

Hofer, Martin et al.

Working Paper

Applying artificial intelligence on satellite imagery to compile granular poverty statistics

ADB Economics Working Paper Series, No. 629

Provided in Cooperation with:

Asian Development Bank (ADB), Manila

Suggested Citation: Hofer, Martin et al. (2020) : Applying artificial intelligence on satellite imagery to compile granular poverty statistics, ADB Economics Working Paper Series, No. 629, Asian Development Bank (ADB), Manila,
<https://doi.org/10.22617/WPS200432-2>

This Version is available at:

<https://hdl.handle.net/10419/246706>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/3.0/igo/>

APPLYING ARTIFICIAL INTELLIGENCE ON SATELLITE IMAGERY TO COMPILE GRANULAR POVERTY STATISTICS

*Martin Hofer, Tomas Sako, Arturo Martinez Jr., Mildred Addawe, Joseph Bulan,
Ron Lester Durante, and Marymell Martillan*

NO. 629

December 2020

ADB ECONOMICS
WORKING PAPER SERIES

Applying Artificial Intelligence on Satellite Imagery to Compile Granular Poverty Statistics

Martin Hofer, Tomas Sako, Arturo Martinez Jr.,
Mildred Addawe, Joseph Bulan,
Ron Lester Durante, and Marymell Martillan

No. 629 | December 2020

Martin Hofer (hofer1991@gmail.com) is a lecturer at the Vienna University of Economics and Business. Tomas Sako (tomassako@gmail.com) is a freelance data scientist. Arturo Martinez, Jr. (amartinezjr@adb.org) is a statistician, Joseph Albert Nino Bulan (jbulan@adb.org) is an associate statistics analyst, Mildred Addawe (maddawe.consultant@adb.org), Ron Lester Durante (rdurante1.consultant@adb.org), and Marymell Martillan (mmartillan.consultant@adb.org) are consultants at the Economic Research and Regional Cooperation Department of the Asian Development Bank.

This paper was prepared as background material for the report Mapping Poverty through Data Integration and Artificial Intelligence: A Special Supplement of the Key Indicators for Asia and the Pacific. The authors thank colleagues from the Philippine Statistics Authority and National Statistical Office of Thailand who closely worked with the project team.



Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO)

© 2020 Asian Development Bank
6 ADB Avenue, Mandaluyong City, 1550 Metro Manila, Philippines
Tel +63 2 8632 4444; Fax +63 2 8636 2444
www.adb.org

Some rights reserved. Published in 2020.

ISSN 2313-5867 (print), 2313-5875 (electronic)
Publication Stock No. WPS200432-2
DOI: <http://dx.doi.org/10.22617/WPS200432-2>

The views expressed in this publication are those of the authors and do not necessarily reflect the views and policies of the Asian Development Bank (ADB) or its Board of Governors or the governments they represent.

ADB does not guarantee the accuracy of the data included in this publication and accepts no responsibility for any consequence of their use. The mention of specific companies or products of manufacturers does not imply that they are endorsed or recommended by ADB in preference to others of a similar nature that are not mentioned.

By making any designation of or reference to a particular territory or geographic area, or by using the term “country” in this document, ADB does not intend to make any judgments as to the legal or other status of any territory or area.

This work is available under the Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO) <https://creativecommons.org/licenses/by/3.0/igo/>. By using the content of this publication, you agree to be bound by the terms of this license. For attribution, translations, adaptations, and permissions, please read the provisions and terms of use at <https://www.adb.org/terms-use#openaccess>.

This CC license does not apply to non-ADB copyright materials in this publication. If the material is attributed to another source, please contact the copyright owner or publisher of that source for permission to reproduce it. ADB cannot be held liable for any claims that arise as a result of your use of the material.

Please contact pubsmarketing@adb.org if you have questions or comments with respect to content, or if you wish to obtain copyright permission for your intended use that does not fall within these terms, or for permission to use the ADB logo.

Corrigenda to ADB publications may be found at <http://www.adb.org/publications/corrigenda>.

The ADB Economics Working Paper Series presents data, information, and/or findings from ongoing research and studies to encourage exchange of ideas and to elicit comment and feedback about development issues in Asia and the Pacific. Since papers in this series are intended for quick and easy dissemination, the content may or may not be fully edited and may later be modified for final publication.

CONTENTS

TABLES AND FIGURES	iv
ABSTRACT	v
I. INTRODUCTION	1
II. LITERATURE REVIEW	2
III. DATA AND METHODS	3
A. Reference Period	3
B. Daytime Satellite Images	3
C. Nighttime Satellite Images	4
D. Sources of Poverty Data Used as Inputs	4
E. Shapefiles	5
F. Predicting Poverty Using Artificial Intelligence	5
IV. KEY FINDINGS	7
A. Convolutional Neural Networks Validation	7
B. Ridge Regression	8
V. ROBUSTNESS ASSESSMENT	12
A. Averaged Features versus Averaged Outputs	12
B. Exploring Data Splitting Strategies	12
C. Do Separate Models for Urban and Rural Areas Improve the Prediction Accuracy?	13
D. Ridge Regression versus Random Forest Estimation	13
E. Comparing Results with a Simpler Model	14
F. Harmonizing Artificial Intelligence-Based Predictions with Published Poverty Rates	15
VI. DISCUSSION AND SUMMARY	17
REFERENCES	19

TABLES AND FIGURES

TABLES

1	Prediction Accuracy of Convolutional Neural Networks	8
2	Root Mean Square Error	12
3	Prediction Accuracy of Convolutional Neural Networks Using Alternative Data Splitting Strategies, Validation Set	13
4	Comparison of Predictive Performance of Ridge Regression and Random Forest	14
5	Root Mean Square Error for Poverty Rate (Validation) Small Area Estimation and Survey Level	14

FIGURES

1	Methodology for Predicting Poverty Using Satellite Imagery	6
2	Scatter Plot of Published and Predicted Poverty Rates	9
3	Plots of Multidimensional Poverty Index and Asset Ownership	11
4	Scatter Plot of Published or Survey-Based Estimates and Artificial Intelligence-Based Predictions	11
5	Calibrated Machine-Learning Poverty Predictions	16

ABSTRACT

The spatial granularity of poverty statistics can have a significant impact on the efficiency of targeting resources meant to improve the living conditions of the poor. However, achieving granularity typically requires increasing the sample sizes of surveys on household income and expenditure or living standards, an option that is not always practical for government agencies that conduct these surveys. Previous studies that examined the use of innovative (geospatial) data sources such as those from high-resolution satellite imagery suggest that such method may be an alternative approach of producing granular poverty maps. This study outlines a computational framework to enhance the spatial granularity of government-published poverty estimates using a deep layer computer vision technique applied on publicly available medium-resolution satellite imagery, household surveys, and census data from the Philippines and Thailand. By doing so, the study explores a potentially more cost-effective alternative method for poverty estimation method. The results suggest that even using publicly accessible satellite imagery, in which the resolutions are not as fine as those in commercially sourced images, predictions generally aligned with the distributional structure of government-published poverty estimates, after calibration. The study further contributes to the existing literature by examining robustness of the resulting estimates to user-specified algorithmic parameters and model specifications.

Keywords: big data, computer vision, data for development, machine learning algorithm, official statistics, poverty, SDG

JEL codes: C19, D31, I32, O15

I. INTRODUCTION

When the 17 Sustainable Development Goals (SDGs) were officially launched in 2015, countries all over the world pledged to leave no one behind, with poverty eradication at the top of the list. Included in the Global Indicator Framework for monitoring progress on poverty reduction is the proportion of a country's population living below its national poverty line, which is typically sourced from household income and expenditure surveys. From the perspective of statisticians and other compilers of official poverty statistics, the SDGs' *leave no one behind* principle which requires data to be disaggregated by geographic location, ethnicity, gender, income class, and other relevant dimensions, presents several challenges. Conventionally, household income and expenditure surveys have sample sizes that are sufficient to provide nationally representative poverty estimates, but not large enough to provide reliable estimates at levels granular enough to meet all disaggregated data requirements of SDG 1. Moreover, the sample sizes are inadequate in providing reliable estimates at levels granular enough to allow development planners efficiently target areas that need immediate poverty intervention.

In addition to issues on granularity, there are also concerns on timeliness and frequency of release of updated poverty data. Conducting household income and expenditure surveys can be expensive and consumes a lot of time particularly in planning, executing and publishing results. This means that in many countries, the lag time for the release of data is 1 year or more, depending on the information being provided. However, financial resources from national governments and international development organizations that support statistical programs such as conducting household income and expenditure surveys are limited and sparse (PARIS21 2017). Hence, it is important for national statistical systems to explore alternative cost-effective methods of meeting the growing demand for SDGs and other types of data for development.

In 2017, the Asian Development Bank (ADB) designed a project which aims to strengthen the capacity of national statistics offices (NSOs) in the Asia and Pacific region to meet the disaggregated data requirements of the SDGs using innovative data sources such as satellite imagery. One component of the project conducted a case study to examine the feasibility of using publicly accessible satellite imagery and computer vision techniques to enhance the granularity of poverty statistics compiled using conventional estimation methods. Using publicly available resources is attractive for NSOs and development institutions who are planning to conduct similar feasibility studies, as they serve as an excellent starting point to learn advanced machine learning techniques without incurring significant financial costs upfront.

The method is applied using data from the Philippines and Thailand. These two countries are good case studies as both are compiling small area poverty estimates which provide enough data that can be used in the training of algorithms capable of predicting poverty. Both countries also have slightly different poverty profiles with Thailand having a lower prevalence of poverty. This variation allows for the examination of the impact of variability in the indicator of interest on the performance of an algorithm designed for image classification-related tasks. The study further contributes to the existing literature by examining robustness of the resulting estimates to user-specified algorithmic parameters and model specifications, which may serve as a means of strengthening confidence in poverty maps compiled through this innovative methodology.

II. LITERATURE REVIEW

As briefly discussed earlier, poverty statistics are typically compiled based on data collected from household surveys. However, sample sizes of these surveys are typically not large enough to provide reliable estimates at more granular levels, and therefore resulting poverty estimates may not be reliable at very granular disaggregation levels. Increasing sample sizes is a way to enhance reliability of survey estimates, but it is often not practical as achieving such increases requires significant additional resources, which are not readily available to NSOs or the organizations that conduct these surveys (ADB 2020).

Alternative methods to compile poverty statistics to address some of these concerns are being explored. For instance, many countries have attempted to use small area estimation (SAE) techniques, in collaboration with development partners like the World Bank, by combining survey results with census and other auxiliary data to produce granular statistics. Conventional small area poverty estimation techniques have limitation which can often lead to increased model errors if the survey and census of population used were conducted years apart. To address this error, the key is to work with variables whose values do not change significantly over time. However, this approach may not always lead to a better model as imposing such criterion limits the number of covariates that can be included in model-building and affect the model's predictive performance. Other limitations of the SAE include its inability to address the issue of timeliness of statistics. If surveys and census data are not collected frequently, this will not lead to updated granular data even if SAE techniques are employed.

On the other hand, there have been attempts to integrate beyond traditional types of data such as those coming from surveys and censuses. A good example is the use of satellite imagery for various development indicators, and there are several reasons why its popularity is increasing. For one, advances in satellite-based socioeconomic measurements have led to an influx of high frequency data for both, data-rich and data-poor environments. One of these measurements is nighttime light intensity which has been increasingly used following the initial works of Chen and Nordhaus (2011) and Henderson, Storeygard, and Weil (2012). This helped mitigate some of the known data shortcomings, including those of the SAE, if enhancing granularity is the main objective. Nighttime light intensity can also be used to estimate values in between surveys and enable nowcasting as well as help illuminate areas that are less covered by surveys and censuses. However, using data on nightlights alone have several drawbacks. The data produced by satellites are top-coded which makes highly developed, urbanized areas hard to differentiate. On the other side of the spectrum, the least developed areas often do not have measurable nighttime lights, and this makes it difficult to obtain estimates for proxy measures of socioeconomic development in such areas.

Building on these developments, the use of daytime satellite images has started to become an important focus of research. Xie et al. (2015) showed that poverty mapping using satellite imagery in combination with transfer learning and convolutional neural networks (CNN) can lead to the predictive performance of survey data collected in the field. Jean et al. (2016) trained a CNN to extract features in high-resolution daytime images using nighttime images as labels. The extracted features were used to predict asset wealth and consumption expenditure across five African countries. Jean et al. (2016) were able to provide that such a model is strongly predictive of both average household consumption expenditure and asset wealth as measured at the cluster level for countries where recent survey data is available. On the other hand, Head et al. (2017) has proven that this method does not generalize in the same way that other measures of development predict access to drinking water and a variety of health

and education-related indicators. It is possible to apply this method in other countries and continents given certain limitations.

The study presented in this manuscript serves as a proof of concept in implementing the techniques used by Jean et al. (2016) using only publicly available satellite data that have lower resolution and are readily available tools for data processing, akin to the objective of Yeh et al. (2020) which used the same kind of satellite imagery to examine spatial distribution of economic well-being in Africa. Whereas some of the follow-up studies that adopted similar techniques (e.g., Engstrom, Hersh, and Newhouse 2017; Tingzon et al. 2019; Yeh et al. 2020) focus in estimating the spatial distribution of wealth, income, or poverty for years or areas that are not covered by conventional data sources, this study focuses in predicting poverty in the same year where conventional data sources are available, albeit, providing more spatially disaggregated numbers. Furthermore, this study contributes to existing literature on using computer vision techniques to map the spatial distribution of poverty by examining the robustness of results to algorithmic training parameters and specifications.

III. DATA AND METHODS

A. Reference Period

Our target reference period coincides with the years when small area poverty estimates are available—2012 and 2015 for the Philippines and 2013, 2015, and 2017 for Thailand. For the Philippines, we also included 2018 even if small area poverty estimates have yet to be compiled and only provincial-level survey estimates of poverty are available. Inclusion of the 2018 data allows for the examination of whether the granularity of the indicator to be used for training a machine learning algorithm has an impact on the extent of granularity at which predictions can be made.

B. Daytime Satellite Images

Publicly accessible satellite images were used for this study, particularly Landsat (15-meter resolution) and Sentinel (10-meter resolution). The higher resolution of Sentinel is preferred, but since Sentinel was launched in 2015, we had to use alternative source of satellite imagery.¹ In particular, we use Landsat satellite images for 2012 and 2013, while Sentinel was used for all later years. The resolution of the satellites dictates the size of images that can be used and the final resolution of artificial intelligence-based poverty predictions. The study used 256 pixel by 256 pixel images for Landsat years and 384 pixel x 384 pixel images for Sentinel years. The research also tested smaller images (128 pixel by 128 pixel) for Thailand, but the CNN did not perform well.² However, Landsat's 256 pixel by 256 pixel images performed well for CNN training. This provided a grid size of 3,840 m (15 meters per pixel multiplied by 256 pixels) at the equator. To reuse the same grid for all years, we use 384 pixels for Sentinel images too. This gave roughly 36,000 images in Thailand and 13,000 in the Philippines. In succeeding analyses, the center of each image was used as reference point when identifying which municipality, city, province, or *tambon* (i.e., subdistrict) it belongs to.

¹ ESA Earth Online. <https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/sentinel-2>.

² The use of smaller-sized images inflates CNN's prediction errors by roughly 10%.

The input data for this study was obtained using georeferenced and tagged image files sourced from Landsat 8 and Sentinel 2. These image files were stored as three-dimensional arrays, with each pixel represented in red, green, and blue color bands.³

One of the intermediate steps undertaken while preparing the satellite images was to collect cloud-free daytime images that covered the entirety of both the Philippines and Thailand. This process entailed running an algorithm to select the best daytime images falling within an acceptable cloudiness or cloud cover during the period of the study. Although it is ideal to set a low cloudiness threshold, setting it too low could lead to having several areas with no image that satisfy the threshold. The cloudiness threshold was determined by experimenting until the algorithm produced a composite image for the whole country with the least amount of cloud cover with adequate number of images to train on.⁴

The use of “pansharpening” was another intermediate data preparation step undertaken to enhance the resolution of the Landsat 8 images. Pansharpening combines high resolution panchromatic images (black and white but sensitive to colors) with lower resolution multispectral band images. The process produces a single red, green, blue (RGB) color, multiband image at the higher resolution of the panchromatic image. From the original 30-meter resolutions, pansharpening was able to produce 15-meter resolutions for the Landsat images.

C. Nighttime Satellite Images

There are many sources of nightlight intensity data. However, the best known and publicly published are datasets based on Defense Meteorological Satellite Program Operational Line-Scan System (DMSP-OLS) and Suomi National Polar-Orbiting Partnership Visible Infrared Imaging Radiometer Suite (SNPP-VIIRS) missions.

Both were conducted by the National Oceanic and Atmospheric Administration (NOAA). It was decided that images from VIIRS will be used for this study because it offered a substantial number of improvements over Operational Line-Scan System as stated in the work of Elvidge et al. (2013).

A cloud-free average radiance value was used to filter out the effects of fires and other transitory events as well as irrelevant background, while unlit areas were set to zero.

D. Sources of Poverty Data Used as Inputs

Since the main objective is to predict the proportion of a country’s population living below its national poverty line using satellite imagery, the study needed ‘ground truth’ data on poverty for training a machine learning algorithm and validation.

For this purpose, the data used were the municipal- or city-level estimates of proportions of population living below the national poverty line in 2012 and 2015 as compiled by the Philippine Statistics Authority (PSA) using SAE technique. The SAE technique adopted by PSA combined data from the Family Income and Expenditure Survey and Census of Population and Housing. The provincial

³ In principle, satellites have multiple bands including infrared. This study focused on RGB color bands only because the CNN has been pretrained on regular RGB images (through ImageNet). In addition, the artificial intelligence framework used in this study required .jpg files as input, which only support the RGB bands as well.

⁴ We use Google Earth Engine’s built-in function for creating annual composite image based on a user-defined cloudiness threshold parameter.

poverty estimates in 2018 were also used as these were directly estimated from Family Income and Expenditure Survey by PSA. The small area poverty estimates for 2018 are not yet available at the time this study was conducted.

In Thailand, this study used the data compiled by the National Statistical Office of Thailand using SAE technique applied on its Socioeconomic Survey and Census of Population. The poverty data estimates were at the *tambon* level for 2013, 2015, and 2017.

For succeeding sections, municipalities, cities, and *tambons* are referred to as SAE-level areas while provinces are referred as survey-level areas. Collectively, SAE-level and survey-level areas are coined using levels where government-published poverty estimates are available. These terminologies are used in succeeding sections.

E. Shapefiles

To combine spatial datasets, *tambon*-level shapefiles for Thailand and municipal- or city-level and provincial-level shapefiles for the Philippines were used. Thailand has more than 7,400 *tambons* while the Philippines have more than 1,500 municipalities and cities and more than 80 provinces. On the average, there are eight (4 kilometer by 4 kilometer) images in a municipality and 152 images in a province in the Philippines. In Thailand, there are five images in a *tambon*, on average.

F. Predicting Poverty Using Artificial Intelligence

Data from satellite images are inherently unstructured, noisy, and hard to process both statistically and computationally. In recent years, CNNs have propelled the field of computer vision forward. This created opportunities to use computer vision on satellite images to gain information from the ground.

Since the main objective of the study was to predict the prevalence of poverty at granular grid-level using satellite imagery⁵, it was initially thought that directly linking poverty and daytime satellite images with computer vision techniques would yield the best results. However, poverty data with the granularity and frequency needed to train large neural networks was rare. To work around this issue, the researchers follow the approach published by Jean et al. (2016), as described below.

First, the researchers trained a CNN to predict the intensity of night lights using daytime images as input. In this study, the CNN was pretrained on ResNet-24 which has a deeper layer neural network than Visual Geometry Group's Face model which was used by Jean et al. (2016). A deeper later CNN could potentially capture more complex features and enhance accuracy of predictions. Intensity of lights at night is commonly used as proxy for wealth and human interaction on the ground when working with satellite imagery (Henderson, Storeygard, and Weil 2012; Addison and Stewart 2015; Keola, Andersson, and Hall 2015). At this stage, the CNN learned to detect general features on daytime images, related to light intensity.⁶ These features, or latent variables, can be used for other tasks, like estimating poverty measures. The condensed image-based information can be averaged on a coarser level (i.e., from image grid level to a level where we have ground truth data). To speed up the training, and reduce the amount

⁵ As explained earlier, the study examined different levels of granularity and chose the level at which CNN performed reasonably well.

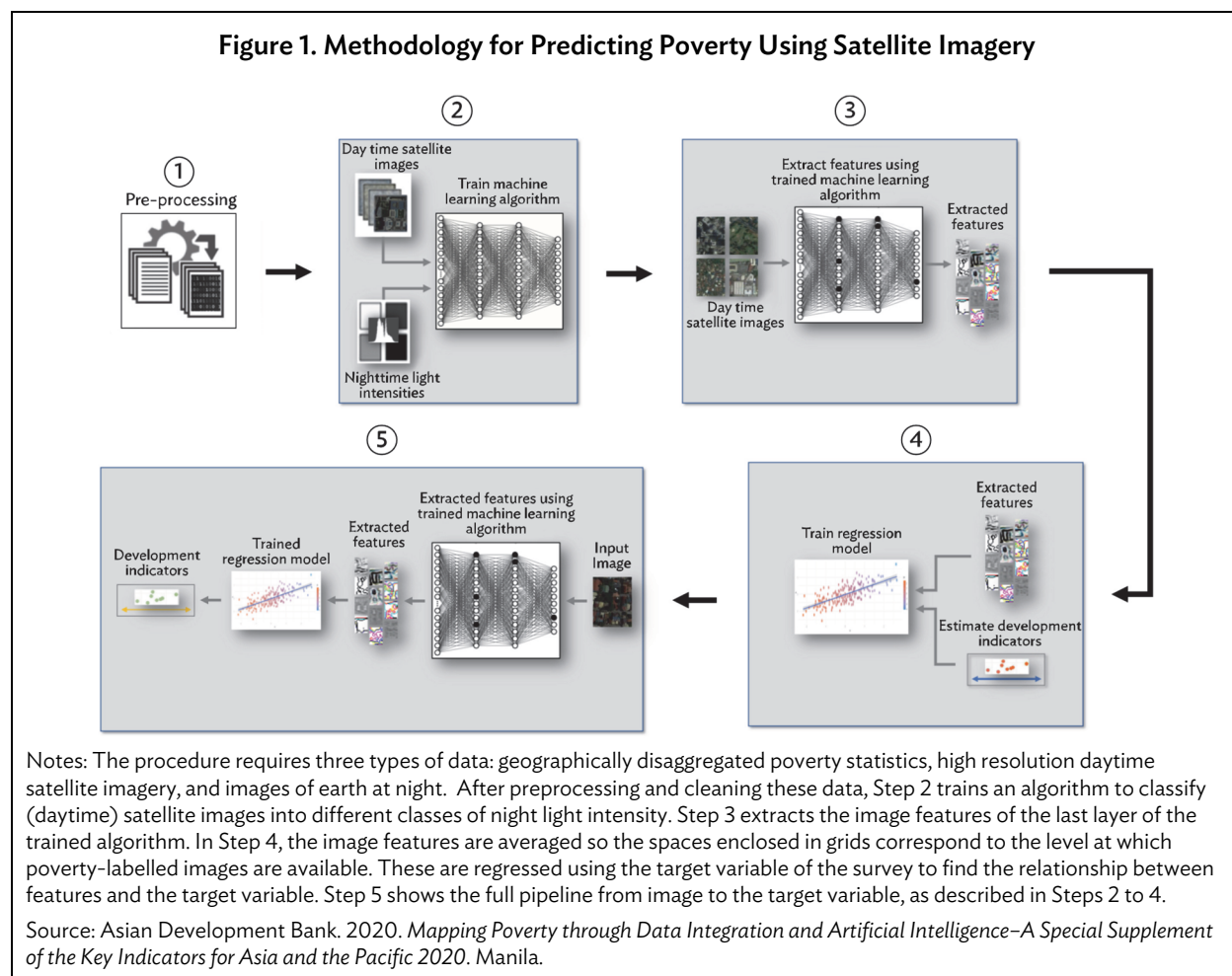
⁶ The distribution of intensity of night lights is grouped into different clusters. Ideally, clusters are determined by a Gaussian mixture model with different nightlight ranges for each country-year combination. However, we also use heuristic methods when Gaussian mixture model does not produce satisfactory results.

of data needed to train, we use a network that has already been trained on the ImageNet database to assign labels to a large database of images (Deng et al. 2009). This is a well-known technique and referred to in the literature as transfer learning (Pan and Yang 2010).

Second, the research team dropped the prediction of nighttime light intensity and solely used the trained CNN as a function that summarizes the complex multidimensional input of image data into a single vector. This vector has 512 elements called “features” with a single feature vector per image. These features are a representation of what the network detected on the image. These had several advantages over the raw pixel values, most notably that convolutional layers scan over the image using kernels so that it does not matter where features will appear on the image.

To combine the grid-based image features with the SAE or survey-based poverty data, the team took the average value of each feature within the given SAE or survey areas. In the final training step, we used ridge regressions to examine the relationship between the image features and. The trained CNN and ridge parameters were then used to predict image level poverty using only a daytime image as input. Finally, we examined whether the results depend on data splitting strategies and type of machine learning algorithm, a feat which has not been thoroughly explored in previous studies.

Figure 1 summarizes the steps described above.



IV. KEY FINDINGS

For each country-year combination, we set aside 90%⁷ of the given SAE and survey areas for training and 10% for validation. The training set is used for model parameter estimation while the validation set is used to assess the model's predictive performance of the final poverty results. This is in addition to the separate validation of both the CNN and the ridge regression, that use 10-fold cross validation to tune hyperparameters and measure performance. The research team also explored alternative data splitting strategies but as this section shows, the researchers did not find significant differences in the results when no other parameters were changed.

A. Convolutional Neural Networks Validation

For CNN training, the study team used ResNet34 pretrained model on ImageNet database.⁸ A weighted cross entropy loss function is also used to mitigate imbalanced prediction classes.

Data augmentation is used for daytime images to get more samples in the training dataset, in order to get a model which generalizes better and to mitigate imbalanced classes in datasets. Specifically, vertical and horizontal flipping, random lighting, and contrast change with 10% probability, dihedral and symmetric warp are applied in order to enhance relatively small datasets. During the experiments, the team found out that data augmentation was needed, as models suffered from significant overfitting. These types of augmentations were used because of their suitability for remote sensing images (Perez and Wang 2017).

The study team also experimented with several parameters, such as learning rate, number of epochs, and weight decay to optimize the CNN model. All experiments were performed on Fastai framework.

After validation of the top losses of CNN predictions, we found that using partly clouded images (with semi-transparent clouds) is a challenge for CNN. For cloud removal on Sentinel images, a cloud masking technique was used. However, further cloud removal will be needed for future research.

Table 1 shows the overall accuracy of CNN predictions. Accuracy rate was computed by dividing the total number of images that were correctly classified according to the class of luminosity of night to total number of images in the validation dataset.

⁷ Roughly 160 Philippine cities and municipalities in 2012 and 2015, and 8 provinces in 2018 (or equivalent to approximately 1,500 images per year), were randomly selected to constitute the validation set. On the other hand, about 700 tambons (or equivalent to approximately 3,500 images per year) comprised the validation set in Thailand.

⁸ In this context, pretraining refers to using an algorithm that can already detect image features, such as lines, edges, etc. instead of developing an algorithm from scratch. The algorithm is trained on a wide range of images. Many features relevant for general image classification tasks overlap, regardless of the source of the image. The database used for pretraining in this study, ImageNet, is regarded as a solid benchmark performer in computer vision predictions.

Table 1. Prediction Accuracy of Convolutional Neural Networks
(%)

Country	Thailand	Thailand	Thailand	Philippines	Philippines	Philippines
Year	Year 1 2013	Year 2 2015	Year 3 2017	Year 1 2012	Year 2 2015	Year 3 2018
Validation set	85.79	85.22	86.43	94.15	93.50	92.91
Full dataset	86.52	87.28	86.98	94.66	93.86	90.56

Source: Calculations generated by the study team.

There is small difference between prediction accuracy of Thailand and Philippine models. This is potentially caused by different distribution of nightlight values in these countries. As the Philippines has low and sparse nightlight values, it was easier for the model to correctly classify areas with low levels of intensity of nightlight class (Deng et al. 2009) where no or only few man-made structures were visible and therefore, easier to get higher prediction accuracy.

B. Ridge Regression

As explained earlier, prediction of night lights intensity is just an intermediate step in the method adopted in this study. The main objective is to compile estimates of poverty that are more granular than government-published numbers. To do this, the research team leveraged on computer vision techniques, particularly CNN, to learn features from satellite images that could be used to predict the levels of poverty in a specific area. Although directly linking poverty and daytime satellite images with CNN would probably yield the best results in line with this objective, poverty data with sufficient granularity needed to train a CNN model is uncommon. The intermediate step taken was to first train the CNN to predict the intensity of night lights. This was necessary because night light images provided more labelled data where the model can be trained to group daytime satellite images into different clusters with features associated with different levels of economic development.

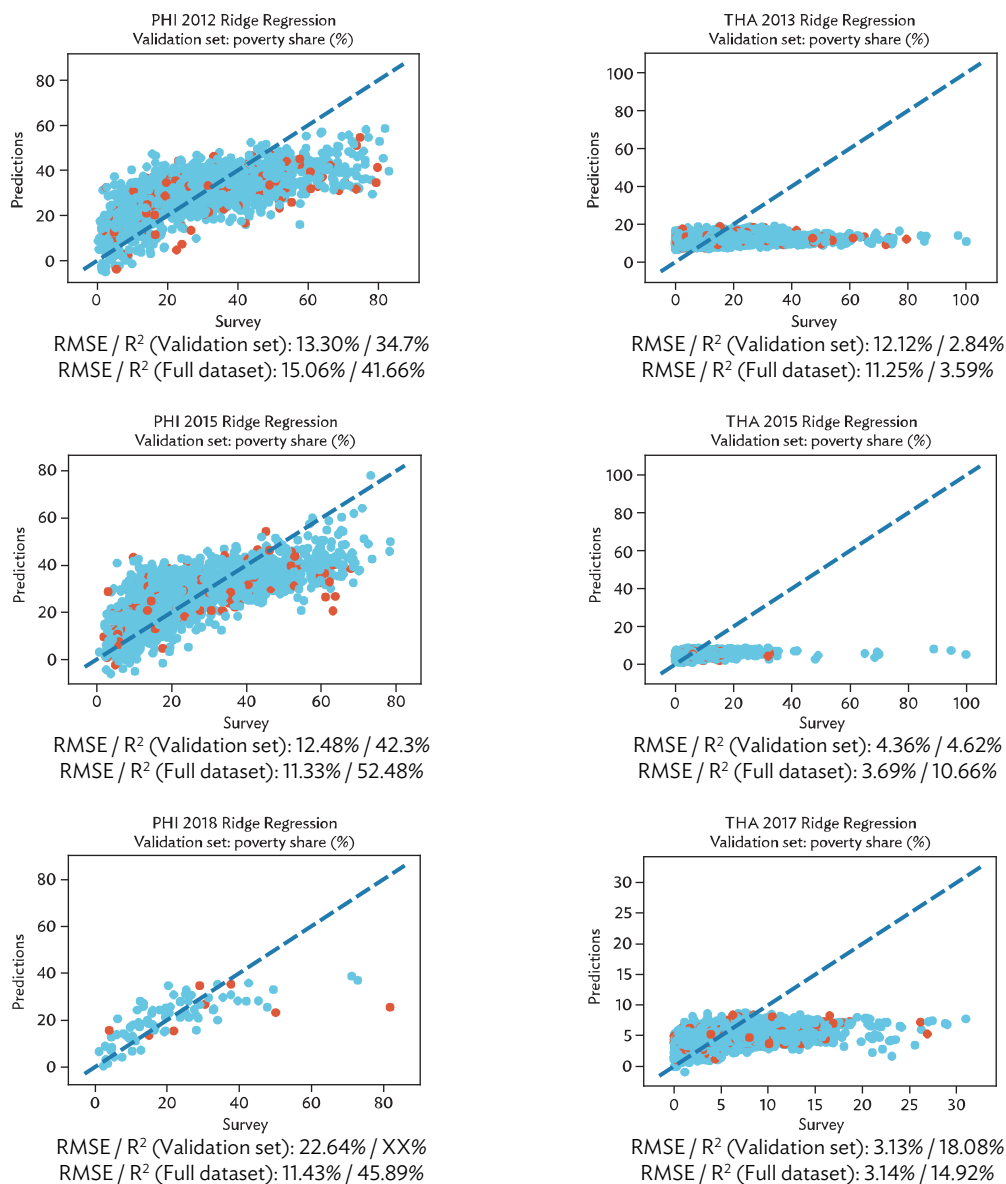
After implementing the CNN, the next step in the study was to extract the features within the satellite images that were used in predicting the night light intensity. This was done by modifying the CNN such that it produced its last hidden layer as an output instead of the regular output layer. In other words, instead of the classification based on intensity of night lights, the study team extracted the feature vectors the CNN uses to determine the intensity of night lights. These data were aggregated by taking the element wise average values of the vectors at the SAE or survey-area level. These were regressed on the aggregated data using ridge regressions. Ridge regression was chosen because compared to Ordinary Least Squares and other regression modelling techniques, it had a better out-of-sample predictive performance when the number of covariates was large. Further, ridge regression penalized large model coefficients to shrink unimportant ones toward zero.⁹

The results of ridge regression are summarized in Figure 2. Using the root mean square error (RMSE) as a metric, the study team observed higher out of sample errors for the Philippines using 2018 data (Philippines 2018), presumably due to this specific country-year combination having a more aggregated poverty data used as input (i.e., provincial level) relative to other country-year combinations.

⁹ The study team experimented on excluding outliers, cutting the CNN at different points with different sizes of feature vectors, and taking the logarithm of poverty but the results did not change dramatically. Alternative modeling strategies—such as random forest estimation, support vector regression, and others—can also be considered for future studies (Puttanapong et al. 2020).

Meanwhile, the lower resolution associated with Landsat images could explain the higher error rates observed in the training sample of PHI 2012. Results for Thailand which showed higher prediction error in 2013 support this claim.¹⁰

Figure 2. Scatter Plot of Published and Predicted Poverty Rates



PHI = Philippines, RMSE = root mean square error, THA = Thailand.

Notes: Blue dots represent the training sample while red dots represent validation data that has not been used to train the parameters.

Source: Calculations generated by the study team.

¹⁰ In future experiments, studies can compare the results using Landsat and Sentinel images with the same input poverty data to have a more conclusive assessment of how the resolution of images affect the method's predictive accuracy.

However, RMSE values do not provide a complete picture, particularly in the case of Thailand since many areas have very low SAE-level poverty rates. This means that even a very basic model can produce low RMSE values simply by predicting a value which is close to the average or mean poverty rate.¹¹

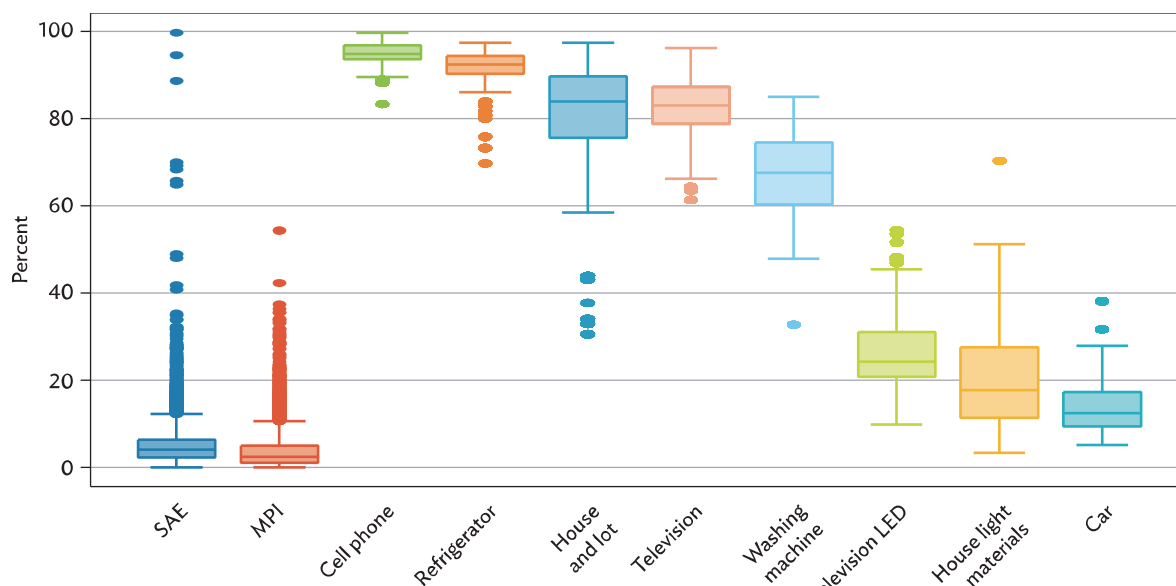
Scatter plots may also be used as an alternative validation tool. In a scatter plot, each dot represents one city and municipality, province, or *tambon*. In the plots above, the x-axis contains the government-published poverty estimates while the artificial intelligence-based predictions are on the y-axis. Despite Thailand having lower RMSE values than that of the Philippines, it can be observed that artificial intelligence-based method does not perform well in predicting higher levels of poverty rates. The estimated values for adjusted R^2 reaffirm this point. One possible reason for this is that there are significantly more *tambons* with low poverty rates than *tambons* with poverty rates exceeding 20%. The lack of variability in the poverty data used in the study may have also contributed to the algorithm's underestimation of poverty distribution in Thailand.

To test this hypothesis, the study team explored other socioeconomic indicators with different levels of variability. For this purpose, the multidimensional poverty index (MPI) was considered as well as the proportion of households owning different types of assets and durable goods. The National Economic and Social Development Council of Thailand compiled *tambon*-level MPI, following a similar methodology developed by the Oxford Poverty and Human Development Initiative (Alkire et al. 2019). In addition to the MPI, the proportion of the population owning durable goods like cellular phone, refrigerator, television, washing machine, house and lot, car and proportion of households were estimated with houses made of light materials for every province represented in the Socioeconomic Survey conducted by the National Statistical Office of Thailand in 2015.

As Figure 3 shows, relative to the distribution of income poverty rates in Thailand, the variability in the distribution of the MPI estimates is comparable with that of income poverty. However, the individual distributions measuring the ownership of assets and durable goods have larger variability.

The research team replicated the step-by-step procedure of artificial intelligence-based method. However, instead of estimating a ridge regression model for income poverty rates, a separate ridge regression model was estimated for each of the alternative socioeconomic indicators cited above. The predictive performance of the model is summarized in the scatter plots below (Figure 4). The results suggest that predictive performance is better when dealing with a variable that has a reasonable amount of variation.

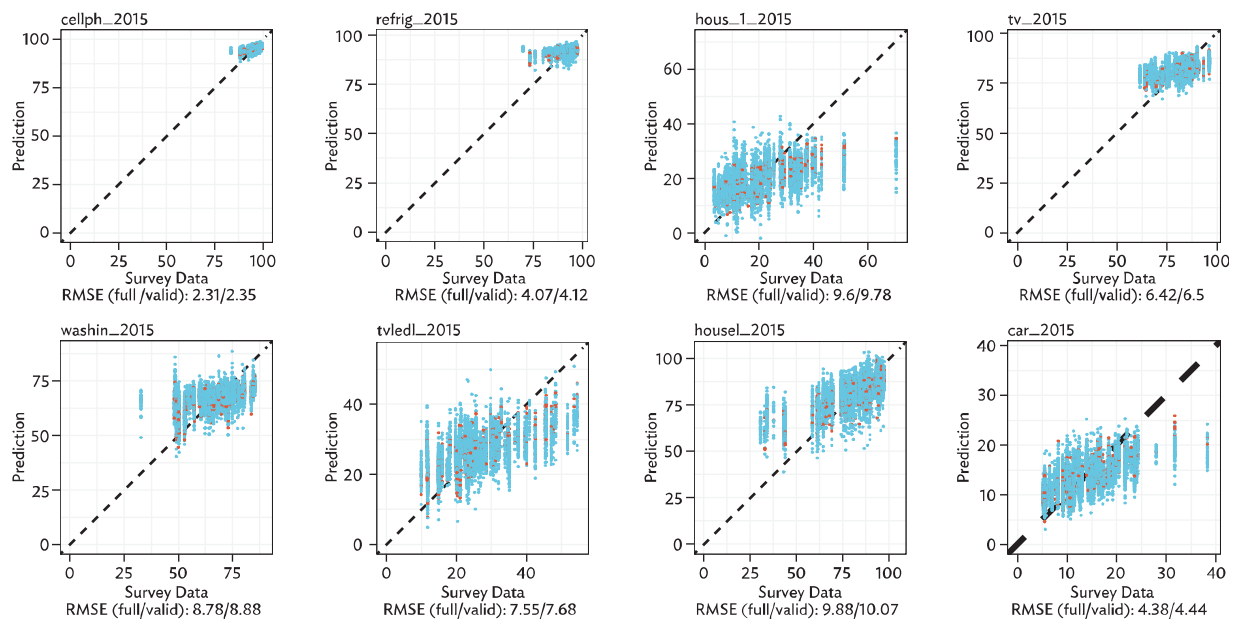
¹¹ Different modeling strategies were considered, but the data fit did not improve significantly. Meanwhile, the appendix presents the results of an experiment that was conducted which examined other socioeconomic indicators, in addition to income poverty rates, some of which have more variability. The results revealed better predictive performance when working on indicators with reasonable amount of variability.

Figure 3. Plots of Multidimensional Poverty Index and Asset Ownership

LED = light-emitting diode, MPI = multidimensional poverty index, Ref = refrigerator, SAE= small area estimates of poverty, TV = television.

Notes: The boxplots show the distribution of income poverty rates, along with other nonincome based proxy measures of poverty. The boxes represent the middle 50% of values of each metric while the lines extending from the boxes represent 1.5 times the interquartile range from the lower and upper quartiles. The dots represent the outliers.

Source: Calculations and graphics generated by the study team.

Figure 4. Scatter Plot of Published or Survey-Based Estimates and Artificial Intelligence-Based Predictions

Source: Calculations and graphics generated by the study team.

V. ROBUSTNESS ASSESSMENT

A. Averaged Features versus Averaged Outputs

To derive artificial intelligence-based poverty rates on the same level where SAE or survey-based estimates of poverty were used, two approaches could be considered. The first one comes from the ridge regression training. The study team used the averaged feature vectors and regressed them on published poverty rates. With this model, the study could predict back the published SAE and survey-level poverty rates. The second approach comes from the image level estimates. In particular, the trained ridge parameters on the image level features were used to get image level poverty rates. These were aggregated back to the SAE and survey level while controlling for gridded population.¹² In general, these two approaches yielded almost the same RMSE values (Table 2), suggesting that the choice of aggregation method was not very critical in our study.

Table 2. Root Mean Square Error
(%)

Year	Philippines		Thailand	
	Averaged Features	Averaged Outputs	Averaged Features	Averaged Outputs
Year 1	15.06%	13.00%	11.25%	11.00%
Year 2	11.33%	11.00%	3.69%	4.00%
Year 3	11.43%	12.00%	3.14%	3.00%

Source: Calculations generated by the study team.

B. Exploring Data Splitting Strategies

In this study, we originally set aside 90% of the given SAE and survey areas for training and 10% for validation. This is in addition to the separate validation implemented for both the CNN and ridge regression, which used tenfold cross validation to tune hyperparameters and measure performance. In existing literature, there is hardly any agreement on the optimal allocation of splitting data between training and validation (Xu and Goodacre 2018). The study team allocated 10% for validation (similar to Babenko et al. 2017; Sohnesen, Fisker, and Malmgren-Hansen 2019) because one of the input datasets specifically, the 2018 poverty rates in the Philippines had few data points. Increasing the allocation for validation would significantly reduce the amount of data that could be used for training and therefore, may yield unreliable results. Nevertheless, the study team also explored alternative data splitting strategies. For the purpose of simplicity, we have reimplemented the methodology for only 1 year. As shown in Table 3, we did not see any significant improvement in the prediction accuracy when we reduced the size of training set and increased the size of validation set.

¹² Grid-level population estimates were separately estimated following the method done by Stevens, et al. (2015).

Table 3. Prediction Accuracy of Convolutional Neural Networks Using Alternative Data Splitting Strategies, Validation Set (%)

Dataset	90–10	80–20	70–30
PHI, Year 2	93.5	92.65	92.42
THA, Year 2	85.22	85.58	86.52

PHI = Philippines, THA = Thailand.

Source: Calculations generated by the study team.

C. Do Separate Models for Urban and Rural Areas Improve the Prediction Accuracy?

When we examined the results more closely by comparing the artificial intelligence-based predictions with government-published estimates, we noted that some of the pronounced differences between the two sets of estimates were observed in cities and highly urbanized areas with large ‘green spaces.’ For such areas, the artificial intelligence-based method tends to overestimate poverty rates than government-published estimates. Initially, we hypothesized that the artificial intelligence model treats those green spaces as agricultural areas, and consequently leads to higher poverty estimates. To examine such hypothesis, we attempted to have separate CNN models for urban and rural areas within each country. This is slightly different from the original approach that we adopted which trained images for urban and rural areas combined. The results suggest that the prediction accuracy of CNN, when trained with urban area images only, declined further. On the other hand, the prediction accuracy of CNN when trained with rural area images only was approximately the same as when we trained all urban and rural area images together. In other words, we did not observe significant improvement in the prediction accuracy by training CNN separately for urban and rural areas.

D. Ridge Regression versus Random Forest Estimation

As discussed earlier, the second part of the adopted method entails estimating a ridge regression on poverty rates using the aggregate features extracted from CNN. In addition to ridge regression, we also checked how the results will change if random forest estimation was used instead. Random forest estimation is one of the most commonly used machine learning models due to its strengths. It has precise learning algorithms, operates well on a large dataset, can manage thousands of input variables, provides estimates on variables that are critical in the classification, and it is capable of estimating missing data (Chakure 2019, Mwit 2020).

Table 4 summarizes the results of our comparison of ridge regression and random forest’s predictive performance. We present the results for only 1 year for simplicity. The results show that the prediction accuracy for the training dataset dramatically improves when random forest estimation is used. This is potentially driven by the ability of the random forest’s modeling framework to yield very precise estimates from large sets of training data. However, when we compare the results for the validation set, the difference between the ridge regression and random forest becomes less remarkable and we expect that it will be much less if the size of validation set is increased further as random forest tends to overfit the model onto the training dataset but may not necessarily perform as well using the validation dataset. Hence, the results based on ridge regression appear to be more robust to the size of training and validation datasets.

Table 4. Comparison of Predictive Performance of Ridge Regression and Random Forest

Data	Metric (%)	Ridge Regression		Random Forest	
		Validation Dataset	Full Dataset	Validation Dataset	Full Dataset
PHI, Year 2	RMSE	12.48	11.33	11.07	5.08
	Adjusted R ²	42.29	52.43	25.96	90.06
THA, Year 2	RMSE	4.36	4.43	3.64	2.19
	Adjusted R ²	3.97	9.25	-0.11	77.67

PHI = Philippines, RMSE = root mean square error, THA = Thailand.

Note: The estimates are based on the original 90–10 training-validation data split.

Source: Calculations generated by the study team.

E. Comparing Results with a Simpler Model

Does the use of daytime imagery and deep learning, artificial intelligence-based method perform better than a simpler structural model? To answer this question, the published SAE or survey-level poverty rates were regressed on the average nightlight intensity and an intercept using ordinary least squares.¹³ The resulting parameter estimates were then used to predict image level poverty rates. The RMSEs are presented below (Table 5).

Table 5. Root Mean Square Error for Poverty Rate (Validation)
Small Area Estimation and Survey Level
 (%)

Year	Philippines		Thailand	
	Artificial Intelligence	Simple Model Night Lights	Artificial Intelligence	Simple Model Night Lights
Year 1	13.30	17.86	12.12	14.69
Year 2	12.48	15.16	4.36	3.96
Year 3	22.64	17.12	3.13	5.51

Source: Calculations generated by the study team.

Overall, the artificial intelligence-based approach outperformed the simpler model with intensity of night lights as the sole predictor with few exceptions (i.e., Philippines 2018 and Thailand 2015). In Thailand's case, this is potentially another evidence that the resolution of daytime images affected the performance of the artificial intelligence-based method. Meanwhile, the occurrence where the simpler model performed better is Philippines 2018 which suggested that the artificial intelligence-based approach could be quite sensitive to the resolution of the input data.

¹³ In principle, one can add more covariates other than night light intensity. However, given that the main objective of this study is to produce poverty numbers that are more granular than published ones, the list of potential covariates should only include variables that are available at a granularity finer than the level at which published poverty numbers were available.

F. Harmonizing Artificial Intelligence-Based Predictions with Published Poverty Rates

In the course of the research, the team saw that the image-level poverty predictions generated from employing an artificial intelligence-based method when aggregated at the same level of the published poverty numbers, the data is not necessarily equal to the official poverty figures. From a practical perspective, this could potentially create confusion among users of poverty statistics in terms of which dataset should be followed. To avoid causing confusion, governments can consider calibration methods.

If it is a reasonable assumption to make that the survey-based or small area poverty estimates are statistically reliable at the disaggregation level where these numbers are published, the more granular artificial intelligence-based predictions can be rescaled or calibrated such that, when they are aggregated at the same level where published poverty statistics are available, the two sets of estimates are consistent. Adopting this strategy is appealing for compilers of official statistics as it avoids confusion for users who rely on official estimates but also use alternative data sources.

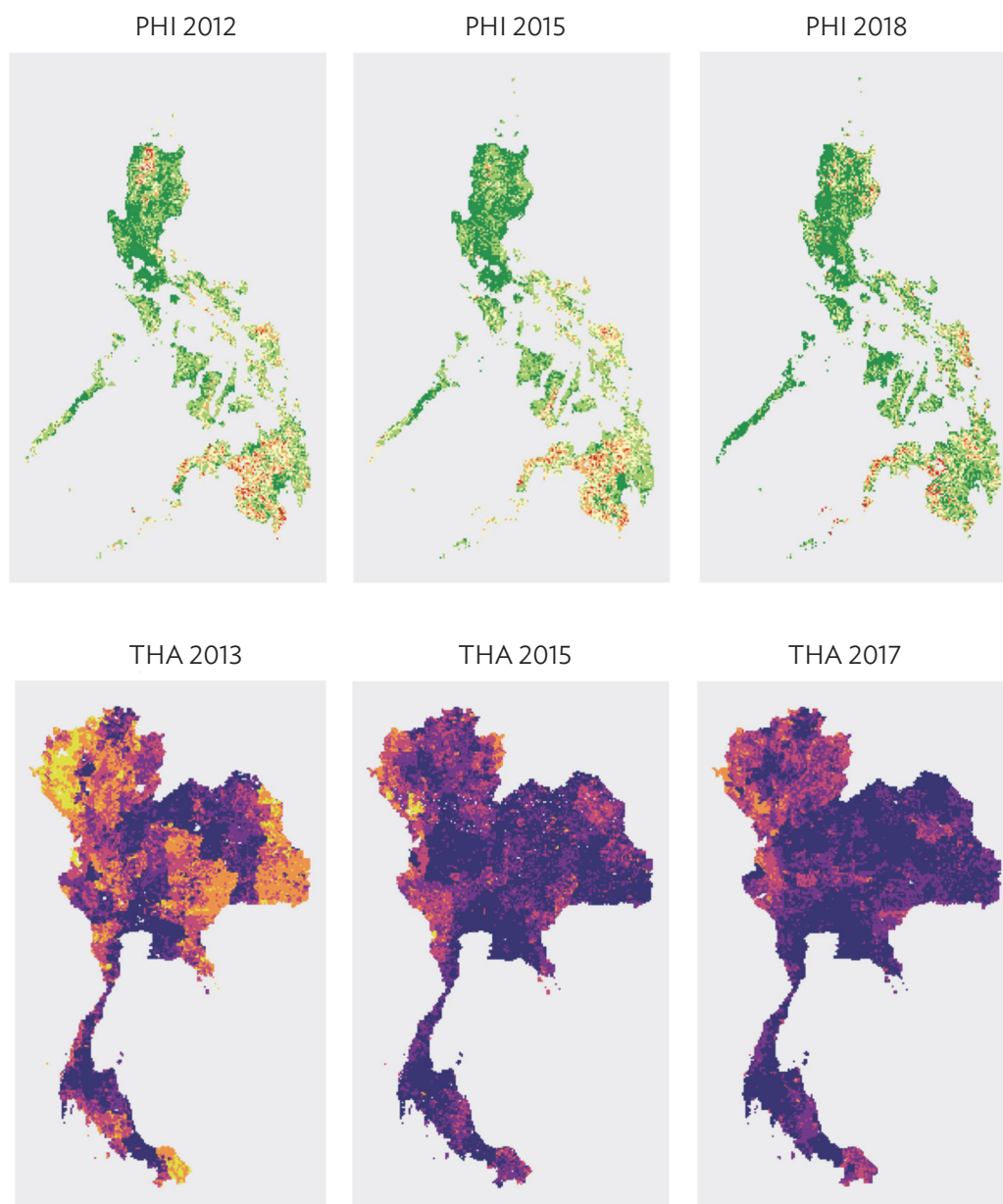
Further, through calibration, the study team was able to address the underestimation of poverty in Thailand which may have been caused by the lack of variability in the input data.

Figure 5 shows the distribution of artificial intelligence-based poverty predictions that have been calibrated for the Philippines and Thailand.

If there would be concerns about the reliability of the survey-based or small area estimates, the uncalibrated artificial intelligence-based predictions may be used as a point of validation.¹⁴ When differences between the uncalibrated predictions and the published estimates are small, there may be cause to have increased confidence in the reliability of the published estimates. However, where there are nontrivial differences between the artificial intelligence-based predictions and the published estimates, further investigation may be merited. In such context, further validation may entail consulting local authorities and key stakeholders who have better appreciation of the socioeconomic realities on the ground.

¹⁴ In general, there are several reasons why there could be concerns about the reliability of an official statistic. In the case of poverty estimation, if the statistics are directly estimated from household surveys, there is a chance that the homeless or those who do not have permanent dwellings are not captured in such surveys. Similarly, if census data are combined with survey data to generate small area poverty estimates, it is possible that the model has suboptimal predictive performance due to lack of covariates that are time invariant.

Figure 5: Calibrated Machine-Learning Poverty Predictions



km = kilometer, PHI = Philippines, THA = Thailand.

Note: The images present the calibrated machine-learning based estimates of poverty rates for every (approximately) 4 km by 4 km grid.

Source: Calculations and graphics generated by the study team.

VI. DISCUSSION AND SUMMARY

Can new sources of data, such as satellite images, which NSOs do not typically use when compiling development statistics, serve as a useful supplementary data source in measuring the Sustainable Development Goals?

This was the question that this study attempted to answer by looking at poverty mapping using satellite imagery as a case study. The team leveraged on recent developments in computer vision techniques and publicly available earth observation data available during the day and at night. With the high costs associated with high resolution imagery, using publicly available satellite images may be attractive to NSOs that are beginning to explore these innovative data sources and methods, and hence, greatly increase the applicability of the approach to other areas where NSOs are also working.

In this study, data from satellite images were integrated with household income as well as expenditure surveys and censuses for two developing countries—the Philippines and Thailand. The results helped produce poverty maps that were more granular than existing maps based on conventional poverty data sources. The methodological approach that was followed was inspired by deep machine learning techniques used by Jean et al. (2016). The results of applying this methodology on specific datasets were encouraging. The use of publicly accessible satellite imagery, whose resolutions were not as refined as proprietary images, still yielded predictions that were generally aligned with the government-published poverty estimates.

Nevertheless, there are aspects of the adopted method which could be further improved. For instance, it can be noted that the resolution of the input imagery has an effect on the quality of the outputs. Higher resolution imagery (i.e., Sentinel's 10-meter resolution versus Landsat's 15-meter resolution) was associated with better predictive performance. Hence, scaling up from exploratory studies to a more rigorous poverty mapping initiative would potentially benefit from high resolution imagery that are commercially available and more sophisticated computing tools.

Another key consideration is the granularity of input data used in training the algorithm. In this study, the team showed that for periods when small area poverty estimates were available, the predictive performance was better than when input data from periods when only household income and expenditure surveys were available was used. This was an important caveat for future research, especially when only household survey data are available as input data for training an algorithm.

Further, validation of the results is warranted particularly for instances when the government published poverty estimates were significantly different from the predictions produced by our adopted method. For this purpose, it would have been ideal to conduct field validation and consult key stakeholders through interviews or workshops. However, this activity was difficult as large gatherings were discouraged at the time of writing of this report, due to the coronavirus disease (COVID-19) pandemic.

To address this issue in this study, the study team employed calibration methods under the assumption that government-published estimates provided an accurate picture of poverty at the level for which statistics are published (i.e., municipal, city, or provincial level in the Philippines and *tambon* level in Thailand). In particular, the resulting machine-learning predictions were rescaled or calibrated such that, when the grid-level predictions were aggregated at the appropriate levels, they aligned more closely with the government-published numbers. The calibration method was designed to preserve the

distributional structure of the grid-level poverty predictions, while paying heed to the estimates published by the government at more aggregated levels. In practice, adopting this type of calibration may be appealing as it avoids confusion for users who traditionally rely on government estimates published by NSOs and other relevant government agencies.

Further, although it was found that the predictive performance of the adopted method was better than a model which predicted poverty rates using only data from the intensity of night lights, combining the extracted image features with other granularly available data could improve on the existing predictions. For example, adding nighttime lights or other remotely sensed information (elevation, precipitation) or other types of readily available socioeconomic data can readily be tested since the image features can be thought of as latent variables and can be added to analyses related to wealth and poverty.

Lastly, this study focused on examining the feasibility of data integration—which entailed the combination of satellite imagery with surveys and census—and computer vision techniques to enhance the granularity of poverty estimates compiled from conventional data sources. Instead of enhancing granularity, other studies focused on predicting poverty or wealth for areas or years where conventional data for poverty estimation were not available. To test whether the methodological principles discussed in this paper still apply for such research objective, additional investigation may be needed. For instance, it is important to examine whether it is safe to assume that the relationships observed for the years wherein data are available are stable over time or not.

REFERENCES

- Addison, Douglas and Benjamin Stewart. 2015. "Nighttime Lights Revisited: The Use of Nighttime Lights Data as a Proxy for Economic Variables." World Bank Policy Research Working Paper No. 7496.
- Alkire, Sabina, Fanni Kovesdi, Corinne Mitchell, Monica Pinilla-Roncancio, and Sophie Scharlin-Pettee. 2019. "Changes over Time in the Global Multidimensional Poverty Index: A Ten-Country Study." OPHI MPI Methodological Note 48. Oxford Poverty and Human Development Initiative, University of Oxford.
- Asian Development Bank (ADB). 2020. *Mapping Poverty through Data Integration and Artificial Intelligence—A Special Supplement of the Key Indicators for Asia and the Pacific 2020*. Manila.
- Babenko, Boris, Jonathan Hersh, David Newhouse, Anusha Ramakrishnan, and Tom Swartz. 2017. "Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, with an Application in Mexico." Proceedings from NIPS 2017: Neural Information Processing Systems Workshop on Machine Learning for the Developing World. Long Beach, CA.
- Chakure, Afroz. 2019. "Random Forest Regression: Along with Its Implementation in Python." <https://medium.com/@aaaanchakure/random-forest-and-its-implementation-71824ced454f>.
- Chen, Xi and William D. Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." *Proceedings of the National Academy of Sciences* 108 (21): 8589–99. doi: 10.1073/pnas.1017031108.
- Deng, Jia, Wei Dong, Richard Socher, Li-jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 2009, 248–55. doi: 10.1109/CVPR.2009.5206848.
- Elvidge, Christopher, Kimberly Baugh, Mikhail Zhizhin, and Feng-Chi Hsu. 2013. "Why VIIRS data are superior to DMSP for mapping nighttime lights." *Proceedings of the Asia-Pacific Advanced Network* 35: 62–69. <http://dx.doi.org/10.7125/APAN.35.7>.
- Engstrom, Ryan, Jonathan Hersh, and David Newhouse. 2017. "Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being." World Bank Policy Research Working Paper No. 8284.
- ESA Earth Online. <https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/sentinel-2>.
- Head, Andrew, Nhat Tran, Melanie Manguin, and Joshua Blumenstock. 2017. "Can Human Development be Measured with Satellite Imagery?" Association for Computing Machinery, 1–11. 10.1145/3136560.3136576.
- Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.
- Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science* 353 (6301): 790–94.

Keola, Souknilanh, Magnus Andersson, and Ola Hall. 2015. "Monitoring Economic Development from Outer Space: Using Nighttime Light and Land Cover Data to Measure Economic Growth." *World Development* 66: 322–34.

Mwiti, Derrick. 2020. "Random Forest Regression: When Does It Fail and Why?" <https://neptune.ai/blog/random-forest-regression-when-does-it-fail-and-why>.

Pan, Sinno Jialin and Qiang Yang. 2010. "A Survey on Transfer Learning." *IEEE Transactions on Knowledge and Data Engineering* 22 (10): 1345–59. doi:10.1109/TKDE.2009.191.

Partnership in Statistics for Development in the 21st Century (PARIS21). 2017. *Partner Report on Support to Statistics Press 2017*. Paris.

Perez, Luis and Jason Wang. 2017. "The Effectiveness of Data Augmentation in Image Classification Using Deep Learning." *arXiv preprint arXiv:1712.04621*.

Puttanapong, Nattapong, Arturo Martinez Jr., Joseph Albert Nino Bulan, Ron Lester Durante, Mildred Addawe, and Marymell Martillan. 2020. "Predicting Poverty Using Geospatial Data in Thailand. ADB Economics Working Paper Series No. 630.

Sohnesen, Thomas, Peter Fisker, and David Malmgren-Hansen. 2019. "Using Satellite Data to Guide Urban Poverty Reduction. Proceedings of Special IARIW-World Bank Conference." Washington, DC. <http://iariw.org/washington/Sohnesen-paper.pdf>.

Stevens, Forrest, Andrea Gaughan, Catherine Linary, and Andrew Tatem. 2015. "Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data." *PLoS One* 10, e0107042.

Tingzon, Isabelle, Ardie Orden, Stephanie Sy, Vedran Sekara, Ingmar Weber, Masoomali Fatehikia, Manuel Garcia-Herranz, and Dohyung Kim. 2019. "Mapping Poverty in the Philippines Using Satellite Imagery, and Crowd-Sourced Geospatial Information." *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W19.

Xie, Michael, Neal Jean, Marshall Burke, David Lobell, and Stephano Ermon. 2015. "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping." *CoRR* abs/1510.00098.

Xu, Yun, and Royston Goodacre. 2018. "On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning." *Journal of Analysis and Testing* 2: 249–62.

Yeh, Christopher, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. 2020. "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa." *Nature Communications* 11: 2583. <https://doi.org/10.1038/s41467-020-16185-w>.

Applying Artificial Intelligence on Satellite Imagery to Compile Granular Poverty Statistics

This study outlines a computational framework to enhance the spatial granularity of government-published poverty estimates using computer vision techniques applied on publicly available medium-resolution satellite imagery, household surveys, and census data from the Philippines and Thailand. The results suggest that even using publicly accessible satellite imagery, in which the resolutions are not as fine as those in commercially sourced images, predictions generally aligned with the distributional structure of government-published poverty estimates, after calibration. The study further examines the robustness of the resulting estimates to user-specified algorithmic parameters and model specifications.

About the Asian Development Bank

ADB is committed to achieving a prosperous, inclusive, resilient, and sustainable Asia and the Pacific, while sustaining its efforts to eradicate extreme poverty. Established in 1966, it is owned by 68 members—49 from the region. Its main instruments for helping its developing member countries are policy dialogue, loans, equity investments, guarantees, grants, and technical assistance.



ASIAN DEVELOPMENT BANK

6 ADB Avenue, Mandaluyong City

1550 Metro Manila, Philippines

www.adb.org