

ANALYSING THE RELATIONSHIP BETWEEN IMAGE-BASED FEATURES AND SOCIO- ECONOMIC VARIATIONS OF SLUMS

ALIREZA AJAMI

Enschede, The Netherlands, February 2018

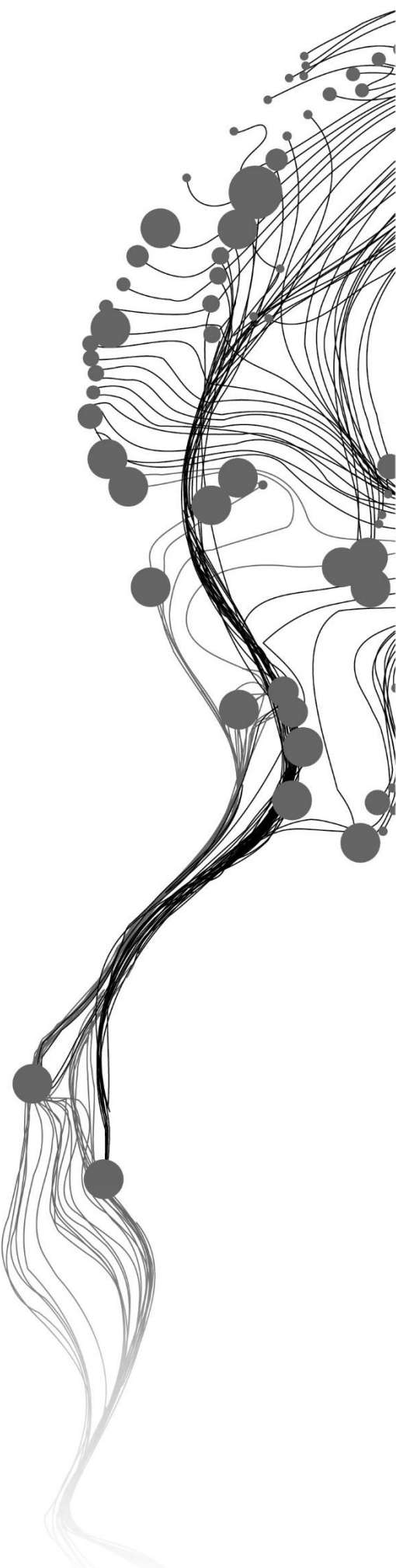
SUPERVISORS:

Dr. M. Kuffer

Dr. C. Persello

ADVISOR:

Prof. dr. K. Pfeffer



ANALYSING THE RELATIONSHIP BETWEEN IMAGE-BASED FEATURES AND SOCIO- ECONOMIC VARIATIONS OF SLUMS

ALIREZA AJAMI

Enschede, The Netherlands, February 2018

Thesis submitted to the Faculty of Geo-Information Science and Earth
Observation of the University of Twente in partial fulfilment of the
requirements for the degree of Master of Science in Geo-Information Science
and Earth Observation.

Specialization: Urban Planning and Management

SUPERVISORS:

Dr. M. Kuffer

Dr. C. Persello

ADVISOR:

Prof. dr. K. Pfeffer

THESIS ASSESSMENT BOARD:

Prof. dr. R.V. Sliuzas (Chair)

Dr. M. Netzband (External Examiner, University of Wuerzburg)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

Slum settlements are growing in the cities of the Global South, but their information is usually hidden in the official documents. Studies have been developed to address the problem of detecting slums from satellite images. However, few studies focused on capturing variations of such settlements from above. This study aims to explore the relationship between image-based features derived from Very High Resolution satellite images and the socio-economic variations of slums in Bangalore, India. The study develops indices based on the Index of Multiple Deprivation and conducts a data-driven approach using Multiple Correspondence Analysis to characterise slums and demonstrate their relative differences in terms of deprivation. The study takes advantage of deeply learned features by developing a system, entirely based on Convolutional Neural Networks to predict deprivation indices. As values of the deprivation indices are known for few samples, a two-step approach is developed to train the CNN. Distinctive features are learned with the aim of classifying slums from formal areas. Then, the trained network is fine-tuned with the limited samples to directly predict the deprivation indices. In parallel, the study implements Principal Component Regression models using manually extracted hand-crafted and GIS-based features to predict the indices and to compare the result with the CNN performance. The ability to combine the CNN, hand-crafted, and GIS features is also examined to create models with a better prediction power. All the models are developed with the ability to capture deprivation even from tiny slums with few settlements. It is found that the physical and contextual domains of deprivation mainly characterise the deprivation levels of the slum settlements in Bangalore. The CNN-based model predicts the deprivation indices with the R^2 of 0.67. Relying only on hand-crafted and GIS features, the study obtains the R^2 of 0.52 to predict the deprivation indices. The non-linear model which is developed by this study combines fully-automated deeply learned with manually extracted features to detect the indices with the R^2 of 0.75. The study finds that using the two-step learning process, a deep CNN can be trained with a limited number of samples to predict variations of slums. The study concludes that hand-crafted features can hardly bring any improvement to the CNN performance, showing the features learned by the CNN are sophisticated enough to be used for the analysis. However, by adding the GIS layers to the CNN features, the performance of the model to predict the deprivation indices is improved. This shows the added value of the GIS layers by bringing the spatial information to the model.

Keywords: deprivation, convolutional neural networks, slum, deep learning, very high resolution satellite imagery

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Dr Monika Kuffer and Dr Claudio Persello, my research supervisors, and Prof. Karin Pfeffer, my research advisor for their patient guidance, enthusiastic encouragement, and useful critiques during this research. I would like to express my great appreciation to Ms Chloe Pottinger-Glass who carried out the fieldwork and collected the data used to build the QS index.

I would like to offer my special thanks to Ms Champaka Rajagopal and Dr Debraj Roy for their help to have a deeper local insight into the available data and slums in Bangalore.

I am particularly grateful for the technical assistance given by Drs Petra Budde to order satellite images used for this research. I wish to acknowledge the advice provided by Dr Javier Martinez and Mr Eduardo Perez Molina which helped for the statistical analyses of this research.

I would like to acknowledge the support from the NWO/Netherlands eScience Centre funded project Dynaslum - Data Driven Modelling and Decision Support for Slums - under the contract number 27015G05.

Lastly, I would like to express my deepest appreciation to my wife, Maryam, for her unfailing support and continuous encouragement throughout this research.

TABLE OF CONTENTS

| | |
|---|-----|
| List of figures | v |
| List of tables | vi |
| Abbreviations | vii |
| 1. Introduction | 1 |
| 1.1. Background and justification..... | 1 |
| 1.2. Research gap identification..... | 2 |
| 1.3. Research objective..... | 3 |
| 1.3.1. General objective | 3 |
| 1.3.2. Specific objectives..... | 3 |
| 1.4. Research questions | 3 |
| 1.5. Hypothesis | 4 |
| 1.6. Thesis structure..... | 4 |
| 2. Literature review | 5 |
| 2.1. Detecting slums from above | 5 |
| 2.2. Mapping deprivation..... | 6 |
| 2.3. Approaches to building an index..... | 8 |
| 2.3.1. Multiple Correspondence Analysis | 9 |
| 2.4. Capturing variations of deprivation | 10 |
| 2.5. Deep learning and Convolutional Neural Networks | 11 |
| 2.6. Research concepts | 13 |
| 3. Study area and data description | 15 |
| 4. Methodology | 17 |
| 4.1. Fieldwork data..... | 18 |
| 4.2. Understanding data and adopting deprivation indicators | 19 |
| 4.3. Understanding slum variations and building deprivation indices | 20 |
| 4.3.1. Relation between HH and QS results | 21 |
| 4.4. CNN-based system to predict deprivation indices..... | 21 |
| 4.4.1. Sample preparation..... | 22 |
| 4.4.2. Image preparation..... | 24 |
| 4.4.3. Patch extraction..... | 26 |
| 4.4.4. Training CNN – simple model..... | 27 |
| 4.4.5. Training CNN – deep models | 27 |
| 4.5. Supplementary hand-crafted and GIS features | 28 |
| 4.5.1. Spectral information..... | 29 |
| 4.5.2. GLCM..... | 29 |
| 4.5.3. LBP..... | 29 |
| 4.5.4. GIS layers | 30 |
| 4.6. Regression models..... | 31 |
| 4.6.1. Regression using CNN output | 31 |
| 4.6.2. Fine-tuning CNN to predict indices..... | 31 |
| 4.6.3. Regression models with hand-crafted and GIS features..... | 31 |
| 4.6.4. Combining results | 32 |
| 5. Result and discussion | 33 |
| 5.1. Build indices using MCA | 33 |
| 5.1.1. HH index..... | 33 |
| 5.1.1.1. Interpreting HH individuals (households)..... | 34 |
| 5.1.1.2. Interpreting HH variables (indicators) | 37 |
| 5.1.2. QS index | 39 |

| | | |
|--------|--|----|
| 5.1.3. | Relationship between HH and QS | 41 |
| 5.1.4. | Discussion on classical indexing | 42 |
| 5.2. | CNN performance | 44 |
| 5.3. | Connecting image-based features to deprivation indices | 45 |
| 5.3.1. | Regression using CNN softmax layer | 46 |
| 5.3.2. | Transfer learning – CNN as a regression model..... | 46 |
| 5.3.3. | Regression with hand-crafted and GIS features..... | 47 |
| 5.3.4. | Combining PCR with CNN features..... | 48 |
| 5.3.5. | Ability to generalize the results..... | 52 |
| 5.4. | Reflection on findings | 53 |
| 6. | Conclusion and Recommendations..... | 55 |
| 6.1. | Recommendations for further studies | 56 |
| | List of references | 57 |
| | Appendix..... | 63 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1: Simple illustration of ANN | 11 |
| Figure 2: Simple illustration of CNN..... | 12 |
| Figure 3: Research concepts | 13 |
| Figure 4: Bangalore and delineated slum locations 2017 | 15 |
| Figure 5: Slum in Bangalore | 15 |
| Figure 6: Extents of available images..... | 16 |
| Figure 7: Research steps | 17 |
| Figure 8: Relation between available data | 18 |
| Figure 9: An example of a slum in 2010 which was re-developed in 2017 | 19 |
| Figure 10: Example of correcting slum polygons..... | 22 |
| Figure 11: Small and large tessellations for sampling formal areas | 23 |
| Figure 13: Samples after generating and erasing buffer..... | 23 |
| Figure 12: Example of formal sample polygon..... | 23 |
| Figure 15: Extent of the eight images after the division | 24 |
| Figure 16: Example of a tiny slum | 24 |
| Figure 17: Original and ground truth raster after doing sampling steps | 25 |
| Figure 18: Steps to generate required data for generating patches | 26 |
| Figure 19: Patch size on a slum sample..... | 26 |
| Figure 20: Architecture of simple CNN..... | 27 |
| Figure 21: Architecture of VGG-like CNN | 28 |
| Figure 22: four directions of GLCM with 1-pixel lag..... | 29 |
| Figure 23: A uniform LBP with two transitions, radius one and eight neighbors..... | 30 |
| Figure 24: PCR combinations | 32 |
| Figure 25: Scatterplot of households in a three-dimensional space. | 34 |
| Figure 26: Plot of households along dimension 1 | 35 |
| Figure 27: Index values aggregated into slums with some examples | 36 |
| Figure 28: Squared correlation of indicators with MCA dimension 1 | 37 |
| Figure 29: Indicator categories along dimension 1 | 38 |
| Figure 30: Plot of slums along dimension 1 | 40 |
| Figure 31: Squared correlation of indicators with MCA dimension 1 | 40 |
| Figure 32: QS slums on map and some examples | 41 |
| Figure 33: Comparing distribution of values of some indicators with MCA dimension 1 along HH slum | 42 |
| Figure 34: Classical index result with some examples | 43 |
| Figure 35: Effect of varying patch size..... | 44 |
| Figure 36: Results obtained by different CNNs | 44 |
| Figure 37: Per class accuracy with some examples of classified patches | 45 |
| Figure 38: Relation between CNN softmax and QS index..... | 46 |
| Figure 39: Predicting indices with CNN | 46 |
| Figure 40: Result of varying number of grey level in GLCM..... | 47 |
| Figure 41: Results of performing PCR to predict HH and QS indices | 48 |
| Figure 42: results of combining different features to predict QS index | 48 |
| Figure 43: Worst-off slums predicted by the model with respective patches..... | 50 |
| Figure 44: Best-off slums predicted by the model with respective patches..... | 51 |
| Figure 45: Predicted values over standardized residual of the created models | 52 |

LIST OF TABLES

| | |
|---|----|
| Table 1: Available data | 16 |
| Table 2: Distribution of patches along images and train/val and test sets | 26 |
| Table 3: Simple CNN hyper-parameters..... | 27 |
| Table 4: Hand-crafted features - band ratios..... | 29 |
| Table 5: Hand-crafted features - GLCM..... | 29 |
| Table 6: Hand-crafted features - LBP..... | 30 |
| Table 7: GIS features | 30 |
| Table 8: Summary of the MCA model for the HH data | 33 |
| Table 9: Summary of the MCA model for QS data | 39 |

ABBREVIATIONS

| | |
|---------------|---|
| AHP | Analytic Hierarchy Process |
| ANN | Artificial Neural Network |
| BNorm | Batch Normalization |
| CDT | Complete Disjunctive Table |
| CNN | Convolutional Neural Network |
| DEM | Digital Elevation Model |
| FA | Factor Analysis |
| FCN | Fully Convolutional Network |
| GBR | Gradient Boost Regressor |
| GIS | Geographic Information System |
| GLCM | Grey Level Co-occurrence Matrix |
| HH | Household |
| HSIC | Hilbert-Schmidt Independence Criterion |
| ICT | Information and Communications Technology |
| ILSVRC | ImageNet Large-Scale Visual Recognition Challenge |
| IMD | Index of Multiple Deprivation |
| JM | Jeffries-Matusita |
| LBP | Local Binary Pattern |
| LOO | Leave One Out |
| LRN | Local Response Normalization |
| MCA | Multiple Correspondence Analysis |
| ML | Machine Learning |
| MRF | Markov Random Field |
| NDVI | Normalized Difference Vegetation Index |
| OBIA | Object Based Image Analysis |
| OSM | Open Street Map |
| PCA | Principal Component Analysis |
| PCR | Principal Component Regression |
| PLSR | Partial Least Square Regression |
| QS | Quick Scan |
| ReLU | Rectified Linear Unit |
| RF | Random Forest |
| RS | Remote Sensing |
| SVM | Support Vector Machine |
| VGG | Visual Geometry Group |
| VHR | Very High Resolution |
| VIF | Variance Inflation Factor |

1. INTRODUCTION

Capturing slum settlements using satellite images has been a challenge due to the inherent complexity of such areas, and studies have been increasingly conducted to address this problem. In addition to studies which have been conducted to find where slums are, recently, some studies have identified local variations of slums with the help of Very High Resolution (VHR) images. This study investigates whether we can connect socio-economic characteristic of a slum area to image features derived from VHR imagery using state-of-the-art machine learning algorithms.

1.1. Background and justification

Urban population in developing countries is rapidly growing, which results in many challenges for the local governments (Duque, Royuela, & Noreña, 2013). Presently, the majority of people live in urban areas, and it is estimated that the proportion of urban dwellers will increase from 54% in 2014 to 66% by 2050 (United Nation, 2015). All of this population growth is expected in urban areas, and most of it will take place in the developing world (United Nation, 2015). According to Weeks et al. (2012), population growth in such countries needs to be absorbed by urban areas as they are the only places in which economic growth and job opportunities would be expected. However, lack of the governments' and planners' capacity to meet the vast volume of housing demands coupled with their inability to provide basic services leads to an increase of poverty rates (Kohli, Sliuzas, Kerle, & Stein, 2012). In contrast with the past, poverty is not concentrated in rural areas, and it is being urbanised in many cities, especially in the Global South (Kombe, 2005). Presently, urban poverty mostly brings about the emergence and expansion of slum areas which are sub-standard shelters for the growing urban population (Kohli et al., 2012).

Slum dwellers are approximately one-quarter of the total urban population (UN-Habitat, 2015). Such areas are generally defined as places deprived from at least one of these five elements; safe water, proper sanitation, durable housing, tenure security and sufficient living space (UN-Habitat, 2003). However, slums are highly diverse and identifying them is a very complex problem, so the indicators need to be contextualised at the local scale (Kohli et al., 2012). Currently, approaches in dealing with slum areas are to upgrade and transform them into a better stage and not just to eradicate them (Arimah, 2010). Therefore, monitoring such areas is vital to understand where to invest and intervene (Duque, Patino, Ruiz, & Pardo-Pascual, 2015; Duque et al., 2013). In contrast, slums and their characteristics are mostly hidden in the official census data (Nijman, 2008). Even if there is census data available, the characteristics of such deprived areas are hidden as a result of aggregating data at administrative boundary levels (Kuffer, Pfeffer, Sliuzas, Baud, & Maarseveen, 2017). As an alternative, remote sensing (RS) is one of the tools with the ability to provide disaggregated information about such areas (Kuffer, Pfeffer, & Sliuzas, 2016).

RS imagery and techniques provide up-to-date information about places with no available or accessible data (Weeks, Hill, Stow, Getis, & Fugate, 2007), in this case, slum areas. Possible questions that RS could answer about such places are "where, when and what?" (Kuffer et al., 2016, p. 7). Therefore, taking advantage of these techniques enables us to analyse slum areas, their dynamics and variations in space and time (Kuffer et al., 2017; Patino & Duque, 2013). Such applications can be motivated by social, economic, environmental, or governance purposes to analyse temporal and morphological dynamics (Kuffer et al., 2016). Ultimately, these analyses could result in a basis for localised and contextualised slum upgrading programs (Olthuis, Benni, Eichwede, & Zevenbergen, 2015). Nevertheless, slums do not have unique

physical characteristics (Kohli et al., 2012) and identifiable spectral differences (Kit, Lüdeke, & Reckien, 2012), so they are not easily detectable from satellite images.

Although it is crucial to detect where the slums are, it is more important to know their conditions and special needs to establish relevant upgrading programmes for each of them. Slums are highly diverse in their socio-economic conditions (Krishna, Sriram, & Prakash, 2014) and it is likely to find impoverished, worst-off slums and more wealthy, formal-like slums in the same context (Kuffer et al., 2017). Thus, advanced methods to consider spectral and spatial information is needed to provide accurate and reliable information about slum settlements (Sandborn & Engstrom, 2016).

A wide range of image analysis methods is available to extract and provide information about slum areas from remotely sensed images. The scale of the analysis starts from pixel to area level and includes manual and visual interpretation to automatic recognition (Kuffer et al., 2016). The following section briefly introduces previous works and methods in this area to show where we are standing now.

1.2. Research gap identification

This study aims to capture the variations of slum areas using VHR images. To be specific, we can see that past studies tried to correlate and connect image features to different socio-economic indicators, but there are some shortcomings. Some studies correlated individual socio-economic indicators to image features, but they did not comprehensively explore these indicators as well as how to characterize them in terms of socio-economic strata. As an example, Sandborn and Engstrom (2016) correlated some individual indicators such as population and housing density to image metrics, but they did not develop a framework or an index to describe different socio-economic groups in the city. Individual indicators cannot describe what is exactly happening on the ground by themselves. Moreover, describing limited individual indicators might not give a comprehensive view of the settlements and might not result in productive policy implications. There are studies which developed indices of poverty, slum, or deprivation and connected them to image-based features; however, the developed indices did not cover all the aspects of deprivation. Arribas-Bel, Patino, and Duque (2017) focused on deprivation of the living environment, Duque et al. (2015), developed an index covering the physical and financial aspects of deprivation, and Engstrom, Newhouse, Haldavanekar, Copenhaver, and Hersh, (2017) related image features to the financial aspect of deprivation¹. Furthermore, all of these studies captured deprivation in delineated boundaries, covering both formal and informal areas. Kuffer et al. (2017) captured the diversity of slum areas and identified four slum sub-categories with image-based features. Therefore, this study showed the capability of extracting detailed information about slums (that relates to socio-economic status). However, the classes were broad and did not include details on socio-economic variations, so this offers potentials for more investigation. Having more detailed information about different dimensions of deprivation could lead to more effective policies in the planning process. Connecting image features to deprivation, could guide us to a better understanding of variations in socio-economic conditions of slum dwellers and provide a better insight about people's status and needs.

Another gap in this research field relates to the overall methodology. On the one hand, one of the burdens and time-consuming processes in the previous works was to extract relevant features from the image. Knowing which features can explain variations of classes is not straightforward and needs a lot of experience and local knowledge (see Duque, Patino, & Betancourt, 2017; Kuffer et al., 2017). Studies which focused on Object Based Image Analysis (OBIA) also needed to invest a lot of time for parameter tuning (see Kohli, Sliuzas, & Stein, 2016). Besides, in many cases, this experience is not transferable from one context to another as some characteristics vary in different areas (Kuffer et al., 2017). Even using advanced classifiers and feature reduction methods, it is not easy to generally match the most relevant features to capture the diversity of urban areas. On the other hand, Convolutional Neural Networks

¹ Different domains of deprivation containing social, human, financial, physical, and contextual domains is discussed elaborated in chapter 2.

(CNNs) have an advantage over other classifiers as they can learn, and extract features itself. CNNs are becoming increasingly used as one of the most advanced methods to solve RS classification problems (see Castelluccio, Poggi, Sansone, & Verdoliva, 2017; Scott, England, Starms, Marcum, & Davis, 2017; Zhang, Zhang, & Du, 2016). Some studies have been developed aiming to map and detect informal settlements using CNN. Mboga (2017) showed that CNN outperformed a Support Vector Machine (SVM) classifier fed by hand-crafted GLCM features to identify informal settlements, although they also mentioned that carefully selected hand-crafted features might generate higher accuracies. Persello and Stein (2017) also developed an efficient deep network to map informal settlements². However, none of the studies analysed the relationship between automatically extracted features and socio-economic variations of such settlements and they treated all slums or informal settlements as a single labelled class.

Although the second gap is more related to the method of the study, the significance of this gap and the capability of the introduced method made the author to use this method as the core of this study. Capturing the socio-economic status using image-based features could be very complicated, as we are capturing abstract concepts by images, and methods used in one application might be considerably different from another application or context. Thus, the general idea is that taking advantage of the capacity of deep learning algorithms might result in new methods of extracting detailed information about slums, which have not been explored yet. To this end, this study will benefit from CNNs to relate image-based features to deprivation and explain variations of slum settlements³.

1.3. Research objective

1.3.1. General objective

Considering above-mentioned discussions, the main aim of this study is:

To analyse the relationship between derived image-based features from VHR satellite images and the socio-economic variations of slum settlements.

1.3.2. Specific objectives

1. To adopt an existing deprivation framework for identifying variations in socio-economic conditions of slums.
2. To train CNNs to distinguish slums from formal areas.
3. To extract hand-crafted and GIS features to be used for predicting deprivation.
4. To build regression models with CNN, hand-crafted, and GIS features and predict deprivation.

1.4. Research questions

1. To adopt an existing deprivation framework for identifying variations in socio-economic conditions of slums.
 - 1.1. What ground-based indicators characterise deprivation and its variations in slum areas based on the literature?
 - 1.2. How to characterise current slum areas based on derived indicators and available secondary data?
 - 1.3. To what extent indicators from the quick scan fieldwork⁴ can explain slums' socio-economic variations?

²They used a deep Fully Convolutional Network (FCN) which is a type of CNNs without any fully connected layer.

³reasons to use deprivation concept to describe slums' variations will be elaborated in chapter 2.

⁴This is the fieldwork conducted for this study to collect data from slums by standing on a point and documenting their visible deprivation-related characteristics. More explanation is provided in section 4.1.

2. To train CNNs to distinguish slums from formal areas.
 - 2.1. What is the best strategy to create patches and distribute them to train, validation, and test sets?
 - 2.2. What is the optimal architecture of the CNN in terms of accuracy and efficiency?
 - 2.3. To what extent pre-trained networks can improve the results?
3. To extract hand-crafted and GIS features to be used for predicting deprivation.
 - 3.1. What kind of hand-crafted features can be extracted to explain deprivation?
 - 3.2. What kind of GIS features can be extracted with a potential of describing deprivation?
4. To build regression models with CNN, hand-crafted, and GIS features and predict deprivation.
 - 4.1. To what extent a CNN can predict variations of deprivation?
 - 4.2. To what extent hand-crafted and GIS features can predict variations of the deprivation framework?
 - 4.3. Can hand-crafted and GIS features improve the predicted values obtained by CNN?

1.5. Hypothesis

The central hypothesis in this study is: observable indicators, either on the ground or in the image (i.e., physical and contextual domains of deprivation) can explain variations of socio-economic status of slums (at least to some extent). We considered this hypothesis in objective 1 when we explore two sets of socio-economic data about slums; one set covers all aspects of deprivation (i.e., social, human, financial, physical, and contextual domains); and another one focuses only on physical and contextual domains. In objective 2 to 4, the main hypothesis is that physical and contextual information extracted from satellite images and GIS features can explain socio-economic variations of slums which contains all aspects of deprivation.

1.6. Thesis structure

The thesis was structured into four main chapters:

- Chapter 2 reviews related literature in the field of slum mapping, deprivation studies, approaches toward creating indices, and Convolutional Neural Networks.
- Chapter 3 briefly describes study area and available primary and secondary data for this study.
- Chapter 4 elaborates the methods have been conducted for this study. It contains methods used for preparing samples, preparing images, training CNNs, extracting hand-crafted and GIS features, and building regressions.
- Chapter 5 presents the results obtained in this study and discusses them to provide a detailed interpretation of them.
- Chapter 6 ends this report by providing a list of remarks and recommendations for further studies.

2. LITERATURE REVIEW

This chapter reviews related literature and clarifies research directions. In the first section, an overview of the efforts in image-based slum identification will be presented. In the second section, related issues and literature about understanding socio-economic patterns will be discussed. The next section provides an overview of approaches to building an index. After that, the chapter reviews deep learning and Convolutional Neural Networks. The chapter ends by summarizing the research concepts.

2.1. Detecting slums from above

There is a broad range of studies, which captured the location and dynamic of slums. According to Kuffer et al. (2016), the term “slum” is not common among all of these studies and the way they call such settlements reflects the local background and the goal of the study. These studies range from health to policy and planning, and to socio-economic studies. Therefore, we can find many other terms like “informal settlement”, “sub-standard area”, “unplanned area”, and “deprived area”. An important point here is that all these studies tried to address a problem connected to settlements which are below a standard level in a specific context. Studies about finding these areas mostly result in classified images into different settlement types followed by assessing the overall accuracy of the classification (see Kohli et al., 2016; Kuffer, Pfeffer, Sliuzas, & Baud, 2016; Zhang et al., 2016)

There is a considerable amount of methods available in capturing slums. Here a range of methods which vary in terms of the level of automation is described to demonstrate this variation. Methods with a great deal of user involvement used manual image interpretation. As an example, Munyati and Motholo (2014) examined the relationship between socio-economic status and features derived from visual image interpretation. Pixel-based classification based on spectral information is a conventional method. For instance, Thomson and Hardin (2000) took advantage of spectral information derived from satellite images in combination with GIS layers to find potential low-income groups. Some studies enriched pixel-based image classification with calculating metrics derived from the classified image. Weeks et al. (2007) used this method to explore the heterogeneity of deprived areas.

More automated studies took advantage of texture and OBIA features. Textures and object-based features are hand-crafted features which can be extracted in addition to original spectral information to perform contextual analysis beyond individual pixels. These methods outperform pixel-based methods especially when we are working with VHR images because there is a massive amount of information in such images and the relation between pixels also becomes essential (Kuffer, Pfeffer, & Sliuzas, 2016). Ignoring this relation could result in a very noisy classification of the VHR image and corrupt further analysis. To explore some studies, Williams, Quincey, and Stillwell (2016) classified roof objects of the informal settlements to estimate the population. Kohli et al. (2016) identified urban slums using Grey Level Co-occurrence Matrix (GLCM) texture features and image metrics derived from OBIA. The study relied on a developed ontology of slums by Kohli et al. (2012), which conceptualized physical characteristics of slums in terms of RS in three levels; i.e., building level, settlement level, and environment level. Looking deeper at texture features, Ella and Wyk (2008), showed that using Local Binary Pattern (LBP) features gives the highest accuracy in detecting slum settlements comparing to other texture features. Recently, machine learning algorithms have brought more capabilities to the image analysis field.

Machine Learning (ML) algorithms are methods, which use training samples to learn how to identify and distinguish different patterns to solve regression and classification problems (Richards & Jia, 2006). The better training samples are, the better they can cover variations of patterns and the higher accuracy the

algorithm can achieve. An example of using machine learning in computer science is handwriting recognition algorithms. Basically, these methods are pixel-based, so they can not consider the pixels' context, and they are unlikely to perform well in classifying VHR images (Vatsavai, 2012). To address this problem, two main approaches exist; using other algorithms like Markov Random Field (MRF) which allow considering pixels' spatial autocorrelation (Graesser et al., 2012); or feeding ML algorithms with contextual image features which is mostly used in the field of mapping deprivation and slums. As an example, Arribas-Bel et al. (2017) derived textural, spectral, and structural information as well as land cover metrics to feed machine learning classifiers and estimate deprivation in Liverpool, UK. Another study with the same approach is from Duque et al. (2017) which was conducted to analyse three cities in Latin American countries using VHR Google Earth images and the SVM classifier. Using mixed methods is also a possibility in the field of image analysis. Kuffer et al. (2017) used the Random Forest (RF) classifier to categorise land covers, then the result, as well as many other image features, were used to capture the variation of slum settlements.

There are also feature reduction methods to select the most relevant features among a broad set of extracted features. These methods use some algorithms to calculate the separability of classes and to select necessary features to distinguish the classes (Richards & Jia, 2006). Hilbert–Schmidt Independence Criterion (HSIC), which is based on statistical dependency of features on class labels; Jeffries-Matusita (JM) Distance, which analyses classes based on probability distances; and regression models, which calculate the contribution of each feature to predict classes are algorithms which are possible to be used to reduce the dimensionality of images (Camps-Va, Mooij, & Scholkopf, 2010; Kuffer et al., 2017; Richards & Jia, 2006). One of the ML algorithms which is recently being explored in the field of satellite image classification is Convolutional Neural Network (CNN) (e.g., Castelluccio et al., 2017; Jean et al., 2016; Marmanis, Datcu, Esch, & Stilla, 2016; Scott et al., 2017). CNNs are deep learning algorithms that can extract features from the original image during its learning process instead of being fed by handcrafted features (Nielsen, 2015). CNNs focus only on the most relevant features and reduce large feature sets used in hand-crafted feature studies. Examples of recent studies adopting this method in the field of deprivation-related studies are by Persello and Stein (2017) to develop a deep Fully Convolutional Network (FCN) to detect informal areas from VHR images⁵, Jean et al. (2016) to predict poverty based on consumption rate and wealth, and by Mboga (2017) to train a CNN to distinguish formal and informal settlements using VHR Quickbird images.

2.2. Mapping deprivation

Slums are a sub-category of deprived areas containing illegal constructions, but there could also be formal areas which are more deprived than slums depending on the context and the way to conceptualize deprivation (Baud, Sridharan, & Pfeffer, 2008; Kohli et al., 2016). The term “deprivation” refers to a wide range of socio-economic aspects which are essential to understanding variations of slums (Baud et al., 2008). In this study, we focus on slum as a deprived area and consider the concept of deprivation in order to explain variations between slum settlements. To clarify dimensions of deprivation, indices from former works are reviewed.

“Poverty” and “deprivation” are sometimes used interchangeably. Pacione (2009) conceptualized multiple deprivation and introduced poverty as a financial component of deprivation. In fact, poverty is measured mostly by income and consumption rate in poverty mapping studies (e.g., Engstrom et al., 2017; Jean et al., 2016). Many other studies also used deprivation as a wider concept than poverty and extended it to other socio-economic domains (e.g., Baud, Sridharan, & Pfeffer, 2008). Considering this, deprived areas are defined as “areas with socio-economic marginalization and limited access to services” (Cabrera-Barona, Wei, & Hagenlocher, 2016, p. 1), which is broader than just lack of income or consumption rate.

⁵ Fully Convolutional Network (FCN) is a type of CNNs without any fully connected layer and it is more efficient for mapping purposes. For more details see Persello and Stein (2017).

The way we look at poverty and deprivation could be “relative” or “absolute”. Absolute poverty is usually defined by assigning a threshold for income or consumption rate and categorizing the society into poor and non-poor groups (Baud et al., 2008; Pacione, 2009). In contrast, related studies seek to find poorer or more deprived areas in relation to other areas in a specific context (Baud et al., 2008). From this perspective, we could argue that poverty and deprivation are relative in time and space and there is no absolute measure for them (Beteille, 2003). Henceforth, this study focuses only on “relative deprivation” as it gives a more comprehensive understanding of sub-standard areas related to the context of the analysis.

One of the most important elements to measure deprivation is the unit of the analysis. Flowerdew, Manley, and Sabel (2008) demonstrated that using smaller alternative analysis units other than administrative boundaries has a significant influence on the spatial pattern of deprivation and it might demonstrate deprived areas which were already hidden by aggregating information to administrative units. Cabrera-Barona et al. (2016) argued that pre-defined boundaries cannot demonstrate real spatial patterns of deprivation and explained two main factors affecting the result; scale effect and zoning effect. Scale effect is related to the size of units and zoning effect is linked to where we draw boundaries. To identify how to operationalize deprivation in each analytical region, prior related works are reviewed.

Many deprivation indices have been established in the field of health-related studies. The logic behind such studies is that health is more degraded in areas with a higher level of deprivation (Flowerdew et al., 2008). The most critical dimensions these studies have been focused on are employment, education, and household conditions like overcrowding (Cabrera-Barona, Murphy, Kienberger, & Blaschke, 2015; Messer et al., 2006). Health outcomes like mortality rates are the dominant dependent variables in such studies (Bell, Schuurman, Oliver, & Hayes, 2007), meaning they investigated the relation between deprivation indices and indicators of health outcome. Some health-related studies introduced new indicators to the deprivation index. Bell et al. (2007) considered the proportion of single-parent families to establish an index. Pampalon, Hamel, Gamache, and Raymond (2009) also included some social indicators like dependency rate and vulnerability of households. One step further, Cabrera-Barona et al. (2016) added availability of basic services and distance to nearest healthcare.

There are also studies focused on deprivation in a broader perspective. Pacione (2009) conceptualized multiple deprivation as a phenomenon that includes several components; income (poverty), housing, education, health, safety, dependency rate, availability of services, employment, and institutional power. Each component contributes to the level of multiple deprivation. Baud et al. (2008) established an Index of Multiple Deprivation (IMD) based on the livelihood asset framework and identified four different capitals, as assets used to improve well-being. Human capital consists of health, education, and employment; financial capital refers to income, savings, and household assets; physical capital relates to housing and availability of services; social capital contains collective social connections. This framework focused on households and did not consider any contextual dimension.

At the national level, the UK countries developed their own relative deprivation indices. They call these indices “relative” as they emphasized that the value resulted from these indices cannot be compared across countries, and they only show relative deprivation variations within each country (e.g., Welsh Government, 2014). The more comprehensive one belongs to the Welsh government, which contains income, employment, health, education, accessibility, community safety, physical environment and safety domains (Welsh Government, 2014). Northern Ireland’s excluded the housing domain, Scotland’s discarded the physical environment domain, and England’s dropped the accessibility domain (Ministry of Housing Communities & Local Government, 2015; Northern Ireland Statistics and Research Agency, 2010; Scottish Executive, 2006).

2.3. Approaches to building an index

After selecting a number of deprivation-related indicators, we usually need to aggregate them to end up with an index with a more abstract definition (in this study, deprivation) and measure it in our region of interests. Here is a brief review of the conventional approaches to building an index. Most of them aimed to end up with a single number for each region and then compare the output. However, some of them also used information of indicators or aggregated information into some domains instead of using a single number as an index which could be helpful for policy making. The range of these approaches is from manual indexing to entirely data-driven approaches. Baud et al. (2008) created the IMD index, which is based on giving equal weights to continuous numeric indicators to build four dimensions and then giving equal weights to dimensions to create an index. Indeed, this method needs a great deal of experience as we need to manually select relevant and meaningful indicators, but the implementation is rather simple, and the interpretation of results is very comprehensible. Drawbacks of such method are assumptions, which might not exactly match what is happening in reality. Having more indicators for some dimensions also might bias the result as the weights for all dimensions are equal.

Data-driven approaches are popular alternatives to overcome the limitations of introducing many assumptions to the index. UK countries used a standard procedure to create their deprivation indices. For instance, the Welsh Government (2014) used continuous indicators and factor analysis (FA) to aggregate indicators into dimensions. In the second step, they considered some weights for dimensions based on the quality and importance of indicators and created their index. Pampalon et al. (2009) also conducted the same approach and used continuous indicators followed by a Principal Component Analysis (PCA) to create two components. Then they named domains and used them as abstract domains of deprivation.

Consider indices explained earlier, it is worth mentioning the difference between PCA and FA as they are sometimes used instead of each other. Both methods reduce the dimensionality and give the contribution of each indicator to each factor (in case of FA) or component (in case of PCA). However, FA is mainly used to find underlying concepts which indicators try to describe, and PCA is used when we would like to only reduce data dimensionality (Field, 2013). Considering these definitions, it seems that the two examples explained in the previous paragraph used these two methods instead of each other.

Commonly, in literature, the first component of PCA is used to reduce the dimensionality of data and aggregate indicators into dimensions or an index. Cabrera-Barona et al. (2016) used continuous indicators and executed a Variance Inflation Factor (VIF) test i.e., a test of multicollinearity between indicators (Field, 2013), then aggregated indicators using the first component of PCA. Messer et al. (2006) also used the first component of a PCA on continuous indicators to build a deprivation index. Other methods of weighting indicators also exist like Analytical Hierarchy Process (AHP) which is a method based on experts' judgment (Cabrera-Barona et al., 2016). Though, detailed review of such methods is out of the scope of this study as the main approach to build an index in this study is by considering as few assumptions as possible and methods like AHP are mainly based on people's judgments and experiences.

However, PCA and FA are not appropriate methods when working with categorical or ordinal data⁶. Rains, Krishna, and Wibbels (2017) worked with household data from different slums including many categorical indicators. They ranked these categorical indicators and calculated descriptive statistics (e.g., mean) over households within each slum to end up with a continuous indicator. Two main disadvantages of this method are assumptions made for ranking categories and for considering equal distances between them. Moreover, by averaging a ranked variable, we are not considering its variation within a settlement. As an example, having ten households with value 1 and ten with value 3 is equal to having twenty households with value 2. In such cases, it seems better to conduct principal component methods, which are designed for categorical data. Multiple Correspondence Analysis (MCA) is a widely used method with the same approach as PCA but uniquely designed for categorical data (Jolliffe, 2002). The following section explains how this method works.

⁶ Actually, we can use PCA or FA for ordinal data when it is assumed that the data is originally continuous (Field, 2013).

2.3.1. Multiple Correspondence Analysis

This section summarises how MCA works and finds variations of categorical data. Having J indicators, each contains K_j categories, and I individuals, MCA creates a Complete Disjunctive Table (CDT). CDT is a table from individuals as rows, and categories as columns, with binary values showing either each category k belongs to each individual i or not (Coulangeon & Lemel, 2007). Note that in this study we will use the term “individual” to analyse and interpret the results of MCA and it means the unit of the analysis. Depends on what is analysing or interpreting, it could be a slum or a household.

MCA creates point clouds of individuals and categories separately, but it transfers and matches them also on each other to find some relationships between categories and individuals. According to Le Roux and Rouanet (2011), having K categories in total, MCA creates a $K - J$ dimensional point cloud of individuals. Distance of two points in the point cloud is defined as follow:

$$d_{i,i'}^2 = \frac{1}{J} \sum_{k=1}^K \frac{1}{p_k} (y_{ik} - y_{i'k})^2$$

Equation 1: Distance of two individuals in the point cloud created by MCA

where $d_{i,i'}$ is the distance between individuals i and i' , p_k is the proportion of individuals having category k , y_{ik} and $y_{i'k}$ are 1 if category k belongs to individual i or i' and 0 otherwise. Therefore, two individuals with exactly the same categories have distance of zero, two individuals which share many categories have a small distance, and two individuals that one of them has a rare category have a large distance. In the other words, individuals with common categories are located around the origin of the point cloud and individuals with rare categories are located at the periphery.

MCA also creates a J dimensional point cloud of categories and locates the centre point of each category. The distance of each category from the origin of the point cloud (i.e., the variance of each category) is defined as follow:

$$d_k^2 = \frac{1}{p_k} - 1$$

Equation 2: Distance of two categories in the point cloud created by MCA

Therefore, rare categories (which have low value of p_k) are located far from the origin. The inertia of each category is defined as follow:

$$Inertia(k) = \frac{p_k}{J} d_k^2 = \frac{1 - p_k}{J}$$

Equation 3: Inertia of each category in MCA

where $\frac{p_k}{J}$ is defined as the weight of each category. Thus, inertia is defined as the weighted variance and according to Equation 3, rare categories have more inertia, means more capability, to explain variations of individuals. Equation 3 also shows that MCA locates rare categories far from the origin but does not exaggerate their effects as the maximum value of inertia for the rarest categories approaches $\frac{1}{J}$.

Then, MCA projects both point clouds in a low dimensional space which explains the maximum variation of individuals (same as PCA). It also transfers point clouds, so we can see some relations between individuals and categories. Total inertia of each extracted dimension is defined as follow:

$$\lambda_s = \frac{1}{J} \sum_{j=1}^J \eta_{s,j}^2$$

Equation 4: Total inertia of each dimension

where λ_s is the total inertia of dimension s and $\eta_{s,j}$ is the correlation coefficient between dimension s and variable j . λ_s shows the percentage of variations explained by dimension s . Finally, extracted dimensions are explained based on the distribution of individuals and categories along them.

As examples of studies used this method, Coulangeon (2017) and Coulangeon and Lemel (2007) conducted MCA on social topics to understand patterns of categorical data. One advantage of this method is that we are not forcing indicators to be ordered as an assumed logic and we can understand existing patterns as well as relative distances between categories. MCA gives importance to variables considering their power to explain the variation of individuals.

2.4. Capturing variations of deprivation

As mentioned earlier, an extensive range of studies have been conducted to detect slums or deprived areas in general, but a few of them were dedicated to investigating variations of such settlements (Kuffer et al., 2017). In many cases, studies concluded that the variations which exist in slum areas were ignored by official maps or data (e.g., Krishna et al., 2014; Kuffer et al., 2017). Furthermore, Martinez, Pfeffer, and Baud (2016) showed the level of aggregation in the analysis could easily hide a considerable amount of information in sub-analytical region levels. To this end, availability of disaggregated data helps to identify deprived areas with higher accuracy and improves urban governance (Henninger & Snel, 2002).

Reviewing recent researches is required to realize what is distinguishable and categorizable by satellite images. Some studies distinguished settlement types by manual interpretation and related them to deprivation related indicators (e.g., Munyati & Motholo, 2014). Another category of studies only used specific analytical regions and compared indices with image-based features. Duque et al. (2015) utilized land cover and textural information and tried to explain the variation in their established slum index by such image features. Weeks et al. (2007) created a slum index based on UN-Habitat definition of the slum (i.e., UN-Habitat, 2003) and correlated it with spectral and textural information of the image to detect the slum-ness level of each location. Weeks et al. (2012) relied only on land cover information to identify patterns of poverty and health. Baud, Kuffer, Pfeffer, Sliuzas, and Karuppannan (2010) related metrics derived from land cover classification to IMD score and explored variations of deprivation within census boundaries.

In addition to textural and spectral information, some studies focused on metrics derived from OBIA, additional statistical analysis, and machine learning classifiers. Sandborn and Engstrom (2016) correlated image metrics in addition to Normalized Difference Vegetation Index (NDVI) with census-derived indicators like population density. Arribas-Bel et al. (2017) correlated a deprivation index with spectral and textural information followed by a PCA and compared Gradient Boost Regressor (GBR) and RF classifiers to classify an image based on extracted features. Kuffer et al. (2017) derived image features based on environment, geometry, density, and texture using RF classifier followed by a logistic regression model to distinguish four deprived area and one formal area types in the image.

Regarding the mentioned related works, there are potentials to develop more advanced methods and extract more detailed information about slums using satellite images. CNNs are the state-of-the-art methods used to solve image recognition problems mostly in the field of computer vision, but they can also be used to analyse satellite images. However, none of the studies in the field of capturing deprivation variations took advantage of such deeply learned features. The next section reviews this method and presents examples in the field of satellite image analysis.

2.5. Deep learning and Convolutional Neural Networks

Artificial Neural Networks (ANNs) are a branch of ML algorithms consisting of three main layer categories; one input layer, one output layer, and one or more hidden layers (Figure 1). Each layer contains some neurons, and each neuron in a layer is connected to all neurons in the next layer. This is an architecture containing fully-connected layers of neurons. In a simple ANN, each neuron's activation value z is defined by all neurons in the previous layer with:

$$z = \sigma \left(\sum_j w_j x_j - b \right)$$

Equation 5: Activation value of a neuron in ANN

where x_j is the input and w_j is the weight of j th neuron in the previous layer and b is the bias of the target neuron (Figure 1) (Nielsen, 2015). In Equation 5, σ is called activation function

i.e., a non-linear function that could be different in each network depends on the application. The most common activation functions used in ANN are sigmoid, tanh, and Rectified Linear Unit (ReLU) (Nielsen, 2015). The number of input neurons depends on the size of one sample to be used in the network and number of possible classes as output reflects the number of output neurons. Generally, ANNs classify vectors (i.e., one-dimensional arrays). As an example, considering Figure 2, it classifies an input vector of size 3 into two possible classes. However, one can also classify two- or three-dimensional arrays (i.e., an image with one or more channels) with an ANN by considering each pixel as an input neuron. For instance, to classify patches of 20x20 pixels into six possible land use classes, we need 400 neurons as input and 6 neurons as output. Note that by classifying images with ANNs, local contextual characteristics of the input images will not be considered. Each ANN has at least one hidden layer which takes decisions. The output neuron with a higher value is selected as the decided class for each input.

Training a network means tuning its weights w and biases b in a way that the network can distinguish different classes. As an example, having six possible classes (six output neurons), if we give an input with class 4 to the network, all weights and biases should be tuned in a way that in the output layer, neuron of class 4 gets the highest activation value than other five neurons. If the network can correctly classify almost all inputs into six classes, then it has been tuned well. To quantify the overall performance of the model an objective function is calculated⁷. Better tuned weights and biases results in lower objective function, so it is a measure to calculate how much the network learned. Therefore, if we tune weights and biases in a way that we get the minimum value of the objective function, the network is trained very well (Nielsen, 2015). To reduce the objective function, the idea of stochastic gradient descend is used, i.e., iteratively choosing a number of inputs which are called batches to compute current error of the network and find the minimum of objective function (Bottou, 2010). After analysing a batch of data, the error of the objective function is distributed to the weights of all neurons using backpropagation algorithm to find change of each weight in the network (Rumelheart, Hinton, & Williams, 1986). After training the CNN using all available training samples in form of small batches, the first epoch of training is completed and samples are distributed into new batches for the next epoch of training (Nielsen, 2015). Any value which is tuned during the learning process is called a parameter of the network. Values which are set on top of the learning process are called hyper-parameters.

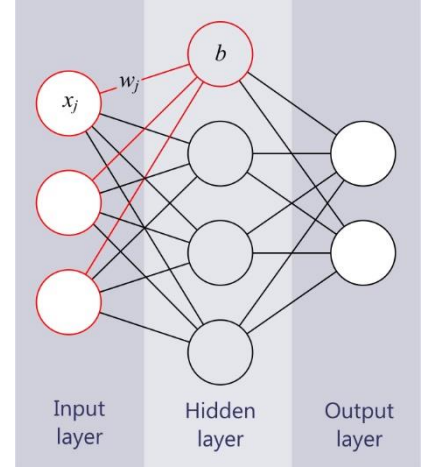


Figure 1: Simple illustration of ANN

⁷ it is sometimes called cost or loss function

Deep networks contain many hidden layers, can extract more abstract information of inputs and can solve more complicated problems. Convolutional Neural Networks (CNNs) are a branch of deep ANNs that have at least one convolutional layer in addition to fully-connected layers. These networks have been designed specifically to solve problems of image recognition.

Figure 2 shows a simple illustration of a CNN. An essential characteristic of these networks is that they use local receptive fields, i.e., a moving window filter, to extract information from images (see Figure 2). All weights and biases in a local receptive field are shared so it dramatically decreases the number of parameters to learn in a CNN. Considering Figure 2, an input of size m with c channels is downsampled by p filters of size k in a convolutional layer. This is followed by a non-linear function and inputs are transformed to an image of size n with p channels (Nielsen, 2015). In fact, this is how a convolutional layer considers the contextual information of an image and extracts important features from it. This process is followed by one or more fully connected layers that make the decision for the output (Bergado, Persello, & Gevaert, 2016). Usually, convolutional layers are followed by a max pooling layer that downsamples the image by extracting the highest value of a given size filter⁸.

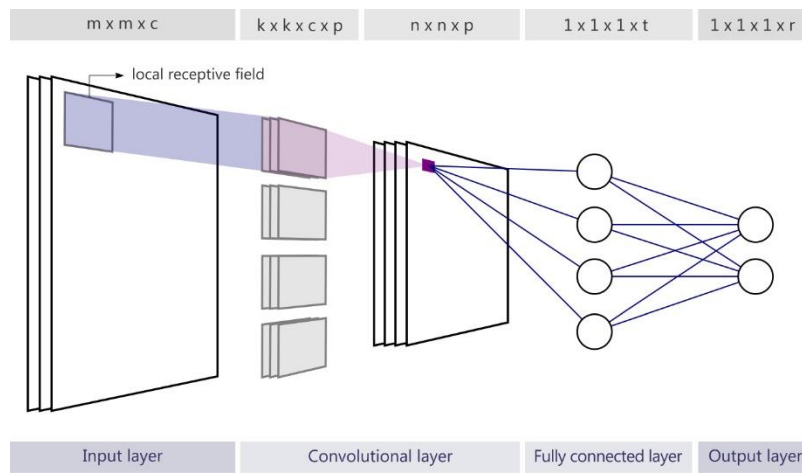


Figure 2: Simple illustration of CNN

Such networks have a tremendous number of parameters, and there is a high risk of overfitting on training data, weakening the availability of a network to be generalized (Nielsen, 2015). To avoid this problem, methods have been used for regularization. L1 and L2 regularizations are additional terms that are added to objective functions to avoid overfitting (Nielsen, 2015). Currently, conventional methods to prevent overfitting are dropout layers (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) that randomly exclude some neurons in a layer in each iteration, and batch normalization (Ioffe & Szegedy, 2015) that normalizes each batch of training samples. Image augmentation (Simard, Steinkraus, & Platt, 2003) is also an advantageous method that increases the number of training samples for more robust learning. Proper weight initialization and early stopping are also other considerations for this problem (Nielsen, 2015).

Many complex networks have been developed for challenging problems like classifying ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). These pre-trained networks are often used to solve other problems, instead of training new networks from scratch. Studies fine-tuned these networks and adopted them to their own problems. Some of these popular networks are AlexNet (Krizhevsky, Sutskever, & Geoffrey E., 2012), GoogLeNet (Szegedy et al., 2015), VGG (Chatfield, Simonyan, Vedaldi, & Zisserman, 2014), VGG-VD (Simonyan & Zisserman, 2014) and ResNet (He, Zhang, Ren, & Sun, 2015).

⁸ There is also another pooling method called average pooling which is less popular than max pooling.

During past few years, many sophisticated studies have been carried out using CNNs, and some focused particularly on analysing satellite images. Castelluccio et al. (2017) used deep pre-trained networks to classify land-use classes. Scott et al. (2017) took advantage of image augmentation and increased their sample size by rotating and transposing each sample. Then they used three different pre-trained networks to classify 21 land use classes. Marmanis et al. (2016) used a pre-trained network and fed it with their inputs, then converted two last fully-connected layers to a two-dimensional array and trained a simpler network with those inputs. Jean et al. (2016) used a pre-trained network to estimate two continuous variables related to poverty with nightlight images.

2.6. Research concepts

Considering the related literature, we can summarize associated concepts and their relations (Figure 3). The author identified five main domains to be considered in mapping and characterizing slums, or deprived areas in general. Financial, human, social, and physical domains were identified based on the livelihood asset framework and the IMD (Baud et al., 2008) which were collected using very detailed surveys from households (HH). One other domain also linked to deprivation is the contextual domain. This looks a settlement from a broader point of view, considering its location, so it was also considered as one of the deprivation domains. Saharan, Pfeffer, and Baud (2017) also emphasized the importance of this domain to define deprivation. From RS point of view, we can identify deprivation regarding three levels of observation; building, settlement, and environment (Kohli et al., 2012). These kinds of information levels are mainly related to physical and contextual domains of deprivation. Moreover, taking advantage of the added value of GIS layers and spatial analysis, we can understand contextual issues (like hazard risk, land use related data, recorded crimes, accessibility issues, slope and elevation data, etc.). Some studies like Thomson and Hardin (2000), Kuffer et al. (2017) as well as deprivation indices belong to the United Kingdom took advantage of such additional layers.

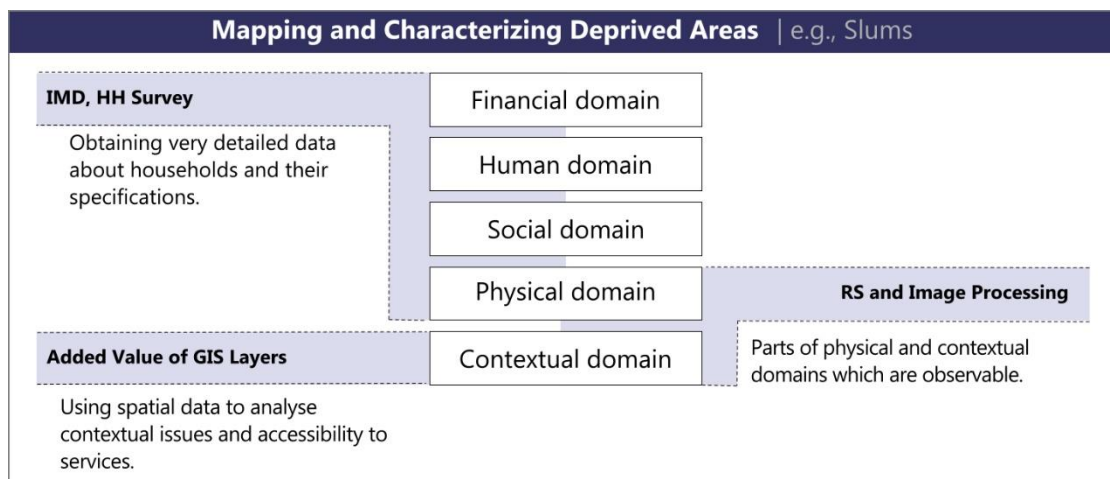


Figure 3: Research concepts

3. STUDY AREA AND DATA DESCRIPTION

This section provides a description of the study area and the available data. The section starts with introducing the case study, and it is followed by explaining the available socio-economic data and satellite images.

Bangalore, one of the biggest cities in India with a population of more than 10 million in 2016 (United Nations, 2016) has been selected as the case study for this research. The city is known as the Silicon Valley of India due to lots of investment in the ICT sector and the presence of many multi-nationality companies (Jayatilaka & Chatterji, 2007). In parallel with growing wealth due to such investments, Bangalore faces poverty

extended throughout the city since people do not benefit equally from the increased wealth (Rains et al., 2017). Consequently, slum settlements have emerged all over the city as a challenge the city should cope with (Rains et al., 2017) (Figure 4). There are two types of slums identified by official agencies: notified slums, where basic services are provided by the government; and non-notified slums (Krishna et al., 2014)(Figure 5). Non-notified slums are not homogeneous settlements. A wide range of variation from very temporary settlements to more permanent, multistorey buildings exist in this group (Krishna et al., 2014). To this end, in order to establish more effective and efficient policies for these settlements, identification of slums' variations is needed (Krishna et al., 2014; Rains et al., 2017). Therefore, this study focuses on the Bangalore context to find potentials of using satellite images in exploring these variations.

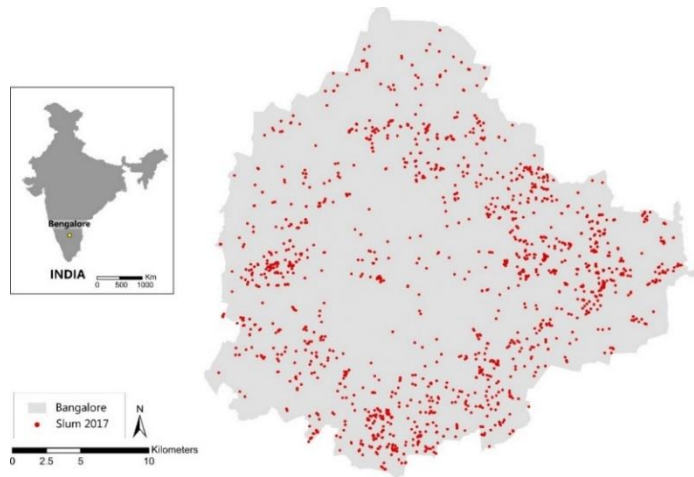


Figure 4: Bangalore and delineated slum locations 2017

Source: Data from DynaSlum project ⁹



Figure 5: Slum in Bangalore
(a) a notified slum; (b) a worst-off non-notified slum.
Source: (Krishna et al., 2014)

Availability of data is also an important factor in choosing the study area (Table 1). This study used two sets of data: a set of socio-economic data and a set of satellite imagery. A detailed survey of 1,114 households living in 37 slums from 2010 in Bangalore was provided within the DynaSlum project⁹. Although locations of the 37 slums were provided, a unique key to connect socio-economic data to their locations was not provided for all of them. So, only the location of 26 slums was known. Furthermore, four Pleiades images, three from March 2016 and one from March 2015 (Figure 6) were available; all were ordered pansharpened with four bands and the spatial resolution of 0.5m. The cloud coverage of the images was 0%, and they had already been corrected radiometrically. Although one of the images was captured on a different date, it helped to have almost a full coverage of the city; therefore, it was also used for the analysis. Moreover, this image was captured in the same month, and we could expect spectral similarities between images. In addition to this data, slum boundaries in 2017 delineated by experts using visual image interpretation and field verification were also available.

In this study, we also conducted fieldwork to collect information about current slums called “Quick Scan” (QS). Details about the fieldwork and data analysis are elaborated in section 4.1.

Table 1: Available data

| Data | Year | Specification |
|--|------|---|
| <i>Household survey of 37 slums + location of the 26 of them Pleiades images</i> | 2010 | Shapefile Survey database |
| | 2016 | Pansharpened |
| | 2015 | Res. 0.5m Cloud coverage: 0% Dynamic range: 12bit Bands: B,G,R,NIR |
| | | |
| <i>Slum boundaries</i> | 2017 | Shapefile 1461 slums delineated |

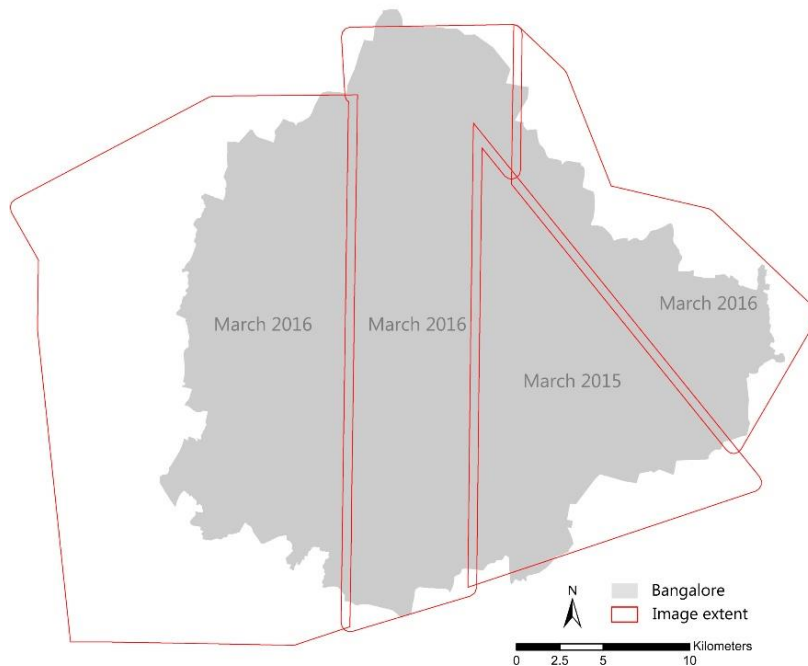


Figure 6: Extents of available images

⁹ <http://www.dynaslum.com/>

4. METHODOLOGY

This section describes the research methodology. It starts with the fieldwork followed by three central section of this study. The first part explains the analysis done using the household survey 2010 and Quick Scan fieldwork to derive deprivation indices. The second part is related to the images analysis, focusing on sampling methods and steps to train CNNs and extracting hand-crafted and GIS features. The last section elaborates methods that have been used to connect image-features with deprivation. Figure 7 summarizes the research steps.

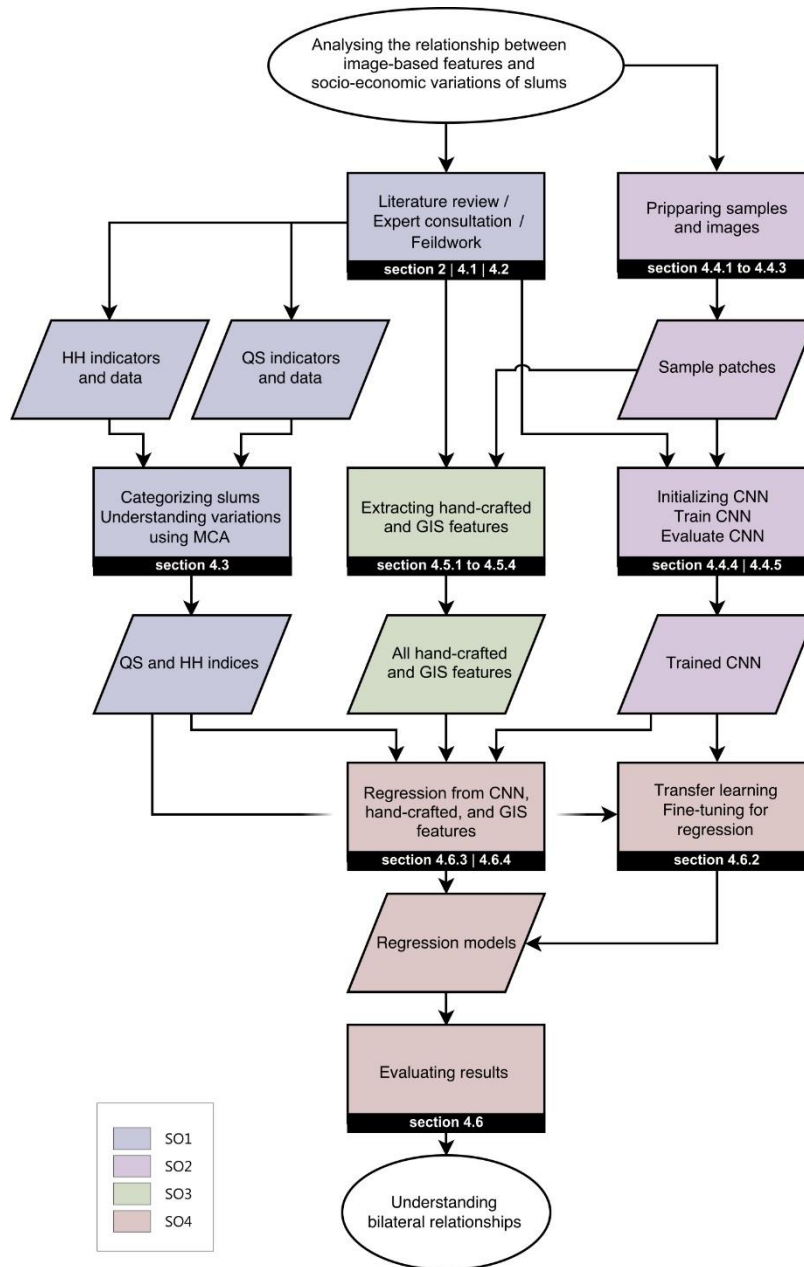


Figure 7: Research steps

4.1. Fieldwork data

This study conducted fieldwork as the first step of the analysis. Regarding Figure 8, detailed household survey data of 2010 was available (HH). Furthermore, VHR images from 2015 and 2016 were also available for the analysis. As the aim of this study was to relate socio-economic data to image features, we could only analyse “A” (in Figure 8) regarding the available dataset because, for 2017, we had only the boundary of slums and no additional socio-economic data. A problem with the available data was that we had only 37 HH samples and we knew the location of only 26 of them. So, if we wanted to model the relationship between HH and VHR with a lot of predictors, the result would not be statistically significant (Field, 2013). In addition, as we had few samples, there was a risk that they were not representative of the population, so the result could not be generalized (this is shown in section 5.1.2). On the other hand, surveying sufficient samples to obtain 2017 data within the very limited time of this study was not feasible. Therefore, a Quick Scan (QS) fieldwork was designed to collect data from 100+ samples within three weeks. The survey collected data on visible deprivation-related indicators. This made it possible to visit more than 100 samples in addition to the HH samples within such limited time. Using QS data, we examined two possibilities: a) to analyse QS data and explore its variations apart from the HH data using 137+ more representative samples (“B” in Figure 8); and b) exploring the relationship between QS data and HH data (if there is no significant change in the structure of slum settlements between 2010 and 2017) (“C” in Figure 8).

A list of indicators was conducted to be used in the survey. The indicators covered three categories, i.e., building, environment, and people. These indicators were selected based on literature as well as experts’ opinions to explain deprivation in slum areas (Annex 1).

An external colleague has done the fieldwork. The rest of this section explains a summary of the sampling strategy. As the time for fieldwork was limited, a two-stage cluster sampling instead of simple random sampling was selected (Stehman, 2009). Doing a simple random sampling would not have allowed reaching the targeted 137+ samples as there were lots of difficulties in transportation in Bangalore and many gridlocks. Therefore, a two-stage cluster sampling was designed as follow:

Since we had 15 days to collect data, we needed 15 clusters and on average 7 samples in each to reach about 105 samples. As the maximum distance of two neighbouring slum settlements was 2.1km, the study area was split into 4km by 4km grids, so each grid contained at least 7 samples theoretically¹⁰. Samples within the clusters located at the city centre were few, but within the clusters located at the periphery were abundant. To increase efficiency, clusters at the periphery with less than 7 samples were removed, but all clusters located at the city centre remained to avoid losing representativeness of samples.

After selecting clusters as explained in the previous paragraph, we needed to select some samples among samples within those clusters. There are two ways of selecting samples, but each has specific disadvantages: 1) By selecting samples systematically, we minimize the effect of spatial autocorrelation, but we also minimize the representativeness of the selected samples. 2) With random sampling, we select more representative samples, but we keep the effect of spatial autocorrelation. To deal with this problem, Vermeiren, Van Rompaey, Loopmans, Serwajja, and Mukwaya (2012) presented a two-stage cluster sampling, first choosing clusters systematically (i.e., dispersed), then choosing random samples within the clusters. The same approach was conducted, a set of spatially dispersed points was created, and 15 clusters were chosen based on points’ location, so 375 samples were selected out of 1461 slums. Using Google

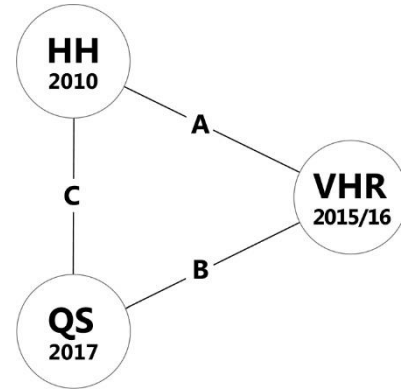


Figure 8: Relation between available data

¹⁰ In a 4km by 4km grid, we can theoretically cover 9 samples with ≥ 2 km distance from each other. Here, by distance, we mean distance of two centre points.

Earth, samples within the chosen clusters were verified, and the ones which had no settlement anymore were removed. After removing samples with no settlement, 208 samples remained out of 375 samples. From 208 samples, 107 samples were randomly selected proportionally based on the number of samples within each cluster. After adding 37 samples, for which we had detailed data, to the 107 samples (144 samples is total), all samples were coded, then an online google map and an offline locus map containing samples as well as an SPSS template were prepared and were introduced to the surveyor.

After completion of the survey, data of all 144 samples were checked with the Pleiades images to be prepared for further analysis. After clearing samples, 121 samples remained. Reasons for removing some samples are as follow: some samples were very remote or were not accessible due to safety reasons, so data about them were missing. As mentioned earlier, samples that we had detailed household survey data about were also supposed to be surveyed again during the fieldwork. However, some of them had been significantly changed or seemed to be changed to formal residential areas, so such samples were also removed to avoid confusion (see Figure 9 as an example). 31 samples out of 37 HH samples (containing the 26 samples which we knew their locations) remained almost unchanged in the period of 2010 to 2017, so all of them were used for the analysis.

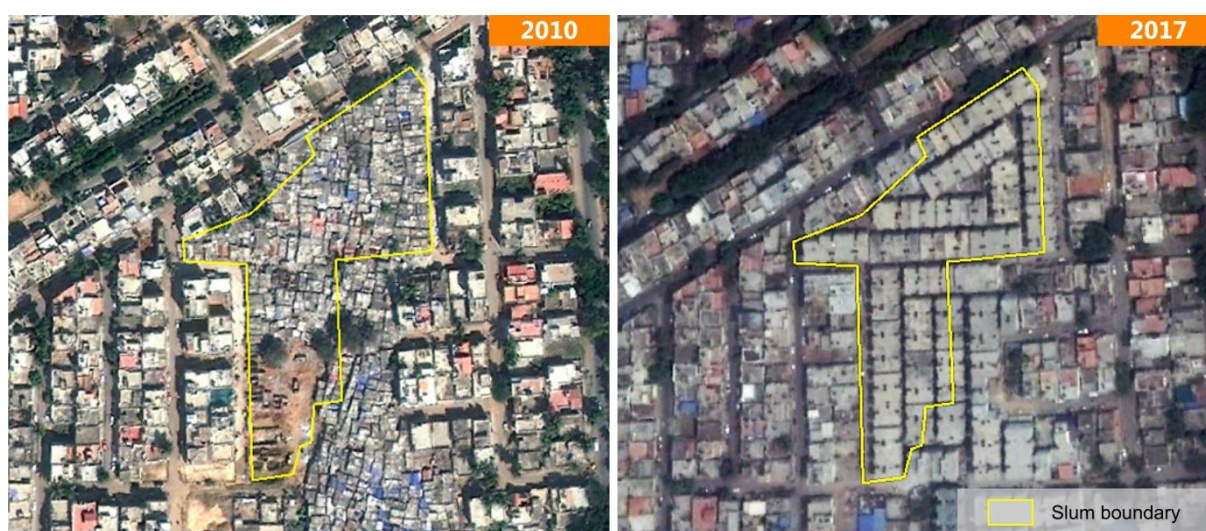


Figure 9: An example of a slum in 2010 which was re-developed in 2017

Source: Google Earth

To prepare the collected fieldwork data for the analysis, three indicators, dominant building type, number of floors, and roof material, that we had each category in percentage were aggregated, and the dominant category was considered for each slum. Other indicators remained unchanged as they were already prepared with multiple choice possibilities, but these three indicators were in percentage, and we needed to aggregate them before using for the analysis.

4.2. Understanding data and adopting deprivation indicators

Available secondary data for this study was a detailed household survey from 1114 households living in 37 slums in Bangalore. The survey contained very detailed socioeconomic data about each member of these households which allows a deep understanding of such settlements. Indicators related to five capitals were selected/constructed based on each capital definition of IMD (explained in section 2.2). Annex 2 lists selected/constructed indicators with respective assumptions and references. Its explanation is as follow:

For the social capital, caste indicator was selected as it was in the original data provided with five classes.

As the human capital, the highest educated person in a household was considered as the highest education level obtained in that particular household, so data of household members were aggregated to find the highest educated person classified in 10 categories. Dependency rate was a constructed indicator based on

Baud et al. (2008), and it was a continuous number between 0 and 1 showing the proportion of workers to household members. Distance to healthcare was used as it was in the original data with three categories. Two indicators as parts of the financial capital were selected. Income, that consisted of 9 categories and ration card with 4 categories, both were used in their original forms.

Physical capital had the maximum number of indicators on our list. Main water source (provided for summer and other seasons separately), toilet facility, and access to electricity were used with their original forms. There was also the second water source for each season, but as the majority of values were missed, it was dropped from the analysis. Crowdedness was a constructed indicator by considering the number of household members and dwelling area, so it was a continuous indicator. Dwelling age was also included in the analysis as it was assumed that better-off slums have older dwellings showing people live there more permanently. Floor/wall/roof materials were also included in their original forms. These materials were related to the first floors of the dwellings. Data about other floors was also provided but again due to many missing values it was dropped.

As this study also considers contextual capital, travel time to work, education and household purposes were also included in the analysis. However, to deal with missing values, each of these indicators was calculated for each household, then three indicators were aggregated, and the mean value was considered as the travel time to services indicator.

To have an insight of the provided data, Annex 3 provides the number and percentage of missing values in each indicator. We can see that excluded variables have a lot of missing values. Also, dwelling age and toilet indicators have many missing values, so they should be used and interpreted with caution.

It is also essential to deal with variations within slums as the unit of the analysis when we want to connect images to socio-economic data is slum and not household. This issue is explained in section 5.1.1.1 and 5.1.4 after constructing index because the method we used to build indices also deals with this issue.

4.3. Understanding slum variations and building deprivation indices

After pre-processing all data, 121 QS samples from the 2017 fieldwork and 1114 samples from 37 slums from 2010 household survey were used for further statistical analysis. The main approach in analysing socio-economic data in this study was to use MCA as it was found very helpful in analysing categorical data with few assumptions. In this sense, we could use all the categorical data we had without making any assumption on their order or their relative importance. In addition, there was no worry that indicators from domains (or capitals) with few indicators have more power than indicators from domains with many indicators. Two deprivation indices were constructed using MCA, one with more focus on physical and contextual capitals (QS index), and another more comprehensive one, covering all the dimensions (HH index).

Considerations of implementing MCA for both indices are as follow. In MCA we can consider also some supplementary variables (i.e., variables that do not contribute to creating the model but are involved to be interpreted on the results) to the analysis. However, as we aimed to find variations of our individuals (households in case of HH and slums in case of QS), we considered all the indicators as primary. Continuous variables were discretised into ten equal interval classes after testing classes from 5 to 15. With ten classes these indicators could be better explained by the final dimensions derived from MCA.

MCA produces several outputs and the most important one among them is the created main dimensions, i.e., projected variables into a lower dimensional space keeping maximum possible variations. Number of dimensions to be extracted from MCA was decided based on Cronbach's alpha more than 0.7, i.e., the most common way to measure dimension's reliability (Field, 2013). As mentioned in section 2.3, MCA creates scatter plots of categories (of indicators) and individuals. These scatter plots are helpful to interpret and understand data. Therefore, scatter plots were used to find relations between individuals and categories and for interpreting the results. Another useful output is the squared correlation between indicators and dimensions showing the extent each dimension can explain each indicator. Indicators with

low values have less variation among individuals. Finally, MCA gives individual scores, i.e., individual coordinates in an n-dimensional space which shows relations between individuals. Farther individuals have more different patterns of indicator categories.

Scatter plots are very helpful for interpreting data, but we needed to end up with one number for each individual¹¹ as its index value to build regression models in further steps, then to connect image-features of each slum to its index value. One option was to mix up values obtained from each dimension, means calculating descriptive statistics between dimension values, or summing up dimension values for each individual. This is not a proper way of using dimensions because some dimensions explain same variations of particular indicators, so by mixing up these dimensions, we will overestimate some of the indicators. As a result, the first dimensions (as the most crucial dimension) of final results were used to build the final indices. This is a common approach in previous studies when using a principal component method (e.g., Rains et al., 2017) (for more details see section 2.3).

In QS index, individuals were slums, and we could connect their values to the image in further steps, but in the case of the HH index, we needed to aggregate household data to end up with one number for each slum. To do this, scores of the HH index for households were averaged to find a value for each slum. Discussion on the internal variations of these slums is provided in section 5.1.1.1.

To verify the results obtained from the built indices, they were checked by the photos captured during the fieldwork to see whether there is a logical relationship. Moreover, an HH index based on ordering indicators (classical method to build an index) was created, then compared with the index constructed using MCA.

4.3.1. Relation between HH and QS results

We explored the correlation between values extracted from HH and QS indices across the 26 samples of the same locations. An important advantage of the QS samples over the HH samples is that they represent the current slums in Bangalore better. Therefore, the aim of connecting the two indices was to see whether an index which is only based on physical and contextual information (QS) can explain deprivation levels¹² and whether we can interpret our results obtained from QS as deprivation levels. If the two indices are correlated, we can have a better insight of deprivation levels in Bangalore using QS samples. A Pearson correlation was calculated using bootstrap (as we had few samples) to see whether there is a relationship between these two indices. Our assumption was these indices must be correlated, as they were describing the same thing from different perspectives. A significant correlation means the two indices describe the same phenomena and obtaining high correlation coefficient R means they can explain variations exist within one another.

4.4. CNN-based system to predict deprivation indices

This section provides an overview of the pre-processing of images and steps in building CNNs. The main goal of this analysis after building deprivation indices was to train a CNN with the ability to predict these indices. “Number of samples” is one of the most issues in training CNNs as these networks have many parameters to tune. In fact, studies usually use tens of thousands of samples to train CNNs (see Chatfield et al., 2014). Nielsen (2015) also stated that having a comprehensive dataset is more important than having a more sophisticated network to obtain a good result. Considering the aim of this study, we had only 26 samples to predict HH index and 121 samples to predict QS index which is very few. On the other hand, we had also boundaries of 1,461 slums, so we took advantage of this abundant data to propose a solution. We initially trained a CNN with the ability to distinguish slums from formal areas using slum boundaries. By training such network, we learned discriminative features to separate slums from formal areas (and consequently, to separate less deprived areas from more deprived areas). After that, we used this trained

¹¹ Individual means household in case of HH data which then we aggregated them into slums, and mean slums in case of QS data.

¹² We assumed that HH index can explain deprivation levels as it covers all the aspects of deprivation.

network (that learned discriminative features) and transformed it to a regression model by changing its objective function from log-likelihood to Euclidean function which changed the behaviour of the network to work like a least square model. We used the limited number of samples with indices values to fine-tune the new CNN parameters and predicted deprivation indices. The process of using a pre-trained network and fine-tuning it for a specific purpose is called transfer learning, and there are many studies in the field of land cover/use classification which took advantage of this idea and used popular pre-trained networks (e.g., Castelluccio et al., 2017). The aim of such studies is to facilitate the training process and invest less time and computational power to train a network. However, we used our pre-trained network and its learned features (and not popular pre-trained networks) to deal with few samples available for our study. Unlike other studies (e.g., Arribas-Bel et al., 2017), we used CNN for deprivation indices predictions with a unique framework trained end-to-end. Although this method is not a standard way of training and using CNN, it allows taking advantage of feature learning capability of deep learning models for prediction using very few samples.

What was explained about training CNN will be elaborated in the following sections. Sections 4.4.1, 4.4.2, and 4.4.3 explain steps to prepare samples and images and section 4.4.4 and 4.4.5 elaborates the process of training networks.

4.4.1. Sample preparation

As mentioned in chapter 3, 1,461 delineated slum polygons were available. These polygons were checked one by one on top of the available images (Pleiades images). Moreover, almost all polygons were not accurate enough to confidently be used for further steps, so they were carefully corrected to match with the images (Figure 10). Many polygons were removed as there was no visible settlement within their boundaries. This could be due to the temporal gap between when the images were captured and when the samples were delineated. Apart from the large number of removed polygons, some new polygons were also added as we could confidently identify them as slum comparing with other slum samples. After correcting, removing, and adding slum polygons, we ended up with 1,121 polygons for further analysis.

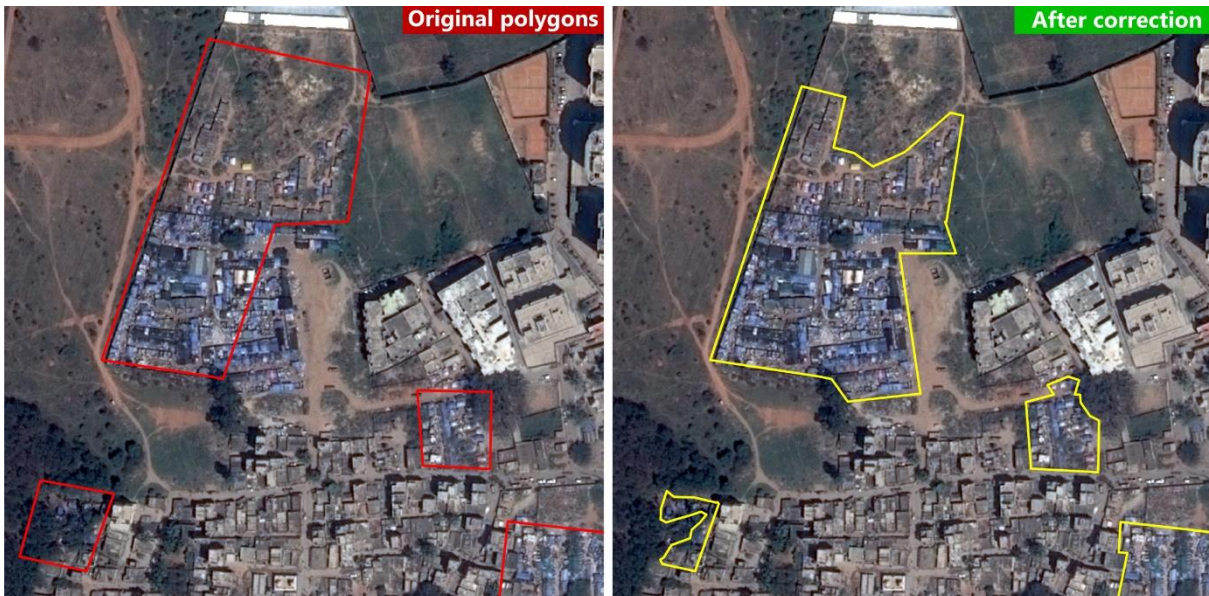


Figure 10: Example of correcting slum polygons

As no sample from formal residential areas was available, a similar strategy as used for the fieldwork was employed. Two sets of tessellations, 250m by 250m and 4km by 4km, were generated covering the entire city. So, 256 small tessellations (250m) were generated within each large tessellation (4km) and 6400 small tessellations in total (Figure 11). Using stratified random sampling, samples were randomly selected from small tessellations in a way to have an equal number of samples in large tessellations. Although Congalton (1991) recommended having 75 to 100 samples to have representative samples for each land use class, 600 tessellations (almost 10% of the population) were selected as we needed a large number of samples to train the CNN. After selecting tessellations, they were zoomed in one by one, and formal residential areas were delineated by drawing polygons around them (Figure 12). Finally, using OSM data, commercial and industrial areas were erased from the delineated polygons. After all these steps, 611 polygons were prepared as formal residential samples for the following steps.

The size of small tessellations was decided in a way that we could easily zoom in and delineate formal residential areas within them. The size of large tessellations was decided that we could divide the whole area into 10 by 10 tessellations. Using these large tessellations reduces the effect of spatial autocorrelation.

To avoid confusion when generating patches, buffers of 150m were generated around slum samples and

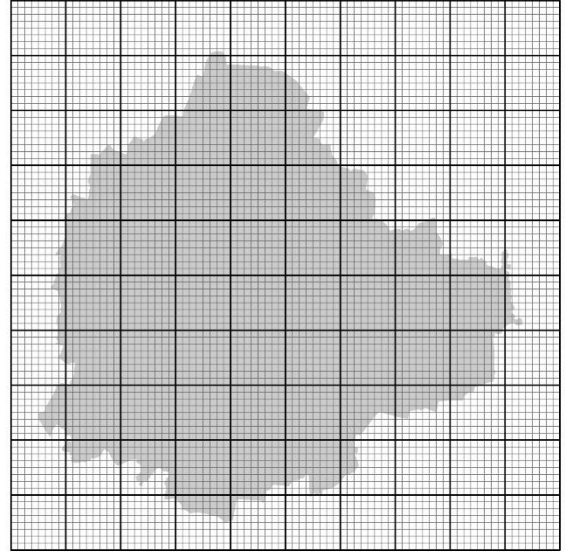


Figure 11: Small and large tessellations for sampling formal areas

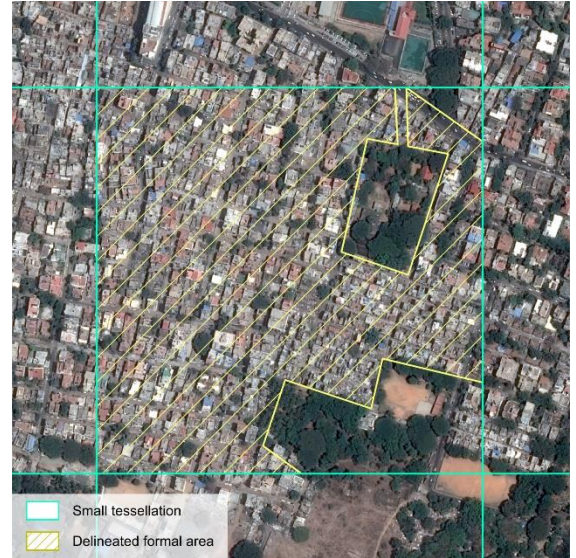


Figure 13: Example of formal sample polygon

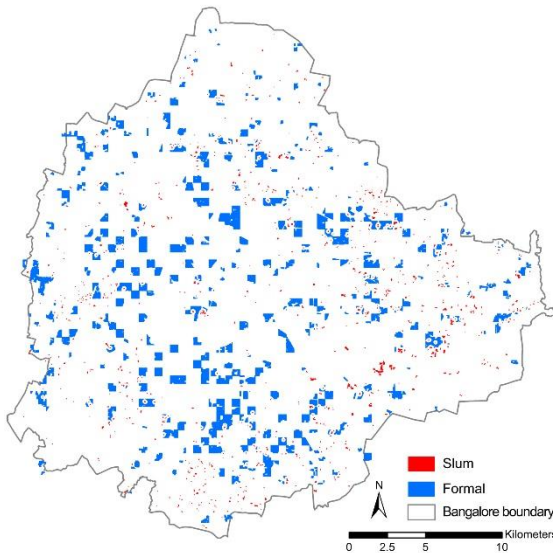


Figure 14: All sample polygons

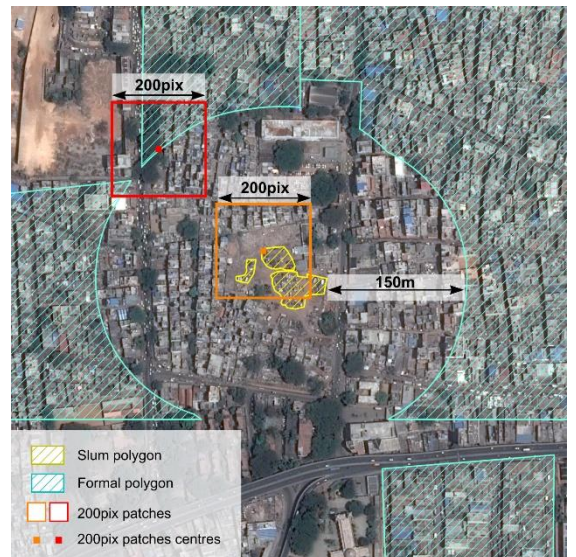


Figure 12: Samples after generating and erasing buffer

erased from formal areas (Figure 13). This allows to generate patches up to 200 pixels (100 meters) on slum and formal areas with no overlap¹³ (see orange and red patches illustrated in Figure 13). Figure 14 shows the location of all final delineated and cleaned polygons.

4.4.2. Image preparation

The study area of this research was almost 713km², and images had the resolution of 0.5m, so we were working with 2.8 billion pixels. Studies that aim to classify slums usually work on smaller areas. For instance, Graesser et al. (2012) worked with 14000 by 14000 tiles, i.e., 196 million pixels, or Kohli et al. (2016) worked on tiles of 3000 by 3000 pixels, i.e., 9 million pixels, and Mboga (2017) used tiles of 2000 by 2000, i.e., 4 million pixels. Creating patches¹⁴ from available polygons was done in MATLAB, and it reads images as numeric arrays. Maximum possible array size to work with is directly related to the amount of RAM and to work with each of four available images of this study, almost 30GB to 40GB RAM was required. This is beyond the capacity of the machine used for this study that had only 16GB of RAM. Therefore, the following method was developed to deal with the computational limitations. Firstly, each available image was divided into two images (see Figure 15 and compare it with Figure 6). This was the simplest way to decrease the size of each image. After that, samples were distributed into eight images ensuring they were not duplicated and images can cover the whole boundary of samples.



Figure 15: Example of a tiny slum

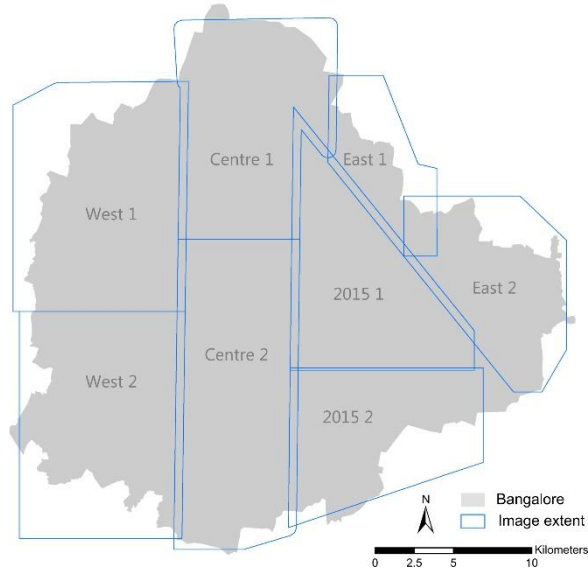


Figure 14: Extent of the eight images after the division

CNN uses a fixed square patch as input, so we cannot have samples of different size as inputs of same networks. However, slums' size varies a lot, and there were a wide range of slums from very tiny ones with only one or two small dwellings (Figure 16) to huge ones with hundreds of dwellings. Studies that aim to detect slums usually focus on large settlements or do the analysis where there are large settlements available and ignore very tiny ones (e.g., Graesser et al., 2012; Kohli et al., 2016). In this study, as we are going to characterize slums and not to detect where they are (in fact we are interested in "What" instead of "Where"), we developed our method in a way that we could keep even very tiny slums in our analysis. Kuffer et al. (2017) considered a radius of 20m to extract contextual patterns. Based on that, a buffer of 20m was created around sample polygons, and this buffer was considered as the context of each sample. Any pixel out of this buffer was set to zero as we were not interested in them. There were two main reasons to do this:

¹³ Note that by generating patches of size 100m, the diagonal of each patch will be almost 150m, so buffer of size 150m was generated.

¹⁴ Steps to generate patches is explained in section 4.4.3.

1) Many slums were located between formal areas; this means when we generate a patch on them, we will have a patch with the majority of formal areas and small slum areas. As the aim of training a CNN to classify slum and formal areas was to generate distinctive features related to each of these two classes, having such patches would bring confusion to the network. This means the network should generate features for slums from patches which are almost formal areas. This is an example for more clarification: suppose we would like to classify images into two possible classes; apple and strawberry. If we have images with a strawberry in the centre and big apples around it, it is more likely that the network classifies these images as apple. The reason is that most of the areas of these images are covered by apples, and the network extracts features which are more related to apples than strawberries. The same logic is what we did with the images. The aim was to create features exclusively for slums and formal areas to be used in the next step for the regression models.

2) Besides CNN, we also extracted some hand-crafted and GIS features to see whether we can improve the results obtained from CNN. We also compared the results obtained from hand-crafted/GIS features with CNN. Using this method, we could use the same patches to generate hand-crafted and GIS features as we cannot use patches with the majority of formal areas to extract features related to slums. Using the same patches for both CNN and hand-crafted/GIS features makes the results obtained from the two models more comparable.

Next step was to randomly distribute samples into two main sets; one for train and validation process and one for evaluating results (i.e., test set). The training set was used to tune parameters (and hyper-parameters) of CNN; the validation set was used to evaluate the tuned network after each epoch. This helps to see the network's behaviour on a dataset apart from the training set and prevent overfitting of the network on the training data. The test set was used to evaluate the final tuned network and to calculate the accuracy of the model on an independent dataset. Samples were randomly distributed into three almost equal sets, so we had 2/3 of our samples for training and validating and 1/3 for testing. In terms of numbers, slum polygons were distributed into two sets of 796 and 325 samples, and formal polygons were distributed into two sets of 429 and 175 sets.

In addition to the original images, a ground truth raster should also be created to show the class of each pixel in the original image. This raster was also created for each image and had three values; 0 meant unlabelled, 1 meant slum, and 2 meant formal.

The process of distributing samples, creating a buffer around polygons, setting all other pixels as zero, and creating ground truth raster was done by creating a model in ArcGIS. Figure 17 illustrates the result of this process on one sample and Figure 18 shows summarized steps of the model. By using this method, train/validation and test sets were completely independent of each other, and we could expect no overlap after creating patches. The model allowed us to easily generate new original images and ground truth raster regarding the selected samples for training/validation or test sets in case of any change in the number of samples for each set, change of any single sample, etc.



Figure 16: Original and ground truth raster after doing sampling steps

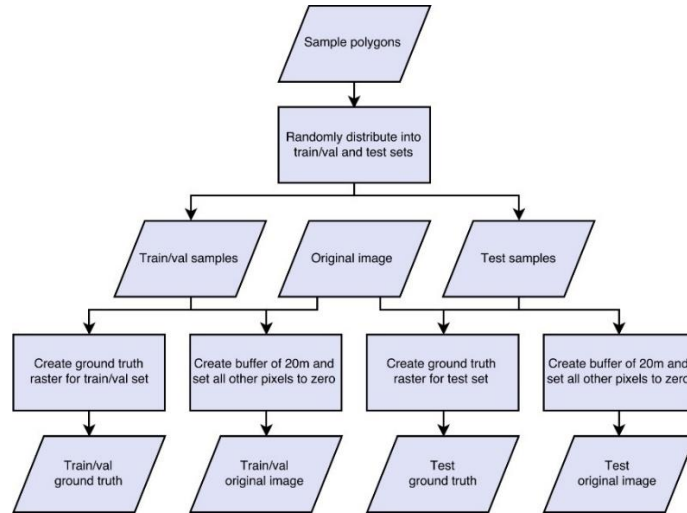


Figure 17: Steps to generate required data for generating patches

Note: This process was repeated once for each image (i.e., 8 times in total)

4.4.3. Patch extraction

By completing the previous steps, we had four inputs (1. Train/validation set original image, 2. Train/validation set ground truth, 3. Test set original image, 4. Test set ground truth) to be used for generating patches for each image. An equal number of patches was generated from each image but with different proportion of slum/formal samples. In each image, the proportion of slum patches depended on the proportion of the area of slum and formal samples in that image in relation to other images, and this also holds true for formal samples. Patches were generated two times. For tuning CNN parameters, 1000 training, 1000 validation, and 400 test samples were generated to find the optimal patch size. Less training and validation samples helped for faster parameter tuning. After tuning the CNN parameters, 2000 training, 2000 validation and 2000 test patches were generated to train and evaluate the final network. Table 2 shows the number of patches generated in each image in case of having 4000 train/validation patches and 2000 test patches. Note that the number of training and validation patches are equal in each image.

Based on Mboga (2017), patches of size 99, 129, and 165 were created to find the optimal size for training CNN (Figure 19). Centre points of patches were randomly generated on the ground truth raster.

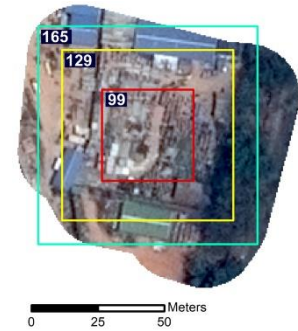


Figure 18: Patch size on a slum sample

Table 2: Distribution of patches along images and train/val and test sets

| Image | Slum train/val pixel | Formal train/val pixel | Slum train/val number | Formal train/val number | Slum test pixel | Formal test pixel | Slum test number | Formal test number | Sum train+val | Sum test |
|--------------|----------------------|------------------------|-----------------------|-------------------------|-----------------|-------------------|------------------|--------------------|---------------|-------------|
| 2015 1 | 699444 | 20123419 | 139 | 111 | 274517 | 4368330 | 167 | 83 | 500 | 250 |
| 2015 2 | 833387 | 10747278 | 184 | 66 | 333846 | 3602236 | 187 | 63 | 500 | 250 |
| West 1 | 392929 | 23547217 | 94 | 156 | 93786 | 8372698 | 66 | 184 | 500 | 250 |
| West 2 | 284144 | 16893297 | 94 | 156 | 106270 | 4090710 | 113 | 137 | 500 | 250 |
| Centre 1 | 445140 | 13706814 | 135 | 115 | 132950 | 9106328 | 80 | 170 | 500 | 250 |
| Centre 2 | 325920 | 36190477 | 61 | 189 | 199612 | 13465628 | 80 | 170 | 500 | 250 |
| East 1 | 146977 | 2127067 | 178 | 72 | 181221 | 2187906 | 181 | 69 | 500 | 250 |
| East 2 | 406691 | 3859710 | 198 | 52 | 172799 | 2629564 | 169 | 81 | 500 | 250 |
| Total | 3534632 | 127195279 | 1083 | 917 | 1495001 | 47823400 | 1043 | 957 | 4000 | 2000 |

4.4.4. Training CNN – simple model

As mentioned earlier, our aim was to train a network which could distinguish slum from formal areas. The reason we did not train a network to directly predict our deprivation indices was that we had too few samples that we knew their index values (for more details see section 4.4). First, a simple CNN network was trained and was evaluated; then a deeper network was used to see whether it can improve the result. The simple network was initialized based on Bergado et al. (2016) and Mboga (2017) with two convolutional layers and one fully connected layer (Figure 20). Only eight filters were used in each convolutional layer, and the width of the fully connected layer was 128. With an input of 129 by 129, this rather simple network had 810,000 parameters to learn. As the activation function, ReLU function was used as it is the most common and effective activation in image recognition problems (e.g., Krizhevsky et al., 2012). The log-likelihood objective function was used in the network accompanied by the last softmax layer with two neurons; each shows the probability of a given input to be one of the specified classes (i.e., slum or formal). In this sense, the network works similar to a logistic regression model. This is also a common architecture in image recognition studies (e.g., Chatfield et al., 2014). To regularize the network and prevent overfitting, drop-out layers were used after each convolutional layer. Moreover, weights were initialized as $\sqrt{2/\text{number of input neurons}}$ based on He et al. (2015) to prevent saturation in the network and increase learning pace. Another approach to speed-up learning was to give higher learning rates for the first epochs and gradually decrease it when the learning curve is converging. Table 3 shows a summary of network's hyper-parameters.



Figure 19: Architecture of simple CNN

Table 3: Simple CNN hyper-parameters

| | |
|---------------|--------------------------|
| Batch size | 64 |
| Learning rate | decrease logarithmically |
| Weight decay | 0.0005 |
| Momentum | 0.9 |

Training of the network was carried out with MATLAB and MatConvNet library (Vedaldi & Lenc, 2015). Networks were compiled on GPU that can significantly improve learning speed (Vedaldi & Lenc, 2015). The training process for this study was done on NVIDIA QUADRO 1000M GPU (NVIDIA, 2018c) with CUDA toolkit (NVIDIA, 2018a) and cuDNN library (NVIDIA, 2018b) that has been developed to accelerate deep learning pace.

The network was trained three times, each time with different patch size to find the optimal result and with maximum 700 epochs. Networks were evaluated on the test set by the percentage of correctly predicted patches. Patch size with a higher accuracy was used for further steps.

4.4.5. Training CNN – deep models

After training a rather simple network, we also took advantage of using popular networks in the field of image recognition. These networks have been developed to solve very complex problems, classifying images with up to 1000 classes. Although they have not been developed specifically to work with satellite images, studies showed their good performance on such problems (e.g., Castelluccio et al., 2017).

Due to the complexity of these networks and the limitation of the hardware used for this study, it was not possible to test all these networks on our problem. Networks from Visual Geometry Group (VGG) from University of Oxford (Chatfield et al., 2014) was selected to be used for this study. These networks were

selected as they obtained relatively good results in comparison with other popular networks to classify ImageNet dataset¹⁵ (MatConvNet Team, 2017). Furthermore, they have a relatively smaller number of parameters, so we could train them on our machine. This set of networks has three versions; VGG Fast (VGG-F), VGG Medium (VGG-M), and VGG Slow (VGG-S). VGG Very Deep (VGG-VD) was also developed that was deeper and more complex (Simonyan & Zisserman, 2014). VGG-F that has fewer parameters was selected for this study. It accepts 224 by 224 inputs with three channels. To use this network, we should slightly change the network, add one channel as input, change the output to two classes, and fine-tune the network. This is a standard way of using a pre-trained network, but we should use the same input size. Due to a large number of parameters in such network, required GPU RAM exceeded the capacity of our machine. Therefore, a VGG-Like model was trained from scratch using our patch size as input (Figure 21). It is worth mentioning that the network had 28,000,000 parameters to learn.

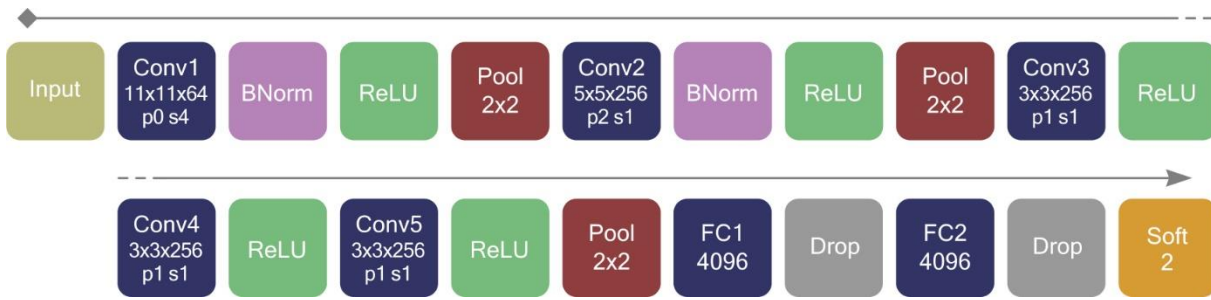


Figure 20: Architecture of VGG-like CNN

The VGG networks use Local Response Normalization (LRN) instead of Batch Normalization (BNorm) (Chatfield et al., 2014), but we used Batch Normalization instead since it is more effective and by using it we can remove LRN (Ioffe & Szegedy, 2015).

After evaluating the VGG-like network, we used the idea of image augmentation (Simard et al., 2003), to increase the number of training samples. Based on Scott et al. (2017), each patch was rotated in seven directions; 7, 90, 97, 180, 187, 270, and 277 degrees with linear interpolation. This is a logical way to increase the number of samples as for instance looking at a slum from any angle does not change what it is. By doing this, our 2000 training patches increased to 16000. The VGG-like network was tuned and trained again to explore any improvement.

4.5. Supplementary hand-crafted and GIS features

In addition to using CNN, some supplementary hand-crafted and GIS features were also added to the features learned by the CNN to see whether they can contribute to improving predictions. Although with using hand-crafted features we are not restricted to fixed patch size, the same patch size as CNN was used with the same excluded areas (see 4.4.2) as the aim was to improve what we already had from CNN. Furthermore, using the same patches make the analysis between CNN and hand-crafted and GIS features comparable.

Section 4.5.1 to 4.5.3 elaborates steps to extract hand-crafted features from patches, and section 4.5.4 looks one step beyond a settlement and considers the spatial location of a patch.

¹⁵ see MatConvNet Team, (2017) for a comparison between the performance of the most popular pre-trained networks.

4.5.1. Spectral information

The most straightforward statistics we can make from the original image bands is to calculate statistics from band values. They are important as they reflect land cover variations exist in different regions and many studies took advantage of using them (e.g., Arribas-Bel et al., 2017; Duque et al., 2017).

Table 4 shows values extracted from each patch as features. In total, 10 features were extracted for each patch.

Table 4: Hand-crafted features - band ratios

| Feature name | # of feature |
|---|--------------|
| Band 1 to 4 mean and standard deviation | 8 |
| NDVI mean and standard deviation | 2 |

4.5.2. GLCM

One of the most common texture features that has been used in image analysis studies is Grey Level Co-occurrence Matrix (GLCM). It describes the relationship between a pair of co-occurrence pixels in an image (Arribas-Bel et al., 2017). We can extract many GLCM features varying lag and direction. Then, many properties can be calculated from the extracted GLCMs. Another option when using GLCM is to choose the number of grey levels to be considered in the image. Lower numbers of grey levels, simplify the patches more.

Four directions (i.e., $[i\ 0]$, $[0\ i]$, $[0\ -i]$, $[-i\ i]$) which presented in Duque et al. (2017) with one to four pixel lags were decided to extract features (Figure 22 shows 1 pixel lag). Based on Kuffer et al. (2016), three properties; entropy, variance, and contrast were calculated on each feature. Table 5 illustrates GLCM features.

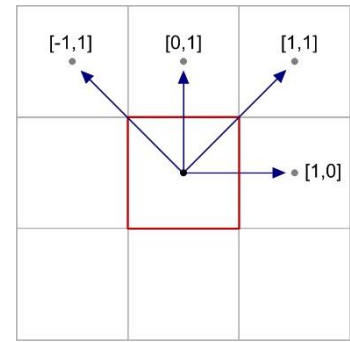


Figure 21: four directions of GLCM with 1-pixel lag

Table 5: Hand-crafted features - GLCM

| Feature name | # of feature |
|--------------------------------|--------------|
| Contrast, 4 lags, 4 directions | 16 |
| Variance, 4 lags, 4 directions | 16 |
| Entropy, 4 lags, 4 directions | 16 |

To decide on the number of grey levels, features with 4, 8, 16, 32, 64, 128, and 256 were extracted and correlated with QS index (as we had more samples there) and the one with the highest correlation coefficient was selected to be used for predictions.

As some patches had many zero values due to buffering procedure as explained in section 4.4.2, first GLCM with 12bit (i.e., same as our image dynamic range) grey level was created, and the number of $[0\ 0]$ co-occurrence pixels was deducted from $[0\ 0]$ GLCMs with lower grey levels. This was done to keep the GLCMs more representative of image features inside the buffers (i.e., settlements with their contextual information). Another point to mention is as we did not have a panchromatic band of our images and they were pansharpened bands, GLCM properties were calculated on each band of a patch, and the mean value was considered as the patch GLCM property value.

4.5.3. LBP

One of the most potent texture features used in studies related to informal settlements is Local Binary Pattern (LBP) (Ella et al., 2008). It is a binary representation of circular neighbouring pixels around a centred pixel showing which have higher and which have lower values than the centre pixel. Going clockwise (or counter-clockwise) across neighbouring pixels, changing value from zero (i.e., lower value

than the centring pixel) to one (i.e., equal with, or higher value than the centring pixel) or vice versa, is called one transition. Patterns with maximum two transitions are called uniform, and they describe the most important textural information about an image (Ojala, Pietikäinen, & Mäenpää, 2002). These patterns were extracted for each patch altering radius and number of neighbours. These patterns are extracted within a given cell size, so they can extract local features of an image. As we were working on patches and not the whole image, the whole patch was considered as a cell.

Based on Abeigne Ella et al. (2008), $LBP_{8,1}^{riu2}$ (i.e., rotation invariant uniform patterns with radius of 1 considering eight neighbours) (Figure 23), $LBP_{16,2}^{riu2}$, and $LBP_{24,3}^{riu2}$ were considered to extract features for each patch (Table 6). For n number of neighbours, we can extract n+2 uniform patterns as there are always two patterns with zero transition (all zero, all one). After extracting LBP of each band, they were averaged to get the value for a patch. In case that a point fell on the boarder of two pixels, its value was linearly interpolated considering other pixel values.

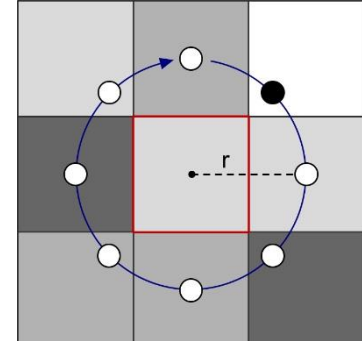


Figure 22: A uniform LBP with two transitions, radius one and eight neighbors

Table 6: Hand-crafted features - LBP

| Feature name | # of feature |
|---------------------|--------------|
| $LBP_{8,1}^{riu2}$ | 10 |
| $LBP_{16,2}^{riu2}$ | 18 |
| $LBP_{24,3}^{riu2}$ | 26 |

4.5.4. GIS layers

The last group of hand-crafted features is related to the context of each patch (contextual capital). Although OSM data are not officially validated, they provide extensive information about roads, and different land uses. The ability of such data to improve model predictions was also examined as they are publicly available. As road data are not consistent enough to perform network analysis in the area, we calculated Euclidean distances to patch's centre points. Distance to different land uses and public services are used to calculate the deprivation level of settlements especially in deprivation indices developed in the UK (e.g., Welsh Government, 2014). The minimum distance from each of the ten layers (see Table 7) was assigned to each patch. Town hall was considered as the centre of the city (and it is also located almost at the centre of the city). Annex 4 provides a map showing all the OSM layers used to create the GIS features. We considered all the facilities and water bodies within a buffer of 6 kilometres around the city to avoid losing information outside the city boundary (see Annex 4).

World elevation data is also publicly provided by ESRI (Nagi, 2014) and downloadable for all regions, so we also used DEM of Bangalore to calculate mean elevation and mean slope within each patch (Kuffer et al., 2017). Indeed, we excluded areas out of the buffer we created around each settlement. DEM raster which was used for this analysis had a resolution of 11 meters. Table 7 summarizes GIS layers used to extract features for each patch.

Table 7: GIS features

| Feature name | # of feature |
|---|--------------|
| Distance to: 1) main road, 2) bus stop, 3) healthcare, 4) leisure activities, 5) pharmacy, 6) railway station, 7) railway, 8) school, 9) waterbody, 10) town hall | 10 |
| Elevation and slope mean | 2 |

4.6. Regression models

After training the CNN and extracting all hand-crafted features, we examined the relationship between socio-economic variations exist in our indices and image-based features. It is worth mentioning that patches from HH and QS samples were created in a way that the centre of the patches was located on the centre of sample polygons. Section 4.6.1 explains initial results of the CNN used to predict indices, section 4.6.2 elaborates fine-tuning steps took to build regression models with our trained CNN, section 4.6.3 explains methods to build regression models using hand-crafted and GIS features, and section 4.6.4 examines combining regression models from CNN and hand-crafted/GIS features to improve prediction.

4.6.1. Regression using CNN output

We first examined the ability of the softmax layer from the CNN which can predict and distinguish formal areas from slums, to predict indices. For a given patch, the network gives two probability numbers, first is the probability of a patch to be a slum area, and second is the probability of a patch to be a formal area. The sum of these two numbers is always equal to 1. Therefore, we had a number between 0 and 1 for each patch shows the probability of a patch to be a slum. Using this value to predict indices makes sense as the assumption is that more deprived slums are morphologically very different from formal areas (Krishna et al., 2014). Linear and polynomial regression models were built to examine the performance of the CNN output to predict indices.

To assess the accuracy of regression models in this study, cross-validation methods were used as we did not have a large number of samples to divide them independently into train and test sets. Two widespread cross-validation methods were used to prevent an overfitting of the model; k-fold cross validation for assessing models on QS index and we set k as 10; and Leave One Out (LOO), that is a specific case of k-fold cross validation when k is equal to number of samples, and we used it to assess models on HH index (Refaeilzadeh, Tang, & Liu, 2009). To predict QS index values, models were trained 10 times (once for each fold) and to predict HH index values, models were trained 26 times (equal to the number of samples). In case of the QS index, we had 191 training samples and 12 test samples, and in case of the HH index, we had 25 training samples and 1 test sample for each model. In this manner, we predicted all samples we had, and our test sets were not involved in the training process. To assess the overall predicting power of the models, the coefficient of determination (R^2) was used as defined in Field (2013):

$$R^2 = 1 - \frac{\sum(y - y_p)^2}{\sum(y - \bar{y})^2}$$

Equation 6: calculating R^2

where y is the observed index value, y_p is the predicted value, and \bar{y} is the average of y values. In this study, all the regression models were assessed using explained cross validation methods (4.6.1 to 4.6.4)

4.6.2. Fine-tuning CNN to predict indices

Next step in building regression models was to use the trained CNN to directly predict index values. To do this, we needed to slightly change the network architecture and fine-tune it with few epochs of training. The input of the network remained unchanged, but we changed the output of the network with only one neuron instead of two. We also changed the objective function from Log-likelihood to Euclidean loss. This allows the network to work like a least square model.

The network was trained 10 times for the QS index and 26 times for the HH index, and we allowed the learning process to run 100 epochs for each network to ensure convergence. Indices were predicted two times, once with and once without image augmentation with the same approach as section 4.4.5.

4.6.3. Regression models with hand-crafted and GIS features

To predict indices using hand-crafted and GIS features, we needed to deal with a large number of variables in relation to the number of samples. Although Knofczynski and Mundfrom (2008), and Austin

and Steyerberg (2015) proved that we can use even two samples per variable in a linear regression for prediction keeping significant coefficients, using a large number of variables can easily end up in an overfitted model. Two main regression methods to deal with this problem are Partial Least Square Regression (PLSR) and Principal Component Regression (PCR)¹⁶. Although there is a risk of losing some non-collinear but important variables in regression models when using PCR, we selected this method as we used cross-validation methods in assessing our models and the risk of overfitting in PLSR is higher than PCR. The reason is, when we use PLSR, components of regression equation were created in a way that they explain training dependent variables, and when we apply it to the test data, it does not necessarily explain them as well. On the other hand, PCR removes multi-collinearity on independent variables and has less biased predictions.

It is also essential to decide on the number and degree of components in the regression model. To find the optimal number of components, each model was created using stepwise forward-backward regression altering number of components and results were compared. Also, each regression model was created with three configurations: linear only; linear and allowing interaction between variables (i.e., the multiplication of a pair of variables was also used as a variable); quadratic with interaction. We also used different combinations of hand-crafted and GIS features to test their prediction performance: hand-crafted + GIS feature; only GIS features; only hand-crafted features. This would sharpen the importance of relying only on single patches or considering spatial configurations. Figure 24 shows different combination of PCRs created. We used all the combinations from step 1 to 4 to build regression models. As an example, we selected HH from step 1, GIS features from step 2, 5 components from step 3, and Quadratic + interaction from step 4, to build a stepwise PCR to predict HH index using GIS features and maximum 5 components in a quadratic regression model.

| | | | | | | | | | | | | | |
|---------------------|---------------------|---|---|---|---|----------------------|---|---|---|-------------------------|----|----|--|
| ① → Index | Predicting HH index | | | | | Predicting QS index | | | | | | | |
| ② → Features | GIS | | | | | Hand-crafted | | | | GIS + Hand-crafted | | | |
| ③ → # of components | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
| ④ → Complexity | Linear | | | | | Linear + interaction | | | | Quadratic + interaction | | | |

Figure 23: PCR combinations

To create each model, first a PCA performed on the training data and components were extracted, then a regression was performed, then components were reconstructed using test data based on the coefficients obtained from performing PCA on the training data, and finally, values were predicted on the components created with test data.

In total, 72 stepwise PCRs to predict HH index and 108 stepwise PCRs to predict QS index were performed¹⁷, each used cross-validation method explained in 4.6.1 to assess its result.

4.6.4. Combining results

The final step in predicting indices was to combine the results derived from CNN and hand-crafted/GIS features to see whether we can improve predictions. For each index, the best-performed regression models created with hand-crafted/GIS features were selected, and their results were put in stepwise regressions. In total, three combinations were tested for each index allowing linear to multi orders polynomial models to track improvements.

¹⁶ In fact, both methods try to shrink many dimensions to few ones keeping maximum variance and reduce multi-collinearity. The difference is, PLSR tries to create components in a way that explains maximum variance of dependent variables, but PCR maximizes the variance of independent variables (Mevik & Wehrens, 2007).

¹⁷ The maximum number of components used to predict each index was based on the behavior of the models. We used maximum 12 components to predict QS index and maximum 8 components to predict HH index.

5. RESULT AND DISCUSSION

This chapter provides the results and discussions of the analyses. Section 5.1 explains the results of the MCA analysis with a comparison of this method and classical indexing. Section 5.2 provides the results of CNNs in detecting and distinguishing slums from formal settlements. Finally, section 5.3 elaborates results of connecting image-based features with deprivation indices.

5.1. Build indices using MCA

As mentioned in chapter 4, MCA is used to analyse our HH and QS data and to build deprivation indices. This section provides results and interprets them to explore variations exist among slums in Bangalore.

5.1.1. HH index

We built a deprivation index using data of 1114 households. Table 8 shows a summary of the MCA on the HH data. The table shows the first three dimensions, which have a Cronbach's Alpha higher than 0.7. These dimensions explain up to 64.1% of the data variations. Eigenvalue shows the sum of squared correlation of indicators with each dimension. Inertia value in Table 8 shows the total inertia of each dimension which is calculated by dividing Eigenvalue by the number of indicators (i.e., 16 in our case).

Table 8: Summary of the MCA model for the HH data

| Dimension | Cronbach's Alpha | Variance Accounted For | |
|--------------|------------------|------------------------|---------|
| | | Total (Eigenvalue) | Inertia |
| 1 | 0.804 | 4.056 | 0.253 |
| 2 | 0.730 | 3.168 | 0.198 |
| 3 | 0.714 | 3.029 | 0.189 |
| Total | | 10.252 | 0.641 |

As mentioned earlier, we selected the first dimension as our HH deprivation index because it makes the results more comprehensible and interpretable. Furthermore, it is a conventional method to use the first dimension of a principal component method (e.g., Rains et al., 2017) (for more details see section 2.3). Dimension 1 explains only 25.3% of the data variations and seems it cannot indicate deprivation across our samples very well. However, we should consider that having J indicators and K categories in total, MCA creates K-J dimensional space (see section 2.3.1). In our case, it created a 102-dimensional space to find patterns of data. If we compare it with an ordinary PCA, on the same data, it creates only 16 dimensions. Therefore, we can expect lower inertia values in MCA in comparison with PCA methods on continuous data. Greenacre (2017) explains that inertia values are artificially low in MCA and underestimate the real ability of created dimensions in explaining variations existing in raw data. Coulangeon (2017) used MCA to analyse a set of 17 social indicators with 44 categories in total (27-dimensional space) on 4570 samples, and the first dimension of their analysis explained 12.8% of their data variations. Coulangeon and Lemel (2007) used MCA to analyse surveyed data with 9 indicators, and 21 categories (12-dimensional space) and the first dimension of their analysis explained 21% of the variations. Considering these studies, it is logical to use the first dimension which explains 25.3% of the data variations as the HH index.

Another critical point is the interpretability of results. Using one dimension, we can interpret individuals along two sides (like having better-off slums at the positive side and worse-off slums at the negative side). By adding the second dimension, individuals should be interpreted along four sides, so it will make the interpretation complicated. The problem becomes more challenging if we add the third dimension, as the individuals should be interpreted along six sides (see Figure 25). Therefore, we used the first dimension as an index as it describes the variations of our data in an acceptable level, and it also keeps the results simple enough for further interpretation.

MCA creates point clouds of variable categories and individuals, and this can help a lot in the interpretation of the result. Although both variables and individuals are plotted on the same diagram, Almeida, Infantosi, Suassuna, and Costa (2009) suggested to interpret them independently. Therefore, we interpret each of them separately but also mention the relation of these two in general.

5.1.1.1. Interpreting HH individuals (households)

Considering the pattern of categories each HH individual (i.e., a household) has, it is located in the point cloud along MCA dimensions. To have an overview of the result, Figure 25, shows the plotted individuals along the three dimensions created by MCA¹⁸. The closer the points, the more similar their patterns, so points that are far away from each other have very different patterns and we can expect very different situations in terms of deprivation. The closer a point to the origin of the point cloud (in this case [0,0,0]) the more average (and common) categories they have. Thus, such a point does not have a very distinct pattern compared to other points. Therefore, points which are very far away from the origin are very different from the common categories existing among all points. Consequently, points around the origin have more mixed patterns and do not have a very distinct character. In Figure 25, we can see that most of the points are gathered around the mean, showing an almost regular pattern of points along dimensions.

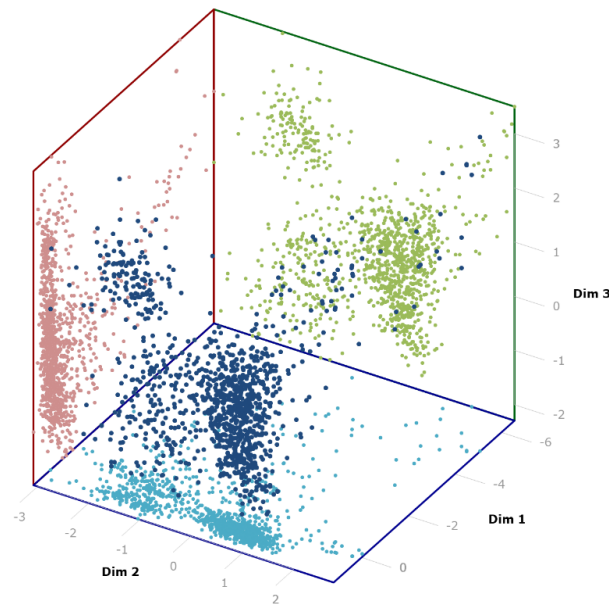


Figure 24: Scatterplot of households in a three-dimensional space.

Note: Light blue, red, and green dots show the projection of the 3D points (i.e., dark blue points) on each plane

To simplify the interpretation, we only used the first dimension created by the MCA (i.e., HH deprivation index) and values of each individual in this plot shows its deprivation index value (Figure 26).

¹⁸ To see how the individuals are located along the MCA dimensions see section 2.3.1.

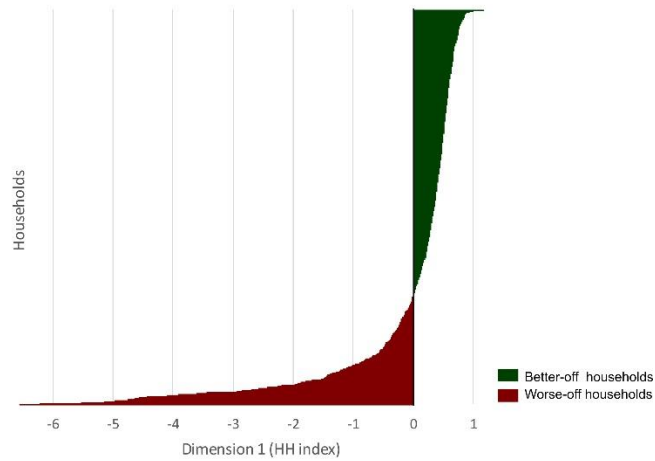


Figure 25: Plot of households along dimension 1

What we can say about this plot is relative, but we can see that better-off households (better than average, green in Figure 26) are abundant, but they are less different from each other and from the average situation. There are fewer households which are worse-off (red in Figure 26) compared to the average situation (i.e., value 0), but they are very different from the average situation. By aggregating households into slums, we can compare these results with photos from the fieldwork.

Figure 27 shows the result of averaging index values across slums. To indicate internal variations, ± 1 standard deviation shown for each slum. Comparing the average value of standard deviations among the better-off and the worse-off slums, we can see that this value is two times more in worse-off slums (see Figure 27). This shows that the internal variations (of deprivation) among the worse-off slums is more than the better-off slums. This means that indicators which more contribute to dimension 1, have more variations among households in the worse-off slums (indicators will be interpreted in section 5.1.1.2). To this end, the better-off slums have less internal variations and are more similar to the average situation, but the worse-off slums have more internal diversity, and many of them are significantly different from the average situation. Comparing average values and standard deviation of aggregated values, we can say averaging values of households for each slum is a reasonable way of aggregating values¹⁹, but we should consider these variations after predicting average values for policy implications²⁰. This means, as an example, if we predict a value of -2.5 (see Figure 27), we should consider that not all the households have situations similar to value -2.5. Instead, many have better, and many have a worse situation, so it is important to consider these facts for policy implications. Moreover, these internal variations could result in more errors when predicting deprivation from images, as different groups of households with different deprivation levels are living in the same slums.

To see what do these numbers mean in reality, we compared the results with photos taken during the Quick Scan fieldwork. Although the data of 37 slum was analysed, we knew the location of 26 of them (and we had photos of 26 of them which can show situations close to the situations they had in 2010). Therefore, we could not provide photos from all slums. We selected photos to show better-off, average, and worse-off situations. The plotted numbers show, better-off slums are less different (see number 3 and 4 in Figure 27), and they are similar to the average situation (number 4 in Figure 27 is a sample close to the average situation), but worse-off slums are very different from the average situation (see number 1 and 2 and compare it with number 4 in Figure 27). Note that what we can see on photos gives a general idea of the overall situation of these slums and are mainly related to the physical and contextual capitals. Other

¹⁹ The result obtained from the MCA is compared with manual indexing and the internal variations of single indicators is compared with the MCA results in section 5.4.1.

²⁰ This study does not cover developing policy recommendations for these areas, but it is a fact which should be considered in further studies.

aspects of deprivation cannot be seen. One question arises here: Does picture 4 (Figure 27) represent the average situation of slums in Bangalore? According to Krishna et al. (2014), settlements like picture 4 are notified, better-off slums in Bangalore. Although we checked all the settlements using Google Earth to confirm that the samples have not experienced significant changes as of 2010, we also confirmed the results with local experts to ensure that such better-off settlements are slums. The local experts also confirmed that these settlements are upgraded, better-off settlements, but officially they are still known as slums. The main question is, do HH samples represent the common situation of slums in Bangalore? We address this question in section 5.1.2.

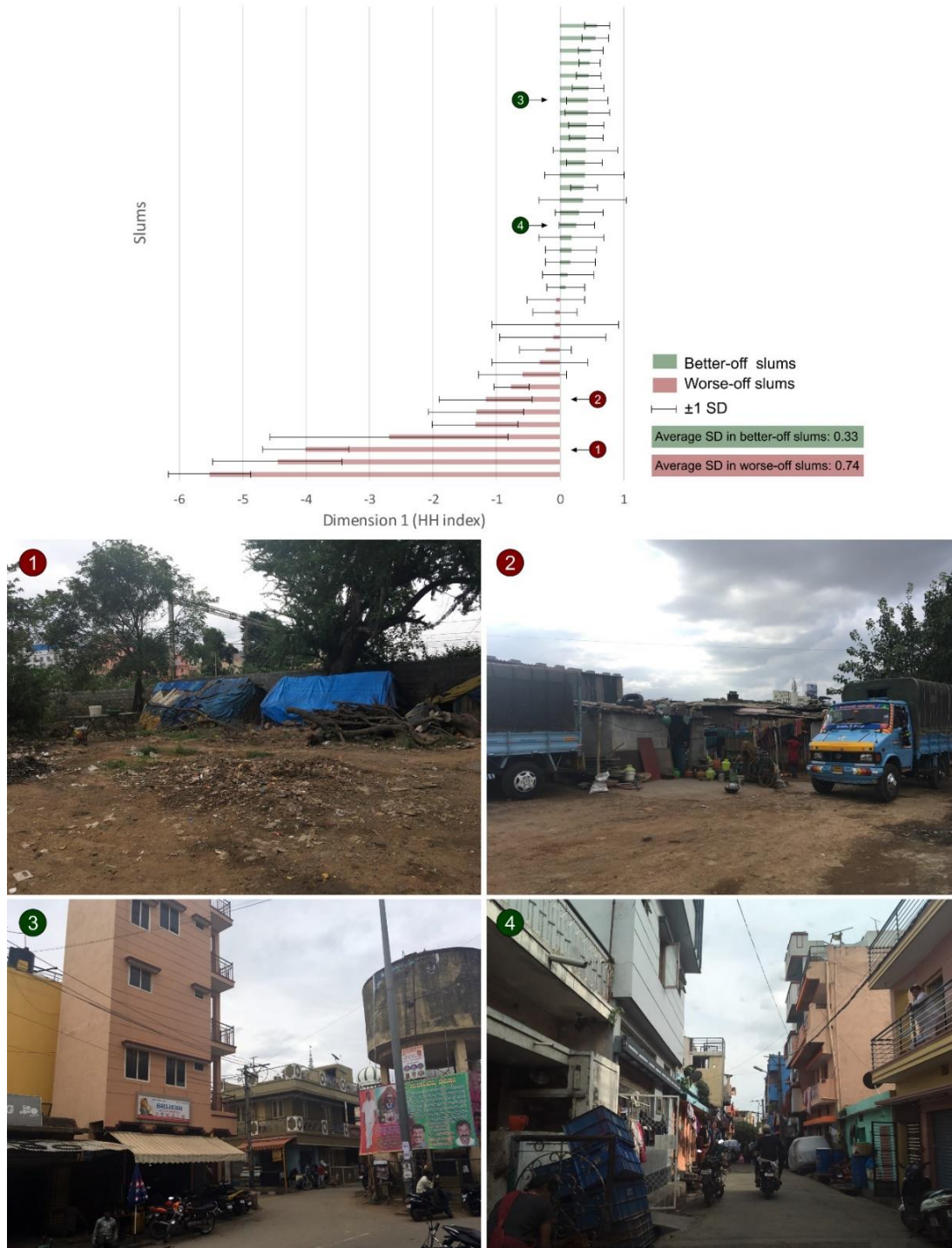


Figure 26: Index values aggregated into slums with some examples

5.1.1.2. Interpreting HH variables (indicators)

Now we focus on indicators and their categories to see their contribution to creating the deprivation index. Figure 28 shows the squared correlations of indicators and MCA dimension 1. The higher the value of this graph, the more variance of indicators are explained by the deprivation index. In this sense, electricity, floor material, and wall material are the most critical indicators in distinguishing households' deprivation. It is interesting that the contribution of indicators related to the physical capital is significantly higher than indicators of other capitals. Thus, we cannot observe distinct patterns related to other domains of deprivation among better-off and worse-off households. To see whether other dimensions of MCA focus on other capitals, Annex 5 provides squared correlation values for all three dimensions. The results show, indicators from other capitals are not explainable well even in other dimensions.

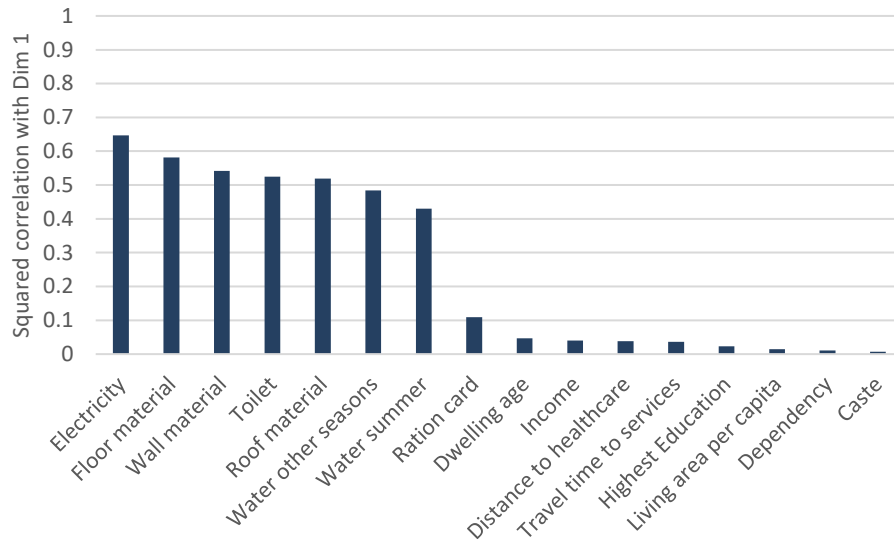


Figure 27: Squared correlation of indicators with MCA dimension 1

Like what we have done in plotting households along dimension 1 (see Figure 26), we also plotted categories to see the most important ones in making distinct patterns of better-off and worst-off households (Figure 29).

According to Figure 29, the most important categories in distinguishing households belong to indicators with highest squared correlation with dimension 1. Most common categories are located near the origin, and rare cases are far away. The explanatory power of categories (i.e., inertia) on the positive side is less than the exploratory power of categories on the negative side to distinguish households from each other. Thus, categories on the negative side make households being very different from the average situation, but on the positive side, households are only slightly different from the average. The reason is, very negative categories are significantly rarer than very positive categories.

We can see relationships between Figure 29 (plot of categories) and Figure 26 (plot of households). As explained in section 2.3.1, the plot of categories and the plot of individuals are in fact created in the same space. Therefore, it is logical that categories and individuals with close values have some relationships. For instance, we can say having surface water makes the worst-off household very distinct from other households, as it is scarce; however, it does not mean that any household with a close value to any category in the plot has that category in its pattern. Location of each household in the plot is the result of overall interactions between all categories. Therefore, we may only predict categories of households that are very extreme cases. Two households near the origin (i.e., 0) for example, do not have precisely the same categories, but mostly have categories with the same importance (have categories located around the origin in Figure 29).

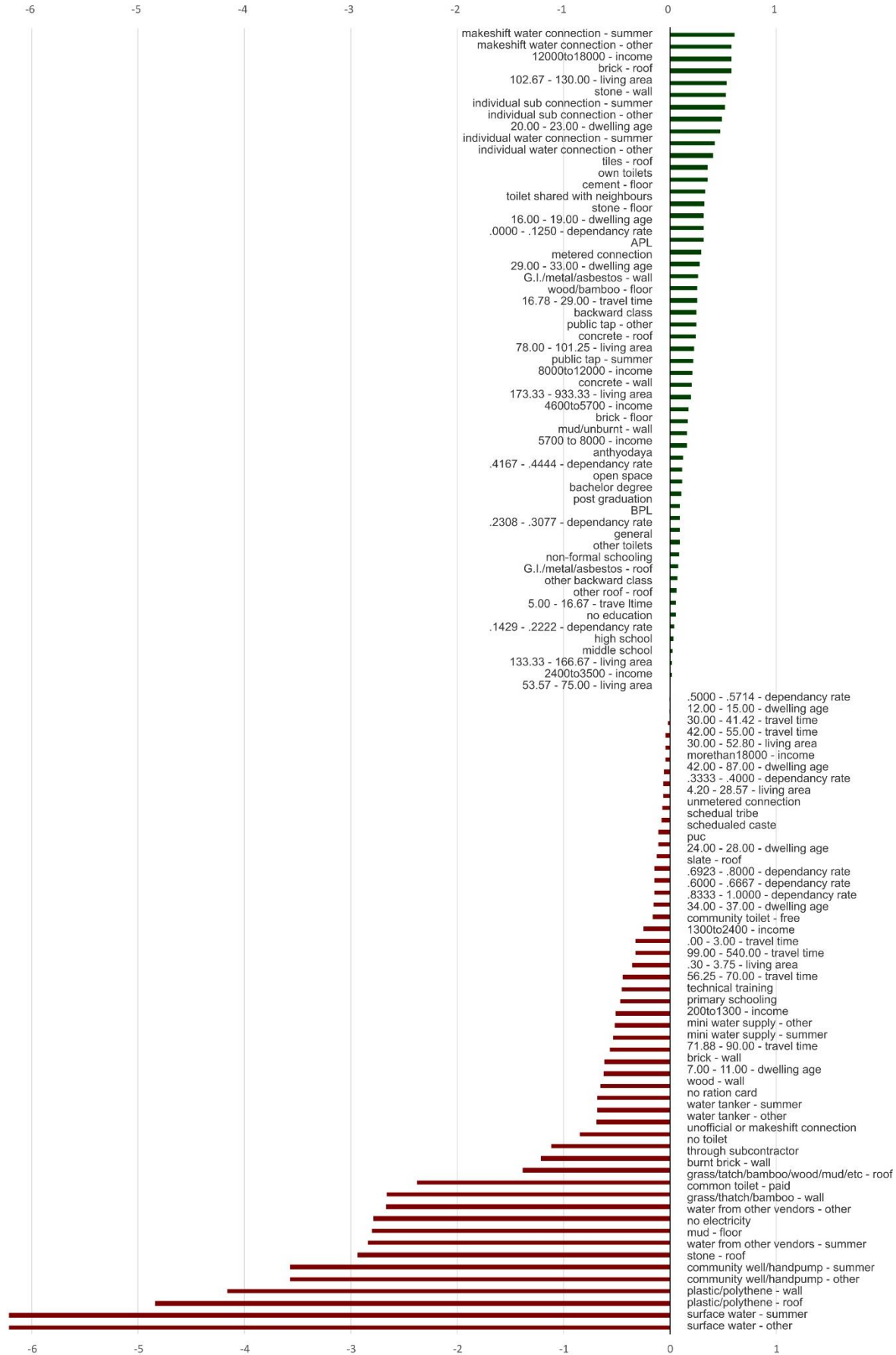


Figure 28: Indicator categories along dimension 1

5.1.2. QS index

The second index is based on the fieldwork conducted for this study. The Open sewers indicator was dropped from the analysis as it had zero variance. 34 indicators with 107 categories used to build the QS index. Thus, MCA created a 73-dimensional space to locate individuals. The first dimension explains 30.9% of variations (Table 9) which is an acceptable number based on section 5.1.1. Similar to the HH index, we focus on the first dimension as QS index and interpret the results based on that.

Table 9: Summary of the MCA model for QS data

| Dimension | Cronbach's Alpha | Variance Accounted For | |
|--------------|------------------|------------------------|---------|
| | | Total (Eigenvalue) | Inertia |
| 1 | 0.932 | 10.505 | 0.309 |
| 2 | 0.821 | 4.919 | 0.145 |
| 3 | 0.757 | 3.765 | 0.111 |
| 4 | 0.749 | 3.656 | 0.108 |
| 5 | 0.733 | 3.467 | 0.102 |
| 6 | 0.708 | 3.201 | 0.094 |
| Total | | 29.514 | 0.868 |

First, we focus on individuals (i.e., slums) to see whether a clear pattern exists within the QS index (Figure 30). As HH slums were also surveyed during the QS fieldwork, we showed them with light colours in the figure to indicate their positions in this index. Most of the slums are located on the negative side, and better-off slums are very different from the average situation. HH slums, which were discussed in section 5.1.1.1 are mostly indicated as better-off, showing that HH slums do not represent the overall slum situations of Bangalore²¹. About the QS samples, we selected slums from different parts of the city for the fieldwork, so they are more representative than HH which its samples are mostly located around the city centre. Considering Annex 6, most of the slums are located at the periphery, but the HH samples are concentrated around the city centre. This is the reason that when we were analysing HH slums, the average situation seemed better than it was (see section 5.1.1.1). Figure 32 shows the QS slums on a map. Yellow points are near the average situation of slums in Bangalore, and we can see the average situation is very different from what we saw before about the HH samples. In Figure 32, picture 2 represents the average situation of slums in Bangalore, which is significantly different from picture 4 in Figure 27. Considering the QS index in Figure 32, the worst-off slum has almost 1 unit difference, and the best-off slum has almost 3 units difference. Thus, the worst-off slum should be less different from the average than the best-off slums. Photos also show the same situation. In Figure 32, picture 1 is less different from picture 2 than picture 4.

²¹ From 37 HH samples, we used only 31 samples beside other QS samples, as 6 HH samples changed significantly in the period of 2010 to 2017 (verified with Google Earth).

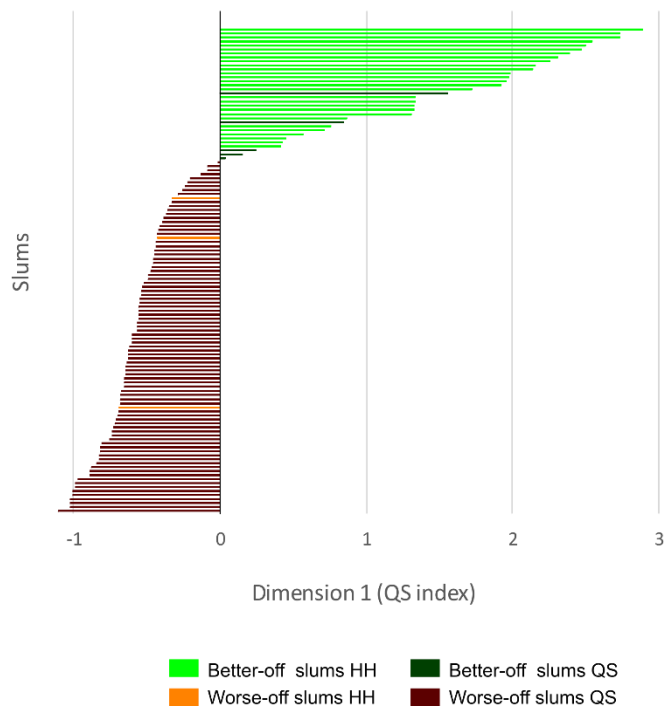


Figure 29: Plot of slums along dimension 1

Figure 31 shows squared correlation of indicators with dimension 1, to provide an overview of important indicators. Dominant roof material and dominant building footprint size are the most important indicators contributing to the QS index. For more details, Annex 7 shows squared correlations of all indicators with all six dimensions created by QS data.

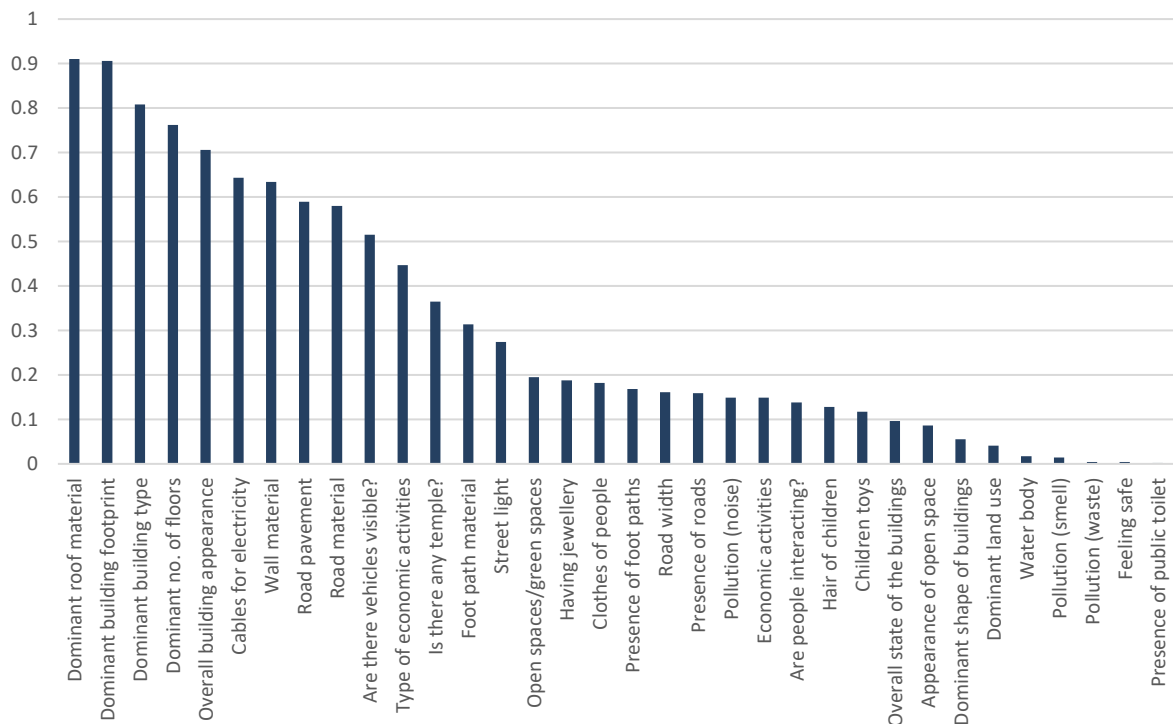


Figure 30: Squared correlation of indicators with MCA dimension 1

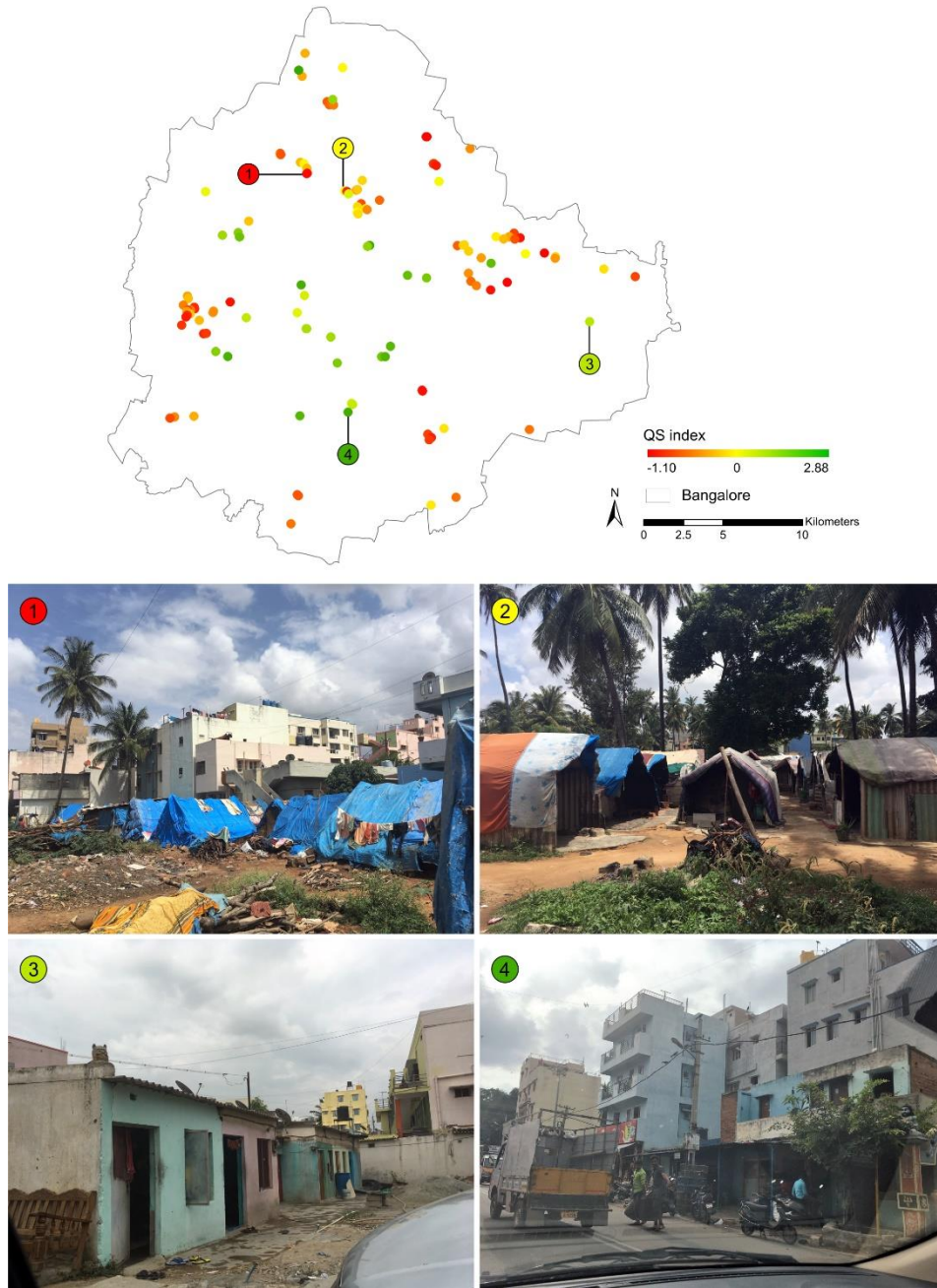


Figure 31: QS slums on map and some examples

5.1.3. Relationship between HH and QS

For analysing the relationship between QS and HH data, the 26 slums were explored. A Pearson correlation using bootstrap was performed between all dimensions derived from HH and QS data to see whether they are correlated (Annex 8). Samples were bootstrapped 1000 times to get confidence intervals. Dimension 1 from HH and QS significantly correlate. This means they were describing the same things about these samples. It means both are describing deprivation (they are correlated) but from different perspectives (one cannot explain variations of the other well, as R is 0.63 [0.28, 0.82]²²²³). We should also consider the temporal gap between data of HH and QS. Although the 26 samples were checked to ensure

²² Numbers in [] show 95% confidence intervals calculated using bootstrapping samples. As we had few samples, the confidence intervals help to find whether the correlation between HH and QS is confidently positive. The confidence intervals also show the range of the correlation coefficient with 95% of confidence.

²³ As R is 0.63 [0.28, 0.82], R^2 is 0.40 [0.08, 0.67], so the two indices cannot explain each other well.

they have not significantly changed since 2010, they were only checked using Google Earth and there is a risk that some of them experienced changes which are not visible from above.

MCA selected physical indicators of HH data as the most important contributors in building index and QS also focuses on the physical and contextual domains. Therefore, the leading indicators which make distinctions among slums are visible, and we can expect relationships between deprivation and satellite images as they can also see physical and contextual characteristics (see the conceptual framework, section 2.6).

5.1.4. Discussion on classical indexing

In this section, we built an index using a classical approach²⁴ for the purpose to compare its result with what we have gotten from MCA. First, we explored what would be the outcome if we use HH categories as ordinal data, then average each indicator similar to Rains et al. (2017) over slums.

Figure 33 shows the distribution of two indicators which are well explained and two indicators which are not explained well by HH index (green and red ones)²⁵. Well-explained indicators have distinct values among slums. However, living area per capita, has almost the same mean value across all slums. Similar, distance to healthcare has a mean value between 1 and 2 for almost all the slums with relatively large amount of variations. Comparing these indicators with what we got from MCA (blue plot), it maximized differences between slums by creating dimensions. Furthermore, Using the most important dimension as the deprivation index, the internal diversity of deprivation was minimized (see standard deviations in Figure 33). This show that although the internal diversity of each indicator is different, the deprivation levels among households within a slum is almost the same for nearly all the samples. One problem of averaging ordinal indicators is that we ignore variations within slums among households. Another approach to aggregated data into slums could be performing ordinary PCA (or FA), but these two methods are only applicable to ordinal data which is assumed to be continuous (Field, 2013).

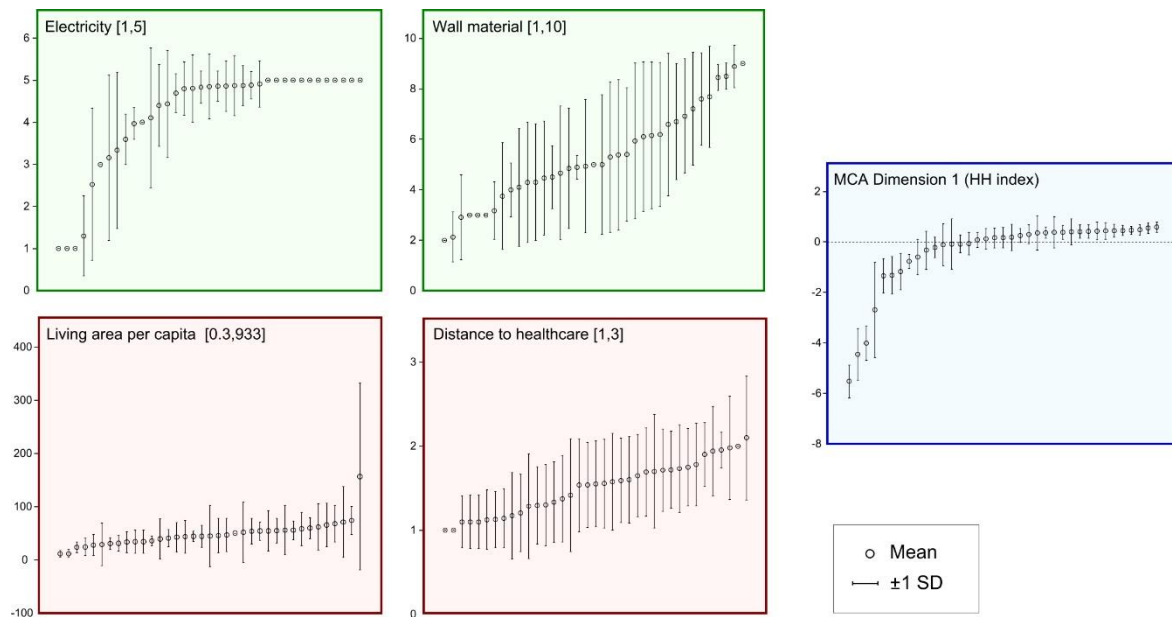


Figure 32: Comparing distribution of values of some indicators with MCA dimension 1 along HH slum

²⁴ By the classical approach, we mean using indicators as ratio or ordinal data types (i.e., numeric values) (and not nominal data) and summing them up with equal weights (or weights coming from other analysis). We called it classical as it is a common way to build an index in social studies (e.g., Noble, Wright, Smith, & Dibben, 2006).

²⁵ Indicators with few missing values were selected.

To test how a classical index works, we aggregated data into slums using the mode value, which seems logical (note that we still ignored intra-variations and assumed the equal difference between categories). All indicators were normalized between 0 and 1, aggregated into capitals with the same weights and finally aggregated into deprivation index again with same weights (we used the same weights as there was no evidence of why one capital or one indicator is more important than the others).

Figure 34 illustrates the result of the classical index with some examples. A first difference between the classical index and MCA is that with the classical index we restricted our data between a minimum and a maximum value, but in MCA we force the common situation to be zero and leave all other to negative or positive sides without limitation. In practice, in MCA we finally know min, max and average, but in the classical approach, we only know min and max. As the MCA analyses categories inside indicators and calculate relative differences between them, it can also indicate differences between slums more accurately than the classical index. Considering Figure 34, samples got very close deprivation values, and it does not indicate relative differences between samples in terms of deprivation (compare Figure 34 with Figure 27). Thus, the result of the classical index is less accurate than the MCA (in our case at least). We had some worst-off slums among the HH samples which were significantly different from the average situation of Bangalore (see Figure 27), but the classical index was unable to identify them.

It could be also an option to predict indicators one by one in a logistic regression; however, the number of samples needed to perform a logistic regression is a complicated topic. Based on Peduzzi, Concato, Kemper, Holford, and Feinstein (1996), the minimum number of samples needed to predict an indicator with k categories using a logistic regression model, is $10k/p$. Here, p is the proportion of individuals belonging to the category with the minimum frequency in an indicator²⁶. Therefore, having an indicator with only two categories, if the two categories have equal frequencies, the minimum number of samples needed is 40. In case of QS which we had 121 samples, even in indicators with few categories, the frequency of categories is different. As an example, to predict an indicator with three categories and the proportional frequencies of 20%, 30% and 50%, we need 150 samples. To this end, it is possible to predict some QS indicators with logistic regression models, and it could be a direction for further studies, but the prediction will not result in understanding deprivation levels. In this study, we focused on predicting deprivation and not individual indicators. However, in case of an interest in this direction for further studies, we recommend predicting indicators with more contribution to creating indices as they have more variations among slums.



Figure 33: Classical index result with some examples

²⁶ There many studies discussed the number of samples needed for a logistic regression model. We introduced one of the most common of such studies. The point is that for further studies, the issue of the number of samples should be considered with caution.

5.2. CNN performance

As explained in chapter 4, We trained CNNs with the ability to distinguish slums from formal areas. This helped to extract distinctive features to be used in a regression model with few samples to predict deprivation indices. We trained CNNs from simple networks to more deep ones. The first step was to tune CNN with different patch size to find the optimal one. We tested patches with 99, 129, and 165 pixels and for each, we tuned CNN parameters to find the optimal result. Figure 35 shows the result of training the CNNs with different patch size and up to 700 epochs. In this stage, we used 1000 training, 1000 validation and 400 test patches^{27 28}. Patch size of 129 had the best performance in classifying slums from formal areas.

We used patch size of 129 for the further analysis and used 2000 training, 2000 validation and 2000 test patches to train CNN again. To have an idea of how the objective function behaves, Annex 9 provides the created curves after training tuned network using patch size of 129 with 600 epochs. The final accuracy we got on the test set with this simple network was 92.50%. This means 92.50% of 2000 test patches were correctly classified either as a slum or a formal area.

Next step was to create a VGG-like network as explained in section 4.4.5 to see whether it can improve our result. Our network converged after 100 epochs, and we got the accuracy of 96.05% on the test set, which is a significant improvement over the simple CNN. Thus, using such deep networks could heavily boost the image classification result. It also reduced the time to tune (many) CNN parameters and made the analysis more efficient. However, it requires high computational power and learns slower as adding more layers to CNNs and make them deeper increase the number of parameters to learn.

The final network we trained for the classification was the VGG-like network, but we increased the training sample size from 2000 to 16000 taking advantage of image augmentation. This time, our network converged after only 50 epochs, and the accuracy on test data reached 98.4%. Again, we got a considerable improvement over our previous network. This confirms Chatfield et al. (2014), stressing that image augmentation can boost the result of CNN. Figure 36 compares accuracies obtained with the different networks. For the further steps of the analysis, we used our best performing network (i.e., VGG-like CNN with image augmentation).

To have an overview of classified patches, Figure 37 shows some slum patches from the test set. All these patches are slums, but some were incorrectly classified as formal. The percentages bellow patches show

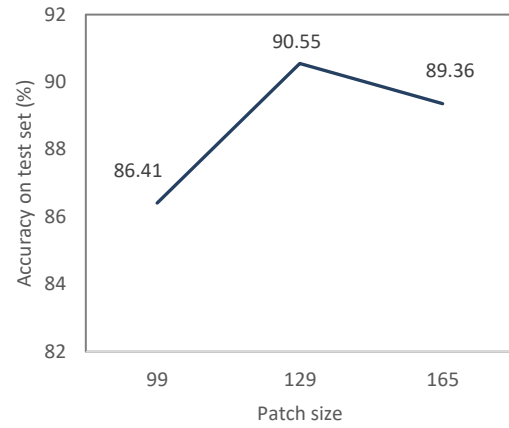


Figure 34: Effect of varying patch size

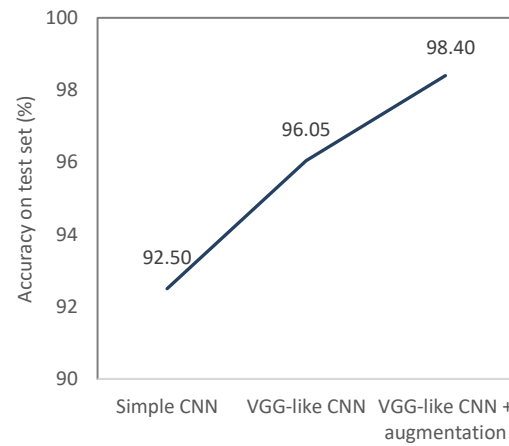


Figure 35: Results obtained by different CNNs

²⁷ According to Congalton (1991), having 75 to 100 test polygons per class can give us a good insight of the classification accuracy. We used 400 test samples to tune patch size as an enough number of samples. However, for the main analyses, we increased the number of test samples equal to the number of training and validation sets to have more robust accuracy assessments.

²⁸ In fact, we conducted the main analyses with 2000 training, 2000 validation, and 2000 test samples. However, to choose the optimal patch size, as we needed to tune and train three different networks (for three different patch size), we decreased number of samples for more efficiency.

the confidence of the network in classifying these patches as slum (derived from the softmax layer). Scores less than 50% result in classifying patches as formal areas. Patches like number 1 were clearly classified as slum. They have very distinct characteristics with small dwellings and irregular patterns, easily distinguishable from formal areas. Slums like number 2 with some regular patterns were classified as slum with less confidence. Patch 3 is a very challenging patch. Patches like this contain small slums between formal areas. Although it is not even easy to identify the slum in between formal areas in this patch, it was also correctly classified. Patch 4 and 5 have almost the same situation but dwellings in patch 5 are tiny and we cannot even confidently recognize them by eyes. Patches like number 6 completely confused the network as they have larger dwellings with some regular patterns. Overall, only 1.92% of slums were classified incorrectly (Figure 37).



Figure 36: Per class accuracy with some examples of classified patches

5.3. Connecting image-based features to deprivation indices

The final section of this chapter elaborates the result of performing regressions to connect image features with derived deprivation indices. First, we will look at the result of using the CNN softmax layer as the predictor of deprivation indices in section 5.3.1. Then, section 5.3.2 describes the results of performing regression with the pre-trained CNN. After that, section 5.3.4 explains the performance of hand-crafted and GIS features as predictors. Finally, we combine results and discuss the capability of models to be generalized.

5.3.1. Regression using CNN softmax layer

Starting from the initial idea of using a CNN, we tried to connect the CNN softmax layer to indices. We used the score of the slum class in the softmax layer to predict values of the QS and HH index. Using 10-fold cross validation for QS and LOO for HH, we got R^2 of -0.01 and -0.08 when predicting QS and HH indices respectively as the best results. To explore why we got these discouraging results, values of the QS index were plotted over CNN softmax scores (Figure 38).

This figure illustrates why CNN softmax values cannot predict the QS index. As our CNN performs well in distinguishing slum from formal areas, most of the points fall precisely on 0 or 1, means the network classifies them with 100% confidence into either slum or formal. Therefore, this situation is not very different from trying to predict a continuous variable with a binary predictor.

Another interesting observation considering Figure 38, is that high QS index values show slums that are similar to formal areas. Thus, most of the misclassified points (have CNN softmax value less than 0.5) have values more than zero. This also confirms what the MCA tells about slums. Slums with an index value more than zero have more variations and have better conditions, which makes them close to formal areas.

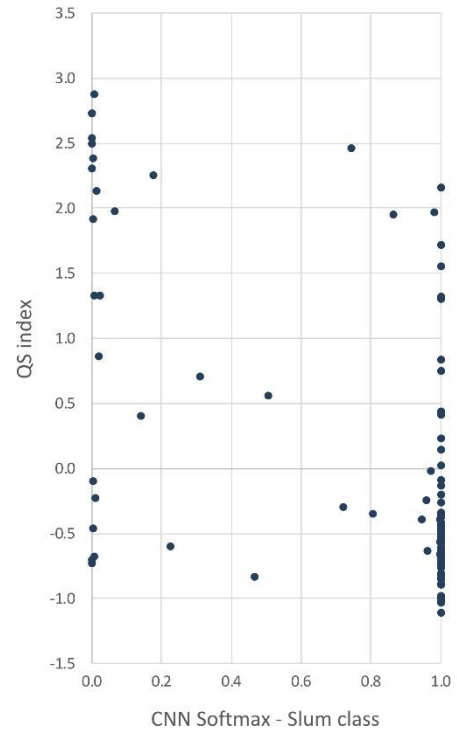


Figure 37: Relation between CNN softmax and QS index

5.3.2. Transfer learning – CNN as a regression model

We developed a fully CNN-based solution to the problem of predicting deprivation with image-based features. The pre-trained CNN (to distinguish slum from formal areas) with its learned distinctive features was fine-tuned to directly predict index values. We changed the objective function from log-likelihood to Euclidean loss. We also changed the number of output neurons from two neurons to one (i.e., index value). Each network was trained for 100 epochs to ensure convergence. Figure 39 shows the result of predicting HH and QS indices directly with the CNN. It shows that the CNN could predict the QS index with R^2 of 0.68, but in terms of the HH index, we got negative R^2 , which shows the model cannot explain the variation of the HH index at all²⁹, although the objective function reached almost zero during training. It means, using very few samples (in our case 26), the network overfitted in the training data and using such few cases, we cannot cover the whole variety of samples. Moreover, the figure shows that image augmentation could not improve the result significantly, so we left it out for our further steps.



Figure 38: Predicting indices with CNN

²⁹ The meaning of the negative R^2 value is described in section 5.3.3.

5.3.3. Regression with hand-crafted and GIS features

In this section, we discuss the result of performing stepwise PCRs to predict deprivation indices. Before performing the regression models, we examined the number of grey levels in GLCM features to find the optimal number of levels. Figure 40 shows the result of this exploration. We chose 8 grey levels as it created the highest correlation with QS index. Back to regression results, as explained in section 4.6.3, we varied the number of component (1 up to 12 components) and the complexity of the model (linear, linear + interaction, quadratic + interaction) to find the optimal result. Figure 41 summarizes the results of these regressions and shows their performance by plotting R^2 . Any R^2 less than -1 was plotted as -1 for better visualization. In general, a

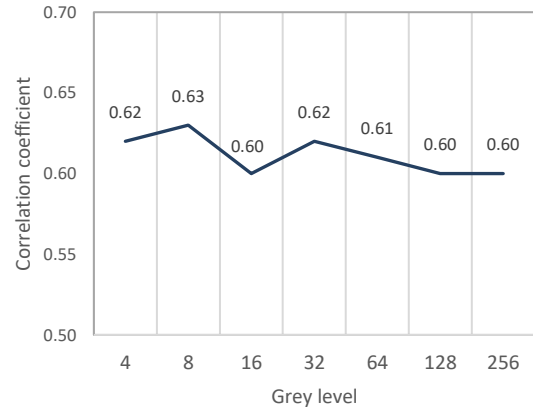


Figure 39: Result of varying number of grey level in GLCM

negative R^2 value means the model cannot explain the variations of the dependent variable at all. In a regression model, the model $\hat{y} = \bar{y}$, where \bar{y} is the mean of y values, has the R^2 of 0 (see Equation 6). In a simple linear regression without using cross validation methods, it is impossible to build a model worse than the mean model because in a linear regression, the model is initialized equal to the mean model, then is improved to reach the optimal result. Accordingly, the R^2 is always a value between 0 and 1, and it is equal to the squared correlation coefficient R between the dependent variable and the predictors (Field, 2013). However, in non-linear models or using cross validation methods, we can also measure whether the model is overfitted. A model with a great amount of overfitting cannot explain the test data and, it is likely to perform worse than the mean model, thus results in a negative R^2 .

Overall, we got a better performance of regression models on the QS data. Using all hand-crafted + GIS features, 11 components, and a linear model, we obtained R^2 of 0.52. Comparing this result with what we have gotten by CNN, we can see CNN with R^2 of 0.67 outperformed our complex PCR models. For a better comparison, we should compare CNN with the result of regression using only hand-crafted features as CNN also does not consider where the patches are located. In this manner, we have R^2 of 0.38 from hand-crafted features using a quadratic model in comparison with R^2 of 0.67 from CNN. This shows the advantage of the CNN over hand-crafted features and the PCR model. Indeed, more sophisticated hand-crafted features as extracted in Kuffer et al. (2017), Duque et al. (2017), and many other studies reviewed in chapter 2 might improve the result.

In case of the HH index, R^2 values are relatively lower. Using only hand-crafted features was disappointing, and we could not obtain any positive R^2 (same result as CNN). However, using GIS layers, three components and a linear model, we obtained R^2 of 0.42. The results show that relying only on patches (and not considering spatial configurations) did not help to predict the HH index values. Thus, adding the GIS layers showed that considering these spatial dimensions to the model can boost the results. This also holds true in case of the QS index, where GIS layers improved R^2 from 0.38 to 0.52. Recall that we used rather simple GIS layers created with Euclidean distance function. More sophisticated GIS layers, considering the impact of all land uses on each patch, using network analysis, creating density maps of land uses, and many other spatial analyses might improve the model further.

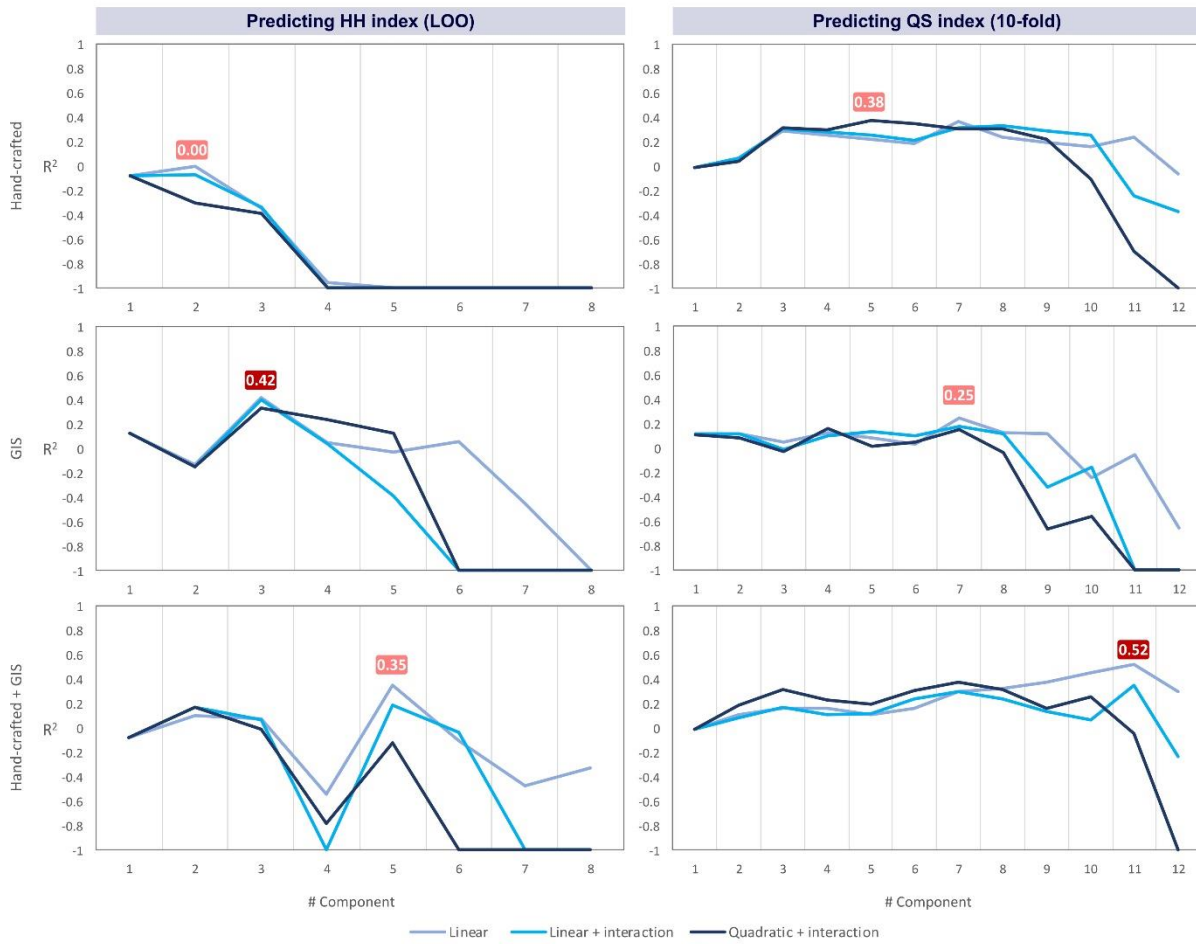


Figure 40: Results of performing PCR to predict HH and QS indices

5.3.4. Combining PCR with CNN features

In the next step, we used results from predicting the QS index and combined them to build a new model with a better ability to predict the QS index. The aim was to test whether there is any improvement on CNN result by using hand-crafted and GIS features. We did not combine results obtained for HH index as only the GIS layers performed well to predict this index. Figure 42 shows the results of combining models: hand-crafted + GIS + CNN; hand-crafted + CNN; and GIS + CNN. One interesting result is that using the combination of hand-crafted + CNN, R^2 dropped for 0.01. This means the hand-crafted features could not contribute to improving what we already had from CNN. In fact, they brought also some confusion to the model. However, GIS layers could improve the CNN results for about 0.04. This means by adding the GIS layers, we included what CNN had not covered before. The best result we obtained was 0.75 by using hand-crafted + GIS + CNN in a third-degree polynomial model, allowing interaction between variables. This shows although

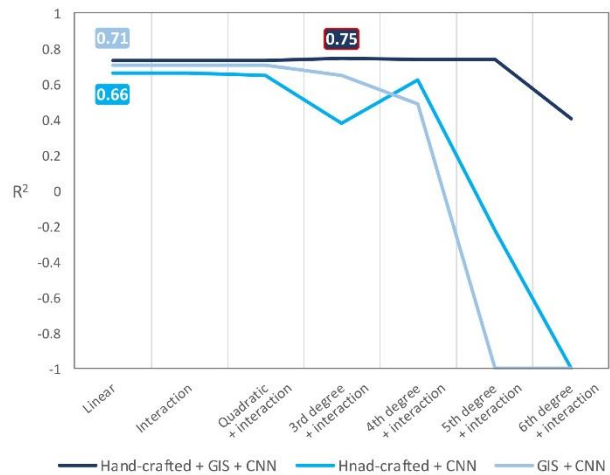


Figure 41: results of combining different features to predict QS index

hand-crafted features could not improve the result of CNN, their interactions with GIS features could bring new improvements to the model. We conclude that using GIS layers in parallel with CNN can bring improvements to the model as basically CNN do not consider the spatial location of patches. Hand-crafted features cannot improve CNN results by themselves, so if we use them in parallel with CNN, there is hardly any improvement. One interesting option to explore could be to create GIS features and add them to the original patches as new image channels. Suppose instead of having a 4-channel image, we use more channels as input to the CNN and let it do the regression using all of them. Although this is recommended for further research, it needs high computational power as adding new channels increases CNN parameters significantly.

To explore our final model (with R^2 of 0.75), we visualized the eight worst-off and the eight best-off slums predicted by the model in Figure 43 and Figure 44. As variation in the negative side of the QS index is low, the model errors did not make illogic predictions at least by looking at the photos in Figure 43. Worst-off slums are mostly tiny with few dwellings and are not easily distinguishable from satellite images. On the positive side, although slums were mostly predicted well, there are still some confusions in the model (see slum number 2 and 3 in Figure 44). Comparing number 2 and 4 in Figure 44, patches are very similar, and indicators that make number 2 worse than number 4 in the index might not be visible from above. It could be also an error coming from the fieldwork as in Quick Scan fieldwork the surveyor was supposed to stand on a point and report what is visible. There is a risk that this point is different from the typical structure of that slum, especially in huge settlements. The other source of error could be using a fixed square patch to extract features from all the slums. Some settlements are enormous, and a patch can cover a small area of them; however, some settlements are tiny, and even if we consider their context (i.e., 20-meter buffer), a patch is bigger than the area of the analysis. Thus, we ignored all the area of the enormous settlements except a patch located at the centre of them, so there is a risk that the patch does not represent the whole area of the settlement. Overall, these kinds of confusions were expected, but we can rely on the model with 75% confidence in its predictions.



Figure 42: Worst-off slums predicted by the model with respective patches

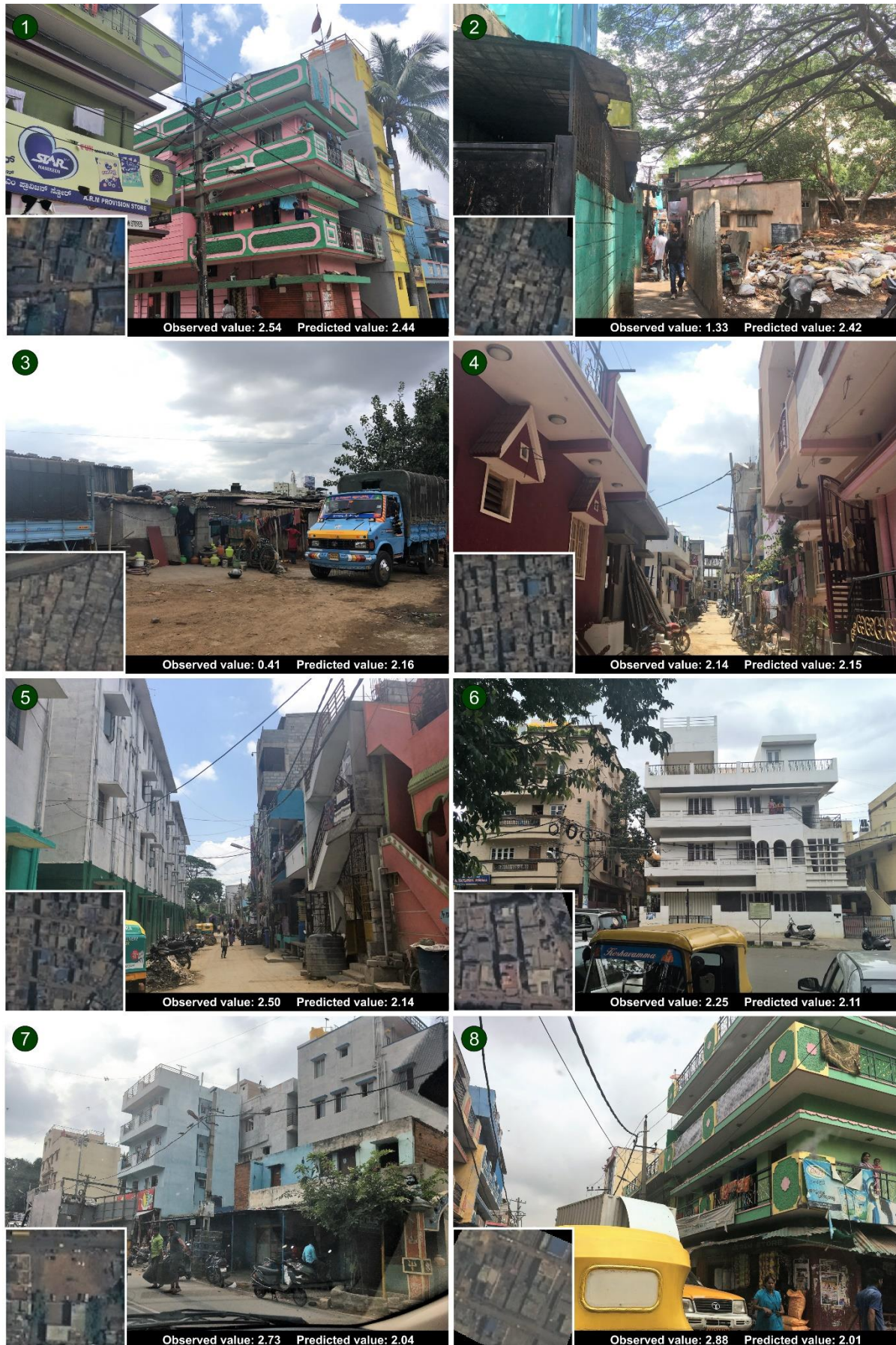


Figure 43: Best-off slums predicted by the model with respective patches

5.3.5. Ability to generalize the results

Although all the models are validated on our samples, it is worth discovering the ability of our models to be generalized. We focus on the model created by CNN with R^2 of 0.67 (see section 5.3.2), the model created with hand-crafted and GIS features with R^2 of 0.52 (see section 5.3.3), and the model created by combining CNN, handcrafted, and GIS features with R^2 of 0.75 (see section 5.3.4). One crucial assumption to consider when we want to generalize a model is homoscedasticity, i.e., “residual at each level if the predictor(s) should have the same variance” (Field, 2013, p. 311). Figure 45 plots standardized residual over predicted values of the three models. In an acceptable plot, we should see scattered point without any systematic pattern; however, our plots violated this assumption. By looking closer at the points, we can find different patterns among negative and positive predicted values. Considering the regression model with CNN, residuals at the negative side are biased in a way that predicted values close to zero have more variance. Values on the positive side are more scattered. In the PCR model, values on the negative side are strongly biased showing the low capability of the model to be generalized. On the positive side, the variance of the residuals is more acceptable. The final model created by combining the CNN model and the PCR model has the least biased residuals across predictors. The plot is similar to the CNN plot but on the negative and positive sides points are more scattered. Nevertheless, in all plots residuals have different patterns on the negative and positive sides. Comparing Figure 45 with Figure 30, less difference (more homogeneity) resulted in less variance in the residuals. The worse-off slums are more similar to each other, so the predictions also had less amount of errors. On the positive side, better-off slums are very different from each other. Comparing picture 2 and 4 in Figure 32, there is a wide difference between the common situation (i.e., value 0) and the best-off slum. Therefore, residuals had more variance, and the predictions were less accurate. Considering picture 2 and 3 in Figure 44, the model had large errors which resulted in predicting very different settlements from the best-off slums (e.g., picture 1 in Figure 44).

Although the predictions of the PCR model were biased, by combining its result with the CNN model, a less biased model was created. Therefore, we can consider combining hand-crafted and GIS features (especially GIS features, as discussed in 5.3.4) even if the aim of the model is to be generalized on a broader population. By dividing slums into two worse-off and better-off groups (i.e., slums with a negative value and slums with a positive value), it is more likely to create models with less biased predictions. This could be a direction for further studies, although it needs more samples, especially with positive index values.

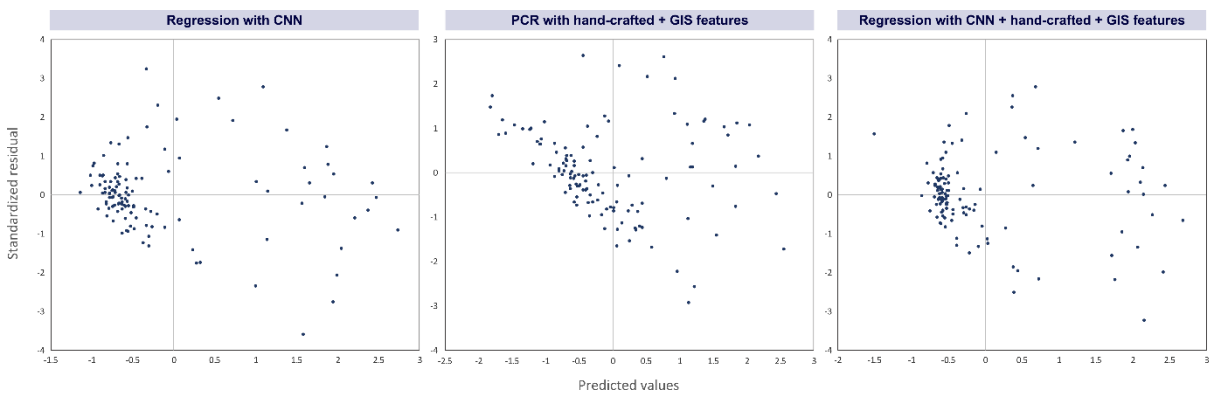


Figure 44: Predicted values over standardized residual of the created models

5.4. Reflection on findings

The section provides a summary of the study results and compares them with recent relevant literature in the field of slum mapping and characterisation. It starts with the way we constructed the deprivation indices. Then, reviews our findings related to classifying slums from formal areas. After that, it explores related literature in the field of connecting socio-economic variations to satellite images and compares them to the results of this study.

We used the MCA to build deprivation indices as a data-driven approach with the least assumptions. Categorical indicators were used without manipulation, and index values were assigned to individuals based on the patterns of categories. We found that relying only on data, without pre-assumptions like ordering categories and assigning pre-defined weights, it is possible to distinguish the better-off, the worse-off and the average situation slum settlements with their relative differences and distinctive categories. We also verified the values of the deprivation indices by photos from the settlements and found a logical relationship between deprivation index values and slum situations in Bangalore. Rains et al. (2017), aggregated categorical data to ordinal and aggregated them using descriptive statistics. This brings many assumptions to the index and may result in a biased result. Saharan et al. (2017) aggregated indicators of a deprivation index with equal weights. Based on the data-driven approach we conducted, the importance of deprivation dimensions is not the same. We showed that the indicators related to the physical capital plays the most crucial role in distinguishing slum types. Instead of assuming the most important indicators to understand the deprivation, we let the data decide on the most important categories and indicators based on our samples. Indeed, our approach is meaningful when there is a set of representative samples is accessible. Non-representative samples, as the HH index samples, results in less meaningful index values³⁰.

We emphasised on both of index creation and image analysis equally in this study. The logic was if we do not develop a meaningful index, even if we develop a model with a very high prediction power, it is not predicting something valuable and meaningful. Therefore, we developed two indices, one covering all the aspects of deprivation, and the other focusing on the physical and the contextual domains, and we built models to predict both. We showed that these two indices are correlated, and both represent the deprivation levels among slums. Engstrom et al. (2017), developed a model to predict urban poverty using image features. They used consumption rate as an indicator of urban poverty that is only the financial aspect of deprivation. Jean et al. (2016) also focused on consumption rate and wealth as indicators to predict poverty which only cover the financial domain of deprivation. Arribas-Bel et al. (2017) focused on the living environment as one of the seven deprivation domains of English index of deprivation (Ministry of Housing Communities & Local Government, 2015) that covers the contextual domain of deprivation only. Duque et al. (2015) built a slum index with four indicators which only covers the physical and financial domains of deprivation. All mentioned studies predicted poverty, deprivation, or slum index for analytical regions (either administrative boundaries or created regions). None of these studies especially focused on slums or informal settlements, and they aggregated and predicted data related to both formal and informal settlements together. We showed the ability of the satellite images to predict deprivation levels even for tiny slum settlements with few dwellings. Our method can capture deprivation of any type of slum in Bangalore regardless of its size.

We developed a fully CNN-based approach to predict deprivation index values using few samples. We trained a CNN with the ability to distinguish slums from formal areas which predicted the test set with 98.4% of the overall accuracy. The CNN was developed to learn distinctive features; then the network was fine-tuned to behave as a regression model and predict the deprivation indices. The accuracy we obtained in the classification task should not be compared to studies with the aim of slum mapping (e.g., Persello & Stein, 2017). We developed a method with the ability to assign slum or formal area to patches³¹. This is

³⁰ As we saw in case of the HH index, that the average situation was deviated through the better-off slums (see section 5.1.1.1).

³¹ This is more similar to studies in the field of computer vision like He et al. (2015).

different from pixel-wise classification studies. We also created buffers around the samples which somehow guided the network to learn more relevant features. We showed that a well-trained CNN can be fine-tuned to predict deprivation levels of slum settlements with the R^2 of 0.67 using few samples.

In parallel with CNN, we developed regression models based on hand-crafted and freely available GIS features to predict deprivation indices. We developed PCRs with different levels of complexity from pure linear models to Quadratic models allowing multiplication of predictors. Studies which predicted deprivation or poverty using hand-crafted features did not include GIS layers, and they only relied on linear regression models (see Arribas-Bel et al., 2017; Duque et al., 2015; Engstrom et al., 2017). We showed that by adding GIS layer to hand-crafted features, the result can be boosted significantly. Furthermore, studies which used CNNs for land cover/use classification usually do not include such spatial information extractable from the GIS layers (see Bergado et al., 2016; Mboga, 2017; Scott et al., 2017). We combined results of the PCR and the CNN and showed that the GIS layers can even improve results obtained from the CNN. Although CNNs learn abstract sophisticated features from the training samples they do not consider the location of the patches in a city, so the ability of the GIS layers to improve the CNN result was expected. It is worth exploring the ability of other ML methods to predict deprivation indices combining the CNN and the GIS features. Arribas-Bel et al. (2017) showed GBR and RF can outperform a linear regression model. To this end, we found a more reliable result from the CNN than PCR with hand-crafted features as residuals were less biased and the model was more capable for generalization purposes.

6. CONCLUSION AND RECOMMENDATIONS

The aim of this study was to analyse the relationship between slums' variations from the perspective of deprivation with image-based features. Many studies have been conducted to detect where the slums are located, but this study tried to see the extent we can understand deprivation level of such areas from above. The study was conducted in two main steps; understanding deprivation in slum areas; and trying to explain variations of deprivation using satellite images.

To understand deprivation, relevant indicators were extracted from available data based on IMD. To build deprivation indices, this study used a data-driven approach. MCA was the primary method to create deprivation indices and interpreting variations exist in slums. We found that MCA as a principal component method could explain variations of deprivation indicators. After verifying values created by the first dimension of MCA with photos taken from the fieldwork, we confirmed that it is a good measure to compute relative deprivation among slums with few assumptions. Using MCA, we could find the common level of deprivation as well as relative difference each settlement has from the common level either on the negative or positive sides. Considering the contribution of indicators to create deprivation indices, we found that indicators related to the physical capital have the most power in distinguishing slum settlements.

CNN was the main approach to analyse satellite images and extracting relevant features. As we had few samples with deprivation index values, we initially trained our CNN to distinguish slums from formal settlements. To do this, we trained a network inspired from VGG with five convolutional layers and two fully connected layers. By such network and taking advantage of image augmentation, we obtained the accuracy of 98.4% on our test set. Therefore, we trained a network which learned distinctive features related to slum and formal areas, and we used it in further steps.

Two main steps were followed in order to create regression models and predict deprivation index values. First, we used a CNN-based approach and fine-tuned our trained CNN to directly predict deprivation indices. Using this method, we could train our network in few epochs and with few samples (i.e., 121 samples) and predict deprivation with R^2 of 0.67. We found that using fewer samples (i.e., 26 samples), the network overfits in the training data. Further studies may add more regularisation terms to mitigate this problem. As a supplementary step, powerful texture features, spectral features, and GIS features were also extracted to investigate their capability to improve results obtained from CNN. Therefore, the second step in performing regression was to use PCR, altering the number of components, to build regression models with hand-crafted and GIS features. Using 121 samples and only hand-crafted features R^2 of 0.38 was the best result obtained. By adding GIS layers, R^2 increased to 0.52. We found that using only 26 samples, it was not possible to predict deprivation using hand-crafted features (same as CNN). However, GIS features could predict deprivation with R^2 of 0.42. This showed the necessity of including spatial characteristics of the settlements to our models and looking beyond the delineated boundaries or the extracted patches. By exploring the capability of built models for generalization, we found that CNN had less biased errors and more capability for generalization.

Finally, we combined the results obtained from CNN, hand-crafted, and GIS features to explore any improvement. We found that adding the results from hand-crafted features to CNN could not improve our model. In fact, hand-crafted features brought confusion into what we had already obtained from CNN and dropped R^2 value. Though, adding the results from the GIS layers to CNN resulted in some improvements to the model. This means GIS layers could bring information which are not extractable from single patches. The best R^2 obtained by using the CNN, hand-crafted, and GIS features using 121 samples, was 0.75. Exploring the predicted values by the model, we could see that the model can

distinguish slums with different deprivation levels. There are also directions which could be followed by further studies. The next section recommends some of the possibilities.

6.1. Recommendations for further studies

The section lists possible directions which can be followed for further studies:

- We used the first dimensions created by MCA and distributed slums along two sides (on a 1-dimensional line). This means assigning one value to each settlement. Further studies could explore the possibility to add the second (or even the third) dimension to the analysis. By considering the second dimension, slums are distributed along four sides (on a 2-dimensional plain) and instead of one index value, each slum will get two index values. Thus, it will be possible to find more slum types and provide a better insight of deprivation levels. As an example, studies might find that each side represents deprivation regarding specific capitals.
- Instead of having more than one value for each sample, another approach of using more dimensions could be to use the values of one dimension for some samples and use the values of another dimension for other samples. As an example, we can choose one dimension for better-off slums and one dimension for worse-off slums. We showed HH households along dimension 1 in Figure 26. However, considering Figure 25, it seems that better-off slums (with value more than 0 in dimension 1), have more variations along dimension two. It means indicators which have more contribution to dimension 2 can distinguish better-off slums well. This could be an idea for further studies to work with more than one dimension, but the extent it can help to understand variations of deprivation among individuals should be explored.
- This study showed the capability of GIS layers to improve the result obtained from the CNN and the regression model with hand-crafted features. It could be a direction to develop methods to combine GIS layers as a new channel (or new channels) of input patches to a CNN.
- CNN used a fixed patch size for all samples so by creating one patch from each sample most of the area of such samples are dropped from the analysis. Further studies could develop methods to generate more than one patch from such settlements and calculate the deprivation in their different locations, then combine the obtained results for a settlement.
- We only explored the variations among slums, but there is also a far-reaching range of literature which developed methods to detect where the slums are. Further studies could develop models that can detect slums and predict their deprivation levels simultaneously. The model could be an end-to-end CNN-based model with the ability to predict slum locations (e.g., Persello & Stein 2017), then assign their deprivation levels. It is also a possibility to develop methods using hand-crafted features (e.g., Kuffer, et al., 2016) and other ML classifiers (e.g., SVM), then apply regression models to predict deprivation indices.

LIST OF REFERENCES

- Almeida, R. M. V. R., Infantis, A. F. C., Suassuna, J. H. R., & Costa, J. C. G. D. (2009). Multiple Correspondence Analysis. *Computer Methods and Programs in Biomedicine*, 95(2), 116–128.
<https://doi.org/10.1016/j.cmpb.2009.02.003>
- Arimah, B. C. (2010). The Face of Urban Poverty: Explaining the Prevalence of Slums in Developing Countries. In *Urbanization and Development* (pp. 143–164). Oxford: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199590148.003.0008>
- Arribas-Bel, D., Patino, J. E., & Duque, J. C. (2017). Remote sensing-based measurement of Living Environment Deprivation: Improving classical approaches with machine learning. *PLoS ONE*, 12(5), 1–25.
<https://doi.org/10.1371/journal.pone.0176684>
- Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of Clinical Epidemiology*, 68(6), 627–636. <https://doi.org/10.1016/j.jclinepi.2014.12.014>
- Baud, I., Kuffer, M., Pfeffer, K., Sliuzas, R., & Karuppannan, S. (2010). Understanding heterogeneity in metropolitan india: The added value of remote sensing data for analyzing sub-standard residential areas. *International Journal of Applied Earth Observation and Geoinformation*, 12(5), 359–374. <https://doi.org/10.1016/j.jag.2010.04.008>
- Baud, I., Sridharan, N., & Pfeffer, K. (2008). Mapping Urban Poverty for Local Governance in an Indian Mega-City: The Case of Delhi. *Urban Studies*, 45(7), 1385–1412. <https://doi.org/10.1177/0042098008090679>
- Bell, N., Schuurman, N., Oliver, L., & Hayes, M. V. (2007). Towards the construction of place-specific measures of deprivation: A case study from the Vancouver metropolitan area. *Canadian Geographer*, 51(4), 444–461.
<https://doi.org/10.1111/j.1541-0064.2007.00191.x>
- Bergado, J. R. A., Persello, C., & Gevaert, C. (2016). A Deep Learning Approach to the Classification of Sub-Decimetre Resolution Aerial Images. *2016 IEEE Geoscience and Remote Sensing Symposium (IGARSS)*, 1516–1519.
<https://doi.org/10.1109/IGARSS.2016.7729387>
- Beteille, A. (2003). Poverty and Inequality. *Economic and Political Weekly*, 38(42), 4455–4463.
<https://doi.org/10.2307/4414161>
- Bima, L., Nurbani, R., Diningrat, R., Marlina, C., Hermanus, E., & Lubis, S. (2017). *Urban Child Poverty and Disparity : The Unheard Voices of Children living in Poverty in Indonesia*. Jakarta. Retrieved from
<http://www.smeru.or.id/sites/default/files/publication/ucpd2017.pdf>
- Bottou, L. (2010). Large-Scale Machine Learning with Stochastic Gradient Descent. In *Proceedings of COMSTAT'2010* (pp. 177–186). Heidelberg: Physica-Verlag HD. https://doi.org/10.1007/978-3-7908-2604-3_16
- Cabrera-Barona, P., Murphy, T., Kienberger, S., & Blaschke, T. (2015). A multi-criteria spatial deprivation index to support health inequality analyses. *International Journal of Health Geographics*, 14, 11.
<https://doi.org/10.1186/s12942-015-0004-x>
- Cabrera-Barona, P., Wei, C., & Hagenlocher, M. (2016). Multiscale evaluation of an urban deprivation index: Implications for quality of life and healthcare accessibility planning. *Applied Geography*, 70, 1–10.
<https://doi.org/10.1016/j.apgeog.2016.02.009>
- Camps-Va, G., Mooij, J., & Scholkopf, B. (2010). Remote Sensing Feature Selection by Kernel Dependence Measures. *IEEE Geoscience and Remote Sensing Letters*, 7(3), 587–591.
<https://doi.org/10.1109/LGRS.2010.2041896>
- Castelluccio, M., Poggi, G., Sansone, C., & Verdoliva, L. (2017). Training convolutional neural networks for semantic classification of remote sensing imagery. *2017 Joint Urban Remote Sensing Event (JURSE)*, (Section 2), 1–4.
<https://doi.org/10.1109/JURSE.2017.7924535>
- Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the Devil in the Details: Delving Deep into Convolutional Nets. *arXiv Preprint arXiv:1405.3531*, 1–12. <https://doi.org/10.5244/C.28.6>
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)

- Coulangeon, P. (2017). Cultural Openness as an Emerging Form of Cultural Capital in Contemporary France. *Cultural Sociology*, 11(2), 145–164. <https://doi.org/10.1177/1749975516680518>
- Coulangeon, P., & Lemel, Y. (2007). Is “distinction” really outdated? Questioning the meaning of the omnivorization of musical taste in contemporary France. *Poetics*, 35(2–3), 93–111. <https://doi.org/10.1016/j.poetic.2007.03.006>
- Deolalikar, A. B. (2005). *Attaining the millennium development goals in India : reducing infant mortality, child malnutrition, gender disparities and hunger-poverty and increasing school enrolment and completion*. Oxford University Press.
- 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), 1–23. <https://doi.org/10.3390/rs9090895>
- Duque, J. C., Patino, J. E., Ruiz, L. A., & Pardo-Pascual, J. E. (2015). Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. *Landscape and Urban Planning*, 135, 11–21. <https://doi.org/10.1016/j.landurbplan.2014.11.009>
- Duque, J. C., Royuela, V., & Noreña, M. (2013). A Stepwise Procedure to Determinate a Suitable Scale for the Spatial Delimitation of Urban Slums. In E. Fernández Vázquez & F. Rubiera Morollón (Eds.), *Defining the Spatial Scale in Modern Regional Analysis* (pp. 237–254). Berlin: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31994-5_12
- Ella, L. P. A., van den Bergh, F., van Wyk, B. J., & van Wyk, M. A. (2008). A Comparison of Texture Feature Algorithms for Urban Settlement Classification. In *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium* (Vol. 1, p. III-1308-III-1311). IEEE. <https://doi.org/10.1109/IGARSS.2008.4779599>
- Engstrom, R., Newhouse, D., Haldavaneekar, V., Copenhaver, A., & Hersh, J. (2017). Evaluating the relationship between spatial and spectral features derived from high spatial resolution satellite data and urban poverty in Colombo, Sri Lanka. In *2017 Joint Urban Remote Sensing Event (JURSE)* (pp. 1–4). IEEE. <https://doi.org/10.1109/JURSE.2017.7924590>
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. (M. Carmichael, Ed.), *SAGE Publications Ltd* (Vol. 53). London. <https://doi.org/10.1017/CBO9781107415324.004>
- Flowerdew, R., Manley, D. J., & Sabel, C. E. (2008). Neighbourhood effects on health: Does it matter where you draw the boundaries? *Social Science and Medicine*, 66(6), 1241–1255. <https://doi.org/10.1016/j.socscimed.2007.11.042>
- Graesser, J., Cheriyyadat, A., Vatsavai, R. R., Chandola, V., Long, J., & Bright, E. (2012). Image based characterization of formal and informal neighborhoods in an urban landscape. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(4), 1164–1176. <https://doi.org/10.1109/JSTARS.2012.2190383>
- Greenacre, M. (2017). *Correspondence Analysis in Practice* (3rd ed.). CRC Press.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter*, 1026–1034. <https://doi.org/10.1109/ICCV.2015.123>
- Henninger, N., & Snel, M. (2002). *Where are the poor? : experiences with the development and use of poverty maps*. Arendal: World Resources Institute Washington, DC, UNEP/GRID- Arendal. Retrieved from <http://pdf.wri.org/wherepoor.pdf>
- Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Jayatilaka, B., & Chatterji, M. (2007). Globalization and Regional Economic Development: A Note on Bangalore City. *Studies in Regional Science*, 37(2), 315–333. <https://doi.org/10.2457/srs.37.315>
- 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 (New York, N.Y.)*, 353(6301), 790–4. <https://doi.org/10.1126/science.aaf7894>
- Jolliffe, I. T. (2002). Principal Component Analysis and Factor Analysis. In *Principal Component Analysis* (pp. 150–166). New York: Springer-Verlag. https://doi.org/10.1007/0-387-22440-8_7

- Kit, O., Lüdeke, M., & Reckien, D. (2012). Texture-based identification of urban slums in Hyderabad, India using remote sensing data. *Applied Geography*, 32, 660–667. <https://doi.org/10.1016/j.apgeog.2011.07.016>
- Knofczynski, G. T., & Mundfrom, D. (2008). Sample sizes when using multiple linear regression for prediction. *Educational and Psychological Measurement*, 68(3), 431–442. <https://doi.org/10.1177/0013164407310131>
- Kohli, D., Sliuzas, R., Kerle, N., & Stein, A. (2012). An ontology of slums for image-based classification. *Computers, Environment and Urban Systems*, 36(2), 154–163. <https://doi.org/10.1016/j.compenvurbsys.2011.11.001>
- Kohli, D., Sliuzas, R., & Stein, A. (2016). Urban slum detection using texture and spatial metrics derived from satellite imagery. *Journal of Spatial Science*, 61(2), 405–426. <https://doi.org/10.1080/14498596.2016.1138247>
- Kombe, W. J. (2005). Land use dynamics in peri-urban areas and their implications on the urban growth and form: The case of Dar es Salaam, Tanzania. *Habitat International*, 29(1), 113–135. [https://doi.org/10.1016/S0197-3975\(03\)00076-6](https://doi.org/10.1016/S0197-3975(03)00076-6)
- Krishna, A., Sriram, M. S., & Prakash, P. (2014). Slum types and adaptation strategies: identifying policy-relevant differences in Bangalore. *And Development (IIED)*, 26(2), 568–585. <https://doi.org/10.1177/0956247814537958>
- Krizhevsky, A., Sutskever, I., & Geoffrey E., H. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25 (NIPS2012)*, 1–9. <https://doi.org/10.1109/5.726791>
- Kuffer, M., Pfeffer, K., & Sliuzas, R. (2016). Slums from Space—15 Years of Slum Mapping Using Remote Sensing. *Remote Sensing*, 8(6), 455. <https://doi.org/10.3390/rs8060455>
- Kuffer, M., Pfeffer, K., Sliuzas, R., & Baud, I. (2016). Extraction of Slum Areas From VHR Imagery Using GLCM Variance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5), 1830–1840. <https://doi.org/10.1109/JSTARS.2016.2538563>
- Kuffer, M., Pfeffer, K., Sliuzas, R., Baud, I., & Maarseveen, M. (2017). Capturing the Diversity of Deprived Areas with Image-Based Features: The Case of Mumbai. *Remote Sensing*, 9(4), 384. <https://doi.org/10.3390/rs9040384>
- Le Roux, B., & Rouanet, H. (2011). *Multiple correspondence analysis*. Thousand Oaks: SAGE Publications, Inc.
- Marmanis, D., Datcu, M., Esch, T., & Stilla, U. (2016). Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks. *IEEE Geoscience and Remote Sensing Letters*, 13(1), 105–109. <https://doi.org/10.1109/LGRS.2015.2499239>
- Martinez, J., Pfeffer, K., & Baud, I. (2016). Factors shaping cartographic representations of inequalities. Maps as products and processes. *Habitat International*, 51, 90–102. <https://doi.org/10.1016/j.habitatint.2015.10.010>
- MatConvNet Team. (2017). Pretrained CNNs. Retrieved February 11, 2018, from <http://www.vlfeat.org/matconvnet/pretrained/>
- Mboga, N. O. (2017). *Detection of Informal Settlements From VHR Satellite Images Using Convolutional Neural Networks (MSc thesis)*. Enschede: University of Twente Faculty of Geo-Information and Earth Observation (ITC). Retrieved from http://www.itc.nl/library/papers_2017/msc/gfm/mboga.pdf
- Messer, L. C., Laraia, B. A., Kaufman, J. S., Eyster, J., Holzman, C., Culhane, J., ... O'Campo, P. (2006). The development of a standardized neighborhood deprivation index. *Journal of Urban Health*, 83(6), 1041–1062. <https://doi.org/10.1007/s11524-006-9094-x>
- Mevik, B.-H., & Wehrens, R. (2007). The pls Package: Principal Component and Partial Least Squares Regression in R, 18(2). Retrieved from <http://hdl.handle.net/2066/36604>
- Ministry of Housing Communities & Local Government. (2015). English indices of deprivation. Retrieved August 11, 2017, from <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015>
- Munyati, C., & Motholo, G. L. (2014). Inferring urban household socio-economic conditions in Mafikeng, South Africa, using high spatial resolution satellite imagery. *Urban, Planning and Transport Research*, 20(June 2014), 1–15. <https://doi.org/10.1080/21650020.2014.901158>
- Nagi, R. (2014). ESRI's World Elevation Services. Retrieved January 25, 2018, from <https://blogs.esri.com/esri/arcgis/2014/07/11/introducing-esris-world-elevation-services/>

- Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. Determination Press.
- Nijman, J. (2008). Against the odds: Slum rehabilitation in neoliberal Mumbai. *Cities*, 25(2), 73–85. <https://doi.org/10.1016/j.cities.2008.01.003>
- Noble, M., Wright, G., Smith, G., & Dibben, C. (2006). Measuring multiple deprivation at the small-area level. *Environment and Planning A*, 38(1), 169–185. <https://doi.org/10.1068/a37168>
- Nolan, B., & Whelan, C. T. (2010). Using Non-Monetary Deprivation Indicators to Analyze Poverty and Social Exclusion: Lessons from Europe? *Journal of Policy Analysis and Management*, 22(2), 305–325. <https://doi.org/10.1002/pam>
- Northern Ireland Statistics and Research Agency. (2010). *Northern Ireland Multiple Deprivation Measure 2010*. Belfast. Retrieved from https://www.nisra.gov.uk/sites/nisra.gov.uk/files/publications/NIMDM_2010_Report_0.pdf
- NVIDIA. (2018a). CUDA Toolkit. Retrieved January 21, 2018, from <https://developer.nvidia.com/cuda-toolkit>
- NVIDIA. (2018b). NVIDIA cuDNN. Retrieved January 21, 2018, from <https://developer.nvidia.com/cudnn>
- NVIDIA. (2018c). QUADRO. Retrieved January 21, 2018, from <http://www.nvidia.com/object/quadro-mobile-features-benefits.html>
- Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971–987. <https://doi.org/10.1109/TPAMI.2002.1017623>
- Olthuis, K., Benni, J., Eichwede, K., & Zevenbergen, C. (2015). Slum Upgrading: Assessing the importance of location and a plea for a spatial approach. *Habitat International*, 50, 270–288. <https://doi.org/10.1016/j.habitatint.2015.08.033>
- Pacione, M. (2009). Poverty and Deprivation in Western City. In *Urban Geography: A Global Perspective* (pp. 308–329). Routledge.
- Pampalon, R., Hamel, D., Gamache, P., & Raymond, G. (2009). A deprivation index for health planning in Canada. *Chronic Diseases in Canada*, 29(4), 178–191. Retrieved from http://publications.gc.ca/collections/collection_2009/aspc-phac/H12-27-29-4E.pdf
- Patino, J. E., & Duque, J. C. (2013). A review of regional science applications of satellite remote sensing in urban settings. *Computers, Environment and Urban Systems*, 37, 1–17. <https://doi.org/10.1016/j.compenvurbsys.2012.06.003>
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373–1379. [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3)
- Persello, C., & Stein, A. (2017). Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images. *IEEE Geoscience and Remote Sensing Letters*, 14(12), 2325–2329. <https://doi.org/10.1109/LGRS.2017.2763738>
- Rains, E., Krishna, A., & Wibbels, E. (2017). *SLUMMIER THAN OTHERS: A Continuum of Slums and Assortative Residential Selection*.
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. In L. LIU & M. TAMER ÖZSU (Eds.), *Encyclopedia of Database Systems* (pp. 532–538). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-39940-9_565
- Richards, J. a, & Jia, X. (2006). *Remote Sensing Digital Image Analysis. Methods* (5th ed.). Berlin/Heidelberg: Springer-Verlag. <https://doi.org/10.1007/3-540-29711-1>
- Rumelheart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(9), 533–536.
- Saharan, T., Pfeffer, K., & Baud, I. (2017). Urban Livelihoods in Slums of Chennai: Developing a Relational Understanding. *European Journal of Development Research*, 1–21. <https://doi.org/10.1057/s41287-017-0095-2>
- Sandborn, A., & Engstrom, R. N. (2016). Determining the Relationship between Census Data and Spatial Features Derived from High-Resolution Imagery in Accra, Ghana. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. <https://doi.org/10.1109/JSTARS.2016.2519843>

- Scott, G. J., England, M. R., Starks, W. A., Marcum, R. A., & Davis, C. H. (2017). Training Deep Convolutional Neural Networks for Land-Cover Classification of High-Resolution Imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(4), 549–553. <https://doi.org/10.1109/LGRS.2017.2657778>
- Scottish Executive. (2006). *Scottish Index of Multiple Deprivation 2006 Technical Report. Comparative and General Pharmacology*. Retrieved from <http://www.gov.scot/Resource/Doc/933/0041180.pdf>
- Simard, P., Steinkraus, D., & Platt, J. C. (2003). Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. *Proceedings of the 7th International Conference on Document Analysis and Recognition*, 958–963. <https://doi.org/10.1109/ICDAR.2003.1227801>
- Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition, 1–14. <https://doi.org/10.1016/j.infsof.2008.09.005>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15, 1929–1958. <https://doi.org/10.1214/12-AOS1000>
- Stehman, S. V. (2009). Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing*, 30(20), 5243–5272. <https://doi.org/10.1080/01431160903131000>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07–12–June, 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Thomson, C. N., & Hardin, P. (2000). Remote sensing/GIS integration to identify potential low-income housing sites. *Cities*, 17(2), 97–109. [https://doi.org/10.1016/S0264-2751\(00\)00005-6](https://doi.org/10.1016/S0264-2751(00)00005-6)
- UN-Habitat. (2003). *The Challenge of Slums - Global Report on Human Settlements*. Earthscan Publications on behalf of UN-Habitat. London: Earthscan Publications Ltd. <https://doi.org/http://dx.doi.org/10.1108/meq.2004.15.3.337.3>
- UN-Habitat. (2009). Urban Indicators Guidelines, (August), 47. Retrieved from <http://www.unhabitat.org>
- UN-Habitat. (2015). *Informal settlements*. New York, NY: UN-Habitat. Retrieved from https://unhabitat.org/wp-content/uploads/2015/04/Habitat-III-Issue-Paper-22_Informal-Settlements.pdf
- United Nation. (2015). *World Urbanization Prospects, The 2014 Revision*. New York, NY: United Nation. Retrieved from <https://esa.un.org/unpd/wup/Publications/Files/WUP2014-Report.pdf>
- United Nations. (2016). *The World's Cities in 2016: Data Booklet. Economic and social affair*. Retrieved from http://www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2016_data_booklet.pdf
- Vatsavai, R. R. (2012). Scalable Multi-Instance Learning Approach for Mapping the Slums of the World. In *2012 SC Companion: High Performance Computing, Networking Storage and Analysis* (pp. 833–837). IEEE. <https://doi.org/10.1109/SC.Companion.2012.117>
- Vedaldi, A., & Lenc, K. (2015). MatConvNet. *Proceedings of the 23rd ACM International Conference on Multimedia - MM '15*, 689–692. <https://doi.org/10.1145/2733373.2807412>
- Vermeiren, K., Van Rompaey, A., Loopmans, M., Serwajja, E., & Mukwaya, P. (2012). Urban growth of Kampala, Uganda: Pattern analysis and scenario development. *Landscape and Urban Planning*, 106(2), 199–206. <https://doi.org/10.1016/j.landurbplan.2012.03.006>
- Weeks, J. R., Getis, A., Stow, D. A., Hill, A. G., Rain, D., Engstrom, R., ... Ofiesh, C. (2012). Connecting the Dots Between Health, Poverty and Place in Accra, Ghana. *Annals of the Association of American Geographers*, 102(5), 932–941. <https://doi.org/10.1080/00045608.2012.671132>
- Weeks, J. R., Hill, A., Stow, D., Getis, A., & Fugate, D. (2007). Can we spot a neighborhood from the air? Defining neighborhood structure in Accra, Ghana. *GeoJournal*, 69(1–2), 9–22. <https://doi.org/10.1007/s10708-007-9098-4>
- Welsh Government. (2014). *Welsh Index of Multiple Deprivation (WIMD) 2014*. Cardiff. Retrieved from <http://gov.wales/docs/statistics/2015/150812-wimd-2014-revised-en.pdf>

- Williams, N., Quincey, D., & Stillwell, J. (2016). Automatic Classification of Roof Objects From Aerial Imagery of Informal Settlements in Johannesburg. *Applied Spatial Analysis and Policy*, 9(2), 269–281. <https://doi.org/10.1007/s12061-015-9158-y>
- Zhang, L., Zhang, L., & Du, B. (2016). Deep learning for Remote Sensing Data. *IEEE Geoscience and Remote Sensing Magazine*, 4(2), 22–40. <https://doi.org/10.1155/2016/7954154>

APPENDIX

Annex 1: Indicators to be collected during the fieldwork

The list of indicators is supposed to be collected during fieldwork. The idea is to briefly scan each slum area (identified by their boundaries) visually and check the most relevant choices in the indicator list. The slum areas are defined by the boundaries (i.e., polygons). Use Locus map or online google map (the links and instructions are expressed in the email) to find them with their respective unique ID. The idea is to find each area and look for a typical location somewhere close to the entrance to assess. If possible, a typical photo of each area could be very helpful for further analysis.

To make the indicators easy to collect, they were designed to be binary or categorized in levels, as many of them are subjective or qualitative indicators. The levels are mostly relative and could have different meanings in the context. The indicators are presented in three categories; building-related, environment-related, and people-related followed by their levels.

Building-Related indicators:

| <i>Indicator</i> | <i>Reference</i> |
|--|---|
| <i>Dominant building type</i> | (Bima et al., 2017; Kohli et al., 2012) |
| <i>Number of floors</i> | (Kohli et al., 2012) |
| <i>Dominant building footprint size</i> | (Kohli et al., 2012; Kuffer et al., 2017; UN-Habitat, 2003) |
| <i>Wall material</i> | (Kohli et al., 2012; Kuffer et al., 2017; UN-Habitat, 2003) |
| <i>Roof material</i> | (Kohli et al., 2012; Kuffer et al., 2017; UN-Habitat, 2003) |
| <i>Dominant shape of buildings</i> | (Kohli et al., 2012; Kuffer et al., 2017) |
| <i>Overall state of the buildings</i> | (Expert knowledge) |
| <i>Overall building appearance</i> | (Expert knowledge) |
| <i>Open spaces/green spaces*</i> | (Kuffer, Pfeffer, & Sliuzas, 2016; Kuffer et al., 2017; UN-Habitat, 2003) |
| <i>Appearance of open space</i> | (Expert knowledge) |

Note: *this means any space available in the neighbourhood in terms of accessible roads for any vehicle type, small vegetation around the buildings or open squares.

Environment-Related indicators:

| <i>Indicator</i> | <i>Reference</i> |
|--|--|
| <i>Presence of roads</i> | (Kohli et al., 2012) |
| <i>Road pavement (if there is road)</i> | (Kohli et al., 2012; UN-Habitat, 2003) |
| <i>Road material</i> | (Kohli et al., 2012; UN-Habitat, 2003) |
| <i>Road width (if there is road)</i> | (Kohli et al., 2012) |
| <i>Cables for electricity</i> | (UN-Habitat, 2003) |
| <i>Presence of foot paths</i> | (UN-Habitat, 2003) |
| <i>Foot path material (if there is foot path)</i> | (UN-Habitat, 2003) |
| <i>Street light</i> | (UN-Habitat, 2003) |
| <i>Pollution (smell)</i> | (Bima et al., 2017; Kuffer et al., 2017; Nolan & Whelan, 2010; UN-Habitat, 2003) |
| <i>Pollution (mechanical or extraordinary traffic noise)</i> | (Bima et al., 2017; Kuffer et al., 2017; Nolan & Whelan, 2010; UN-Habitat, 2003) |
| <i>Pollution (waste)</i> | (Bima et al., 2017; Kuffer et al., 2017; Nolan & Whelan, 2010; UN-Habitat, 2003) |
| <i>Open sewers</i> | (UN-Habitat, 2003) |
| <i>Presence of public toilet</i> | (Expert knowledge) |
| <i>Water body</i> | (Expert knowledge) |
| <i>Economic activities</i> | (UN-Habitat, 2009) |
| <i>Type of economic activities (if there is any)</i> | (UN-Habitat, 2009) |
| <i>Dominant land use next to the neighborhood</i> | (Kuffer, Pfeffer, & Sliuzas, 2016; Kuffer et al., 2017) |
| <i>Feeling safe</i> | (Kuffer et al., 2017) |
| <i>Are people interacting or chatting?</i> | (Nolan & Whelan, 2010) |
| <i>Are there vehicles visible within the area?</i> | (Bima et al., 2017) |
| <i>Is there any temple?</i> | (Expert knowledge) |

People-Related indicators:

| <i>Indicator</i> | <i>Reference</i> |
|--------------------------|---------------------|
| <i>Clothes of people</i> | (Bima et al., 2017) |
| <i>Having jewelry</i> | (Bima et al., 2017) |
| <i>Hair of children</i> | (Expert knowledge) |
| <i>Children toys</i> | (Bima et al., 2017) |

General Information

1. Survey Date: _____

2. Unique ID: G ____ S ____

3. X coordinate _____

4. Y coordinate _____

| Indicator | Level |
|---|---|
| Dominant building type | <input type="radio"/> Single-story <input type="radio"/> Single-story with garden <input type="radio"/> Multi-story <input type="radio"/> Multi-story with balcony |
| In case of a mix of building types specify the approximate % | _____% _____% _____% _____% |
| Number of floors | <input type="radio"/> One <input type="radio"/> Two <input type="radio"/> Three <input type="radio"/> Four <input type="radio"/> Five + |
| In case of a mix of number of floors specify the approximate % | _____% _____% _____% _____% _____% |
| Dominant building footprint size | <input type="radio"/> Very small <input type="radio"/> Small <input type="radio"/> Medium <input type="radio"/> Large <input type="radio"/> Very large |
| Wall material | <input type="radio"/> Temporary <input type="radio"/> Permanent <input type="radio"/> Mix |
| Roof material | <input type="radio"/> Plastic <input type="radio"/> Metal <input type="radio"/> Asbestos <input type="radio"/> Tile <input type="radio"/> Concrete <input type="radio"/> Others please specify: _____ |
| In case of a mix of roof material specify the approximate % | _____% _____% _____% _____% _____% _____% |
| Dominant shape of buildings | <input type="radio"/> Simple <input type="radio"/> Complex |
| Overall state of the buildings | <input type="radio"/> Not maintained well <input type="radio"/> Well-maintained |
| Overall building appearance | <input type="radio"/> Simple <input type="radio"/> Some decorations <input type="radio"/> Many decorations |
| Open spaces/green spaces | <input type="radio"/> Not available <input type="radio"/> Some <input type="radio"/> Many |
| Appearance of open space | <input type="radio"/> Clean without vegetation <input type="radio"/> Clean with vegetation cover <input type="radio"/> Not clean without vegetation cover <input type="radio"/> Not clean with vegetation cover |

| Indicator | Level |
|---|---|
| Presence of roads | <input type="radio"/> No <input type="radio"/> Yes |
| Road pavement (if there is road) | <input type="radio"/> Not paved <input type="radio"/> Mostly unpaved <input type="radio"/> Mostly paved <input type="radio"/> All paved |
| Road material | <input type="radio"/> Asphalt <input type="radio"/> Gravel <input type="radio"/> Sand <input type="radio"/> Cobble <input type="radio"/> Mix <input type="radio"/> Other, please specify: _____ |
| Road width (if there is road) (meter) | <input type="radio"/> [1-1.5] <input type="radio"/> (1.5-2.5] <input type="radio"/> (2.5-4] <input type="radio"/> (4-6] <input type="radio"/> More, please specify: _____ |
| Cables for electricity | <input type="radio"/> Not exist <input type="radio"/> Exist |
| Presence of foot paths | <input type="radio"/> Not exist <input type="radio"/> Exist |
| Foot path material (if there is foot path) | <input type="radio"/> Asphalt <input type="radio"/> Gravel <input type="radio"/> Sand <input type="radio"/> Cobble <input type="radio"/> Mix <input type="radio"/> Other, please specify: _____ |

| <i>Indicator</i> | <i>Level</i> | | | |
|--|-------------------------------------|---|---|---|
| <i>Street light</i> | <input type="radio"/> Not exist | <input type="radio"/> Exist | | |
| <i>Pollution (smell)</i> | <input type="radio"/> Yes | <input type="radio"/> No | | |
| <i>Pollution (mechanical or extraordinary traffic noise)</i> | <input type="radio"/> Yes | <input type="radio"/> No | | |
| <i>Pollution (waste)</i> | <input type="radio"/> Yes | <input type="radio"/> No | | |
| <i>Open sewers</i> | <input type="radio"/> Yes | <input type="radio"/> No | | |
| <i>Presence of public toilet</i> | <input type="radio"/> Yes | <input type="radio"/> No | | |
| <i>Water body</i> | <input type="radio"/> No water body | <input type="radio"/> Polluted water body | <input type="radio"/> Clean water body | |
| <i>Economic activities</i> | <input type="radio"/> Yes | <input type="radio"/> No | | |
| <i>Type of economic activities (if there is any)</i> | <input type="radio"/> Agriculture | <input type="radio"/> Small commercial | <input type="radio"/> Animal husbandry | <input type="radio"/> Manufacturing |
| <i>Dominant land use next to the neighborhood</i> | <input type="radio"/> Industrial | <input type="radio"/> Agricultural | <input type="radio"/> Residential | <input type="radio"/> Commercial |
| <i>Feeling safe</i> | <input type="radio"/> Not safe | <input type="radio"/> Relatively safe | | |
| <i>Are people interacting or chatting?</i> | <input type="radio"/> No | <input type="radio"/> Yes | | |
| <i>Are there vehicles visible within the area?</i> | <input type="radio"/> Nothing | <input type="radio"/> Bikes | <input type="radio"/> Motor bikes (or scooters and Rickshaws) | <input type="radio"/> Cars <input type="radio"/> Trucks |
| <i>Is there any temple?</i> | <input type="radio"/> No | <input type="radio"/> Yes, Hindu | <input type="radio"/> Yes, mosque | <input type="radio"/> Yes, other |

| <i>Indicator</i> | <i>Level</i> | | |
|--------------------------|---|----------------------------------|------------------------------------|
| <i>Clothes of people</i> | <input type="radio"/> Torn and shabby | <input type="radio"/> Basic | <input type="radio"/> Well-dressed |
| <i>Having jewellery</i> | <input type="radio"/> Almost no | <input type="radio"/> Some | <input type="radio"/> Many |
| <i>Hair of children</i> | <input type="radio"/> Not maintained well | <input type="radio"/> Maintained | |
| <i>Children toys</i> | <input type="radio"/> No toy | <input type="radio"/> Basic toys | <input type="radio"/> Good toys |

Annex 2: Extracted indicators from HH survey 2010 related to IMD with respective categories, assumptions, and references

| Deprivation Dimension | Indicator form HH survey 2010 | Categories | Assumption | Reference |
|------------------------------|---|--|--|---|
| Social Capital | Caste | 1. Scheduled caste 2. Scheduled tribe 3. Backward class 4. Other backward class 5. General caste | Belonging to scheduled caste causes systematic differences in access to education and health services | (Deolalikar, 2005) |
| Human Capital | Highest Educational Obtained | 1. non-formal schooling 2. some formal schooling 3. Primary school 4. Middle school 5. High school 6. Pre-university college (puc) 7. Technical training 8. Bachelor's degree 9. Post-graduation 10. No education | Higher educational level reflects accessing to higher-skilled occupation and better livelihood | (Rains et al., 2017) |
| | Dependency rate | Proportion of workers in relation to all HH members (a continuous number between 0 and 1) | Having more workers enables more possibility to have better livelihood | (Baud et al., 2008) |
| | Distance to healthcare | 1. Less than 1km 2. 1 to 5km 3. more than 5km | More accessible healthcare facility potentially provides better healthcare services to HHs and decreases deprivation | (Baud et al., 2008) |
| Financial Capital | Income (Rupee/month) | 1. [200, 1300) 2. [1300, 2400) 3. [2400, 3500) 4. [3500, 4600) 5. [4600, 5700) 6. [5700, 8000) 7. [8000, 12000) 8. [12000, 18000) 9. More than 18000 | More income results in less poverty and deprivation | (Pacione, 2009) |
| | Ration Card | 1. Anthyodaya 2. BPL 3. APL 4. No ration card | More deprived areas have more ration cards. Anthyodaya was assumed as the ration card for worse-off people | (Rains et al., 2017) (Expert consultation) |
| Physical Capital | Water Source quality (Provided for summer and other seasons separately) | 1. Individual water connection 2. Makeshift water connection 3. Individual sub-connection 4. Mini water supply 5. Public tap 6. Community well / hand pump | More private and in-building water sources have better quality | (Rains et al., 2017) |

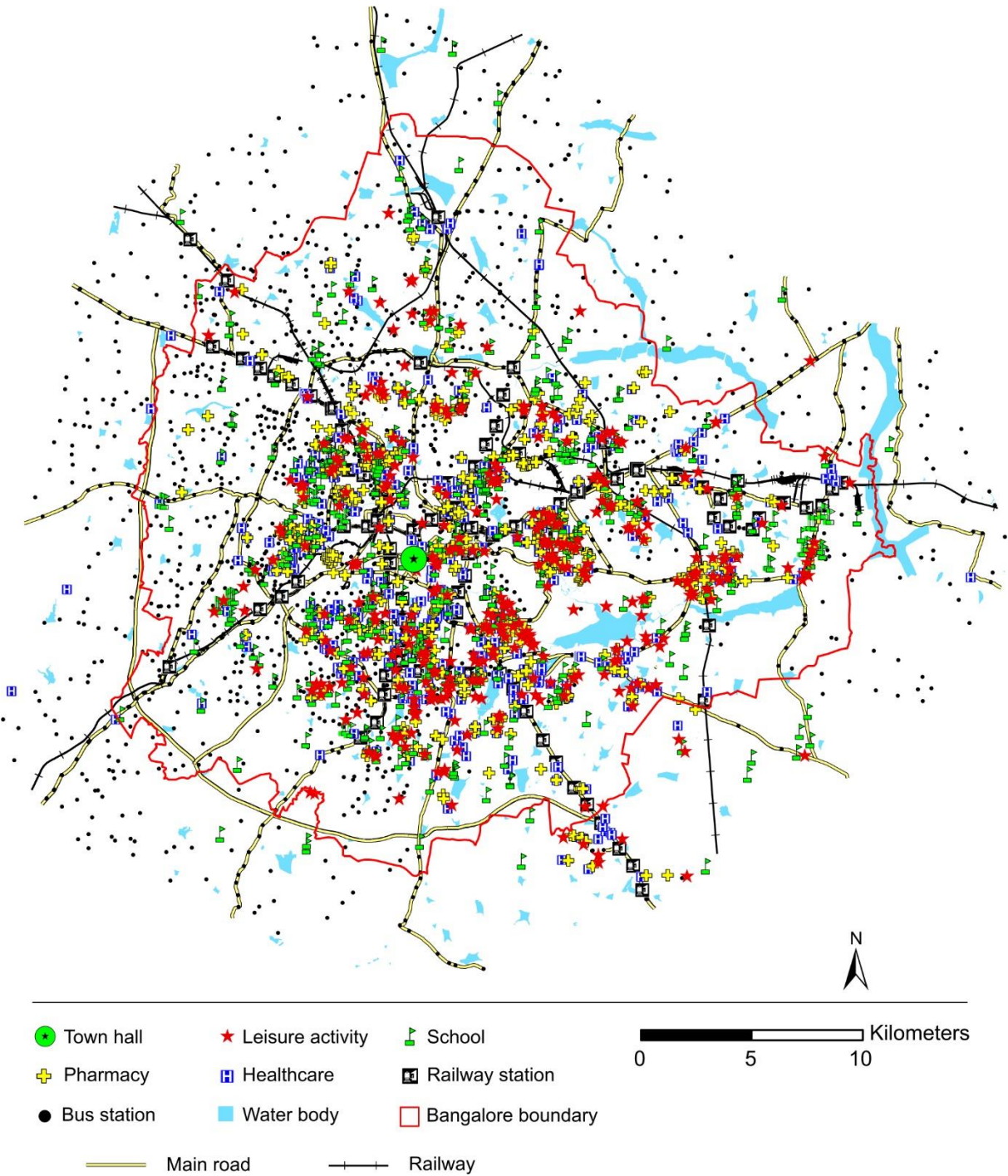
| Deprivation Dimension | Indicator form HH survey 2010 | Categories | Assumption | Reference |
|---------------------------|---|--|--|------------------------------------|
| | | 7. Water tanker 8. Surface water 9. Other vendors | | |
| | Toilet facility | 1. No toilet 2. Toilet shared with neighbours 3. Community toilet – free 4. Open space 5. Community toilet – paid 6. Other toilets 7. Own toilet | More private sanitation types have better quality and are more hygienic | (Rains et al., 2017) |
| | Access to Electricity | 1. Metered connection 2. Unmetered connection 3. Unofficial or makeshift connection 4. Through sub-contractor 5. No electricity | More official connection leads to better quality of electricity and less deprivation | Expert knowledge |
| | Crowdedness (pop/m ²) | Living area per capita | More living space shows less slum-ness and less deprivation | (Baud et al., 2008) |
| | Dwelling Age | Continuous variable of dwelling ages by year | Better-off slum dwellers live in older dwellings | (Krishna, Sriram, & Prakash, 2014) |
| | Floor material | 1. Mud 2. Wood/Bamboo 3. Brick 4. Stone 5. Cement 6. Mosaic/Tiles 7. Other floor materials | | (Rains et al., 2017) |
| | Wall material | 1. Grass/Thatch/Bamboo 2. Plastic/Polythene 3. Mud/Unburnt 4. Brick 5. Wood 6. G.I./Metal/Asbestos 7. Burnt brick 8. Stone 9. Concrete 10. Other wall materials | | (Rains et al., 2017) |
| | Roof material | 1. Grass/Thatch/Bamboo/Wood/Mud 2. Plastic/Polythene 3. Tiles 4. Slate 5. G.I./Metal/Asbestos 6. Brick 7. Stone 8. Concrete 9. Other roof materials | | (Rains et al., 2017) |
| Contextual capital | Travel time to services (Education/Work/Household purposes) | Average minutes take to get to education/work/household purpose in a household | | (Welsh Government, 2014) |

Annex 3: Number and percentage of missing values in HH data

Excluded indicators were highlighted.

| | Number | Percent |
|-----------------------------------|--------|---------|
| Caste | 391 | 35.1 |
| Highest education | 156 | 14.0 |
| Dependency rate | 0 | 0.0 |
| Distance to healthcare | 2 | 0.2 |
| Income | 32 | 2.9 |
| Ration card | 20 | 1.8 |
| Water summer first | 117 | 10.5 |
| Water summer second | 930 | 83.5 |
| Water other seasons first | 67 | 6.0 |
| Water other seasons second | 926 | 83.1 |
| Toilet | 454 | 40.8 |
| Electricity | 63 | 5.7 |
| Crowdedness | 120 | 10.8 |
| Dwelling age | 748 | 67.1 |
| Floor material first | 16 | 1.4 |
| Floor material second | 1103 | 99.0 |
| Wall material first | 18 | 1.6 |
| Wall material second | 1095 | 98.3 |
| Roof material first | 23 | 2.1 |
| Roof material second | 1099 | 98.7 |
| Travel time to services | 123 | 11.0 |
| Travel time to education | 735 | 66.0 |
| Travel time to work | 342 | 30.7 |
| Travel time to household purposes | 412 | 37.0 |

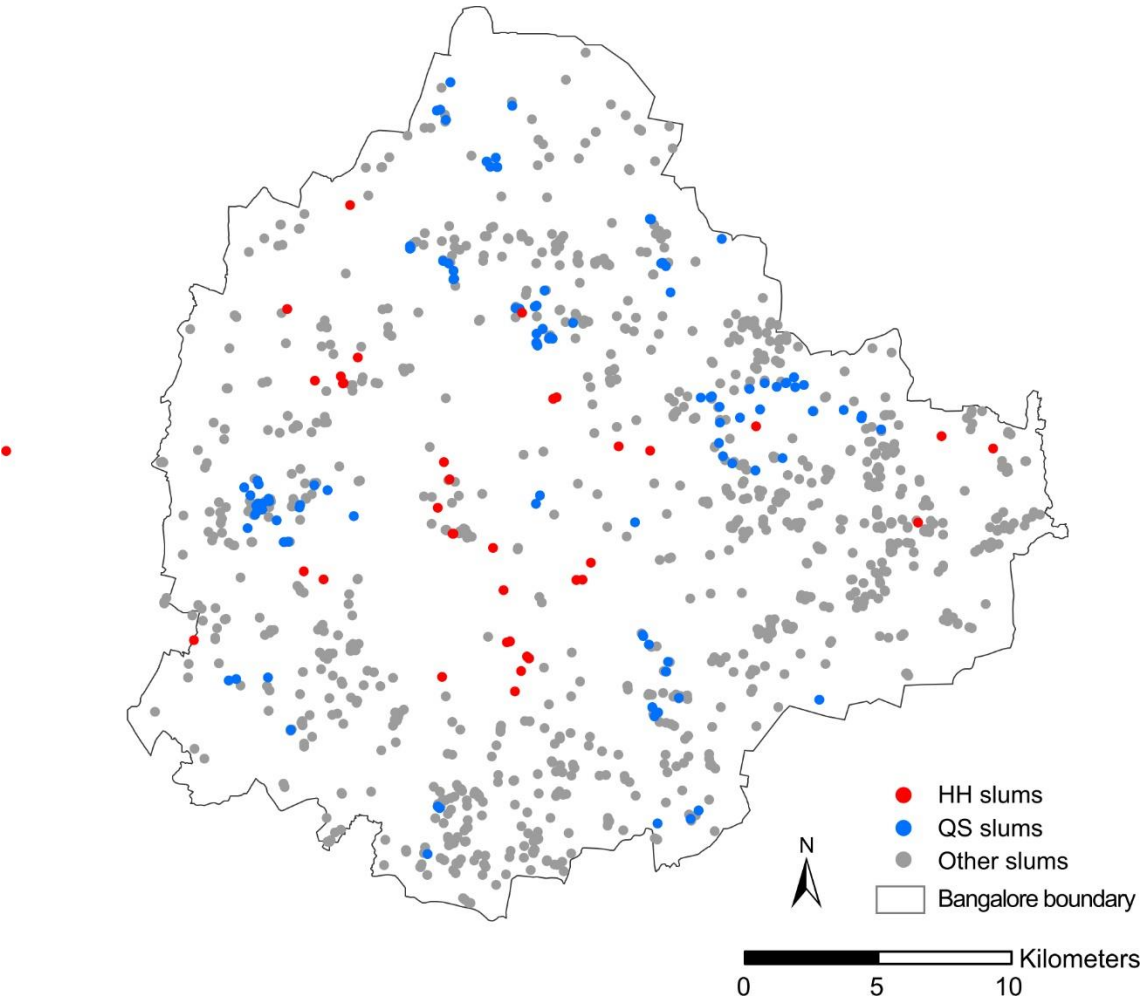
Annex 4: OSM data used to create GIS layers



Annex 5: Squared correlation of indicators and dimensions of HH data

| | Dimension | | |
|-------------------------|-----------|-------|-------|
| | 1 | 2 | 3 |
| Caste | 0.007 | 0.023 | 0.017 |
| Highest Education | 0.023 | 0.023 | 0.022 |
| Distance to healthcare | 0.038 | 0.044 | 0.012 |
| Income | 0.040 | 0.101 | 0.020 |
| Ration card | 0.109 | 0.048 | 0.007 |
| Water summer | 0.430 | 0.934 | 0.733 |
| Water other seasons | 0.484 | 0.940 | 0.733 |
| Toilet | 0.525 | 0.212 | 0.172 |
| Electricity | 0.647 | 0.149 | 0.290 |
| Floor material | 0.582 | 0.014 | 0.033 |
| Wall material | 0.542 | 0.251 | 0.264 |
| Roof material | 0.519 | 0.082 | 0.190 |
| Travel time to services | 0.036 | 0.168 | 0.151 |
| Living area per capita | 0.015 | 0.095 | 0.284 |
| Dwelling age | 0.047 | 0.046 | 0.057 |
| Dependency | 0.011 | 0.038 | 0.042 |

Annex 6: HH, QS, and other slum locations on a map



Annex 7: Squared correlation of indicators and dimensions of QS data

| | Dimension | | | | | |
|---|-----------|-------|-------|-------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Dominant building footprint size | 0.906 | 0.059 | 0.016 | 0.075 | 0.142 | 0.201 |
| Wall material | 0.634 | 0.003 | 0.023 | 0.109 | 0.196 | 0.002 |
| Dominant shape of buildings | 0.055 | 0.006 | 0.061 | 0.014 | 0.094 | 0.019 |
| Overall state of the buildings | 0.096 | 0.197 | 0.051 | 0.005 | 0.001 | 0.007 |
| Overall building appearance | 0.706 | 0.024 | 0.001 | 0.002 | 0.007 | 0.042 |
| Open spaces/green spaces | 0.195 | 0.010 | 0.116 | 0.152 | 0.012 | 0.126 |
| Appearance of open space | 0.086 | 0.284 | 0.030 | 0.326 | 0.091 | 0.172 |
| Presence of roads | 0.159 | 0.500 | 0.243 | 0.056 | 0.372 | 0.148 |
| Road pavement (if there is road) | 0.589 | 0.347 | 0.184 | 0.094 | 0.460 | 0.254 |
| Road material | 0.580 | 0.352 | 0.174 | 0.179 | 0.460 | 0.086 |
| Road width (if there is road) (meter) | 0.161 | 0.297 | 0.118 | 0.192 | 0.365 | 0.123 |
| Cables for electricity | 0.643 | 0.039 | 0.005 | 0.179 | 0.102 | 0.054 |
| Presence of foot paths | 0.168 | 0.025 | 0.001 | 0.014 | 0.021 | 0.003 |
| Foot path material | 0.314 | 0.320 | 0.617 | 0.115 | 0.070 | 0.105 |
| Street light | 0.274 | 0.014 | 0.038 | 0.003 | 0.001 | 0.104 |
| Pollution (smell) | 0.014 | 0.014 | 0.019 | 0.027 | 0.001 | 0.061 |
| Pollution (mechanical or extraordinary traffic noise) | 0.149 | 0.110 | 0.217 | 0.059 | 0.002 | 0.021 |
| Pollution (waste) | 0.004 | 0.291 | 0.001 | 0.123 | 0.001 | 0.040 |
| Presence of public toilet | 0.002 | 0.037 | 0.004 | 0.002 | 0.013 | 0.033 |
| Water body | 0.017 | 0.002 | 0.001 | 0.014 | 0.019 | 0.143 |
| Economic activities | 0.149 | 0.327 | 0.032 | 0.022 | 0.058 | 0.025 |
| Type of economic activities (if there is any) | 0.447 | 0.425 | 0.571 | 0.218 | 0.133 | 0.187 |
| Dominant land use next to the neighbourhood | 0.041 | 0.354 | 0.144 | 0.254 | 0.099 | 0.274 |
| Feeling safe | 0.004 | 0.158 | 0.270 | 0.002 | 0.051 | 0.032 |
| Are people interacting or chatting? | 0.138 | 0.074 | 0.015 | 0.071 | 0.039 | 0.045 |
| Are there vehicles visible within the area? | 0.515 | 0.255 | 0.267 | 0.188 | 0.008 | 0.184 |
| Is there any temple? | 0.365 | 0.087 | 0.099 | 0.136 | 0.066 | 0.123 |
| Clothes of people | 0.182 | 0.005 | 0.216 | 0.017 | 0.003 | 0.142 |
| Having jewellery | 0.188 | 0.020 | 0.027 | 0.046 | 0.019 | 0.103 |
| Hair of children | 0.128 | 0.030 | 0.009 | 0.120 | 0.027 | 0.017 |
| Children toys | 0.117 | 0.034 | 0.022 | 0.115 | 0.007 | 0.053 |
| Dominant roof material | 0.910 | 0.098 | 0.013 | 0.549 | 0.119 | 0.083 |
| Dominant no floors | 0.762 | 0.085 | 0.109 | 0.122 | 0.297 | 0.106 |
| Dominant building type | 0.808 | 0.035 | 0.050 | 0.057 | 0.113 | 0.083 |

Annex 8: Correlation between the first dimensions of QS and HH data with bootstrap sampling

| | | | D1QS | D2QS | D3QS | D4QS | D5QS | D6QS |
|-------------|---------------------|-------------------------|--------|--------|--------|--------|--------|--------|
| D1HH | Pearson Correlation | | .630** | -0.147 | 0.005 | -0.303 | 0.011 | -0.102 |
| | Sig. (2-tailed) | | 0.001 | 0.473 | 0.981 | 0.132 | 0.957 | 0.621 |
| | N | | 26 | 26 | 26 | 26 | 26 | 26 |
| | Bootstrap | Bias | -0.007 | -0.006 | 0.011 | -0.044 | -0.013 | -0.021 |
| | | Std. Error | 0.137 | 0.177 | 0.195 | 0.169 | 0.159 | 0.193 |
| | | 95% Confidence Interval | Lower | 0.275 | -0.511 | -0.344 | -0.690 | -0.376 |
| | | | Upper | 0.824 | 0.234 | 0.471 | -0.053 | 0.257 |
| D2HH | Pearson Correlation | | -0.092 | 0.040 | -0.102 | 0.038 | 0.143 | -0.154 |
| | Sig. (2-tailed) | | 0.655 | 0.848 | 0.620 | 0.853 | 0.487 | 0.453 |
| | N | | 26 | 26 | 26 | 26 | 26 | 26 |
| | Bootstrap | Bias | 0.007 | -0.004 | 0.021 | 0.004 | -0.022 | 0.011 |
| | | Std. Error | 0.162 | 0.205 | 0.207 | 0.151 | 0.247 | 0.182 |
| | | 95% Confidence Interval | Lower | -0.395 | -0.359 | -0.448 | -0.256 | -0.424 |
| | | | Upper | 0.225 | 0.429 | 0.358 | 0.351 | 0.522 |
| D3HH | Pearson Correlation | | -0.047 | 0.214 | -0.325 | -0.229 | 0.158 | -0.232 |
| | Sig. (2-tailed) | | 0.821 | 0.293 | 0.105 | 0.260 | 0.441 | 0.255 |
| | N | | 26 | 26 | 26 | 26 | 26 | 26 |
| | Bootstrap | Bias | 0.004 | -0.005 | 0.003 | 0.004 | -0.011 | 0.013 |
| | | Std. Error | 0.165 | 0.200 | 0.174 | 0.160 | 0.197 | 0.161 |
| | | 95% Confidence Interval | Lower | -0.368 | -0.187 | -0.648 | -0.516 | -0.256 |
| | | | Upper | 0.288 | 0.596 | 0.033 | 0.109 | 0.577 |

Note: ** means $p < 0.01$

Annex 9: Objective function and error curves resulted from the simple CNN

