Article
Predicting Poverty Using Geospatial Data in Thailand

Nattapong Puttanapong [1,*], Arturo Martinez, Jr. [2], Joseph Albert Nino Bulan [2], Mildred Addawe [2], Ron Lester Durante [2] and Marymell Martillan [2]

1   Faculty of Economics, Thammasat University, Bangkok 10200, Thailand
2   Asian Development Bank (ADB), Mandaluyong City 1550, Metro Manila, Philippines; amartinezjr@adb.org (A.M.J.); jbulan@adb.org (J.A.N.B.); maddawe.consultant@adb.org (M.A.); rdurante1.consultant@adb.org (R.L.D.); mmartillan.consultant@adb.org (M.M.)
*   Correspondence: nattapong@econ.tu.ac.th

Abstract: Poverty statistics are conventionally compiled using data from socioeconomic surveys. This study examines an alternative approach to estimating poverty by investigating whether readily available geospatial data can accurately predict the spatial distribution of poverty in Thailand. In particular, the geospatial data examined in this study include the intensity of night-time light (NTL), land cover, vegetation index, land surface temperature, built-up areas, and points of interest. The study also compares the predictive performance of various econometric and machine-learning methods such as generalized least squares, neural network, random forest, and support-vector regression. Results suggest that the intensity of NTL and other variables that approximate population density are highly associated with the proportion of an area's population that are living in poverty. The random forest technique yielded the highest level of prediction accuracy among the methods considered in this study, primarily due to its capability to fit complex association structures even with small-to-medium-sized datasets. This obtained result suggests the potential applications of using publicly accessible geospatial data and machine-learning methods for timely monitoring of the poverty distribution. Moving forward, additional studies are needed to improve the predictive power and investigate the temporal stability of the relationships observed.

Keywords: poverty; Thailand; geospatial; machine learning

1. Introduction

Over the past three decades, the real GDP per capita of Thailand has grown by 3.35% annually. Prior to the COVID-19 pandemic, the country's impressive economic growth was accompanied by declining household poverty rates, dropping from 61.41% in 1988 to 5.04% in 2019 (Figure 1). However, significant pockets of poverty still exist, particularly in rural areas where about 6.76% of households are considered poor [1]. Furthermore, the pandemic brought about by COVID-19 may undermine poverty reduction gains over the years.

Hence, poverty monitoring remains an essential task for the country's development practitioners. At present, the National Economic and Social Development Council (NESDC) and National Statistics Office (NSO) are responsible for compiling poverty statistics in Thailand. Poverty statistics are based on the Household Socioeconomic Survey (HSES), which collects household income data every two years. The survey's sample size provides estimates that fall within tolerable levels of reliability when presented at the national and provincial levels, but are typically not large enough to provide reliable estimates at more granular levels. On the other hand, increasing the frequency and survey sample sizes is often not practical due to the high cost [2].

Given the need for more timely and granular poverty data that can be used to target population segments that have the greatest need for intervention, researchers and development practitioners have explored alternative methodological approaches. For instance, small area estimation (SAE) techniques that combine surveys with census and other types

of administrative data have been widely used to facilitate estimation at levels more granular than can be afforded when working with surveys alone. However, since SAE requires census data that are not frequently available, the obtained poverty maps are not necessarily timely. There are also initiatives to use innovative data from call detail records, social media, digital transactions, and remotely sensed data to compile more granular and timely poverty statistics [3–9].
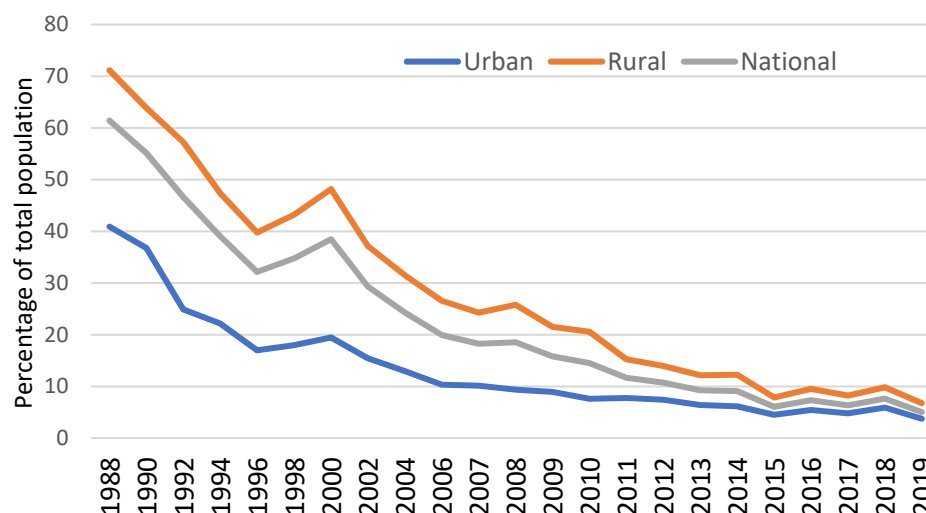

Figure 1. Proportion of Thailand's population living below its national poverty line. Source: Thailand's NSO. Note: Thailand's poverty line is calculated based on the minimum standard required by an individual to fulfill one's basic food and non-food commodities. Details are provided in the NESDC report (2015).

Mapping Thailand's spatial distribution of poverty is an area that could greatly benefit from the integration of remote sensing data. In this context, two types of analytical frameworks are worth pointing out. First, by capitalizing on ongoing developments in computer vision techniques and satellite imagery, several researchers have shown that it is feasible to develop an algorithm that can automatically predict survey-based estimates of poverty with satisfactory levels of accuracy [2–5]. Such approach is quite attractive for instances in which collecting survey data, particularly in remote and/or hard-to-reach areas, is onerous, and no other types of supplementary data are readily available. It is also useful when increasing the survey's sample size is prohibitively costly. However, since the features extracted by computer vision techniques are relatively abstract [2], it is difficult to manually pinpoint exactly which features are being picked up by the computer when predicting poverty. Consequently, it is also difficult to validate what could have triggered an unexpectedly low or high estimate of poverty if such instances arise. Wider adoption of these new poverty compilation techniques may also be hampered if they do not generate features that are interpretable to policymakers [4].

Alternatively, if structured geospatial data are readily available, one can develop a more tractable and interpretable econometric model for predicting poverty. This can be achieved by leveraging interpretable geospatial data that have already been pre-compiled or are passively collected. This approach facilitates a more interpretable computational framework for predicting poverty. This study explores the second approach, where poverty is predicted by identifying correlates from pre-compiled geospatial data. It contributes to the existing literature by assessing whether it is feasible to develop a model with satisfactory predictive performance, even if we solely depend on pre-compiled geospatial datasets, which theoretically can be considered as just a fraction of the number of covariates that the first approach can potentially generate, a feat that has not been explored thoroughly in the context of Thailand, in previous studies. Here, we also compare the performance of different machine-learning techniques, a topic that has not been well explored in previous

studies of poverty estimation using non-traditional data sources. In doing so, the study aims to contribute to the literature that explores other cost-effective methods of predicting poverty using an interpretable computational framework applied to geospatial data.

The rest of this paper is structured as follows. The second section reviews related literature, while the third and fourth parts introduce the data and research methodologies, respectively. The fifth section presents the key findings of the econometric and machine-learning methods adopted in this study. The last section summarizes lessons learned and draws brief recommendations for future studies.

2. Literature Review

2.1. Using Pre-Compiled Geospatial Data to Predict Socioeconomic Indicators

The existing literature offers a wide range of case studies showcasing various applications of satellite imagery and geospatial data for development-related analyses. For instance, data on the intensity of NTL compiled through Defense Meteorological Satellite Program (DMSP)/Operational Linescan System (OLS) and Suomi National Polar-orbiting Partnership (SNPP)—Visible and Infrared Imaging/Radiometer Suite (VIIRS) are widely used. Many studies found a statistically significant relationship between the intensity of NTL and various ground data such as GDP, electricity consumption, inequality, and infant mortality rate [6–14].

In addition to NTL intensity, Landsat, National Oceanic and Atmospheric Administration (NOAA)—Polar Orbiting Environmental Satellites (POES) and Terra Moderate Resolution Imaging Spectroradiometer (MODIS) satellites have scanned the Earth's surface with multi-spectral sensors. These multi-spectral data have been used by various researchers to compile a number of geospatial indicators such as the building density, water coverage, Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), NDWI (Normalized Difference Water Index), NDSI (Normalized Difference Snow Index) NDSI (Normalized Difference Soil Index), and NDBI (Normalized Difference Built-up Index). Specifically, NDVI represents the spatiotemporal pattern of forest and cultivated areas and is considered one of the conventional indices commonly used in remote-sensing analysis of vegetation. NDVI is calculated by measuring the difference between near-infrared (which vegetation reflects) and red light (which vegetation absorbs). For applications in socioeconomic studies, a correlation between urban expansion and decreasing NDVI has been documented [15–17]. Similarly, the statistical relationship between NDVI and the spatial distribution of income inequality has been statistically verified [18–21].

Data on land surface temperature is another type of pre-compiled geospatial information which researchers are using to predict income. For instance, a statistically significant relationship between land surface temperature and income has been statistically validated [22–28]. Additionally, many studies found a statistically significant correlation between rainfall on income, human capital, and economic activity in developing countries [29–34].

In addition, the forecast models using both temperature and NDVI for predicting drought and, in turn, forecasting the loss of agricultural output and its effect on farmers' incomes have been formulated [35,36].

Efforts to crowdsource geospatial data are also expanding. A good example is OpenStreetMap, a collaborative project producing a crowdsourced geographic database, and one of the major platforms promoting the use of geospatial data in the fields of global humanitarian action and community development. The OpenStreetMap database also features other types of geospatial data such as the presence of roads, rivers, built-up areas, and points of interest (POI), enabling investigation of the association between the geographical characteristics and socioeconomic conditions. Studies such as those by Hu et al. (2016), Ye et al. (2019), and Deng et al. (2019) [37–39] demonstrate that OpenStreetMap can provide details of the spatial distribution of population and economic activities.

2.2. Poverty Mapping in Thailand

For many countries, poverty was a critical development challenge even before the COVID-19 pandemic struck, with pre-pandemic trends in poverty reduction showing a relatively slower decline compared to what has been observed in the past. In Developing Asia, for instance, about 203 million people were living below USD 1.9 a day as of 2017, and there is evidence suggesting that the pandemic might have further turned back the region's poverty clock.

Spatial disparities in poverty have been well documented in a number of empirical studies. In general, geography can act as either a gateway to better living standards, especially when a specific location has greater access to richer natural resources, or to poverty when an area is too remote, has limited job creating-economic activities, and has limited access to various social services. On the other hand, severe climatic events such as rainfall shocks and even modest changes in temperature may make it difficult for poor and vulnerable people who have limited access to social safety-nets to escape poverty as their ability to accumulate assets and invest in human capital is hampered.

As an upper-middle-income country, Thailand is considered as one of Asia's great development success stories. In less than a generation, it was able to move away from being a low-income country. However, its development trajectory is constrained by spatial income disparities, among other development challenges. The concentration of poverty in rural areas is possibly driven by Bangkok's high agglomeration force and the fact that most economic activities are concentrated in Bangkok and its suburbs [40]. Since rural provinces have a limited variety of economic activities, they have a constraint of creating non-agriculture jobs. Trends in non-pecuniary indicators of development are also concerning (Figure 2). For instance, half of the country's working population is still in precarious employment. There is also ample room for improvement in the education sector as rural migrants and the urban poor generally lack the skills demanded by modern jobs.
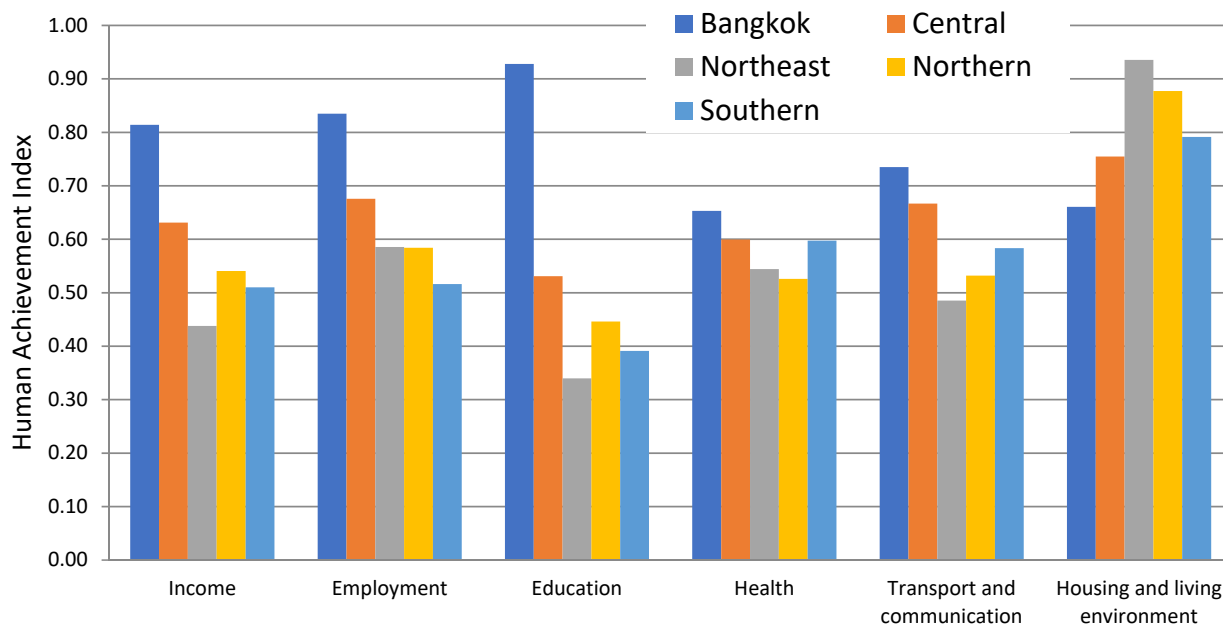
Figure 2. The Human Achievement Index in 2017. Source: NESDC (2017). Note: The Human Achievement Index value ranges from 0 (worst outcome) to 1 (best outcome).

As mentioned earlier, official poverty statistics in Thailand are based on the Household Socioeconomic Survey, which provides reliable estimates from national down to provincial levels. However, recognizing the importance of having more geographically disaggregated poverty data as inputs for policy targeting, NSO of Thailand started compiling small area (Tambon or subdistrict-level) poverty estimates in 2003 in collaboration with other

development partners such as NESDC, Thailand Development Research Institute (TDRI), and the World Bank. Since then, small area poverty estimates in the country have been compiled for the following years: 2005, 2007, 2008, 2011, 2012, 2015, and 2017. The outputs in 2003 and 2005 were jointly prepared by three local institutions, namely, NESDC, NSO, and TDRI, together with the technical advisory from the World Bank. In 2015, the World Bank provided further technical assistance to NSO to build its capacity to implement small area estimation among more NSO staff. Additional technical details on the process of compiling poverty maps are documented by Jitsuchon (2004) and Jitsuchon and Richter (2007) [41,42].

However, despite the availability of analytical tools for compiling granular estimates of poverty, it is important to identify alternative methods due to limitations associated with the conventional poverty mapping technique, which heavily relies on the availability of census data. For instance, since censuses are usually conducted every five to ten years only, poverty-mapping models that use covariates derived from census have restrictively strong assumptions [43].

Poverty statistics compilation presents exciting opportunities to blend traditional and innovative data sources, particularly information extracted from satellite imagery [3]. As discussed earlier, conducting detailed household surveys with a sample size large enough to reflect all geographic areas and different population groupings may not be a practical option due to the high cost. Moreover, the importance of poverty statistics for policy targeting requires that granular data are available regularly. Incorporating innovative data can potentially address the restrictions that conventional data sources are associated with.

In a study published recently, researchers from ADB extended the conventional small-area poverty-estimation framework by tapping geospatial data extracted from daytime imageries and NTL through machine-learning algorithms to create granular poverty maps of the Philippines and Thailand [2,4]. The adopted method was inspired specifically by Jean et al. (2016) [3], which was further used and/or enhanced in subsequent studies [44–47]. These studies fall under the strand of literature that broadly aims to explore applications of artificial intelligence and computer vision techniques for estimating poverty. However, as hinted earlier, this methodology has several technical issues. First, validating aberrant or unexpected predictions becomes challenging because the features being used to correlate poverty are abstract. Second, instead of directly predicting poverty, the method employs an intermediate step wherein an algorithm is first trained to predict the intensity of NTL. The intermediate step is necessary in this context because sources of NTL data, particularly satellite imagery, are readily accessible and can cost-effectively provide large volumes of labeled images on which to train a computer vision algorithm, something that cannot be easily achieved if we were to predict poverty outright, since readily available poverty data are not quite granular. Using data on NTL intensity as a proxy for poverty during the intermediate step is arguably valid if it is assumed that places that are brighter at night are less poor than those places that are less well-lit. However, if there are places that are equally lit but show varying levels of poverty on the ground, such an intermediate step could potentially lead to loss of vital information by not predicting poverty outright. Third, having abstract satellite image features as model covariates which cannot be intuitively understood by development practitioners and policymakers makes adoption of such techniques less appealing, as it is not straightforward to draw insights why a given location is associated with a specific level of poverty [4].

This study contributes to the existing literature of poverty measurement in Thailand by developing a predictive model whose correlates were derived from pre-compiled geospatial data, which are interpretable. By doing so, we aim to assess whether it is feasible to develop a model with satisfactory predictive performance even if we are solely depending on pre-compiled geospatial dataset(s) instead of applying computer vision techniques to automatically extract satellite image features that are potentially correlated with poverty, a feat that has not been thoroughly explored in previous studies.

The outcomes of interest in this study are income and multidimensional poverty indices. Our model specifications include several covariates. First, we consider intensity

of NTL, which serves as proxy measure of level of economic activity in a given area. To capture the level of urbanization, we also consider land surface temperature, land use, and vegetation index. Measures of density of points of interest are used to capture the accessibility of services as well as the level of economic activity. Rainfall data capture climatic factors which may amplify poverty risk in a given location. Details are provided in the next section.

3. Data

3.1. Satellite Data

Data Obtained from Google Earth Engine

Google Earth Engine is an open cloud-based data storage and computing platform which provides access to satellite imagery for free. In this study, we extracted the following information from Google Earth Engine:

1. Intensity of night-time lights (NTL)
2. Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS') rainfall data
3. Land Surface Temperature (LST)
4. Normalized Difference Vegetation Index (NDVI)

Table 1 summarizes the range of data that can be obtained from Google Earth Engine.

Table 1. Technical specification of data obtained from Google Earth Engine.

Data	Satellite	Data Full Name	Resolution	Area (Approximate)	Frequency
Rainfall		CHIRPS Pentad: Climate Hazards Group InfraRed Precipitation with Station Data (version 2.0 final)	0.05 arc degree	110 m^2/Pixel	Monthly
NTL (Old)	DMSP/OLS	Nighttime Lights Time Series Version 4, Defense Meteorological Program Operational Linescan System	30 arc s	1 km^2/Pixel	Monthly
NTL (New)	VIIRS/DNB	VIIRS Nighttime Day/Night Band Composites Version 1	15 arc s	375 m^2/Pixel	Monthly
Land Surface Temperature (day)	MODIS	MOD11A1.006 Terra Land Surface Temperature and Emissivity Daily Global 1 km	30 arc s	1 km^2/Pixel	Daily
Land Surface Temperature (night)	MODIS	MOD11A2.006 Terra Land Surface Temperature and Emissivity Daily Global 1 km	30 arc s	1 km^2/Pixel	Daily
Normalized Difference Vegetation (NDVI)	MODIS	MODIS Combined 16-Day NDVI	15 arc s	375 m^2/Pixel	18 Day

Source: Google Earth Engine. Global Urban Footprint (GUF).

The Global Urban Footprint (GUF) project by the German Remote Sensing Data Center (DFD) of the German Aerospace Center (DLR) compiles geocoded data which identify urban areas, land surface, and water bodies. Geocoded data on built-up and non-built-up areas are also available from the GUF.

Global Human Settlement Layer (GHSL)

Mainly supported and supervised by the Directorate General Joint Research Centre of the European Commission, the Global Human Settlement Layer project has produced a fully open and free geospatial spatial dataset. The generated geospatial database provides informative evidence and the broadened insight of global human presence.

USGS

This geospatial dataset has been generated based on the ten years (2001–2010) collection of MODIS-based Global Land Cover maps (MCD12Q1 land cover type data). There are 16 classifications for each pixel, identifying the type of land cover based on the method of highest confidence during 2001–2010 [48].

European Space Agency Land Cover (ESA-LC)

Initially, the main objective of the European Space Agency (ESA)'s Climate Change Initiative is to produce an accurate land cover classification that can serve the climate modeling community. This project has developed the Essential Climate Variable (ECV) spatial dataset based on the extensive archives of remote-sensing data. The database covers time series from 1992 to 2017 and contains 38 land cover classes, which are based on the UN Land Cover Classification System.

3.2. Crowdsourced Geospatial Database (OpenStreetMap)

OpenStreetMap features crowdsourced data on locations of infrastructures, human settlements and economic activities. In this study, we extracted the following information from OpenStreetMap: road count, road length, Point of Interest (POI), and built-up area. We categorized POIs into 16 types based on its economic activity, matched to the official classifications of 16 production and service sectors published by NESDC.

Figures 3–6 exemplify the spatial distributions of NTL, NDVI, Land Surface Temperature (Day), and rainfall obtained from Google Earth Engine. Figure 7 exhibits the transportation network and the distribution of POI derived from OpenStreetMap. For further details, Tables A1 and A2 of the Appendix A provide the list of all variables obtained from geospatial data of 2015 and 2017, respectively.
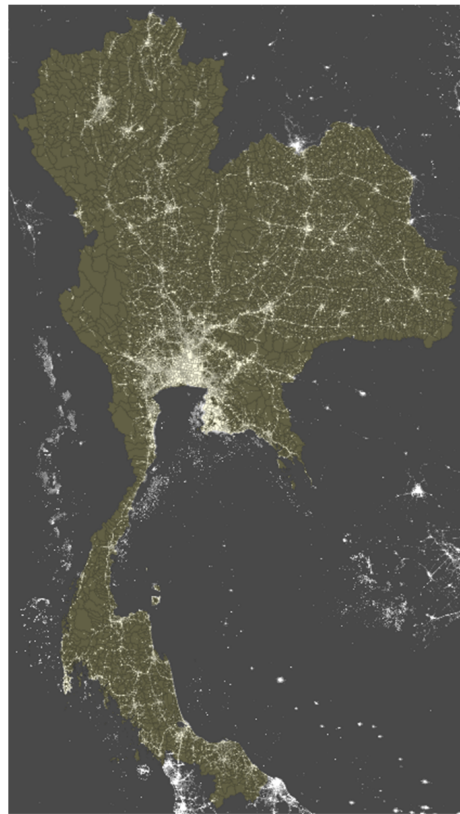
Figure 3. Spatial distribution of NTL in 2017. Source: Google Earth Engine.
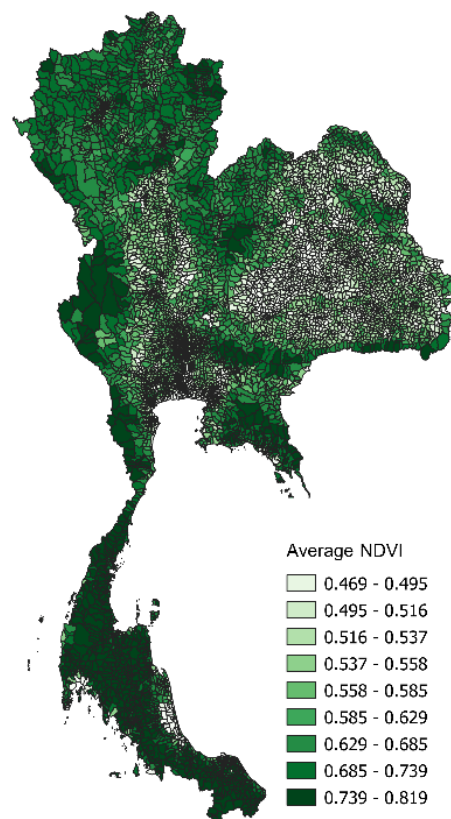

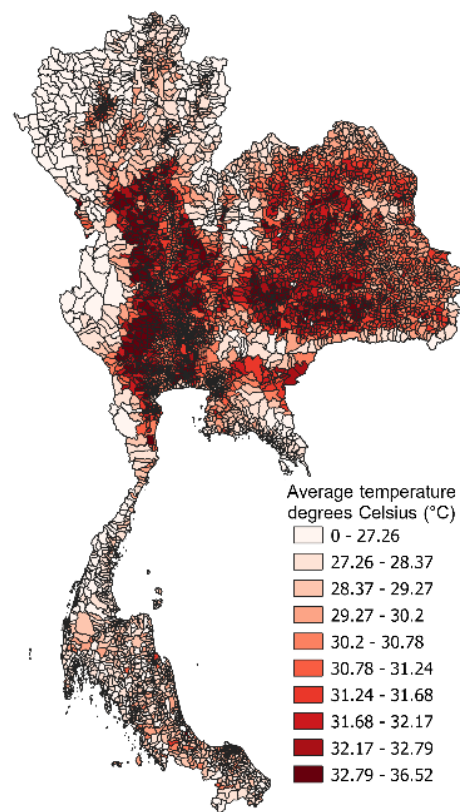**Figure 4.** Spatial distribution of NDVI in 2017. Source: Google Earth Engine.

Figure 5. Spatial distribution of Land Surface Temperature (Day) in 2017. Source: Google Earth Engine.
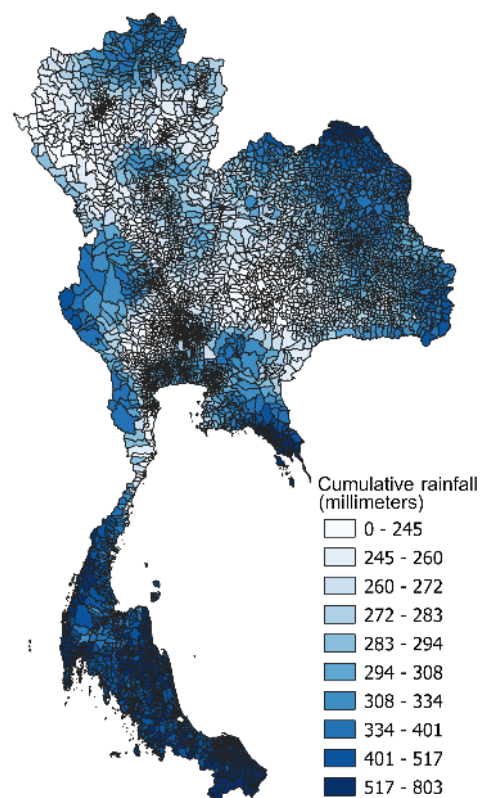

Figure 6. Spatial distribution of rainfall in 2017. Source: Google Earth Engine.
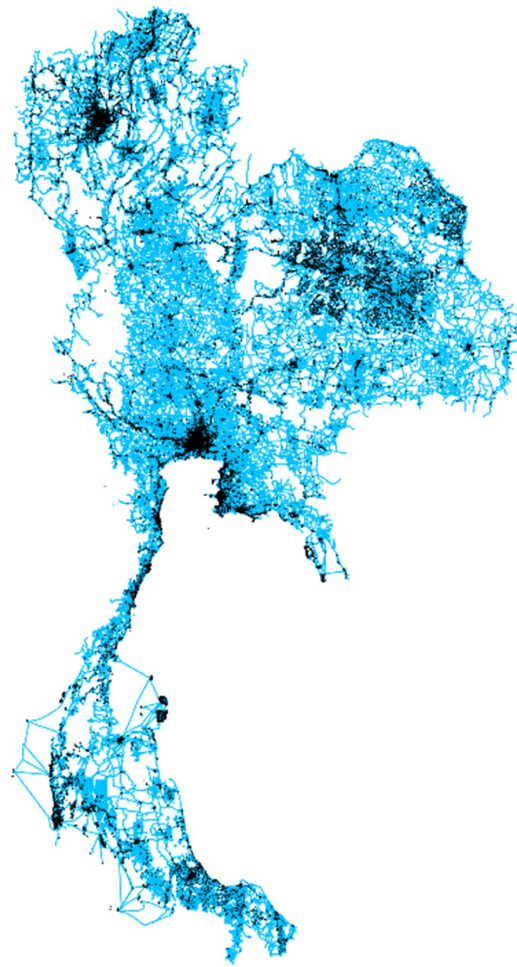
Figure 7. Spatial distribution of transportation network and POI in 2017. Source: OpenStreetMap.

3.3. Poverty Data

3.3.1. Income-Based Poverty

As mentioned earlier, poverty mapping is a regular initiative conducted by the Thai government. In this study, the ratio of the population living below the national poverty line per total population in each Tambon (i.e., subdistrict) is used as one of the dependent variables in our computations.

3.3.2. Multidimensional Poverty Index

As an alternative metric of poverty, NESDC and National Electronics and Computer Technology Center (NECTEC) also compile statistics on prevalence of multidimensional poverty index (MPI) starting 2017. Dimensions included in the calculation of MPI include education, healthy living, living conditions, and financial security (Table 2).

The data are based on:

(1) Census-based Basic Minimum Need (BMN) data, supervised by the Community Development Department, Ministry of Interior, which includes a population of approximately 36 million.

(2) A register-based data source of approximately 11.4 million individuals gathered by the Ministry of Finance through the national welfare card program.

Table 2. List of indicators included in MPI calculation.

Dimensions	Indicators	Deprivation Cut-Off
Education	Year of education	A household is deprived if at least one member of the household (1) aged 15–29 has not attained grade 9-level education or (2) aged 30–50 years and has not attained grade 6-level education.
	Late attendance	Households with at least one child aged 6–17 years who does not go to school or is up to 2 years behind the grade they should be for their age, unless graduated from grade 9.
	Living with parents	Households with at least one child aged 0–6 years who does not live with their father and/or mother. (In cases where the father and/or mother are still alive.)
Healthy living	Drinking water	Households drink water from (1) indoor wells or (2) outdoor wells or (3) rivers/streams/canals/waterfalls/mountains or (4) rainwater or (5) other sources.
	Taking care of yourself	At least one household member aged more than 15 years is unable to take care of himself in daily life without help and is unable to travel outside the residential area without a guardian.
	Food poverty	Households' food expenditures are below the food poverty line, which is calculated from the minimum nutrient (calorie) needs that people of each age and gender require per day.
Living conditions	Garbage disposal	Households dispose of waste by (1) burning, or (2) landfills, or (3) dumping into a river, canal, or (4) dumping in a public space, or (5) other.
	Internet access	No household members use the Internet at all.
	Asset owner	Household does not own at least four small objects (radio, TV, air conditioner, bicycle, phone, and refrigerator) and one large object (car and boat).
Financial security	Savings	Households do not have financial assets to save.
	Financial burden	In the past 12 months, households have difficulty paying home rent, water, electricity, or tuition.
	Pensions	At least one household member aged 60 and over has no pension and allowances.

Source: NESDC.

The criteria used in identifying a multidimensionally poor person is inspired by the Multidimensional Poverty Index method developed by the Oxford Poverty and Human Development Initiative and United Nations Development Program.

Figure 8 exemplifies the spatial distribution of poverty headcount in 2017, obtained from Thailand's NSO. Similarly, Figure 9 illustrates the distribution of MPI in 2017 derived from TPMAP data.

3.4. Reference Period

Our target reference period coincides with two most recent years where Tambon-level estimates of poverty in Thailand are available: 2015 and 2017 for income poverty, and 2017 for the multidimensional poverty index.
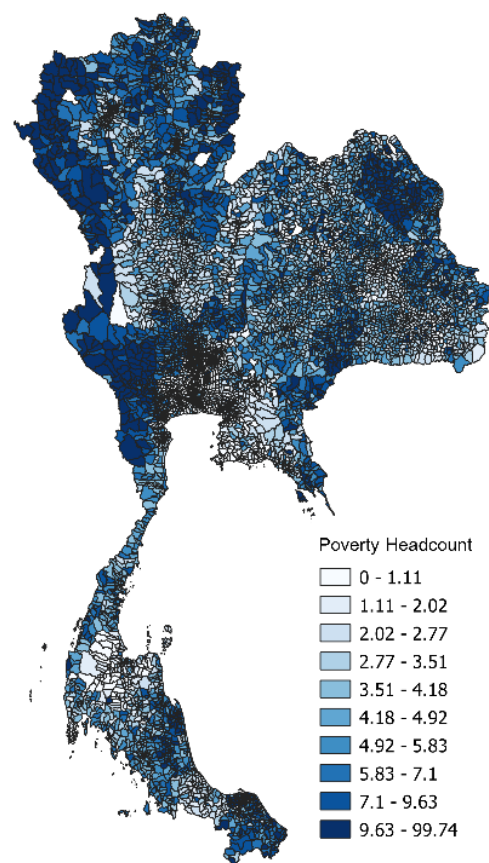
Figure 8. Spatial distribution of poverty headcount in 2017. Source: Thailand's NSO.
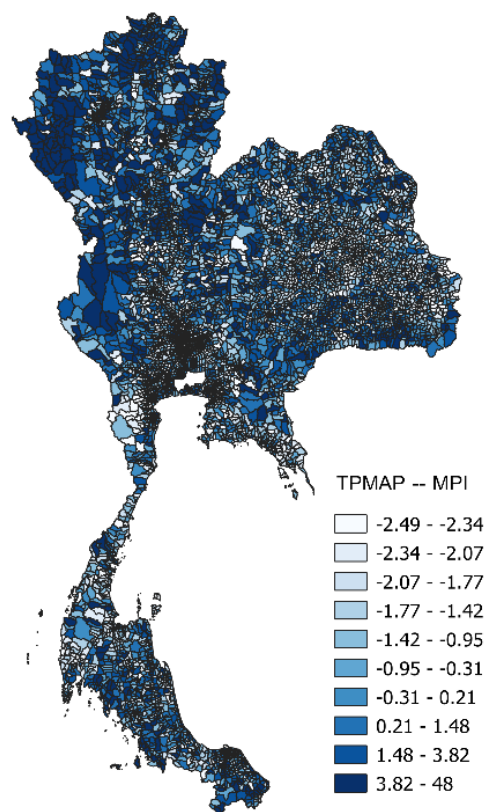

Figure 9. Spatial distribution of MPI (from TPMAP data) in 2017. Source: NESDC.

4. Methods

Figure 10 summarized the analytical framework applied in this study. Data pre-processing was undertaken, including the transformation of spatial resolution, variable normalization [49], and data integration. Then, we applied four computational methods, Generalized Least Squares (GLS) method, and three other widely used machine-learning algorithms: neural network (NN), random forest (RF) estimation, and support-vector regression (SVR). Following the technical approaches suggested by McBride and Nichols (2018) and Hu et al. (2022) [50,51], 50% of the data were allocated for training, while the remaining 50% constituted the validation set. Based on this allocation, we resampled the data 100 times. The values of metrics used to compare machine-learning algorithms are based on averages from these 100 datasets.
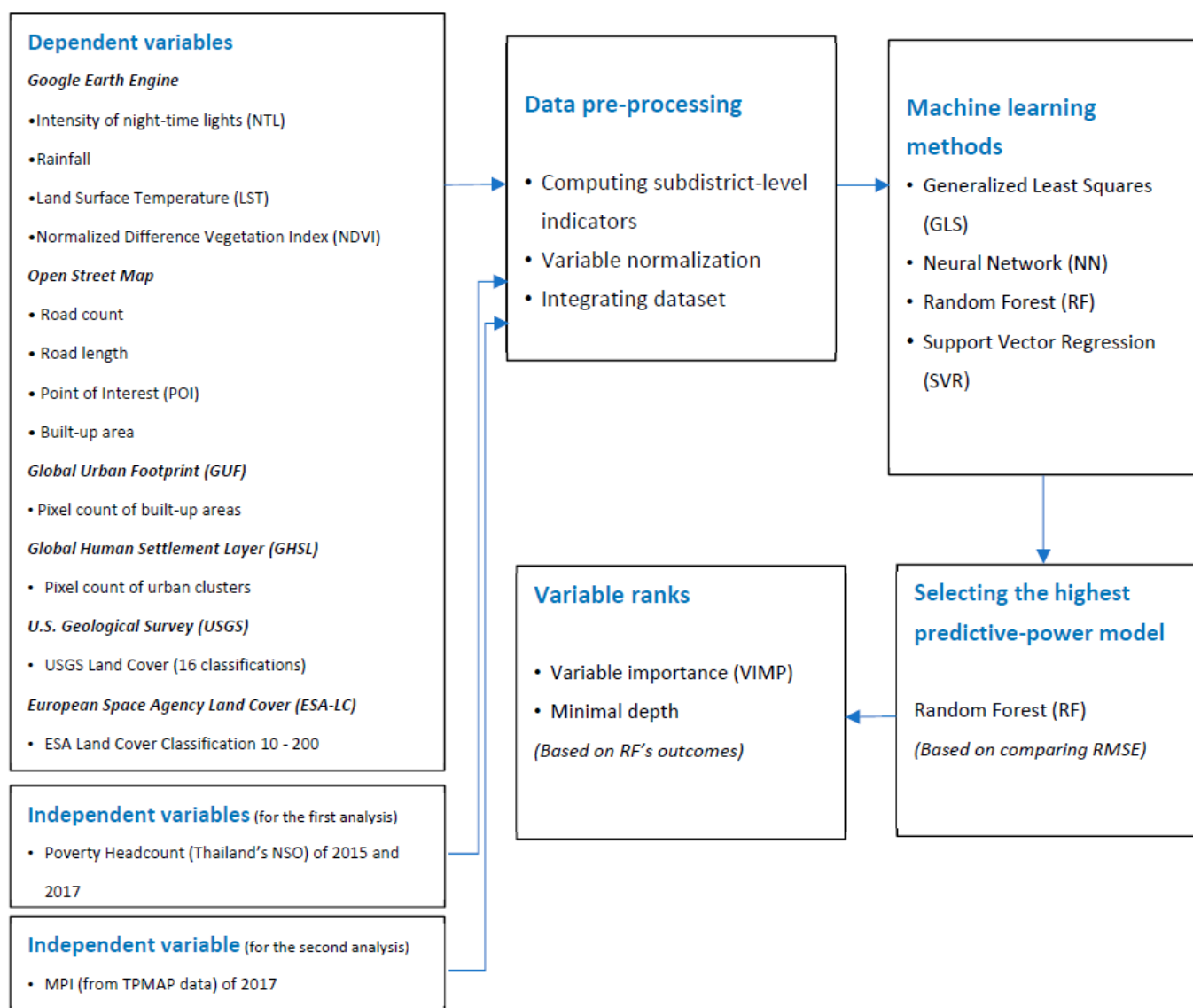
Figure 10. Analytical framework of this study. Source: Graphics generated by authors.

4.1. Generalized Least Squares

GLS is considered a modification of the Ordinary Least Squares (OLS) as it relaxes the assumption that the variance of observation is homogeneous regardless of the explanatory variables associated with it. Mathematically speaking, the issue of inconsistent variance of residuals is corrected by imposing the weight matrix derived from Cholesky decomposition. In particular, applying the weight matrix throughout the regression equation yields the

modified variance-covariance matrix with independent and identically distributed property (i.i.d.), subsequently leading to unbiased, consistent, and efficient regression coefficients. The result obtained from GLS is the benchmark for comparing the predictive power between the conventional statistical method and those of other machine-learning algorithms.

4.2. Neural Network

A neural network (NN) is an example of a machine-learning model inspired by the biological neural network that constitutes the human brain. As with other types of machine-learning models, a neural network can learn to perform different tasks without being explicitly programmed to do so.

Structurally, a neural network is composed of numerous nodes and edges. A node can be a variable or a mathematical function connected by edges. These nodes combine together to form different layers within the neural network. The input layer takes in the raw data. In the hidden layers, each node or neuron serves as a filter and is activated each time it detects a specific pattern or feature. The output layer simply organizes the identified features into an appropriate category. As alternatively described and compared by Anesti et al. (2021) [52], the conventional OLS method is a special case of a neural network comprised of only the single input and output layers, where input nodes are regressors, and the output layer generates the predicted value as the weighted sum of regressors. Thus, the neural network is considered the extended structure of OLS, incorporating a multi-layer process of weighted sums.

As introduced by Ciaburro and Venkateswaran (2017) [53], this study used the R package of *nnet* as the main tool for performing prediction applying the neural network algorithm. All parameters followed the default of *nnet* package [54].

4.3. Random Forest

Originally, Breiman (2001) [55] introduced the prediction method using a set of "de-correlated" decision trees. Hastie et al. (2009) [56] subsequently suggested the algorithm to formulate a large number of decision trees. This development constituted the random forest (RF) method, an ensemble-tree-based technique, with each tree building on a random subset of the training data and a random subset of the independent variables. This method can perform classification- and/or prediction-related tasks by averaging the outcomes. It can also improve a model's predictive accuracy and control over-fitting.

This study used *randomForestSRC* package in R [57]. Following Alsharkawi et al. (2021) [58], all parameters were set to the default values. In particular, a total number of 1000 trees is sufficient, as shown in Hu et al. [51]. In addition to poverty prediction, Variable Importance (VIMP) and Minimal Depth (MD) analyses were conducted. These metrics use the main features obtained from all decision trees to assess the relative significance of explanatory variables in selecting the final predictors in the model.

4.4. Support Vector Regression (SVR)

Typically, the main objective in a linear regression framework is to minimize a specific loss function. For instance, OLS method aims to minimize the sum of squared errors. Methods such as lasso or ridge regression extend this framework by introducing additional penalty parameters to minimize complexity and/or reduce the number of covariates that marginally contribute to the model's predictive performance.

On the other hand, Vapnik (1998) [59] introduced the support vector regression (SVR), providing an alternative framework. Instead of minimizing a specific loss function, SVR is only concerned about reducing it to a certain degree. This gives greater flexibility in the estimation and helps in dealing with the limitations pertaining to distributional properties of the variables included in the analyses. As shown in the case of predicting the city-level poverty rate in Indonesia [60], flexibility with allowable error renders SVR superior to other conventional estimation methods that are fixated on minimizing a loss function. This study

used *e1071* package in R [61] to conduct the SVR-based prediction. All parameters were set to the default values.

5. Results

5.1. Preliminary Analysis

As preliminary estimation tools, we first estimated a full model and various model specifications using OLS and stepwise regression. In general, we found that the proportion of people living below the income-based poverty line and the value of the multidimensional poverty index are negatively associated with geospatial indicators that represent the degree of an area's urbanization, i.e., the intensity of NTL, building density, and a number of points of interest which are associated with the manufacturing and utility sectors. On the other hand, poverty outcomes are positively correlated with rainfall, NDVI, and other land cover classes that are typically associated with rural areas. While the directions of these correlations align with our expectations, the resulting adjusted R-square values are relatively low, ranging from 0.13 to 0.33.

5.2. Using Machine-Learning Algorithms to Predict Income-Based Poverty Rate

As previously described, all four predictive models were constructed by using the training datasets. Then, the poverty headcount values of 2015 and 2017 were predicted by applying the test datasets to the constructed models. Subsequently, the comparison of predictive power was based on the goodness-of-fit of the predicted outcomes. Figure 11 exhibits the comparison of the root mean square error (RMSE) (averaged across 100 trials) from the four computational methods, indicating that random forest yielded the lowest RMSE values (0.067 and 0.084 for 2015 and 2017, respectively). Under the same criterion, SVR performed second (with RMSE of 0.129 and 0.161 for 2015 and 2017, respectively), and GLS yielded the third lowest RMSE (0.133 and 0.170 for 2015 and 2017, respectively). Notably, the neural network generated the highest RMSE for 2015 and 2017 (with RMSE 0.419 and 0.549, respectively). Alternatively, the graphical illustration of the goodness-of-fit (Figures 12 and 13) also confirms that the random forest has the best predictive performance among the four methods that we have considered, generating predicted values closest to the actual ones.
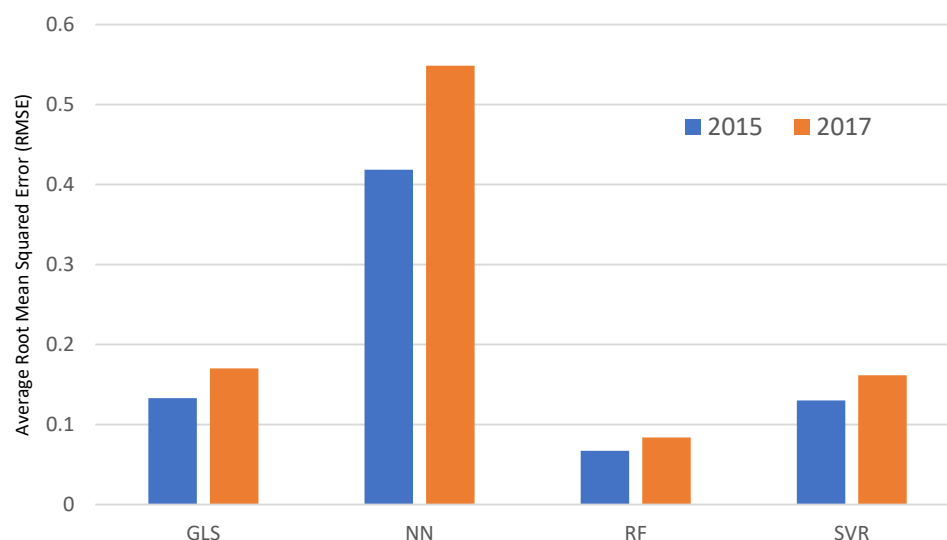

Figure 11. Comparison of average root mean squared error (RMSE) obtained from four machine-learning algorithms (on predicting poverty headcount rates of 2015 and 2017). Source: Calculation and graphics generated by authors.
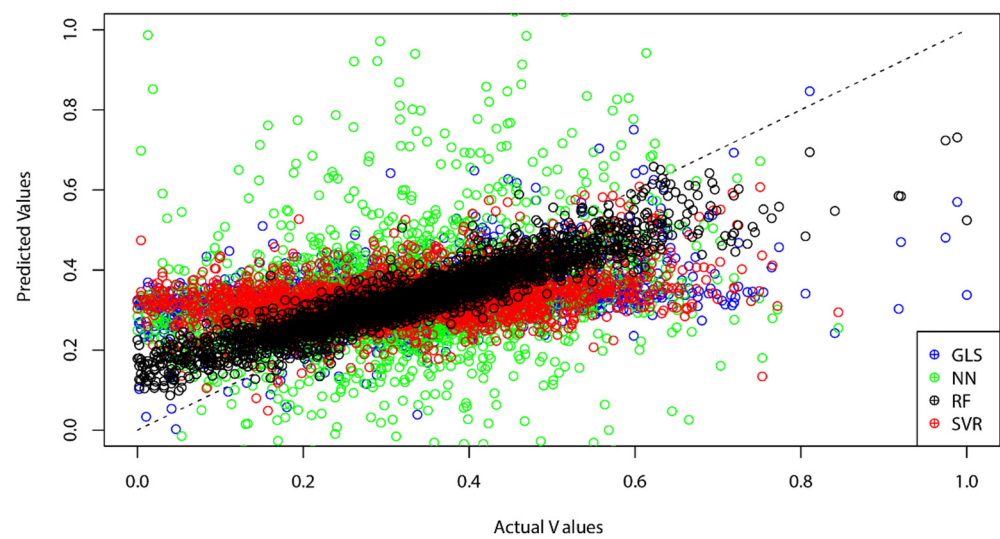
Figure 12. Scatter plot comparing the actual and predicted poverty headcount rates of 2015 (obtained from four machine-learning algorithms). Source: Calculation and graphics generated by authors.

Figure 13. Scatter plot comparing the actual and predicted poverty headcount rates of 2017 (obtained from four machine-learning algorithms). Source: Calculation and graphics generated by authors.
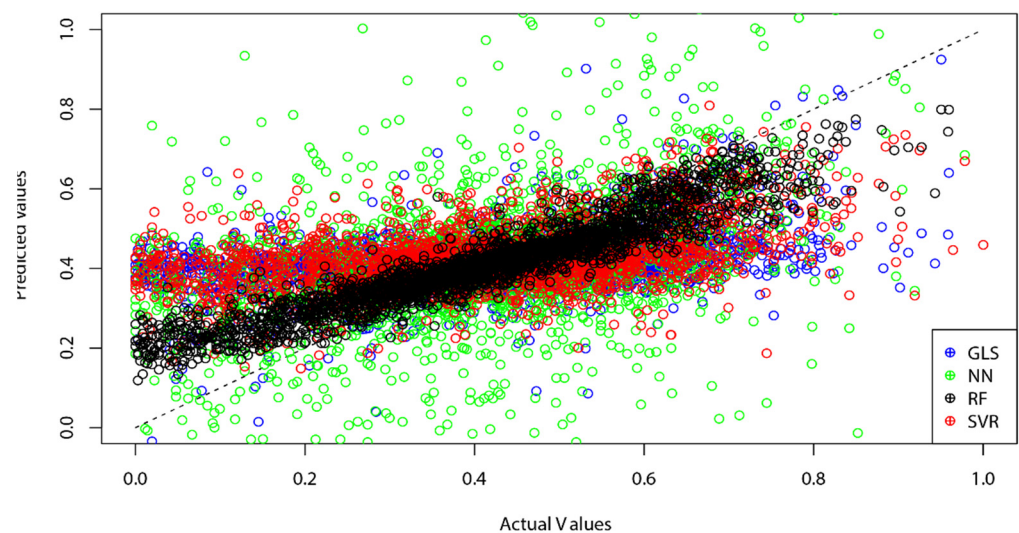
Based on the outcomes of random forest, Variable Importance (VIMP) and Minimal Depth (MD) were further conducted to prioritize the significance of each variable. Figures 14 and 15 show the result of VIMP for 2015 and 2017, while Figures 16 and 17 illustrate the outcomes of computing MD.

VIMP identified the intensity of NTL and population-density-related variables as the biggest contributors to the model. Meanwhile, five variables were identified as false positive in the VIMP's results for 2015 and 2017, indicating the irrelevance of these variables in predicting the poverty headcount rate. Alternatively, it is also possible that the information provided by these variables is already captured by other variables. The 'unimportant' variables are the area covered by tree or shrub (ESALC_12), the area covered by tree (broadleaved and deciduous more than 40%) (ESALC_61); the area covered by mosaic herbaceous (more than 50%) (ESALC_110); the area covered by tree, flooded, fresh, or brackish water (ESALC_160); and the bare areas (ESALC_200). The variable ranks indicated by VIMP are equally important, exhibiting the power-law distribution of relative magnitudes. Specifically, the magnitudes of top three variables are approximately three times higher than those of the fifth and lower ones.

Figure 14. Result of Variable Importance (VIMP) analysis on predicting poverty headcount rates of 2015 (based on RF's outcome). Source: Calculation and graphics generated by authors.

Figure 15. Result of Variable Importance (VIMP) analysis on predicting poverty headcount rates of 2017 (based on RF's outcome). Source: Calculation and graphics generated by authors.

Figure 16. Result of Minimal Depth (MD) analysis on predicting poverty headcount rates of 2015 (based on RF's outcome). Source: Calculation and graphics generated by authors.

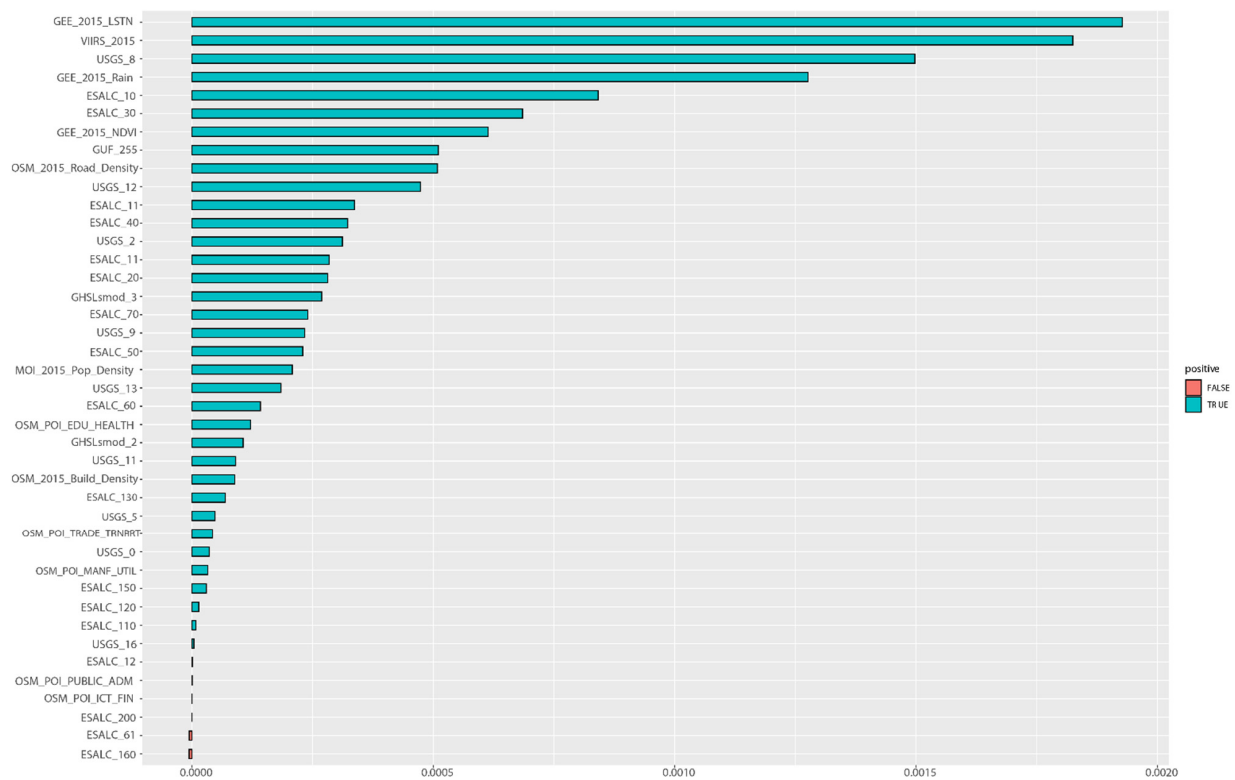**Figure 17.** Result of Minimal Depth (MD) analysis on predicting poverty headcount rates of 2017 (based on RF's outcome). Source: Calculation and graphics generated by authors.
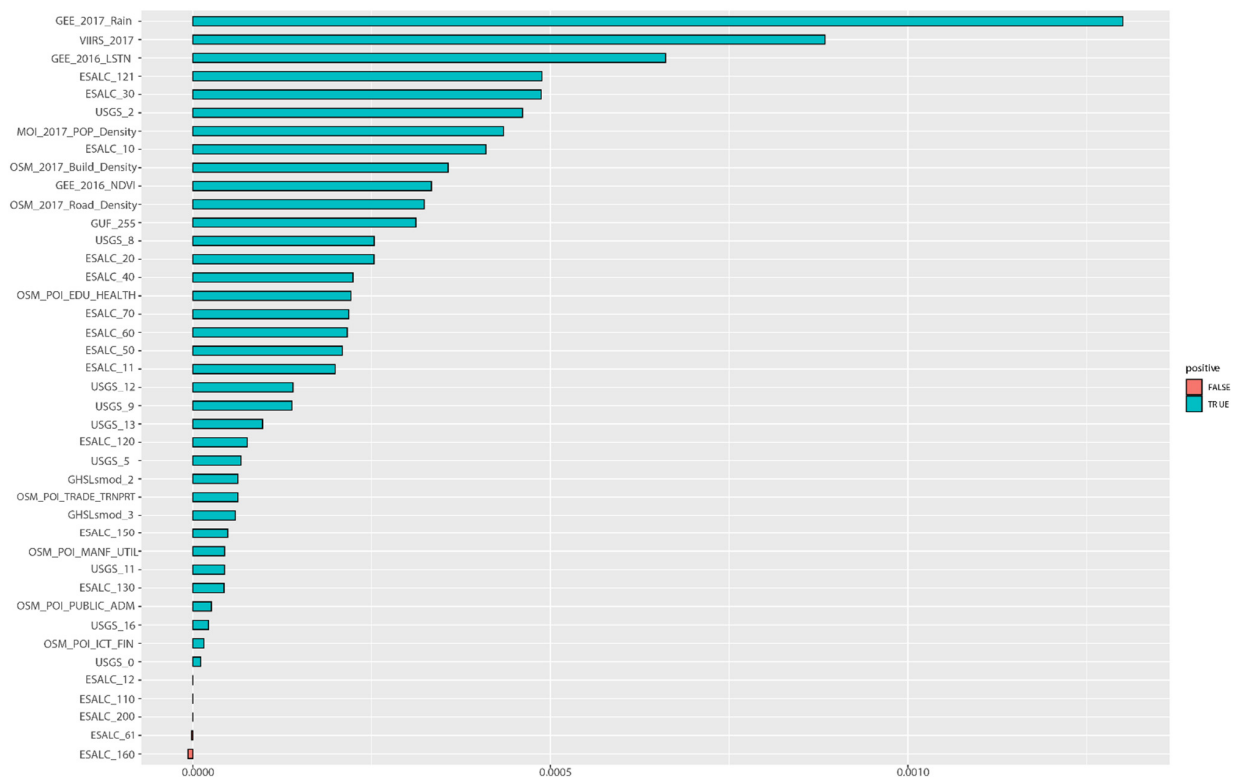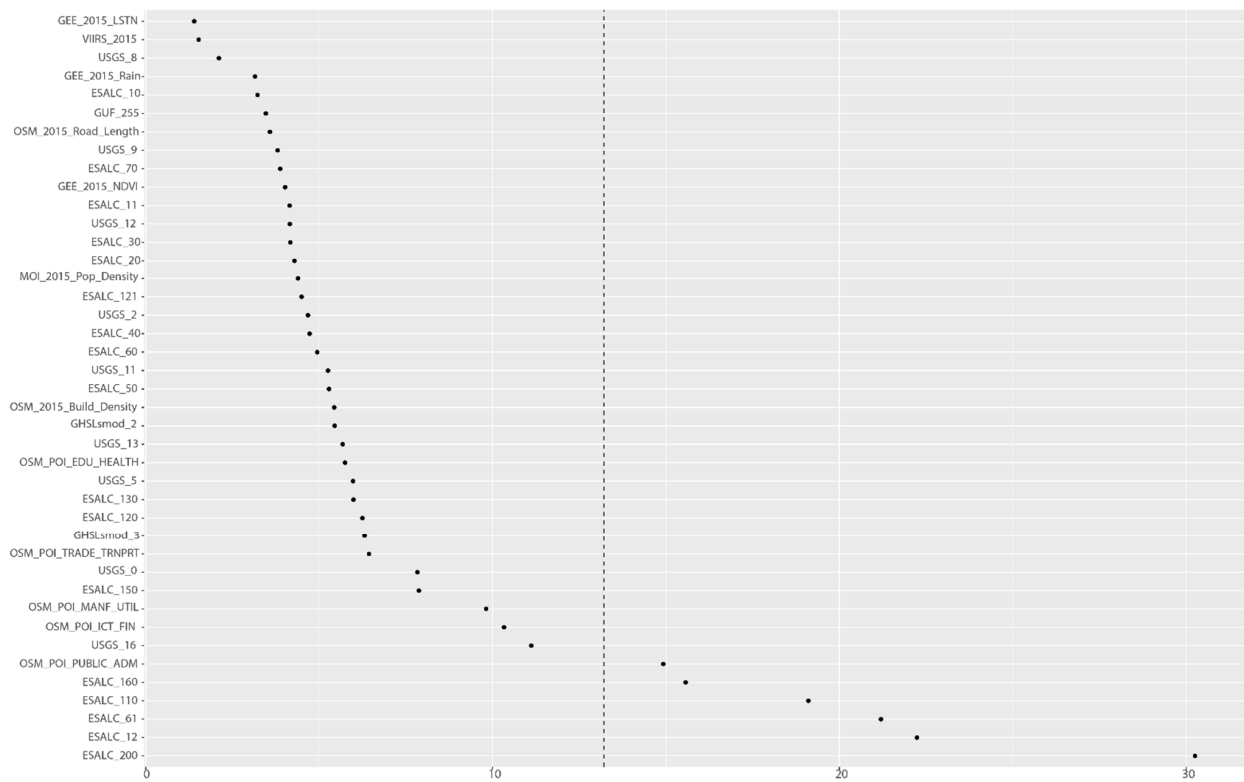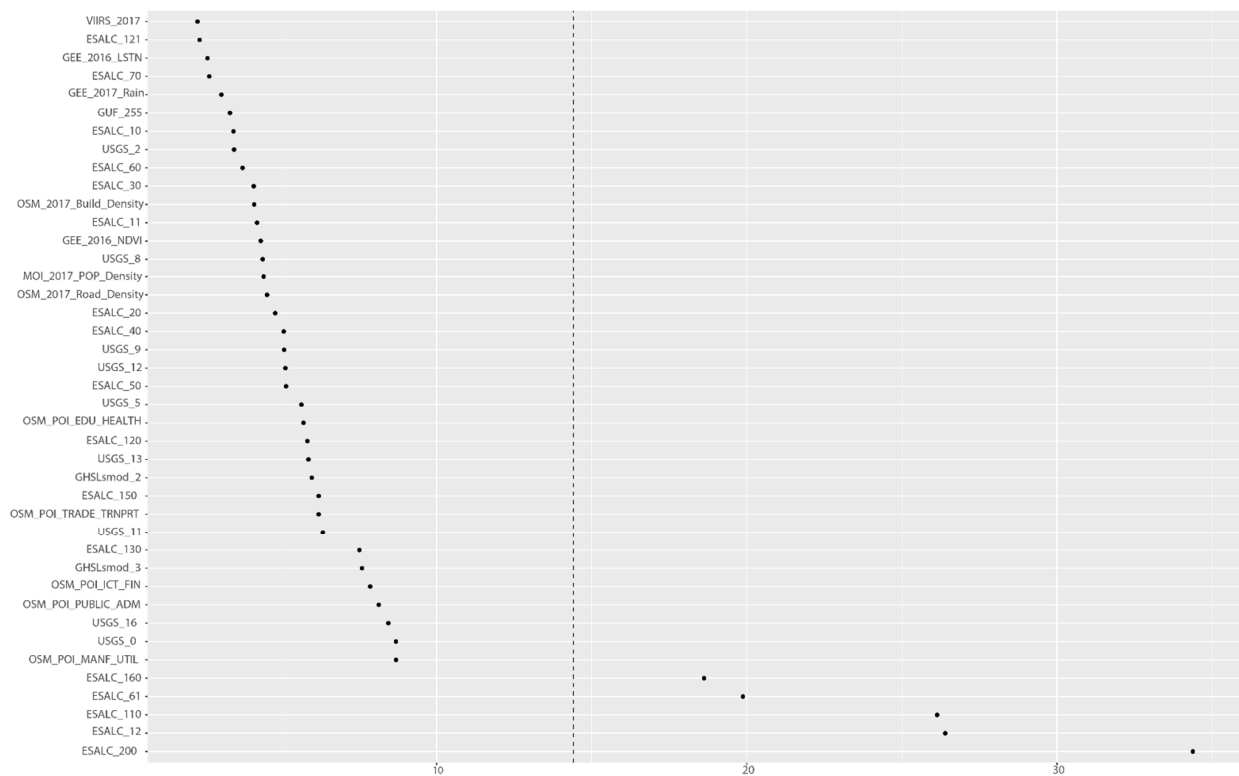
The results obtained from Minimal Depth (MD) calculation generated similar outcomes, confirming that intensity of NTL and population-density-related variables are highly associated with poverty headcount. In addition, the variables located on the right-hand side of the dashed line in Figures 16 and 17 are considered as having low explanatory power. The results based on this criterion show that five variables possess very low predictive power—the same five variables identified by VIMP result as irrelevant to the model. These variables can therefore be excluded from the model in further analysis.

5.3. Using Machine-Learning Algorithms to Predict Multidimensional Poverty Index (MPI)

In addition to the income-based poverty rate, we also applied GLS, neural network, random forest, and SVM to predict the MPI.

Figure 18 depicts the comparison of Root Mean Squared Error (RMSE) obtained from four machine-learning methods. Similar to the case of income poverty rates, the random forest method yielded the lowest RMSE (0.0877), while those of the SVR and GLS are almost identical (0.1631 and 0.1634, respectively). The neural network prediction produced the largest RMSE (0.2998). The scatterplot in Figure 19 compares the actual MPI and the predicted values. It shows that most predicted values generated by the random forest are located closest to the 45-degree line, suggesting that it has the best fit among the four methods considered in this study.
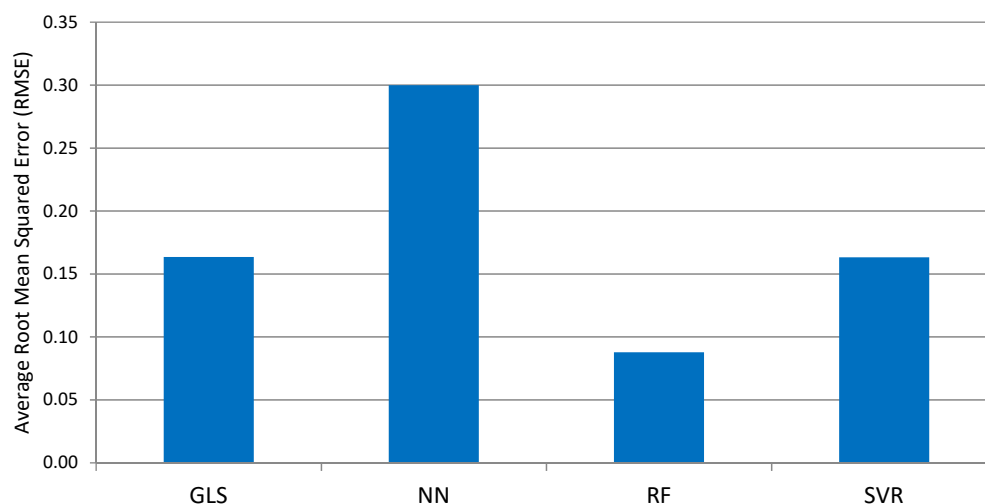
Figure 18. Comparison of average Root Mean Squared Error (RMSE) obtained from four machine-learning algorithms (on predicting MPI of 2017). Source: Calculation and graphics generated by authors.

Again, based on the outcome of random forest, we examined the degree of the explanatory power of each variable by calculating VIMP and MD. Figure 20 exhibits that based on VIMP, variables related to population density such as NTL, rainfall, Land Surface Temperature (LST), and road density have a high degree of contribution to predict the variation in the poverty rate. Similar to VIMP results of poverty headcount prediction (Figures 14 and 15), the order of magnitudes shows the power-law distribution of explanatory power. In particular, the magnitudes of the top four variables are approximately two times larger than the fifth and lower ones.

The result obtained from MD, as illustrated in Figure 21, also shows qualitatively similar results, revealing that NTL, LST, rainfall, road density, and the area covered by woody Savannas (USGS8) are key geographical features associated with the value of MPI.
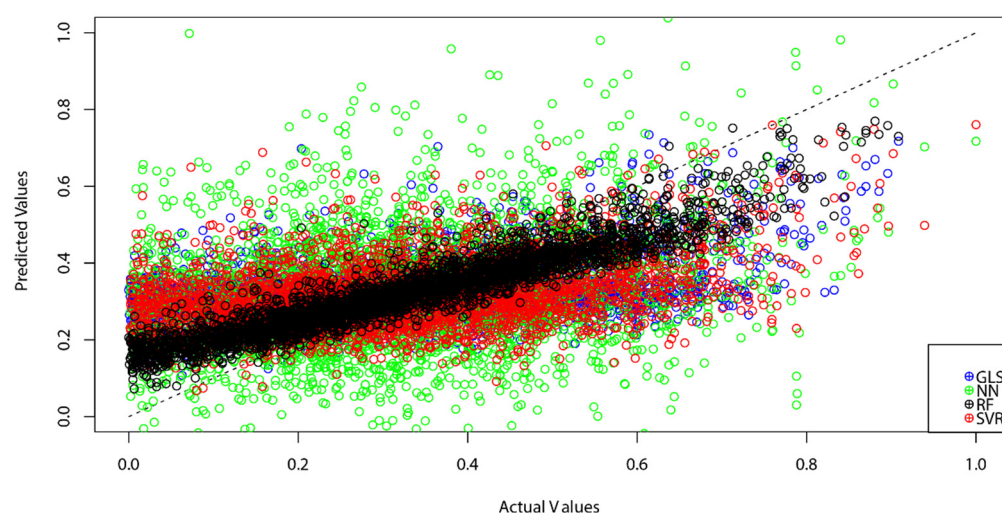
Figure 19. Scatter plot comparing the actual and predicted MPI of 2017 (obtained from four machine-learning algorithms). Source: Calculation and graphics generated by authors.
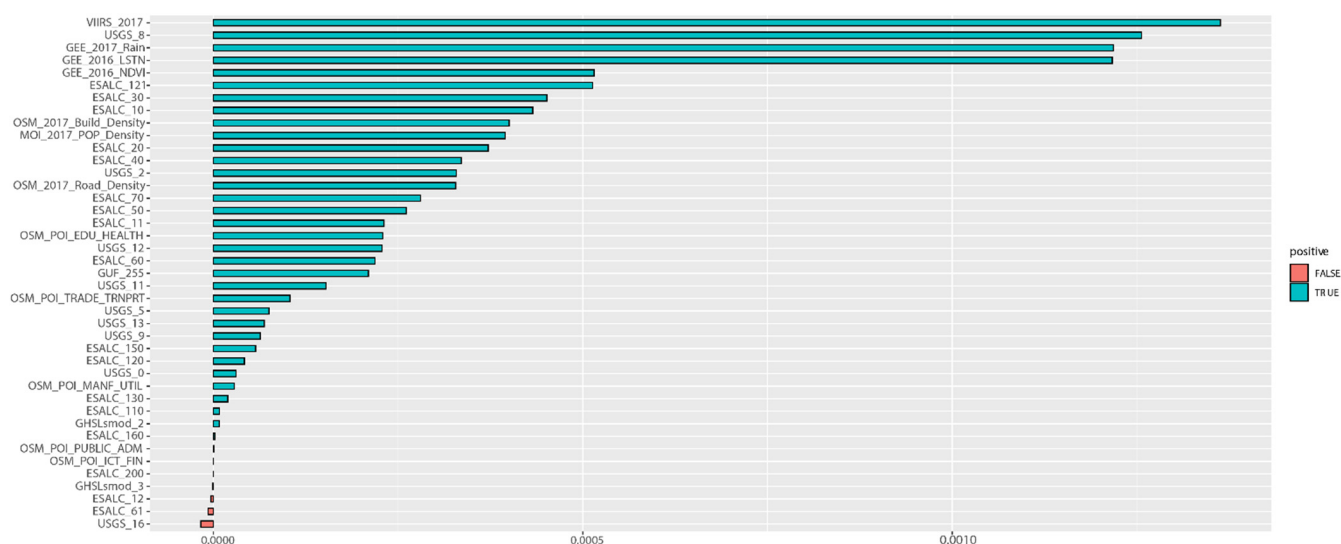
Figure 20. Result of Variable Importance (VIMP) analysis on predicting MPI of 2017 (based on RF's outcome). Source: Calculation and graphics generated by authors.

In summary, among the methods applied in this study, the random forest technique yielded the highest accuracy when predicting both the income poverty rate and the multidimensional poverty index. Furthermore, the resulting random forest models fit the datasets well, as suggested by the adjusted R-square values presented in Table 3.

Table 3. Adjusted R^2 values for RF-based models.

Variable Being Predicted	Adjusted R^2
NSO's poverty headcount of 2015	0.8526
NSO's poverty headcount of 2017	0.8459
TPMAP's poverty rate of 2017	0.8632

Source: Calculation and graphics generated by authors.
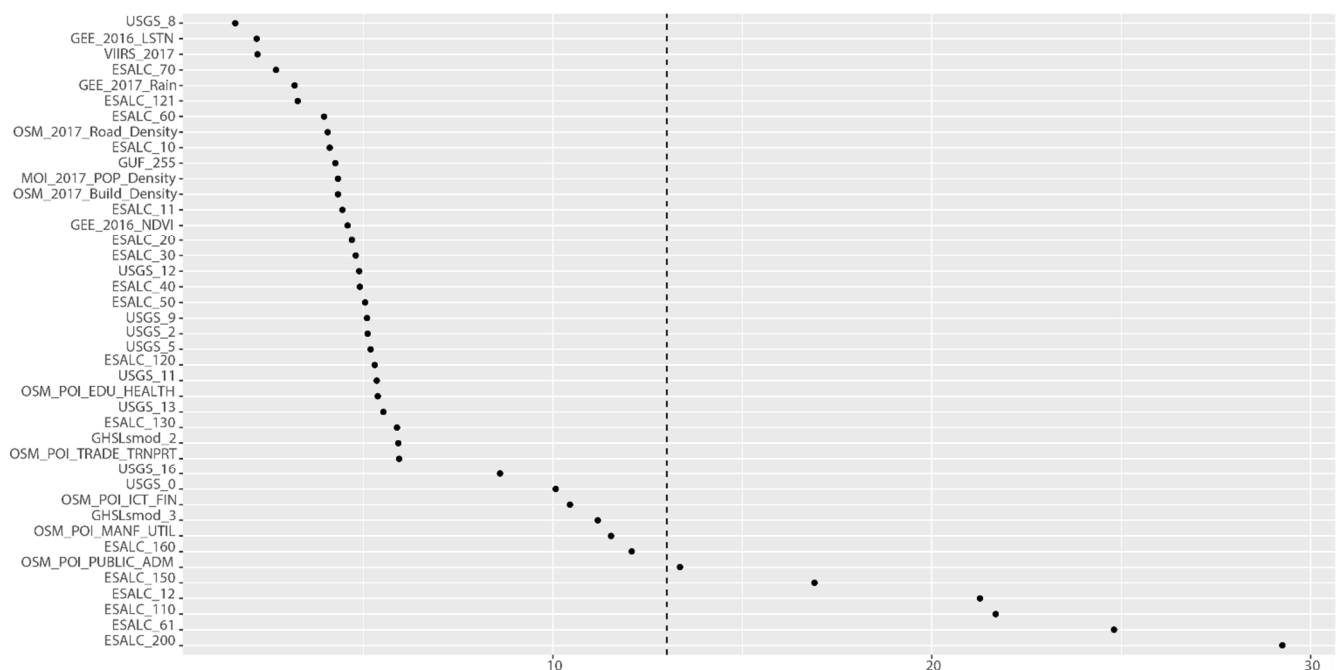
Figure 21. Result of Minimal Depth (MD) analysis on predicting MPI of 2017 (based on RF's outcome). Source: Calculation and graphics generated by authors.

6. Discussion

Globally, poverty and inequality have been major concerns for researchers and policymakers [62–66]. These development challenges have been similarly addressed at the regional level [67,68]. Hence, the spatiotemporal accuracy of data indicating socioeconomic conditions is invaluable for monitoring and formulating development programs. In the case of Thailand, the spatial analysis of poverty is crucial because the economic development has been geographically disproportionate for decades [69–71]. With the increasing accessibility of open data, the applications of geospatial indicators for poverty analysis have been recommended [72–74]. The poverty and inequality mapping have been developed using single-satellite data [75–79]. Alternatively, the estimation of spatial distribution can be enhanced by using a combination of remote sensing and geospatial indicators [80–83]. Following many publications' technical progress and data availability, this study integrated data obtained from open platforms such as Google Earth Engine [84], OpenStreetMap [85,86], and Point of Interest (POI) [87]. Similarly, guided by the international experience of implications, the machine-learning techniques were applied to predict poverty indicators [50,51,88–94].

The obtained results of this study show that random forest yielded the highest predictive power, which is in accordance with findings of several previous publications [51,88,90–92,95]. In particular, the obtained accuracy is higher than 70%, similar to other studies using the random forest to predict poverty [51,89,90]. Likewise, a review conducted by the World Bank [93] suggested that random forest could contribute to a highly accurate predictor of poverty. Fundamentally, one of the unique features of the random forest method is the combination of decision trees, allowing the discrete and continuous explanatory variables to predict the output jointly. Moreover, the random forest enables extended analyses of VIMP and MD, empowering the ranking of the explanatory power of independent variables. The outcomes obtained from the two methods show that NTL has a very high predictive power, which is similar to the case of predicting county-level poverty in China [96] and Bangladesh [90]. The obtained results are in accordance with those of previous literature, suggesting that several geospatial characteristics are associated with poverty, such as travel accessibility [90,97,98], proximity to important public services [99], and land-use types, as well as household surroundings [51,100].

The outcome of this study emphasized the significance of geographical conditions as crucial factors influencing the socioeconomic status of households. Similar to the cases of Latin America [101], Africa [99], and neighboring countries in Southeast Asia [102], agriculture and resource-based manufacturing are the main economic activities of low-income families. Thus, geographical features related to those activities represent the high concentration of poverty. Furthermore, the geographical isolation from markets and other infrastructures is the major constraint to access economic opportunities, implying the extreme poverty condition [103]. On the contrary, proximity to urban areas and infrastructures provides access to job opportunity, healthcare, education, and the city's agglomeration force [94]. Therefore, development policies can alleviate poverty by either expanding infrastructures or relocating low-income families [104].

There are several areas which merit further investigation. Firstly, the analysis can be extended to include other data sources (e.g., mobile phone data and texture features), enabling the multidimensional examination of spatial associations [79,80,105,106]. Secondly, the temporal coverage of survey-based data should be lengthened, allowing larger datasets for training models [107,108]. Thirdly, the spatial resolution should be enhanced in order to identify the urban poor (i.e., slums), which would broaden insights on intra-city inequality [109–111].

7. Conclusions

The contribution of this study is twofold. Firstly, it introduces the integration of data, composed of the nationwide survey, register-based data, geospatial information, and satellite imagery. Secondly, this paper has applied computational techniques to examine the relationship between geospatial features such as intensity of NTL, land cover, land use, etc., and the proportion of people living below the poverty line as measured using the conventional method of estimating poverty. It is shown that the Random Forest is the best prediction method, yielding an accuracy of more than 80%. Furthermore, the results obtained from Variable Importance (VIMP) and Minimal Depth (MD) reveal the associations between geospatial covariates such as intensity of NTL, population density, and poverty rates. These contributions suggest the potential of applying the open data and open-source computational tools for timely analysis of the spatial distribution of poverty, especially for developing countries which conventionally have data-compilation constraints.

Author Contributions: Conceptualization, Nattapong Puttanapong and Arturo Martinez, Jr.; methodology, Nattapong Puttanapong; software, Nattapong Puttanapong; validation, Joseph Albert Nino Bulan, Mildred Addawe, Ron Lester Durante and Marymell Martillan; formal analysis, Nattapong Puttanapong; data curation, Nattapong Puttanapong, Arturo Martinez, Jr., Joseph Albert Nino Bulan, Mildred Addawe, Ron Lester Durante and Marymell Martillan; writing—original draft preparation, Nattapong Puttanapong and Arturo Martinez, Jr.; writing—review and editing, Arturo Martinez, Jr., Joseph Albert Nino Bulan, Mildred Addawe, Ron Lester Durante and Marymell Martillan; visualization, Nattapong Puttanapong; supervision, Arturo Martinez, Jr.; project administration, Arturo Martinez, Jr. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. List of variables obtained from geospatial data of 2015.

Variable	Definition
VNTL2015f_sum	VIIRS cloud mask—outlier removed—NTL average DNB radiance, year 2015
log_VIIRS_2015	Logarithm (based 10) of VIIRS cloud mask—outlier removed—NTL average DNB radiance, 2015
log_VIIRS_density_2015	Logarithm (based 10) of VIIRS cloud mask—outlier removed—NTL average DNB radiance, 2015, per area
log_POP_2018	Logarithm (based 10) of population size, 2018
log_Total_Pop_density	Logarithm (based 10) of population density, year 2015
log_LST_2015	Logarithm (based 10) of land surface temperature, 2015
log_Rain_2015	Logarithm (based 10) of amount of rainfall, 2015
GUF_255	Global Urban Footprint, pixel count of built-up areas (2011–12)
GHSLsmod2015_2	Global Human Settlement Layer, pixel count of "urban clusters" or low-density clusters, 2015
GHSLmod2015_3	Global Human Settlement Layer, pixel count of "urban centres" or high-density clusters, 2015
log_NDVI_2015	Logarithm (based 10) of normalized difference of vegetation index, 2015
USGS_0	USGS Land Cover, pixel count of Water (2001–2010 data)
USGS_2	USGS Land Cover, pixel count of Evergreen Broadleaf Forest (2001–2010 data)
USGS_5	USGS Land Cover, pixel count of Mixed Forests (2001–2010 data)
USGS_8	USGS Land Cover, pixel count of Woody Savannas (2001–2010 data)
USGS_9	USGS Land Cover, pixel count of Savannas (2001–2010 data)
USGS_11	USGS Land Cover, pixel count of Permanent Wetland (2001–2010 data)
USGS_12	USGS Land Cover, pixel count of Croplands (2001–2010 data)
USGS_13	USGS Land Cover, pixel count of Urban and Built-up (2001–2010 data)
USGS_16	USGS Land Cover, pixel count of Barren or Sparsely Vegetated (2001–2010 data)
ESALC2015_10	ESA Land Cover, pixel count of Cropland, rainfed, 2015
ESALC2015_11	ESA Land Cover, pixel count of Herbaceous cover, 2015
ESALC2015_12	ESA Land Cover, pixel count of Tree or shrub cover, 2015
ESALC2015_20	ESA Land Cover, pixel count of Cropland, irrigated or post-flooding, 2015
ESALC2015_30	ESA Land Cover, pixel count of Mosaic cropland (>50%)/natural vegetation (tree, shrub, herbaceous cover) (<50%), 2015
ESALC2015_40	ESA Land Cover, pixel count of Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%)/cropland (<50%), 2015
ESALC2015_50	ESA Land Cover, pixel count of Tree cover, broadleaved, evergreen, closed to open (>15%), 2015
ESALC2015_60	ESA Land Cover, pixel count of Tree cover, broadleaved, deciduous, closed to open (>15%), 2015
ESALC2015_61	ESA Land Cover, pixel count of Tree cover, broadleaved, deciduous, closed (>40%), 2015
ESALC2015_70	ESA Land Cover, pixel count of Tree cover, needleeaved, evergreen, closed to open (>15%), 2015
ESALC2015_110	ESA Land Cover, pixel count of Mosaic herbaceous cover (>50%)/tree and shrub (<50%), 2015
ESALC2015_120	ESA Land Cover, pixel count of Shrubland, 2015
ESALC2015_121	ESA Land Cover, pixel count of Evergreen shrubland, 2015
ESALC2015_130	ESA Land Cover, pixel count of Grassland, 2015
ESALC2015_150	ESA Land Cover, pixel count of Sparse vegetation (tree, shrub, herbaceous cover) (<15%), 2015
ESALC2015_160	ESA Land Cover, pixel count of Tree cover, flooded, fresh or brackish water, 2015
ESALC2015_200	ESA Land Cover, pixel count of Bare areas, 2015
Density_2015_Road_Count	Number of road paths per area, 2015
Density_2015_Road_Length	Total length of road paths per area, 2015
Density_2015_POI	Number of Point of Interest (POI) per area, 2015
NESDC5_3	Number of POIs in 2015 of this type: manufacturing
NESDC5_7	Number of POIs in 2015 of this type: wholesale and retail trade and repair of motor vehicles
NESDC5_8	Number of POIs in 2015 of this type: transportation and storage
NESDC5_9	Number of POIs in 2015 of this type: accommodation and food-service activities
NESDC5_10	Number of POIs in 2015 of this type: information and communication
NESDC5_11	Number of POIs in 2015 of this type: financial and insurance activities
NESDC5_13	Number of POIs in 2015 of this type: professional, scientific, and technical activities
NESDC5_14	Number of POIs in 2015 of this type: administrative and support service activities
NESDC5_15	Number of POIs in 2015 of this type: public administration and defense; compulsory social security
NESDC5_16	Number of POIs in 2015 of this type: education

Table A1. *Cont.*

Variable	Definition
NESDC5_17	Number of POIs in 2015 of this type: human health activities
NESDC5_industry	Number of POIs in 2015 of this type: mining and quarrying/manufacturing/electricity, gas, steam, and air-conditioning supply/water supply, sewerage, waste management, and remediation activities/construction
NESDC5_svcs1	Number of POIs in 2015 of this type: wholesale and retail trade and repair of motor vehicles/transportation and storage/accommodation and food service activities
NESDC5_svcs2	Number of POIs in 2015 of this type: information and communication/financial and insurance activities/real-estate activities/professional, scientific, and technical activities/administrative and support-service activities
NESDC5_svcs3	Number of POIs in 2015 of this type: public administration and defense; compulsory social security/education/human health activities/arts, entertainment and recreation/other service activities
Density_POI_Area_2015	Total area of Point of Interest per area, 2015
Density_NESDC5_industry	Number of POIs per sq km in 2015 of this type: mining and quarrying/manufacturing/electricity, gas, steam, and air-conditioning supply/water supply, sewerage, waste management, and remediation activities/construction
Density_NESDC5_svcs1	Number of POIs per sq km in 2015 of this type: wholesale and retail trade and repair of motor vehicles/transportation and storage/accommodation and food-service activities
Density_NESDC5_svcs2	Number of POIs per sq km in 2015 of this type: information and communication/financial and insurance activities/real-estate activities/professional, scientific, and technical activities/administrative and support service activities
Density_NESDC5_svcs3	Number of POIs per sq km in 2015 of this type: public administration and defense; compulsory social security/education/human health activities/arts, entertainment and recreation/other service activities

Table A2. List of variables obtained from geospatial data of 2017.

Variable	Definition
VNTL2017f_sum	VIIRS cloud mask—outlier removed—NTL average DNB radiance, year 2017
log_VIIRS_2017	Logarithm (based 10) of VIIRS cloud mask—outlier removed—NTL average DNB radiance, 2017
log_VIIRS_density_2017	Logarithm (based 10) of VIIRS cloud mask—outlier removed—NTL DNB radiance, 2017, per area
log_POP_2018	Logarithm (based 10) of population size, 2018
log_Total_Pop_density	Logarithm (based 10) of population density, year 2017
log_LST_2017	Logarithm (based 10) of land surface temperature, 2017
log_Rain_2017	Logarithm (based 10) of amount of rainfall, 2017
SYNMAP_46	Synergetic Land Cover, pixel count of urban, 2000
USGS_13	USGS Land Cover, pixel count of urban and built-up areas (2001–2010 data)
GUF_255	Global Urban Footprint, pixel count of built-up areas (2011–12)
log_GUF_255	Logarithm (based 10) of Global Urban Footprint, pixel count of built-up areas (2011–12
GHSLsmod2017_2	Global Human Settlement Layer, pixel count of "urban centres" or low-density clusters, 2017
GHSLmod2017_3	Global Human Settlement Layer, pixel count of "urban centres" or high-density clusters, 2017
log_NDVI_2017	Logarithm (based 10) of normalized difference of vegetation index, 2017
log_NDVI_density_2017	Logarithm (based 10) of normalized difference of vegetation index, 2017, per area
USGS_0	USGS Land Cover, pixel count of Water (2001–2010 data)
USGS_1	USGS Land Cover, pixel count of Evergreen Needle Leaf Forest (2001–2010 data)
USGS_2	USGS Land Cover, pixel count of Evergreen Broadleaf Forest (2001–2010 data)
USGS_3	USGS Land Cover, pixel count of Deciduous Needle Leaf Forest (2001–2010 data)
USGS_4	USGS Land Cover, pixel count of Deciduous Broadleaf Forest (2001–2010 data)
USGS_5	USGS Land Cover, pixel count of Mixed Forests (2001–2010 data)
USGS_6	USGS Land Cover, pixel count of Closed Shrublands (2001–2010 data)
USGS_7	USGS Land Cover, pixel count of Open Shrublands (2001–2010 data)
USGS_8	USGS Land Cover, pixel count of Woody Savannas (2001–2010 data)
USGS_9	USGS Land Cover, pixel count of Savannas (2001–2010 data)
USGS_10	USGS Land Cover, pixel count of Grasslands (2001–2010 data)

Table A2. *Cont.*

Variable	Definition
USGS_11	USGS Land Cover, pixel count of Permanent Wetland (2001–2010 data)
USGS_12	USGS Land Cover, pixel count of Croplands (2001–2010 data)
USGS_13	USGS Land Cover, pixel count of Urban and Built-up (2001–2010 data)
USGS_14	USGS Land Cover, pixel count of Cropland/Natural Vegetation Mosaic (2001–2010 data)
USGS_16	USGS Land Cover, pixel count of Barren or Sparsely Vegetated (2001–2010 data)
USGS_PCA	USGS Land Cover, First Principal Component of USGS_0—USGS_16
ESALC2017_10	ESA Land Cover, pixel count of Cropland, rainfed, 2017
ESALC2017_11	ESA Land Cover, pixel count of Herbaceous cover, 2017
ESALC2017_12	ESA Land Cover, pixel count of Tree or shrub cover, 2017
ESALC2017_20	ESA Land Cover, pixel count of Cropland, irrigated or post-flooding, 2017
ESALC2017_30	ESA Land Cover, pixel count of Mosaic cropland (>50%)/natural vegetation (tree, shrub, herbaceous cover) (<50%), 2017
ESALC2017_40	ESA Land Cover, pixel count of Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%)/cropland (<50%), 2017
ESALC2017_50	ESA Land Cover, pixel count of Tree cover, broadleaved, evergreen, closed to open (>15%), 2017
ESALC2017_60	ESA Land Cover, pixel count of Tree cover, broadleaved, deciduous, closed to open (>15%), 2017
ESALC2017_61	ESA Land Cover, pixel count of Tree cover, broadleaved, deciduous, closed (>40%), 2017
ESALC2017_70	ESA Land Cover, pixel count of Tree cover, needleaved, evergreen, closed to open (>15%), 2017
ESALC2017_80	ESA Land Cover, pixel count of Tree cover, needleaved, deciduous, closed to open (>15%), 2017
ESALC2017_100	ESA Land Cover, pixel count of Mosaic tree and shrub (>50%)/herbaceous cover(<50%), 2017
ESALC2017_110	ESA Land Cover, pixel count of Mosaic herbaceous cover (>50%)/tree and shrub (<50%), 2015
ESALC2017_120	ESA Land Cover, pixel count of Shrubland, 2017
ESALC2017_121	ESA Land Cover, pixel count of Evergreen shrubland, 2017
ESALC2017_122	ESA Land Cover, pixel count of Deciduous shrubland, 2017
ESALC2017_130	ESA Land Cover, pixel count of Grassland, 2017
ESALC2017_150	ESA Land Cover, pixel count of Sparse vegetation (tree, shrub, herbaceous cover) (<15%), 2017
ESALC2017_160	ESA Land Cover, pixel count of Tree cover, flooded, fresh or brackish water, 2017
ESALC2017_170	ESA Land Cover, pixel count of Tree cover, flooded, saline water, 2017
ESALC2017_180	ESA Land Cover, pixel count of Shrub or herbaceous cover, flooded, fresh/saline/brackish water, 2017
ESALC2017_190	ESA Land Cover, pixel count of Urban areas, 2017
ESALC2017_200	ESA Land Cover, pixel count of Bare areas, 2017
ESALC2017_210	ESA Land Cover, pixel count of Water bodies, 2017
Density_2017_Road_Count	Number of road paths per area, 2017
Density_2017_Road_Length	Total length of road paths per area, 2017
Density_2017_POI	Number of Point of Interest (POI) per area, 2017
NESDC7_3	Number of POIs in 2017 of this type: manufacturing
NESDC7_7	Number of POIs in 2017 of this type: wholesale and retail trade and repair of motor vehicles
NESDC7_8	Number of POIs in 2017 of this type: transportation and storage
NESDC7_9	Number of POIs in 2017 of this type: accommodation and food service activities
NESDC7_10	Number of POIs in 2017 of this type: information and communication
NESDC7_11	Number of POIs in 2017 of this type: financial and insurance activities
NESDC7_13	Number of POIs in 2017 of this type: professional, scientific, and technical activities
NESDC7_14	Number of POIs in 2017 of this type: administrative and support service activities
NESDC7_15	Number of POIs in 2017 of this type: public administration and defense; compulsory social security
NESDC7_16	Number of POIs in 2017 of this type: education
NESDC7_17	Number of POIs in 2017 of this type: human health activities
NESDC7_industry	Number of POIs in 2017 of this type: mining and quarrying/manufacturing/electricity, gas, steam, and air conditioning supply/water supply, sewerage, waste management and remediation activities/construction
NESDC7_svcs1	Number of POIs in 2017 of this type: wholesale and retail trade and repair of motor vehicles/transportation and storage/accommodation and food service activities
NESDC7_svcs2	Number of POIs in 2017 of this type: information and communication/financial and insurance activities/real estate activities/professional, scientific, and technical activities/administrative and support service activities

Variable	Definition
NESDC7_svcs3	Number of POIs in 2017 of this type: public administration and defense; compulsory social security/education/human health activities/arts, entertainment and recreation/other service activities
Density_POI_Area_2017	Total area of Point of Interest per area, 2017
Density_NESDC7_industry	Number of POIs per sq km in 2017 of this type: mining and quarrying/manufacturing/electricity, gas, steam, and air-conditioning supply/water supply, sewerage, waste management and remediation activities/construction
Density_NESDC7_svcs1	Number of POIs per sq km in 2017 of this type: wholesale and retail trade and repair of motor vehicles/transportation and storage/accommodation and food-service activities
Density_NESDC7_svcs2	Number of POIs per sq km in 2017 of this type: information and communication/financial and insurance activities/real estate activities/professional, scientific, and technical activities/administrative and support-service activities
Density_NESDC7_svcs3	Number of POIs per sq km in 2017 of this type: public administration and defense; compulsory social security/education/human health activities/arts, entertainment, and recreation/other service activities
log_House_density_2017	Logarithm (based 10) of registered house, 2017, per area
Density_Building_Area	Total sq. meter of building per area, year 2017
log_F2017_Buil_Density	Logarithm (based 10) of build-up (square meter), 2017, per area

References

1. National Statistical Office. *Key Statistical Data*; National Statistical Office: Bangkok, Thailand, 2000.
2. ADB. Mapping Poverty through Data Integration and Artificial Intelligence: A Special Supplement of the Key Indicators for Asia and the Pacific. In *A Special Supplement of the Key Indicators for Asia and the Pacific 2020*; ADB: Manila, Philippines, 2020. [CrossRef]
3. Jean, N.; Burke, M.; Xie, M.; Davis, W.; Lobell, D.; Ermon, S. Combining satellite imagery and machine learning to predict poverty. *Science* **2016**, *353*, 790–794. [CrossRef] [PubMed]
4. Hofer, M.; Sako, T.; Martinez, A., Jr.; Addawe, M.; Bulan, J.; Durante, R.L.; Martillan, M. Applying Artificial Intelligence on Satellite Imagery to Compile Granular Poverty Statistics. In *Asian Development Bank Economics Working Paper Series*; Asian Development Bank: Manila, Philippines, 2020. [CrossRef]
5. Piaggesi, S.; Gauvin, L.; Tizzoni, M.; Cattuto, C.; Adler, N.; Verhulst, S.; Young, A.; Price, R.; Ferres, L.; Panisson, A. Predicting City Poverty Using Satellite Imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 16–20 June 2019; pp. 90–96.
6. Elvidge, C.D.; Baugh, K.E.; Kihn, E.A.; Kroehl, H.W.; Davis, E.R.; Davis, C.W. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *Int. J. Remote Sens.* **1997**, *18*, 1373–1379. [CrossRef]
7. Doll, C.N.H.; Jan-Peter, M.; Elvidge, C.D. Night-Time Imagery as a Tool for Global Mapping of Socioeconomic Parameters and Greenhouse Gas Emissions. *Ambio* **2000**, *29*, 157–162. [CrossRef]
8. Doll, C.N.H.; Muller, J.-P.; Morley, J.G. Mapping regional economic activity from night-time light satellite imagery. *Ecol. Econ.* **2006**, *57*, 75–92. [CrossRef]
9. Sutton, P.C.; Elvidge, C.D.; Ghosh, T. Estimation of Gross Domestic Product at Sub-National Scales using Nighttime Satellite Imagery. *Int. J. Ecol. Econ. Stat.* **2007**, *8*, 5–21.
10. Henderson, J.V.; Storeygard, A.; Weil, D.N. Measuring Economic Growth from Outer Space. *Am. Econ. Rev.* **2012**, *102*, 994–1028. [CrossRef]
11. Bickenbach, F.; Bode, E.; Nunnenkamp, P.; Söder, M. Night lights and regional GDP. *Rev. World Econ.* **2016**, *152*, 425–447. [CrossRef]
12. Forbes, D.J. Multi-scale analysis of the relationship between economic statistics and DMSP-OLS night light images. *GISci. Remote Sens.* **2013**, *50*, 483–499. [CrossRef]
13. Li, X.; Xu, H.; Chen, X.; Li, C. Potential of NPP-VIIRS Nighttime Light Imagery for Modeling the Regional Economy of China. *Remote Sens.* **2013**, *5*, 3057–3081. [CrossRef]
14. Li, D.; Zhao, X.; Li, X. Remote sensing of human beings—A perspective from nighttime light. *Geo-Spat. Inf. Sci.* **2016**, *19*, 69–79. [CrossRef]
15. Sun, J.; Wang, X.; Chen, A.; Ma, Y.; Cui, M.; Piao, S. NDVI indicated characteristics of vegetation cover change in China's metropolises over the last three decades. *Environ. Monit. Assess.* **2011**, *179*, 1–14. [CrossRef] [PubMed]
16. Li, G.; Chen, S.S.; Yan, Y.-H.; Yu, C. Effects of Urbanization on Vegetation Degradation in the Yangtze River Delta of China: Assessment Based on SPOT-VGT NDVI. *J. Urban Plan. Dev.-Asce* **2015**, *141*, 05014026. [CrossRef]
17. Jin, X.M.; Wan, L.; Zhang, Y.K.; Schaepman, M. Impact of economic growth on vegetation health in China based on GIMMS NDVI. *Int. J. Remote Sens.* **2008**, *29*, 3715–3726. [CrossRef]

18. Kristjanson, P.; Radeny, M.; Baltenweck, I.; Ogutu, J.; Notenbaert, A. Livelihood mapping and poverty correlates at a meso-level in Kenya. *Food Policy* **2005**, *30*, 568–583. [CrossRef]

19. Bhattacharya, H.; Innes, R.D. Is There a Nexus between Poverty and Environment in Rural India. In Proceedings of the American Agricultural Economics Association Annual Meeting, Long Beach, CA, USA, 23–26 July 2006.

20. Morikawa, R. Remote Sensing Tools for Evaluating Poverty Alleviation Projects: A Case Study in Tanzania. *Procedia Eng.* **2014**, *78*, 178–187. [CrossRef]

21. Aburas, M.M.; Abdullah, S.H.; Ramli, M.F.; Ash'aari, Z.H. Measuring Land Cover Change in Seremban, Malaysia Using NDVI Index. *Procedia Environ. Sci.* **2015**, *30*, 238–243. [CrossRef]

22. Weng, Q. A remote sensing?GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, China. *Int. J. Remote Sens.* **2001**, *22*, 1999–2014. [CrossRef]

23. Buyantuyev, A.; Wu, J. Urban heat islands and landscape heterogeneity: Linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landsc. Ecol.* **2009**, *25*, 17–33. [CrossRef]

24. Huang, G.; Zhou, W.; Cadenasso, M.L. Is everyone hot in the city? Spatial pattern of land surface temperatures, land cover and neighborhood socioeconomic characteristics in Baltimore, MD. *J. Environ. Manag.* **2011**, *92*, 1753–1759. [CrossRef]

25. Ruthirako, P.; Darnsawasdi, R.; Chatupote, W. Intensity and Pattern of Land Surface Temperature in Hat Yai City, Thailand. *Walailak J. Sci. Technol.* **2015**, *12*, 83–94. [CrossRef]

26. Youneszadeh, S.; Amiri, N.; Pilesjö, P. The effect of land use change on land surface temperature in the Netherlands. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 745–748. [CrossRef]

27. Cooper, L.A.; Ballantyne, A.P.; Holden, Z.A.; Landguth, E.L. Disturbance impacts on land surface temperature and gross primary productivity in the western United States. *J. Geophys. Res. Biogeosci.* **2017**, *122*, 930–946. [CrossRef]

28. Dissanayake, D.; Morimoto, T.; Murayama, Y.; Ranagalage, M.; Handayani, H.H. Impact of Urban Surface Characteristics and Socio-Economic Variables on the Spatial Variation of Land Surface Temperature in Lagos City, Nigeria. *Sustainability* **2019**, *11*, 25. [CrossRef]

29. Richardson, C.J. How Much Did Droughts Matter? Linking Rainfall and GDP Growth in Zimbabwe. *Afr. Aff.* **2007**, *106*, 463–478. [CrossRef]

30. Maccini, S.; Yang, D. Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall. *Am. Econ. Rev.* **2009**, *99*, 1006–1026. [CrossRef] [PubMed]

31. Arezki, R.; Brückner, M. Rainfall, financial development, and remittances: Evidence from Sub-Saharan Africa. *J. Int. Econ.* **2012**, *87*, 377–385. [CrossRef]

32. Thiede, B.C. Rainfall Shocks and Within-Community Wealth Inequality: Evidence from Rural Ethiopia. *World Dev.* **2014**, *64*, 181–193. [CrossRef]

33. Sarsons, H. Rainfall and conflict: A cautionary tale. *J. Dev. Econ.* **2015**, *115*, 62–72. [CrossRef]

34. Gilmont, M.; Hall, J.; Grey, D.; Dadson, S.; Simpson, M.; Abele, S. Analysis of the relationship between rainfall and economic growth in Indian states. *Glob. Environ. Change* **2018**, *49*, 56–72. [CrossRef]

35. Leroux, L.; Baron, C.; Zoungrana, B.; Traoré, S.B.; Seen, D.L.; Bégué, A. Crop Monitoring Using Vegetation and Thermal Indices for Yield Estimates: Case Study of a Rainfed Cereal in Semi-Arid West Africa. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 347–362. [CrossRef]

36. Sruthi, S.; Aslam, M.A.M. Agricultural Drought Analysis Using the NDVI and Land Surface Temperature Data; a Case Study of Raichur District. *Aquat. Procedia* **2015**, *4*, 1258–1264. [CrossRef]

37. Hu, T.; Yang, J.; Li, X.; Gong, P. Mapping Urban Land Use by Using Landsat Images and Open Social Data. *Remote Sens.* **2016**, *8*, 151. [CrossRef]

38. Ye, T.; Zhao, N.; Yang, X.; Ouyang, Z.; Liu, X.; Chen, Q.; Hu, K.; Yue, W.; Qi, J.; Li, Z.; et al. Improved population mapping for China using remotely sensed and points-of-interest data within a random forests model. *Sci. Total Environ.* **2019**, *658*, 936–946. [CrossRef] [PubMed]

39. Deng, Y.; Liu, J.; Liu, Y.; Luo, A. Detecting Urban Polycentric Structure from POI Data. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 283. [CrossRef]

40. Council, N.E.a.S.D. *Human Achievement Index Report 2017*; National Economic and Social Development Council: Bangkok, Thailand, 2017.

41. Jitsuchon, S. Small Area Estimation Poverty Map for Thailand. In Proceedings of the SMERU Research Institute and Ford Foundation International Seminar, Jakarta, Indonesia, 1–2 December 2004.

42. Jitsuchon, S.; Richter, K. Thailand's Poverty Maps from Construction to Application. In *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*; Bedi, T., Coudouel, A., Simler, K., Eds.; The World Bank: Washington, DC, USA, 2007.

43. Bedi, T.; Coudouel, A.; Simler, K. *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*; The World Bank: Washington, DC, USA, 2007.

44. Babenko, B.; Hersh, J.; Newhouse, D.; Ramakrishnan, A.; Swartz, T. Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, With an Application in Mexico. *arXiv* **2017**, arXiv:1711.06323.

45. Tingzon, I.; Orden, A.; Sy, S.; Sekara, V.; Weber, I.; Fatehkia, M.; Herranz, M.; Kim, D. Mapping Poverty in the Philippines Using Machine Learning, Satellite Imagery, and Crowd-Sourced Geospatial Information. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, *XLII-4/W19*, 425–431. [CrossRef]

46. Heitmann, S.; Buri, S. *Poverty Estimation with Satellite Imagery at Neighborhood Levels: Results and Lessons for Financial Inclusion from Ghana and Uganda*; International Finance Corporation—The World Bank Group: Washington, DC, USA, 2019; Available online: https://www.ifc.org/wps/wcm/connect/industry_ext_content/ifc_external_corporate_site/financial+institutions/resources/poverty+estimation+with+satellite+imagery+at+neighborhood+levels (accessed on 1 January 2022).

47. Yeh, C.; Perez, A.; Driscoll, A.; Azzari, G.; Tang, Z.; Lobell, D.; Ermon, S.; Burke, M. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nat. Commun.* **2020**, *11*, 2583. [CrossRef]

48. Broxton, P.D.; Zeng, X.; Sulla-Menashe, D.; Troch, P.A. A Global Land Cover Climatology Using MODIS Data. *J. Appl. Meteorol. Climatol.* **2014**, *53*, 1593–1605. [CrossRef]

49. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*; Elsevier: Waltham, MA, USA, 2012. [CrossRef]

50. McBride, L.; Nichols, A. Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning. *World Bank Econ. Rev.* **2018**, *32*, 531–550. [CrossRef]

51. Hu, S.; Ge, Y.; Liu, M.; Ren, Z.; Zhang, X. Village-level poverty identification using machine learning, high-resolution images, and geospatial data. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102694. [CrossRef]

52. Anesti, N.; Kalamara, E.; Kapetanios, G. *Forecasting UK GDP Growth with Large Survey Panels*; Bank of England: London, UK, 2021.

53. Ciaburro, G.; Venkateswaran, B. *Neural Networks with R*; Packt Publishing: Birmingham, UK, 2017.

54. Ripley, B. Feed-Forward Neural Networks and Multinomial Log-Linear Models. 2022. Available online: https://cran.r-project.org/web/packages/nnet/nnet.pdf (accessed on 29 April 2022).

55. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

56. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2009.

57. Kogalur, U. randomForestSRC: Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC). 2022. Available online: https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf (accessed on 29 April 2022).

58. Alsharkawi, A.; Al-Fetyani, M.; Dawas, M.; Saadeh, H.; Alyaman, M. Poverty Classification Using Machine Learning: The Case of Jordan. *Sustainability* **2021**, *13*, 1412. [CrossRef]

59. Vapnik, V.N. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998.

60. Wijaya, D.; Paramita, N.L.P.S.P.; Uluwiyah, A.; Rheza, M.; Zahara, A.; Puspita, D. Estimating city-level poverty rate based on e-commerce data with machine learning. *Electron. Commer. Res.* **2022**, *22*, 195–221. [CrossRef]

61. Meyer, D. Support Vector Machines: The Interface to Libsvm in Package e1071. 2017. Available online: http://web.mit.edu/~{}r/current/arch/i386_linux26/lib/R/library/e1071/doc/svmdoc.pdf (accessed on 29 April 2022).

62. Sachs, J.; Kroll, C.; Lafortune, G.; Fuller, G.; Woelm, F. *Sustainable Development Report 2021*; Cambridge University Press: Cambridge, UK, 2021.

63. Robles Aguilar, G.; Sumner, A. Who are the world's poor? A new profile of global multidimensional poverty. *World Dev.* **2020**, *126*, 104716. [CrossRef]

64. Alesina, A.; Michalopoulos, S.; Papaioannou, E. Ethnic Inequality. *J. Political Econ.* **2016**, *124*, 428–488. [CrossRef]

65. Milanovic, B. *Global Inequality: A New Approach for the Age of Globalization*; Harvard University Press: Cambridge, MA, USA, 2016.

66. Group, W.B. *Poverty and Shared Prosperity 2016: Taking on Inequality*; World Bank Publications: Washington, DC, USA, 2016.

67. Deutsch, J.; Silber, J.; Wan, G.; Zhao, M. Asset indexes and the measurement of poverty, inequality and welfare in Southeast Asia. *J. Asian Econ.* **2020**, *70*, 101220. [CrossRef]

68. Wan, G.; Wang, C.; Zhang, X. The Poverty-Growth-Inequality Triangle: Asia 1960s to 2010s. *Soc. Indic. Res.* **2021**, *153*, 795–822. [CrossRef]

69. Kudo, T.; Satoru, K. *Two-Polar Growth Strategy in Myanmar: Seeking "High" and "Balanced" Development*; Institute of Developing Economies—Japan External Trade Organization (IDE-JETRO): Chiba, Japan, 2012; Available online: https://www.ide.go.jp/library/English/Publish/Reports/Brc/PolicyReview/pdf/08.pdf (accessed on 1 January 2022).

70. ADB. *Asian Development Bank Sustainability Report 2015: Investing for an Asia and the Pacific Free of Poverty*; Asian Development Bank: Manila, Philippines, 2015.

71. Puttanapong, N. *Monocentric Growth and Productivity Spillover in Thailand*; Institute of Developing Economies—Japan External Trade Organization (IDE-JETRO) (Bangkok Office): Bangkok, Thailand, 2018; Available online: https://www.ide.go.jp/library/English/Publish/Reports/Brc/pdf/23_03.pdf (accessed on 1 January 2022).

72. Guo, H. Big Earth data: A new frontier in Earth and information sciences. *Big Earth Data* **2017**, *1*, 4–20. [CrossRef]

73. Lee, J.-G.; Kang, M. Geospatial Big Data: Challenges and Opportunities. *Big Data Res.* **2015**, *2*, 74–81. [CrossRef]

74. Kansakar, P.; Hossain, F. A review of applications of satellite earth observation data for global societal benefit and stewardship of planet earth. *Space Policy* **2016**, *36*, 46–54. [CrossRef]

75. Ivan, K.; Holobâcă, I.-H.; Benedek, J.; Török, I. Potential of Night-Time Lights to Measure Regional Inequality. *Remote Sens.* **2020**, *12*, 33. [CrossRef]

76. Kemper, T.; Pesaresi, M.; Ehrlich, D.; Schiavina, M. *Detecting Spatial Pattern of Inequalities from Remote Sensing towards Mapping of Deprived Communities and Poverty*; European Union: Luxembourg, 2018. [CrossRef]

77. Galimberti, J.; Pichler, S.; Pleninger, R. *Measuring Inequality using Geospatial Data*; Auckland University of Technology, Department of Economics: Auckland, New Zealand, 2020.

78. Mirza, M.U.; Xu, C.; Bavel Bas, v.; van Nes Egbert, H.; Scheffer, M. Global inequality remotely sensed. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e1919913118. [CrossRef] [PubMed]

79. Blumenstock Joshua, E. Fighting poverty with data. *Science* **2016**, *353*, 753–754. [CrossRef] [PubMed]

80. Duque, J.C.; Patino, J.E.; Ruiz, L.A.; Pardo-Pascual, J.E. Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. *Landsc. Urban Plan.* **2015**, *135*, 11–21. [CrossRef]

81. Klemens, B.; Coppola, A.; Shron, M. *Estimating Local Poverty Measures Using Satellite Images: A Pilot Application to Central America*; The World Bank: Washington, DC, USA, 2015.

82. Watmough, G.R.; Atkinson, P.M.; Hutton, C.W. Exploring the links between census and environment using remotely sensed satellite sensor imagery. *J. Land Use Sci.* **2013**, *8*, 284–303. [CrossRef]

83. Watmough, G.R.; Atkinson, P.M.; Saikia, A.; Hutton, C.W. Understanding the Evidence Base for Poverty–Environment Relationships using Remotely Sensed Satellite Data: An Example from Assam, India. *World Dev.* **2016**, *78*, 188–203. [CrossRef]

84. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]

85. Foody, G.; Fritz, S.; Fonte, C.; Bastin, L.; Olteanu Raimond, A.-M.; Mooney, P.; See, L.; Antoniou, V.; Liu, H.-Y.; Minghini, M.; et al. Mapping and the Citizen Sensor. In *Mapping and the Citizen Sensor*; Ubiquity Press: London, UK, 2017; pp. 1–12.

86. Goodchild, M. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *Int. J. Spat. Data Infrastruct. Res.* **2007**, *2*, 24–32.

87. Gao, S.; Janowicz, K.; Couclelis, H. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Trans. GIS* **2017**, *21*, 446–467. [CrossRef]

88. Hersh, J.; Engstrom, R.; Mann, M. Open data for algorithms: Mapping poverty in Belize using open satellite derived features and machine learning. *Inf. Technol. Dev.* **2021**, *27*, 263–292. [CrossRef]

89. Tian, F.; Wu, B.; Zeng, H.; Watmough, G.R.; Zhang, M.; Li, Y. Detecting the linkage between arable land use and poverty using machine learning methods at global perspective. *Geogr. Sustain.* **2022**, *3*, 7–20. [CrossRef]

90. Zhao, X.; Yu, B.; Liu, Y.; Chen, Z.; Li, Q.; Wang, C.; Wu, J. Estimation of Poverty Using Random Forest Regression with Multi-Source Data: A Case Study in Bangladesh. *Remote Sens.* **2019**, *11*, 375. [CrossRef]

91. Browne, C.; Matteson, D.S.; McBride, L.; Hu, L.; Liu, Y.; Sun, Y.; Wen, J.; Barrett, C.B. Multivariate random forest prediction of poverty and malnutrition prevalence. *PLoS ONE* **2021**, *16*, e0255519. [CrossRef] [PubMed]

92. Xu, Y.; Mo, Y.; Zhu, S. Poverty Mapping in the Dian-Gui-Qian Contiguous Extremely Poor Area of Southwest China Based on Multi-Source Geospatial Data. *Sustainability* **2021**, *13*, 8717. [CrossRef]

93. Sohnesen, T.; Stender, N. Is Random Forest a Superior Methodology for Predicting Poverty? An Empirical Assessment: Predicting Poverty. *Poverty Public Policy* **2017**, *9*, 118–133. [CrossRef]

94. Liu, M.; Hu, S.; Ge, Y.; Heuvelink, G.B.M.; Ren, Z.; Huang, X. Using multiple linear regression and random forests to identify spatial poverty determinants in rural China. *Spat. Stat.* **2021**, *42*, 100461. [CrossRef]

95. Wang, S.; Aggarwal, C.; Liu, H. Random-Forest Inspired Neural Networks. *ACM Trans. Intell. Syst. Technol.* **2018**, *9*, 69. [CrossRef]

96. Yu, B.; Shi, K.; Hu, Y.; Huang, C.; Chen, Z.; Wu, J. Poverty Evaluation Using NPP-VIIRS Nighttime Light Composite Data at the County Level in China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 1217–1229. [CrossRef]

97. Engstrom, R.; Hersh, J.; Newhouse, D. *Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being*; Policy Research Working Paper 8284; The World Bank Group: Washington, DC, USA, 2017. [CrossRef]

98. Wang, Z.; Han, Q.; de Vries, B. Land Use/Land Cover and Accessibility: Implications of the Correlations for Land Use and Transport Planning. *Appl. Spat. Anal. Policy* **2019**, *12*, 923–940. [CrossRef]

99. Okwi Paul, O.; Ndeng'e, G.; Kristjanson, P.; Arunga, M.; Notenbaert, A.; Omolo, A.; Henninger, N.; Benson, T.; Kariuki, P.; Owuor, J. Spatial determinants of poverty in rural Kenya. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16769–16774. [CrossRef]

100. Watmough Gary, R.; Marcinko Charlotte, L.J.; Sullivan, C.; Tschirhart, K.; Mutuo Patrick, K.; Palm Cheryl, A.; Svenning, J.-C. Socioecologically informed use of remote sensing data to predict rural household poverty. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 1213–1218. [CrossRef] [PubMed]

101. Vakis, R.N.; Rigolini, J.; Lucchetti, L. *Left Behind: Chronic Poverty in Latin America and the Caribbean*; The World Bank: Washington, DC, USA, 2016.

102. Cook, S.; Pincus, J. Poverty, Inequality and Social Protection in Southeast Asia: An Introduction. *J. Southeast Asian Econ.* **2014**, *31*, 1–17. [CrossRef]

103. Sunderlin, W.D.; Dewi, S.; Puntodewo, A. *Poverty and Forests Multi-Country Analysis of Spatial Association and Proposed Policy Solutions*; Center for International Forestry Research: Bogor, Indonesia, 2007.

104. Yang, Y.; de Sherbinin, A.; Liu, Y. China's poverty alleviation resettlement: Progress, problems and solutions. *Habitat Int.* **2020**, *98*, 102135. [CrossRef]

105. Pokhriyal, N.; Jacques Damien, C. Combining disparate data sources for improved poverty prediction and mapping. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E9783–E9792. [CrossRef] [PubMed]

106. Steele, J.E.; Sundsøy, P.R.; Pezzulo, C.; Alegana, V.A.; Bird, T.J.; Blumenstock, J.; Bjelland, J.; Engø-Monsen, K.; de Montjoye, Y.-A.; Iqbal, A.M.; et al. Mapping poverty using mobile phone and satellite data. *J. R. Soc. Interface* **2017**, *14*, 20160690. [CrossRef] [PubMed]

107. Burke, M.; Driscoll, A.; Lobell David, B.; Ermon, S. Using satellite imagery to understand and promote sustainable development. *Science* **2021**, *371*, eabe8628. [CrossRef]

108. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [CrossRef]

109. Engstrom, R.; Pavelesku, D.; Tanaka, T.; Wambile, A. Mapping Poverty and Slums Using Multiple Methodologies in Accra, Ghana. In Proceedings of the 2019 Joint Urban Remote Sensing Event (JURSE), Vannes, France, 22–24 May 2019; pp. 1–4.

110. Wang, J.; Kuffer, M.; Pfeffer, K. The role of spatial heterogeneity in detecting urban slums. *Comput. Environ. Urban Syst.* **2019**, *73*, 95–107. [CrossRef]

111. Müller, I.; Taubenböck, H.; Kuffer, M.; Wurm, M. Misperceptions of Predominant Slum Locations? Spatial Analysis of Slum Locations in Terms of Topography Based on Earth Observation Data. *Remote Sens.* **2020**, *12*, 2474. [CrossRef]