

A review of machine learning and satellite imagery for poverty prediction: Implications for development research and applications

Ola Hall¹  | Francis Dompae² | Ibrahim Wahab¹ | Fred Mawunyo Dzanku² 

¹Department of Human Geography, Lund University, Lund, Sweden

²Institute of Statistical, Social and Economic Research, University of Ghana, Legon, Ghana

Correspondence

Ola Hall, Department of Human Geography, Lund University, Lund, Sweden.

Email: ola.hall@keg.lu.se

Funding information

Swedish Research Council, Grant/Award Number: 2019-04253; Riksbankens Jubileumsfond, Grant/Award Number: MXM19-1104:1

Abstract

The field of artificial intelligence is seeing the increased application of satellite imagery to analyse poverty in its various manifestations. This nascent but rapidly growing intersection of scholarship holds the potential to help us better understand poverty by leveraging big data and recent advances in machine vision. In this study, we statistically analyse the literature in the expanding field of welfare and poverty predictions from the combination of machine learning and satellite imagery. Here, we apply an integrative review method to extract key data on factors related to the predictive power of welfare. We found that the most important factors correlated to the predictive power of welfare are the number of pre-processing steps employed, the number of datasets used, the type of welfare indicator targeted and the choice of AI model. Studies that used stock measure indicators (assets) as targets achieved better performance—17 percentage points higher—in predicting welfare than those that targeted flow measures (income and consumption) ones. Additionally, we found that the combination of machine learning and deep learning significantly increases predictive power—by as much as 15 percentage points—compared to using either alone. Surprisingly, we

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Journal of International Development* published by John Wiley & Sons Ltd.

found that the spatial resolution of the satellite imagery used is important but not critical to the performance as the relationship is positive but not statistically significant. These findings have important implications for future research in this domain and for anyone aspiring to use the methodology.

KEYWORDS

deep learning, machine learning, poverty analysis, satellite imagery, welfare

1 | INTRODUCTION

There is a need to measure the welfare of people and nations. The world's population should be counted, measured, weighed and evaluated (Jerven, 2017). Behind these rather raw statements hides a more humanistic perspective where measuring and counting people are in support of humanitarian and developmental efforts, targeting, mapping and monitoring people at risk of food insecurity, famine, poverty and disease (McBride et al., 2021). The effort is supported by the 17 Sustainable Development Goals along with their 169 targets that were adopted by member states of the United Nations as part of the 2030 Agenda (United Nation, 2015). To balance the agenda's economic, social and environmental aspects, more timely, reliable and appropriate ways of collecting and interpreting information on a broad range of human development outcomes are needed (Head et al., 2017). Traditional approaches such as household surveys, often rich in detail but infrequent in time and space, especially in the poor regions of the world, have served the development community with data for a long time are now augmented with new methods and new types of data (Burke et al., 2021). The new approaches are predominantly digital in nature, for example, using cell phone data, harvesting social media including Internet, crowd sourced data, imagery including google street view and other forms of remotely sensed imagery, for measuring welfare and poverty and monitoring the progress towards the attainment of these SDGs. Many are computationally intensive and advanced and many times qualify as 'big data'.

For close to a decade now, the most interesting approach in this area of research has been the combination of satellite imagery (SI) with different machine learning (ML) algorithms, including deep learning (DL) for the estimation of human outcomes. When put to work, wealth and poverty can be estimated from a single satellite image almost as good as achieved from surveys. A recent review of 12 studies evinces that the methodology can predict, for example, the Demographic and Health Survey (DHS) welfare asset index with R^2 between 0.45 and 0.80 (Burke et al., 2021).

This study focuses on studies at the intersection between ML/DL/TL, SI and poverty analysis. The SIML (Satellite Image Machine Learning) methodology combines some of the recent achievements from computer science and object recognition research and applies them to the field of human development research. At a conceptual level, the SIML approach shares many similarities with well-known applications such as learning different object categories from imagery, for example, distinguishing dogs from cats in photographs. For several reasons, it is more complicated to train algorithms to estimate, for example, poverty from imagery. This may be partly because labelled training data are less abundant for human development targets than for everyday objects typically found in image databases such as ImageNet or AlexNet. Another is that poverty and welfare and how to measure them are strongly contested issues (Gibson, 2016), a fact that is not yet adequately reflected in recent works on SIML. Still, this is a rapidly growing area of scholarship.

Like in most other fields of study, context matters in this domain. Some forms of data types perform better than others in different contexts. For example, in poor and extremely poor regions, night-time light (NTL) satellite data have been shown to underperform compared to daytime SI as the former is noted not to vary significantly in such regions (Yeh et al., 2020). This is particularly true for rural regions where the presence of economic activity, including

large farms, does not necessarily mean increasing intensity or even the presence of illumination. At the fundamental level, for a dataset to be useful in this endeavour, there must be sufficient variability in the input variable at different welfare levels and daytime imagery tend to provide that.

The conceptualization and measurement of poverty are a complex endeavour. The increasing success rate of the SIML approach in this domain is influenced by the kind of indicator or measure of poverty that is being targeted. In this regard, it is important to note the different forms and manifestations of poverty or welfare so that in measuring or predicting it, both the so-called 'soft' measures (income, nutrition, food consumption, literacy rates, etc.) and the 'hard' measures (mainly physical assets) are taken into consideration or at least acknowledged. The use of the latter in machine vision processes would be expected to yield higher accuracy than the former in poverty prediction. That is, physical indicators of welfare such as roofing quality and type, infrastructure and farm sizes can more effectively be detected and classified from high-resolution SI than the quantity of meat consumed by a household in the last week. This constrains its applicability to certain types of poverty, which are place based and have a physical manifestation, rather than those measures that are transient in character.

The present study introduces this paradigm of SI/ML/poverty analysis to a wider audience in the humanities, the social sciences and the development community. For anyone interested in this topic, the available literature is difficult to navigate and filled with discipline specific notations. The aim of this study is to quantitatively synthesize the findings and methods in the field, from the perspective of a potential user.

1.1 | Shifting the frontiers of poverty analysis

A major strand of poverty geography studies the distributional characteristics—identification of poor areas and impoverished populations (Zhou & Liu, 2022). Spatial identification of poverty is important to the extent that it can reveal the spatial heterogeneity and geographical character of poverty and thus aid the prioritization of poverty alleviation efforts and resource allocation activities (Erenstein et al., 2010). This is critical given the relative and multi-dimensional character of poverty. The current standard approaches to quantifying welfare—based on face-to-face household interviews—can deliver detailed estimates of poverty, gender roles, the experience of hunger and many other important indicators of poverty. Such surveys on poverty often measure incomes, consumption, expenditure or assets. However, surveys are expensive, time-consuming, error-prone and difficult to scale up beyond the community or site level without substantial financial investment. Two of the more well-known survey programmes are the World Bank's flagship household Living Standards Measurement Survey and the USAID's DHS. With their roots in the 1980s, they have evolved substantially in terms of coverage, methods and technology. For example, GPS receivers are nowadays part of the standard equipment when visiting sample villages. While unprecedentedly rich in indicators, most large-scale surveys, such as these two, tend to suffer from long lags between surveys, limited spatial coverage and high aggregation levels (nation or region), which impede effective monitoring and evaluation of poverty and welfare indicators at the village and household levels where they are most needed (Burke et al., 2021).

Again, given the multidimensional character of poverty, poverty indices are often constructed from these surveys as a single variable would not suffice as an indicator of poverty. A major determinant in the construction of a poverty index is the kinds of data that are available (Zhou & Liu, 2022). Increasingly, the data requirements, as well as approaches to the measurement of poverty, are becoming more sophisticated as the scale of measurement shifts from national and regional levels to the district, town, household and even individual levels. Work in recent decades has explored poverty estimation using remote techniques (Blumenstock, 2016). New forms of data are aiding this transition towards finer resolutions though some have met with limited success. Applications based on NTLs such as Keola et al. (2015) were a step forward. While initial efforts in this domain relied on NTL data—given the strong correlation between nightlight luminosity and traditional measures of economic growth—the application has seen limited application at finer resolutions and in the poorest of regions (Blumenstock, 2016; Jean et al., 2016). Other major approaches include the use of high-resolution daytime satellite data (Head et al., 2017; Jean et al., 2016), mobile phone metadata (Aiken, Bedoya, et al., 2021; Blumenstock et al., 2015), internet search history and social media

activities (Choi & Varian, 2012; Fatehikia et al., 2020; Llorente et al., 2015) and various combinations of these data sources (Pokhriyal & Jacques, 2017; Steele et al., 2017). These applications have been possible largely because of the proliferation of big data as well as newer methods in ML to process them.

1.2 | Application of artificial intelligence to poverty analysis

The introduction of ML, especially DL algorithms, has been instrumental in the application of these new 'big data' sources in poverty studies. An array of ML approaches exists, but the most utilized in this domain of research could be classified as feature extraction algorithms or feature selection ones. Feature extraction methods include Principal Component Analysis, Linear Discriminant Analysis and Singular Value Decomposition. The most popular feature extraction algorithms include Neural Networks, Random Forests, Gradient Boosting, Decision Trees, Naïve Bayes, Gaussian models, Vector Machines and Linear and Logistic Regressions. These algorithms can also be categorized into *unsupervised*—using information that is not labelled to guide the learning of patterns, similarities and differences; *supervised*—using labelled data to train the machine to produce correct outcomes; and *semi-supervised*—working with both labelled and unlabelled data so that the former is used to train the algorithm while the latter is used to make predictions. In this domain of research, the outcome could be whether specific locations, areas, households or individuals are poor or not as well as estimate the degree of deprivation. The choice, utility and propriety of each technique are largely informed by the structure of the dataset and the task at hand.

While NTLs, by themselves, have proven less accurate at differentiating between the poor and the ultra-poor, especially in less-developed regions (Yeh et al., 2020), they have proven useful when used in conjunction with daytime SI, which tends to have higher spatial resolutions. Jean et al. (2016), for instance, successfully trained the DL algorithm Convolutional Neural Network (CNN) on detail-rich daytime SI in which paved roads and metal roofs are visible. In doing so, they developed a technique that shows the relationship between features from daytime SI and night-time imagery—for the latter, lighted areas are indicative of economic activity. Through this approach, the authors were able to predict indicators of poverty at the regional level.

With regard to mobile phone metadata, there is a strong association between mobile phone use and regional distribution of wealth. Eagle et al. (2010), for instance, found that network diversity alone accounted for more than three quarters of the variance in a region's economic status in the United Kingdom. The application of this approach is not limited to developed countries though, given the increasingly high mobile penetration rates even in developing countries. Blumenstock et al. (2015), for instance, constructed a composite wealth index using principal components of various wealth indicators gleaned from the 2007 and 2010 DHS for Rwanda as well as a phone survey and CDR such as calls and text messages. Through this, the authors demonstrate that a mobile phone subscriber's wealth status can be inferred from their historical phone use pattern, with cross-validated correlation coefficients of 0.68. The authors accomplish this through a combination of feature engineering and feature selection to transform phone users' transaction logs into metrics that are then winnowed through various dimension reduction techniques. The authors also demonstrate how alternative supervised learning models, including decision tree-based regressors and classifiers, can produce comparable results. In addition to call and text messaging history, other key CDR features include top-up patterns, handset type and user mobility between and among cell towers (Steele et al., 2017). Indeed, merely owning a mobile device is indicative of a certain level of welfare.

Exponential growth in computing power and more effective and efficient learning algorithms are providing conducive conditions for combining disparate data sources to predict and estimate poverty and welfare. Pokhriyal and Jacques (2017), for example, employed a Bayesian Gaussian Process regression on CDR, SI and environmental data to accurately predict poverty even at the individual level in Senegal. They demonstrate superior prediction accuracy when using such disparate data sources compared to using single datasets. Similarly, Steele et al. (2017) employed multi-

ple data sources—CDR, poverty data and remote sensing covariates such as NTLs, vegetation indices and metrics on distances to roads or urban centres—to predict poverty levels in Bangladesh using hierarchical Bayesian Geostatistical Models.

It is important to note that CDRs are usually not easily accessible from mobile network operators due to commercial rights and privacy concerns. Even with scanty CDR, enough insights can be gleaned when such data are combined with other data sources. Njuguna and McSharry (2017), for instance, demonstrate how such sparse CDR can be combined with per capita mobile handset ownership and call volume per handset, normalized NTL from SI, and population density to estimate a multi-dimensional poverty level in Rwanda, with a cross-validated correlation coefficient of 0.88. The accuracy of these studies is not overly negatively impacted by applying them to multiple countries either. For example, using a variety of datasets including the DHS malnutrition and asset poverty data, remotely sensed solar-induced chlorophyll fluorescence, and precipitation and conflict data, Browne et al. (2021) demonstrate how random forest models can be used to estimate malnutrition and poverty prevalence across 11 countries—Bangladesh, Ethiopia, Ghana, Guatemala, Honduras, Kenya, Mali, Nepal, Nigeria, Senegal and Uganda. Here, we hypothesize that while the combination of methods and data sources can improve the accuracy of prediction of welfare, performance reduces as the application is spread over multiple countries and regions. This may be due mainly to the multidimensional nature of poverty and the different ways in which poverty spatially manifests. Thus, while it is useful to assess how accurately human development can be predicted, it is equally important to evaluate the performance of the different approaches and datasets.

While the performance of models for poverty analysis in this area of scholarship is reported to be on a general ascendancy (Burke et al., 2021), certain factors play a role in the predictive performance. Some important indicators can be measured with higher success than others. Model performance may, thus, be influenced by the type, spatial resolution and nature of the data used in the modelling. Poverty data, for example, are usually sourced from variables based on income, consumption and/or assets. The predictive power of assets has been shown to be higher than consumption and income-based variables (Jean et al., 2016). This is partly because consumption and income levels tend to vary more significantly within shorter periods of time as they relate more directly to harvest outcomes, job losses and gains, and even household size (Steele et al., 2017). Assets such as ownership of mobile phones, tractors and the type and quality of roofing of buildings are more durable and tend to vary less often over time. Even more importantly, some poverty indicators, such as assets, are potentially more easily discernible in high-resolution SI—the so-called hard indicators—than other indicators such as consumption and income.

Furthermore, despite the increasing access to big data, higher resolution datasets are usually more difficult or expensive to access and computationally more taxing to process. Some studies employ pre-processing steps such as pan-sharpening to improve the spatial resolution of satellite data. For example, NTLs, which tend to have lower spatial resolution of about 1 km, can be fused with higher resolution daytime SI to produce a higher resolution composite dataset. Still, ML algorithms have been used to infer individual subscribers' socioeconomic status directly from their individual phone use habits (highest resolution) and then aggregate such predictions to town, district and regional levels (lower resolutions) (Blumenstock et al., 2015). We hypothesize that certain types of assets and features can more effectively be measured in higher resolution imagery than others. This will mean the spatial resolution of the dataset that is used for the analysis becomes important for model performance. Given the trade-off between increasing spatial resolution and accessibility of data, a clearer understanding of the cut-off point at which ML algorithms most accurately predict welfare is most relevant for the further development of the SIML approach.

For the wider user community—social science researchers and the development practitioners—concerned with tackling poverty, inequality and improving welfare, understanding which ML approaches best predict welfare, its accuracy and even at what spatial resolution this can most optimally be done are fundamental questions. Shedding light on these, based on the state of the art in this domain of scholarship, is not only important for future

methodological development of the field but also has long-term policy implications. It can, for instance, help to understanding why there are currently scanty downstream applications of this approach in development (Burke et al., 2021), with a few recent exceptions (Aiken, Bedoya, et al., 2021; Aiken, Bellue, et al., 2021; Blumenstock et al., 2021).

2 | METHODOLOGY

2.1 | Inclusion criteria

Our review method can be described as integrative rather than systematic (Snyder, 2019). This body of knowledge, mixing preprints, working papers, technical reports, peer-reviewed papers and conference papers with contributions of various disciplines is notoriously difficult to capture with one single approach. We used Xie et al. (2015)—one of the first publications to examine the application of ML and SI for measuring poverty and economic well-being—as the benchmark and narrative for our analysis. On this basis, papers completed prior to 2014 and do not apply ML to study socioeconomic wellbeing from SI were excluded from the study. We include literature from published journal articles, grey literature such as working papers and validation studies that have clear empirical application; that is, we excluded reviews of literature. We however limited our inclusion criteria for the year of publication or completion of the drafts of the grey literature. For study design, we included any study that sufficiently describes the application of AI, ML and DL on SI. On selection criteria based on population and geographical location, we had no restrictions, meaning that studies from high-, middle- and low-income countries were eligible for inclusion. For thematic focus, we included studies that explicitly describe or propose either conventional or new ways of measuring the welfare or poverty levels of populations or proxies for doing so within social science disciplines.

We gathered papers from multiple sources using different search words, phrases and topics related to the subject of the study. We focused on the use of SI or data, prediction of socioeconomic welfare indicators within the timeframe specified earlier. Since our interest was on both peer-reviewed papers and grey literature, we did not restrict our search to any specific search engines. However, we accessed papers on the former from Google Scholar and ScienceDirect. Our final database from this search comprised 60 papers from peer-reviewed journal articles, preprints, conference presentations and working papers and other grey literature.

2.2 | Data preparation, regression model specification and description of variables

2.2.1 | Data preparation

Some studies report multiple results for the same metric used in analysing the target outcome, making it difficult to record results for such studies as a single variable. The multiple results may be estimated and reported for different datasets, models, study locations, years or a combination of these. To identify the actual performance of models, where separate results are presented for training and validation datasets (e.g., Hofer et al., 2020), we recorded the results for the latter. For results of different models, we recorded the results of the model with the highest precision (e.g., Mahabir et al., 2020). Where separate results are reported for satellite and ground truth data, the former is used (e.g., Bruederle & Hodler, 2018). Where several models are run to observe control effects, the results of the full model are used (e.g., Bruederle & Hodler, 2018).

A number of studies looked at several target outcome variables. Where studies report multiple indicators, we captured the main target welfare outcome of interest specified or inferred by the authors. Unless otherwise specified, we report the composite target where there is one among the targets of a study. These are usually related to indices of poverty, inequality and related measures or proxies of welfare. And in the absence of such explicit targets as

TABLE 1 Description and summary statistics of study variables

Variable	Min	Max	Mean	Std. Dev.
Proportion of variability explained by models	0.36	0.96	0.75	0.16
Spatial resolution of satellite imagery (metres)	0.05	1000	198.82	2.81
Number of preprocessing methods	0	6	2.25	1.61
Number of datasets used	1	8	3.67	1.68
Target welfare indicator studied (0 if soft, 1 if hard)	0	1	28.33	-
Proportion using Deep Learning (DL)	0	1	0.38	-
Proportion using Machine Learning (ML)	0	1	0.50	-
Proportion using both DL and ML	0	1	0.12	-
Number of countries studied	1	57	5.90	11.45
Year of study or publication	2014	2021	2018	1.81

Note: $N = 60$.

described, we capture data on the target closest or related to the indices of economic wellbeing. The primary targets include asset wealth index, poverty rates, socioeconomic status and slum mapping. Others target socioeconomic indicators (which were rarely captured in our study because of our focus on the main indicators), including access to electricity, NTL, access to water, access to a toilet, educational attainment, monetary income, body mass index and population (see, e.g., Lee et al., 2021; Steele et al., 2017; Tingzon et al., 2019). These are used as proxies for measuring poverty and inequality and yet are not readily observable from daytime SI.

We define target as the outcome variable or what each paper tries to estimate or predict. The target indicators are measured at different levels: individual, household, neighbourhood, village, enumeration area, and so forth. Some authors used multiple indicators for measuring their poverty/welfare outcome variable. In such cases, we extracted the main target outcome reported in such papers for the purpose of our regression model. These were referenced in the title of the study, reported in the abstract or in the main text of the papers.

2.2.2 | Regression model specification and description of variables

What determines the predictive power of the various SIML approaches employed for studying welfare and poverty? Answering this question helps shed light on the relevance for human development and the current methodological complexities and data requirements for predicting welfare using methods other than ground truthing and surveys. We answer the above question using regression analysis. Our dependent variable is the explained proportion of variability in the welfare variables measured using the SIML methods applied by the 60 papers. With our dependent variable measured as a proportion, our interest is in the conditional expectation of the share of variability explained by paper i , y_i , conditional on a vector of covariates, \mathbf{x}_i . We therefore specify the following fractional probit model:

$$E(y_i|\mathbf{x}_i) = \alpha + \beta' \mathbf{x}_i + \varepsilon_i,$$

where $y_i \equiv 0 < y_i < 1$, α is the intercept, β is the coefficient vector associated with the explanatory variables described in Table 1 and ε_i is the random error term. The vector \mathbf{x}_i has seven variables, which are as follows.

Spatial resolution

The first is spatial resolution of the satellite images. The papers reported satellite spatial resolutions in centimetres, metres and kilometres. A majority (about 67%) of the resolutions were reported in metres, and as a result, we

converted all resolutions to metres. We expect the predictive power of a study to be positively correlated with spatial resolution.¹

Number of pre-processing methods

The second variable in the vector \mathbf{x}_i is the number of pre-processing methods used by the various publications. Making meaning and preparing satellite images for analytical purposes as well as linking them to ground truth data requires additional data transformation steps, which are referred to as pre-processing. Depending on the number and complexity of SI involved, one or more methods of pre-processing the data may be required. Therefore, we expect the proportion of variability explained by the papers to be increasing with the number of pre-processing methods used.

Number of datasets used

The vector \mathbf{x}_i also contains the number of datasets used by each paper for predicting welfare. While some of the studies use SI and related datasets, others use a combination of SI methods and ground truth data. We expect the power of the predictions to be increasing as more datasets are used for training.

Target welfare indicator

Different welfare indicators were used in the sample of published papers included in this study. These included household-level measures such as poverty and inequality indices, mobile phone use and expenditure; community and neighbourhood-level indicators such as mapping slums, infrastructure quality and areas lit at night in square kilometres; and city and country-level indicators such as economic development, gross domestic product and employment rates. Given the sample size and the concentration of indicators, we constructed a binary variable that took on the value one for 'soft' welfare indicators (e.g., income, expenditure and quantity of proteins consumed) and zero for 'hard' welfare indicators (e.g., physical assets such as type of roof, rail tracks and bridges). We expect that SI and related methods used by the published papers would more accurately predict 'hard' indicators than 'soft' indicators; that is, we expect this variable to be positively correlated with the predictive power of the models estimated by the papers.

Type of method applied

Fifth, the vector \mathbf{x}_i captures methods of Artificial Intelligence (AI). The 60 papers apply two AI methods—DL and/or ML. Our hypothesis is that the type of method matters for the predictive power of the models presented in the papers. Therefore, we constructed three dummy variables, which are where the study applied (a) DL only, (b) ML only or (c) a combination of DL and ML. We expect that combining both approaches would increase the predictive power of the models over and above using either one of the approaches.

Number of countries

Penultimate, \mathbf{x}_i , also contains a binary covariate that captures whether the datasets used for the studies covered one or more countries. We expect that there would be an inverse relationship between the number of countries included and the performance of the model as spatial poverty tends to manifest differently in various countries.

Year study was published

Lastly, our regression model includes year of publication as a covariate. As knowledge improves and better approaches and methods become available, one could expect prediction accuracy to rise. Thus, we expect prediction power to increase with time, which is why we included the year of publication, expecting a positive and significant association with the proportion of variability explained by the models.

¹We inverted the spatial resolution variable so that high values would intuitively be interpreted as better resolution.

Table 1 shows the summary statistics of the dependent variable and explanatory variables. The predictive power of the models in the 60 articles included in our analysis ranged between 36% and 96% with a mean of 75% and a median of about 79%. Spatial resolution of satellite images reported in the papers ranges from just about 5 cm to 1000m; the mean and median were about 199 and 3.8 m respectively, meaning that the resolutions reported are highly skewed to the right. In fact, if we use 5 m/pixel as benchmark for high resolution, we observe that about 53% of the papers report using high-resolution imagery. Some studies reported no pre-processing; the mean number of pre-processing methods was about two. On average, the studies used nearly four different datasets for their analysis, and less than half of the studies (about 28%) targeted 'hard' welfare indicators.

On type of AI method applied, about half of the studies reported using ML and 38% used DL, and 12% used both. About 58% of the papers were based on country-specific studies, but the mean number of countries covered per study was approximately six. The 60 papers considered in this article were published between 2014 and 2021.

3 | REGRESSION RESULTS

The regression results are presented in Table 2, and the key insights are as follows. First, the mean spatial resolution of SI has a positive but statistically insignificant effect on the predictive power of the published welfare analyses. This is a surprising result because our a priori assumption was that imagery with a higher spatial resolution would be significantly associated with a better prediction of welfare. After examining several functional forms of the variable, the aggregate results from the pool of studies in our sample do not show a significant effect.

Second, the number of preprocessing methods used has a significant positive effect on predictive power at the 0.10 level. An additional method of preprocessing is associated with almost two percentage point increase in the proportion of variation in welfare explained by the papers included in this article.

TABLE 2 Determinants of welfare prediction performance

Variables	(1)	(2)
	Coefficient	Average marginal effects
Inverse spatial resolution of satellite images	0.008 (0.011)	0.002 (0.003)
Number of preprocessing methods	0.052* (0.030)	0.016* (0.009)
Number of datasets used	0.125*** (0.032)	0.038*** (0.010)
Hard vs. soft target welfare indicator	0.607*** (0.130)	0.167*** (0.033)
AI method (reference is deep learning)		
Machine learning	0.212* (0.122)	0.064* (0.036)
Machine learning + deep learning	0.534*** (0.142)	0.138*** (0.032)
Cross-country vs. country-specific studies	-0.036 (0.103)	-0.011 (0.031)
Year of publication	0.015 (0.034)	0.004 (0.010)
Intercept	-30.071 (69.005)	
Observations	60	
Pseudo R-squared	0.050	
Model Chi-squared	63.587	
p value for model test	0.000	

Note: Robust standard errors in parentheses.

Abbreviation: AME, average marginal effect.

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$.

Third, the number of datasets used by a study has a positive and highly significant (p value = 0) correlation with the predictive power of the welfare estimates. This typically means that studies that use a combination of ground truth and satellite data have higher welfare prediction power. Specifically, an additional dataset increases the predictive power of the welfare estimate by about four percentage points.

Fourth, the type of welfare indicator targeted by SI and the AI methods matters. As could be expected, studies targeting 'hard' welfare indicators have higher predictive power than those targeting 'soft' indicators, and the difference in predictive power is significant at the 1% level. Compared with 'soft' indicators, targeting 'hard' indicators is associated with about 17 percentage points higher predictive power; this is a large magnitude of difference.

Fifth, we find that the choice of tool for a study (whether ML, DL or a combination of the two) matters for the predictive power of the welfare models. Using ML increases predictive power by about seven percentage points compared with using DL, but the effect is statistically significant only at the 0.10 level; using a combination of the two increases predictive power by 15 percentage points relative to using DL, and this effect is significant at the 1% level. Similarly, Figure 1 shows that using both machine and DL increases the proportion of explained variation in welfare indicators by about eight percentage points above what could be realized using ML alone (p value = 0.022). This means that combining DL and ML should be preferred to either of them as a single tool in predicting welfare.

Sixth, although we observe, as expected, that the number of countries included in a study is negatively associated with the predictive power of the welfare estimates, the effect is not statistically significant at conventional levels. One reason that may be responsible for this is the lack of standardization of welfare measurements and indicators across countries.

Finally, although we expected a priori that the predictive power of the welfare estimates would have improved over time, we find no evidence of a statistically significant effect even though the expected positive sign is observed.

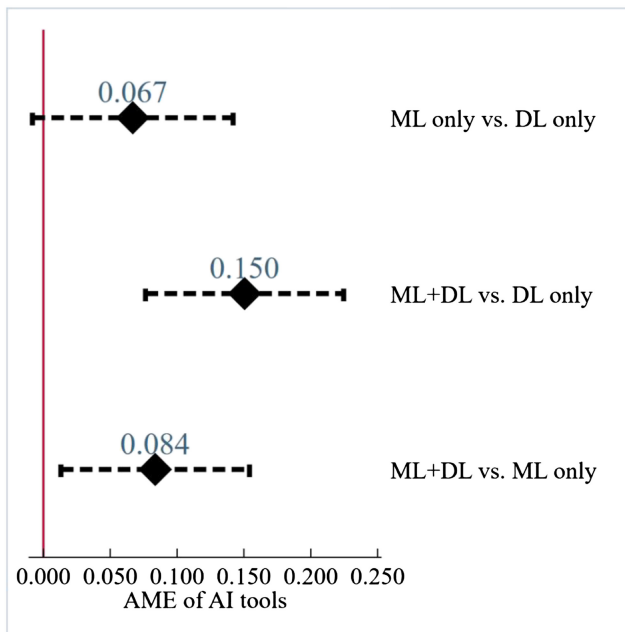


FIGURE 1 Difference in average marginal effects (AME) of AI tools on predictive power. Note: The line caps represent 95% confidence intervals.

4 | DISCUSSIONS

The present paper set out to analyse the state of the art at the intersection of the application of cluster of ML and DL tools on SI to predict poverty or welfare. Key findings from this nascent but rapidly growing field suggests the following. First, our finding that the relationship between the mean spatial resolution of the individual studies and their predictive power is not statistically significant was quite surprising. Conventional wisdom holds that higher resolution SI would contain more abundant information about the landscape and its features that could be correlated with economic activity (Jean et al., 2016). It then stands to reason that training datasets based on such higher resolution imagery would produce more accurate prediction and produce models that have higher predictive power (Engstrom et al., 2022; Head et al., 2017). Our result suggesting a positive but statistically insignificant relationship between spatial resolution and accuracy has important implications. It suggests, for instance, that previous poor results achieved were down to other factors than the unavailability of higher spatial resolution satellite data per se. For researchers, this implies that going forward, additional resources would not need to be expended to acquire higher resolution imagery, which are often only commercially available at high cost (Ayush et al., 2020) and that publicly available SI would suffice in most cases.

While we do not find any evidence of a statistically significant effect that prediction performance using this approach increases over time, Burke et al. (2021) find the contrary. It must be noted, however, that they assessed the performance of these approaches within the specific domains of smallholder agriculture, economic livelihoods, population and informal settlements. They also attribute the improving performance they measure to three main factors: more creative application of advances in computer visions, more abundant and higher quality SI, and more numerous and accurate training datasets. The latter jives with our findings, which suggest that the number of datasets used is positively and statistically significant for prediction performance. In this vein, the increasing proliferation of more accurate and higher quality training datasets portends well for this field of scholarship. Most studies in this area previously relied more heavily on NTLs datasets with coarse spatial resolutions (1 km/pixel) for estimating the level of welfare or development. The cluster of ML approaches recently applied in this intersection has proven to significantly improve predictions that could be achieved using NTL. For example, one of the most important findings from Yeh et al. (2020) was that NTLs tend to perform relatively poorly compared to daytime imagery in predicting asset wealth, largely because the former does not vary sufficiently in poor regions. The review also notes the limited downstream application, which it attributes, in part, to the novelty of the approaches and their lack of interpretability. With regards to the latter, explainable AI is the next rung in the ladder of applying ML to everyday social development issues such as poverty analysis. This requires transparency in model building (Hall et al., 2022). We argue that a necessary, even if insufficient, condition for the development of transparent, explainable and interpretable rather than black box (Rudin, 2019) ML models is adequate domain knowledge, which, in turn, requires co-option of development researchers and practitioners.

An important consideration in the use of SI is the type and number of pre-processing operations that are employed to format images before they are fed to the models for training. In the reviewed papers, the main pre-processing operations include radiometric correction, rotation and flipping, channel re-scaling, normalization, cropping and pan-sharpening. The most-used pre-processing step—pan-sharpening—entails enhancing the lower spatial resolution of multispectral band images by combining them with higher resolution panchromatic images (Hofer et al., 2020). That it is the most employed pre-processing operation is unsurprising given the conventional view that higher spatial resolution datasets invariably translate into better training data and more accurate predictions. However, if it turns out that the relationship between the spatial resolution of the SI and accuracy of the result is positive but statistically insignificant, then pan-sharpening might become a redundant operation. This notwithstanding, other pre-processing steps would remain critical to ensure more accurate prediction of welfare. In the reviewed papers, radiance correction remains key for filtering out ephemeral light sources in DMSP data (Kim et al., 2016), especially in studies that rely on NTLs (Bruederle & Hodler, 2018; Rybnikova & Portnov, 2020; Zhao et al., 2019). Given how fundamental some

of these pre-processing operations are, some SI datasets such as PlanetScope imagery are often radiometrically corrected before delivery to users (Warth et al., 2020), while for others, users need to implement the correction (Duque et al., 2017; Leonita et al., 2018).

Key takeaways from our results are that the total number of datasets used in training of the model, the nature of the target welfare indicator and the specific learning model contribute the most to explaining the level of welfare that can be predicted. With regard to capability of the various learning models, our results suggest that a combination of more conventional ML models and those using DL approaches has the most predictive power in welfare estimation studies using SI data. A vast majority of the reviewed studies employed either ML or DL alone (24 and 30, respectively), with only 7 of them (Gram-Hansen et al., 2019; Hofer et al., 2020; Lee & Braithwaite, 2020; Li et al., 2019; Mboga et al., 2017; Puttanapong et al., 2022) combining both ML and DL in their analyses. These papers combine the approaches differently though. While Li et al. (2019) compare their performance concurrently to determine which performs relatively better, Lee and Braithwaite (2020) combine them in an iterative manner. The latter use the ML algorithm of eXtreme Gradient Boosting (XGBoost) to estimate welfare levels for all populated places in 25 countries and then use this predicted welfare level to train the DL model of CNN. This was done to circumvent the need to use night-time luminosity. Lee and Braithwaite (2020) then fed the featurized information that was the output from the CNN model back to the XGBoost model. They then apply transfer learning from the second iteration onwards to augment learning and speed up the process. Combining these different models in the iterative manner through transfer learning tends to generate better training data which contribute to prediction accuracy (Burke et al., 2021; Head et al., 2017; Hofer et al., 2020).

Other ML models might also contribute to better results in this field. Long Short-Term Memory (LSTM) networks, a class in the recurrent neural networks (RNNs) family, have been shown to be, capable of learning order dependence in sequence prediction problems. This quality makes them most useful in complex machine vision tasks such as predicting poverty from SI. However, thus far, LSTMs have seen limited application in this domain as none of our 60 reviewed papers employed this type of network; CNNs are the most common. Among our final 60 papers, 48% (29 of 60) of the studies employed some form of NN, with 21 of these studies using CNNs as the main model. The limited application of LSTM networks in studies at the intersection of poverty, SI and ML is not too surprising though, as CNNs are predominantly useful for spatial predictions while RNNs such as LSTM networks are more effective at capturing temporal predictions without suffering optimization hurdles, which tend to plague other RNNs (Greff et al., 2016). This, among other advantages, makes LSTM networks ideal candidate models for higher dimensional data analysis tasks such as handwriting recognition, and video and imagery analysis.

More recently, LSTM networks are being used to further enhance the already impressive prediction results achieved by traditional NNs. For example, LSTM models have been applied in the field of infectious diseases to predict the spread of the coronavirus in Bangladesh (Absar et al., 2022). Similarly, in an innovative approach of combining CNNs and LSTMs, Yang et al. (2022) used a hybrid approach to achieve even better prediction results for surface erosion rates prediction at a significantly reduced time and computational costs. This holds great promise for the poverty-satellite data-ML domain of research, and we look forward to more studies adopting this hybrid ML approach in this niche area of research.

Our finding of significantly higher predictive power of models that are based on visible features, the so-called 'hard' indicators, is instructive even if unsurprising given that 'soft' indicators such as income levels, expenditure or the quantity of meat consumed by a household, for example, are more difficult to estimate from an SI than the existence and size of buildings or quality of roofing in a scene. In this sense, these models may be grouped into feature-based algorithms—those that rely on quantifiable geospatial features such as the number of building, length of road, number of junctions and image-based models—those that can recognize the qualitative characteristics of these features (Lee & Braithwaite, 2020). The choice of either of these then comes down to the resolution of the satellite data available since lower resolution SI tends to provide more information about the spatial context such as whether the data are from a rural or urban landscape while higher resolution SI is more useful for extracting the

qualitative characteristics of the features (Kim et al., 2016). This suggests that the increasingly more accurate results obtained by studies in this area of scholarship could be driven more by a combination of improving the spatial resolution of readily available SI data and, perhaps more importantly, the potency and effectiveness of new tools, and approaches as well as the computational power to implement these as Burke et al. (2021) contend. As an illustration, Lee et al. (2021) show the importance of 'hard' indicators such as infrastructure (rail tracks and bridges) as well as other physical features like vehicles, street lights and billboards as indicators of the existence of services and development, and by extension, welfare. It is little wonder that the recent trend is one of combining feature-based and image-based approaches in either sequential or complementary manner to predict welfare (Abitbol & Karsai, 2020; Lee & Braithwaite, 2020; Warth et al., 2020).

5 | CONCLUSIONS

The paper sets out to analyse the state of the art of studies that fall at the intersection of SI (operationalized broadly as all remotely sensed data), poverty analysis and using artificial intelligence tools of machine, deep and transfer learning. Studies in this nascent domain of scholarship are reported to be seeing consistently improving accuracy over the last couple of years, though our results suggest a weak but positive relationship between the spatial resolution of the SI in use and prediction accuracy. The strong explanatory power of models that are based on 'hard' indicators, which are, in turn, more accurate on higher resolution SI data suggests that marked improvements in the tools and computational capabilities are instrumental in these continuous improvements in prediction accuracy in the last few years. It also points to the importance of intermediate data pre-processing steps, especially those related to improving SI resolution such as pan-sharpening. Further progress in algorithms as well as less expensive access to more accurate and numerous training datasets portends well for the field of application of welfare measurement using SI and ML tools. With regard to the specific models and their efficacy in predicting poverty and welfare, we found that a combination of ML and DL algorithms has the best performance, compared to either individual group of models in their own right. This further supports our view that rather than the sheer improvements in the resolution of SI, it is the increasing efficacy of newer tools and models that could be spurring any improving results that we may see in this area. Here, we note the recent advances in LSTM ML approaches (Absar et al., 2022; Rußwurm & Korner, 2017). This holds a promise as LSTMs seem to be specifically well suited for keeping track of long-term dependencies in data, which resonates well with the multi-temporal characteristics of man-made landscapes.

While the application of the SIML approach continues to see improving model performance, more transparency is needed to achieve the next target—explainable AI. This is quite a daunting task given the multidimensional nature, place-based character of poverty. Thus, the differences in spatial manifestation of poverty and its markers in different locations and regions or countries, from a ML model perspective, further complicate this approach. It is mainly for this reason that the combination of multiple datasets or data sources enhances the performance of these models.

ACKNOWLEDGEMENTS

The authors would like to thank financial support during the project from the Swedish Research Council 2019-04253 and Riksbankens Jubileumsfond MXM19-1104:1. We would also like to thank the reviewers for valuable input.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supporting information of this article.

ORCID

Ola Hall  <https://orcid.org/0000-0002-9231-4028>

Fred Mawunyo Dzanku  <https://orcid.org/0000-0002-2271-7876>

REFERENCES

- Abitbol, J. L., & Karsai, M. (2020). Socioeconomic correlations of urban patterns inferred from aerial images: Interpreting activation maps of Convolutional Neural Networks. arXiv preprint arXiv:2004.04907.
- Absar, N., Uddin, N., Khandaker, M. U., & Ullah, H. (2022). The efficacy of deep learning based LSTM model in forecasting the outbreak of contagious diseases. *Infectious Disease Modelling*, 7(1), 170–183. <https://doi.org/10.1016/j.idm.2021.12.005>
- Aiken, E., Bedoya, G., Blumenstock, J. E., & Coville, A. (2021). Program targeting with machine learning and mobile phone data: Evidence from an anti-poverty intervention in Afghanistan.
- Aiken, E., Bellue, S., Karlan, D., Udry, C. R., & Blumenstock, J. (2021). Machine learning and mobile phone data can improve the targeting of humanitarian assistance.
- Ayush, K., Uz Kent, B., Tanmay, K., Burke, M., Lobell, D., & Ermon, S. (2020). Efficient poverty mapping using deep reinforcement learning. arXiv preprint arXiv:2006.04224.
- 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., Karlan, D., & Udry, C. (2021). Using mobile phone and satellite data to target emergency cash transfers. CEGA Blog Post. poverty-action.org/study/using-mobile-phone-and-satellite-data-target-emergency-cash-transfers-togo
- Blumenstock, J. E. (2016). Fighting poverty with data. *Science*, 353(6301), 753–754. <https://doi.org/10.1126/science.aah5217>
- Browne, C., Matteson, D. S., McBride, L., Hu, L., Liu, Y., Sun, Y., Wen, J., & Barrett, C. B. (2021). Multivariate random forest prediction of poverty and malnutrition prevalence. *PLoS ONE*, 16(9), e0255519. <https://doi.org/10.1371/journal.pone.0255519>
- Bruederle, A., & Hodler, R. (2018). Nighttime lights as a proxy for human development at the local level. *PLoS ONE*, 13(9), e0202231. <https://doi.org/10.1371/journal.pone.0202231>
- Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (2021). Using satellite imagery to understand and promote sustainable development. *Science*, 371(6535), eabe8628. <https://doi.org/10.1126/science.abe8628>
- 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>
- Duque, J. C., Patino, J. E., & Betancourt, A. (2017). Exploring the potential of machine learning for automatic slum identification from VHR imagery. *Remote Sensing*, 9(9), 895. <https://doi.org/10.3390/rs9090895>
- Eagle, N., Macy, M., & Claxton, R. (2010). Network Diversity And Economic Development. *Science*, 328(5981), 1029–1031. <https://doi.org/10.1126/science.1186605>
- Engstrom, R., Hersh, J., & Newhouse, D. (2022). Poverty from space: Using high resolution satellite imagery for estimating economic well-being. *The World Bank Economic Review*, 36(2), 382–412. <https://doi.org/10.1093/wber/lhab015>
- 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>
- Fatehikia, 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), 22. <https://doi.org/10.1140/epjds/s13688-020-00235-w>
- Gibson, J. (2016). "Poverty measurement: we know less than policy makers realize." *Asia & the Pacific Policy Studies*, 3(3), 430–442.
- Gram-Hansen, B. J., 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. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society.
- Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- Hall, O., Ohlsson, M., & Rögnvaldsson, T. (2022). Satellite image and machine learning based knowledge extraction in the poverty and welfare domain. *Patterns*, 3(10), 1–15 <https://doi.org/10.2139/ssrn.4102620>
- 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.
- Hofer, M., Sako, T., Martinez, A. Jr., Addawe, M., Bulan, J., Durante, R. L., & Martillan, M. (2020). Applying Artificial Intelligence On Satellite Imagery To Compile Granular Poverty Statistics. In *Asian development bank economics working paper series (Vol. Working Paper No 629)*. Asian Development Bank.
- 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>
- Jerven, M. (2017). How much will a data revolution in development cost? *Forum for Development Studies*, 44(1), 31–50. <https://doi.org/10.1080/08039410.2016.1260050>
- Keola, S., Andersson, M., & Hall, O. (2015). Monitoring economic development from space: using nighttime light and land cover data to measure economic growth. *World Development*, 66, 322–334. <https://doi.org/10.1016/j.worlddev.2014.08.017>

- Kim, J. H., Xie, M., Jean, N., & Ermon, S. (2016). Incorporating spatial context and fine-grained detail from satellite imagery to predict poverty.
- Lee, J., Grosz, D., Uzcent, B., Zeng, S., Burke, M., Lobell, D., & Ermon, S. (2021). Predicting livelihood indicators from community-generated street-level imagery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35, 268–276. <https://doi.org/10.1609/aaai.v35i1.16101>
- Lee, K., & Braithwaite, J. (2020). High-resolution poverty maps in Sub-Saharan Africa. arXiv preprint arXiv:2009.00544. <https://doi.org/10.48550/arXiv.2009.00544>
- Leonita, G., Kuffer, M., Sliuzas, R., & Persello, C. (2018). Machine learning-based slum mapping in support of slum upgrading programs: The case of Bandung City, Indonesia. *Remote Sensing*, 10(10), 1522. <https://doi.org/10.3390/rs10101522>
- Li, G., Cai, Z., Liu, X., Liu, J., & Su, S. (2019). A comparison of machine learning approaches for identifying high-poverty counties: Robust features of DMSP/OLS night-time light imagery. *International Journal of Remote Sensing*, 40(15), 5716–5736. <https://doi.org/10.1080/01431161.2019.1580820>
- Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2015). Social media fingerprints of unemployment. *PLoS ONE*, 10(5), e0128692. <https://doi.org/10.1371/journal.pone.0128692>
- Mahabir, R., Agouris, P., Stefanidis, A., Croitoru, A., & Crooks, A. T. (2020). Detecting and mapping slums using open data: A case study in Kenya. *International Journal of Digital Earth*, 13(6), 683–707. <https://doi.org/10.1080/17538947.2018.1554010>
- Mboga, N., Persello, C., Bergado, J. R., & Stein, A. (2017). Detection of informal settlements from VHR images using convolutional neural networks. *Remote Sensing*, 9(11), 1106. <https://doi.org/10.3390/rs9111106>
- McBride, L., Barrett, C. B., Browne, C., Hu, L., Liu, Y., Matteson, D. S., Sun, Y., & Wen, J. (2021). Predicting poverty and malnutrition for targeting, mapping, monitoring, and early warning. *Applied Economic Perspectives and Policy*, 44, 879–892. <https://doi.org/10.1002/aep.13175>
- Njuguna, C., & McSharry, P. (2017). Constructing spatiotemporal poverty indices from big data. *Journal of Business Research*, 70, 318–327. <https://doi.org/10.1016/j.jbusres.2016.08.005>
- Pokhriyal, N., & Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences*, 114(46), E9783–E9792. <https://doi.org/10.1073/pnas.1700319114>
- 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), 293. <https://doi.org/10.3390/ijgi11050293>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Rußwurm, M., & Korner, M. (2017). Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. Proceedings of the IEEE conference on computer vision and pattern recognition workshops.
- Rybnikova, N., & Portnov, B. A. (2020). Testing the generality of economic activity models estimated by merging night-time satellite images with socioeconomic data. *Advances in Space Research*, 66(11), 2610–2620. <https://doi.org/10.1016/j.asr.2020.09.003>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- 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), 20160690. <https://doi.org/10.1098/rsif.2016.0690>
- Tingzon, I., Orden, A., Go, K., Sy, S., Sekara, V., Weber, I., Fatehikia, M., García-Herranz, M., & Kim, D. (2019). Mapping poverty in the Philippines using machine learning, satellite imagery, and crowd-sourced spatial information. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, XLII, XLII-4/W19, 425–431. <https://doi.org/10.5194/isprs-archives-XLII-4-W19-425-2019>
- United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development. (A/RES/70/1). New York: United Nations Organization Retrieved from sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf
- Warth, G., Braun, A., Assmann, O., Fleckenstein, K., & Hochschild, V. (2020). Prediction of socio-economic indicators for urban planning using VHR satellite imagery and spatial analysis. *Remote Sensing*, 12(11), 1730. <https://doi.org/10.3390/rs12111730>
- Xie, M., Jean, N., Burke, M., Lobell, D., & Ermon, S. (2015). Transfer learning from deep features for remote sensing and poverty mapping. arXiv preprint arXiv:1510.00098.

- Yang, S. D., Ali, Z. A., Kwon, H., & Wong, B. M. (2022). Predicting complex erosion profiles in steam distribution headers with convolutional and recurrent neural networks. *Industrial & Engineering Chemistry Research*, 61(24), 8520–8529. <https://doi.org/10.1021/acs.iecr.1c04712>
- Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, 11(1), 2583. <https://doi.org/10.1038/s41467-020-16185-w>
- Zhao, X., Yu, B., Liu, Y., Chen, Z., Li, Q., Wang, C., & Wu, J. (2019). Estimation of poverty using random forest regression with multi-source data: A case study in Bangladesh. *Remote Sensing*, 11(4), 375. <https://doi.org/10.3390/rs11040375>
- 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>

How to cite this article: 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*, 1–16. <https://doi.org/10.1002/jid.3751>