Targeting the Poor using Big Data

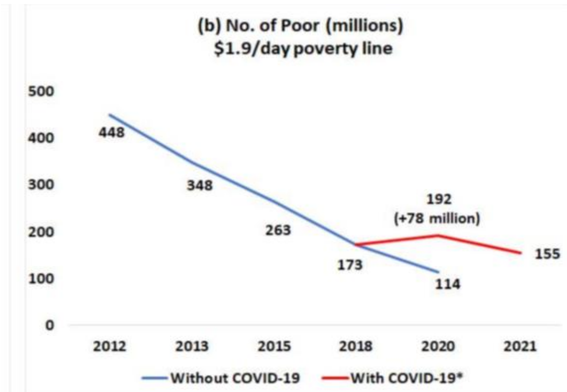
Arturo M. Martinez Jr.
Statistics and Data Innovation Unit
Economic Research and Regional Cooperation Department

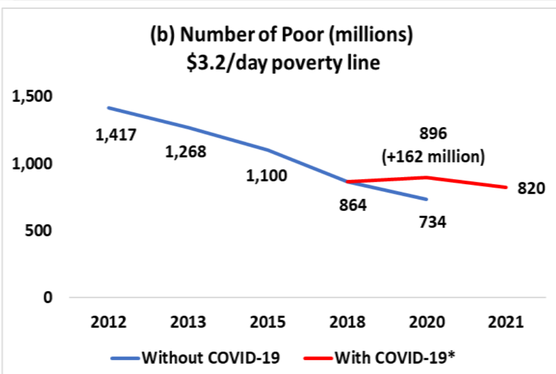
Wolfgang Fengler
World Bank / World Data Lab

Outline

- Finding Alternative Sources of Data for Poverty Mapping
- Using Computer Vision Algorithms to Map the Spatial Distribution of Poverty
- Key Findings
- Moving Forward

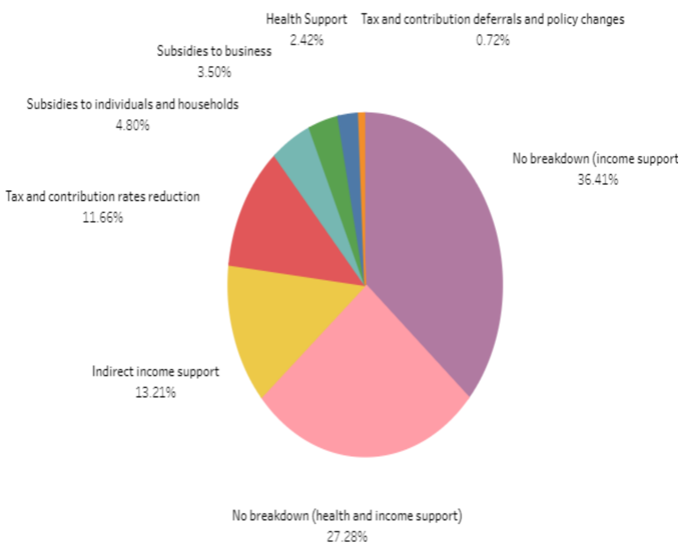
After decades of poverty reduction, COVID-19 threatens to turn back Developing Asia's poverty clock.





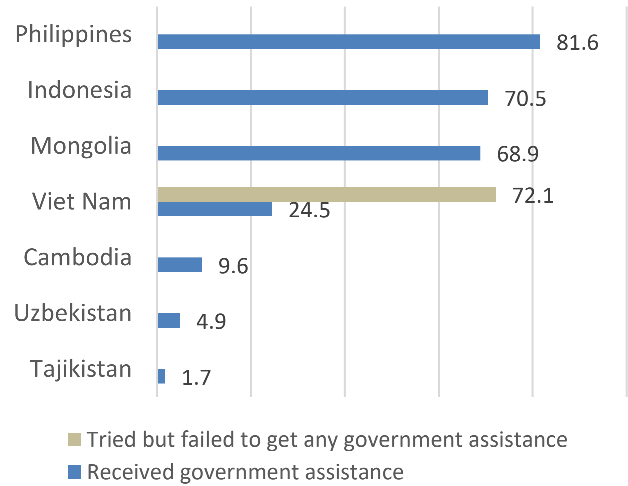
Having access to social safety nets is critical for the poor, as many of them were already struggling even before the pandemic broke out.

Health and Income Support, by Subcategories (%)
ADB's Developing Members



Source: ADB COVID-19 Policy Database

Households that received any form of government assistance & households that tried but could not obtain any government assistance since the start of the pandemic (%)



Source: World Bank COVID-19 High-Frequency Monitoring Dashboard

Effectiveness of poverty targeting partly depends on data availability regime.

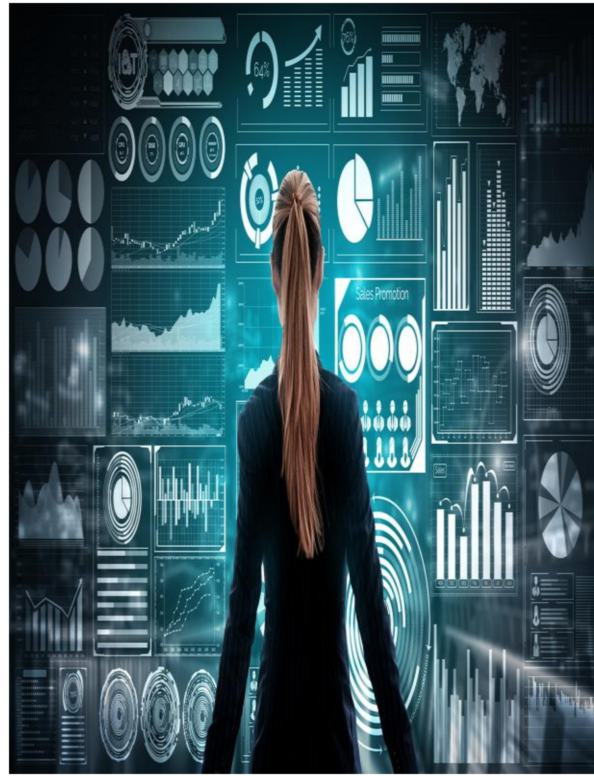
DATA RICH

Availability of regularly updated hhld / individual-level data on socioeconomic status (poverty, employment, etc)

Availability of sparse and irregularly updated information about the hhld / individual's socioeconomic status

Don't have sufficient and reliable information on areas where significant pockets of poverty exist

DATA POOR



Maturity of data and statistical systems affects availability of data.

<u>Availability of regularly updated hhld / individual-level data on socioeconomic status (poverty, employment, etc)</u>

Countries with mature data and statistical systems have capacity to integrate multiple data sources such as administrative data and registration systems on eligibility to social protection benefits, employment / unemployment benefits, etc with survey data, and other non-traditional sources of data for development

<u>Availability of sparse and irregularly updated information about the hhld / individual's socioeconomic status</u>

Countries that heavily depend on traditional data sources like surveys, censuses, etc. Administrative data are available but are not usually updated and harmonized.

<u>Don't have sufficient and reliable information on areas where significant pockets of poverty exist</u>

Countries with weak investments on national data and statistical systems; heavily depend on external resources

Often, DMCs' national data and statistical systems encounter challenges in compiling granular and timely poverty data. Innovative data sources can help address such challenges.

Example: Poverty Data in the Philippines

PHL's official poverty data have been based on the **Family Income and Expenditure Survey** (FIES). These are designed to provide reliable data at the national, regional, and more recently provincial data – <u>useful for broad monitoring purpose</u> but not much for targeting because of insufficient granularity.

Providing more granular poverty data by depending on FIES alone is very costly as it will entail significantly higher sample sizes -> initiative to combine FIES with census data (small area estimation) to estimate poverty at municipal- / city-level to respond to clamor for more granular poverty data – useful for spatial targeting, but not much for hhld-level targeting

Eventually, PHL's Department of Social Welfare and Development developed the <u>**Listahanan**</u> – <u>meant for identifying target hhld beneficiaries for various social protection programs</u>. Coverage of Listahanan started in areas where there were significant pockets of poverty based on PSA's estimates

Example: Poverty Data in the Philippines

FIES (and SAE) – assess whether poverty is increasing / decreasing and which areas and population groups are progressing / lagging behind; available at national, regional, provincial, municipal / city-level; conducted every two to three years

Listahanan – identify eligible hhld beneficiaries of government's social protection programs; ongoing update which started last year; last update was in 2015.

ADB's Bayan-bayanihan Program

The program needed granular data (ideally hhld-level). At that time, however, there were bottlenecks on getting the necessary information from Listahanan

The poverty maps available through FIES / SAE were not granular enough/

We used more granular ADB poverty maps by integrating information from satellite imagery. These were also triangulated with other info (e.g., distance from markets, etc)

EXPLAINER

How Satellite Data Helped Get Food to the Hungry during COVID-19

A COVID-19 emergency food program in the Philippines offered an opportunity to design a targeting program based on granular poverty maps that were compiled using traditional and innovative data sources and artificial intelligence. Photo credit: ADB.

'e on:

lished: 21 December 2020

Poverty maps derived from satellite images helped target the nost vulnerable households in pandemic-affected areas in the Philippines.

troduction

y day, the world generates an estimated 2.5 quintillion bytes of data. Applications of data come digital transactions, telecommunications records, social media, remote sensing, to name a which permeate almost every aspect of daily life.

when the coronavirus disease (COVID-19) struck, everyone was blindsided by the lack of mation on the novel virus. Data have since been collected and analyzed, providing actionable hts on how to navigate through this crisis.

Ask the Experts

Arturo Martinez, Jr.
Statistician, Economic Research and Regional Cooperation Department, Asian Development Bank

Anouj Mehta
Unit Head, Green and Innovative Finance and the ASEAN Catalytic Green Finance Facility, Southeast Asia Department, Asian Development Bank

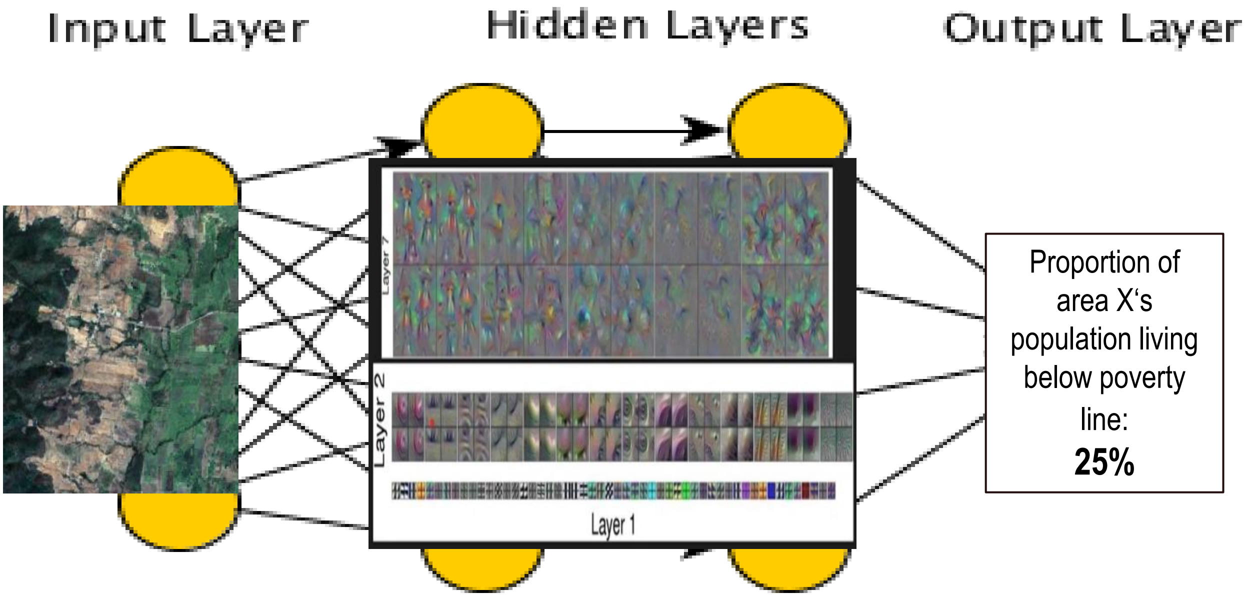
ADB

Asian Development Bank (ADB)

[How Satellite Data Helped Get Food to the Hungry during COVID-19 | Development Asia]

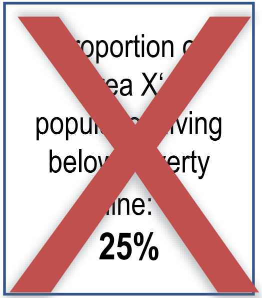
Methods: Using AI for Poverty Mapping

Our objective is to leverage on state-of-the-art computer vision technique, i.e., Convolutional Neural Network, and train it to predict the level of poverty by learning abstract patterns or features from satellite imagery.



Methods: Using AI for Poverty Mapping

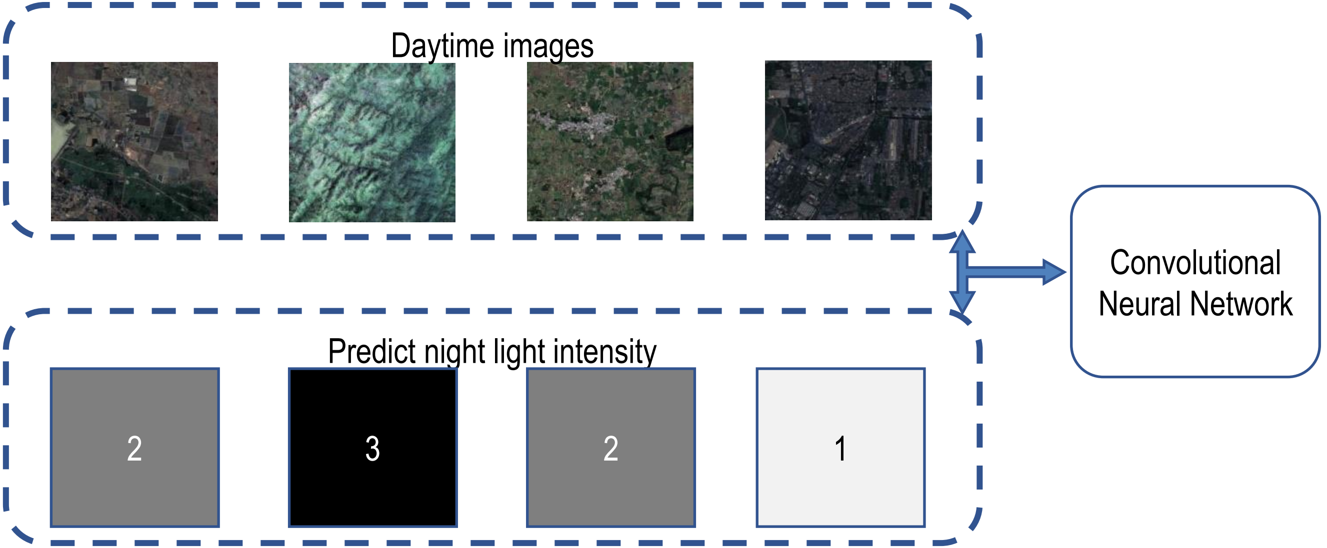
CNN requires volumes of poverty-labelled images as input data for training which we do not have!



Poverty statistics are typically derived from hhld surveys which are designed to provide reliable estimates at national, regional, or provincial-levels only.

Other countries using small area estimation techniques by combining hhld survey with census to provide village or district-level estimates but SAE has technical complications too, and in some cases, even the number of SAEs is not enough to train a CNN.

Methods: Using AI for Poverty Mapping

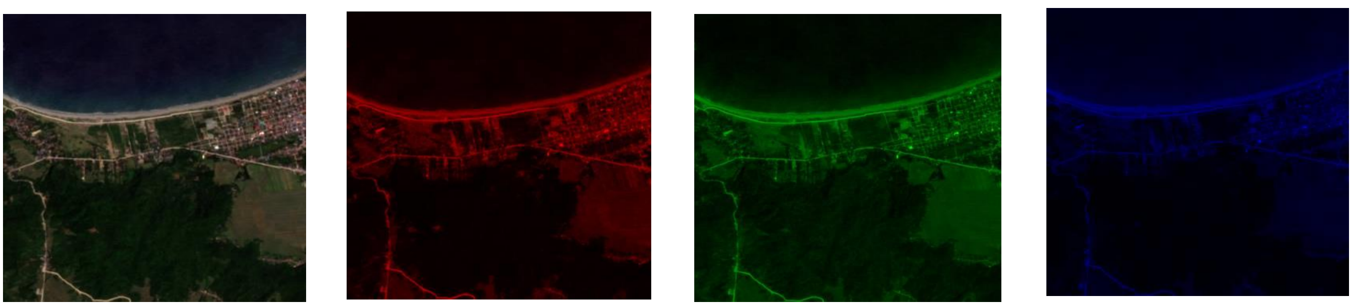


Applying on PHI data

We use the estimated proportion of population living below the national poverty line as compiled by the Philippine Statistics Authority and National Statistical Office of Thailand through SAE techniques: approx. 1600 *Municipal / city-level in the Philippines*

The input data were obtained using georeferenced and tagged image files. These image files are stored in three-dimensional arrays, with each pixel represented in red, green, and blue color bands.

Figure. Image Color Bands within a Georeferenced Image File

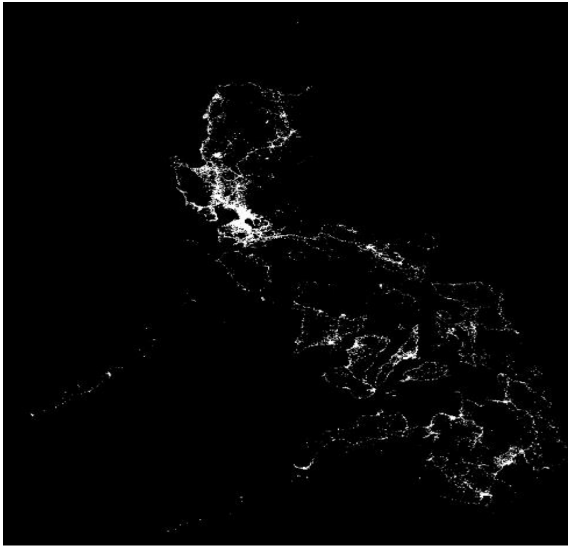


Source: Sentinel Images

Applying on PHI and THA data

We use data on night lights compiled by the Visible Infrared Imaging Radiometer Suite (VIIRS). The intensity levels were categorized into discrete groups using combination of Gaussian Mixed Models and heuristic methods.

Figure. Intensity of Night Lights



Applying on PHI data

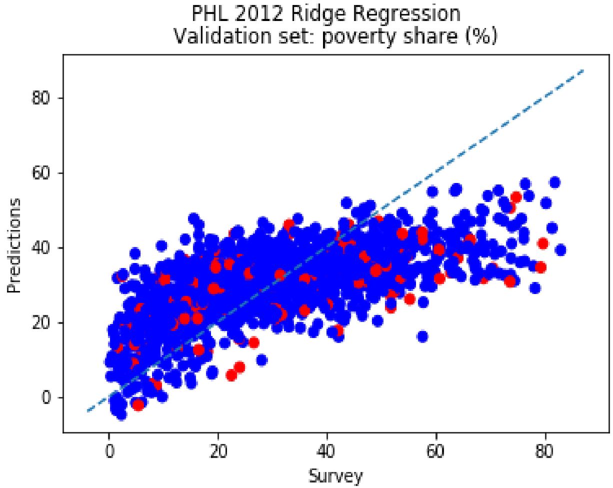
To avoid starting from scratch, we use an off-the-shelf CNN called ResNet34. This algorithm has been pretrained using the ImageNet database to ensure that it is capable of identifying simple features. ImageNet is regarded as a solid benchmark performer in computer vision predictions.

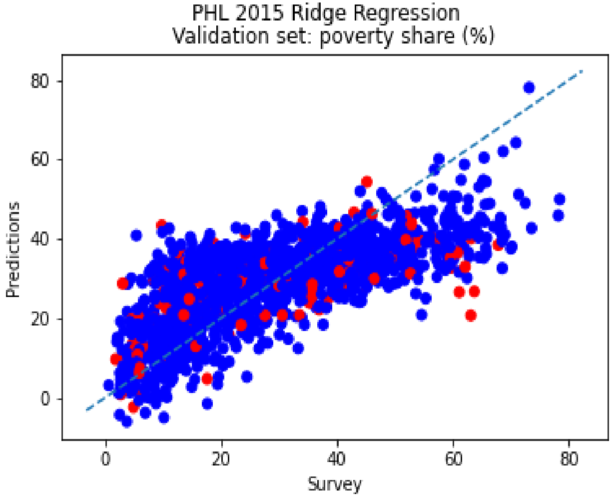
We also leveraged on readily available analytical platforms:

Key Findings

For each country-year combination, we set aside 90% of the areas for which government-published estimates of poverty are available to constitute the training set. The remaining 10% were used for validation. Within the 90%, we did a further split wherein we used 10-fold cross validation to tune hyperparameters.

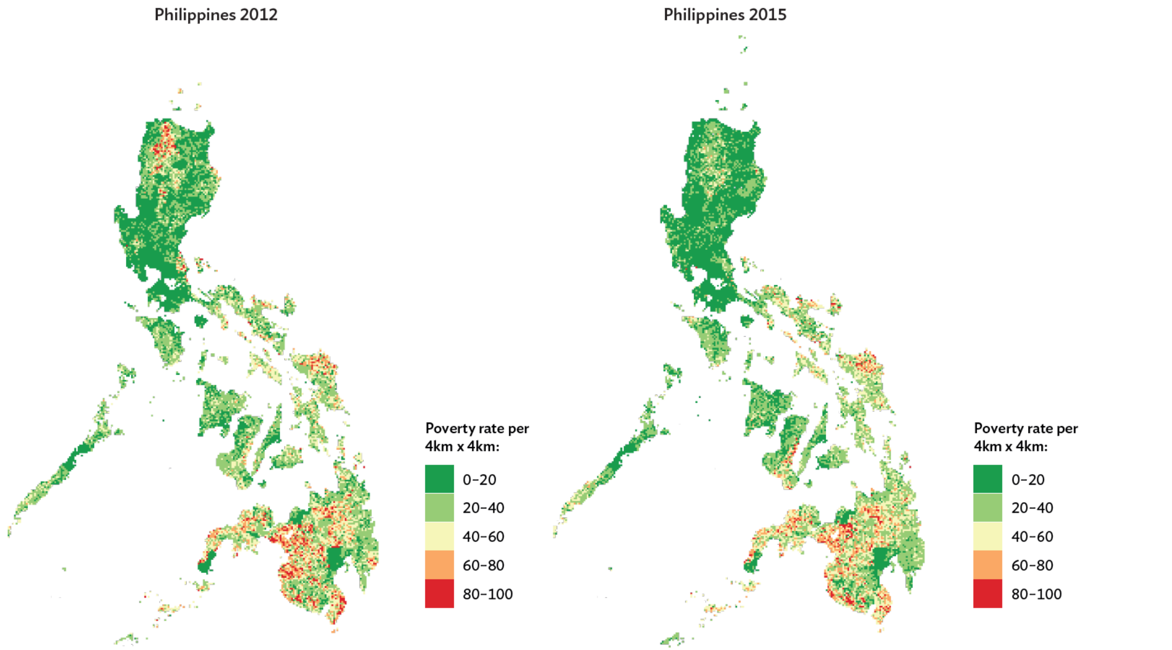
CNN's accuracy rate is about 93.5 to 94%



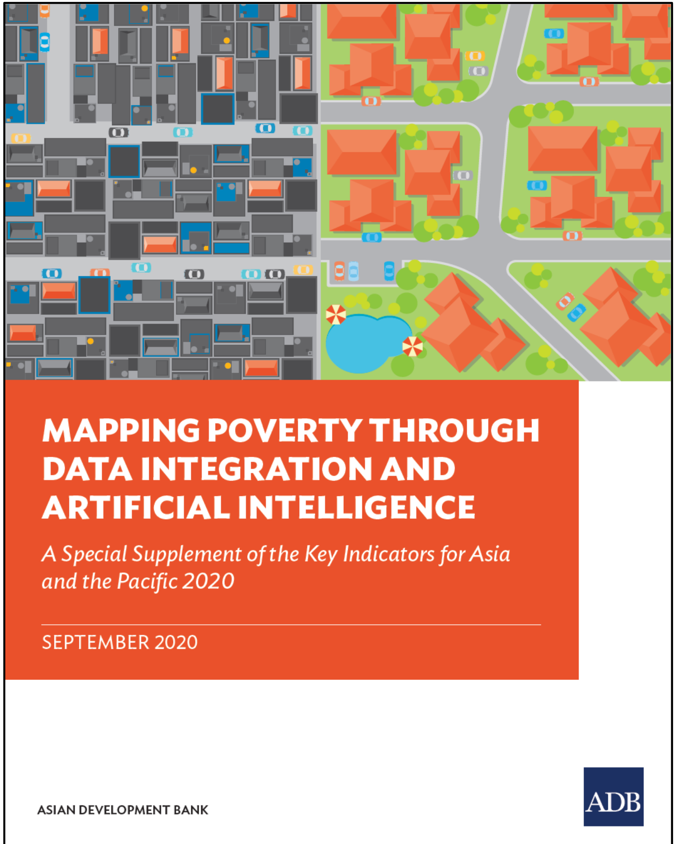


Key Findings

Calibrated Poverty Maps



References

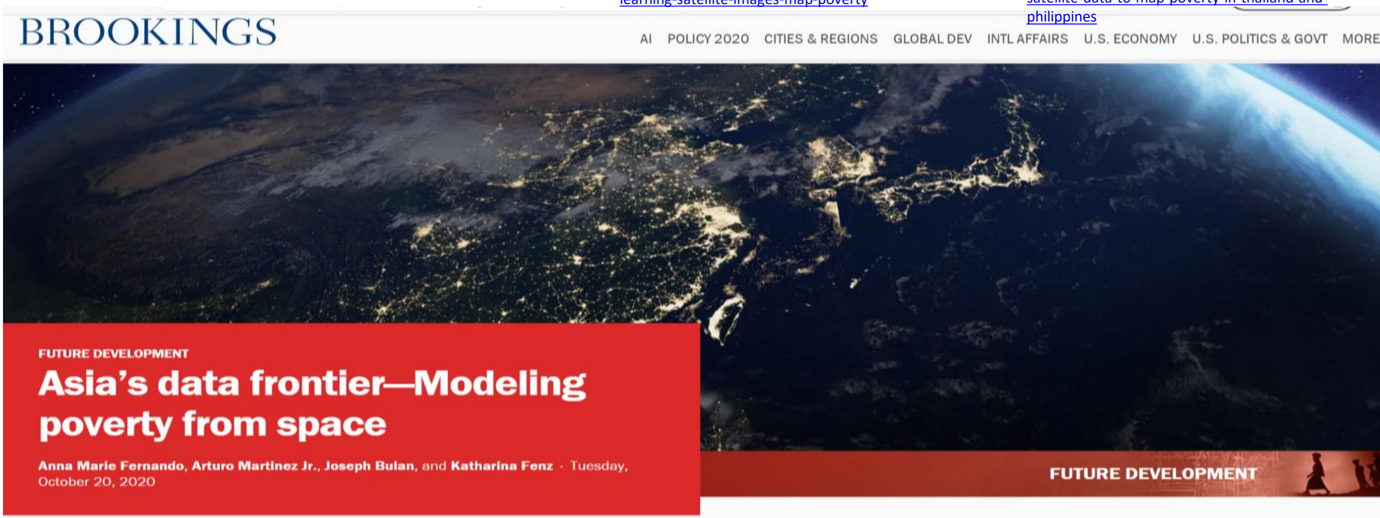


[Mapping Poverty through Data Integration and Artificial Intelligence: A Special Supplement of the Key Indicators for Asia and the Pacific (adb.org)]

https://www.adb.org/news/events/webinar/asian-impact

https://development.asia/insight/using-machine-learning-satellite-images-map-poverty

https://blogs.adb.org/blog/here-s-how-we-used-satellite-data-to-map-poverty-in-thailand-and-philippines



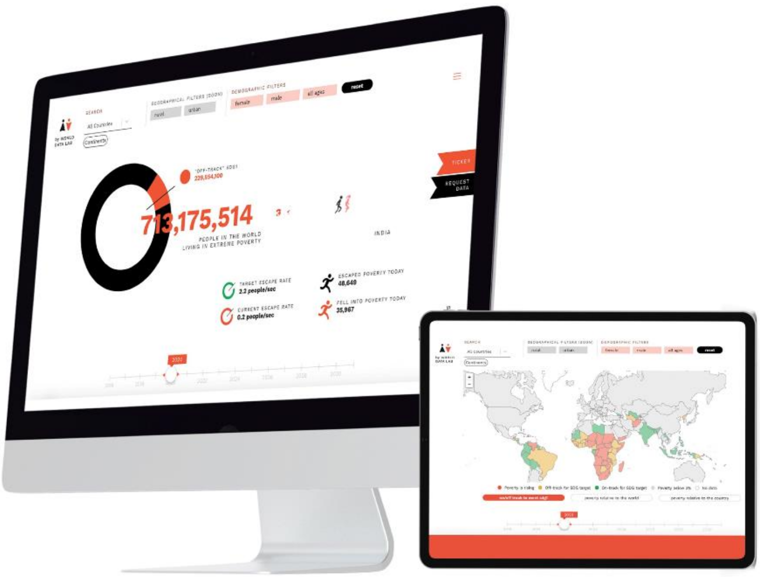
https://www.brookings.edu/blog/future-development/2020/10/20/asias-data-frontier-modeling-poverty-from-space/

Real-time Big Data Model:

World Poverty Clock

WORLD DATA LAB

The World Poverty Clock



WORLD DATA LAB

The African Poverty Clock

AFRICAN POVERTY CLOCK by WORLD DATA LAB

489,744,549

PEOPLE LIVE IN EXTREME POVERTY IN AFRICA

Explore the data

2021

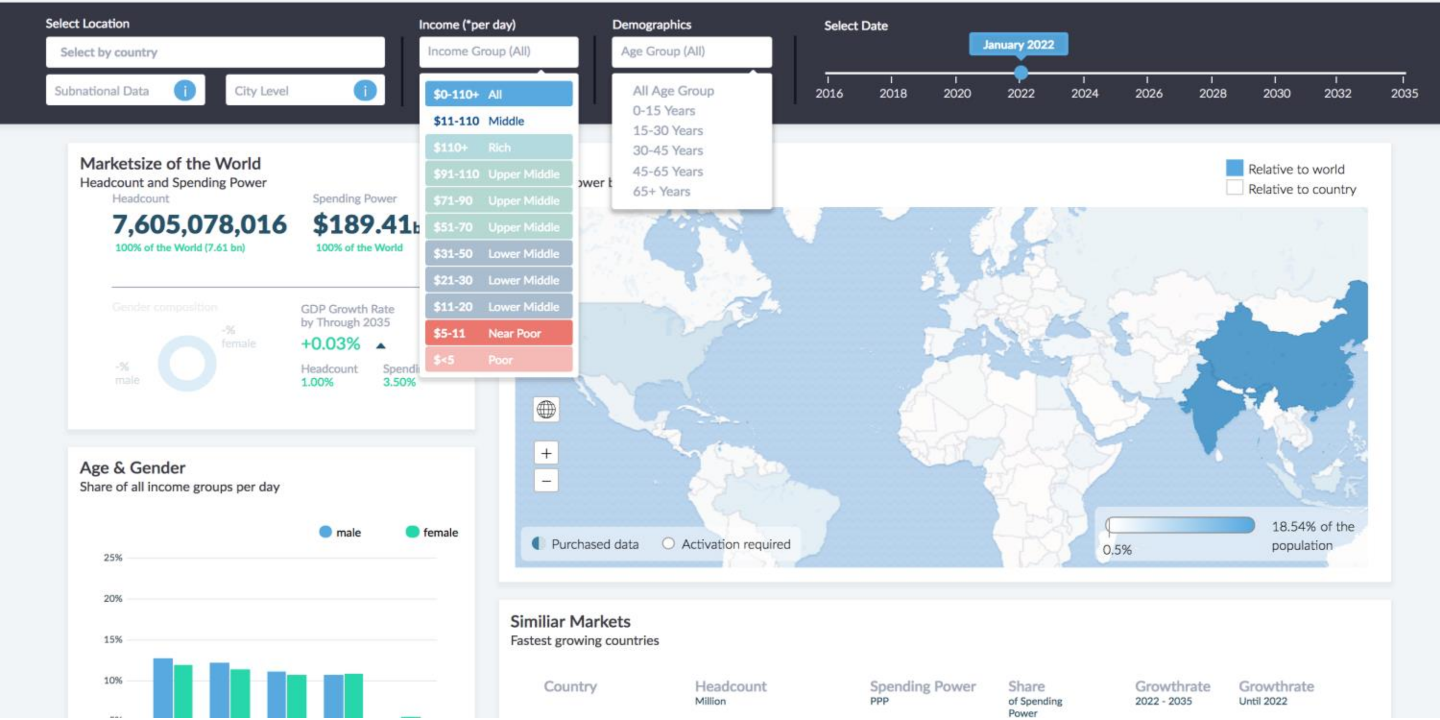
2016 2018 2020 2022 2024 2026 2028 2030

WORLD DATA LAB

MarketPro

CONSUMER SPENDING TIME MACHINE

MarketPro: A new income and demographic time machine



Summary and Moving Forward

Having granular data is important. It helps facilitate more efficient allocation of resources (e.g., Granular poverty estimates produced through the project were used by SERD when implemented ADB's Emergency Food Program for NCR)

Achieving granularity does not necessarily have to prompt data compilers to redesign existing data collection systems and incur significant costs. This can be achieved by embracing the principle of data integration (conventional + innovative data source)

Using publicly accessible satellite imagery is a good starting point, especially for NSOs who are at the exploratory stage. However, scaling up from exploratory studies to more rigorous poverty mapping initiatives could potentially benefit from using higher resolution and adding other types of big data (e.g., call detail records, top up credits from telco, social media data, etc) to provide data beyond spatially disaggregated information.

Summary and Moving Forward

For countries with database of target social protection beneficiaries but are not regularly updated, explore how other data set can complement existing information

- *big data-based poverty estimates can be used to validate the data from the social protection database; areas with significant differences between the two sets of estimates may be prioritized for 'updating'*

For data poor countries with no social protection database, and seldomly collects survey data

- *poverty estimates can be derived from non-traditional data sources, while building capacity in developing the fundamental building blocks for poverty estimation*

THANK YOU.

Email: amartinezjr@adb.org / wfengler@worlddata.io

Calibration

Calibrating Poverty Maps