution images. In total, 608 images were downloaded at zoom level 18. That corresponds to a pixel size of about 0.6 meters.

The extracted images were then fed into the web-based survey platform ². A pilot for the survey ran for two weeks using a group of Master's students at Lund University to test the resilience of the platform as well as the validity and reliability of our experimental design before the main survey went live between March 1 and May 31, 2022. Our approach was to send personalized invitations to experts within our networks who have fieldwork experience in Africa in the area of development research,

²<https://predict.gis.lu.se/>

broadly defined. We had contemplated sending a general invitation to our networks or asking respondents to forward invitations to other experts they deem qualified but eventually decided against these two options. First, we supposed that a general, unsolicited email might not garner the same response rate as a personal one. Secondly, we wanted to control the sample of human experts who rated the images. Each expert is first presented with a batch of 30 images in a sequential manner and asked to rate each one as either *poorer*, *poor*, *average*, *wealthy* or *wealthier*, to be comparable with the DHS quintiles. Raters were allowed to rate additional batches of 30 images if they so wished. Overall, the web portal received 2,174 ratings from 102 experts. The median rating for each cluster image was also taken as representing the relative poverty score for the cluster.

S.3 Domain Experts

As Table S1 shows, the sampled domain experts come with significant experience. This shows in not only their educational attainment and age distribution – 43% are 45 or more years old and 50% hold a Ph.D. – but even more importantly, that half of the respondents (N=51) have more than 10 years of fieldwork experience in SSA (South Saharan Africa). In terms of the region of experience, 48% have Tanzania, Eastern, or Southern Africa as their primary area of fieldwork experience.

Table S1: Summary statistics of web-based survey respondents (a total of 102 respondents)

Variable	Categories	Distribution
Age of respondents	Up to 35 yrs old	31
	36 - 55 yrs old	53
	56 yrs old and above	18
Gender	Females	23
	Males	77
	Non-binary	2
Educational qualifications	Bachelors	17
	Masters	34
	Ph.D.	51
Length of experience (years)	1 - 5 yrs	24
	6 - 10 yrs	27
	11 - 20 yrs	31
	Above 20 yrs	20
Region of experience	Tanzania	14
	Central and West Africa	53
	Southern and Eastern Africa	35

Of the 608 clusters for the 2015 TDHS dataset, 428 were classified as rural with the remaining 180 classified as urban. By our iterative approach to geolocate the clusters, we estimate average dislocation of 3.09 (SD=2.66) km for rural areas and 1.15 (SD=0.77) for urban clusters. These fall within the DHS geographical displacement

guidelines. Fig. 2 (b) and (a) show the distribution of the wealth quintiles based on ratings by our domain experts and the DHS dataset, respectively. While comparable in terms of the quintile categorisation, it must be noted that the unit of analysis for the DHS survey is the household (N=12,563) which then had to be aggregated to the cluster level to match the web-based survey where the unit of analysis is the cluster level (N=608).

S.4 Expert Defined features

To shed light on the correlates of the ratings by experts, a criterion was developed, based on literature and discussions with experts with extensive knowledge of the Tanzanian context to classify and grade the images based on pertinent spatial features which could be discerned. On this, we rated the images on three broad criteria, namely, housing features (building types, roofing materials, roofing condition, and buildings' size); landscape features (settlement structure, building density, greenery coverage, dominant land use, and image colour scale) and assets and infrastructure (roads surface quality, roads width, roads coverage, vehicles presence, and farm sizes. Table S2 presents a detailed description of the sub-criteria used for the classification. Each image was manually inspected for these qualities and scored accordingly to find the correlates of the ratings by the domain experts.

Table S2: Literature and expert-based correlate features in images.

Broad criteria	Spatial features	Categories	Codes
HOUSING FEATURES	Size of buildings	single units	1
		multiple units	2
		unclassified	99
	Roofing Material	thatch roofing	1
		aluminium roofing	2
		slate roofing	3
		unclassified	99
	Roofing Condition	uncompleted	1
		old	2
		new	3
	Presence of tall buildings	no	1
		yes	2
		unclassified	99
LANDSCAPE FEATURES	Structure of settlements	clustered	1
		scattered	2
		gridded	3
	Density of built environment	low coverage (less than 25%)	1
		medium coverage (between 25–74%)	2
		high coverage (greater than 75%)	3
	Presence of trees/greenery	low (below 25%)	1
		medium (25–75%)	2
		high (above 75%)	3
	Dominant land use	bare land	1
		agricultural	2

ASSETS AND INFRASTRUCTURE		built up	3
		industrial	4
		commercial	5
		unclassified	99
	Image colour	brownish	1
		yellowish	2
		greenish	3
	Presence of vehicles	no vehicles visible	0
		few vehicles (1–2 visible)	1
		many vehicles (more than 2 visible)	2
	Road surface quality	no roads	0
		untarred roads	1
		tarred roads	2
	Road width	small	1
		medium	2
		large	3
		unclassified	99
	Farm sizes	small farms	1
		large farms	2
		no farms	3
	Road coverage	low	1
		medium	2
		high	3

All three datasets – the DHS survey data, experts’ rating of relative poverty, and researchers’ scoring of discernible features in images – were merged into one dataset using the cluster code as the link.

S.5 Feature Correlation and Importance

S.5.1 Correlation between feature values

Figure S2 shows how related the expert defined features are to each other in terms of Pearson correlation coefficients, calculated using the 608 clusters in our Tanzania dataset. The highest correlation is between *building density* and *vehicle presence*, and *building density* and *road surface quality*. After that, the highest correlation is between *image colour* and *greenery*. In the case of *image colour* and *greenery* the causal dependence is obvious hence we suggest that *image colour* is an auxiliary feature.

S.5.2 Logistic regression weights

The logistic regression model is linear, which means that the weights (the parameters) of the model reflect the importance of the features. Because the five wealth categories are imbalanced regarding the number of samples we applied class weights to equalize them for training and also classification accuracy.

Here, care must be taken as to what logistic regression method to use. A multinomial logistic regression separates each class from the others and thus produces the feature weights that are most unique for each particular class (in this case, five sets of weights). This may not reflect what is important overall for wealth and poverty. On

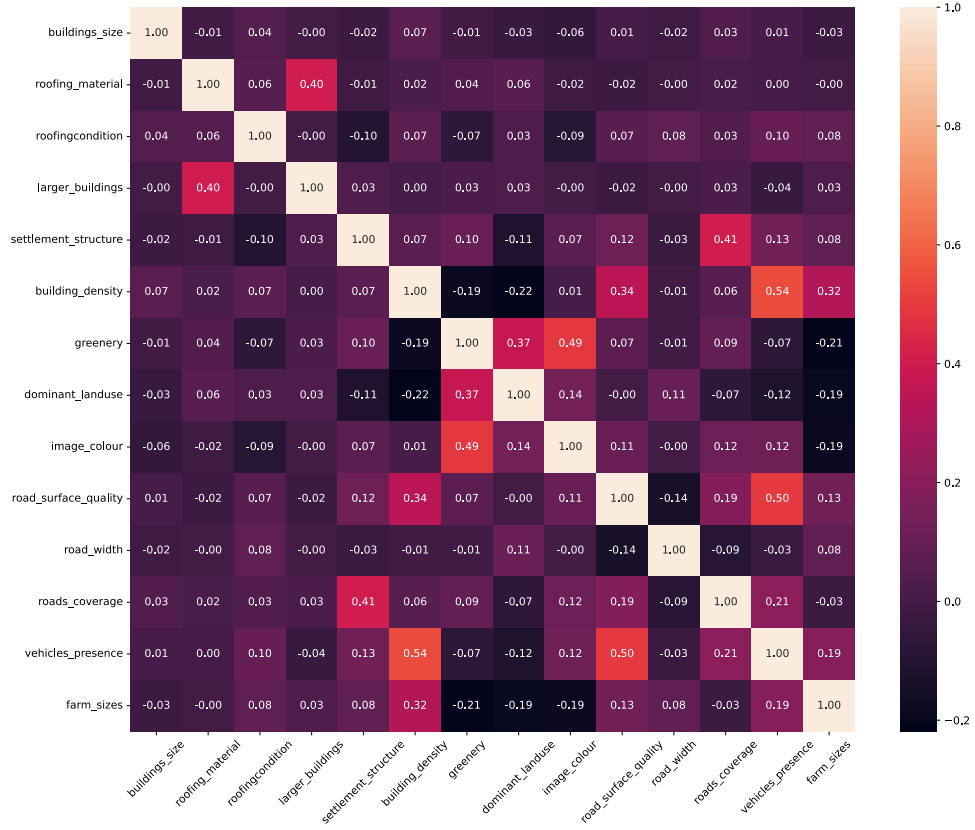


Fig. S2: Correlations between manually defined features extracted for all the test clusters.

the other hand, an ordinal logistic regression [3] will yield only one set of weights, considering that the classes are ordinal. This will be an overall set of weights, which also reflects the wealth and poverty relation better than the multinomial logistic regression. For this experiment, we opted for a choice in-between the multinomial and ordinal, which both gave weights that reflect wealth and poverty and more detailed information. This was achieved by doing binomial logistic regression for each of the four dichotomies described earlier (poorest vs. the others, poorest and poor vs the others, etc.).

For each dichotomy, a total of 100 logistic regression experiments were done, with variations in the data samples, and the feature weights were averaged over the experiments. All features were normalized to have the same scale. The average weights are presented in Fig. S3a, which shows that *roofing material* is the most influential feature for all dichotomies except separating the richest sites from the others (1-4 vs. 5), where *building density* is the most influential feature. A general observation is that the extreme dichotomies (1 vs. 2-5 and 1-4 vs. 5) are different from the more central dichotomies. The feature *roofing material* is more prominent in the separation of the

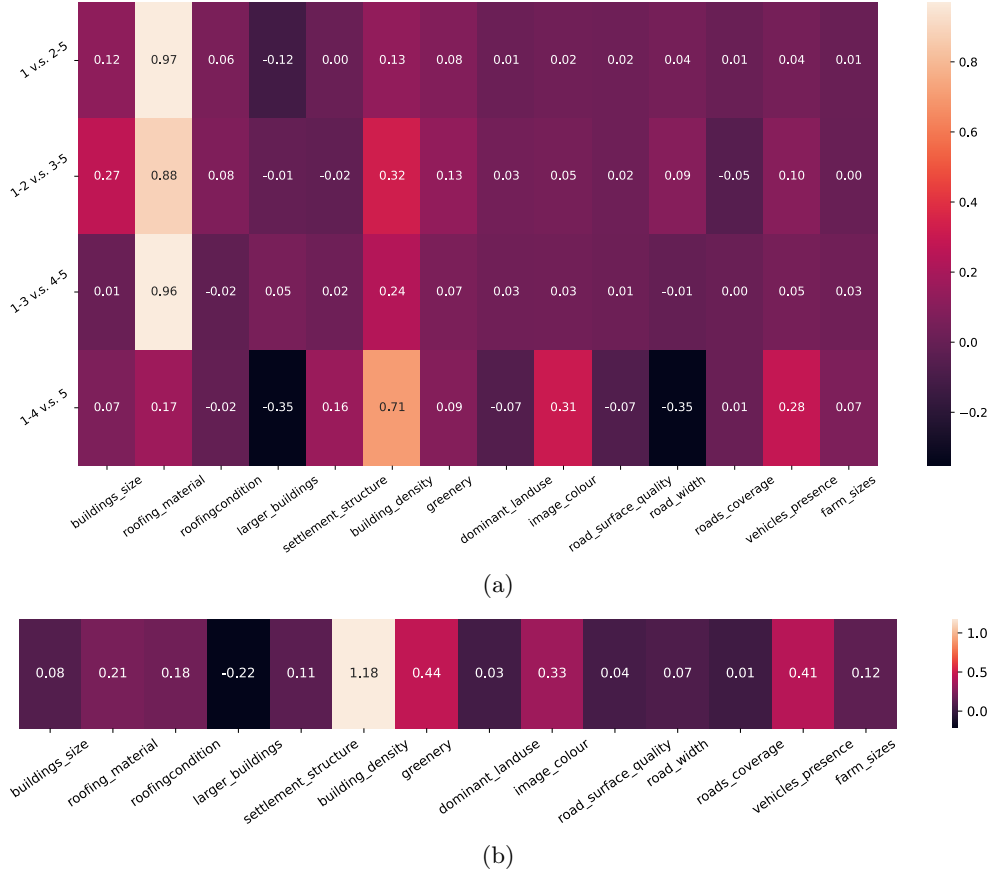


Fig. S3: Average weight for each feature in the (a) multinomial and (b) ordinal logistic regression classifiers across 100 experiments.

poorest from the others. These results can be compared to those achieved with an ordinal logistic regression classifier, in Fig. S3b where *building density* appears to be overall most influential.

S.5.3 Backward elimination of features

To further analyze the features' relative importance we used backward elimination for the logistic regression classifier on the dichotomies. The process involves removing the worst performing feature, one at a time, to see which features remain at later stages. The *image colour* feature was removed since it gave very similar results to *greenery*. The results are presented in Fig. S4, where Fig. S4a shows the average elimination order, and Fig. S4b shows the average classification accuracy achieved with that number of features. A feature that has a value of 12 for average elimination

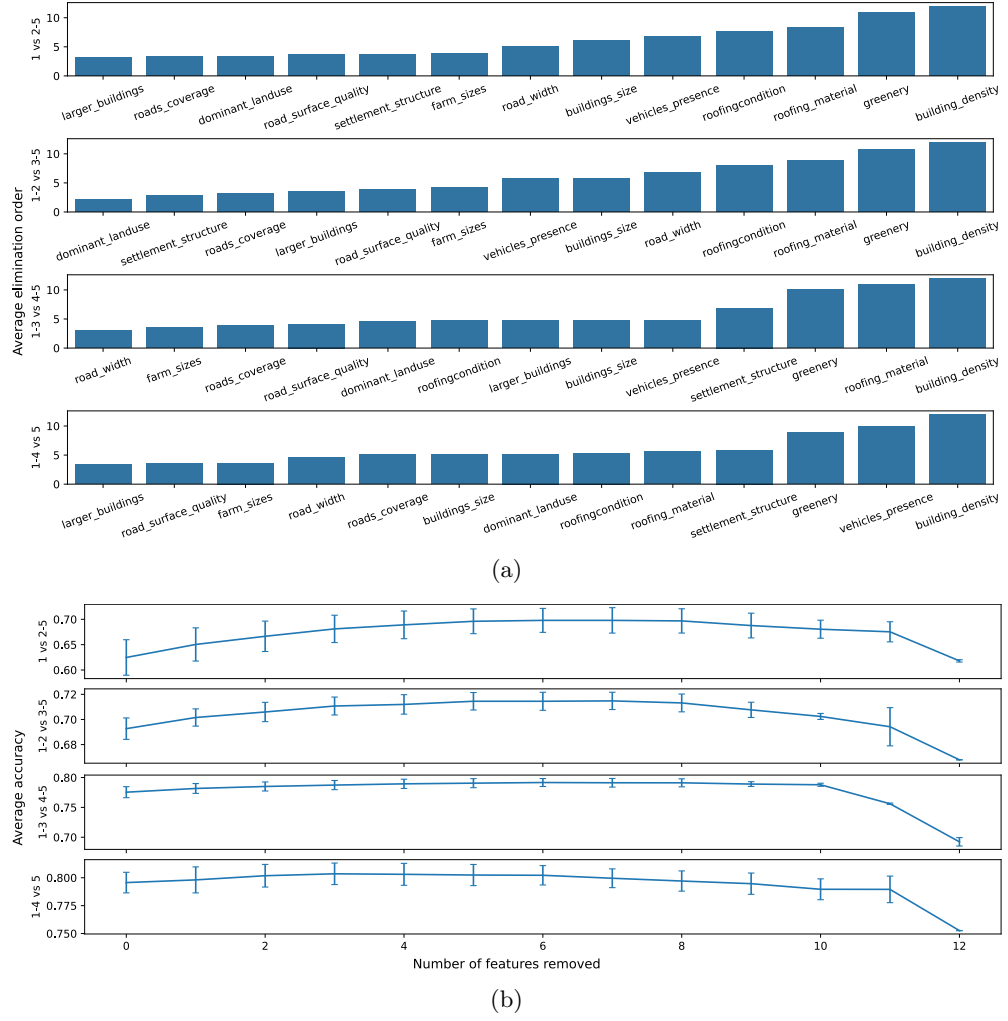


Fig. S4: (a) Average elimination feature ranking and (b) change in classification accuracy from the backward elimination procedure done on the dichotomy classifiers. The results are the average of 100 separate experiments with 5-fold cross-validation.

order survives to the end in all backward elimination experiments, a feature that has a value of 11 is usually the second last to be removed, and so on.

The *building density* feature is the most important feature across the board, which agrees with the ordinal logistic regression weights in Fig. S3b. The feature *greenery* is second most important for separating the poor classes from the others, followed by *roofing material* and *roofing condition*. For the richest, the feature *vehicles presence* appears very important, together with *greenery*.

The accuracy plots in Fig. S4b show that it is often enough with 2-3 features to get good accuracy. It appears that the separation of rich and richest from the other sites requires most features (3), whereas the extreme dichotomies (poorest vs. others, and richest vs. others) get good accuracy with only two features.

References

- [1] Clara R. Burgert and Debra Prosnitz. “Linking DHS household and SPA facility surveys: Data considerations and geospatial methods”. en. In: (Sept. 2014). Number: DHS Spatial Analysis Reports No. 10. URL: <https://dhsprogram.com/publications/publication-SAR10-Spatial-Analysis-Reports.cfm> (visited on 02/23/2024).
- [2] Qing Li et al. “Is poverty predictable with machine learning? A study of DHS data from Kyrgyzstan”. In: *Socio-Economic Planning Sciences* 81 (June 2022), p. 101195. ISSN: 0038-0121. DOI: [10.1016/j.seps.2021.101195](https://doi.org/10.1016/j.seps.2021.101195). URL: <https://www.sciencedirect.com/science/article/pii/S0038012121001877> (visited on 02/23/2024).
- [3] Peter McCullagh. “Regression Models for Ordinal Data”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 42.2 (1980). Publisher: [Royal Statistical Society, Wiley], pp. 109–142. ISSN: 0035-9246. URL: <https://www.jstor.org/stable/2984952> (visited on 04/17/2024).
- [4] Badiea Shaukat, Sajid Amin Javed, and Waqas Imran. “Wealth Index as Substitute to Income and Consumption: Assessment of Household Poverty Determinants Using Demographic and Health Survey Data”. In: *Journal of Poverty* 24.1 (Jan. 2020). Publisher: Routledge eprint: <https://doi.org/10.1080/10875549.2019.1678550>, pp. 24–44. ISSN: 1087-5549. DOI: [10.1080/10875549.2019.1678550](https://doi.org/10.1080/10875549.2019.1678550). URL: <https://doi.org/10.1080/10875549.2019.1678550> (visited on 02/23/2024).
- [5] Sarah Staveteig and Lindsay Mallick. “Intertemporal comparisons of poverty and wealth with DHS data: A harmonized asset index approach”. en. In: (Sept. 2014). Number: DHS Methodological Reports No. 15. URL: <https://dhsprogram.com/publications/publication-mr15-methodological-reports.cfm> (visited on 02/23/2024).