

Mapping Poverty in Bangladesh

with Satellite Images and Deep Learning

Dang Ngoc Huy

Thesis for Master of Public Policy

Class of 2020

Supervised by Dr. Slava Jankin

24th May 2020

Abstract

Mapping poverty to better target aid and development is a difficult undertaking for a number of reasons: the reliance on the collection of individual household data which is time-consuming and costly; household surveys are often not available for many countries of interest and when they are, it is usually not at the desired frequency and this time-lag creates problems in the decision making process for policy makers as well as development and aid agencies; to get a sense of a region's current development, it is often essential to use predictions to fill in the gap in the time-series data of a region's economic indicators, which can be questionable. Recent researches and development in the field of Computer Vision and Deep Learning have displayed the effectiveness of employing satellite imagery to map out impoverished areas in the sub-Saharan African region (Jean et al., 2016). This study seeks to expand further on this research and investigate the extent to which this methodology can be applied in the context of Bangladesh to predict wealth distribution of households within the country. Three different non-survey unstructured data sources and five distinct models have been explored to identify the best approach for poverty estimation in the face of survey data scarcity. In this respect, the state-of-the-art model utilizing daytime and nighttime satellite images captures the most of the welfare variation in Bangladeshi household clusters at 72 percent. However, the model relying on open-source data of nightlight intensity and geo-spatial mapping also presents a promising alternative that can achieve similar result at 70 percent while maintaining ease of access and no incurring cost for acquisition^{*}.

^{*}Notebooks and code for the project can be accessed here: [Research Github Repository](#)

Contents

5	Chapter 1 Executive Summary
8	Chapter 2 Literature Review
12	Chapter 3 The Story of Bangladesh
14	Chapter 4 Data Sources
18	Chapter 5 Methodology
32	Chapter 6 Results and Analysis
40	Chapter 7 Conclusion

1

Executive Summary

In 2015, the United Nations adopted 17 Sustainable Development Goals, one of which is the complete eradication of poverty by 2030. To realize such an ambition, the first step is essentially to locate the areas most affected by poverty, in order to more effectively target aid, financial support and development programs. However, identifying areas of poverty is a costly and intricate endeavor with complex politics, as well as limited and unreliable availability of information and resources. To provide a granular picture of economic well-being of a region, oftentimes an elaborate and comprehensive household survey is required which prompts a series of complications from funding accessibility, monitoring and evaluation procedures to government accountability. This undertaking becomes an even more convoluted affair when the governments presiding over the affected population want to conceal their inadequate engagement with their poorest citizens, and thus deliver inferior data and records quality that interfere with the process of poverty mapping.

In recent years, a number of identification strategies adopting Machine Learning algorithms has proven to be an effective alternative to the costly traditional household surveys. The use of publicly available data, particularly nighttime and daytime satellite imagery, combined with Computer Vision and Deep Learning methodologies, has been heralded as a relatively straightforward, simple and low cost solution for measuring poverty, while bypassing the issue of institutional complications (Jean et al., 2016). Theoretically, nighttime luminosity yields a reliable proxy for welfare as it contains signals for electricity usage, which in turn is highly correlated with levels of economic activities, infrastructure investment and prosperity (Ghosh et al., 2013). Daytime satellite images, on the other hand, contain other geographical features that might also indicate connections with socio-economic well-being of the population residing around them. A variety of research has been conducted in this promising direction to gauge the effectiveness of this new approach in extracting information from satellite images in helping poverty eradication, with both successes as well as limitations. However, the bulk of these studies still remains focused on impoverished nations in Africa (Heitmann and Bur, 2019; Jean et al., 2016).

This research has been implemented with the goal of building up on what has been accomplished thus far and providing another narrative on whether this method could achieve a favorable outcome in a completely different setting such as Bangladesh, a country in the South Asian region. With a vastly different typography of lush green landmass, intersected with swift rivers and diverse mountain ranges, Bangladesh supplies a stark contrast to the geography of Africa that could help to assess the validity of the methodology with an unfamiliar data distribution. To expand further on the approach, apart from the state-of-the-art model utilizing satellite images, other Machine Learning models with open-source data such as Open Street Map, which offers unstructured geospatial mapping of landscape features, are also brought into service to ascertain the most sensible strategy for real-world use case.

The comparison of these different poverty estimation techniques illustrates that while the model employing daytime and nighttime satellite imagery analysis still reigns with top performance, open-source data combining nighttime light intensity and Open Street Map is also able to create predictions that achieve substantially close to the state-of-the-art result, explaining up to 70 percent of the variation in welfare in Bangladesh households. These non-

proprietary and publicly available data also have the added benefits of requiring neither cost nor time complexities for implementation, offering ease of access with frequently updated granular information - all the essential qualities for a sustainable data streaming source for research, policy evaluation and development practice.

This paper will first provide an overview of the current literature in the area of poverty measurement with the use of satellite images analysis. Then, the proposed method relevant to the research will be outlined along with several concerns regarding its implementation as well as a brief overview of the poverty and inequality solution in Bangladesh to provide the basis of ground truth for analysis. Finally, an elaboration of the experiments will be followed by an examination of the results and discussion on the potential policy implications of the findings.

2

Literature Review



Summary

2.1

Poverty from space: a new hope, 9

2.2

Supporting research, 10

2.1 Poverty from space: a new hope

To resolve the problem of poverty identification in the scarcity of available data, Neal Jean and his team of researchers at Stanford University's Department of Earth System Science had proposed the novel approach of analyzing satellite images of the planet's surface to pinpoint regions and cluster areas that might be regarded as affected by impoverishment in their seminal work: *Combining satellite imagery and machine learning to predict poverty* (Jean et al., 2016).

To train a model that can identify impoverished areas, Jean and his team utilized three different data types: images of luminosity at night time which could signal economic level; images of daytime from which useful information on landscape features and socioeconomic data could be extracted; and finally survey data on household consumption expenditure and asset wealth as validation for the findings extrapolated from the analysis of satellite images.

A number of challenges plagued this pursuit. The theoretical framework employed by the authors rested on the assumption that there is a correlation between the nightlight luminosity of a region reflected in their satellite images and their levels of economic activities since less intensity could potentially indicate less consumption and wealth level in an area. However, as the authors observed, places hidden in darkness also carry hidden information. Consequently, this method is unable to detect granular economic activities in regions that are near and below the international poverty line, which makes it less useful to distinguish the poverty-stricken areas of interest. In addition, satellite images are notoriously unstructured, with little to no labels available, which makes supervised learning to classify and extract features difficult to implement.

To accommodate for these drawbacks, the learning process of their research occurred in three consecutive phases, each of which supplemented one another to improve the tracking of clusters of people living in poverty. In the first phase, a transfer learning approach was employed with a convolutional neural network (CNN) model that has been trained on ImageNet to detect features such as edges and corners. In the second phase, this CNN model was further trained to estimate nighttime luminosity level based on the daytime satellite images on a global scale, with the idea that the model could discern certain features in the daylight images such as the segmentation of land use, signs of human assets and activities as well as other useful geographic information that were indicative of the variation in nightlight luminosity - an acceptable proxy for economic activities. In the final phase, economic survey data on clusters of households was combined with the features extracted by the neural network to train and learn the patterns and connection between the different features that it had identified from the images and the household spending and asset possession using a Ridge Regression model, ultimately to estimate the poverty indicators of interest. Focusing on five countries from Africa - Nigeria, Tanzania, Uganda, Malawi, and Rwanda, the researchers were able to explain up to 75 percent of the variation in economic outcomes at the local level in cross-validated sets just based on the satellite images.

One particular ingenious implementation was the analysis of nightlight luminosity and how

it added to the ability of the model to extract information. Although nightlight information was a flawed proxy for economic levels since it demonstrated little correlation with the variation for expenditure spending at a lower and more fine-grained level, daytime imagery features and the socioeconomic information that could be extracted such as materials used on roof of housing to distance from the cluster of households of interest to urban areas were potentially good estimators for the expenditure even at a granular level. Therefore, by training the CNN architecture on both information in nightlight and daytime imagery, the model could learn and identify features that might capture variations across the entire expenditure and household consumption distribution, leading to a more accurate picture of areas affected by poverty and of lower economic-well-being in comparison to others. In this approach of training on daytime satellite images to estimate nightlight luminosity, the model could learn to identify landscape features without needing explicit annotations and labelling, reducing the need for direct supervision while still able to detect useful information relevant to the research question.

With a model that now had an understanding of the correlation between the features present in the satellite images and poverty level, the research could then map out predictions on regions where there were gaps in the survey data, helping to identify previously unknown areas for aid targeting.

2.2 Supporting research

Following a similar investigation, Babenko et al. also sought to understand the usefulness of satellite image analysis with deep learning in identifying poverty through their paper *Poverty mapping using convolutional neural networks trained on high and medium resolution satellite images, with an application in Mexico* (Babenko et al., 2017). Using a different CNN model and architecture setup with GoogleNet, also finetuned from ImageNet, their research was able to successfully estimate land usage through analysing satellite images and subsequently utilized this as a predictor for poverty indicators. Their best model, thus, could explain up to 57 percent of the poverty variation in poverty in a sample set of 10 percent of municipalities in Mexico. Their conclusion, similar to Jean et al., 2016, was that a CNN architecture learning from satellite images could be an end-to-end solution to understand and estimate poverty, in the face of unavailable survey data. However, they also underscored that the training process and how it influenced out of sample validation needed to be further researched and understood.

Another research that helps to solidify the confidence of this approach is a World Bank funded policy research paper from Ryan Engstrom, Jonathan Hersh and David Newhouse looking at how highly spatial satellite images can estimate poverty and economic level in Sri Lanka in *Poverty from Space Using High-Resolution Satellite Imagery for Estimating Economic Well-Being* (Engstrom et al., 2017). In a similar vein as previous studies, Engstrom et al. extracted features through the analysis of satellite view to identify landscape information such as agricultural practices, density of building and infrastructures, density and lengths

of roads, the number of cars, roofing materials, as well as other textual details. A simple linear regression with these features as input as predictors was able to explain up to 60 percent of poverty headcount rates and average log consumption for 1,291 administrative units in Sri Lanka.

These studies and their successes have demonstrated that the use of satellite images and deep learning is certainly a promising avenue to help understand the distribution of wealth and poverty. However, their fluctuating results also call into question the reliability of the approach and how it can be translated to other settings and environments. As any scientific endeavors, the more evidence there is, the more confident we can be of its effectiveness. This research, therefore, seeks to supplement the credence that this approach can work in different settings and countries, focusing on Bangladesh as a nation that has encountered its fair share of poverty and inequality issues.

3

The Story of Bangladesh

Bangladesh is one of the world's poorest nations. A part of East Pakistan following the border separation at the end of the British rule over India in 1947, the nation only achieved its independence recently after the Bangladesh Liberation War in 1971. The following years after the struggle for sovereignty witnessed a country ravaged by political upheaval, military coups, natural disasters, and profound poverty. The hostile historical foundation to establish political freedom was the major factor contributing to the lack of developmental progress in Bangladesh. Today, the country is home to 165.6 million people (2019), with a 20.5 percent poverty rate (Hill and Genoni, 2019).

In recent years, after the stability brought about by the return to democracy in 1991, the nation has accomplished remarkable progress, with rising economic growth of as much as 8 percent in 2018 and 7.9 percent in 2019, and 8 million Bangladeshis lifted out of poverty in only 6 years from 2010 to 2016 (Hill and Genoni, 2019). Apart from economic advancement, literacy rates, citizen's life expectancy, and in-country food production have also improved considerably. It is also on track to graduate from the United Nation's Least Developed Countries List in 2024 (before the Covid-19 crisis of 2020).

Yet, while the overall economic health of the country might be sustained, inequality and disparate distribution of wealth and opportunities are still largely a persisting problem that places a wedge in Bangladesh' society. One out of four people lives below the national poverty line. The pace of poverty reduction also has slowed down, despite the increasingly prosperous economy. The majority of alleviation transpired in rural areas, whereas the portion of people facing extreme conditions in urban spaces continues to be unaffected. On the other hand, the expeditious economic expansion seems to reserve resources mainly for the big cities such as the capital of Dhaka, instigating a massive exodus of people to metropolitan areas, placing further strains on rural regions. The challenge, therein lies in addressing this new frontier of poverty - managing urban poverty and wealth distribution while ensuring those who were recently lifted out of hardship do not face the risk of regressing.

The geography of Bangladesh is also an intriguing element, with a diverse physical landscape and a densely populated topography delineated by three distinct characteristics: a vast flatland, a region of hills intertwined with a swift river system and a long coastline. Gifted with the largest delta in the world, the Ganges Delta, Bangladesh is considered to possess one of the world's most fertile and arable soils, which provides most of the necessary sustenance for its population. These features, some of which can be extracted through our analysis, undoubtedly, will influence how Bangladeshis build and use their land, as well as their way of living and thriving.

For these reasons, Bangladesh presents a fascinating and challenging case study for the task of poverty identification in the face of survey data scarcity. Mapping out the regions of disparity in wealth distribution could help contributing to locating areas of improvement and potentially a better understanding of how different available resources or services could be correlated to inequality.

4

Data Sources



Summary

- 4.1 Demographic and Health Survey (DHS), 15
- 4.2 Global Map of Bangladesh, 15
- 4.3 Nighttime Luminosity Data, 15
- 4.4 Google Static Map Images, 16
- 4.5 Open Street Map Data, 17

4.1 Demographic and Health Survey (DHS)

To provide the ground truth for distribution of wealth across the country, the Bangladesh Demographic and Health Survey (DHS) is utilized. The survey was conducted by the National Institute of Population Research and Training (NIPORT) and Mitra and Associates of Dhaka which sampled roughly 17300 households from June 2014 to November 2014. DHS is a nationally representative survey that focuses on encapsulating the nation's socioeconomic indicators and thus can be used as a poverty benchmark with which to compare the models built from open-source data. The households covered in the survey are grouped into 600 different clusters with varying degrees of wealth based on their GPS locations, with a measure of random noise to preserve the privacy and identity of the participants in the survey (National Institute of Population Research and Training (NIPORT) and Mitra and Associates of Dhaka, 2014). The household wealth for each of the cluster is averaged and held as the principal indicator for socioeconomic well-being as it represents possession of common assets for these households.

4.2 Global Map of Bangladesh

To map out the boundaries of Bangladesh for the acquisition of nighttime and daytime satellite imagery data, a shapefile of the country, provided by Survey of Bangladesh is employed. The file contains vectorized and rasterized layers of geographical information that constitutes the main mapping features of the countries, including transport routes, administrative sectors boundaries, land usage, vegetation among other useful characteristics. These rasterized layers will help to bound the geo-spatial limit of images received from other data sources to focus mainly on the physical location of Bangladesh.

4.3 Nighttime Luminosity Data

Data on nightlight is collected from the NOAA National Centers for Environmental Information (NCEI) which provides continuous mapping of nighttime earth images (National Centers for Environmental Information, 2013). Global nightlight for the calendar year of 2013 represents a close alignment with the information derived from DHS. From the NOAA methodology, the data is procured from the visible infrared band digital number of cloud-free light detections 30 arc-second grids, meaning that the image will be able to capture nighttime luminosity of any region in the world with clarity, since it records the aggregated information of infrared wavelength throughout the year.

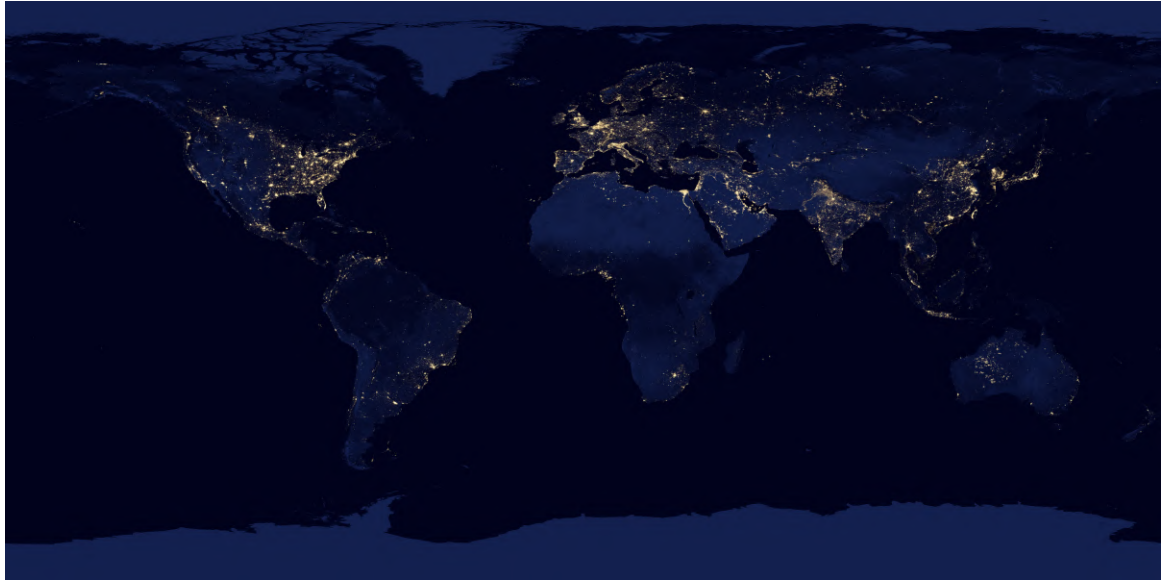


Figure 4.1: Earth At Night 2012. Courtesy of NASA/NOAA Suomi NPP. Released on April 5, 2013.

4.4 Google Static Map Images

The daylight satellite imagery data is collected using Google Static Map API. For a land mass area of approximately $130,170 \text{ km}^2$, 414,757 images in total have been collected to form the full picture of the country, with a zoom level of 16, pixel resolution 2.5m, image size 400x400 pixels and each image matching a single pixel of the nighttime luminosity data, covering about 0.25 km^2 . The daytime satellite images are organized so that they correspond to their nighttime images counterparts in terms of light intensity. This would accelerate the training and geographical features extraction tasks during the unsupervised training.



(a) Low intensity imagery



(b) Medium intensity imagery.



(c) High intensity imagery.

Figure 4.2: Daytime satellite imagery of Bangladesh from Google Static Map by light intensity

4.5 Open Street Map Data

Despite its relative ease of access, Google Static Map is still a proprietary enterprise product and thus incurs a cost for acquiring images. Even though this price is reasonable in comparison to an expensive survey that seeks to represent a national population, it still constitutes a barrier to access information and a time lag for analysis (Google places a limit of 25,000 Static Map image requests per day, which for a country of substantial land mass would create a delay in the analysis of poverty identification). Therefore, an additional source of data to provide geographical feature mapping from Open Street Map for Bangladesh is also adopted. As a crowd-sourced community, Open Street Map (OSM) provides a free alternative to Google Static Image that offers geo-mapping of physical elements in the national landscape of the country of choice, ranging from natural to man-made features in the real world. The features extracted from OSM data might be able to indicate the levels of the wealth distribution across the country, as a stand-alone feature space, or in combination with the nightlight luminosity data to increase the predictive power of the models constructed from these resources.

5

Methodology



Summary

- 5.1 Target Identification, 19
- 5.2 Feature Space Engineering, 19
- 5.3 Feature extraction with Deep Learning, 23
- 5.4 Machine Learning Model Design, 30
- 5.5 Model Evaluation, 31

5.1 Target Identification

The research is an endeavor to discover answers to the following questions:

1. In the face of survey data scarcity, can other publicly available data sources serve as an alternative pathway to accurately measure and map poverty and wealth distribution?
2. Which data sources or permutations of their combinations can be best utilized as predictor for variation in economic outcomes?

To address these challenges, the **Wealth Index** of the DHS survey was employed as the target variable that represents and measures poverty and welfare levels. This metrics has been utilized as a frequent indicator of household-level wealth in country-level surveys and is constituted by reducing the high dimensionality of household asset data through principal components analysis (Rutstein, 1999).



Wealth Index: DHS construction method

1

1. "A subset of indicators common to urban and rural areas is used to create wealth scores for households in both areas. Categorical variables are transformed into separate dichotomous (0-1) indicators. These indicators and those that are continuous are then examined using a principal components analysis to produce a common factor score for each household" (National Institute of Population Research and Training (NIPORT) and Mitra and Associates of Dhaka, 2014, p.16)
2. "Separate factor scores are produced for households in urban and rural areas using area-specific indicators" (National Institute of Population Research and Training (NIPORT) and Mitra and Associates of Dhaka, 2014, p.16)
3. "The third step combines the separate area-specific factor scores to produce a nationally applicable combined wealth index by adjusting area-specific scores through a regression on the common factor scores" (National Institute of Population Research and Training (NIPORT) and Mitra and Associates of Dhaka, 2014, p.16)

According to DHS, this method of formulating the wealth index is able to adapt and accurately convey the socioeconomic status of both urban and rural areas. The households interviewed in the DHS survey were also chosen to be representative of the overall population of Bangladesh, and therefore, their economic statistics can present a balanced picture of the wealth distribution within the country.

5.2 Feature Space Engineering

There are three main features, together with different permutations of their combinations, that serve as potential predictors for the target variable:

i Features and their formats

1

- Nighttime luminosity images (.tiff format);
- Daytime satellite images (.jpeg format);
- Open Street Map geo-spatial data (.shapefile format);

As with real world data, they are all notoriously unstructured and noisy with varying degree of difficulty for handling. The bulk of the research and the main challenge rest in the transformation of these diverse and thorny data types into inputs that can be interpreted by the models.

5.2.1 Nighttime Luminosity as a stand-alone predictor

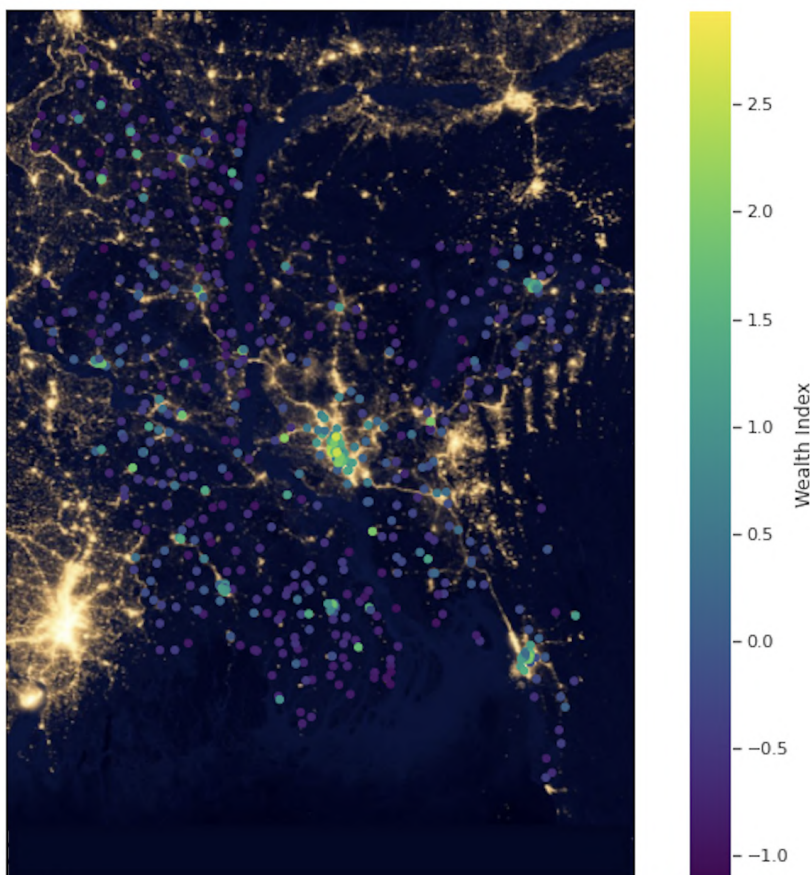


Figure 5.1: Nighttime light intensity correlation with average wealth index for 600 Bangladeshi household clusters.

Theoretically, nightlight can be utilized as a proxy for wealth, to indicate wealth distribution within a country, with higher level of luminescence corresponding to more electricity usage, infrastructure investment and thus economic activity and affluence.

Luminosity features for each of the household clusters location, including level, maximum, minimum, mean, median, standard deviation of luminosity will be extracted and trained on the wealth level of these clusters to learn of their relationship. Different Machine Learning models with K-fold Cross Validation that fit average wealth as a function of these nightlight

features will be utilized to estimate to what extent nighttime luminosity can explain the variation in wealth levels within these clusters of households.

As we can observe from the scatter plot of the households' location and their asset-based wealth index across Bangladesh layered on top of nighttime light emission, regions with more intensity of light seem to have clusters of more wealthy households, particularly in the capital city of Dhaka at the center of the map. At the opposite end of the spectrum, households with low average wealth seem to congregate in regions of darkness.

5.2.2 Daytime Satellite Images as a stand-alone predictor

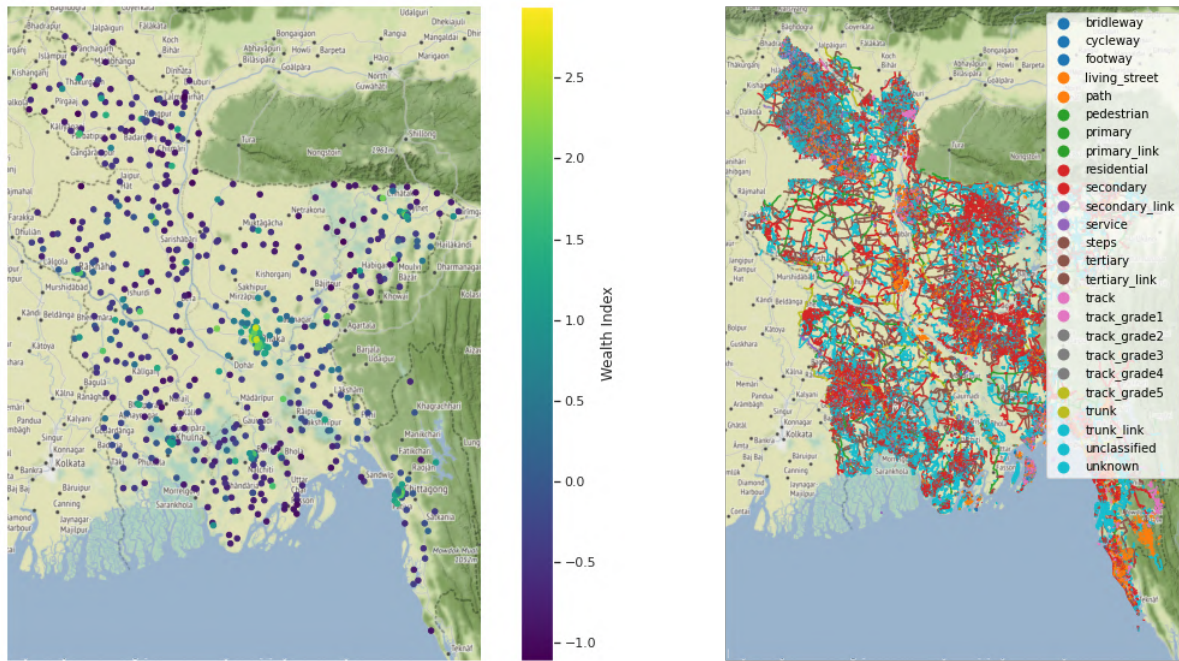
Similar to the nighttime luminosity model, simple characteristics can also be extracted from the daytime satellite images such as the max, min, mean, median and standard deviation of the flattened RGB color channels of the images for locations close to the household clusters. These features can then be used to fit a model of wealth as a function of these basic metrics to identify how much variation of socioeconomic well-being can be explained. This will provide another helpful baseline to compare other methods that require further feature engineering of the daytime images.

The next step in this procedure is to extract deep features and insights from the daytime images using Convolutional Neural Networks (CNN), taking advantage of the existing model pretrained on the ImageNet dataset. This pretrained model presumably have already learnt to detect distinct features such as lines, curves and shapes. The last layer of the model will be frozen and retrained on the Google daytime images data to theoretically capture the features indicative of socioeconomic status such as roads, rivers, land usage, or roofing materials. These features will subsequently be tabulated and utilized as predictors for wealth level within the clusters to discern whether they can improve upon the baseline of basic features computed previously.

5.2.3 OSM data as a stand-alone predictor

Apart from Google Image data, OSM also contains powerful geo-mapping of specific physical features that might also provide an useful approximation of wealth levels in different location. As we expect rural and urban areas to differ in terms of surrounding features, a 5km radius buffer is affixed to the rural clusters and a 2km one for the urban clusters to capture surrounding elements. These buffers have the added benefit of accounting for the random noise offset in the locational data of DHS cluster to protect the confidentiality of the residents.

Features relevant to the prediction task will be engineered to be used for training, including: **road features** (types of roads; length of each road types within the cluster buffer zones; distance of clusters' centroids to the nearest road, etc.), **buildings** (types of buildings; area



(a) Distribution of clusters based on wealth in Bangladesh.

(b) All mapped roads in Bangladesh OSM data.

Figure 5.2: Geo-mapping of clusters against roads in Bangladesh

of buildings; mean distance to clusters), **land-use** (types of land-use, area size of point-of-interest within cluster buffer, proportion of area compared to buffer area), **points-of-interest** (count of each type of point-of-interests within the buffer zone).

Our hypothesis is that there exists a spatial structure within the OSM data information that is indicative of the level of wealth. A naive conjecture suggests that within an affluent neighborhood, many commercial, retail, residential area and public infrastructures would be present. On the other hand, for poorer neighborhood, low-cost housing structures would be more prevalent. Transforming and combining all these OSM data should, therefore, represent a good approach to engineer features that can approximate the wealth distribution within the country. Furthermore, a fairly accurate model can also help to contribute an understanding of how infrastructure, availability of public services or land usage might be correlated with the disparity of resources and wealth allocation.

From simple visual inspection of the sample geo-mapping of wealth distribution of household clusters and the different road types in Bangladesh above, we can also observe that for household clusters that belong to the upper echelon of wealth levels (light green to yellow end of the spectrum), there are also dense groupings of residential and secondary roads in the surrounding area. These connections will be captured more comprehensively through the models that can inform us of the shape of relationship between these features and the target.

5.2.4 Nighttime Luminosity & Daytime Images

A transfer learning approach will be used with the previous CNN being retrained to predict nighttime luminosity using features of daytime image. Theoretically, these features will be able to add compound effect of possessing the deep characteristics of nighttime images which can then contribute more predictive power for wealth estimation.

5.2.5 Nighttime Luminosity & OSM data

This combination brings together two data sources that are readily available and can be collected quickly without any cost incurred. Each data have been engineered in previous steps and the relevant features extracted for analysis. They only need to be appended to each other to work as the combined feature space.

5.3 Feature extraction with Deep Learning

To extract the deep features of the satellite images, a Convolutional Neural Network architecture (CNN) is implemented. CNN is a type of neural network that can be viewed as constituted of two main components: a front-end component that executes the feature learning that pre-processes the image data, and a back-end that carries out the classification task, i.e using the information of the input of the frontend to categorize the wealth level of the household clusters for our use case (Ferlitsch, 2020).

The usage of Neural Networks will be discussed in further details in the next section. We will first focus on the front-end of the feature extraction step, namely the convolutional layers for the task of image processing.

5.3.1 Satellite Image Processing for Computer Vision

To transform the satellite images into inputs that the computer model can understand and process, they need to be converted into numerical format. The key to implement image processing is recognizing that a normal RGB image is typically comprised of three independent channels that correspond to varying levels of intensity in terms of red, green and blue light, compatible with the color receptors in human. Each layer in these images is further represented by pixel values ranging from 0 (none of the color appears) to 255

(highest level of color is present in the pixel). Therefore, these satellite images can be processed in accordance to their pixel values as input into the models.

However, the main obstacle for image processing lie in the size of the input. For our use case, for example, each satellite image has 400x400 pixels across 3 RGB channels. Taking just one image, we have 400x400x3 equalling 480,000 pixels, requiring an input vector of similar length with 480,000 elements. In addition, we have in total 414,757 images to process, resulting in a combined input vector of over 160 billion elements. Assuming that the input layer of a traditional Deep Neural Network for this task has 4096 input nodes to learn from the data, the number of weights that we would need to update would fall into the trillions at just the input layer - an impossible task for our normal computer hardware to process.

5.3.2 The Convolutional Layer

Training and processing real-world image data, thus, was infeasible using neural network before the implementation of convolutional layers. As stated previous, it is a preprocessing front-end of a neural network, helping to reduce the high dimensionality of a pixel-based image to a more agreeable lower dimensionality feature-based representation of that image. The latter can then be loaded as the input vector for a Deep Neural Networks for further processing and learning.

Translational Variance

Another issue with processing image through its pixel representation is translational variance - image recognition through different perspective. Pixels are very dependant on their position in the image to project the correct image. However, for computer vision, this limit the ability for learning as any model that endeavors to train on these pixel representation would overfit on the position of the objects within the image (Ferlitsch, 2020). It would be harder for the machine to recognize a cat that jumps, and rolls around if it only has the ability to train on images of that cat lying on a specific location of the image. Training on every single position that a cat would have in an image is also an equally impossible task to undertake.

Feature Detection

For these pixel-based images that are complex to process, the more sensible approach is not to to implement recognition through pixel positions, but by identifying and detecting features, shapes and objects present in the image. The methodology simulates how human recognizes objects in the dark: from far away, we can only recognize the shapes of an object (whether it has round or sharp edges); getting closer and we can start to have an idea of its profile and shapes; and finally, our eyes identifies what that object is, thanks to the logical conclusion followed through the previously recognized features.

Similarly, a convolutional layer detects features in an image by employing a set of filters - n -by- m size matrices with values that are tailored to identify certain features such as edges and shapes present within an image. These filters perform a sliding window technique across the different pixels of the image, picking up pixel values that signal the likely presence of a feature in that location and record those in a feature map. This map, thus, would hold the information on all the features available within that image. Instead of having to process the entirety of all the pixels of an image, we now only need to process the feature map as the input sequence, which is a significant reduction in size compared to the former.



Figure 5.3: Feature map created by sliding window of filters across the satellite image

In the convolutional layer, we can denote the size of the filters as well as the number of filters and how far each move of these filters is as they slide across the image. By choosing a good size filter (usually 3×3 or 5×5), we can help to transform the image into its lower dimension representation. However, each filter that moves across an image produces its own feature map that capture a certain aspect of the objects within that image. That means that with bigger filter and higher stride, we can have less data, but the more filters with which we can learn from the image, the more feature maps (and thus more data) we would still have to process. This additional problem is resolved by the usage of the pooling layers.

Pooling

Multiple feature maps leads to more data to handle. However, this can be managed by pooling, which is similar to the process of down-sampling. With this strategy, we can set a pooling layer and its stride across the individual feature map, much like how we employ a filter on the original image. For a 2×2 pool size with a stride of 2, we can reduce about 75 percent of the pixel data, while still retaining the detected features along with

their intertwined spatial relationships (Ferlitsch, 2020). Feature information captured in the feature maps are condensed within the pooling layers in two different types of operations that need to be specified:

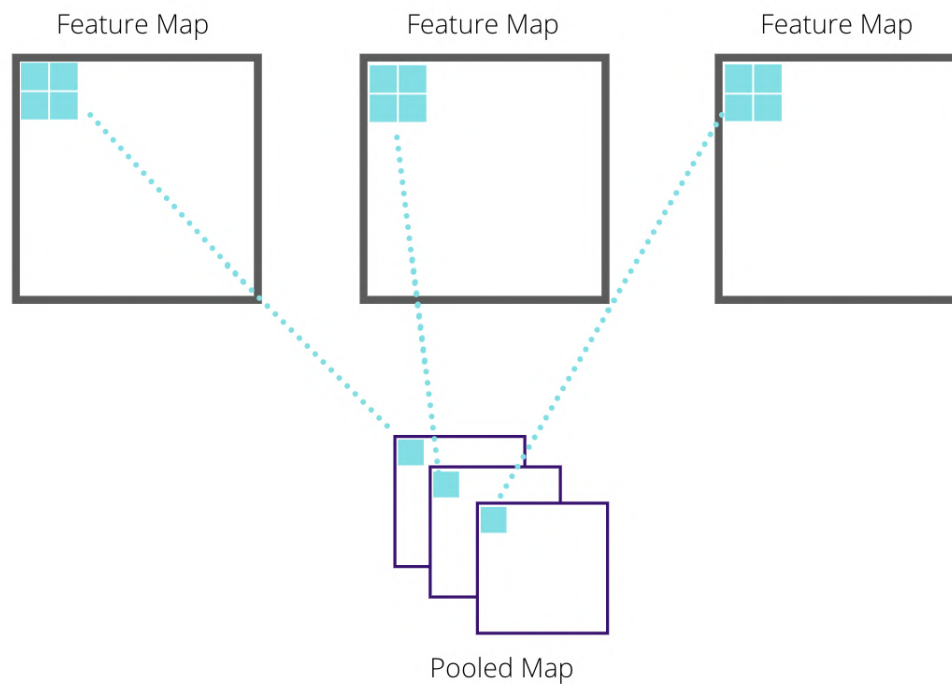


Figure 5.4: Pooled map created from feature maps

- **Max Pooling:** extracts the highest values for each patch of the pooling stride on the individual feature map;
- **Average Pooling:** extracts the average of all values for each path of the pooling stride on the individual feature map.

The max pooling operation rejects most of the information in favor of obtaining only the most important features in the input space such as edges while average pooling, in comparison, takes all information from all the features present in the input. Pictorially, the max pooled layer will theoretically carry high contrast images, whereas average pooled layer will include smoother and less grainy images.

The pooling layers also help to further reduce the sensitivity of the feature maps on the location of these features in the input. This is achieved in the process of down-sampling with pooling: the higher resolution input is reduced to a lower resolution representation, and thus, refined features in the upstream process is transformed into a less detailed and more abstract version that still retains the structural components. The lower representation of the features can therefore rely less on the position of the feature in the input as it no longer has to cast detailed and granular information on the feature. Essentially, pooling is a building block within the architecture to summarize the features detected in the feature maps.

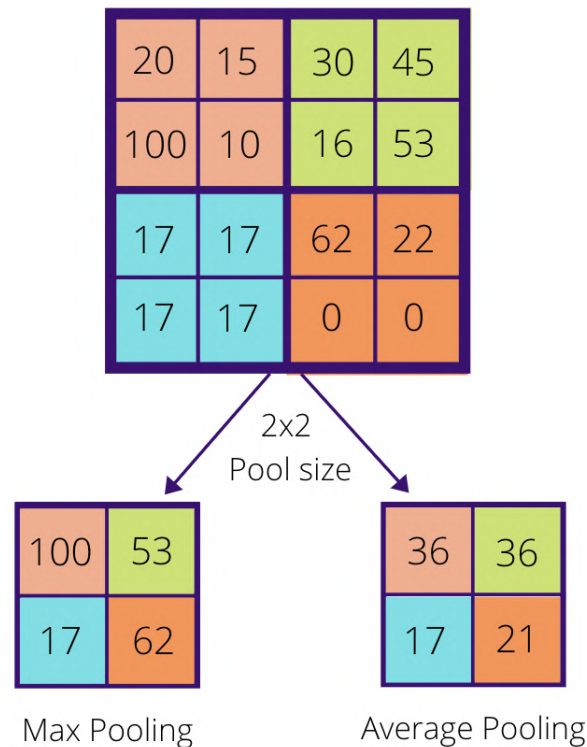


Figure 5.5: Difference between max pooling (taking maximum value) and average pooling (taking the average of all values)

Flattening

The final step of the front-end process of the Convolutional layer is flattening the information obtained from the feature map and the pooling operation (in this case 2D matrices of the pooled feature maps) into a 1D vector that can be used as input for the Deep Neural Network for training and learning. Each pooled map is appended to one another sequentially to preserve the spatial relationship present in the pooled information, ultimately forming the long 1D vector for the input space.

5.3.3 Deep Neural Network

The back-end of the deep learning feature extraction is a Neural Network, a statistical mechanism of approximating the output based on a large number of inputs. Just as other statistical methods that construct probability distribution to produce inference from data, Neural Network trains a model on a vast quantity of data points, without the need for explicitly programming procedural instructions. The model is built to learn the patterns in the distribution of the data, which can subsequently be exploited to draw conclusions about samples unseen by the training process. The large number of data points in the sampling distribution used in the training is assumed to be representative of the population, and thus the patterns within the data can be utilized for capturing the relationships between

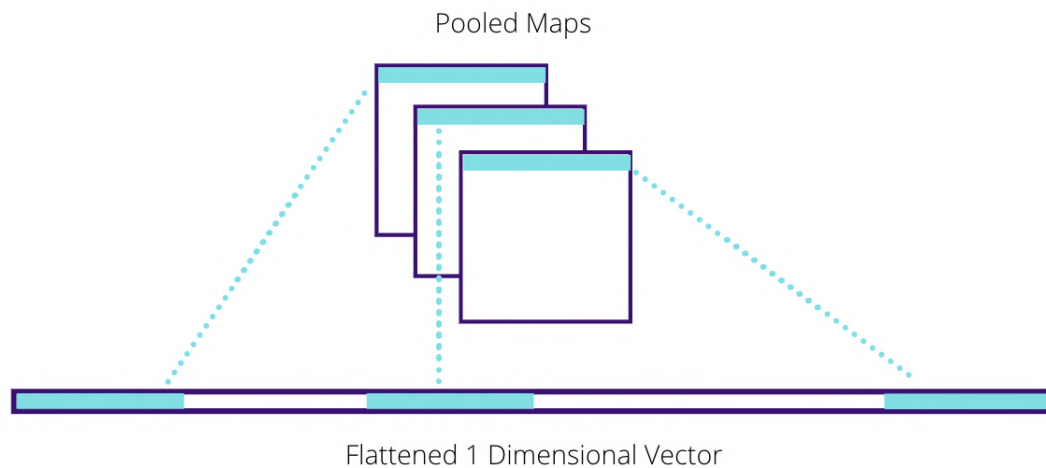


Figure 5.6: Flattened pooled maps for model input

variables in the population distribution.

The model is constructed with three main components:

- **Input layer:** the input data needs to be converted into numerical form and fed into the input layer. These numerical forms can either be vectors, matrices or tensors. The input layer contains a number of nodes (neurons) that connect to other nodes in subsequent layers of the network to simulate the interconnecting neural circuits of the human brain. Each node takes in the input data, and feed it to the next layer through a channel;
- **Hidden layers:** sitting between the root and the terminal nodes of the network are the hidden layers, which receive data from the input layers as parameters and perform a nonlinear transformations of these inputs with specified weights and biases before directing them through the activation function, which determines whether a node is activated or not, for the output layer;
- **Output layer:** the data finally enters the output layers, and through the activation function, a final output result is calculated. This output is compared with the ground truth data, and the loss is determined to identify the model error and how far off it is from the target. The optimizer, then, takes this information and adjust the weights and biases of the model and continues to feed them back through the network in an iterative process called back-propagation until the training concludes or the intended goal for the chosen accuracy metrics is achieved.

For our use case, the Deep Neural Network is particularly useful since it can untangle the non-linearity of the high dimensionality of the input space (input sequence generated from our satellite images) to the targeted output (wealth distribution of Bangladesh household clusters). We hypothesize that the geographical and landscape features of Bangladesh might be able to explain some of the variation in the wealth of Bangladesh households. However, it is unlikely to be a simple linear relationship that can be easily captured with a simple model. The implemented convolutional layers assist in the extraction of features

unique to the landscape of Bangladesh, and the Deep Neural Network at the back-end can help to find the boundaries that divide the input space of these features. Therefore, the model can learn the different linear relationships that each specific feature within certain segments of the input space has with the targeted output. Simply put, there can be many features that the model can identify such as land usage, buildings, roads, infrastructure, roofing materials of housing, etc. thanks to the convolutional layers. Taken as a whole, the feature space will have a non-linear relationship with the average wealth index of Bangladesh households. However, individually, each of these features might have a pronounced linear connection with wealth and poverty level and the Deep Neural Network can distinguish both the relationships and the segmentation among these different features.

5.3.4 Transfer Learning

In the Deep Learning discipline, Transfer Learning has become a popular approach to help speed up the training process of models by re-purposing existing weights and biases of models that have been developed for a different task as the starting point for a new model. The pre-trained model aided the learning process of tasks that are similar to the original use case by reducing the computing time and resources needed for the training process. This is particularly helpful for cases in Computer Vision, such as our task for satellite image processing, when for the first hidden layers of the Neural Networks, the model is only learning the rudimentary representation of the features such as edges and shapes. By reusing an existing models that have extensively trained on detecting these features, and attaching our task as the last layers, we can significantly decrease the time needed for accomplishing the task at hand as well as improving the performance of our model. Our research, therefore, employs the VGG16 model as the base model for transfer learning on our endeavor.

5.3.5 VGG16

VGG16 is a Convolutional Neural Network model developed by the Visual Geometry Group at Oxford. Proposed by K. Simonyan and A. Zisserman in their paper "Very Deep Convolutional Networks for Large-Scale Image Recognition"(Simonyan and Zisserman, 2014). The model was designed to compete in the ILSVRC-2014 contest for image recognition. It achieved 92.7 percent accuracy in ImageNet, a famous dataset of over 14 million images classified into 1000 classes. The model consists of stacks of convolutional layers, with max pooling layer to reduce the size of the feature maps. The design of the model directs the learning of the coarse features of the images in the prior layers while the subsequent ones could learn the fine-grain features by increasing the number of filters with max-pooling wedged in-between to ensure that the data and parameters transferred between these layers does not grow exponentially. The Deep Neural Network component has two hidden layers, each with 4096 nodes and the final output layer has 1000 nodes for the 1000 classes.

5.3.6 Convolutional Neural Network Architecture

With all these elements in place, we can now construct the full graphical model of our own Convolutional Neural Network Architecture. With the goal to use the features extracted from the model for poverty mapping, our model would be fine-tuned from the VGG16 to classify three levels of nighttime luminosity: low, medium and high intensity.

The weights of the VGG16 model pre-trained on ImageNet is first loaded to get the prior rudimentary input features from the satellite images. Subsequently, a second full Convolutional Neural Network is constructed. There are 2 hidden layers, each with 4096 nodes to learn deeper features from the data already extracted by VGG16, with 50 percent Dropout rate to prevent the model from overfitting on the training data. The final output layer has 3 nodes for the 3 classes of nighttime luminosity with a softmax activation. The Adam optimizer with initialized learning rate at $1e-2$, with batch size of 100 and the number of epochs at 10 are employed to fine-tune the network.

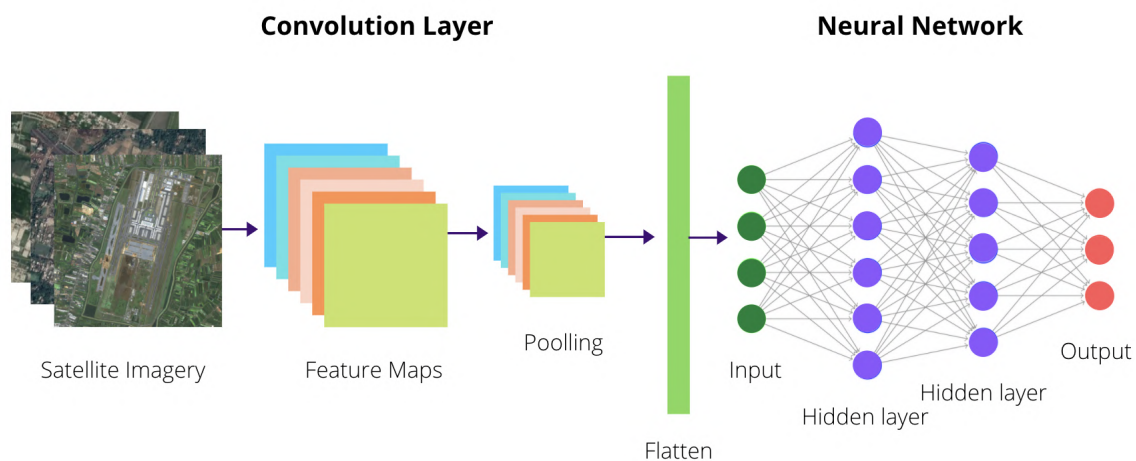


Figure 5.7: Graphical model of Convolutional Neural Network Architecture

70 percent of the satellite images is utilized for training the model and the rest is reserved for validation. The images are processed and fed sequentially first into the pre-trained VGG16 model, then the new CNN model and validated for accuracy.

5.4 Machine Learning Model Design

With the target variable identified and the feature space engineered from the data sources, models can now be built to estimate wealth distribution and identify areas of poverty:

- **Ridge Regression:** since we can expect that some of the features are more important than others in predicting the wealth level of a household cluster, an OLS model with a huge number of parameters might not be ideal as it usually has low bias and high

variance, meaning our model might overfit and have less ability to generalize. Ridge Regression, by introducing a small amount of bias (penalty term) into how the model might fit the data, can help to decrease variance, creating a model that perform slightly worse in training but can provide better long-term prediction;

- **Lasso Regression:** in the same vein of reducing the complexity of the models with a large number of parameters, Lasso helps to shrink coefficients of unimportant variables to zero, excluding them from the equation, resulting in a simpler final model that is also easier to interpret, which helps to pinpoint which variables have more power in explaining the variation of wealth level in the data;
- **ElasticNet:** a combination of Ridge and Lasso Regressions, ElasticNet groups and shrinks parameters associated with correlated variables and leave them in or eliminates them altogether based on their importance. This helps to simplify the model, especially when we suspect that there are correlations between the parameters;
- **Random Forest:** utilizing a wisdom of crowds approach, a Random Forest model is comprised of a large number of low-correlated individual decision trees that operate as an ensemble committee. Each decision tree in the random forest delivers a class prediction and the class with the most votes from all the trees becomes the model's final output. The ensemble learning helps to maximize the accuracy of the model as the decision trees shield each other from error at the aggregate level;
- **XGBoost:** similar to Random Forest, XGBoost is an ensemble decision tree-based algorithm that is considered best-in-class for tabular and structured data. Building up from Gradient Boosting model which produces prediction by building a series of weaker models to learn and improve upon the mistakes of one another, XGBoost marries software and hardware optimization to generate accurate output with less resources and time;

5.5 Model Evaluation

As our research seeks to understand how much of the variation in wealth distribution of Bangladesh household clusters can be explained by the features extracted from the data sources, R-square has been chosen as the main evaluation metrics with which to measure the success of the fitted models.

6

Results and Analysis



Summary

- 6.1 Nighttime Luminosity Model, 33
- 6.2 Daytime Satellite Image Model, 34
- 6.3 OSM Model, 35
- 6.4 Nighttime Luminosity & Daytime Satellite Images Model, 37
- 6.5 Nighttime Luminosity & OSM Model, 38
- 6.6 Poverty map for Bangladeshi household clusters, 38

6.1 Nighttime Luminosity Model

After extracting relevant features in light luminosity from the nighttime satellite view data, there appears to be a substantial correlation between the features and the target of wealth index level in household clusters, which further reinforces the argument for nighttime light intensity as a proxy for estimating wealth and socio-economic wellbeing. The data points are clusters at the lower end of the spectrum of light emission and wealth level as expected.

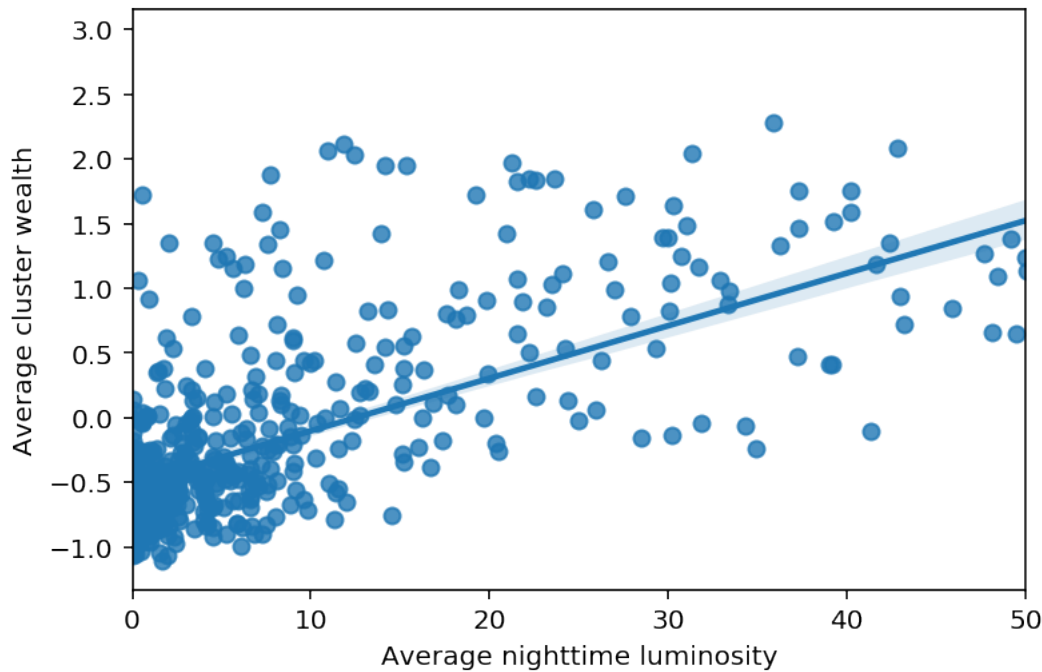


Figure 6.1: Average Nighttime Luminosity Correlation with Wealth Distribution in Household Clusters

To gauge further the depth of this relationship, five separate models were constructed including Ridge Regression, Lasso Regression, ElasticNet, Random Forest Regressor, and XGBoost using 10-fold cross validation to evaluate the predictive powers of these models. The results are relatively close to one another with the XGBoost model being able to explain the most variation in socio-economic status at 66.28 percent R-square score.

Method	R-squared
Ridge Regression	60.02%
Lasso	59.69%
ElasticNet	58.68%
Random Forest	59.32%
XGBoost	66.28%

Table 6.1: Model R-squared for Nighttime Luminosity Predictor

6.2 Daytime Satellite Image Model

6.2.1 Basic Features

Basic features such as the max, min, mean, median, and standard deviations of the RGB pixel color information are first extracted from the daytime satellite images obtained from Google. These features are subsequently fitted to the Machine Learning models to estimate the wealth level of the different household clusters. The results demonstrate that the rudimentary information of the satellite images is only able to explain 36-47.87 percent of the variation in wealth levels, depending on the algorithm used. This outcome is expected as the simple luminosity of daytime satellite image does not have the benefit of signifying the usage of electricity of the population, and thus would not be able to serve as an equally reliable proxy for socio-economic wellbeing estimation.

Method	R-squared
Ridge Regression	40.28%
Lasso	36.53%
ElasticNet	38.51%
Random Forest	39.97%
XGBoost	47.87%

Table 6.2: Model R-squared for Basic Daytime Image Features

6.2.2 Deep Features

To improve upon the basic features of daytime satellite images, we extract their deep features with the Convolutional Neural Network architecture, using pre-trained weights of VGG16 model. The architecture can acquire the specific geographic feature representations surrounding the household clusters within all the collected satellite images, and create tabular data as the input for the model fitting.

To provide sanity check for the model, we can utilize Principal Component Analysis to reduce the dimensionality of the feature representations through the daytime satellite images. PCA1, which accounts for the majority of the total variance of the feature data, can be observed in connection with the average wealth index of household clusters as above. The bulk of the data concentrates on the lower end of the wealth level, as expected with the data points on poor households constituting the predominant data points.

The models fitted with the deep features extracted from the daytime satellite images improve considerably from those of basic features, explaining up to 60.33 percent of the overall variance in average wealth index level of the households by employing the XGBoost method. This result is comparable to the model using nighttime luminosity as a stand-alone predictor. Each models utilizing different types of features input seem to be able to

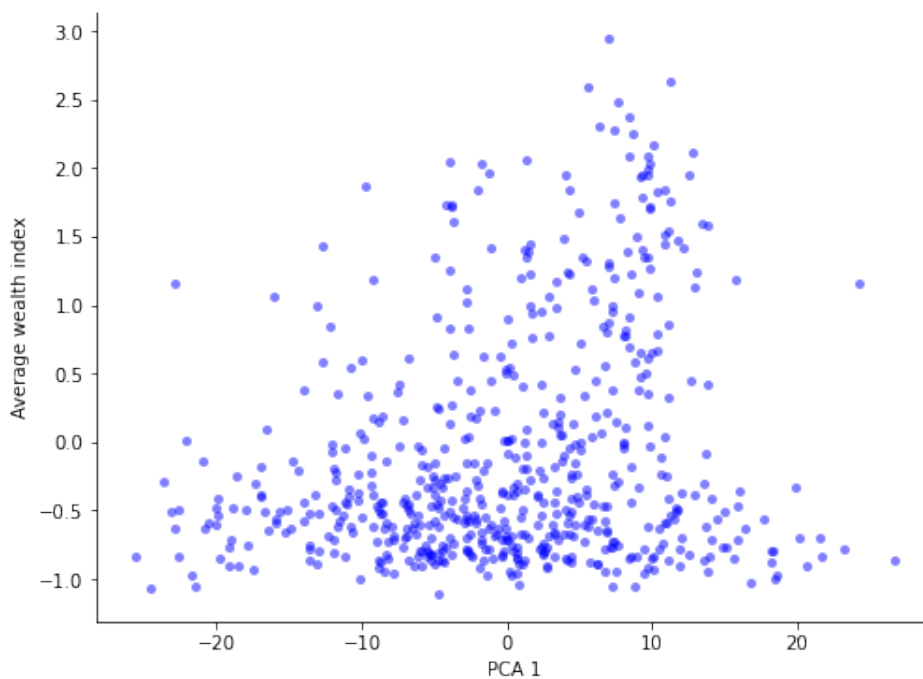


Figure 6.2: Average Wealth Index against PCA

successfully measure their own strength of the relationship with the target.

Method	R-squared
Ridge Regression	51.7%
Lasso	42.9%
ElasticNet	46.3%
Random Forest	42.2%
XGBoost	60.33%

Table 6.3: Model R-squared for Deep Daytime Image Features

6.3 OSM Model

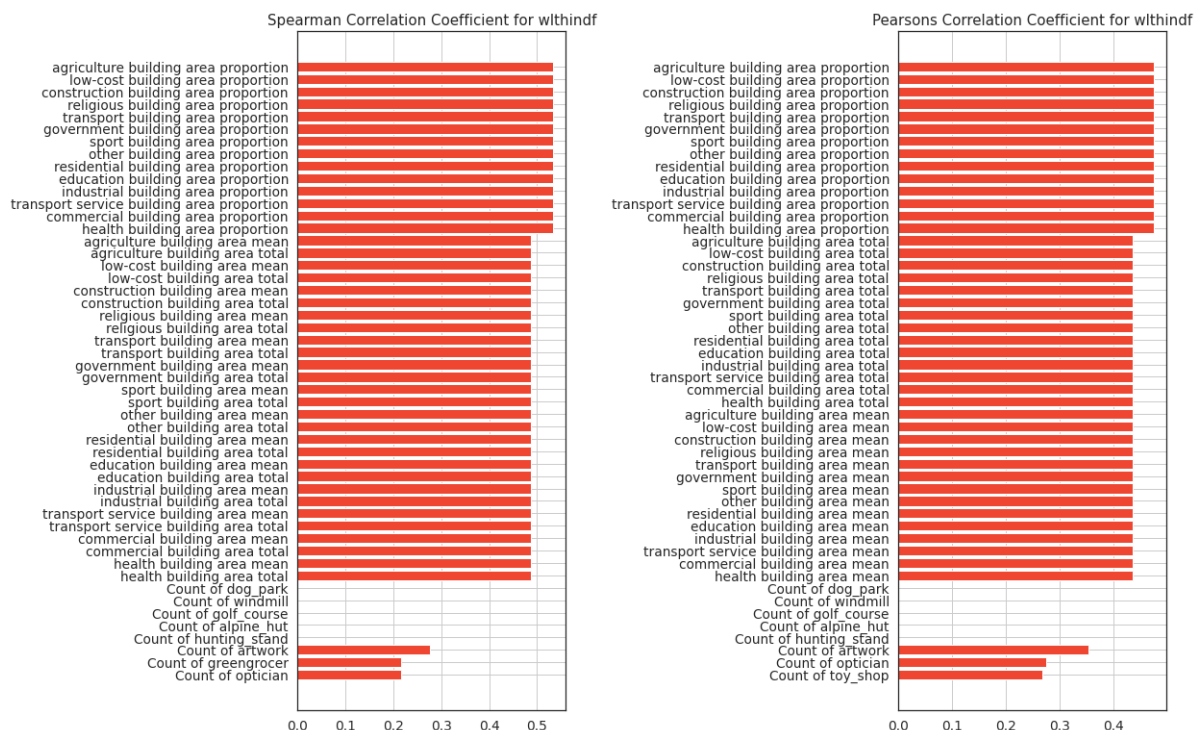
Geographic features on the landscape of Bangladesh were extracted from the geometries and points information of OSM data. As mentioned in the Data Sources and Methodology section, all of these features need to be manually engineered to be meaningful for use of analysis. Over 260 unique features surrounding the buffer zones created around the individual household cluster were procured, and categorized into four main groups:

- **Road Features:** within the buffer zones of urban and rural clusters, information on the number of different roads, their varying types (pedestrian, service, primary, secondary, tertiary, residential, track roads, etc.) the length of each type as well as the distance from the center of the cluster to the nearest road were collected;

- **Buildings Features:** similarly, the particulars of different types of buildings (residential, commercial, governmental, low-cost buildings, etc.) within the border of the buffer zones could signal the level of development and wealth of an area;
- **Land Use Features:** details on how the land was used by the Bangladeshi population for parks, forest, commerce, the military, as well as for industrial, residential, recreational and other purposes could also shed light on their relationship with socio-economic well-being of an area;
- **Point-of-interest Features:** specific point locations that people may find useful such as hospitals, schools, supermarkets, public attractions, etc. is also a telling feature that can pinpoint the infrastructure, services investment and affluence of a local area.

Open Street Map also provides a more comprehensive collection of other landscape features including places of worship, natural features, waterways, bodies of water, etc. The four categories above have been chosen as the main feature representations for the OSM data as they are predominantly human-made and thus logically will entail certain relationship with the population in their surroundings.

As visual sanity check, Spearman and Pearson's rank correlation coefficients are implemented to evaluate the potential relationship between the features and the target variable. As can be observed from the graph, building features appear to hold the most correlation with average wealth index, with information on the total area proportion of buildings related to low-cost structure, agriculture, transportation, governmental, residential and commercial activities taking the lead.



(a) Spearman Correlation between Feature and Target. (b) Pearson Correlation between Feature and Target.

Figure 6.3: Correlation Ranking of Features and Target

Similar to the range of the basic features of daytime satellite images, the OSM data as a stand-alone feature space is able to explain about 43-56.99 percent of the total variance in socio-economic wellbeing. This result is compatible with the visual check from Spearman and Pearson's correlation coefficients. When acting as complementary features, it can help to boost further the signal of other indicators as evident in the subsequent Nighttime Luminosity and OSM hybrid model.

Method	R-squared
Ridge Regression	43.33%
Lasso	44.02%
ElasticNet	44.48%
Random Forest	48.67%
XGBoost	56.99%

Table 6.4: Model R-squared for OSM Predictor

6.4 Nighttime Luminosity & Daytime Satellite Images Model

The hybrid model using both nighttime luminosity and daytime satellite images' deep features achieves the best performance out of all permutations of features combination at 72 percent for wealth index R-squared score. This is not surprising since this transfer learning approach is currently the state-of-the-art at the moment for poverty and wealth level prediction. This result is also consistent with the results of Neal Jean et al. and other research for African countries (Jean et al., 2016). However, it is notable that the data collection process for this model poses the most complications as in order to acquire the daytime satellite images, we need to rely on the proprietary-software services of Google Maps which induces both a monetary cost and a tricky time-lag.

Method	R-squared
Ridge Regression	61.53%
Lasso	59.69%
ElasticNet	60.73%
Random Forest	59.95%
XGBoost	72.19%

Table 6.5: Model R-squared for Nightlight and Daytime Satellite Images Predictor

6.5 Nighttime Luminosity & OSM Model

When combining features from both nighttime luminosity and engineered features of OSM data, the collection of predictors is able to explain from 65-70 percent of the variance in average wealth level distribution of the household clusters. This performance is comparable to the state-of-the-art transfer learning approach using satellite images. In addition, both the nightlight intensity and OSM data in this hybrid model are open-source and thus do not require any auxiliary cost for acquisition such as the satellite imagery obtained from Google Static Map API. Therefore, in terms of both performance, ease of access and cost, the hybrid model incorporating nightlight and OSM data sources serve as the best for estimating wealth level for the Bangladesh household clusters.

Method	R-squared
Ridge Regression	65.39%
Lasso	65.06%
ElasticNet	64.80%
Random Forest	66.68%
XGBoost	70.21%

Table 6.6: Model R-squared for OSM Predictor

6.6 Poverty map for Bangladeshi household clusters

Finally, a map of poverty estimation constructed by the best model utilizing nighttime light luminosity and OSM data can display the model prediction for existing household cluster in the survey data. With a deployed model, we can obtain a fairly accurate reading of average wealth for any geographical location within the country of Bangladesh, as long as the right coordinates of the location of interest are fed into the model. Therefore, even in places where there are gaps in the survey data, the model can extrapolate and provide estimation to fill in the missing information for unknown areas, helping the process of aid targeting and financial inclusion schemes.

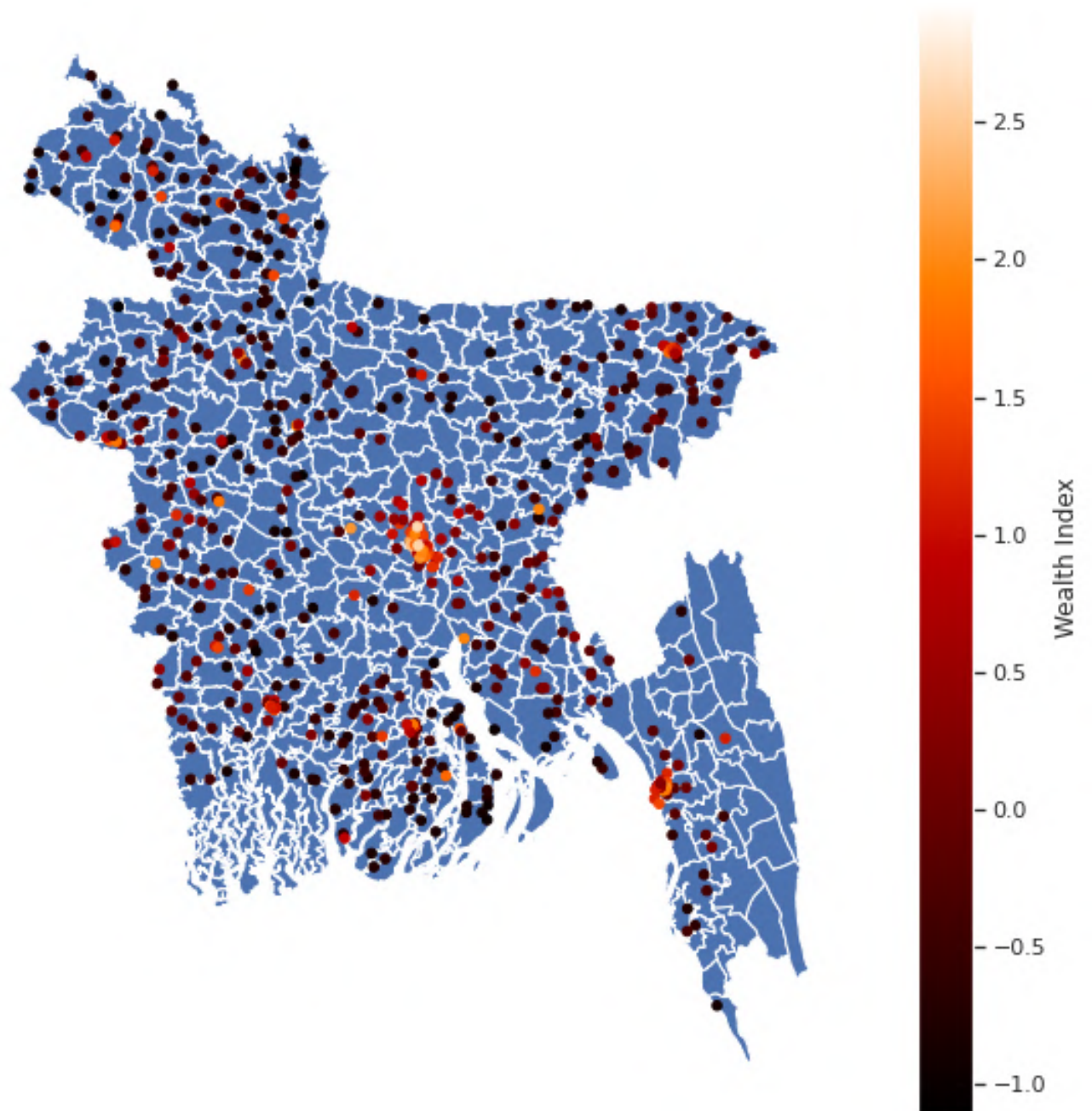


Figure 6.4: Map of average wealth index prediction for household clusters layered with Bangladesh administrative boundaries by best model with nightlight emission and geo-mapping features from OSM

7 Conclusion



Summary

- 7.1 Policy Implication, 41
- 7.2 Conclusion, 43

7.1 Policy Implication

This study further substantiates the usage of open-source data and Machine Learning algorithms in poverty estimation without the need for implementation of costly and complex household surveys. Nighttime luminosity, daytime satellite imagery and crowd-source geo-mapping such as Open Street Map have proven to deliver meaningful features that can offer a fairly accurate explanation of variation in wealth and poverty distribution at household cluster level for Bangladesh.

This methodology, therefore, can be an useful part of the toolkit for researchers, policy makers, and development practitioners in identifying areas most affected by impoverishment for aid and financial inclusion programs. The ability to map out wealth distribution within the country with a certain level of granularity can also assist governments in locating communities and neighborhood that need extra support, devising strategy for local and regional economic growth to improve public engagement and reduce inequality. By employing geo-mapping of landscape attributes in Machine Learning prediction models for poverty, policy makers can also grasp another layer of understanding of the correlations between the availability of public services and infrastructure in a local area and its wealth level, leading to opportunities for a discussion on fair resources distribution scheme.

Aside from poverty mapping, the successful usage of these open data sources might also reveal other patterns in population statistics that can inform other public policy decisions. The thorough assessment of natural as well as artificial landscape features in geo-mapping data and satellite images can convey more insights into the patterns of development, population mobility, the evolution of land use, agricultural practices and infrastructure expansion or degradation. These applications can contribute considerably to the evidence-based approach of policy-making.

However, the employment of these data sources and approach is not without its challenges. It would be amiss not to discuss the shortcomings and areas of improvement that are yet to be achieved, especially in their application for public policy:

- **The lack of variation in nightlight information in rural areas:** Although the scarcity or abundance of nightlight in itself is a signal for predicting welfare distribution, the complete absence or deficiency of variation of nighttime luminosity in rural neighborhoods makes it harder to create a more fine-grained understanding of its relation with poverty in these areas. Low electricity availability in the broad sense creates a sense of uncertainty in determining the variation in welfare level within the poor areas, despite forming a clear segmentation with the higher income zones. We cannot be sure of what information lies in complete darkness. Effectively, the models utilizing nightlight emission would theoretically extrapolate more accurately in urban areas where variation in artificial light is more pronounced and likely. This is why the combination of nightlight information with other types of signals, including satellite images and geo-mapping data, would form a better feature space for estimating welfare. However, choosing the right signals and how to engineer them to obtain the optimal efficacy is a challenging pursuit that requires continuous experimentation and

integration of resources.

- **Combining other data sources for feature engineering:** Besides the three main data sources employed for this research, other features, especially those available exclusively or more easily for governments and policy makers, can illuminate further on the gap of equality. Recently mobile and internet usage and traffic have been identified as another useful source of information that can communicate population characteristics. Access to mobile and internet usage, particularly in developing countries, can be a compelling proxy for socio-economic status. Similarly, remote-sensing data, which can provide a highly refined degree of granularity in the detection and monitoring of physical attributes of a region through the measurement of its ground-emitted radiation, could convey a deeper level of connection with electricity usage and thus, welfare estimation. A research implemented by the World Bank, for example, has shown that purple gradients in pictorial map of remote-sensing data can be interpreted as correlated to dark areas with low electrification whereas bright yellow zones are related to intense artificial light usage (Heitmann and Bur, 2019). Albeit without its own limitations, the method has been adopted for a number of application including agricultural monitoring, natural resource deposit exploration, environmental and climate change study and undoubtedly could prove valuable in the field of poverty prediction;
- **The limitation of geo-mapping features:** Geospatial mapping data on man-made structure such as roads and buildings do not come with ease. These artificial elements and infrastructure takes time to materialize and thus their information is hardly feasible to be updated regularly. There is logically a certain imprecision in the signals of the features extracted from this type of data. However, it is still an acceptable source of direction for ascertain a general understanding of income distribution, given a threshold, for better targeting of aid and development program outreach.
- **The relevance of survey data:** The experience of feature engineering and data processing of satellite images for Bangladesh shows that there are signals specific to the country of research that is not easily replicated for studies on other nations. Therefore, despite the fairly accurate estimation of wealth distribution based on just open-source and freely available data, ground truth information from household surveys and other official data sources is still a necessary component to build and evaluate Machine Learning models for welfare segmentation. Finding the right balance between the frequency of such surveys and how to incorporate open-sourced data to alleviate the cost and bottlenecks of a formal and bureaucratic procedure is also a challenge worthy of further exploration;
- **Causal analysis of features and target:** The methodology could also gain more credence through a more robust analysis to untangle the different causal pathways that the different extracted features could affect wealth distribution. This assessment could lead to a more refined understanding of poverty and the different factors connecting to its variation, and thus could inform relevant decision makers of the direction for potential campaigns and programs to relieve and prevent one of the great challenges facing humanity.

Finally, big data, as useful as it can be as a tool to represent the population and tell a good story of statistical learning, is not people. The human aspect in the science of data cannot be ignored and presents just as much significance for the understanding of poverty and how to counter it as a collection of granular data (Blumenstock, 2018). Surveys and interviews to understand the struggles of people facing impoverishment in different parts of the world, and how they can be helped to overcome their hardships may impart valuable insights to the domain of knowledge for narrowing the gap between mankind and global poverty eradication.

7.2 Conclusion

The research has demonstrated that it is feasible to estimate poverty and welfare level, specifically for Bangladeshi households, through the use of open-source data such as nightlight intensity, daytime satellite imagery and geo-mapping of landscape features. While the model created with daytime satellite images and nighttime luminosity information still performs the best in capturing the variation in average asset-based wealth of household clusters, the combination of nighttime data and geo-mapping from Open Street Map is also able to accomplish comparable result with much greater ease of access and cost effectiveness.

Although further development and expansion of the methodology are essential to improve its efficacy and reliability, the successes of these models constructed from publicly available data and Machine Learning algorithms have illustrated that they can serve as a viable alternative for the task of poverty measurement, facing scarcity and difficulty of complex household survey data.

This strategy for mapping poverty can provide opportunities for researchers, policy makers and development agencies to further enhance population targeting for financial inclusion programs, locating areas with vulnerable communities that should receive more support and engagement, and better understand the connection of the availability of public services, infrastructure and resources in welfare distribution, with the ultimate goal of inducing poverty alleviation and eradication.

Bibliography

- Babenko, B., Hersh, J., Newhouse, D., Ramakrishnan, A., & Swartz, T. (2017). *Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, With an Application in Mexico*. <https://arxiv.org/abs/1711.06323>
- Blumenstock, J. (2018). Don't forget people in the use of big data for development. *Nature*, 170-172, 561.
- Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting Poverty and Wealth from Mobile Phone Metadata. *Science*, 350, 1073–1076.
- Elvidge, C., Sutton, P., Ghosh, T., Tuttle, B., Baugh, K., Bhaduri, B., & Bright, E. (2009). A global poverty map derived from satellite data. *Computers Geosciences*, 35, 1652–1660.
- Engstrom, R., Hersh, J., & Newhouse, D. (2017). Poverty from Space Using High-Resolution Satellite Imagery for Estimating Economic Well-Being. *World Bank Group*.
- Ferlitsch, A. (2020). *Deep Learning Design Patterns Primer*. Google Cloud AI.
- Geofabrik GmbH and OpenStreetMap. (n.d.). *Geofabrik Server for Bangladesh OSM Data*. <https://download.geofabrik.de/asia/bangladesh.html>
- Ghosh, T., Anderson, S., Elvidge, C., & Sutton, P. (2013). Using Nighttime Satellite Imagery as a Proxy Measure of Human Well-Being. *Sustainability*, 5, 4988–5019.
- Heitmann, S., & Bur, S. (2019). Poverty Estimation with Satellite Imagery at Neighborhood Levels: Results and Lessons for Financial Inclusion from Ghana and Uganda. *International Finance Corporation*.
- Hill, R., & Genoni, M. (2019). Bangladesh Poverty Assessment: Assessing progress from 2010 to 2016/17. *World Bank Group*.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining Satellite Imagery and Machine Learning to Predict Poverty. *Science*, 353, 790–794.
- NASA/NOAA Suomi NPP. (2013). *Earth At Night 2012*. <https://svs.gsfc.nasa.gov/30028>
- National Centers for Environmental Information. (2013). *Nighttime Light Luminosity Data*. <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>
- National Institute of Population Research and Training (NIPORT) and Mitra and Associates of Dhaka. (2014). *Bangladesh - Demographic and Health Survey 2014*. <https://dhsprogram.com/pubs/pdf/FR311/FR311.pdf>
- Rutstein, S. (1999). Wealth versus Expenditure: Comparison between the DHS Wealth Index and Household Expenditures in Four Departments of Guatemala. *Calverton, Maryland: ORC Macro*.
- Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. <https://arxiv.org/abs/1409.1556>
- Tingzon, I., Orden, A., Sy, S., Sekara, V., Weber, I., Fatehkia, M., Herranz, M., & Kim, D. (2019). Mapping Poverty in the Philippines Using Machine Learning, Satellite Imagery, and Crowd-sourced Geospatial Information. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W19*.

Statement of Authorship

I hereby confirm and certify that this master thesis is my own work. All ideas and language of others are acknowledged in the text. All references and verbatim extracts are properly quoted and all other sources of information are specifically and clearly designated.

DATE

Berlin, 24.05.2020

NAME

Dang Ngoc Huy

A handwritten signature in blue ink, consisting of a stylized 'H' followed by a long, sweeping horizontal line that curves upwards at the end.