Mapping Poverty in the Philippines using Machine Learning, Satellite Imagery, and Crowdsourced Geospatial Information

Isabelle Tingzon[1,‡], Ardie Orden[1,‡], Stephanie Sy[1], Vedran Sekara[2], Ingmar Weber[3], Masoomali Fatehkia[3], Manuel Garcia Herranz[2], Dohyung Kim[2]

[1]Thinking Machines Data Science, [2]UNICEF, [3]Qatar Computing Research Institute

Motivation

- Around 22 million Filipinos live below the national poverty line.
- A major challenge in fighting poverty is the lack of reliable socioeconomic data, which is often expensive, time-consuming, and labor-intensive to collect.
- Traditional Filipino household surveys can cost up to 1.5M USD, are conducted every 3-5 years, and are often aggregated to the provincial or regional level.
- **Goal**: *Faster*, *cheaper*, and *more granular* estimation of poverty measures and socioeconomic indicators in the Philippines.

Research Objectives

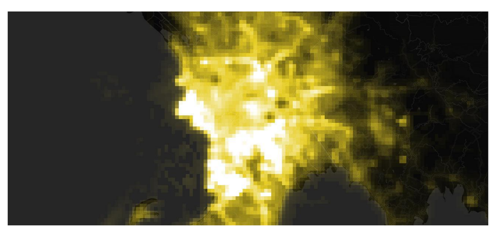
We seek to answer the following questions:

1. Are satellite-based methods developed for poverty prediction in other countries applicable within the Philippine context?
2. How well do predictive models trained on free and publicly available crowdsourced geospatial information compare against state-of-the-art methods for Philippine poverty estimation?

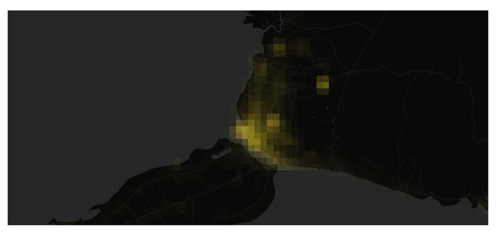
Datasets

We used the following datasets in this study:

- Philippine Demographic and Health Survey (DHS) 2017 (Ground-truth)
- Nighttime lights (NTL) data taken from Visible Infrared Imaging Radiometer Suite Day/Night Band (VIIRS DNB) 2016 [1]
- Daytime satellite images accessed via Google Static Maps API
- High Resolution Settlement Layer (HRSL) by Tiecke et al. [2]
- Volunteered geographic information (VGI) from OpenStreetMap (OSM)



(a) Nightlights in Metro Manila



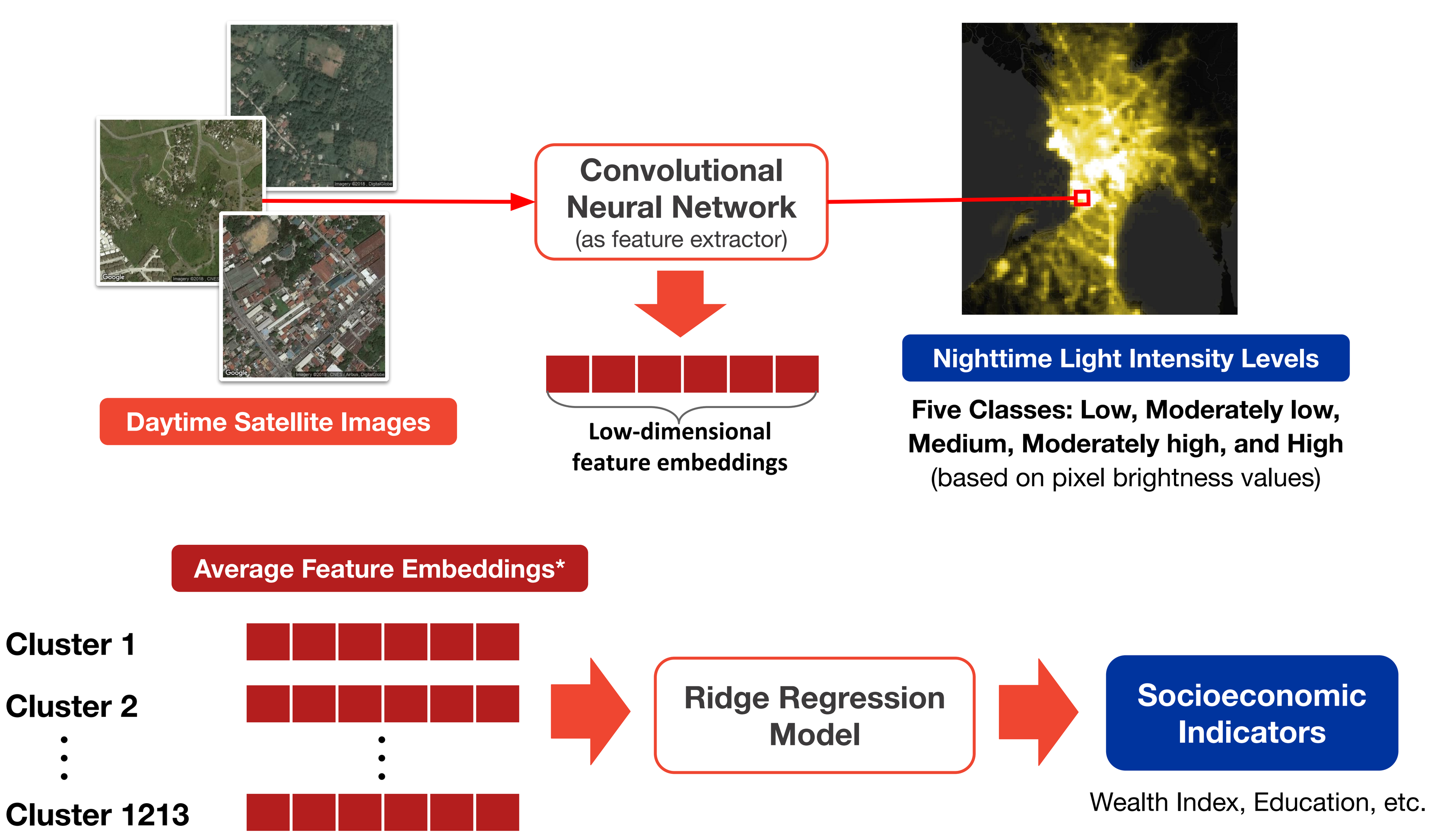
(b) Nightlights in Tagbilaran, Bohol (rural area)

The DHS contains 1,213 clusters, where each cluster consists of up to 44 households. A cluster location is defined by the mean latitudes and longitudes of households within a 2 km (urban) or 5 km (rural) radius.

Methodology

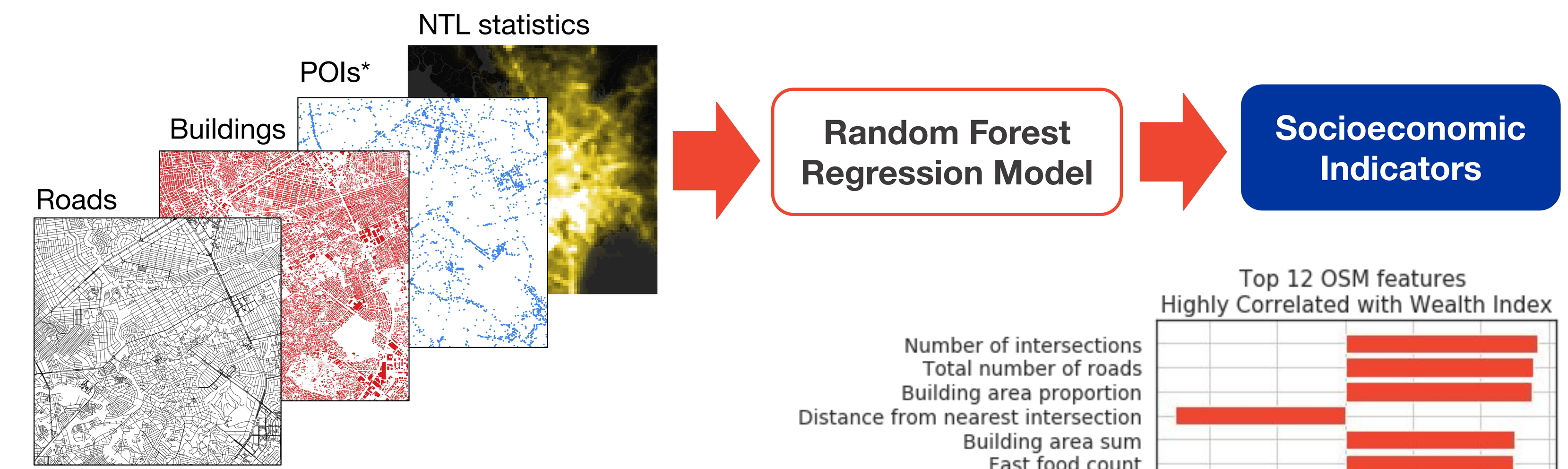
We describe the different methods used in predicting socioeconomic well-being. All models were evaluated using a five-fold nested cross validation scheme.

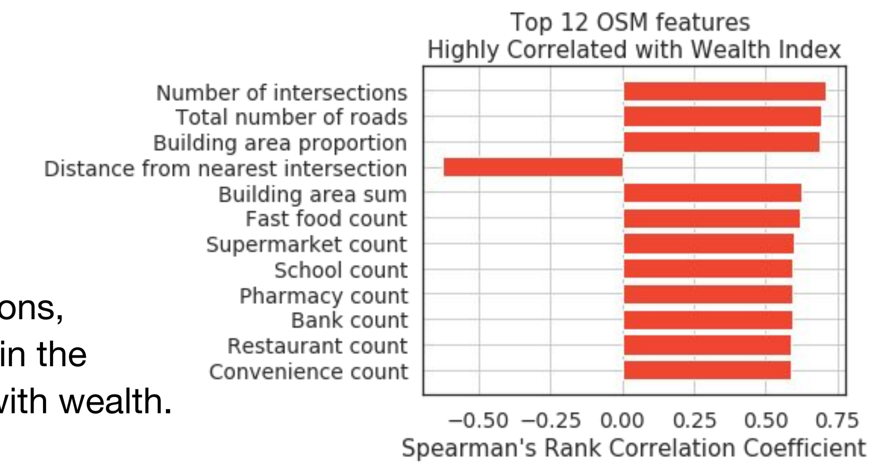
1. Satellite-based deep learning approach as described by Jean et al. [3]



*Average feature embeddings of up to 400 image tiles within a 2km (urban) or 5 km (rural) radius from the cluster centroid. We removed images containing no human settlements using the HRSL dataset by Tiecke et al. [2]

2. Cost-effective alternative approach using volunteered geographic information from OpenStreetMap and nighttime light summary statistics from VIIRS DNB [1]

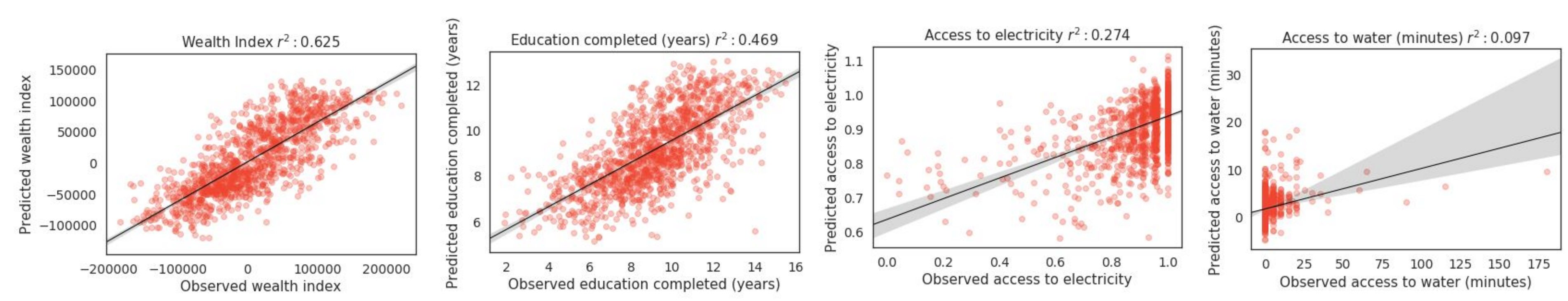




OSM features such as number of road intersections, building area proportion, and count of POIs within the cluster area were found to be highly correlated with wealth.

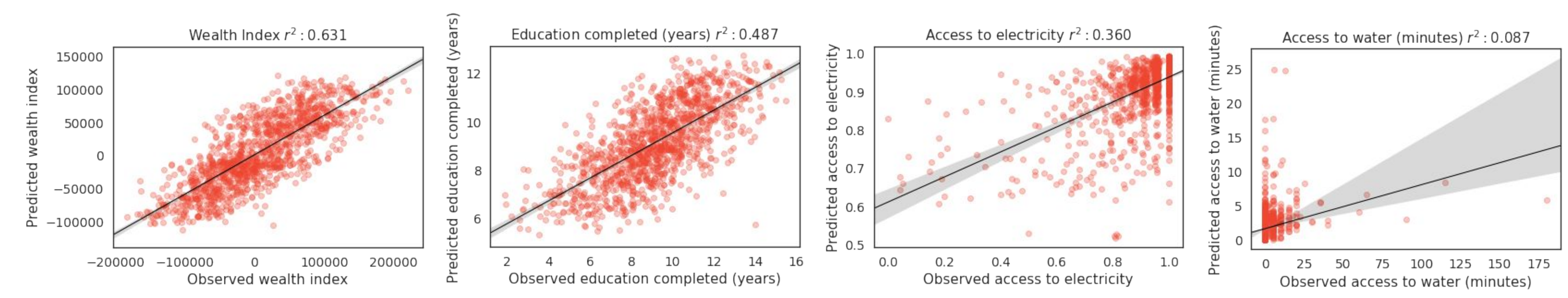
Results

We present the cross-validated predictions and r-squared scores for four different socioeconomic indicators.



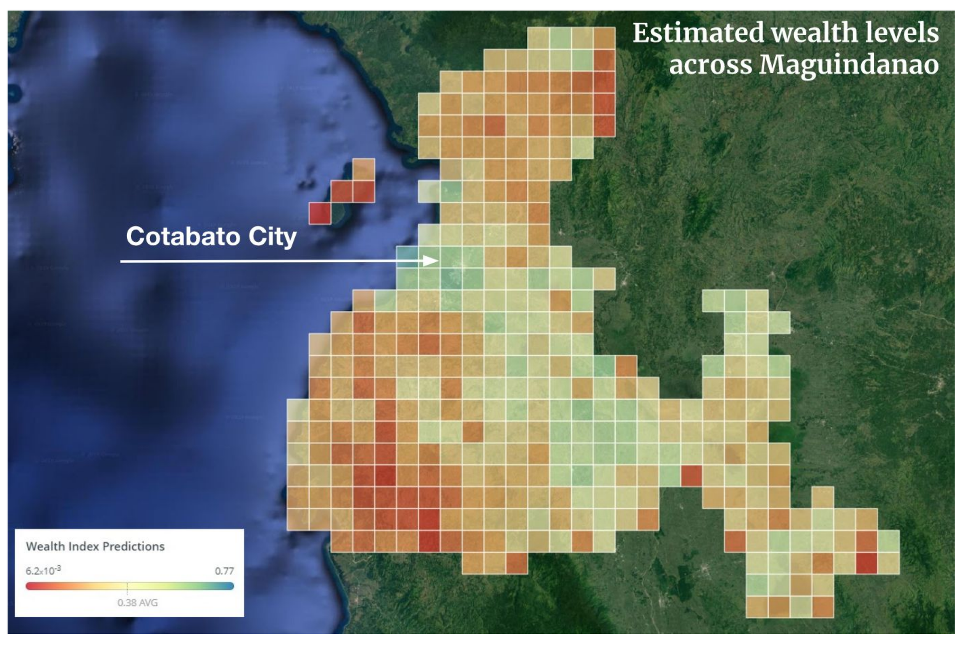
(a) Wealth Index (b) Education (c) Access to Electricity (d) Access to Water

Ground-truth and cross-validated Philippine poverty predictions using the satellite-based deep learning model* by Jean et al. [3].

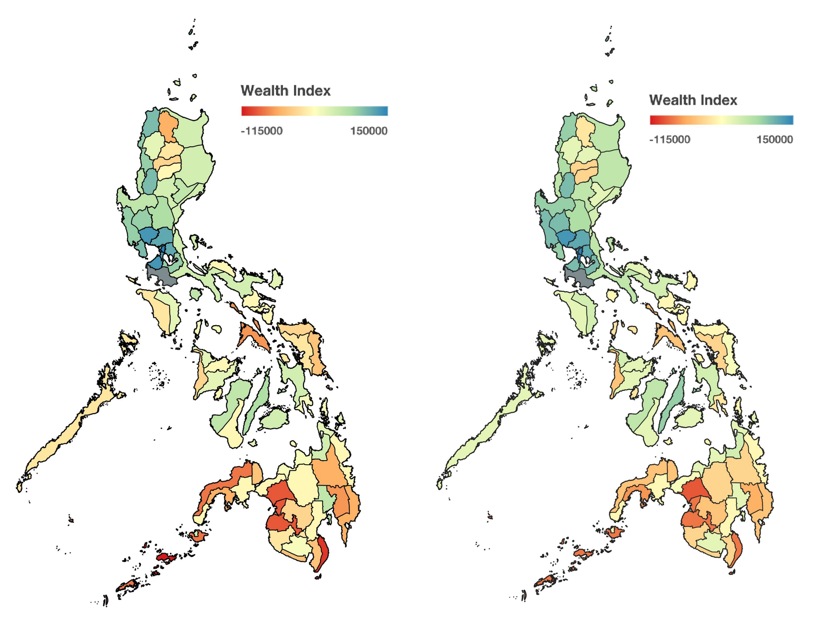


(a) Wealth Index (b) Education (c) Access to Electricity (d) Access to Water

Ground-truth and cross-validated Philippine poverty predictions using the hybrid OpenStreetMap + nighttime lights model*.



High resolution (18 sq. km) wealth estimates for Maguindanao, a conflict area in Southern Mindanao



(*Left*) Ground-truth wealth indices and (*Right*) Cross-validated wealth predictions aggregated to the provincial level.

***Note**: Model incorporates binary regional indicators as input features.

Conclusions

- We investigated the extent to which state-of-the-art remote-sensing methods can be applied in the Philippine context to predict different socioeconomic indicators.
- We also proposed a cost-effective approach that leverages a combination of volunteered geographic information from OpenStreetMap and nighttime lights satellite imagery for estimating socioeconomic indicators.
- The best models explain approximately 63% of the variation in asset-based wealth.
- Our findings indicate that models trained on free and publicly available volunteer-curated geographic data achieve roughly the same predictive performance as that of models trained using proprietary satellite images.

References

[1] NOAA National Centers for Environmental Information. VIIRS DNB Dataset. https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html, 2016. Online; accessed 18 December 2018.

[2] T.G. Tiecke, X. Liu, A. Zhang, A. Gros, N. Li, G. Yetman, T. Kilic, S. Murray, B. Blankespoor, E.B. Prydz, H.H. Dang. Mapping the world population one building at a time.. *arXiv 1712.05839*, 2017.

[3] N. Jean, M. Burke, M. Xie, W.M. Davis, D.B. Lobell, S. Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301): 790-794, 2016. DOI: doi.org/10.1126/science.aaf7894

‡{issa, ardie}@thinkingmachin.es

https://github.com/thinkingmachines/ph-poverty-mapping