

MIT Open Access Articles

A data-driven approach to mapping multidimensional poverty at residential block level in Mexico

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Zea-Ortiz, M., Vera, P., Salas, J. et al. A data-driven approach to mapping multidimensional poverty at residential block level in Mexico. *Environ Dev Sustain* (2024).

As Published: 10.1007/s10668-024-05230-z

Publisher: Springer Science and Business Media LLC

Persistent URL: <https://hdl.handle.net/1721.1/155801>


Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution





A data-driven approach to mapping multidimensional poverty at residential block level in Mexico

Marivel Zea-Ortiz¹ · Pablo Vera¹ · Joaquín Salas^{1,4}  · Roberto Manduchi³ · Elio Villaseñor¹ · Alejandra Figueroa² · Ranyart R. Suárez²

Received: 26 July 2022 / Accepted: 1 July 2024
© The Author(s) 2024

Abstract

Accurate, inexpensive and granular human poverty assessments are critical for data-driven policy decision-making. This research proposes a novel approach to computing poverty scores utilizing multispectral satellite images and indices calculated from census reference values. We show how this approach can leverage standard and sparse survey-based multidimensional poverty assessments at the municipal level to develop a deep learning architecture to obtain poverty scores at the residential block level. This method has the distinctive feature that the obtained inference corresponds to Multidimensional Measurement of Poverty generated by CONEVAL, the Mexican agency responsible for measuring poverty. We provide a reliable alternative to survey-based approaches with an R^2 of 0.802 ± 0.022 for the *lack of housing quality and spaces* dimension. A convolutional neural network trained on multispectral satellite images and the *lack of housing quality and spaces* dimension, which is regressed from census reference variables corresponding to lack of water, electricity, sewage, concrete floor, toilet and occupancy level obtains an R^2 of 0.753. These results represent a significant step forward in including machine learning techniques to provide reliable information at reduced costs and a higher spatiotemporal frequency than traditional person-to-person surveys.

Keywords Human poverty assessment · Sustainable development goals · Computational intelligence for sustainability

✉ Joaquín Salas
jsalasr@ipn.mx

¹ CICATA Querétaro, Instituto Politécnico Nacional, Cerro Blanco 141, 76090 Querétaro, Mexico

² Laboratorio de Ciencia de Datos y Métodos Modernos de Producción de Información, Instituto Nacional de Geografía y Estadística, Héroe de Nacozari 2301, Aguascalientes, 20276 Aguascalientes, México

³ Department of Computer Science and Engineering, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA

⁴ Earth Signals and Systems Group, Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

1 Introduction

Over the past two centuries, we have witnessed a sustained upward trajectory in global prosperity. This greater economic well-being and societal progress has not been without challenges and setbacks. With persistent inequalities and socio-economic disparities, numerous people have managed to break free from extreme poverty, which is defined as living on less than \$2.15 USD per day. However, despite the overall progress in global prosperity, a significant portion of the world's population lives in impoverished conditions, deprived of access to necessities crucial for a dignified life. At the core of this issue lies the deprivation of food security, hunger and malnutrition, education, adequate healthcare, electricity, clean water, and sanitation. All these situations are experienced by those living in impoverished conditions (Ravallion, 2015).

Unfortunately, the convergence of multiple global challenges, such as the COVID-19 pandemic, climate change, and ongoing conflicts, casts a gloomy forecast for the future. It is now anticipated that the number of individuals living in extreme poverty will surge to approximately 860 million people, representing over 10% of the world's population (Oxfam, 2022). This forecast derives from the profound impacts these interconnected crises can have on societies, exacerbating vulnerability and compromising progress toward poverty eradication and sustainable development. The recognition of this situation, coupled with the establishment and adoption of the United Nations' Sustainable Development Goals (SDGs), has served as a catalyst for researchers to precisely measure and understand global economic poverty.

Human vulnerability is a multifaceted and intricate assortment of risks, threats, and weaknesses that are inherent in the human condition. It encompasses a wide range of physical, social, economic, and environmental aspects, all of which interact and influence an individual's susceptibility to harm, adversity, or adverse outcomes. In a more specific definition, UN-habitat (2003) suggests that human settlements poverty may be defined in terms of clean drinking water and proper drainage, the quality and adequacy of housing structures that are not overcrowded, and housing security. Governments and other organizations measure human poverty primarily through censuses, which can be time-consuming and expensive processes that may only occur every few years. However, the effectiveness of this approach has been the subject of debate (Bajotto et al., 2017; Kakareka, 2013; Mackenzie et al., 2014). Some researchers have suggested alternatives to censuses, including surveys. However, this approach can be prone to errors due to declining participation and a lack of representativeness in the sample.

A promising approach to assessing poverty is using multispectral satellite images (Elvidge et al., 2009; Xie et al., 2016). For instance, our recent research developed a methodology to evaluate structural poverty at the residential block level in Mexico based on the UN-habitat definition (Salas et al., 2021). Recently, Owusu et al. (2021) proposed incorporating local needs and end-user requirements into Earth Observations (EO) poverty assessments.

This research definition of poverty closely follows the multidimensional model introduced by the National Council for the Evaluation of Social Development Policy (CONEVAL, 2016), the Mexican institution with a law mandate to measure poverty in the country. CONEVAL expresses poverty in terms of income and social rights. Subsequently, social rights include education, health, housing adequacy, food security, social security, and basic services. Overall, the agency delivers 16 poverty dimensions at the municipal level every five years and at the state level every two years.

In this approach (see Fig. 1), we find the mapping between the census reference values and the poverty dimensions, expressing the latter at the level of residential blocks. We then find the mapping between satellite images and the poverty dimensions at that level of resolution. In the former mapping, we use classical Machine Learning techniques. In the latter, we employ Convolutional Neural Networks (CNNs). This manuscript further details the indicator of poverty related to the lack of housing quality and spaces. This approach of mapping multispectral satellite images to a direct reference of human poverty, as defined by a national agency, holds the potential to advance our understanding of poverty dynamics and inform evidence-based decision-making. This approach enhances the accuracy and reliability of poverty assessments and enables comprehensive poverty mapping on a broader scale.

Distancing from prior research, the contributions of this paper include the following:

- An approach to relating census reference values describing poverty to the dimension constructed by CONEVAL, the Mexican agency responsible for measuring poverty, with a significant coefficient of determination.
- An empirical demonstration showing that it is possible to construct a mapping between multispectral satellite images and the referenced poverty indicator at the residential block level for the entire country of Mexico using CNNs with high-performance levels.
- Publicly releasing computer code and required data enabling other researchers and practitioners to verify the results and serving as a stepping stone for further research.

The remainder of the article is organized as follows. In Sect. 2, we provide an overview of previous works in assessing poverty by remote sensing techniques and the concept of multidimensional poverty employed by CONEVAL. Later on, in Sect. 3, we introduce an approach to transforming census reference values into standardized poverty dimensions. We also detail the methods employed to evaluate the mapping between poverty dimensions and multispectral satellite images. The results of implementing the models, including their use in tracking poverty in census gap years, are presented in Sect. 4. We conclude the paper in Sect. 5 by summarizing the key findings and outlining potential avenues for future research.

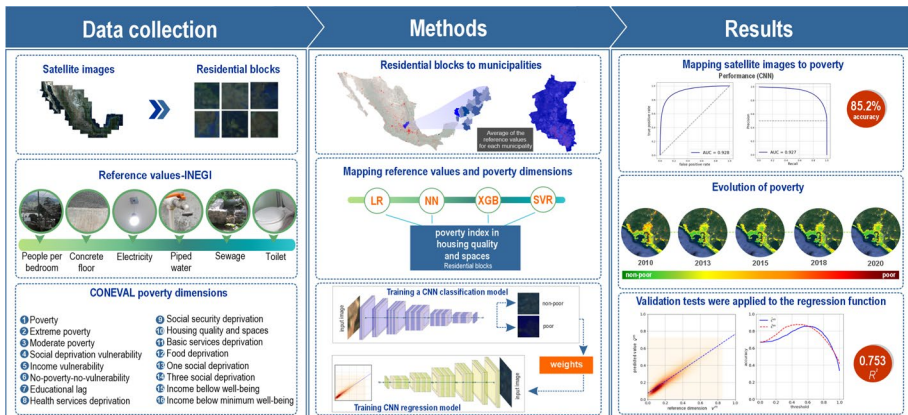


Fig. 1 Fine-grained mapping of satellite images to poverty dimensions (refer to the text for details)

approach materializes an intuition relating to development and energy consumption. Coincidentally, the National Oceanic and Atmospheric Administration (NOAA) has been running the Defense Meteorological Program (DMSP) Operational Line-Scan System (OLS) since 1992, generating a large historical archive of gray-scale images covering the entire Earth. More recently, the LuoJia1 satellite publicly provides a new generation of nightlight satellite images with higher resolution (Lin et al., 2022; Xu et al., 2021a). Complementary to nightlight images, many researchers have carried out studies using multispectral images (Ajami et al., 2019; Engstrom et al., 2017; Fisher et al., 2021; Gram-Hansen et al., 2019; Hersh et al., 2021; de Mattos et al., 2021; Salas et al., 2021; Stark et al., 2020; Suel et al., 2021; Wurm et al., 2019). In some cases, multispectral images have been combined with nightlight images (Jean et al., 2016; Tingzon et al., 2019; Xu et al., 2021b). These techniques have been used to identify poverty, food insecurity, and other indicators of human vulnerability and have shown promising results.

The use of remote sensing technology for poverty assessment is becoming increasingly accepted, primarily due to publicly available data coming from satellites such as Landsat (Bansal et al., 2020; Salas et al., 2021) and the higher relative resolution Sentinel-2 (Fisher et al., 2021; de Mattos et al., 2021). Recent research has focused on exploring even higher resolution satellite data (Ajami et al., 2019; Engstrom et al., 2017; Roy et al., 2019; Stark et al., 2020; Suel et al., 2021), often in conjunction with their lower resolution counterparts (Agarwal et al., 2019; Gram-Hansen et al., 2019; Li et al., 2021; Verma et al., 2019). Particularly noteworthy is the combination of high-resolution multispectral and synthetic aperture radar images with low-resolution images (Hersh et al., 2021; Wurm et al., 2019), or multispectral images with nightlight time data (Elvidge et al., 2009).

Another significant factor is the growing availability of reliable methods to extract poverty indicators. Classical methods such as Extreme Gradient Boosting (XGB) (Bansal et al., 2020; Dorji et al., 2019; Hersh et al., 2021; Xu et al., 2021b), Support Vector Machine (SVM) (Li et al., 2021; Xu et al., 2021a), Canonical Correlation Forest (Gram-Hansen et al., 2019; de Mattos et al., 2021), Random Forest (RF) (Engstrom et al., 2017; Hersh et al., 2021; Lin et al., 2022; Yin et al., 2021) have been extensively explored with researchers finding their proposed methods to outperform others. However, in a comparison study, Li et al. (2019) found RF, SVM, XGB, Neural Networks (NN), and Gaussian Processes (GP) to perform equally well in assessing poverty. Recently, Deep Learning methods in the form of CNNs (Agarwal et al., 2019; Ajami et al., 2019; Fisher et al., 2021; Gram-Hansen et al., 2019; Hersh et al., 2021; Jean et al., 2016; Salas et al., 2021; Stark et al., 2020; Suel et al., 2021; Tingzon et al., 2019; Verma et al., 2019; Wurm et al., 2019; Xie et al., 2016) have been gaining traction due to their formidable performance and their requirement for large datasets, which the problem of assessing human poverty from EO provides ample and global examples.

2.2 Multidimensional poverty

We aim to develop a method for predicting human poverty at the residential block level over time using satellite imagery to capture relevant information. To achieve this goal, we first examine the *poverty indicators* produced by CONEVAL, currently defined only at the municipal level and based on sporadic survey data obtained from the National Institute of Statistics and Geography (INEGI). While these indicators offer valuable insights, they need to provide more granular data to capture the nuances of poverty at the residential block level. To address this limitation, we employ *census reference values*, which are surveyed at the residential block

level and can be considered a proxy for poverty indicators. We then outline our approach to extrapolating poverty indicators at the residential block level from the census reference values. By combining these data sources and leveraging them with CNN models, we aim to create an accurate and reliable predictive tool for assessing poverty at the residential block level.

CONEVAL's approach to assessing poverty is based on a multidimensional space consisting of three axes: economic well-being, social rights, and territorial context. Each axis represents different dimensions of poverty, including unsatisfied basic needs, fundamental human rights, and relational and community aspects (CONEVAL, 2016). CONEVAL recognizes that these dimensions are not static and can change over time in response to social conditions. CONEVAL identifies individuals experiencing multidimensional poverty as those who cannot exercise at least one social right and cannot afford the goods and services necessary to meet their basic needs. CONEVAL has developed specific criteria for each axis, such as income poverty and extreme income poverty for economic well-being, access to education, access to health services, social security, housing quality and spaces, essential housing services, and access to quality and nutritious food for social rights.

The information generated by CONEVAL is crucial for identifying and addressing the needs of the poorest groups in Mexico. The National Development Plan of Mexico naturally incorporates this information as it is an essential reference for the state secretariats in charge of social development and the governments of states and municipalities. Well-being programs that serve the neediest rely on this information as an obligated reference. CONEVAL's multidimensional approach provides a comprehensive understanding of the factors contributing to poverty and enables policymakers to design poverty reduction strategies.

CONEVAL provides a range of dimensions to measure social and economic well-being at various levels of government in Mexico. At the national scale, the council produces measurements every two years for state governments and every five years for the entire country, state governments, and municipalities (Czarnecki, 2013). At the municipal level, CONEVAL's methodology estimates the proportion of poor populations exposed to different types of risks. CONEVAL's database contains statistics for 2010, 2015, and 2020 for the 2456 municipalities in Mexico (CONEVAL, 2022). It includes 16 dimensions of social well-being, economic deprivation, and territorial context (Czarnecki, 2013). These dimensions measure the proportion of the population experiencing poverty, extreme poverty, moderate poverty, income vulnerability, and vulnerability to social deprivation. The database also includes the proportion of the population facing one or three types of social deprivation: lack of access to education, health services, social security, essential household services, nutritious and quality food, and housing quality and spaces. The dataset further registers the proportion of the population with income below the poverty income or extreme poverty income thresholds. Additionally, CONEVAL includes dimensions that capture the ratio between the population in extreme poverty and those not in extreme poverty or vulnerable in the territorial context.

3 Methodology

The Mexican census (INEGI, 2012) is a valuable source of information for assessing poverty, as it provides data on various factors such as access to water, electricity, sewage availability, and household structural information. These factors, which we call *reference values*, are used by UN-habitat (2003) to define poverty. To identify poor areas, we extracted information from the census corresponding to the residential blocks and used reference values suggested by UN-Habitat. Weighted by the residential block population, we aggregated

the residential block reference values to the municipal level to map them to the CONEVAL poverty dimensions. Using this mapping, we interpolated poverty indicators at the residential block level to establish a second regression mapping, this time with multispectral satellite images. We obtained the best regression results when the initial CNN weights for the poverty indicators' response variable were first obtained through training a classifier. This section provides details of such a process. It is worth noting that the census records lack geographic coordinates, such as longitude and latitude, when a person's identity can be established. Please refer to Table 2 for the symbology to the discussion that follows.

3.1 Census reference values

We use various reference values for the j -th residential block in the m -th municipality to assess its living conditions. These reference values include the average number of people per bedroom in a residential block $\hat{x}_{j,o}^m$ and the proportion of households in the block without basic amenities such as concrete floor $x_{j,f}^m$, electricity $x_{j,e}^m$, piped water $x_{j,w}^m$, sewage $x_{j,s}^m$, and toilet $x_{j,t}^m$. Each variable, except $\hat{x}_{j,o}^m$, takes a value between 0 and 1. To normalize the range of $\hat{x}_{j,o}^m$, we set $x_{j,o}^m$ to be the minimum of $\hat{x}_{j,o}^m$ and a predefined value τ_o divided by τ_o , or $x_{j,o}^m = \min(\hat{x}_{j,o}^m, \tau_o) / \tau_o$. Experimentally, we have established $\tau_o = 4$ as the maximum average occupancy per bedroom. Additionally, we obtain the population q_j^m of each block from the census. To summarize the different reference values for the same block, we group them into a vector $\mathbf{x}_j^m = [x_{j,o}^m, x_{j,f}^m, x_{j,e}^m, x_{j,w}^m, x_{j,s}^m, x_{j,t}^m]^T$.

To establish a relationship between the reference values x_j^m , associated with a residential block, and a CONEVAL dimension, we need to aggregate the x_j^m values of all blocks in the same municipality. This aggregation can be achieved by computing the average of the x_j^m values for $j = 1, \dots, J(m)$, weighted by the block population. Thus, we obtain the value x^m for the m -th municipality using

$$x^m = h\left(x_1^m, \dots, x_{J(m)}^m\right) = \frac{\sum_{j=1}^{J(m)} q_j^m x_j^m}{\sum_{j=1}^{J(m)} q_j^m}, \tag{1}$$

where $J(m)$ is the number of residential blocks in the m -th municipality. Suppose we are working with n -dimensional vectors \mathbf{x}_j^m of features defined for each block in a municipality. In that case, we can replace the scalar values x_j^m by \mathbf{x}_j^m in (1) to obtain the average vector \mathbf{x}^m that corresponds to the m -th municipality. Thus, we can compute the averages of the six reference variables weighted by the population of the block q_j^m by using \mathbf{x}_j^m as x_j^m in (1) as

$$\mathbf{x}^m = h\left(\mathbf{x}_1^m, \dots, \mathbf{x}_{J(m)}^m\right). \tag{2}$$

This process enables us to derive a representative value for each municipality, which can then be used to analyze the relationship between the characteristics of the residential blocks and the CONEVAL dimension.

Note that the function h aggregates scalars or vectors associated with residential blocks (see Fig. 2), such as \mathbf{x}_j^m vectors, at the municipality level. Since we use an average weighted by the block population q_j^m for this aggregation, h provides an estimate of the proportion of people in a municipality that would be associated with the characteristics represented by x^m . Then, we analyze the relationship between the traits aggregated by h and a CONEVAL dimension.

Table 2 Symbology. Subindex j refers to the j -th residential block and super index m to the m -th municipality

α	Parameters for the convolutional neural network
\tilde{C}^m, C^m	Aggregated and actual class label, i.e., poverty or no-poverty at municipal level
\tilde{C}_j^m, C_j^m	Predicted and actual class label, i.e., poor or non-poor at the residential block level
\mathbf{I}_j^m	Satellite image
q_j^m	Number of inhabitants
R^2	Coefficient of determination
σ	Reference value standard deviation
τ_o	Occupancy threshold used for normalization
τ_v	Threshold assigned to the poverty value to distinguish between poor and non-poor
θ^k	Parameters corresponding to the k poverty dimension
v_j^m, v^m	Poverty index at residential block and municipal level
$\hat{v}_j^{k,m}, \hat{v}_j^{k,m}$	Actual and predicted k -th poverty index at residential block
$\tilde{v}^{k,m}, \hat{v}^{k,m}$	Estimated k -th poverty index at municipality level aggregated from actual and predicted residential block indices
\mathbf{w}	Weights assigned to the reference values
x_j^m, x^m	Generic reference value with aggregation at residential block and municipal level
$\hat{x}_{j,o}^m$	Unnormalized census reference value for occupancy
$x_{j,o}^m, x_{j,f}^m, x_{j,e}^m, x_{j,w}^m, x_{j,s}^m, x_{j,t}^m$	Census reference values for occupancy (o), floor (f), electricity (e), water (w), sewage (s), and toilet (t)
\mathbf{x}_j^m	A vector containing the census reference values
y_i, \hat{y}_i, \bar{y}	Ground truth, estimate, and mean value for the reference response variable
$F(\cdot)$	Regression function from reference values to poverty indices
$G(\cdot)$	Convolutional neural network
$h(\cdot)$	Aggregating function
$J(m)$	Number of residential blocks at the m -municipality
$T(\cdot)$	Binary classification rule

3.2 Poverty indicators

In previous work (Salas et al., 2021), we established a baseline for a poverty index v_j^m at the residential block j level for the municipality m . We used a linear combination of six reference variables related to UN-Habitat, with a uniform weight \mathbf{w} as

$$v_j^m = \mathbf{w}^T \mathbf{x}_j^m. \tag{3}$$

where \mathbf{x}_j^m represents the vector of the six reference variables. To transform the problem into a classification one, we followed (Dorji et al., 2019) and introduced a threshold τ_v at the median value of the poverty index values as

$$C_j^m = T(v_j^m, \tau_v) = \begin{cases} \text{poor} & \text{if } v_j^m \geq \tau_v, \\ \text{non-poor} & \text{otherwise.} \end{cases} \tag{4}$$

After establishing the poverty index for residential blocks, we trained a CNN to map multispectral satellite images and the corresponding poverty classes. Our approach was based on the assumption that the six reference variables used to define the poverty index were

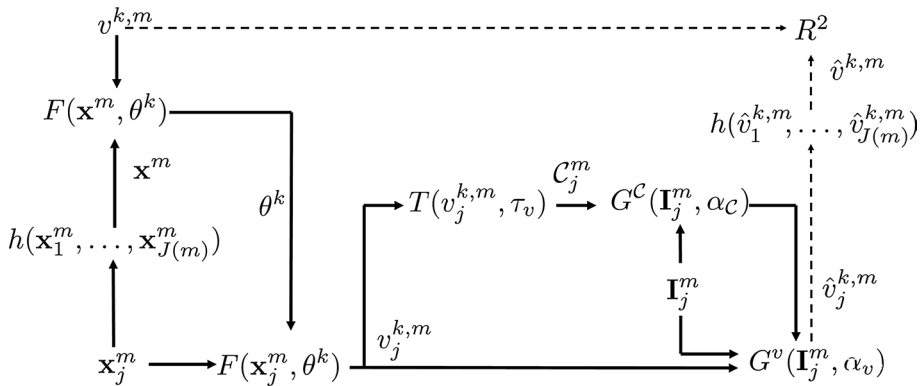


Fig. 2 Mapping reference values and satellite images to poverty (refer to Table 2 and the text for details)

equally important. However, it may be preferable to reference v_j^m to a standard poverty indicator, such as CONEVAL (Czarnecki, 2013). Unfortunately, CONEVAL’s measurements are provided at the municipal level, so we cannot directly associate them with the surveyed residential blocks.

To obtain poverty indicators at the residential block level, we use census reference values and compute the parameters θ^k for a function $v_j^{k,m} = F(\mathbf{x}_j^m, \theta^k)$, which for a geolocation \mathbf{x}_j^m maps these values to the k -th poverty dimension for the j -th block in the municipality m . We assume that census reference values are correlated with poverty indicators.

To ensure consistency with the CONEVAL poverty dimension $v^{k,m}$, defined at the municipality level, we minimize the squared norm of the difference between $v^{k,m}$ and the weighted average $\tilde{v}^{k,m}$ of $v_j^{k,m} = F(\mathbf{x}_j^m, \theta^k)$ computed using (1):

$$\tilde{v}^{k,m} = h(v_1^{k,m}, \dots, v_{J(m)}^{k,m}). \tag{5}$$

We experimented with evaluating various design choices for $F(\mathbf{x}_j^m, \theta^k)$. We started with a simple linear regressor and then considered more complex non-linear models, including NN (Lichtner-Bajjaoui, 2021), XGB (Chen & Guestrin, 2016), and Support Vector Regressors (SVR) (Smola & Schölkopf, 2004). These models were chosen to represent different classes of machine learning algorithms, such as graphical, tree-based, and kernel-based approaches. To train each model, we used the CONEVAL poverty dimension as the ground truth and sought to minimize the Mean Square Error (MSE) of the predicted values for each poverty dimension.

3.3 Validating poverty inferences

To assess poverty at the residential block level, we first train a regression function F using a reference indicator at the municipal level. Then, we apply F to obtain poverty indices for residential blocks. We conduct four validation tests to ensure that using the regression function at a different scale is appropriate.

In the first test, we use the regression function to process the reference values $F(\mathbf{x}_j^m)$ and obtain the weighted average of the values for $v_j^{k,m}$ using (1). This produces $\tilde{v}^{k,m} = h(v_1^{k,m}, \dots, v_{J(m)}^{k,m})$. In the second test, we apply the threshold τ_v to the values of $v_j^{k,m}$ to obtain C_j^m and then aggregate these values using (1) to produce $\tilde{C}^m = h(C_1^m, \dots, C_{J(m)}^m)$. We compare \tilde{C}^m with the actual values of C^m . Similarly, in the third test, we use a CNN $G(\mathbf{I}_j^m, \alpha)$ to generate a class prediction \hat{C}_j^m for each residential block image. We aggregate these predictors using (1) to obtain $\hat{C}^m = h(\hat{C}_1^m, \dots, \hat{C}_{J(m)}^m)$, which we compare to the actual values of C^m . In the fourth test, we generate poverty index predictions $\hat{v}_j^{k,m}$ using the CNN regression model and aggregate these predictions to obtain $\hat{v}^{k,m} = h(\hat{v}_1^{k,m}, \dots, \hat{v}_{J(m)}^{k,m})$ (see Fig. 2). Next, $\tilde{v}^{k,m}$ and $\hat{v}^{k,m}$ are compared with $v^{k,m}$, the reference value given by CONEVAL.

To obtain C^m for each municipality, we apply the threshold τ_v to the $v_j^{k,m}$ values as in (4). We calculate the MSE and the determination coefficient R^2 for the values of $\tilde{v}^{k,m}$ and $\hat{v}^{k,m}$, and the accuracy for the values of \tilde{C}^m and \hat{C}^m as a function of the threshold. Note that \tilde{C}^m corresponds to the aggregation of the class values used for labeling the images \mathbf{I}_j^m , while \hat{C}^m corresponds to the predictions made by the CNN.

3.4 Mapping satellite images to poverty

Previously (Salas et al., 2021), we trained CNNs using image patches to predict poverty. We analyzed and compared several architectures, including LeNet-5 (LeCun et al., 1989), ResNet-50 v2 (He et al., 2016), ResNeXt-50 (Xie et al., 2017), and EfficientNet-B3 (Tan & Le, 2019). The EfficientNet-B3 architecture outperformed the others in classifying images as belonging to poor or non-poor regions.

For this research, we selected the EfficientNet-B3 architecture as it performed the best and used a relevant measurement of poverty (Czarnecki, 2013) as the response variable. Initially, we trained a CNN model for classification as in (Salas et al., 2021). We threshold the poverty indices $v_j^{k,m}$ using (4) to obtain the class values of C_j^m with which we train the CNN classifier that maps satellite images to residential block-level poverty. Due to the low resolution of the images, we considered this approach appropriate as the network only has to learn to identify the features that distinguish between two classes. However, the model discards information as it only predicts whether a region is poor. To gain more insight into the poverty inferences, we trained three additional CNN models, based on EfficientNet-B3, for regression to predict the continuous values of $v_j^{k,m}$. These models differ only in the initiation stage of their training. Transfer learning from the previous CNN classifier significantly improved the regression model performance.

Note that $v_j^{k,m}$ is an estimation of the proportion of people within a residential block who are deemed poor according to a CONEVAL dimension (see Sects. 2.2 and 3.2). However, this definition does not encompass varying degrees of poverty among individuals. We assume that the CONEVAL methodology defines specific criteria between poor and non-poor.

We conducted experiments using three regression models to obtain predictions of the poverty dimension from satellite images. During the experiments, we discovered that

initializing the weights of a regression model's layers with a pre-trained classification model improved performance. This approach, known as transfer learning, proved beneficial for the regression model. It is important to note that when training a classification model, the image patches should be labeled using the categorical variable C_j^m defined by (4), rather than using the continuous values $v_j^{k,m}$ of the poverty index. Below, we provide a detailed explanation of the training process for these models.

3.5 Approximation error

The error we incur in approximating a regressor with a classifier can be evaluated in several forms. Among the possible options, we employ the following approach. Suppose we have a set of poverty index values $\mathbf{v} = \{v_1, \dots, v_N\}$, where $v_i \in [0, 1]$. To frame the regression problem as a classification, we need to establish a threshold value τ_v that separates poor and non-poor instances. To determine the loss associated with this approach, we calculate the difference between the poverty dimension values and the mean values, M_0 and M_1 . First, we compute the mean values as

$$M_0 = \frac{1}{N_0} \sum_{v_i \leq \tau_v} v_i, \text{ and } M_1 = \frac{1}{N_1} \sum_{v_i > \tau_v} v_i. \tag{6}$$

Here, we define $N_0 = \sum_{i=1}^N \mathbb{1}(v_i \leq \tau_v)$ and $N_1 = \sum_{i=1}^N \mathbb{1}(v_i > \tau_v)$, where $\mathbb{1}$ is a function that is one when its argument is true and zero otherwise. Let us define a classification rule $\mathbb{T}(v_i)$ as

$$\mathbb{T}(v_i) = \begin{cases} M_0 & \text{if } v_i \leq \tau_v, \\ M_1 & \text{otherwise.} \end{cases} \tag{7}$$

The approximation error can be evaluated as

$$\epsilon = \frac{1}{N} \sum_{i=1}^N (v_i - \mathbb{T}(v_i))^2.$$

This expression quantifies the deviation of the poverty dimension values from the corresponding mean values for each class.

3.6 Performance assessment

This research aims to develop indicators that closely resemble the reference indicators. This section describes the procedure employed to assess performance. Let y_i denote the reference response variable and \hat{y}_i be its estimate obtained through a function $F(\mathbf{x}_i, \theta)$, where \mathbf{x}_i represents the predictors and θ defines the corresponding parameters for F . We evaluated the performance of the regressors using metrics, including the ratio between the root-mean-square error and the standard deviation of the response variable. These are defined as

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \text{ and } \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}. \tag{8}$$

Here, \bar{y} denotes the response variable's mean value, and σ is its standard deviation. When the ratio RMSE/σ exceeds one, the mean value of the response variable should be used for better performance. We also employ the determination coefficient R^2 , which measures how much better the regression model is compared to using the response variable mean value as an estimate. It is defined as

$$R^2 = 1 - \left(\frac{\text{RMSE}}{\sigma} \right)^2, \text{ for } \text{RMSE} \leq \sigma. \quad (9)$$

R^2 takes values between zero and one, with higher values indicating better performance. It represents the proportion of residual variance the model explains, which is the difference between the observed and predicted values. Finally, we employ the Kullback-Leibler (KL) divergence to measure the difference between the probability distributions of the reference and predicted values. It is defined as

$$\text{KL}(\mathbf{p} \parallel \mathbf{q}) = \sum_{i=1}^n p_i \log \left(\frac{p_i}{q_i} \right), \quad (10)$$

where $\mathbf{p} = (p_1, \dots, p_m)^T$ and $\mathbf{q} = (q_1, \dots, q_m)^T$ are the probability distributions corresponding to the reference and prediction values and KL is evaluated as a representation for \mathbf{p} .

4 Results

This section presents the results of implementing the methods described earlier to construct regressors for mapping census reference values \mathbf{X} and poverty dimensions \mathbf{v} . We also evaluate the performance of these regressors in predicting poverty at the residential block level using multispectral satellite images.

The dataset consists of 2456 municipalities in Mexico, each with six reference values corresponding to the proportion of houses in a residential block with piped water, sewage, electricity, toilet, concrete floor, and occupancy less than $\tau_o = 4$. Residential blocks per municipality vary widely. Santiago Tepetlapa in Oaxaca has the minimum number of residential blocks at 13, whereas Tijuana in Baja California has the maximum at 20,932. On average, municipalities have 887.01 residential blocks, with a standard deviation of 1744.11. The response variable $\mathbf{v}_{2456 \times 1}^k$ corresponds to the k -poverty indicator. To evaluate the performance of the regressors, we randomly partition the dataset $\{\mathbf{X}, \mathbf{v}^k\}$ into a training set $\{\mathbf{X}', \mathbf{v}'^k\}$ (75%) and a testing set $\{\mathbf{X}'', \mathbf{v}''^k\}$ (25%). We repeat this procedure 30 times for each poverty dimension and compute performance evaluations p_i at each iteration (see Fig. 3). We report the mean value and the standard deviation of these performance metrics across all splits.

4.1 Tuning up the regressors

To determine the most effective regression scheme for mapping census reference values \mathbf{X} and poverty indicators \mathbf{v} to multispectral satellite images, we tested four different methods: Linear Regression (LR), Neural Networks (NN), Extreme Gradient Boosting (XGB), and Support Vector Regression (SVR).

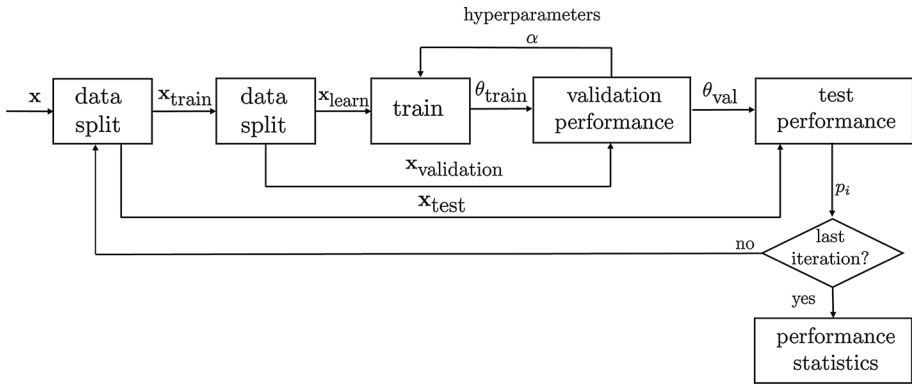


Fig. 3 Performance assessment for regressors (refer to the text for details)

Linear Regression. We used Linear Regression as a baseline. We added a column to both the training \mathbf{X}' set and testing \mathbf{X}'' set to account for bias, resulting in the augmented sets $\mathcal{X}' \leftarrow [\mathbf{X}', \mathbf{1}_{1,842 \times 1}]$ and $\mathcal{X}'' \leftarrow [\mathbf{X}'', \mathbf{1}_{614 \times 1}]$, respectively. We solved the linear system $\mathbf{w}^k = (\mathcal{X}'^T \mathcal{X}')^{-1} \mathcal{X}'^T \mathbf{v}^k$ to obtain the weights for the k -th poverty dimensions during training. We then used these weights to predict the response variable \mathbf{v}''^k for the testing set as $\mathbf{v}''^k = \mathcal{X}'' \mathbf{w}^k$.

Neural Networks. We tested four networks, two with one hidden layer and two with three hidden layers. The first two had 10 or 30 neurons. The other two had 10, 8, and 5 neurons, and the other had 30, 16, and 8 neurons in the first, second, and third layers. We divided the training set into learning and validation subsets using a 75%/25% split. We used the Levenberg-Marquardt backpropagation algorithm (Hagan & Menhaj, 1994) to train the networks. We searched for the optimal hyperparameters, including the maximum number of epochs, the initial value of the learning rate, and the maximum validation failures for which the search space ranged from 100 to 2,000, 0.0001 to 0.01, and 4 to 20, respectively.

XGBoosting. We used Extreme Gradient Boosting to train the models and split the training dataset into learning and validation sets 30 times (75%/25%). We employed random search to fine-tune the hyperparameters of interest, including the learning rate, maximum tree depth, and percentage of data used to grow trees, in the range 0.0001-1, 2-10, and 0.1-1.0, respectively. We generated 2000 of these triplets for each iteration during fine-tuning, evaluated them using the validation partition, and kept the best-performing one. We then used the testing set to obtain the values for the hyperparameters.

Support Vector Regression. We tested Support Vector Regression using polynomial kernels of orders one to four and Gaussian kernels. We searched for the optimal hyperparameters, including the maximal number of iterations, tolerance for gradient difference between the current and previous iterations, and box constraint for the trained SVR coefficients. The search ranged from 10^6 - 10^8 , 0.0-0.01, and 0.1-10.0, respectively.

4.2 From reference values to poverty dimensions

Figure 4 summarizes the results obtained for the mapping between reference values and poverty dimensions using linear and non-linear regression models. Figure 4a illustrates the performance for each model in terms of R^2 . XGB outperforms LR for poverty, moderate

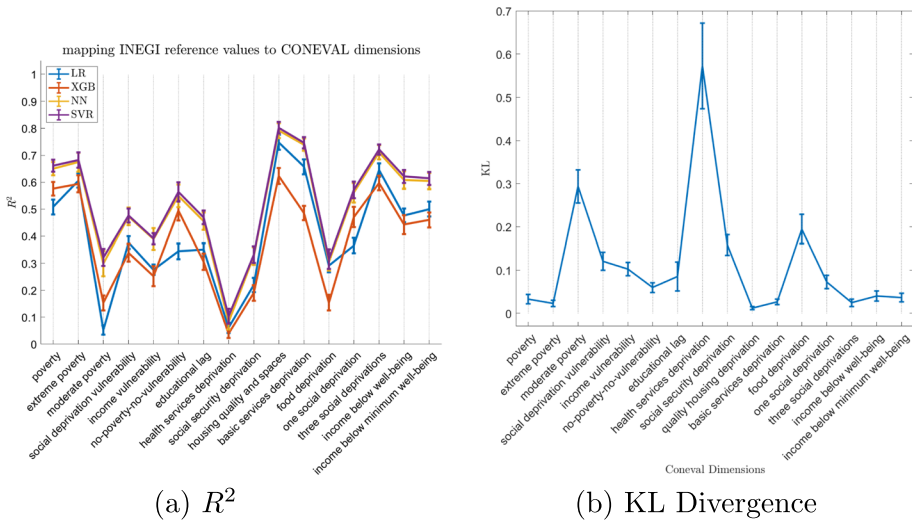


Fig. 4 Mapping census reference values to poverty dimension using linear (LR) and non-linear regressors (NN, SVR, XGB) (refer to the text for details)

poverty, no-poverty-no-vulnerability, and one social deprivation. At the same time, LR does better for indicators such as lack of housing quality and spaces, basic services deprivation, and three social deprivations, all resulting in an $R^2 > 0.6$. However, the non-linear models NN and SVR provide the best performance, with SVR having a slightly higher mean R^2 than NN but overlapping uncertainty bars at one standard deviation. The best mean R^2 performance of 0.802 ± 0.022 (at one standard deviation) was obtained using SVR for the lack of housing quality and spaces indicator. Figure 4b reveals that the regression process has a varying impact on the distributions of the poverty indicator values. Certain response variables, such as health services deprivation, moderate poverty, and food deprivation, exhibit more pronounced divergences than others, such as extreme poverty and lack of housing quality and spaces. This result shows that variables yielding high R^2 values also tend to exhibit less divergence, as the model predictions are more accurate. We hypothesize that the observed discrepancies are attributable to the predictive limitations of the reference values used as predictors. This conjecture necessitates further investigation and acquiring additional predictors to enhance the mapping accuracy for these specific response variables.

We evaluated the performance of various regressors on all the poverty indicators. We selected the SVR with Gaussian kernel as the best option for mapping the lack of housing quality and spaces dimension to the poverty indices (v) used in subsequent experiments. When we refer to poverty, we mean lack of quality and space in housing. Next, we validated the outputs ($v_j^{k,m}$) of the function $F(\mathbf{x}_j^m, \theta^k)$ (see Sect. 3.3) by comparing $\tilde{v}^{k,m}$ values to the corresponding $v^{k,m}$ values of the chosen indicator for all municipalities. The validation test yielded an RMSE of 0.078 and an R^2 of 0.794, which were slightly smaller than those obtained using the function $F(\mathbf{x}^m, \theta^k)$, using the aggregated reference values \mathbf{x}^m , for the test samples. Figure 6a presents a density plot of the values of $\tilde{v}^{k,m}$ versus $v^{k,m}$ for the 2456 municipalities. Next, we describe the mapping of satellite images to the lack of housing quality and spaces dimension.

4.3 From satellite images to poverty inference

We utilized satellite images from the Mexico Geospatial Data Cube for Landsat-7, a sun-synchronous satellite with a 16-day temporal resolution (Irons et al., 2012). Its radiometric resolution consists of eight bits for each band. Based on previous analysis (Salas et al., 2021), we selected six spectral bands: blue (450–515 nm), green (525–605 nm), red (630–690), near-infrared (775–900 nm), SW1 (short-wavelength infrared, 1550–1750 nm), and SW2 (2080–2350 nm) at a spatial resolution of 30 m/pixel. We utilized either the median or the geomedian (Roberts et al., 2017) to aggregate the information yearly. In both cases, the yearly collections of Mexico's images were summarized and combined in a single mosaic of 5000×5000 pixels free of clouds and shadows. For the experiments, we used non-overlapping 20×20 pixel image patches. We reviewed the patches to identify and remove those containing not-a-number (NaN) values, resulting in a database containing 2,164,553 images.

4.3.1 Training a CNN classification model

We fine-tuned EfficientNet-B3, initially trained on ImageNet, employing the satellite images corresponding to Mexican residential blocks. Using bicubic interpolation, we resized the 20×20 pixel image patches to 224×224 pixels. The CNN was trained using the Adam optimizer with a learning rate of 10^{-4} and included a dropout layer with a 0.5 probability, $\beta_1 = 0.9$ and $\beta_2 = 0.999$, during 20 epochs. We initialized the network inputs with Xavier and incorporated a dense layer, including a softmax activation function for two classes. First, we labeled the image patches as either poor or non-poor, denoted by C_j^m , using the $v_j^{k,m}$ indices, setting the threshold $\tau_v = 0.15053$ as in (4). This threshold is the global maximum in the distribution of v , and both classes were roughly balanced, with 52.7% and 47.3% of poor and non-poor samples, respectively. This threshold was also the most frequent value of v when $\|x_j^m\|_1 = 1$, equivalent to all households in a residential block lacking a basic service or having a maximum occupancy. To achieve a balanced dataset, we randomly removed some images corresponding to positive regions, resulting in 50% of regions of each class. We randomly partitioned the image set into training, validation, and test subsets with sample sizes of 50%, 25%, and 25%, respectively.

We evaluated the performance of the model using ROC and Precision-Recall curves (see Fig. 5) with corresponding area under the curve (AUC). The AUC was 0.928 and 0.927, respectively, obtained for the test subset using the model configuration and training hyperparameters described. The accuracy for predicting the classification of an image patch as poor or non-poor was 85.2% in the test dataset at a threshold value of 0.5.

4.3.2 Training CNN regression models

The model described above aims to predict the likelihood that an image corresponds to a poor or non-poor region. To accomplish this, we utilized the softmax function as the output layer, which provides probabilities for each class. We implemented three regression models based on the EfficientNet-B3 architecture. The configuration included a dropout layer having a 0.5 probability, followed by a dense layer with a single output unit. To ensure that the predicted values $\hat{v}_j^{k,m}$ lie in the range of 0–1, the same as the reference values $v_j^{k,m}$, we used

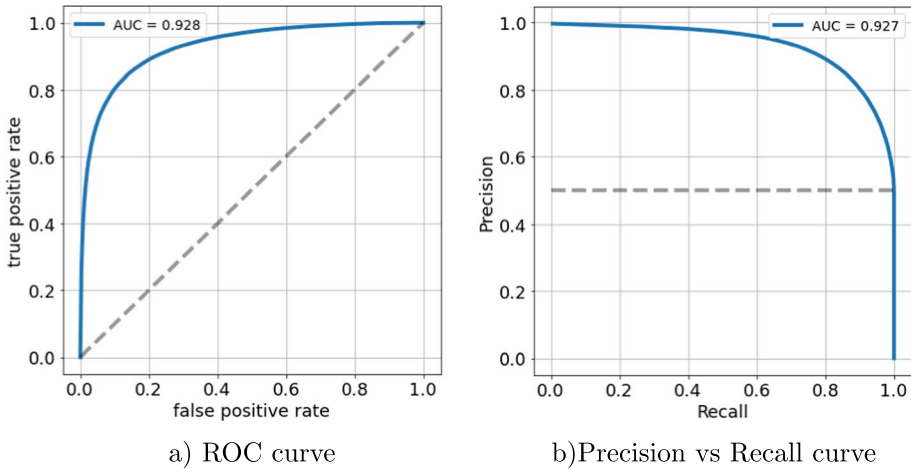


Fig. 5 Performance assessment of convolutional neural networks (refer to the text for details)

the sigmoid activation function for the output layer. The mean squared error (MSE) was the loss function during the training phase.

We trained each CNN network for ten epochs using the Adam optimizer. We used the same partition of the image set to maintain consistency in subsequent comparisons with the CNN classification model. The networks differed in the initialization of their layer weights and learning rates. The first model (A) utilized Xavier initialization and a learning rate of 10^{-4} . For the second model (B), we copied the weights from an EfficientNet-B3 model pre-trained on ImageNet to the layers with the same shape. The non-compatible layers used Xavier initialization, and the learning rate remained 10^{-4} . In the third model (C), we transferred the weights from the CNN trained for classification, excluding the output layer due to its different shape. We used a learning rate of 10^{-5} . The learning rates for each model were selected after several trial runs, where we assessed the best evolution behavior of the loss function during training for both the training and validation sets.

After training the models, we obtained predictions for the poverty index from each model for the test image subset. We compared the predicted values, $\hat{v}_j^{k,m}$, against the corresponding ground truth values, $v_j^{k,m}$, using the coefficient of determination R^2 . Additionally, we determined the classification accuracy of the models by applying the same threshold ($\tau_v = 0.15053$) used for the classification model to both $v_j^{k,m}$ and $\hat{v}_j^{k,m}$ values. The results of these evaluations are presented in Table 3. With an R^2 value of 0.628 for Model C, they demonstrate the advantage of training a classification model initially and subsequently transferring its weights to the regression model. This finding suggests that assessing poverty as a continuous variable poses a more significant challenge for a network than distinguishing between poor and non-poor regions, especially when working with low-resolution images. Furthermore, the corresponding MSE of 0.0088 for the regressor reduces the approximation error in (), which was 0.0103.

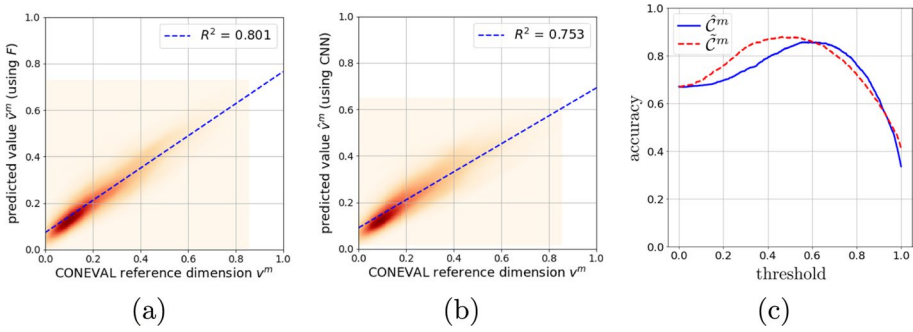


Fig. 6 Results of the validation tests applied to the regression function $F(\mathbf{x}_j^m, \theta^k)$ outputs and the Efficient-Net-B3 inferences (refer to the text for details)

Table 3 Performance of the CNN regression models evaluated with the R^2 and MSE metrics (refer to the text for details)

Model	R^2	MSE	Accuracy ($\tau_v = 0.15053$)
A	0.382	0.0142	75.3%
B	0.511	0.0112	76.8%
C	0.628	0.0088	84.9%

4.4 Aggregating inferences

In this study, we aimed to verify whether the aggregation value preserved the mapping between reference values and the poverty dimension. To accomplish this, we first obtained the outputs $v_j^{k,m}$ of the regression function $F(\mathbf{x}_j^m, \theta^k)$, the referenced class values C_j^m by applying a threshold τ_v to the $v_j^{k,m}$ values, the outputs \tilde{C}_j^m of the trained CNN classification model for each residential block, which corresponded to the probability that people living in a block were poor based on the *lack of housing quality and spaces* dimension, and the outputs $\hat{v}_j^{k,m}$ of the trained CNN regression model. We then computed the weighted averages \tilde{C}^m and \hat{C}^m , for each municipality using (1) and compared them to the corresponding C^m values (see Sect. 3.3). To obtain C^m from $v^{k,m}$, we used as threshold $\tau_v = 0.15053$. We computed the aggregated values $\tilde{v}^{k,m}$ of the regression function outputs and the aggregated predictions $\hat{v}^{k,m}$ of the best CNN regression model (Model C) for each municipality in Mexico. Figure 6a shows the plot of $\tilde{v}^{k,m}$ versus $v^{k,m}$, and Fig. 6b shows the plot of $\hat{v}^{k,m}$ versus $v^{k,m}$, where $v^{k,m}$ represents the CONEVAL dimension of *lack of housing quality and spaces*. We obtained an R^2 of 0.801 for $\tilde{v}^{k,m}$ and 0.753 for $\hat{v}^{k,m}$.

To determine the accuracy, we applied a threshold τ_a to the \tilde{C}^m and \hat{C}^m values (see Fig. 6c). The accuracy was defined as the proportion of correct classifications of the municipalities, either as poor or non-poor. Maximum accuracies of 0.879 and 0.857 were obtained at $\tau_a = 0.463$ and $\tau_a = 0.555$, for \tilde{C}^m and \hat{C}^m , respectively. Although the computation of \tilde{C}^m does not involve the images, determining its accuracy was still

relevant because it was limited only by the MSE of the fitted values \hat{y}_i , and thus represented the best case scenario for the classification problem. From Fig. 6c, we observed that for some threshold values, the accuracy of \hat{C}^m was greater than that of \tilde{C}^m , even though we used the class values C_j^m to train the CNN. However, we averaged the probabilities of the CNN inferences in the first case, while in the second one, we averaged the class values.

4.5 Median versus geomedian

When using satellite optical images, it is common for clouds to occlude a portion of the observation swath. Therefore, it is necessary to aggregate a collection of images to obtain cloud-free area coverage. In this study, we investigate whether summarizing a collection of images with the geomedian or median offers any advantages in determining poverty. To do so, we compared the performance of EfficientNet-B3 using Landsat-7 geomedian and median temporal aggregation of images.

For this experiment, we constructed datasets with a maximum cloud cover of 10% during 2010. The geomedian and median datasets were created using 20×20 pixel patches from the same geographic location. However, due to the presence of null values in Landsat-7 images caused by the zigzag pattern in its scanned area, we had to exclude these patches. The median dataset comprised 2,115,343 patches, with balanced classes accounting for 2,023,204 records. This dataset was then divided into 50% for training (1,011,602 records), 25% for validation (505,801 records), and 25% for testing (505,801 records).

The geomedian dataset contained 2,164,553 patches, with balanced classes comprising 1,707,166 records. This dataset was also divided into 50% for training (853,583 records), 25% for validation (426,791 records), and 25% for testing (426,792 records). We used the same number of epochs and hyperparameters for the median and geomedian datasets. We used the images in the testing dataset to evaluate the CNN model's performance. We obtained an ROC AUC of 0.861 and a Precision-Recall AUC of 0.868 for the median dataset. For the geomedian dataset, we obtained a higher ROC AUC of 0.928 and Precision-Recall AUC of 0.927 (see Table 4).

4.6 Assessing the evolution of poverty

This study on human poverty benefits from the availability of satellite images over time. We employed annual Landsat-7 geomedian images between 2010 and 2020 to generate 20×20 pixel patches for selected states in Mexico, representing all the registered residential blocks in those states according to the 2010 census database. Using the trained EfficientNet-B3 model, we obtained residential block classifications per year. Figure 7 shows

Table 4 The evaluation of the CNNs performance can be measured by ROC AUC and Precision-Recall (refer to the text for details)

Image type	ROC	Precision-recall	Accuracy
Geomedian	0.928	0.927	85.2%
Median	0.861	0.868	78.5%

Bold values highlight better performance

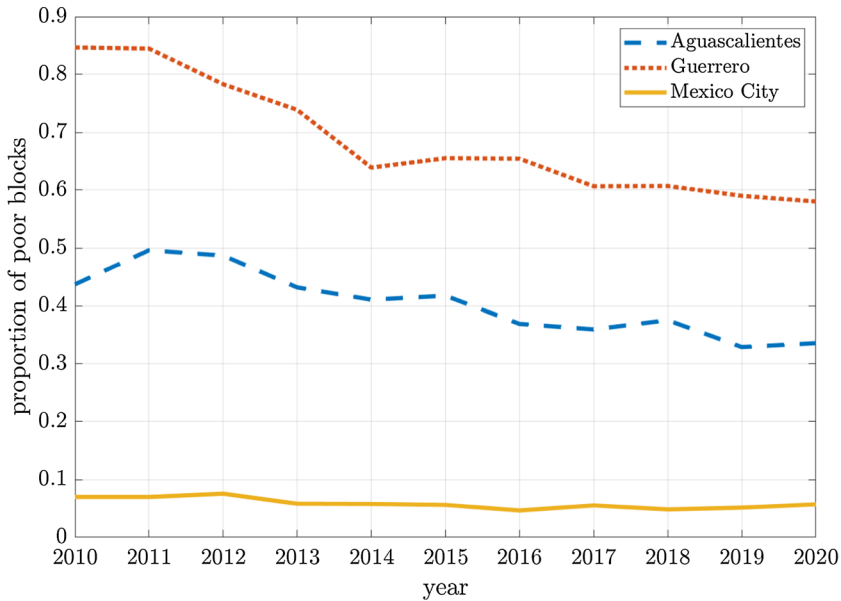


Fig. 7 Changes in poverty for three Mexican states (refer to the text for details)

Table 5 Comparison of the poverty evolution between 2010 and 2020 for Aguascalientes, Guerrero, and Mexico City, represented by the changes in the proportion of poor blocks $\hat{\mathcal{P}}^s$, predicted by the EfficientNet-B3 model for each state

State	$\hat{\mathcal{P}}^s$ (average 2010–2020)	Reduction of 2010–2020 poverty year	Highest	Lowest poverty year
Aguascalientes	38.1%	21.4%	2011	2019
Guerrero	68.6%	31.5%	2010	2020
Mexico City	5.8%	18.7%	2012	2016

the evolution of poverty for Mexico City and the states of Aguascalientes and Guerrero as proportions of poor blocks from 2010 to 2020. The results indicate differences in overall poverty among these states, with average proportions of poor blocks ranging from 5.8% for Mexico City, 38.1% for Aguascalientes, and 68.6% for Guerrero. The model predicted reductions in poor block proportions from 2010 to 2020 of 18.7%, 21.4%, and 31.5% for Mexico City, Aguascalientes, and Guerrero, respectively. However, the years with the highest and lowest poverty values varied among the states. Table 5 summarizes these results. Figure 8 shows the evolution of poverty maps for Acapulco, Guerrero from 2010 to 2020, where the EfficientNet-B3 predictions $\hat{v}_j^{k,m}$ are represented by color.

To ensure the reliability of the yearly poverty assessments from 2010 to 2020 (censuses are programmed every 10 years in Mexico), we evaluated the performance of the EfficientNet-B3 model using image patches from the 2020 geomedians obtained via Landsat-8. We labeled these patches using the poverty dimension inference that resulted from evaluating

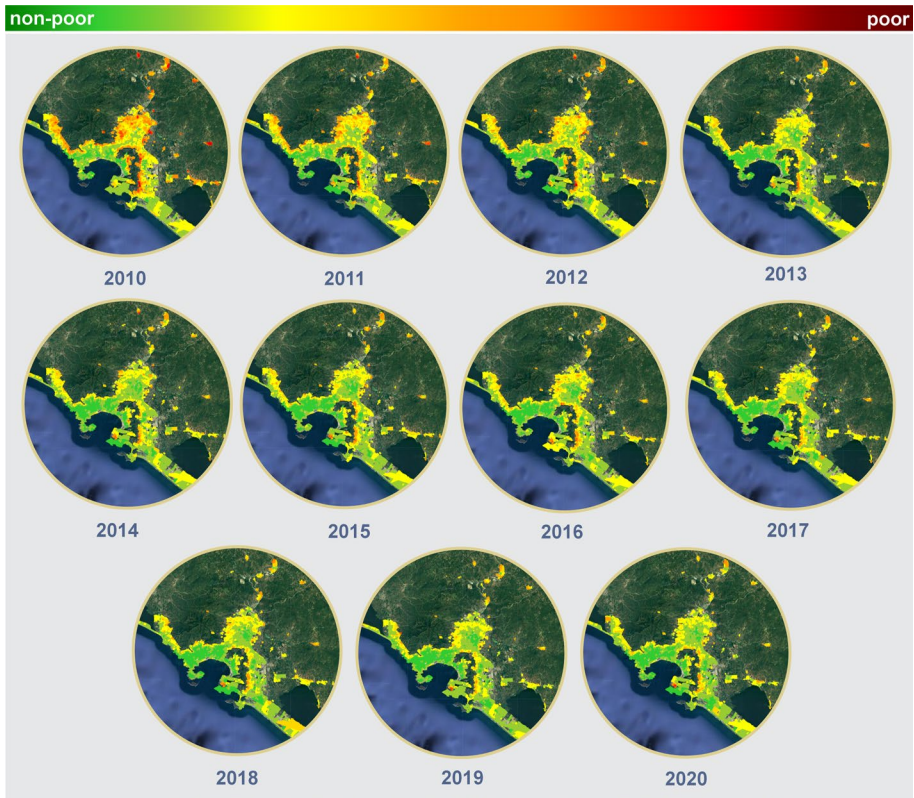


Fig. 8 Evolution of poverty for the city of Acapulco, Guerrero. Spot colors correspond to different EfficientNet-B3 prediction values $\hat{v}_j^{k,m}$ (refer to the text for details). Maps created using Google Earth Engine (GEE, 2024)

the regressors on the 2020 census database, which contained 1,070,046 records. As an intermediate step to obtain the regressor, we trained a classifier by splitting the dataset using a threshold τ_v (see 4). However, since the classes were unbalanced, we randomly selected a subset of 365,576 records to ensure equal positive and negative representation. The ROC AUC and Precision-Recall AUC values were 0.828 and 0.826, respectively, and the model achieved an accuracy of 75.9%. Compared to the performance results on the 2010 census database, the test on the 2020 database showed a reduction of 0.100, 0.101, and 9.3% in ROC AUC, Precision-Recall AUC, and accuracy, respectively.

5 Discussion and conclusion

Remote sensing-based poverty assessments using satellite imagery have gained widespread acceptance. Many studies have employed nightlight (Elvidge et al., 2009), multispectral (de Mattos et al., 2021), high-resolution (Suel et al., 2021),

low-resolution (Salas et al., 2021), SAR images (Wurm et al., 2019), employing classic (Yin et al., 2021) and deep learning (Xie et al., 2016) based techniques. However, as pointed out by Owusu et al. (2021), it is crucial to ensure that the mapping from remote sensing corresponds to actual poverty dimensions. This study focuses on the poverty dimension related to the lack of housing quality and spaces, which factors in aspects directly observable from Earth observations. Other variables related to social welfare and economic deprivation will not create a directly observable causal effect and should be studied with other means.

This research aligns the inference values the machine learning algorithm provided to the reference dimension provided by CONEVAL, the Mexican national agency in charge of measuring poverty (CONEVAL, 2022). The approach achieves higher temporal and spatial resolution and cost reduction.

Using aggregated poverty indicators at the municipal level can be less effective in accurately locating poverty, but it is a step towards addressing privacy concerns (Gai et al., 2022). Since the raw dataset provides reference values at the residential block level, choosing an additive aggregation function was critical in defining aggregation schemes that limit its statistical impact (Gründler & Krieger, 2022). However, the potential differential effect at the local level, especially in sparsely populated areas, requires further consideration to ensure it does not affect subregions significantly. This issue has been highlighted in previous studies on using census data for poverty mapping (Engbo et al., 2020), and it should be addressed to enhance the accuracy of poverty assessments.

This study demonstrates the potential of machine learning techniques to provide reliable and cost-effective information for decision-making at a higher frequency than traditional survey approaches. The results are consistent with previous studies (Salas et al., 2021) and offer a complementary perspective for assessing poverty in Mexico, even though our methodology differs from that of CONEVAL. It is essential to acknowledge that the model's ability to identify the spatial features driving image classification may only capture some factors contributing to poverty, even though the model has been trained with diverse images. Therefore, additional sources of information may be necessary to predict changes in housing quality or changes over time accurately. Combining other sources of information can provide a more comprehensive understanding of poverty.

Identifying settlements with significant deficiencies in access to basic services and precarious housing conditions is crucial for decision-making and the development of effective public policies. These policies strive to alleviate poverty and reduce disparities in communities experiencing high levels of marginalization. However, obtaining this information in a promptly manner has been challenging, as traditional methods involve costly field operations and continuous monitoring. This research proposes an affordable and efficient alternative with improved spatiotemporal resolution, enabling faster tracking of the conditions in these settlements. Consequently, this approach provides more timely information for evaluating and designing public policies.

This research introduces a novel approach for mapping poverty dimensions at the residential block level using multispectral satellite images and census reference values. This approach is particularly valuable because it aligns with the official poverty measurement standards established by the Mexican agency CONEVAL. As a result, it offers a reliable and cost-effective means of obtaining updated poverty assessments by leveraging the newest satellite images and inference from the trained models.

In future work, we aim to improve the accuracy of our poverty assessment for census gap years by leveraging higher-resolution images, such as those from the Sentinel-2

satellite, and validating our results. Additionally, we plan to explore incorporating other sources of information, such as nightlight time images, vegetation indices, and other predictors related to poverty that can be obtained from field surveys, to enhance the performance of the models further.

Acknowledgements Thanks to the anonymous reviewers for their insightful comments. Thanks also to the GEO-Google Earth Engine (GEE) License Programme for facilitating their cloud computing capabilities. Thanks to the National Institute of Statistics and Geography (INEGI), the National Statistical and Geographical Information System (SNIEG), and the INEGI microdata laboratory for facilitating the Population and Housing Census use. The conclusions and opinions expressed in this document are the authors' responsibility and do not represent the official statistics or positions of SNIEG or INEGI.

Funding This research received financial support from from UCMexus-CONACYT for Roberto Manduchi and Joaquín Salas. Also, it was partially supported by SECTEI under grant 201/2021, SIP-IPN 20240619 for Joaquín Salas, and SIP-IPN 20240660 for Pablo Vera. Thanks to CONAHCYT for the scholarship received by Marivel Zea-Ortiz.

Code and data availability The code and data employed in this paper is publicly available at <https://git.inegi.org.mx/laboratorio-de-ciencia-de-datos/vulnerability>

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

Consent to participate Ethics approval was obtained from the Ethics Committee of the Facultad de Ciencias Políticas y Sociales de la Universidad Autónoma de Querétaro.

Consent for publish Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agarwal, P., Garg, N., & Singh, P. (2019). Predicting poverty index using deep learning on remote sensing and household data. *International Journal of Recent Technology and Engineering*, 8(3), 164–168.
- Ajami, A., Kuffer, M., Persello, C., & Pfeffer, K. (2019). Identifying a slums' degree of deprivation from VHR images using convolutional neural networks. *Remote Sensing*, 11(11), 1282.
- Andreato, M. S., Benedetti, R., Piersimoni, F., & Savio, G. (2020). Mapping poverty of latin American and Caribbean countries from heaven through night-light satellite images. *Social Indicators Research*, 156, 533–562.
- Bajotto, A. P., Garcia, L., & Goldim, J. R. (2017). What is vulnerability? A qualitative study about the perception of vulnerability in adults and older adults. *Journal of Clinical Research and Bioethics*, 8(2), 1–5.
- Bansal, C., Jain, A., Barwaria, P., Choudhary, A., Singh, A., Gupta, A. & Seth, A. (2020). Temporal prediction of socio-economic indicators using satellite imagery. *Comad* (pp. 73–81). ACM.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- CONEVAL. (2016). Metodología para la medición multidimensional de la pobreza en México.

- CONEVAL. (2022). *Anexo estadístico de pobreza a nivel municipio 2010 y 2015*. https://www.coneval.org.mx/Medicion/Paginas/AE_pobreza_municipal.aspx.
- Czarnecki, L. (2013). El CONEVAL. La institucionalización del concepto oficial de la pobreza en México. *Barataria*, 16, 177–190.
- de Mattos, A. C., McArdle, G. & Bertolotto, M. (2021). Mapping slums with medium resolution satellite imagery: A comparative analysis of multi-spectral data and grey-level co-occurrence matrix techniques. [arXiv:2106.11395](https://arxiv.org/abs/2106.11395)
- Dorji, U. J., Plangprasopchok, A., Surasvadi, N. & Siripanpornchana, C. (2019). A machine learning approach to estimate median income levels of sub-districts in Thailand using satellite and geospatial data. *Acm Sigspatial International Workshop on AI for Geographic Knowledge Discovery* (pp. 11–14).
- Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., & Bright, E. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, 35(8), 1652–1660.
- Engbo, S., Bull, J. C., Börger, L., Stringell, T. B., Lock, K., Morgan, L., & Jones, O. R. (2020). Census data aggregation decisions can affect population-level inference in heterogeneous populations. *Ecology and Evolution*, 10(14), 7487–7496.
- Engstrom, R., Hersh, J., & Newhouse, D. (2017). *Poverty from space: Using high-resolution satellite imagery for estimating economic well-being*. The World Bank.
- Fisher, T., Gibson, H., Salimi-Khorshidi, G., Hassaine, A., Cai, Y., Rahimi, K. & Mamouei, M. (2021). Deep learning with uncertainty quantification for slum mapping using satellite imagery.
- Gai, N., Xue, K., Zhu, B., Yang, J., Liu, J., & He, D. (2022). An efficient data aggregation scheme with local differential privacy in smart grid. *Digital Communications and Networks*, 8, 333–342.
- GEE. (2024). *Google earth engine developers*. Retrieved May 30, 2024
- Gram-Hansen, B., Helber, P., Varatharajan, I., Azam, F., Coca-Castro, A., Kopackova, V. & Bilinski, P. (2019). Mapping informal settlements in developing countries using machine learning and low resolution multi-spectral data. *Aaa/acm conference on ai, ethics, and society* (pp. 361–368).
- Gründler, K., & Krieger, T. (2022). Should we care (more) about data aggregation? In *European Economic Review*, 104010
- Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6), 989–993.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. In *Ieee Conference on Computer Vision and Pattern Recognition* (pp. 770–778).
- Hersh, Jonathan, Engstrom, Ryan, & Mann, Michael. (2021). Open data for algorithms: Mapping poverty in Belize using open satellite derived features and machine learning. *Information Technology for Development*, 27(2), 263–292. <https://doi.org/10.1080/02681102.2020.1811945>
- INEGI. (2012). Censo Nacional de Poblacion y Vivienda 2012. *INEGI*, 1 (2016).
- Irons, James R., Dwyer, John L., & Barsi, Julia A. (2012). The next landsat satellite: The landsat data continuity mission. *Remote Sensing of Environment*, 122, 11–21. <https://doi.org/10.1016/j.rse.2011.08.026>
- Jean, N., Burke, M., Xie, M., Davis, M., Lobell, D., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790–794.
- Kakareka, A. (2013). What is vulnerability assessment? In J. R. Vacca (Ed.), *Managing Information Security* (2nd ed., pp. 201–221). Boston: Syngress.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4), 541–551. <https://doi.org/10.1162/neco.1989.1.4.541>
- Li, Guie, Cai, Zhongliang, Liu, Xiaojian, Liu, Ji., & Su, Shiliang. (2019). A comparison of machine learning approaches for identifying high-poverty counties: Robust features of DMSP/OLS nighttime light imagery. *International Journal of Remote Sensing*, 40(15), 5716–5736. <https://doi.org/10.1080/01431161.2019.1580820>
- Li, Guie, Cai, Zhongliang, Qian, Yun, & Chen, Fei. (2021). Identifying urban poverty using high-resolution satellite imagery and machine learning approaches: Implications for housing inequality. *Land*, 10(6), 648. <https://doi.org/10.3390/land10060648>
- Lichtner-Bajjaoui, A. (2021). A mathematical introduction to neural networks. Universitat de Barcelona, Master's thesis.
- Lin, J., Luo, S. & Huang, Y. (2022). Poverty estimation at the county level by combining luojia1-01 nighttime light data and points of interest. *Geocarto International*, 37(12), 3590–3606.
- Mackenzie, C., Rogers, W., & Dodds, S. (Eds.). (2014). Introduction: What is vulnerability and why does it matter for moral theory. In *Vulnerability New Essays Ethics Feminist Philos* (pp. 1–29). Oxford: Oxford University Press.

- Owusu, M., Kuffer, M., Belgiu, M., Grippa, T., Lennert, M., Georganos, S., & Vanhuysse, S. (2021). Towards user-driven earth observation-based slum mapping. *Computers, Environment and Urban Systems*, 89, 101681.
- Oxfam (2022). *First Crisis, then Catastrophe* (Tech. Rep.). Oxfam Media Briefing.
- Ravallion, M. (2015). *The economics of poverty: History, measurement, and policy*. Oxford University Press.
- Roberts, Dale, Mueller, Norman, & McIntyre, Alexis. (2017). High-dimensional pixel composites from earth observation time series. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11), 6254–6264. <https://doi.org/10.1109/TGRS.2017.2723896>
- Roy, D., Bernal, D., & Lees, M. (2019). An exploratory factor analysis model for slum severity index in Mexico City. *Urban Studies*, 57, 789–805.
- Salas, Joaquín, Vera, Pablo, Zea-Ortiz, Marivel, Villaseñor, Elio-Atenogenes., Pulido, Dagoberto, & Figueroa, Alejandra. (2021). Fine-grained large-scale vulnerable communities mapping via satellite imagery and population census using deep learning. *Remote Sensing*, 13(18), 3603. <https://doi.org/10.3390/rs13183603>
- Shi, K., Chang, Z., Chen, Z., Wu, J., & Yu, B. (2020). Identifying and evaluating poverty using multi-source remote sensing and point of interest (POI) data: A case study of Chongqing, China. *Journal of Cleaner Production*, 255, 120245.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222.
- Stark, T., Wurm, M., Zhu, X. X., & Taubenböck, H. (2020). Satellite-based mapping of urban poverty with transfer-learned slum morphologies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5251–5263.
- Suel, E., Bhatt, S., Brauer, M., Flaxman, S., & Ezzati, M. (2021). Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. *Remote Sensing of Environment*, 257, 112339.
- Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning* (pp. 6105–6114).
- Tingzon, I., Orden, A., Sy, S., Sekara, V., Weber, I., Fatehikia, M. & Kim, D. (2019). Mapping poverty in the philippines using machine learning, satellite imagery, and crowd-sourced geospatial information. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(4/W19).
- UN-habitat. (2003). *The challenge of slums: Global report on human settlements, 2003*. UN-HABITAT.
- Verma, D., Jana, A., & Ramamritham, K. (2019). Transfer learning approach to map urban slums using high and medium resolution satellite imagery. *Habitat International*, 88, 101981.
- Wurm, M., Stark, T., Zhu, X. X., Weigand, M., & Taubenböck, H. (2019). Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150, 59–69.
- Xie, M., Jean, N., Burke, M., Lobell, D. & Ermon, S. (2016). Transfer learning from deep features for remote sensing and poverty mapping. In *Aaai Conference on Artificial Intelligence*.
- Xie, S., Girshick, R., Dollár, P., Tu, Z. & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Ieee Conference on Computer Vision and Pattern Recognition* (pp. 1492–1500).
- Xu, J., Song, J., Li, B., Liu, D., & Cao, X. (2021). Combining night time lights in prediction of poverty incidence at the county level. *Applied Geography*, 135, 102552.
- Xu, Y., Mo, Y., & Zhu, S. (2021). Poverty mapping in the dian-gui-qian contiguous extremely poor area of southwest china based on multi-source geospatial data. *Sustainability*, 13(16), 8717.
- Yin, J., Qiu, Y., & Zhang, B. (2021). Identification of poverty areas by remote sensing and machine learning: A case study in Guizhou, Southwest China. *ISPRS International Journal of Geo-Information*, 10(1), 11.