Nakhara: Journal of Environmental Design and Planning (2024)
Volume 23(3), Article 421
DOI: https://doi.org/10.54028/NJ202423421
Article Type: Research Article

Predicting Household Expenditure Using Machine Learning Techniques: A Case of Cambodia

Nattapong Puttanapong[1*], Siphat Lim[2]

[1] Faculty of Economics, Thammasat University, Thailand
[2] CamEd Business School, Phnom Penh, Cambodia

*Corresponding author e-mail: nattapong@econ.tu.ac.th

Received 2024-03-18; Revised 2024-10-18; Accepted 2024-11-21

ABSTRACT

This study aimed to predict household expenditure using a combination of survey and geospatial data. A web-based application operating on the Google Earth Engine platform has been specifically developed for this research, providing a set of satellite-based indicators. These data were spatially averaged at the district level and integrated with household nonfood expenditures, a proxy of socioeconomic conditions, derived from the World Bank's 2019 Living Standards Measurement Study (LSMS). Four machine learning algorithms were applied. By using root mean square error as the goodness-of-fit criterion, a random forest algorithm yielded the highest forecasting precision, followed by support vector machine, neural network, and generalized least squares. In addition, variable importance and minimal depth analyses were conducted, indicating that the geospatial indicators have moderate contributive powers in predicting socioeconomic conditions. Conversely, the predictive powers of variables derived from the LSMS were mixed. Some asset ownership yielded a high explanatory power, whereas some were minimal. The attained results suggest future development aimed at enhancing accuracy. Additionally, the findings revealed an association between economic activity density and household expenditure, recommending regional development promotion through urbanization and transition from agriculture to other economic sectors.

Keywords: Cambodia, household expenditure, google earth engine, machine learning, prediction

INTRODUCTION

Contextual Overview of Cambodian Poverty

As of 2021, Cambodia's population is approximately 16.59 million, with 38% residing in urban areas and 62% in rural locations. Urbanization has seen a significant uptick since 2015, rising 17% to constitute 38% of the population. Concurrently, the rural population has correspondingly declined by 17% (National Institute of Statistics, 2021). Economic metrics also demonstrate growth; the GDP per capita has more than doubled from $741 in 2009 to $1,619 in 2021 (Asian Development Bank, 2022).

As shown in Table 1, along with the provincial map detected in Figure 1, poverty rates have also improved markedly. In 2009, about 4.7 million people were impoverished, the majority residing in rural areas. A decade later, this figure dropped by 40.43% to 2.8 million, lifting nearly 1.9 million Cambodians above the poverty threshold. This improvement coincides with an average economic growth rate of 7.7% over two decades. Projections indicate that Cambodia will transition from a lower-middle-income to an upper-middle-income country by 2030 (World Bank, 2022).

In addition to the poverty rates, Human Development Index (HDI) can provide a comprehensive view summarizing the status of healthcare, education, and individual income. Particularly, Cambodia's HDI has experienced a slight but consistent decline from 0.598 in 2019 to 0.593 in 2021 (United Nations Development Programme, 2022).

Cambodia has made significant progress in economic development, poverty reduction, and social indices. Nuanced metrics such as the HDI indicate that room for improvement remains, emphasizing the need for a multifaceted approach to assessing well-being and development. Essentially, the innovation of geographic information system (GIS) analysis and data availability, as introduced in the subsequent sections, have enabled the investigation at the district level, enhancing spatial accuracy and predictive capabilities of poverty analysis in the case of Cambodia.

Table 1

Poverty Rate (%) by Cities/Provinces in 2011 and 2020

No.	Provinces	2011	2020
1	Koh Kong	23	28
2	Mondul Kiri	26	27
3	Prey Veng	17	23
4	Kep	17	23
5	Battambang	20	23
6	Preah Sihanouk	12	21
7	Kratie	23	21
8	Pailin	18	20
9	Pursat	22	19
10	Stung Treng	25	19
11	Kampong Chhnang	20	18
12	Svay Rieng	14	18
13	Preah Vihear	25	18
14	Ratanak Kiri	29	18

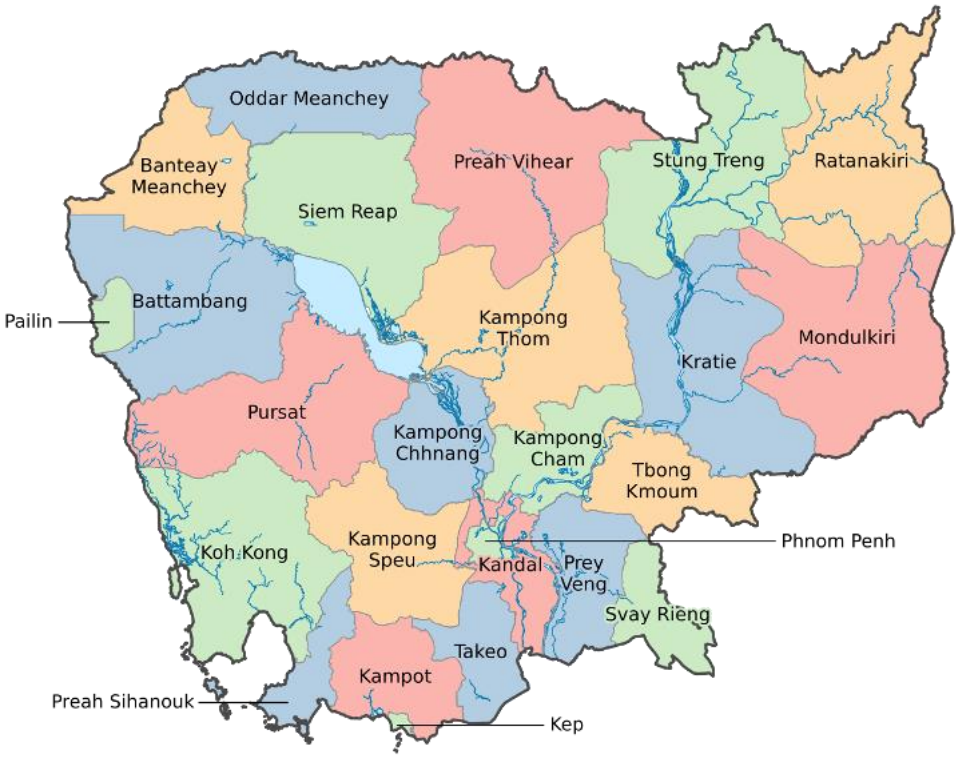
Table 1 (Continued)

No.	Provinces	2011	2020
15	Kampot	20	17
16	Kampong Thom	23	17
17	Kampong Speu	18	16
18	Tbong Khmum	NA	16
19	Takeo	16	15
20	Kampong Cham	19	14
21	Siem Reap	21	13
22	Banteay Meanchey	21	11
23	Kandal	15	10
24	Otdar Meanchey	NA	10
25	Phnom Penh	3	7

Note. From "Educational Administration: Theory and Practice," by R. Eng, and S. Lim, 2024, *The economic development and level of poverty in Cambodia, 30*(6), 3693–3701 (https://doi.org/10.53555/kuey.v30i6.5806). Copyright 2024 by Eng & Lim.

Figure 1

Cambodia's Provincial Map



Note. From *Cambodia's Provincial Map, by* Wikimedia Commons, 2020, Wikimedia Commons (https://commons.wikimedia.org/wiki/File:Provincial_Boundaries_in_Cambodia.svg). CC-BY-SA-3.0.

Innovative Frameworks for Developmental Surveillance through Spatial Data

With the evolution of information technology, open geographical data and open-source software tools have become significantly more accessible. Many online platforms (e.g., Google Earth Engine) have been providing public access to satellite data, as well as simplifying data extraction and computational tasks. These platforms also facilitate the creation of task-specific web applications, enriching the tailored development to serve specific needs. The rise of open software tools offers users, particularly those in developing nations, an equitable opportunity to engage in spatial and AI-driven computations without financial constraints.

In view of nations such as Cambodia, where a predominant portion of the populace is engaged in agriculture and resides in rural locales, the innovative methodologies enable new opportunities to monitor regional development. This study, therefore, aims to introduce a new analytical framework integrating satellite-based indicators, survey data, and machine learning to examine the household socioeconomic condition proxied by non-food expenditure.

Technically, in this study, a web application powered by the Google Earth Engine platform was developed and launched to facilitate satellite data acquisition. To investigate the association between the household expenditure and explanatory variables sourced from satellite indicators and the ground survey, our research methodology employed four machine learning algorithms (i.e., generalized least squares (GLS), neural network (NN), Random Forest (RF), and support vector regression (SVR)) using the R software suite. Subsequent analyses were conducted using two feature analysis methods: variable importance (VIMP) and minimal depth, allowing the prioritization of explanatory variables in predicting household non-food expenditure.

The remainder of this paper is structured as follows. Section 2 surveys related literature. Sections 3 and 4 detail data sources and methodological approaches, respectively. Section 5 presents findings derived from machine learning and feature analysis. Finally, Section 6 summarizes key insights and suggests directives for future exploration and improvement.

LITERATURE REVIEW

Evolution of Spatial Analysis of Poverty

As comprehensively reviewed by Hall et al. (2023), spatial examinations of poverty traditionally relied on face-to-face household surveys. These methods, though established, face challenges, especially when the majority of impoverished populations inhabit remote or rugged terrains. Such locales can lead to increased costs, errors, and scalability challenges, often resulting in sporadic updates and restricted spatial reach (Burke et al., 2021; Puttanapong et al., 2022). In addition, the multifaceted nature of poverty has necessitated the formulation of specific indices, reinforcing that a single measure cannot sufficiently represent the breadth of poverty.

Recent shifts in poverty analysis underscore the importance of granularity, with emphasis on district, household, and individual levels. Such a nuanced approach demands the assimilation of advanced data sources and techniques (Blumenstock, 2016). Innovations in this domain encompass high-resolution satellite imagery (Head et al., 2017; Jean et al., 2016), mobile phone metadata (Aiken et al., 2022; Blumenstock et al., 2015), and digital footprints, such as online search trends and social media behaviors (Choi & Varian, 2012; Fatehkia et al., 2020; Llorente et al., 2015). The emergence and integration of these data sources can be attributed to technological progress, specifically the proliferation of big data and the enhancement of machine learning algorithms (Pokhriyal & Jacques, 2017; Steele et al., 2017).

Such novel techniques are instrumental in identifying areas and communities grappling with poverty, a critical component for targeted resource allocation in poverty mitigation efforts, given the complex dimensions of poverty (Aiken et al., 2022; Blumenstock et al., 2021; Erenstein et al., 2010; Zhou & Liu, 2022).

Geospatial Approaches to Poverty Analysis in Cambodia: A Historical Overview

Initiated in 1997 as a collaborative effort between the National Institute of Statistics and international agencies, such as UNDP, World Bank, and SIDA, the Cambodia Socioeconomic Survey (CSES) aims to assess living standards across diverse geographic segments. Covering nine thematic areas from demographics to household assets, CSES data serve various stakeholders, including NGOs and government bodies (National Institute of Statistics, 1997). Since 2008, the CSES has been conducted annually, with the latest one completed in 2021 (National Institute of Statistics, 2021).

After the Khmer Rouge era, Cambodia's first national census took place in 1998, registering 11.44 million individuals. Subsequent censuses were in 2008 and 2019, covering a comprehensive set of demographic and socioeconomic indicators. Some populations in conflict-affected areas were omitted from the 1998 census, affecting the total count (Huguet et al., 2000).

In 2002, a pioneering poverty measurement technique was employed in a study by the Ministry of Planning, the United Nations World Food Program, and the World Bank, which utilized community-level data from multiple sources, including CSES and the 1998 census (Elbers et al., 2002). According to the World Bank report, the national poverty rate in 2019/20 was 17.8 percent (World Bank, 2022).

Nutritional mapping techniques were introduced in 2003 in a study involving the World Food Program and the Ministry of Health. This technique, however, indicated a weak correlation between poverty and malnutrition in children, a finding later addressed through methodological refinements (Fujii, 2007; Fujii, 2010).

Research Gaps

Although traditional methods such as CSES are robust, their limitations include high costs and infrequent data collection, impeding the timely evaluation of poverty alleviation efforts. Smaller surveys offer more frequent data but also come with limitations, such as cost and time constraints. With advancements in remote sensing and geospatial technology, modern poverty mapping techniques provide cost-effective, timely data, making them increasingly relevant for nations such as Cambodia. These innovations pave the way for developing an alternative approach to integrating multiple data sources for predicting and monitoring socioeconomic status, thereby better-informing policy interventions. To our knowledge, this exists only in the United Nations Development Programme's initiative[1] exploring the application of big data and AI in mapping poverty in Cambodia. Therefore, this study aimed to bridge the knowledge gap by introducing the analytical framework and applying machine learning methods to a combination of remote sensing and survey data.

METHODOLOGY

Data

Following the sets of remote-sensing data applied in existing literature Ayush et al., 2021, Engstrom et al., 2017, Jean et al., 2016 and Yeh et al., 2020, this study utilizes two sets of data. The first was derived from the official nationwide survey, namely, LSMS Plus (LSMS+). The second dataset consists of satellite indicators sourced from Google Earth

[1] https://cambodiapovertymapping.sig-gis.com/en/about/

Engine. To facilitate the data extraction process, a tailor-made application has been developed for this study. Technical attributes of each dataset are outlined in subsequent subsections.

Survey Data

The National Institute of Statistics conducted the 2019–2020 Cambodia LSMS+ Survey in collaboration with the World Bank LSMS+ program (National Institute of Statistics, 2019). This national survey targeted private interviews with every adult (aged 18 and above) in the selected households. The primary emphasis of the data collection was on (i) asset ownership, (ii) employment status, and (iii) nonfarm business activities. Subsequently, this study utilized household nonfood expenditure as the dependent variable. In addition, other characteristics of each household (such as ownership of assets and dwelling) obtained from LSMS+ Survey were merged with geospatial indicators derived from satellite data, yielding 1,257 observations representing nationwide samplings of households. Details of these satellite-based indicators are explained in the following section.

Satellite Data

In this study, all remote-sensing data were sourced from Google Earth Engine, a public cloud service. Google Earth Engine merges cloud storage, offering an array of satellite data collections, with a computing platform designed for satellite data analysis. Below are the technical specifications for each dataset.

NDVI

This study utilized vegetation index data derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors aboard the Terra and Aqua satellites. These sensors are adept at detecting various terrestrial features, incorporating surface and ground temperatures, clouds, ocean hues, and biogeochemical components. Technically, this index is computed by using data obtained from two sensors, as shown as follows:

$$NDVI = \frac{(NIR-Red)}{(NIR+Red)} \quad (1)$$

where *NIR* represents near-infrared reflectance, and *Red* indicates reflectance in the red spectrum. This equation mimics the absorption and reflectance properties of chlorophyll in vegetation, leading to the green appearance of leaves as perceived by the human eye.

The MODIS sensor has 36 spectral bands, spanning wavelengths from 0.4 to 14.4 μm, with spatial resolutions of 250 m (Bands 1–2), 500 m (Bands 3–7), and 1 km (Bands 8–36). The NDVI value lies between −1 and 1: nearing 1 for dense vegetation, around 0 for unhealthy vegetation, and close to −1 for water surfaces. The GIS-based data of NDVI are depicted in Figure 2.

Several studies have featured the utility of the Normalized Difference Vegetation Index (NDVI) as a vegetation index, enabling analysis of vegetative coverage. For instance, positive correlations between GDP and NDVI were highlighted by Jin et al., (2008), Chen et al., (2022), and Guo et al. (2021). Another research pointed to a link between socioeconomic conditions and NDVI (Li et al., 2015).

Research has also verified the NDVI's relationship with poverty. For Kenya and China, a higher poverty rate corresponded to lower NDVI values (Kristjanson et al., 2005; Shi et al. 2020). A similar negative association was observed in Tanzania (Morikawa, 2014). However, some studies have noted bidirectional ties between rural impoverishment and environmental determinants, including the NDVI (Bhattacharya & Innes, 2006).

Normalized Difference Water Index (NDWI)

Following the computational technique of the NDVI, Gao (1996) and McFeeters (1996) introduced the NDWI to monitor water bodies. The NDWI is calculated as follows:

$$NDWI = \frac{(NIR-SWIR)}{(NIR+SWIR)} \quad (2)$$

where *NIR* stands for near-infrared reflectance, and *SIWR* represents short-wave infrared reflectance. Similar to the NDVI, the NDWI value ranges from −1 to +1. A value exceeding 0.5 typically signifies the presence of water bodies. Figure 3 exhibits the 2019 NDWI map of Cambodia.

Normalized Difference Drought Index (NDDI)

Gu et al. (2007) developed the NDDI using a similar mathematical framework to map and track drought features. The NDDI is calculated as

$$NDDI = \frac{(NDVI-NDWI)}{(NDVI+NDWI)} \quad (3)$$

The NDDI values begin at 0 for no drought, with values exceeding 1.0 indicating intense drought scenarios.

LST

Using Bands 20–23 and 30–31, MODIS sensors measure global surface temperatures and thermal radiation, encompassing bandwidths of 3.66–4.080 nm and 10.780–11.280 μm. Data, updated daily at a 1 km resolution, is retrievable from the Google Earth Engine from March 5, 2000 onward. This study employed average daytime LST and nighttime LST. Figure 4 illustrates the 2019 spatial distribution of average LST in Cambodia.

A strong link has been identified between socioeconomic factors and LST, with areas of high industrial and commercial activity typically showing higher LST values in contrast to those mainly agricultural or forested (Huang et al., 2011). A US-focused study detected a positive trend between LST and increasing per capita income (Buyantuyev & Wu, 2010). In emerging economies, variables, including infrastructural developments, industrial progression, and demographic growth, have influenced LST variations (Dissanayake et al., 2019; Ruthirako et al., 2015).

In addition, the mutual relationship between LST and NDVI values has been the subject of numerous studies (Liaqut et al., 2019; Li et al., 2014; Wan Mohd Jaafar et al., 2020; Sruthi & Aslam, 2015; Youneszadeh et al., 2015). These studies have consistently identified that vegetated areas maintain cooler temperatures than their urban or industrial counterparts.

Rainfall Data

This study utilized data sourced from the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS), which integrates satellite-based rainfall measurements with data from rain gauge stations. Accessible via Google Earth Engine, the CHIRPS dataset extends back to 1981 and offers a spatial granularity of 0.05 arc degrees, equivalent to around 110 m per pixel. Figure 5 shows the geographical distribution of cumulative rainfall within Cambodia in 2019.

A notable study by Barrios et al. (2010) identified a positive link between rainfall and real GDP per capita in African nations. Conversely, in developed nations, rainfall typically negatively influences economic activities, especially within the service sector. Arezki and Brückner (2012) integrated rainfall data with historical financial transactions and observed a positive correlation in nations where the financial sector holds a minor GDP share, yet a negative association emerged in countries with a dominant financial sector. An interesting finding is the concave relationship between GDP growth and rainfall in developing regions (Damania et al., 2020).

Rainfall variability's connection with inequality has also garnered attention. Several studies have highlighted a negative association (Brown & Lall, 2006; Richardson, 2007). Investigations using agricultural yields as a proxy for indirect inequality—such as those from Ethiopia (Thiede, 2014), Nigeria (Amare et al., 2018), and India (Gilmont et al., 2018)—all reported negative correlations.

Population

Annual population data were acquired from the WorldPop Global Project, an open-access

resource for population distribution datasets. Utilizing a machine learning algorithm and various geospatial layers, the WorldPop creates detailed population spatial distributions with a granularity of 100 m. Figure 6 exhibits this population map of Cambodia in 2019.

Leaf Area Index (LAI)

As comprehensively surveyed by Zheng and Moskal (2009), the LAI is a biophysical metric of vegetation. Conventionally, it is scientifically defined as the one-sided green leaf area per unit of ground area. This metric is especially significant in monitoring forest conditions and cultivation activities. The LAI can be employed with other environmental indicators to calculate crop yield. Therefore, the LAI is particularly an important determinant of income for farmers in the study areas (Mourad et al., 2020). Figure 7 shows the LAI map of Cambodia in 2019.

Gross Primary Productivity (GPP)

GPP is a metric quantifying the rate at which solar energy is converted into sugar molecules via photosynthesis, measured per unit area per unit time. GPP is a significant indicator of vegetative density and efficacy in a specific locale, particularly relevant to light energy capture and photosynthetic activity. In addition to its role in environmental science, GPP has practical applications in agriculture, specifically in the estimation of crop yields. Thus, GPP can be implemented as the satellite-based determinant of farmers' incomes (Li et al., 2022; Liu et al., 2022). The spatial distribution of GPP in Cambodia is shown in Figure 8.

Evapotranspiration (ET)

ET represents the aggregate of mechanisms through which water is transferred from the terrestrial surface to the atmosphere, encompassing evaporation and transpiration processes. ET's MODIS remote sensing data are estimated using an algorithm based on the Penman–Monteith equation, incorporating daily meteorological and environmental data, including vegetation characteristics, albedo, and land cover categorizations. Similar to GPP, ET can be integrated with other variables to predict crop yield, which is the proxy for farmers' economic livelihoods (Pandit et al., 2022; Mulovhedzi et al., 2020).

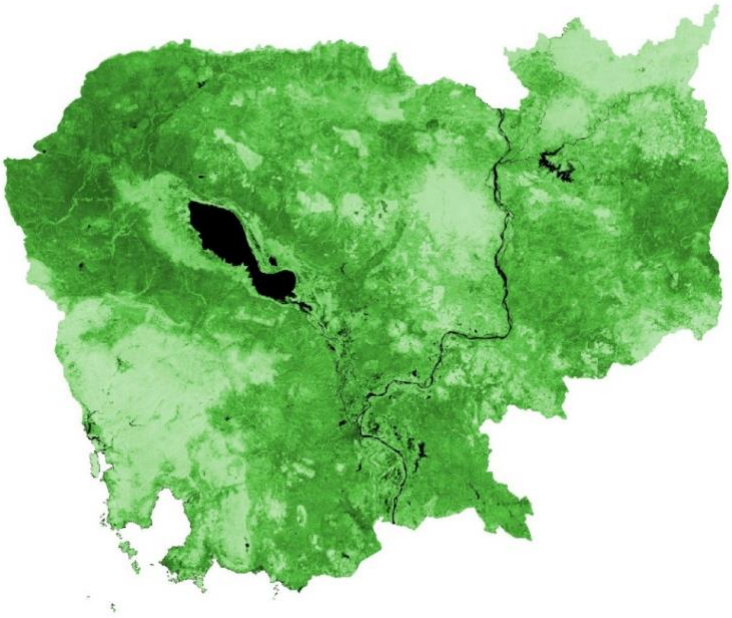
Urban Area

Global land cover types were derived using supervised classification techniques on Terra and Aqua MODIS reflectance data. This approach produces an annual global map detailing 17 distinct land use categories. For this study, Type 14, representing urban areas, was specifically extracted. Figure 9 illustrates the geographical distribution of land use in Cambodia in 2019.

All key technical specifications of satellite indicators are summarized in Table 2.

Figure 2

NDVI Map of Cambodia (Average Value of 2019)



Note. Dark color represents high value. Adapted from *NDVI Map of Cambodia*, by Google Earth Engine, 2019. Google Earth Engine. (https://code.earthengine.google.com/88b38d332a95a32330e39d26edc44edb). Copyright 2019 by Google LLC.

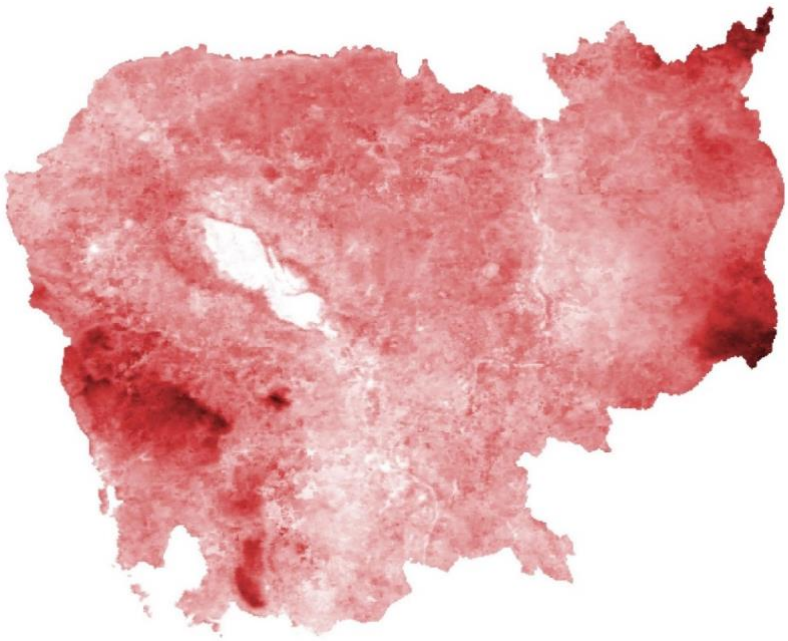
Figure 3

NDWI Map of Cambodia (Average Value of 2019)

Note. Blue color represents water bodies. Adapted from *NDWI Map of Cambodia*, by Google Earth Engine, 2019. Google Earth Engine. (https://code.earthengine.google.com/125e1b8a896197d87bec0c6ab5d941f3). Copyright 2019 by Google LLC.

Figure 4

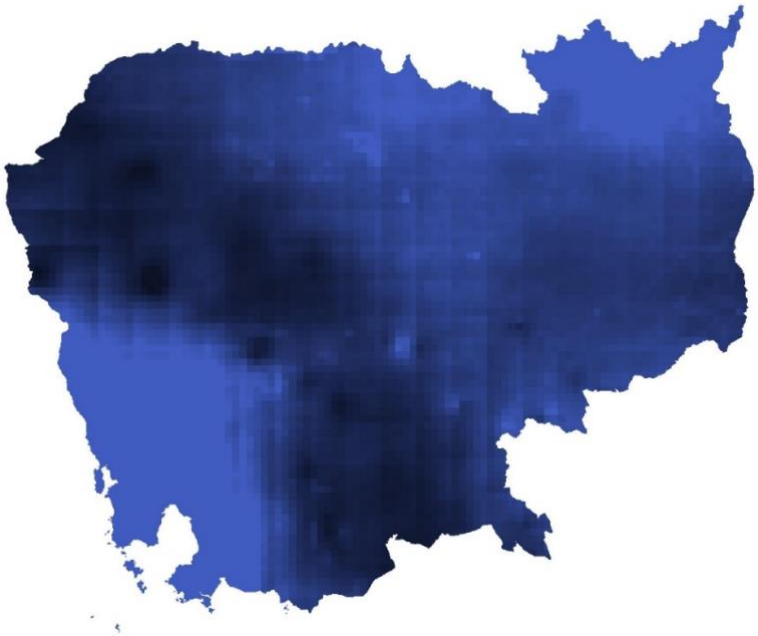
LST Map of Cambodia (Average Value of 2019)



Note. Dark color represents high temperature. Adapted from *LST Map of Cambodia*, by Google Earth Engine, 2019. Google Earth Engine. (https://code.earthengine.google.com/cf0873d0f938d697b898c3f2febc8d92). Copyright 2019 by Google LLC.

Figure 5

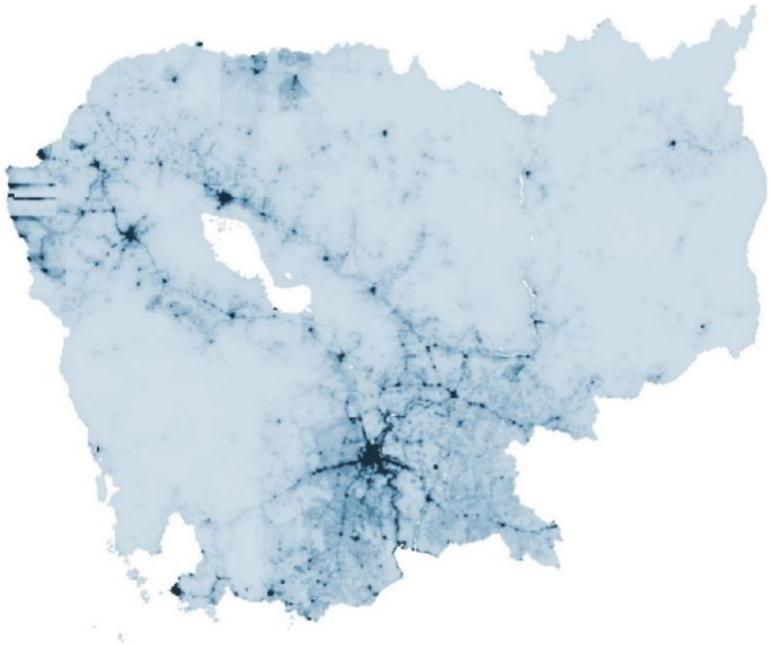
Rainfall Map of Cambodia (Average Value of 2019)



Note. Dark color represents high value. Adapted from *Rainfall Map of Cambodia*, by Google Earth Engine, 2019. Google Earth Engine. (https://code.earthengine.google.com/7da4894e7f83f1f26526a0ffb0999218). Copyright 2019 by Google LLC.

Figure 6

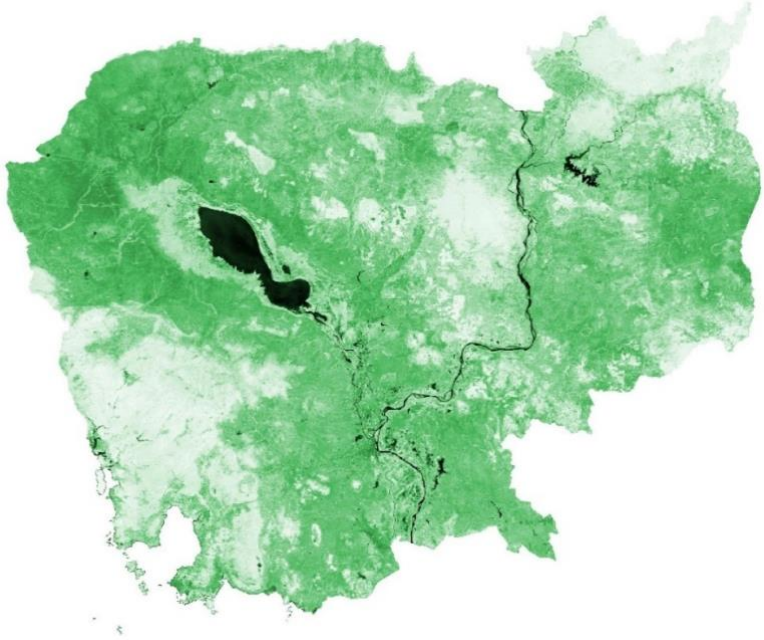
Population Map of Cambodia (2019)



Note. Dark color represents high population density. Adapted from *Population Map of Cambodia*, by Google Earth Engine, 2019. Google Earth Engine. (https://code.earthengine.google.com/89e180bc9a64dae03067f3b047e70d61). Copyright 2019 by Google LLC.

Figure 7

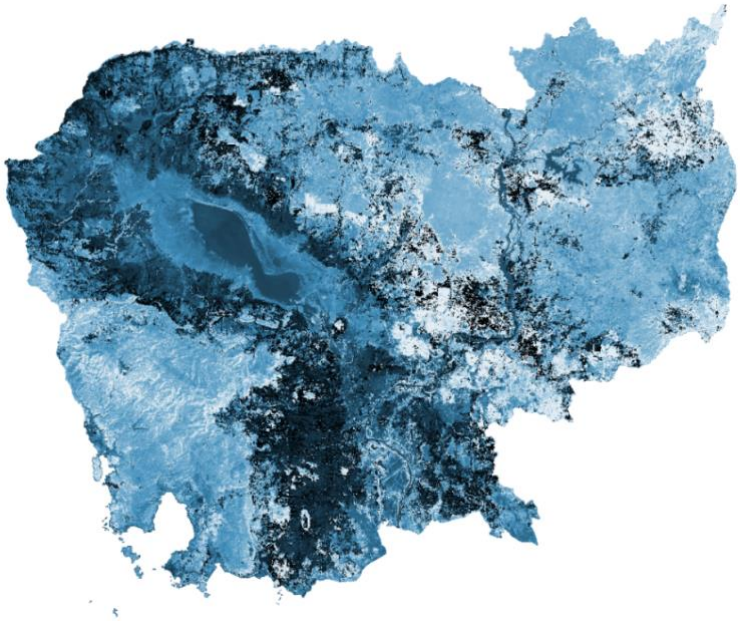
LAI Map of Cambodia (2019)



Note. Dark green color represents high density of LAI. Adapted from *LAI Map of Cambodia*, by Google Earth Engine, 2019. Google Earth Engine. (https://code.earthengine.google.com/e67af9da1e4eea8a579264b8e3cb89ac). Copyright 2019 by Google LLC.

Figure 8

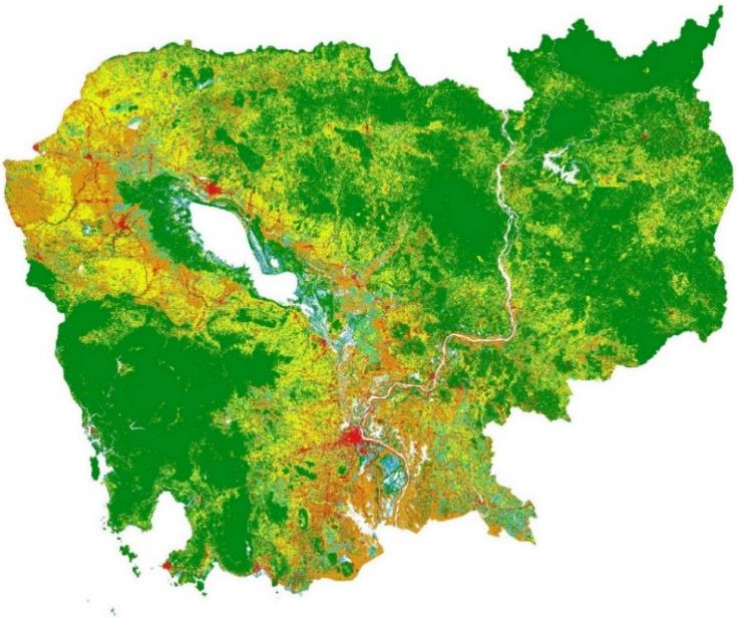
GPP Map of Cambodia (2019)



Note. Dark color represents high GPP. Adapted from *GPP Map of Cambodia*, by Google Earth Engine, 2019. Google Earth Engine. (https://code.earthengine.google.com/3536b4517cd3148f55726f76ae689d9d). Copyright 2019 by Google LLC.

Figure 9

Land-Use Map of Cambodia (2019)



Note. Red color represents the urban area, yellow color indicates the cropland, orange color identifies the flooded vegetation zone and green color denotes the forest. Adapted from *Land-Use Map of Cambodia*, by Google Earth Engine. Google Earth (https://code.earthengine.google.com/88b38d332a95a32330e39d26edc44edb). Copyright 2019 by Google LLC.

Table 2

Main Specifications of Geospatial Data

Indicator	Satellite/Dataset	Resolution	Frequency	Technical reference
Normalized Difference Drought Index (NDDI)	Terra MODIS	500 m	8 days	https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD09A1
Normalized Difference Vegetation Index (NDVI)	Terra MODIS	500 m	8 days	https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD09A1
Normalized Difference Water Index (NDWI)	Terra MODIS	500 m	8 days	https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD09A1
Land Surface Temperature (Daytime)	Terra MODIS	1 km	8 days	https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD11A2
Land Surface Temperature (Nighttime)	Terra MODIS	1 km	8 days	https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD11A2
URBAN (urban area)	MODIS Land Cover Type	500 m	Annual	https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MCD12Q1
Rainfall	CHIRPS	~5 km	Daily	https://developers.google.com/earth-engine/datasets/catalog/UCSB-CHG_CHIRPS_DAILY
Population	WorldPop	~100 m	Annual	https://developers.google.com/earth-engine/datasets/catalog/WorldPop_GP_100m_pop
Leaf Area Index (LAI)	Terra MODIS	500 m	8 days	https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_HARMONIZED
Gross Primary Productivity (GPP)	Terra MODIS	500 m	8 days	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD17A2H
Evapotranspiration (ET)	Terra MODIS	500 m	8 days	https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD16A2

Note. From *Main Specifications of Geospatial Data*, by Google Earth Engine, 2019. Google Earth. Copyright 2019 by Google LLC.

Developing a Web-Based Application on Google Earth Engine

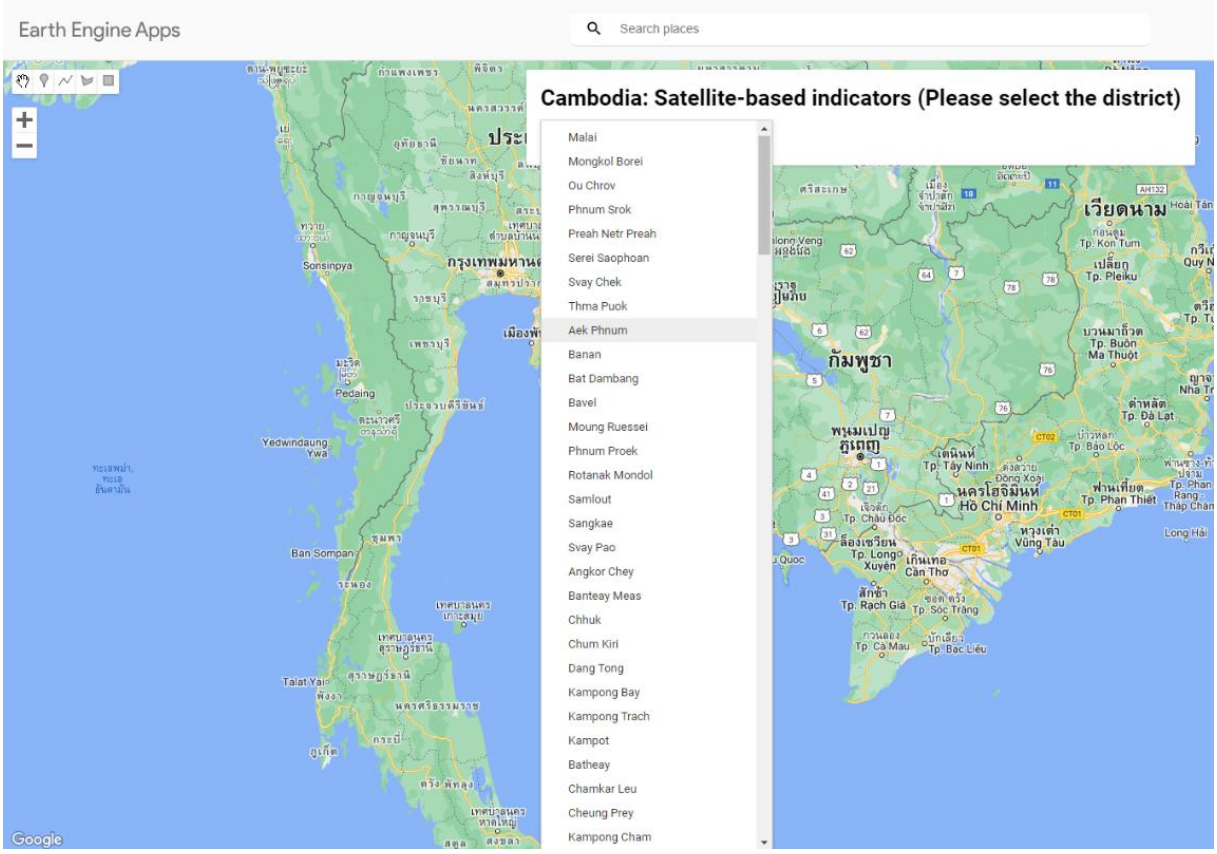
We developed a web application tailored for satellite data analysis using the combined power of Google Earth Engine's cloud services. This application, central to our study, facilitated the transformation of geospatial data into district-level metrics. As depicted in Figure 10, the user-friendly graphical interface of the application lets users easily navigate and choose their desired district. Further, as highlighted in Figure 11, the left-hand panel showcases the indicator, allowing users to export this data graphically and as spreadsheets. Specifically, the 2019 average of each indicator was used in this study.

This constructed web-based application[2] is publicly accessible at: https://nattapong.users.earthengine.app/view/cambodia---districtdata----version-1.

The dataset used in this study was obtained by spatially merging the survey data with satellite indicators at the district level. The average nonfood expenditure is the predicted outcome, and 11 satellite-based indices are independent variables.

Figure 10

The User Interface of the Application Developed Specifically for This Study

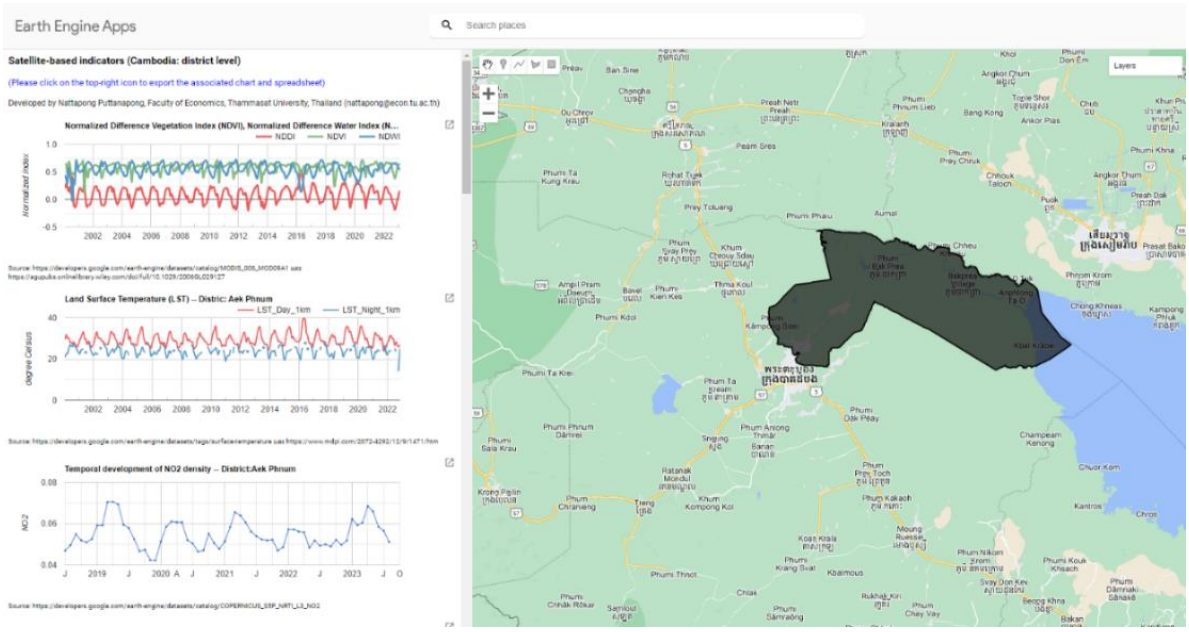


Note. From *Combodia: Satellite-based indicators*, by Google Earth Engine Apps, 2024 (https://nattapong.users.earthengine.app/view/cambodia----districtdata----version-1).

[2] We have developed the alternative version of this web-based application, generating the satellite indicators at the provincial level. The examples of this application and its access link are shown in the appendix.

Figure 11

Satellite Data at the District Level Extracted by the Application



Note. From *Satellite-based indicators (Combodia: district level)*, by Google Earth Engine Apps, 2024 (https://nattapong.users.earthengine.app/view/cambodia----districtdata----version-1).

Methods

Figure 12 illustrates the main process of undertaking quantitative analyses. As previously stated, the dataset was generated by conducting a spatial integration technique. Then, four machine learning methods were applied. In this study, RF yielded the highest accuracy. Therefore, the analysis was furthered by using two feature analysis techniques. Technically, VIMP and minimal depth are extended algorithms specifically based on the RF framework. The theoretical concepts of each method applied in this study are elaborated in the following subsections.

Machine Learning Techniques

GLS Regression

GLS is an enhanced regression method developed based on the ordinary least squares framework. Technically, the main enhancement addresses issues such as heteroscedasticity and residual correlation. Mathematically, the Cholesky decomposition of the variance–covariance matrix is employed to formulate the weight matrix, enabling the transformation of the initial regression model. This procedure results in unbiased, consistent, and asymptotically normal estimators with improved efficiency.

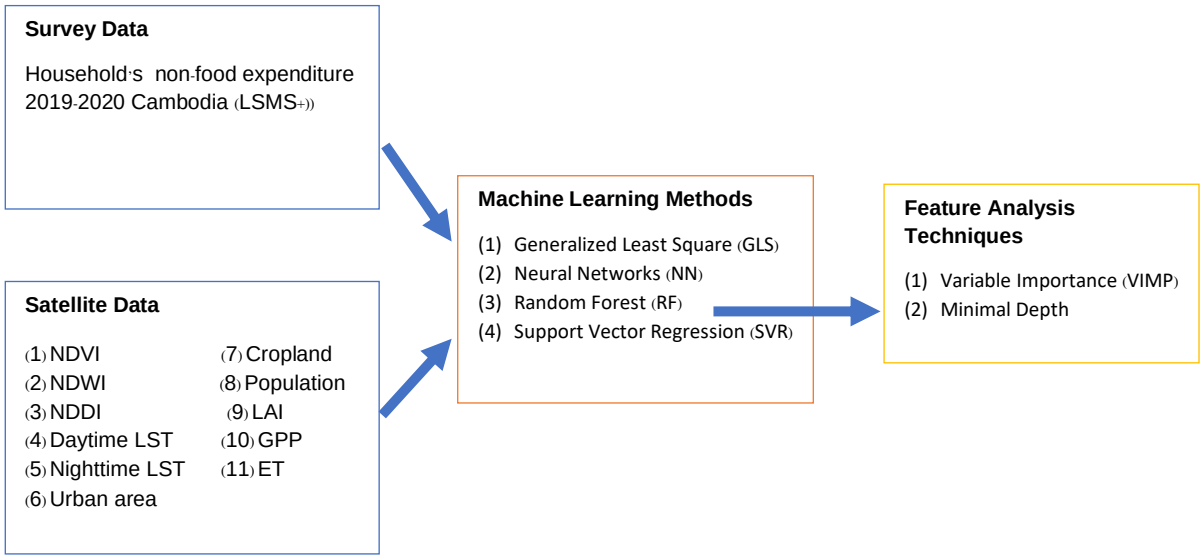
NN

As summarized by Ciaburro and Venkateswaran (2017) and Anesti et al. (2021), NNs are a computational model inspired by the human brain's architecture. These networks consist of interconnected nodes or neurons and are designed to uncover latent patterns within data. With this configuration, NNs can perform various tasks, including classification, regression, and clustering. The standard structure of an NN incorporates three categories of layers, namely, (1) input, (2) hidden, and (3) output layers.

The architecture of an NN is the hierarchy of layers, formulating the structural interconnections among neurons. The input layer takes in feature variables, whereas the output layer generates predictions. Adding hidden layers between the input and output layers introduces nonlinear transformations, augmenting the model's capabilities to handle complexity. Within this architecture, every connection linking nodes

Figure 12

Outline of Research Methods



carries a specific weight. Moreover, each neuron applies an activation function (e.g., sigmoid) on its input, allowing the network to capture complicated relationships within the data.

Training an NN is an iterative procedure that adjusts weights to minimize a designated loss function. This objective reduces the disparity between the model's predictions and target values. The backpropagation technique is usually utilized in this process, leveraging optimization algorithms, such as gradient descent. By employing the chain rule, backpropagation computes the loss function's gradient concerning each weight, thereby facilitating efficient weight updates throughout the network.

RF

RF is a prevalent ensemble learning algorithm for classification and regression tasks. It employs a fundamental approach, constructing an extensive array of decision trees and amalgamating their outcomes to yield predictions. This method leverages multiple trees to counteract overfitting, a common pitfall found in individual decision trees.

Breiman (2001) introduced RF. Its framework involves crafting numerous decision trees, each trained on a randomized subset of the training data. In addition, at every split, a random subset of features is selected, imbuing variety into the trees. The final prediction is derived from the average forecast across all trees. This aggregation process, termed bagging and advocated by Hastie et al. (2009), reduces variance without augmenting bias, thereby enhancing prediction robustness.

RF boasts a notable capability in evaluating feature significance, which sheds light on latent data relationships. Sections 4.2.1 (VIMP) and 4.2.2 (Minimal Depth) will explain these distinct attributes.

SVR

Vapnik (1998) initially introduced SVR, a supervised learning algorithm tailored for regression tasks. On the basis of the principles of support vector machine, SVR can predict continuous outcomes by identifying the optimal hyperplane that effectively captures the link between input variables and the output.

The core computational procedure of SVR is the mathematical process of numerically optimizing a hyperplane that accommodates data within a predefined error threshold denoted as ε. An important distinction is the penalty imposed on errors outside the margin. Exceptionally, SVR can handle linear and nonlinear relationships facilitated by diverse kernel functions, such as linear, polynomial, and radial basis functions. These kernels mathematically project the input

space into higher dimensions, enabling the algorithm to uncover intricate data connections.

Technically, SVR transforms its computational process into a constrained quadratic optimization problem, with the solution manifesting as support vectors—data points situated outside or on the ε-margin boundary. These support vectors crucially contribute to establishing the optimal hyperplane, a concept underscored by Zhang et al. (2010) and Wang et al. (2012). The selection of parameters such as cost parameter and kernel function profoundly influences model performance and its adaptability to particular data characteristics.

Feature Analysis Methods

The machine learning models introduced in the preceding section provide a quantitative approach to exploring the associations between geographical attributes and socioeconomic progress. However, those methods have limitations in explaining the predictive contribution of each variable. Therefore, this study expanded the computational efforts by implementing feature analysis techniques. This extension enhanced the machine learning capabilities to gauge the explanatory strength of individual variables. The following discussion outlines the main theoretical background of each feature analysis method.

VIMP

In the context of RF, a robust ensemble learning technique, VIMP emerges as a pivotal element, offering insights into the individual predictors' significance within the model. As discussed by Díaz et al. (2015), VIMP can prioritize the features that contribute the most to predictive accuracy. It also enables model interpretation and enhances data collection strategies.

A methodology, initially introduced by van der Laan (2006) and Ishwaran (2007) and subsequently elaborated by Strobl et al. (2007), revolves around assessing VIMP within RF. This calculation method predominantly relies on quantifying the rise in prediction error subsequent to randomly permuting the values of a particular variable. By perturbing a variable's values, this approach gauges the extent to which the model's accuracy diminishes due to the disruption of the variable's connection with the response. This comprehensive procedure is applied to all trees within the forest, ultimately resulting in an average importance score for each variable.

Minimal Depth

In the context of RF, the minimal depth technique is an investigative tool that quantifies the importance of variables. This assessment is based on the variables' positioning within individual decision trees forming the forest (Ishwaran et al., 2010; Seifert et al., 2021). On the basis of the foundational structure of decision trees, this method discerns that variables of greater importance are prone to emerge closer to the trees' root. By contrast, those of lesser significance tend to surface nearer to the leaves. Consequently, the minimal depth metric denotes the average depth at which a variable initially appears across all trees within the ensemble.

Each decision tree in the RF is traversed to compute the minimal depth of a variable, and the depth at which the variable is initially introduced in a split is noted. This depth is defined as the number of edges from the root to the node where the variable is utilized. This process is repeated for all trees, and the mean depth across the entire forest is then taken as the minimal depth for that specific variable. Lower minimal depth values correspond to higher importance, indicating a consistent involvement in early tree splits.

The minimal depth approach offers a distinct perspective that can uncover insights not captured by other methods. In contrast to metrics relying on variable permutation, minimal depth concentrates on a variable's structural role within decision trees. It evaluates the variable's contribution to the hierarchical division of the data space. Thus, the minimal depth outcome accurately reflects the variable's contribution in the context of RF-based prediction.

All machine learning algorithm computations and feature analysis procedures were executed using R software. Table 3 provides the key technical attributes of the R packages employed for each calculation.

Table 3

List of R Packages Used in Machine-Learning Computations

Method	Package's name	Technical reference
NN	Nnet	https://cran.r-project.org/web/packages/nnet/nnet.pdf
RF	randomForestSRC	https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf
SVR	e1071	https://cran.r-project.org/web/packages/e1071/e1071.pdf
VIMP	randomForestSRC	https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf
Minimal Depth	randomForestSRC	https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf

Note. From *List of R Packages Used in Machine-Learning Computations*, by The Comprehensive R Archive Network (CRAN), 2024.

RESULTS

Machine Learning Results

By utilizing the framework of classical regression analysis, the GLS approach was employed to produce estimated coefficients that quantify the effects of individual variables on predicted output variability (i.e., household nonfood consumption). These coefficients, along with their levels of statistical significance, are exhibited in Table 4.

Some geospatial indicators are statistically significant. The daytime LST and NDDI values are negatively associated with the predicted outcome. The results are in line with many publications indicating that high temperatures (Buyantuyev & Wu, 2010; Dissanayake et al., 2019; Huang et al., 2011) and drought can affect households' economic status (Amare et al., 2018; Richardson, 2007; Thiede, 2014). Moreover, the size of the population is negatively correlated with household consumption capability. Conversely, urban density positively correlates with the household socioeconomic condition, thereby affirming the agglomeration force inducing the higher income of households in highly urbanized areas.

Table 6 exhibits that variables obtained from the ground survey have mixed outcomes in explaining the household socioeconomic status. Particularly, one group of the assets owned by the household is a powerful predictor, whereas the other is statistically insignificant. The ownership of a car, a mobile phone, and a motorcycle and the floor area of residence (i.e., Flr_area) have high explanatory powers, as indicated by their p values. However, owning a tuktuk (i.e., motor tricycles) has a low explanatory capability, and other variables are statistically insignificant.

These results indicate that the geospatial characteristics extracted from satellite data can be integrated with the ground survey as the independent variables for predicting the household socioeconomic condition. However, the R-square value of the model is 0.260, explaining only 26% of the variance in household expenditure. Thus, machine learning methods are included to alternatively examine the relationship between socioeconomic status, satellite-based indicators, and survey data.

Table 4

Coefficients Obtained from GLS Regression [Dependent variable: Household nonfood expenditure]

Variable	Coefficient and Standard Error
Intercept	0.398*** (0.096)
Cropland	-0.008 (0.037)
Urban	0.163* (0.093)
ET	0.022 (0.028)
GPP	0.009 (0.048)
LAI	-0.033 (0.082)
LST_D	-0.056* (0.029)
LST_N	0.033 (0.051)
NDVI	-0.051 (0.053)
NDWI	-0.039 (0.087)
NDDI	0.057* (0.033)
POP	-0.178* (0.095)
Bicycle	0.023 0.030
Boat	0.036 (0.048)
Car	0.192*** (0.041)
Cellphone	0.154*** (0.038)
Computer	0.087 (0.057)
Motorcycle	0.132*** (0.024)
Tractor	0.022 (0.036)
Tuktuk	0.066* (0.032)
Flr_area	0.168*** (0.030)

Note. Standard Error in parentheses; * p<0.10, ** p<0.05, *** p<0.01

In addition to GLS, three machine learning algorithms—NN, RF, and SVR—were applied. Figures 13 and 14 reveal that these machine learning techniques outperformed the GLS model. Specifically, RF yielded the lowest root mean square error (RMSE), equivalent to an R-square value of 0.406. This superior fit suggests that these machine learning approaches can capture the intricate and multifaceted relationships among the variables affecting household socioeconomic status. However, these methods fall short of detailing the specific contributory effects of individual variables, necessitating a follow-up feature analysis to explore their respective roles in influencing household expenditure and income.

Figure 13

Comparison of Root Mean Square Error (RMSE) [Dependent variable: Household nonfood expenditure]

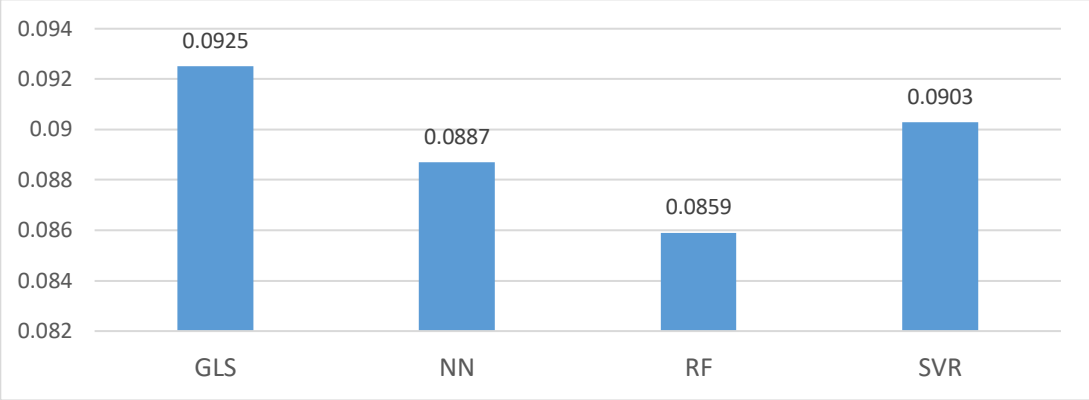
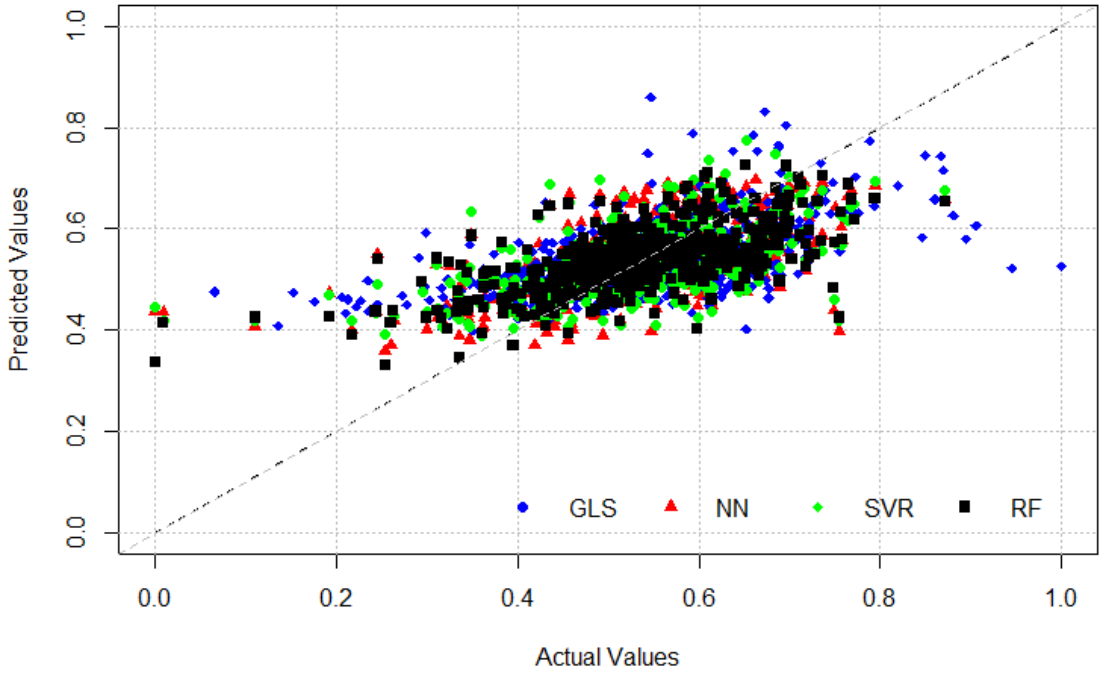


Figure 14

A Scatter Plot Comparing the Actual Values and Predicted Ones of an Independent Variable



Feature Analysis Results

The results generated by the two distinct feature analysis techniques are depicted in Figures 15 and 16. Table 5 ranks each variable's contribution to the model based on these outcomes. The key insights derived from this variable ranking can be summarized as follows:

- Car, motorcycle, cellphone, and the floor area of residence (Flr_area) are ranked as the most contributive variables explaining the variation in a household's socioeconomic status. These variables obtained from the survey represent households' purchasing behavior on high-priced electronic devices. These ownerships are directly associated with purchasing power and income level. Therefore, they were categorized by VIMP and minimal depth methods as the most significant variables. These findings align with the existing literature (Mika et al., 2021; Noeurn, 2020; Asongu, 2013; Wong & Shuaibim, 2023).

- VIMP and minimal depth classified urban area, GPP, population, LST_N, LST_D, LAI, NDWI, NDDI, and NDVI as the factors that moderately contributed to the households' socioeconomic condition variations. These results indicated that the geospatial characteristics have the midrange explanatory power in predicting household expenditure and income. Especially, these results are similar to the conclusions of previous publications (Arezki & Brückner, 2012; Barrios et al., 2010; Damania et al., 2020; Gilmont et al., 2018; Liaqut et al., 2019; Sruthi & Aslam, 2015)

- The rest of the variables included in the survey (i.e., bicycle, tuktuk, tractor, and boat) were categorized as the lowest contributive factors influencing the socioeconomic status of households in Cambodia. The ownership of these assets was not statistically correlated with the households' income and expenditure. Sharma et al. (2016) highlighted the anomalies in poverty distribution in Cambodia, and natural resources significantly contribute to household incomes within Cambodia. In particular, the rate of poverty reduction is unevenly distributed across regions. Rural locales heavily rely on directly consuming products (food and nonfood) harvested from nature. As a result, a direct relationship between income poverty and consumption poverty is not observed in certain areas, suggesting that some asset ownership might not be related to consumption (Hansen & Neth, 2006; Jiao et al., 2015; McKenney & Tola, 2002; Sophal & Acharya, 2002; Tong & Sry, 2011).

The main results obtained from the VIMP and minimal depth methods signify the contribution of integrating survey data and satellite-based indicators in poverty analysis in Cambodia. The survey data can identify household-specific consumption behavior. The satellite-based indices complementarily reveal the location-specific physical conditions influencing infrastructure accessibility, occupational opportunities, and related risks. These obtained outcomes indicate that integration of satellite indicators, survey data and machine learning methods can enable the future enhancement of poverty analysis and effective policy formulation.

Figure 15

Variable Importance (VIMP) Result

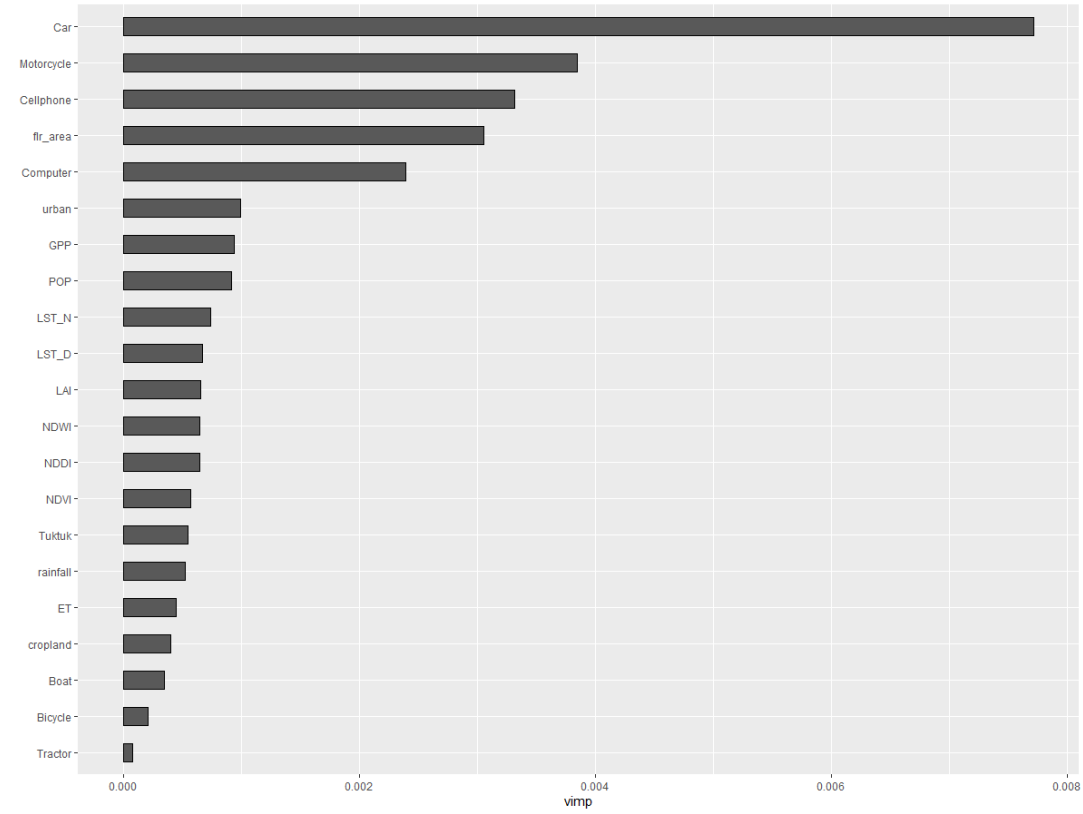


Figure 26

Minimal Depth Result

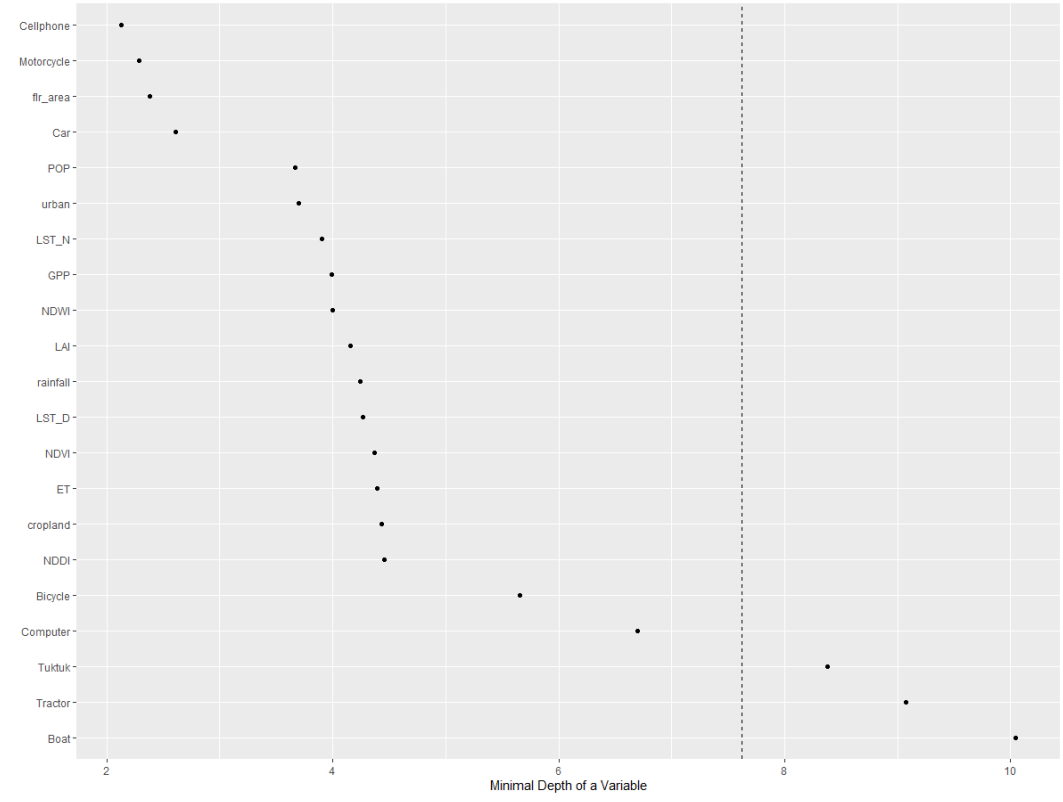


Table 5

Ranking of Contributive Power of Each Variable [Dependent variable: Household nonfood expenditure]

Rank	Variable Importance	Minimal Depth
1	Car	Cellphone
2	Motorcycle	Motorcycle
3	Cellphone	Flr_area
4	Flr_area	Car
5	Computer	Population
6	Urban	Urban
7	GPP	LST_N
8	Population	GPP
9	LST_N	NDWI
10	LST_D	LAI
11	LAI	Rainfall
12	NDWI	LST_D
13	NDDI	NDVI
14	NDVI	ET
15	Tuktuk	Cropland
16	Rainfall	NDDI
17	ET	Bicycle
18	Cropland	Computer
19	Boat	Tuktuk
20	Bicycle	Tractor
21	Tractor	Boat

Limitations

We acknowledge that our RF model achieves lower accuracy compared to existing studies (Ayush et al., 2021; Engstrom et al., 2017; Jean et al., 2016; Puttanapong et al., 2022; Puttanapong et al., 2023; Yeh et al., 2020), which is a major drawback of our study. Satellite-based indicators such as NDDI, NDWI, LAI, and GPP are widely used in geospatial research, especially in environmental and ecological studies. However, as documented by Running et al. (2004), Myneni et al. (2002), Gao (1996), and McFeeters (1996), the explanatory power of these indicators can be diminished in specific contexts, particularly when applied to localized or complex systems. Based on key findings from the existing literature, we have identified the following technical limitations that influence these discrepancies.

High heterogeneity terrain: The capabilities of capturing geographical features, especially vegetation and water bodies, are lowered in mountainous regions or complex landscapes.

Threshold sensitivity and underlying assumption: These remote-sensing indices are sensitive to the thresholds used to classify some

localized conditions. Specifically, generalizing thresholds over wide areas can lead to misinterpretations. Moreover, assumptions about light-use efficiency and climatic factors can lead to errors in estimating agricultural outputs.

Semi-arid and arid regions: These satellite indicators often have a lower correlation with actual agricultural yields in the areas experiencing drought. Technically, soil reflectance and sparse vegetation make it difficult to differentiate between wet and dry conditions in such regions, leading to reduced accuracy and subsequently generating discrepancies in predicting the agricultural income of rural households.

Atmospheric noise: Atmospheric noise, such as clouds or aerosols, diminishes the explanatory powers of many satellite-based indices.

As previously stated, we recognize that the predictive power is a significant limitation of our study. Therefore, future improvements should focus on enhancing the model's accuracy by addressing three key aspects. First, computational methodology could benefit from integrating advanced approaches, including the latest machine learning techniques and deep learning algorithms (Tochaiwat & Pultawee, 2024). Second, enhancing the model with diverse data sources, ranging from additional satellite indicators to social media metrics and mobility indices, could refine its predictive precision. Furthermore, factors representing local environmental conditions and city planning should be included (John et al., 2019; Thammapornpilas, 2015). Third, applying pre-processing methods—such as Principal Component Analysis (PCA)—can significantly enrich the data quality. PCA helps reduce dimensionality by transforming the original variables into a smaller set of uncorrelated components that capture most of the variance in the data. This eliminates redundant information and highlights the most significant features contributing to the predictive model.

In addition to enhancing accuracy, incorporating survey data from various time periods would facilitate robust spatiotemporal analysis. This would enable extended monitoring focused on the geographical distribution of poverty dynamics.

CONCLUSION

This study introduced a new analytical model integrating satellite indices, survey data, and machine learning to monitor developmental progress. An application on Google Earth Engine has been specifically developed to extract satellite-based indicators. Among several machine learning methodologies utilized, RF yielded the most accurate prediction. Moreover, the results obtained from feature analyses identified the significant association between economic activity density, settlement pattern, and household socioeconomic status—as proxied by non-food expenditure.

The obtained findings essentially unveiled the intimate correlation between economic activity density and population patterns, emphasizing the necessity for governments to promote regional development. Strategic investments in infrastructure and the diversification from agricultural activities could be critical, catalyzing job creation and fostering alternative economic opportunities for enhanced income generation.

For further refinement of this analytical approach, its predictive power must be enhanced. Integrating more variables and exploring additional machine learning models are advised to achieve a more accurate and robust analysis.

REFERENCES

Aiken, E., Bellue, S., Karlan, D., Udry, C. R., & Blumenstock, J. (2022). Machine learning and mobile phone data can improve the targeting of humanitarian assistance. *Nature, 603*, 864–870. https://doi.org/10.1038/s41586-022-04484-9

Amare, M., Jensen, N. D., Shiferaw, B., & Cissé, J. D. (2018). Rainfall shocks and agricultural productivity: Implication for rural household consumption. *Agricultural Systems, 166*, 79–89. https://doi.org/10.1016/j.agsy.2018.07.014

Anesti, N., Kalamara, E., & Kapeta, G. (2021). *Forecasting with machine learning methods and multiple large datasets. Bank of England Staff Working Paper No. 923.* https://www.bankofengland.co.uk/working-paper/2021/forecasting-uk-gdp-growth-with-large-survey-panels

Asongu, S.A. (2013). *The impact of mobile phone penetration on African inequality*. AGDI Working Paper, No. WP/13/021, African Governance and Development Institute (AGDI). https://www.econstor.eu/bitstream/10419/123599/1/agdi-wp13-021.pdf

Arezki, R., & Brückner, M. (2012). Rainfall, financial development, and remittances: Evidence from Sub-Saharan Africa. *Journal of International Economics, 87*(2), 377–385. https://doi.org/10.1016/j.jinteco.2011.12.010

Asian Development Bank. (2022). *Cambodia, key indicators* [Dataset]. ADB Data Library. https://data.adb.org/dataset/cambodia-key-indicators

Ayush, K., Uzkent, B., Tanmay, K., Burke, M., Lobell, D., & Ermon, S. (2021). Efficient poverty mapping from high resolution remote sensing images. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*(1), 12–20. https://doi.org/10.1609/aaai.v35i1.16072

Barrios, S., Bertinelli, L., & Strobl, E. (2010). Trends in rainfall and economic growth in Africa: A neglected cause of the African growth tragedy. *Review of Economics and Statistics, 92*, 350–366. https://www.jstor.org/stable/pdf/27867541.pdf

Bhattacharya, H., & Innes, R. (2006). Is there a nexus between poverty and environment in rural India? *Proceedings of the American Agricultural Economics Association Annual Meeting*, *July 23-26, Long Beach, CA, USA* (pp. 23–26).

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science, 350*(6264), 1073–1076. https://doi.org/10.1126/science.aac4420

Blumenstock, J. E. (2016). Fighting poverty with data. *Science, 353*(6301), 753–754. https://doi.org/10.1126/science.aah5217

Blumenstock, J., Karlan, D., & Udry, C. (2021). *Using mobile phone and satellite data to target emergency cash transfers*. CEGA Blog Post. https://poverty-action.org/study/using-mobile-phone-and-satellite-data-target-emergency-cash-transfers-togo

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brown, C., & Lall, U. (2006). Water and economic development: The role of variability and a framework for resilience. *Natural Resources Forum, 30*(4), 306–317. https://doi.org/10.1111/j.1477-8947.2006.00118.x

Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (2021). Using satellite imagery to understand and promote sustainable development. *Science, 371*(6535), Article eabe8628. https://doi.org/10.1126/science.abe8628

Buyantuyev, A., & Wu, J. (2009). Urban heat islands and landscape heterogeneity: Linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landscape Ecology, 25*, 17–33. https://doi.org/10.1007/s10980-009-9402-4

Chen, X., Liu, C., & Yu, X. (2022). Urbanization, economic development, and ecological environment: Evidence from provincial panel data in China. *Sustainability, 14*(3), Article 1124. https://doi.org/10.3390/su14031124

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record, 88*(s1), 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Ciaburro, G., & Venkateswaran, B. (2017). *Neural networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles*. Packt Publishing.

Damania, R., Desbureaux, S., & Zaveri, E. (2020). Does rainfall matter for economic growth? Evidence from global sub-national data (1990–2014). *Journal of Environmental Economics and Management, 102*, Article 102335. https://doi.org/10.1016/j.jeem.2020.102335

Díaz, I., Hubbard, A., Decker, A., & Cohen, M. (2015). Variable importance and prediction methods for longitudinal problems with missing variables. *PLoS ONE, 10*(3), Article e0120031. https://doi.org/10.1371/journal.pone.0120031

Dissanayake, D., Morimoto, T., Murayama, Y., Ranagalage, M., & Handayani, H. H. (2019). Impact of urban surface characteristics and socio-economic variables on the spatial variation of land surface temperature in Lagos city, Nigeria. *Sustainability, 11*(1), 25. https://doi.org/10.3390/su11010025

Elbers, C., Lanjouw, J. O., & Lanjouw, P. F. (2002). *Micro-level estimation of welfare*. Policy Research Working Paper 2911, World Bank, Washington, DC. http://documents.worldbank.org/curated/en/362131468739473297/Micro-level-estimation-of-welfare

Eng, R., & Lim, S. (2024). The economic development and level of poverty in Cambodia. *Educational Administration: Theory and Practice, 30*(6), 3693–3701. https://doi.org/10.53555/kuey.v30i6.5806

Engstrom, R., Hersh, J., & Newhouse, D. (2017). Poverty from space: Using high-resolution satellite imagery for estimating economic well-being. *PLOS ONE, 12*(9), Article e0184396. https://doi.org/10.1371/journal.pone.0184396

Erenstein, O., Hellin, J., & Chandna, P. (2010). Poverty mapping based on livelihood assets: A meso-level application in the Indo-Gangetic Plains, India. *Applied Geography, 30*(1), 112–125. https://doi.org/10.1016/j.apgeog.2009.05.001

Fatehkia, M., Tingzon, I., Orden, A., Sy, S., Sekara, V., Garcia-Herranz, M., & Weber, I. (2020). Mapping socioeconomic indicators using social media advertising data. *EPJ Data Science, 9*(1), Article 22. https://doi.org/10.1140/epjds/s13688-020-00235-w

Fujii, T. (2007). To use or not to use?: Poverty mapping in Cambodia. In T. Bedi, A. Coudouel, & K. Simler (Eds.), *More than a pretty picture: Using poverty maps to design better policies and interventions* (pp. 125–142). https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=1601&context=soe_research

Fujii, T. (2010). *Micro-level estimation of child malnutrition indicators in Cambodia*. Oxford University Press.

Gao, B. C. (1996). NDWI—A normalized difference vegetation index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*(3), 257–266. http://dx.doi.org/10.1016/S0034-4257(96)00067-3

Gilmont, M., Hall, J. W., Grey, D., Dadson, S. J., Abele, S., & Simpson, M. (2018). Analysis of the relationship between rainfall and economic growth in Indian states. *Global Environmental Change*, *49*, 56–72. https://doi.org/10.1016/j.gloenvcha.2018.01.003

Gu, Y., Brown, J. F., Verdin, J. P., & Wardlow, B. (2007). A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central great plains of the United States. *Geophysical Research Letters, 34*(6), Article L06407. https://doi.org/10.1029/2006GL029127

Guo, Y., Zeng, J., Wu, W., Hu, S., Liu, G., Wu, L., & Bryant, C. R. (2021). Spatial and temporal changes in vegetation in the Ruoergai region, China. *Forests, 12*(1), 76. https://doi.org/10.3390/f12010076

Hall, O., Dompae, F., Wahab, I., & Dzanku, F. M. (2023). A review of machine learning and satellite imagery for poverty prediction: Implications for development research and applications. *Journal of International Development, 35*(7), 1–16. https://doi.org/10.1002/jid.3751

Hansen, K., & Top, N. (2006). *Natural forest benefits and economic analysis of natural forest conversion in Cambodia*. Working Paper (Vol. 33), Cambodia Development Resource Institute, Phnom Penh. https://cdri.org.kh/storage/pdf/wp33e_1617794583.pdf

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Random forests. *The Elements of Statistical Learning* (pp. 587–604). Springer.

Head, A., Manguin, M., Tran, N., & Blumenstock, J. E. (2017). Can human development be measured with satellite imagery? *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*, *Lahore, Pakistan* (Article 8). https://doi.org/10.1145/3136560.3136576

Huang, G., Zhou, W., & Cadenasso, M. L. (2011). Is everyone hot in the city? Spatial pattern of land surface temperatures, land cover and neighborhood socioeconomic characteristics in Baltimore, MD. *Journal of Environment Management*, *92*(7), 1753–1759. https://doi.org/10.1016/j.jenvman.2011.02.006

Huguet, J. W., Chamratrithirong, A., Rao, N. R., & Than, S. S. (2000). Results of the 1998 population census in Cambodia. *Asia-Pacific Population Journal, 15*(3), 3–22. https://repository.unescap.org/rest/bitstreams/a945053c-b447-4754-ac33-01c15a647d35/retrieve

Ishwaran, H. (2007). Variable importance in binary regression trees and forests. *Electronic Journal of Statistics, 1*, 519–537. https://doi.org/10.1214/07-EJS039

Ishwaran, H., Kogalur, U. B., Gorodeski, E. Z., Minn, A. J., & Lauer, M. S. (2010). High-dimensional variable selection for survival data. *Journal of the American Statistical Association, 105*(489), 205–217. https://doi.org/10.1198/jasa.2009.tm08622

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

Jiao, X., Smith-Hall, C., & Theilade, I. (2015). Rural household incomes and land grabbing in Cambodia. *Land Use Policy*, *48*, 317–328. https://doi.org/10.1016/j.landusepol.2015.06.008

Jin, X., Wan, L., Zhang, Y. K., & Schaepman, M. (2008). Impact of economic growth on vegetation health in China based on GIMMS NDVI. *International Journal of Remote Sensing, 29*(13), 3715–3726. https://doi.org/10.1080/01431160701772542

John, A., Allison, M., Amadi, D. E., & Allison, C. (2019). Anti-democratic spaces and impoverishment: Role of roads in low-income residential areas. *Nakhara: Journal of Environmental Design and Planning, 16*, 15–32. https://doi.org/10.54028/NJ2019161532

Kristjanson, P., Radeny, M., Baltenweck, I., Ogutu, J., & Notenbaert, A. (2005). Livelihood mapping and poverty correlates at a meso-level in Kenya. *Food Policy, 30*(5–6), 568–583. https://doi.org/10.1016/j.foodpol.2005.10.002

Li, L., Tan, Y., Ying, S., Yu, Z., Li, Z., & Lan, H. (2014). Impact of land cover and population density on land surface temperature: Case study in Wuhan, China. *Journal of Applied Remote Sensing, 8*(1), Article 084993. https://doi.org/10.1117/1.JRS.8.084993

Li, G. Y., Chen, S. S., Yan, Y., & Yu, C. (2015). Effects of urbanization on vegetation degradation in the Yangtze River Delta of China: Assessment based on SPOT-VGT NDVI. *Journal of Urban Planning and Development, 141*(4), Article 05014026. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000249

Li, M., Wu, T., Wang, S., Sang, S., & Zhao, Y. (2022). Phenology–gross primary productivity (GPP) method for crop information extraction in areas sensitive to non-point source pollution and its influence on pollution intensity. *Remote Sensing, 14*(12), Article 2833. http://dx.doi.org/10.3390/rs14122833

Liu, F., Xiao, X., Qin, Y., Yan, H., Huang, J., Wu, X., Zhang, Y., Zou, Z., & Doughty, R. (2022). Large spatial variation and stagnation of cropland gross primary production increases the challenges of sustainable grain production and food security in China. *Science of the Total Environment, 811*, Article 151408. https://doi.org/10.1016/j.scitotenv.2021.151408

Liaqut, A., Younes, I., Sadaf, R., & Zafar, H. (2019). Impact of urbanization growth on land surface temperature using remote sensing and GIS: A case study of Gujranwala City, Punjab, Pakistan. *International Journal of Economic Environment Geology, 9*, 44–49. https://www.researchgate.net/publication/330441884_Impact_of_Urbanization_Growth_on_Land_Surface_Temperature_using_remote_sensing_and_GIS_A_Case_Study_of_Gujranwala_City_Punjab_Pakistan

Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2015). Social media fingerprints of unemployment. *PLoS ONE, 10*(5), Article e0128692. https://doi.org/10.1371/journal.pone.0128692

Mika, K., Minna, M., Noora, V., Jyrki, L., Jari, K. O., Anna, A., Eliyan, C., Dany, V., Maarit, K., & Nicholas, H. (2021). Situation analysis of energy use and consumption in Cambodia: household access to energy. *Environment, Development and Sustainability*, *23,* 18631–18655. https://doi.org/10.1007/s10668-021-01443-8

McFeeters, K. (1996). The use of the normalized difference water index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing, 17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

McKenney, B., & Tola, P. (2002). *Natural resources and rural livelihoods in Cambodia: A baseline assessment*. Working Paper (Vol. 23), Cambodia Development Resource Institute, Phnom Penh. https://cdri.org.kh/storage/pdf/wp23e_1617794774.pdf

Morikawa, R. (2014). Remote sensing tools for evaluating poverty alleviation projects: A case study in Tanzania. *Procedia Engineering, 78*, 178–187. https://doi.org/10.1016/j.proeng.2014.07.055

Mourad, R., Jaafar, H., Anderson, M., & Gao, F. (2020). Assessment of leaf area index models using harmonized landsat and sentinel-2 surface reflectance data over a semi-arid irrigated landscape. *Remote Sensing, 12*(19), Article 3121. http://dx.doi.org/10.3390/rs12193121

Mulovhedzi, N., Araya, N., Mengistu, M., Fessehazion, M., du Plooy, C., Araya, H., & van der Laan, M. (2020). Estimating evapotranspiration and determining crop coefficients of irrigated sweet potato (Ipomoea batatas) grown in a semi-arid climate. *Agricultural Water Management, 233*, Article 106099. https://doi.org/10.1016/j.agwat.2020.106099

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., & Running, S. W. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment, 83*(1–2), 214–231. https://doi.org/10.1016/S0034-4257(02)00074-3

National Institute of Statistics. (2017). *Report of Cambodia socio-economic survey 2017*. Ministry of Planning. https://www.nis.gov.kh/nis/CSES/Final%20Report%20CSES%202017.pdf

National Institute of Statistics. (2019). *Cambodia Living Standards Measurement Study - Plus 2019-2020*. The World Bank. https://doi.org/10.48529/agcn-nn81

National Institute of Statistics. (2021). *Report of Cambodia socio-economic survey 2021*. Ministry of Planning. https://www.nis.gov.kh/nis/CSES/Final%20Report%20of%20Cambodia%20Socio-Economic%20Survey%202021_EN.pdf

Noeurn, V. (2020). Factors affecting electricity consumption of residential consumers in Cambodia*. IOP Conf. Series: Earth and Environmental Science, 746*, Article 012034. https://doi.org/10.1088/1755-1315/746/1/012034

Pandit, P., Krishnamurthy, K., & Bakshi, B. (2022). Chapter 22 - Prediction of crop yield and pest-disease infestation. In A. Abraham, S. Dash, J. J.P.C. Rodrigues, B. Acharya, & S. K. Pani (Eds.), *Intelligent Data-Centric Systems, AI, Edge and IoT-based Smart Agriculture* (pp. 375–393). Academic Press. https://doi.org/10.1016/B978-0-12-823694-9.00021-9

Pokhriyal, N., & Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences of the United States of America, 114*(46), E9783–E9792. https://doi.org/10.1073/pnas.1700319114

Puttanapong, N., Prasertsoong, N., & Peechapat, W. (2023). Predicting provincial gross domestic product using satellite data and machine learning methods: A case study of Thailand. *Asian Development Review*, *40*(2), 39–85. https://doi.org/10.1142/S0116110523400024

Puttanapong, N., Martinez, A., Bulan, J. A. N., Addawe, M., Durante, R. L., & Martillan, M. (2022). Predicting poverty using geospatial data in Thailand. *ISPRS International Journal of Geo-Information, 11*(5), Article 293. http://dx.doi.org/10.3390/ijgi11050293

Richardson, C.J. (2007). How much did droughts matter? Linking rainfall and GDP growth in Zimbabwe. *African Affairs, 106*(424), 463–478. https://www.jstor.org/stable/4496463

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *BioScience, 54*(6), 547–560. https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2

Ruthirako, P., Darnsawasdi, R., & Chatupote, W. (2015). Intensity and pattern of land surface temperature in Hat Yai City, Thailand. *Walailak Journal of Science and Technology, 12*(1), 83–94. https://wjst.wu.ac.th/index.php/wjst/article/view/977

Seifert, S., Gundlach, S., & Szymczak, S. (2019). Surrogate minimal depth as an importance measure for variables in random forests. *Bioinformatics, 35*(19), 3663–3671. https://doi.org/10.1093/bioinformatics/btz149

Sharma, R., Nguyen, T. T., Grote, U., & Nguyen, T. T. (2016). *Changing livelihoods in rural Cambodia: Evidence from panel household data in Stung Treng*. ZEF Working Paper Series (No. 149), Center for Development Research (ZEF), University of Bonn. https://www.econstor.eu/bitstream/10419/144856/1/857348353.pdf

Shi, K., Chang, Z., Chen, Z., Wu, J., & Yu, B. (2020). Identifying and evaluating poverty using multisource remote sensing and point of interest (POI) data: A case study of Chongqing, China. *Journal of Cleaner Production, 255*, Article 120245. https://doi.org/10.1016/j.jclepro.2020.120245

Sophal, C., & Acharya, S. (2002). *Facing the challenge of rural livelihoods: A perspective from nine villages in Cambodia*. Cambodia Development Resource Institute, Phnom Penh.

Sruthi, S., & Aslam, M. M. (2015). Agricultural drought analysis using the NDVI and land surface temperature data; a case study of Raichur district. *Aquatic Procedia, 4*, 1258–1264. https://doi.org/10.1016/j.aqpro.2015.02.164

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. (2017). Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society Interface, 14*(127), Article 20160690. https://doi.org/10.1098/rsif.2016.0690

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*(1), Article 25. https://doi.org/10.1186/1471-2105-8-25

Takada, S., Morikawa, S., Idei, R., & Kato, H. (2021). Impacts of improvements in rural roads on household income through the enhancement of market accessibility in rural areas of Cambodia, *Transportation,* 48:2857–2881. https://doi.org/10.1007/s11116-020-10150-8

Thammapornpilas, J. (2015). Urban spatial development to mitigate urban heat Island effect in the inner area of Bangkok. *Nakhara: Journal of Environmental Design and Planning*, *11*, 29–40. https://ph01.tci-thaijo.org/index.php/nakhara/article/view/104849

Thiede, B. C. (2014). Rainfall shocks and within-community wealth inequality: Evidence from rural Ethiopia. *World Development, 64*, 181–193. https://doi.org/10.1016/j.worlddev.2014.05.028

Tochaiwat, K., & Pultawee, P. (2024). House type specification for housing development project using machine learning techniques: A study from Bangkok metropolitan region, Thailand. *Nakhara: Journal of Environmental Design and Planning, 23*(1), Article 403. https://doi.org/10.54028/NJ202423403

Tong, K., & Sry, B. (2011). *Poverty and environmental links: The case of rural Cambodia*. Working Paper (Vol. 64), Cambodia Development Resource Institute, Phnom Penh. https://cdri.org.kh/storage/pdf/wp64e_1617793884.pdf

United Nations Development Programme. (2022). *Human development report 2021/2022*. RR Donnelley Company. https://hdr.undp.org/system/files/documents/global-report-document/hdr2021-22overviewen.pdf

van der Laan, M. J. (2006). Statistical inference for variable importance. *International Journal of Biostatistics, 2*(1), Article 2. https://doi.org/10.2202/1557-4679.1008

Vapnik, V. (1998). *Statistical learning theory*. John Wiley & Sons, Inc.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., & Lobell, D. (2020). Using pu.blicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications, 11*, Article 2583. https://doi.org/10.1038/s41467-020-16185-w

Wan Mohd Jaafar, W. S., Abdul Maulud, K. N., Muhmad Kamarulzaman, A. M., Raihan, A., Md Sah, S., Ahmad, A., Saad, S. N. M., Mohd Azmi, A. T., Jusoh Syukri, N. K. A., & Razzaq Khan, W. (2020). The influence of deforestation on land surface temperature: A case study of Perak and Kedah, Malaysia. *Forests, 11*, 670. https://doi.org/10.3390/f11060670

Wang, Y., Wang, B., & Zhang, X. (2012). A new application of the support vector regression on the construction of financial conditions index to CPI prediction. *Procedia Computer Science, 9*, 1263–1272. https://doi.org/10.1016/j.procs.2012.04.138

Wikimedia Commons. (2020). *Provincial boundaries in Cambodia* [Map]. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Provincial_Boundaries_in_Cambodia.svg

Wong, G., & Shuaibim, A. (2023). Model selection and optimization for poverty prediction on household data from Cambodia. *Journal of Emerging Investigators*, *6,* 1–11. https://doi.org/10.59720/22-290

World Bank. (2022). *Cambodia poverty assessment 2022: Toward a more inclusive and resilient Cambodia*. https://www.worldbank.org/en/country/cambodia/publication/cambodia-poverty-assessment-2022-toward-a-more-inclusive-and-resilient-cambodia

Youneszadeh, S., Amiri, N., & Pilesjo, P. (2015). The effect of land use change on land surface temperature in the Netherlands. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-1/W5*, 745–748. https://doi.org/10.5194/isprsarchives-XL-1-W5-745-2015

Zhang, X., Hu, L., & Wang, Z. (2010). Multiple kernel support vector regression for economic forecasting. *International Conference on Management Science & Engineering 17th Annual Conference Proceedings* (pp. 129-134). IEEE. https://doi.org/10.1109/ICMSE.2010.5719795

Zheng, G., & Moskal, L. M. (2009). Retrieving leaf area index (LAI) using remote sensing: Theories, methods and sensors. *Sensors, 9*(4), 2719–2745. https://doi.org/10.3390/s90402719

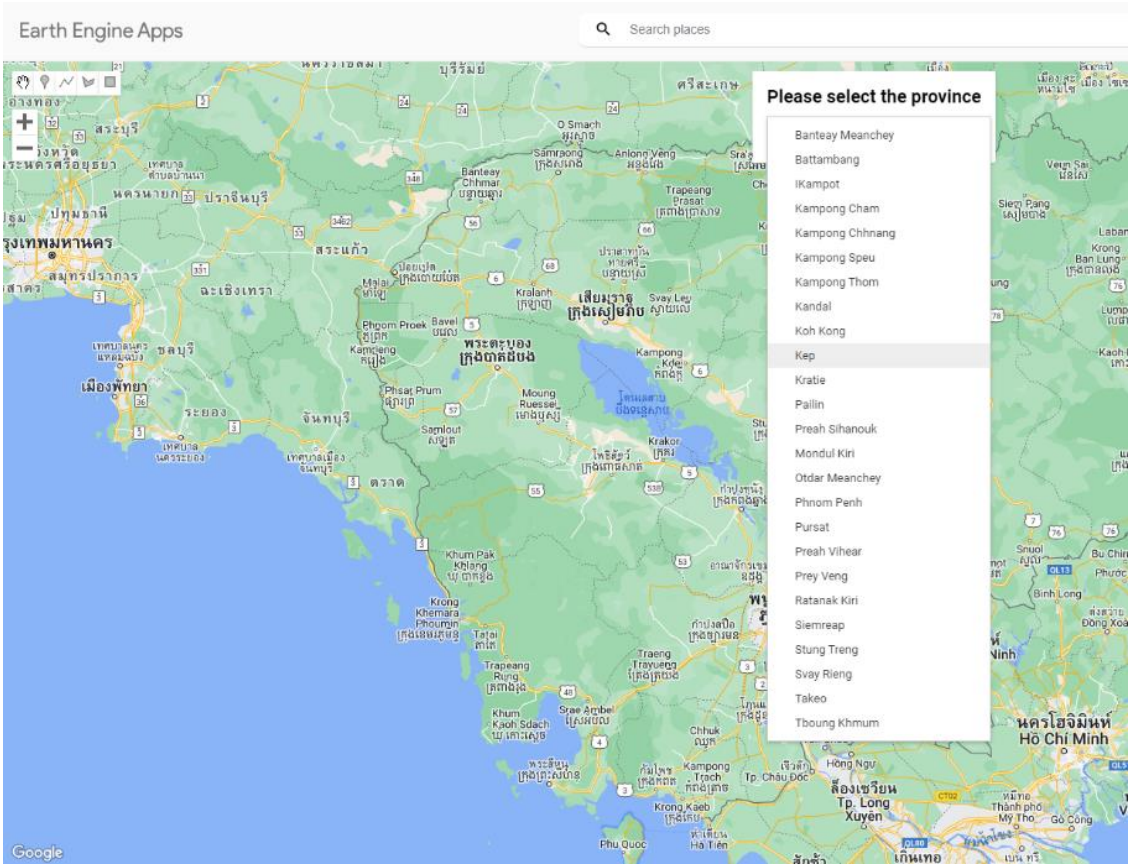
Zhou, Y., & Liu, Y. (2022). The geography of poverty: Review and research prospects. *Journal of Rural Studies, 93*, 408–416. https://doi.org/10.1016/j.jrurstud.2019.01.008

APPENDIX

The alternative version of the application has been developed and launched, publicly accessible at: https://nattapong.users.earthengine.app/view/cambodia---provincial-time-series---satellite-data. This application provides a collection of satellite data, which is the annual average at the provincial level. Figures A1 and A2 exemplify the user interface and the result generated by this application.

Figure A1

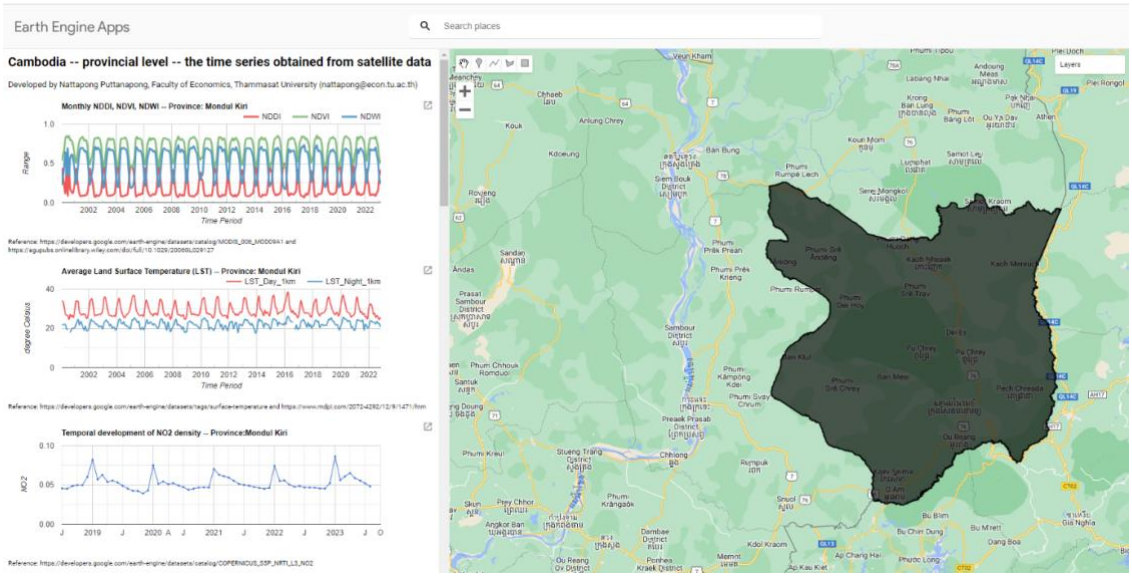
The User Interface of the Alternative Version of Web-Based Application



Note. From *Combodia: Satellite-based indicators*, by Google Earth Engine Apps, 2024 (https://nattapong.users.earthengine.app/view/cambodia----districtdata----version-1). Copyright 2024 by Google LLC.

Figure A2

Satellite Data at the Provincial Level Extracted by the Application



Note. From *Satellite-based indicators (Combodia: district level)*, by Google Earth Engine Apps, 2024 (https://nattapong.users.earthengine.app/view/cambodia----districtdata----version-1). Copyright 2024 by Google LLC.