# TOWARDS UNCOVERING SOCIO-ECONOMIC INEQUALITIES USING VHR SATELLITE IMAGES AND DEEP LEARNING

*Claudio Persello and Monika Kuffer*

Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands



## ABSTRACT

In many cities of the Global South, informal and deprived neighborhoods, also commonly called slums, continue to proliferate, but their locations and dwellers' socio-economic status are often invisible in official statistics and maps. Very high resolution (VHR) satellite images coupled with deep learning allow us to efficiently map these areas and study their socio-economic and spatio-temporal variability to support interventions. This paper investigates a deep transfer learning approach based on convolutional neural networks (CNN) to identify the socio-economic variability of poor neighborhoods in Bangalore, India. Our deep network, pre-trained on a slum classification data set, is tuned towards the prediction of a continuous-valued socio-economic index capturing multiple levels of deprivation. Experimental results show that the CNN-based regression model can explain the socio-economic variability with an $R^2$ of 0.75. The use of additional publicly available geographic information layers allow us to spatially extend the analysis beyond the surveyed deprived area data samples to uncover city-wide patterns of socio-economic inequalities.

***Index Terms*—** urban deprivation, slums, deep learning, convolutional neural networks, remote sensing.

## 1. INTRODUCTION

The rapid urbanization and population growth lead often to the proliferation of deprived neighborhoods in low- and middle-income countries, i.e., settlements with limited or no access to basic infrastructure, sanitization, and adequate housing. UN-Habitat estimates that currently one billion people live in slums, informal and inadequate housing areas (here called deprived areas); this number is expected to grow more and faster if national and local governments do not urgently take counter measures. An information gap that hampers pro-poor policy definition and the planning of interventions in cities of the Global South is about the spatial distributions of deprived neighborhoods and the socio-economic status of their dwellers.

The remote sensing literature [1], shows that VHR images offer us the opportunity to map deprived areas, extract their morphological characteristic and their spatio-temporal variability [2]. To this aim, machine learning techniques combined with various textural features have been investigated in [3], [4]. The recent introduction of deep learning techniques, such as CNNs and fully convolutional networks (FCNs), have showed great potentials for automatically learning the spatial, textural and morphological characteristics of deprived areas and to produce accurate classification maps within an end-to-end learning framework [5], [6]. Most papers cast the mapping problem as a binary classification one, assuming that a sharp boundary can be drawn to separate formal and deprived settlements (in both the feature space and the 2D image domain) and assuming substantial homogenous characteristics among deprived areas. However, there is a large uncertainty and variability associated with the definition of what constitute a deprived area, which characteristics encompass multiple forms of deprivation, including not only the physical (e.g., poor house material) and financial level (e.g., low-income residents), but also human, social and contextual variables (such as accessibility to healthcare, education and other services or social exclusion factors). A framework conceptualizing the multi-dimensional nature of deprivation has been developed in [7] based on the asset vulnerability framework [8]. Ajami *et al.* adopted such a framework to introduce a data-driven approach to summarize multiple deprivation variables (both categorical and real-valued) into a single real-valued socio-economic index, named data-driven index of multiple deprivation (DIMD) [9]. As such, this approach overcomes the limitations of previous deprivation indices, which relied on weighted indictors that are sensitive to the choice of individual weights.

This paper builds upon the work of Ajami *et al.*, investigating a deep-transfer-learning-based approach to map the DIMD socio-economic index using VHR images and geographic information layers. Moreover, we extend the spatial extend of our analysis beyond the available sample data using publicly available data, e.g., OpenStreetMap (OSM), VIIRS night-time lights, WorldPop data, census, and data from the Demographic and Health Surveys (DHS), to extrapolate the socio-economic variability at city scale.

## 2. MATERIALS AND METHODS

The method consists of three steps. The first step is about the analysis of socio-economic data, including a household data set (HH) and an in-situ quick-scan (QS) survey for the development of the DIMD index. The second step consists in a CNN-based transfer learning approach to predict DIMD values for the QS data set based on VHR images and GIS-extracted features. The third step consists in the extrapolating of the socio-economic variability over the entire urban area.

### 2.1. Study site and data set description

Our method is applied to Bangalore, a large Indian city with a population of more than ten millions inhabitants. Bangalore is growing rapidly, attracting considerable investments from the ICT sector. However, citizens are not equally benefitting from these investments and the economic growth has been accompanied by the proliferation of deprived areas. According to the census, around 8% of the population is living in slums, which is a large under reporting of the actual deprived population in around 1,500 slums (as compared to the 600 officially recorded slums).

The data for our study consist in two sets of socio-economic data: 1) a detailed data set of 1114 households survey from 37 notified slums in 2010 (HH data), and 2) a less detailed quick-scan (QS) data set with primary data collected from 121 slums in August 2017 covering physical and contextual domains of deprivation. The QS surveys are based on 35 categorical indicators (Table 1). Moreover, we have access to delineated boundaries of 1461 slums from 2017. The study uses also four Pleiades pan-sharpened satellite images with a spatial resolution of 0.5 m, three of them are acquired in March 2016 and one in March 2015 (see Figure 1). Furthermore, we use freely available spatial data and features extracted from OSM, WorldPop, census and the DHS data sets.

**Table 1. QS indicators.**

| |
|---|
| Dominant building type, Number of floors, Dominant building footprint size, Wall material, Roof material, Dominant shape of building, Overall state of buildings, Overall building appearance, Open spaces/green spaces, Appearance of open space, Presence of roads, Road pavement, Road material, Road width, Cables for electricity, Presence of footpaths, Footpath material, Streetlight, Pollution (smell, noise, waste), Open sewers, Presence of public toilet, Waterbody, Economic activities, Type of economic activities, Dominant land use around the slum, Feeling safe?, Are people interacting?, Are there vehicles visible?, Temple, Clothes of people, Having jewelry?, Hair of children, Children toys |

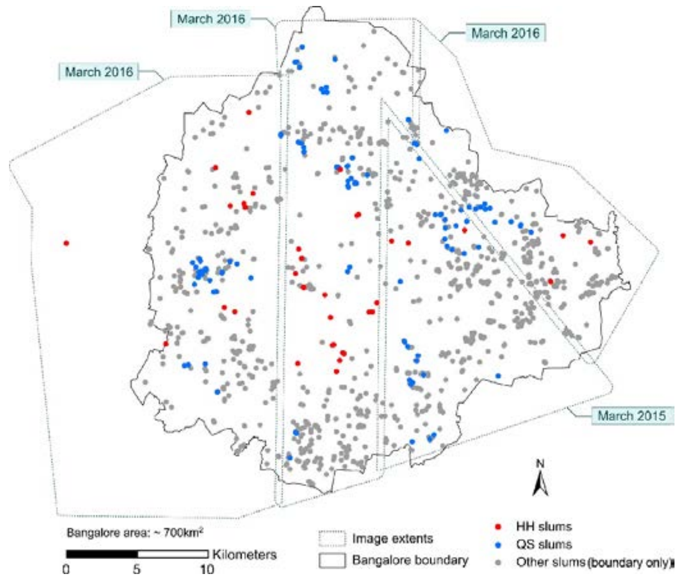


Figure 1. Available Pleaiades images over Bangalore city. Source [9].

### 2.2. DIMD socio-economic index

To aggregate and summarize the indicator values of both HH and QS, we adopt the DIMD approach, which is based on transforming the input indicators using multiple correspondence analysis (MCA). MCA operates on categorical data extracting the underlying structure in the data set. It represents the data as points in low-dimensional Euclidean space and extract the most meaningful directions similarly to principal component analysis. The considered DIMD index is calculated as the first dimension created by MCA. See [9] for more details.

### 2.3. CNN-based transfer learning

We built a deep-transfer-learning model to learn informative spatial-contextual features and predict the DIMD socio-economic index from VHR image patches. However, the 121 QS samples with corresponding target DIMD values are not sufficient to train a deep CNN. We therefore adopted a two-stage transfer learning approach, taking advantage of the larger set of available slum boundaries (1461 samples). We pre-trained our CNN to address the binary classification problem to separate deprived areas from formal settlements using the 1461 available slum boundaries and additional 611 polygons of formal areas. We extracted 2000 labeled patches of various sizes (99, 129, and 165 pixels) from the VHR images and reference polygons and used them for training our CNN with a cross-entropy loss function. Our network architecture is inspired by VGG-16 [10], but it is lighter, it is extended to accept 4-channel multispectral images as input and uses batch normalization instead of local response normalization. More details about the adopted network architecture are shown in Figure 2, including number of layers and filter sizes.


Figure 2. CNN architecture. Source [9].

After optimizing the hyper-parameters (e.g., patch size, learning rate) and training our CNN for the binary classification problem, we turned it into a regression model by changing the loss function to Euclidean (L2). In this second training phase, the CNN is trained to predict the continuous-valued DIMD index for the available 121 samples. We finally built an ensemble regression model using CNN-extracted features together with additional hand-crafter and GIS features.

### 2.4. Spatial extrapolation

To extrapolate the socio-economic patters observed in the area of the QS data to the entire urban area of Bangalore, a set of covariates was selected from publicly available data to predict DIMD values. The selection of covariates was based on previous studies [11], [12], relying only on open-access data (extracted from OSM, freely available Earth Observation (EO) data and products as well as DHS data) the specific local geography and includes, Visible Infrared Imaging Radiometer Suite (VIIRS) night-time light (NTL) data, roads, facility and services data, location of water bodies, health related and population data. First, the covariates are checked whether they correlate with the DIMD socio-economic index. Second, a multiple regression model is built based on the most correlating covariates. Third, the regression model is used to predict the socio-economic patters for the entire city of Bangalore. For this prediction, a gridded system of 200 × 200 m cell size is used. This allow us to generate a city-level map of the socio-economic status.

## 3. RESULTS

This section presents the results of 1) the DIMD socio-economic index for the areas of the QS data, 121 samples of all deprived areas in the city of Bangalore, 2) the predictions of the CNN-based transfer learning model for the QS data, and 3) the result of the spatial extrapolation to generate a city-wide map of socio-economic variability.

### 3.1. DIMD socio-economic index

The DIMD index (Figure 3), representing the first component of the MCA, shows that negative values indicate the most deprived areas (often the temporary settlements),

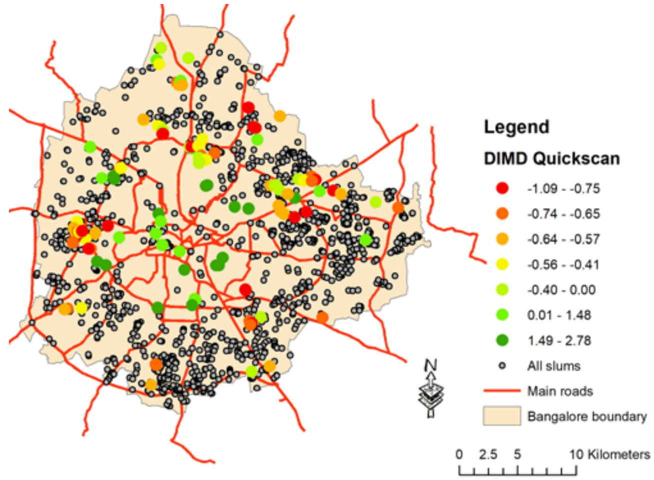


Figure 3. DIMD socio-economic index for the QS locations.

while positive values indicate the less deprived areas (often upgraded slums). Figure 4 shows ground photos of two areas, having the lowest and the highest DIMD values, respectively. The example of the negative DIMD shows a temporary settlements with low accessibility and centrality, while the positive DIMD sample is an upgraded slum with much higher accessibility and centrality. The covariates with highest correlation with the DIMD index are service density, road density, population density, amount of NTL (VIIRS) and distance to center.

### 3.3. Results of the CNN-based transfer learning model

The accuracy of our CNN-based regression model is assessed through a 10-fold cross-validation, calculating the coefficient of determination $R^2$. Our best model, using a 3rd order polynomial function reaches an $R^2 = 0.75$, showing that CNN-based image features can explain 75% of the DIMD variability (Table 2).

### 3.3. Results of the spatial extrapolation

The predictions of the multiple regression model using the public data sets (with selected covariates) have a moderate $R^2$ of 0.47. Table 3 shows that centrality is most important features (i.e., highest value of standardized coefficient), followed by the amount of (NTL), density of services (health and education), population density and road density being the least important. Taking the example of centrality and NTL, as closer to the center and as more NTL, the larger the DIMD index.

DIMD index = -1.09 DIMD index = 2.72

Figure 4. Ground photos and corresponding DIMD values.

Table 2. CNN-based regression of DIMD index.

| $R^2$ | RMSE | BIAS |
|---|---|---|
| 0.75 | 0.53 | 0.20 |

Based on the result of the regression model, the obtained coefficients are used to make a prediction of the socio-economic conditions based on the five covariates for the entire city. Fig. 5 shows the result of the model, which allows a prediction of the non-surveyed slums but also provides a general understanding of all areas across the city.

Table 3. Regression coefficients of the selected covariates.

| | Coeff. | Stand. Coeff. | Var. Infl. Fact. (VIF) |
|---|---|---|---|
| **(Constant)** | 0.28710 | | |
| **Services** | 0.07737 | 0.166 | 2.091 |
| **Road Density** | 0.02273 | 0.114 | 1.724 |
| **Pop Density** | -0.00193 | -0.146 | 2.372 |
| **VIIRS_NTL** | 0.00019 | 0.239 | 1.644 |
| **Centrality** | -0.00019 | -0.485 | 1.876 |

## 4. DISCUSSION AND CONCLUSION

Previous studies have shown that EO data allow one to map the location, extent and physical appearance of deprived neighborhoods. Our study goes beyond that, showing that deep-learning-based models using EO data can capture also the socio-economic conditions of deprived areas at city scale. Our deep-transfer-learning approach shows the capability to explain 75% of the variation in the DIMD index. This allows to predict socio-economic conditions for urban areas without a detailed (and expensive) ground survey. EO data combined with open-access spatial data allow to extrapolate socio-economic conditions over large urban areas.

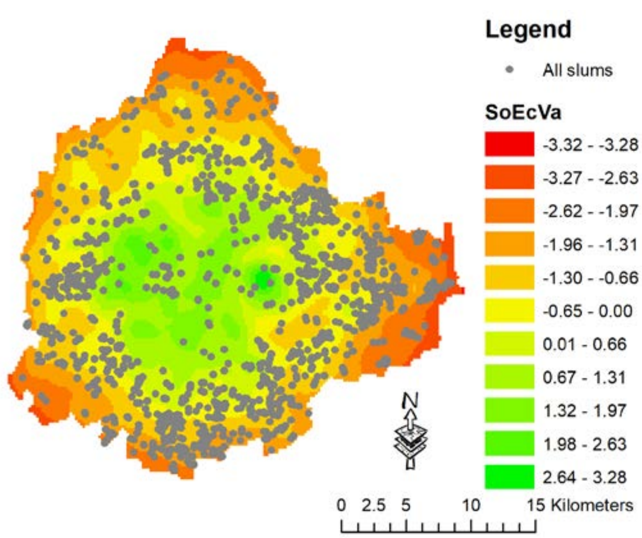


Figure 5. Socio-economic variability (SoEcVa) index

Overcoming issues of transferability and scalability, is in fact a major limitation of current urban studies in deprived areas. Reliable and updated city-wide information on the socio-economic conditions of deprived neighborhoods and their variability is an important piece of information for planning interventions and allocating resources. This information is also essential for analyzing vulnerabilities in the context of disaster preparedness and serves as contextual information for health studies as well as for the monitoring of the progress and supporting the achievement of the sustainable development goals (SDG) 1 and 11.

## 5. REFERENCES


[1] M. Kuffer, K. Pfeffer, and R. Sliuzas, "Slums from space-15 years of slum mapping using remote sensing," *Remote Sens.*, vol. 8, no. 6, 2016.

[2] R. Liu, M. Kuffer, and C. Persello, "The Temporal Dynamics of Slums Employing a CNN-Based Change Detection Approach," *Remote Sens.*, vol. 11, no. 23, p. 2844, Nov. 2019.

[3] M. Kuffer, K. Pfeffer, R. Sliuzas, and I. Baud, "Extraction of Slum Areas From VHR Imagery Using GLCM Variance," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 5, pp. 1830–1840, 2016.

[4] L. P. A. Ella, F. Van Den Bergh, B. J. Van Wyk, and M. A. Van Wyk, "A comparison of texture feature algorithms for urban settlement classification," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2008, vol. 3, no. 1.

[5] C. Persello and A. Stein, "Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2325–2329, Dec. 2017.

[6] N. Mboga, C. Persello, J. Bergado, and A. Stein, "Detection of Informal Settlements from VHR Images Using Convolutional Neural Networks," *Remote Sens.*, vol. 9, no. 11, p. 1106, Oct. 2017.

[7] I. Baud, N. Sridharan, and K. Pfeffer, "Mapping Urban Poverty for Local Governance in an Indian Mega-City: The Case of Delhi," *Urban Stud.*, vol. 45, no. 7, pp. 1385–1412, Jun. 2008.

[8] C. O. N. Moser, "The asset vulnerability framework: Reassessing urban poverty reduction strategies," *World Dev.*, vol. 26, no. 1, pp. 1–19, 1998.

[9] A. Ajami, M. Kuffer, C. Persello, and K. Pfeffer, "Identifying a Slums' Degree of Deprivation from VHR Images Using Convolutional Neural Networks," *Remote Sens.*, vol. 11, no. 11, p. 1282, May 2019.

[10] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[11] R. Lilford *et al.*, "Because space matters: Conceptual framework to help distinguish slum from non-slum urban areas," *BMJ Glob. Heal.*, vol. 4, no. 2, Apr. 2019.

[12] R. Mahabir, P. Agouris, A. Stefanidis, A. Croitoru, and A. T. Crooks, "Detecting and mapping slums using open data: a case study in Kenya," *Int. J. Digit. Earth*, pp. 1–25, Dec. 2018.