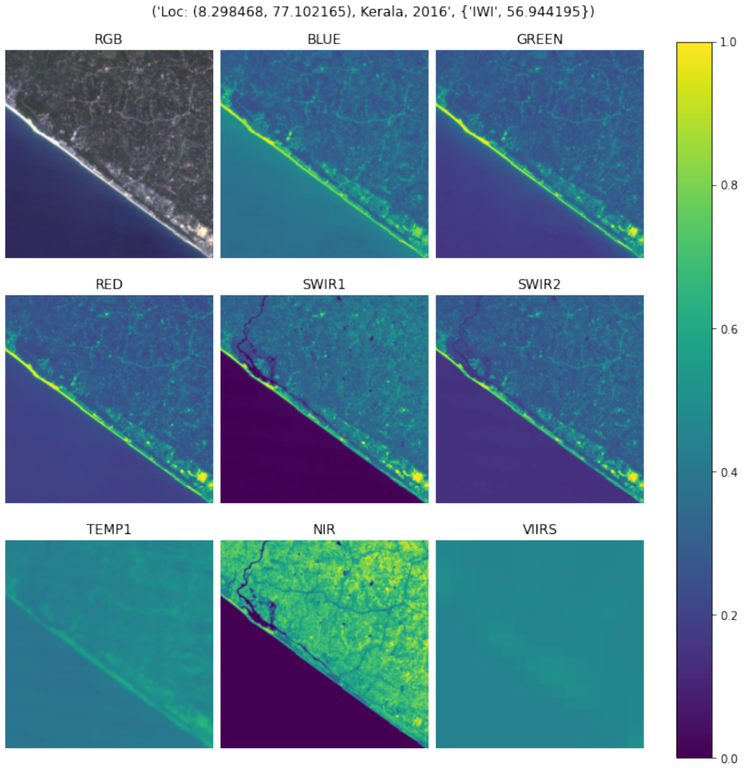
CHALMERS UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG


('Loc: (8.298468, 77.102165), Kerala, 2016', {'IWI', 56.944195})
RGB
BLUE
GREEN
RED
SWIR1
SWIR2
TEMP1
NIR
VIIRS
1.0
0.8
0.6
0.4
0.2
0.0

Predicting Health and Living Standards of India using Deep Learning

Master's thesis in Applied Data Science

Sarath Mookola Raveendran

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2022

Master's thesis 2022

Predicting Health and Living Standards of India using Deep Learning

Sarath Mookola Raveendran

UNIVERSITY OF GOTHENBURG

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2022

Predicting Health and Living Standards of India using Deep Learning
Sarath Mookola Raveendran

Supervisor: Adel Daoud, Department of Sociology and Work Science
Supervisor: Fredrik Johansson, Department of Computer Science and Engineering.
Examiner: Devdatt Dubhashi, Department of Computer Science and Engineering.

Master's Thesis 2022
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Satellite Image with multispectral bands and nightlight band

Typeset in LaTeX
Gothenburg, Sweden 2022

Predicting Health and Living Standards of India using Deep Learning
Sarath Mookola Raveendran
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Poverty eradication is an inexorable process in human growth [21], with poverty estimation being the first and most important stage. Identifying strategies for poverty reduction programs and distributing resources appropriately requires determining the poverty levels of distinct places throughout the world. However, trustworthy data on global economic livelihoods are scarce, particularly in poor countries, making it difficult to provide programs and track and evaluate success. This is partly since this information is gathered through time-consuming and costly door-to-door surveys. Furthermore, survey data includes large gaps, especially in densely populated countries like India. Therefore, we use overhead satellite imagery that contains characteristics that make it possible to estimate the region's poverty level along with the survey data. In this work, we develop deep learning models that can predict a region's poverty level from both DHS survey data and overhead satellite images. This study makes use of both daytime and nighttime imagery in different combinations and analyzes the performance. Poverty prediction studies are mostly focused on datasets from Africa, and very few studies have used a dataset from India. Therefore, in this, thesis, we train a Single Frame model with two deep CNNs having ResNet-18 architecture to predict the average cluster wealth index which is an indicator of poverty given a satellite image of the cluster using DHS survey data and satellite imagery.

Keywords: IWI index, Deep CNN, poverty, ResNet-18, Deep Learning, multispectral images, nightlight images, India, health, and living standards.

Acknowledgements

First, I would like to express my sincere gratitude to my main supervisor Adel Daoud for the continuous support of my thesis work and related research, and his patience, support, motivation, and immense knowledge. His guidance helped me throughout the study, especially his insightful feedback and advice pushed me to sharpen my thinking and brought my work to a higher level. I could not have imagined having a better advisor and mentor for my Master's thesis. Second, I would like to thank my co-advisor, Fredrik D. Johansson. His immense expertise in developing and evaluating DL models was invaluable throughout the study. I would like to thank everyone at Chalmers University who has been involved in the process of this research, especially my examiner, Devdatt Dubhashi. In addition to him, I wish to express my gratitude to Birgit Grohe, our dissertation coordinator who helped me to complete all the practicalities and made this process smooth. I wish to express my gratitude to Mohammad Kakooei and Markus Pettersson for giving insights into the thesis work. Last but not the least, I would like to extend my sincere gratitude to all my family members who gave a supporting environment for conducting the study.

Sarath Mookola Raveendran, Gothenburg, June 2022

Contents

List of Figures

List of Tables

1

Introduction

Poverty is defined as a state or condition in which a person or a group lacks the financial resources and requirements for a comfortable living [46]. Poverty-stricken individuals and families may lack enough shelter, safe drinking water, nutritious food, and medical services. When labour wages are insufficient to meet basic human needs, it leads to a high prevalence of illiteracy, poor healthcare, and a lack of financial resources. Poverty has a significant impact on people's health due to a lack of adequate food, decent clothing, medical services, and sanitary living circumstances [40]. Poverty is associated with undesirable circumstances such as substandard housing, homelessness, limited food and nutrition insecurity, poor child care, lack of access to health care, dangerous neighborhoods, and overcrowded classrooms, all of which harm children [19].

Widespread chronic poverty and less economic development than other countries are the problems with the underdeveloped countries [12]. The phrase "underdeveloped country" is unofficial, however, the United Nations classifies countries that qualify as underdeveloped as developing countries or least-developed countries (LDCs) [7]. India is one of the developing countries listed by the United Nations [1] and home to 26% of the world's poor population in 2012 [36]. Poverty reduction, more work opportunities, and reduced income disparities are all critical conditions for development[44]. The United Nations has defined 17 Sustainable Development Goals (SDGs). The first goal is to eliminate extreme poverty for all people by 2030.

For years, international organizations like The World Bank and developing countries like India spent a significant amount of resources on poverty measurement and analysis [20]. Measuring and analyzing the poverty level in various developing and underdeveloped countries can help government and non-government organizations to interpret which poverty reduction strategy has succeeded and which are not successful. These developing and underdeveloped countries have a rapidly changing economic situation [15]. Poverty measurement aids developing countries in evaluating the efficacy of their programs and guiding their development plan. One of the typically used indices for measuring the economic positions of households is wealth indices. To measure the economic condition of a household, a scale called International Wealth Index (IWI) is used. IWI is a reliable and easy-to-understand metric for comparing the economic condition of a household. It also helps to understand the wealth distribution across the country. IWI ranges from 0 to 100, with 0 indicating households with no assets and the poorest housing quality and 100 representing households with all assets and the best housing quality [47]. Demographic and Health Surveys (DHS) help us to get a wide range of monitoring and impact evaluation indicators nationally with help of representative household surveys [50]. DHS survey provides IWI for all developing countries. DHS surveys for India from 2015 to 2016 are available. However, these surveys are not complete and are done only for a small sample size of the country (30797 samples across India). To overcome this limitation, AI can be used by researchers and policy-makers to get reliable data on poverty and wealth in developing countries [55]. Technological innovations such as machine learning-powered by the Big Data revolution has opened unimaginable possibilities for government organization and countries for planning, analyzing, and reviewing policy decisions, as well as directing humanitarian activities [30]. The main objective of this thesis is to predict IWI for the missing areas using Deep Learning and Satellite Images.

Poverty prediction through DL using satellite images is a less explored field compared to survey-based predictions. Perez et al. [39] use CNN models trained on open-source multi-spectral daytime satellite images of the African continent from the Landsat 7. Jean et al. [30] had a different architecture, which is a CNN model pre-trained on ImageNet to identify low-level image features such as edges and corners. Then, CNN is fine-tuned to predict the nighttime light intensities corresponding to input daytime satellite imagery. Further, they estimated cluster-level expenditures or assets with ridge regression models trained on mean cluster-level values from the survey data and corresponding image features extracted from daytime imagery by CNN [30]. Yeh et al. [55] used publicly available multispectral satellite images, and a DHS survey and created deep learning models to predict survey-based estimates of asset wealth in over 20,000 African Villages. Pandey et al. [38] used a dataset from India and developed multitask fully convolutional model to predict the material of the roof, source of lighting, and source of drinking water from the satellite imagery of a village. Then, a second model is built to predict the income levels (a direct indicator of poverty) using the predicted developmental parameter outputs of the first model.

Most of the studies on poverty prediction work with the data from Africa. India also faces similar poverty problems, at least in some parts. Moreover, the gaps in DHS data exist, especially in the remote areas of India. However, very few studies have focused on the dataset from India. Therefore, in this study, we focus on the prediction of the IWI index for the dataset from India. Our objective of this study is to train models to predict the average cluster wealth index given a satellite image of the cluster. The contributions of this thesis are three-fold. First, it develops four DL models that can predict the IWI indices of missing regions in India, thereby filling the gaps in the DHS survey data. Second, it evaluates the difference in performances of these four models for two types of data splits. Third, it analyses the performance of all four models and identifies the best among them for predicting the IWI of Indian administrative units. We use two deep CNN models with ResNet-18 architecture, where one is used for training multispectral images and the other one is for training nightlight images. As a baseline, we have trained the model with only nightlight images. We have also trained the ResNet-18 model with multispectral images and multispectral together with nightlight images. For performance comparison, we have split the data randomly into 5 folds according to the states, which are called out-of-state splits and not according to states which are called in-state splits. For evaluating the performance of the model, we have used metrics such as r^2 and R^2.

Finally, we have also done a performance comparison of different models on different types of data splits.

2

Literature Review

This thesis studies the application of Deep Learning models, for predicting the health and living standards of India. In order to provide the reader with the necessary information needed to better understand the remainder of the thesis, this section provides background information and describes the related works and state-of-the-art poverty prediction using various DL models.

2.1 Benchmarks for monitoring the sustainable development goals

SUSTAINBENCH is a collection of benchmarks for monitoring the sustainable development goals with Machine Learning [54]. The Sustainable Development Goals are a framework for a better, more sustainable future for everyone. Poverty, inequality, climate change, environmental degradation, peace, and justice are among the worldwide concerns they address. Lack of ground data/survey data is one of the major challenges toward progress in United Nations Sustainable Development Goals (SDGs). The progress toward the SDGs is measured by censuses, surveys and civil registration. However, these surveys and censuses are expensive, time-consuming and are not conducted regularly. Yeh et al. [54] use these available survey data and abundant, cheap data such as satellite imagery, social media posts, and/or mobile phone activity to predict the gaps in the data. For using non-traditional data sources that are cheap, globally available, and constantly updated to fill in data gaps, ML algorithms tailored for monitoring SDGs are critical. The main contributions of this study are towards enabling machine learning to measure and achieve the SDGs, establishing standard benchmarks for evaluating machine learning models, and building unique machine learning strategies where enhanced model performance promotes progress toward the SDGs. However, the approach has the following limitations. Ground surveys may not be totally replaced by machine learning algorithms. Imperfect ML model predictions may induce biases that spread through downstream policy decisions, resulting in detrimental societal consequences. Privacy concerns may arise from the usage of survey data, high-resolution remote sensing photos, and street-level images. In future work, the authors plan to expand datasets and benchmarks as new data sources [54].

2.2 Multi-Task Deep Learning for Predicting Poverty from Satellite Images

Shailesh et al. [38] suggest a two-step technique for using satellite images to predict poverty in rural areas. In this study, data from the most populous state of India, Uttar Pradesh, is collected from The 2011 Census of India. Further, Google Geocoding API was queried to obtain coordinates of the centre of a village as well as the box-bounding latitudes and longitudes(geocodes) from its address in the census data. Then the Google Static Maps API is utilized to extract images for the villages from the determined geocodes. First, the authors created a multitask fully convolutional deep network capable of predicting roof material, illumination source, and drinking water source from satellite pictures. Second, they measure poverty using the predicted developmental statistics. The models were able to learn significant features such as highways, water bodies, and farm areas using full-size satellite imagery as input and without pre-trained weights, and attain near-optimal performance. The performance of predicting income levels on the basis of the multi-task model has an accuracy of 0.969. The multi-task fully convolutional model was able to distinguish task-specific and independent feature representations, in addition to speeding up the training process. The authors also observed that the model trained on the predictions of the multi-task model performs close to the optimum model (model trained on Census data, and significantly better than their baseline model trained for majority class prediction.

2.3 Livelihood indicators from community-generated street-level imagery

Measurement of the populace's well-being is taken into account by the government and other organizations while taking a decision. However, these measurements at a large scale are expensive. So these measurements are often taken rarely in developing countries. Lee et al. [32] propose an approach to measure these predict key livelihood indicators which are less expensive, interpretable and scalable. Street-level imagery is used as input to this approach, which is less expensive compared to ground-level surveys. Lee et al. suggest two approaches in their studies. In the first approach, multi-household cluster representations are detected from the informative images from street-level imagery. In the second approach, the relationship between images is captured by a graph-based approach. The study mainly focused on three indicators, poverty, population, and women's body mass index. The main contribution of the study is that it provides a less expensive, effective and scalable approach than the traditional surveying approach [32]. The study is validated by predicting indicators of poverty, population, and health and its scalability by testing in two different countries, India and Kenya.

2.4 Economic well-being using satellite images

Measurements of human well-being at the local level are critical for governments to make informed decisions about public service delivery and policy, for governmental and non-governmental organizations to target and evaluate livelihood programs, and for the private sector to develop and deploy new products and services. However, data for measuring the accurate and comprehensive economic well-being of many developing and underdeveloped countries are missing. Yeh et al. [55] in this study uses a deep learning model to predict wealth assets in different countries in Africa. ResNet-18 architecture is used for this study. The model uses two ResNet-18 architecture: one trained for night light images and the other for multispectral images. This approach outperforms the other approaches used in the literature, such as scalar nightlights. However, the policy community finds CNN based approach hard to adopt as the information CNN used for prediction is less interpretable compared to the simpler approaches. As a future extension of this work, the authors plan to improve the interpretability of deep learning models in this context and develop approaches to navigate the performance-interpretability tradeoff. Further, they also plan to improve the approach by the incorporation of higher-resolution optical and radar imagery now becoming available at near-daily frequency, or in combination with data from other passive sensors such as mobile phones or social media platforms.

2.5 Combining satellite imagery and machine learning to predict poverty

In the developing as well as the underdeveloped world, reliable data on economic livelihoods are limited, making it difficult to investigate these results and implement policies to improve them [30]. For much of the developing world, data on crucial indicators of economic development is scarce. This data gap is impeding efforts to detect and analyze variance in these outcomes, as well as to effectively deliver assistance to the most vulnerable locations. The lack of data on the African continent is extremely limiting. Jean et al. propose a novel machine learning-based approach for collecting socioeconomic data from high-resolution daylight satellite photos. We then test our method in five African countries that have recently georeferenced local-level data on economic outcomes. Although the method is successful in assessing economic well-being at the cluster level, it is unable to examine the ability to distinguish disparities within clusters, since public-domain survey data assigns identical coordinates to all households in a particular cluster to protect respondent privacy.

2.6 Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping

In developing an underdeveloped country, a lack of reliable data is a serious impediment to long-term growth, food security, and disaster relief. Data on poverty are often limited, little covered, and time-consuming to gather. On the other hand, remote sensing data such as high-resolution satellite images is becoming increasingly accessible and affordable. Nevertheless, such data is very unstructured, and hence there were no techniques for automatically extracting relevant insights to help guide legislative decisions and humanitarian efforts. Xie et al. proposed a machine learning approach to extract large-scale socioeconomic indicators from high-resolution satellite imagery [53]. However, the scarcity of training data made it difficult to apply to CNNs. Therefore, Xie et al. introduced a novel transfer learning approach for analyzing satellite imagery that leverages recent deep learning advances and multiple data-rich proxy tasks to learn high-level feature representations of satellite images. A major advantage of this approach is its generalizability, and therefore has great potential to help solve global sustainability challenges.

2.7 Using machine learning tools for predicting poverty in rural India

In this paper [48], To predict rural poverty on a state-by-state basis, Subash et al. used satellite night light data and machine learning methods (Artificial Neural Network). The authors used night light data as a predictor for the constructed model, comparing it to per capita domestic product. The data used for this study was gathered from the University of Michigan's open access nightlight data. The 'India Lights API' collection contains data on rural nightlights for around 6,00,000 villages across India over a 20-year span, from 1993 to 2013. The data was gathered from satellite images of the world taken every night by the US Department of Defense's Defense Meteorological Satellite Program (DMSP). The authors also used available data on rural poverty estimates at the state level. Further, they used data on GDP to predict poverty at the state level, using the same algorithm used for predicting poverty with night light data. Per capita income was calculated by dividing the gross domestic product of a state by the population of the state, as the per capita income can be used as benchmarking data. Due to the negative correlation relation between income and poverty, it could also be used as a predictor of poverty. Authors came to the conclusion that nightlight data is a stronger predictor of poverty than per capita GDP.

All except one of the aforementioned studies focused on the African dataset. However, Africa is not the only developing country in the world. Therefore, our study is conducted using an Indian dataset. Although India is a country unlike Africa, the data points are more compared to the African dataset. The study which was validated using the Indian dataset was using economic indicators other than IWI. Further, the successful implementations of AI poverty prediction use a CNN model and transfer learning approach. Hence, in our study, we make use of CNN as a deep learning model for predicting the health and living standards of various parts of India.

3

Background

This chapter presents the background that provides the foundation for this thesis report. It helps to interpret and understand the results that are obtained. The theory presented in this chapter is referred to in the discussion/conclusion part of the report as well to show how the research connects to existing research.

3.1 Deep learning

Deep Learning (DL) is a subset of Machine Learning (ML). In DL, hierarchical architectures are used to learn high-level abstractions from the data [24]. DL models were successful in various tasks such as image classification, object detection, natural language processing and information retrieval [41]. For this study, we use Convolutional Neural Networks and satellite images are given as input to the CNN as shown in Fig 3.1 [14].

3.1.1 Convolutional Neural Networks (CNN)

Regular grid-like topology means that each node in data is connected with two neighbours along one or more dimensions. Time series data and images are examples of this kind of data. Time series data can be considered as a one-dimensional grid with regular time intervals, and images can be considered as a two-dimensional grid of pixels. CNN's are a type of neural network which can process data with grid-like topology [22]. Consequently, CNN has high performance in practical applications [22]. The convolutional layer, non-linearity layer, pooling layer, and fully connected layer are the layers of CNN. Pooling and non-linearity layers do not have parameters, whereas convolutional and fully connected layers have parameters.[11]

The input image is transformed using a convolution layer to extract features from it. The image is convolved with a kernel in this transformation. The output of this convolution operation is called a feature map or activation map [2]. A kernel, also known as a convolution matrix or convolution mask, is a small matrix that is smaller in height and width than the image to be convolved. This kernel slides across the image input's height and breadth, computing the kernel's and image's dot product at each spatial place.

The convolution layer is followed by a non-linear transformation. Non-linear transformation can adjust or reduce the generated output. The major functions like sigmoid function, tanh are used as activation functions. However, the most popular function used recently is rectified linear activation function or ReLU [11]. The rec-

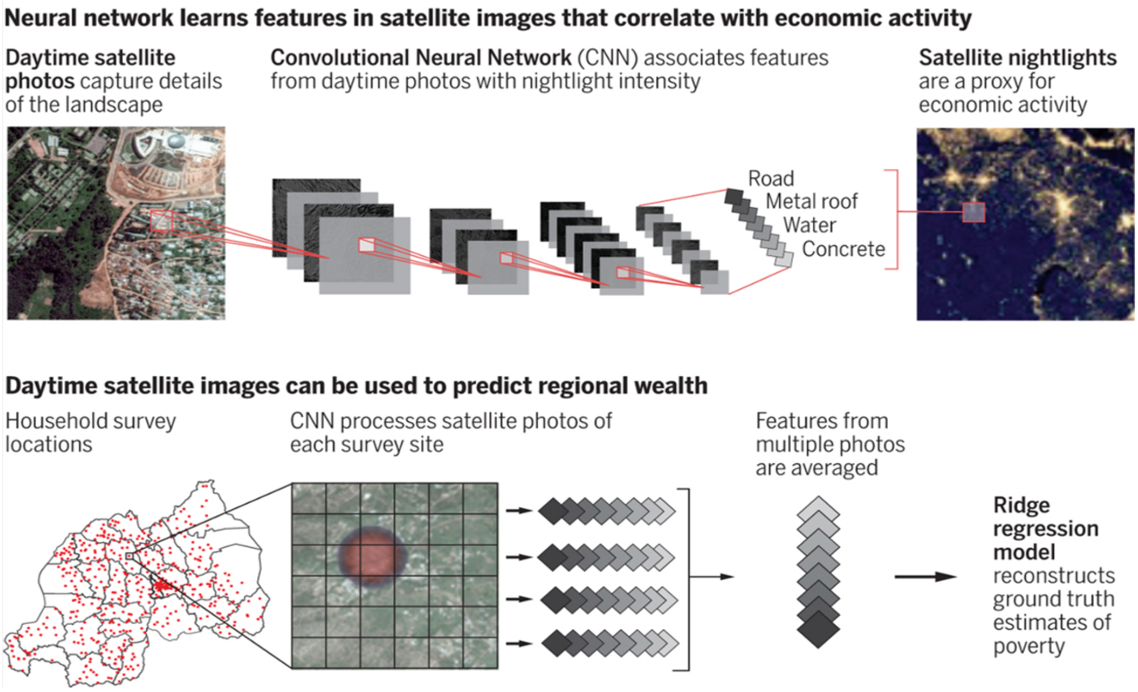

Neural network learns features in satellite images that correlate with economic activity
Daytime satellite photos capture details of the landscape
Convolutional Neural Network (CNN) associates features from daytime photos with nightlight intensity
Satellite nightlights are a proxy for economic activity
Road
Metal roof
Water
Concrete
Daytime satellite images can be used to predict regional wealth
Household survey locations
CNN processes satellite photos of each survey site
Features from multiple photos are averaged
Ridge regression model reconstructs ground truth estimates of poverty

Figure 3.1: CNN illustration [14]

tified linear activation function (ReLU) is a piecewise linear function that outputs the input directly if it is positive and zero otherwise [6].

The main aim of the pooling layer is to downsample, which reduces the complexity for further layers [11]. After a convolution layer, a pooling layer is usually applied. Pooling split images to rectangular subregions. Pooling is based upon the assumption that changing the input by a little amount has no effect on the pooled outputs. Max and average pooling are different kinds of pooling used. In max pooling, the max value of the subregion is taken while in average pooling, the average value of the subregion is taken. Max pooling provides better performance compared to min or average pooling [2].

Fully connected layers are usually found towards the end of a CNN architecture. Hence, each node in a fully connected layer is linked to every node in the previous and subsequent layers. The vectorization of the features map created by the previous layers is sent through a fully connected layer, which captures complicated interactions between high-level features. This layer produces a one-dimensional feature vector.

Fig 3.1 [14] illustrates that using the ReLU activation function, filters or feature detectors are applied to the input image to build feature maps of activation maps. Feature detectors or filters aid in the identification of various features in an image, such as edges, vertical lines, horizontal lines, bends, and so on. A non-linear transformation is applied after the convolution layer to limit the values of the generated output. The feature maps are then pooled to ensure translation invariance. The features map generated by the pooling layer is vectorized and sent through a fully connected layer, which captures complicated interactions between high-level features. This layer produces a one-dimensional feature vector as its output.

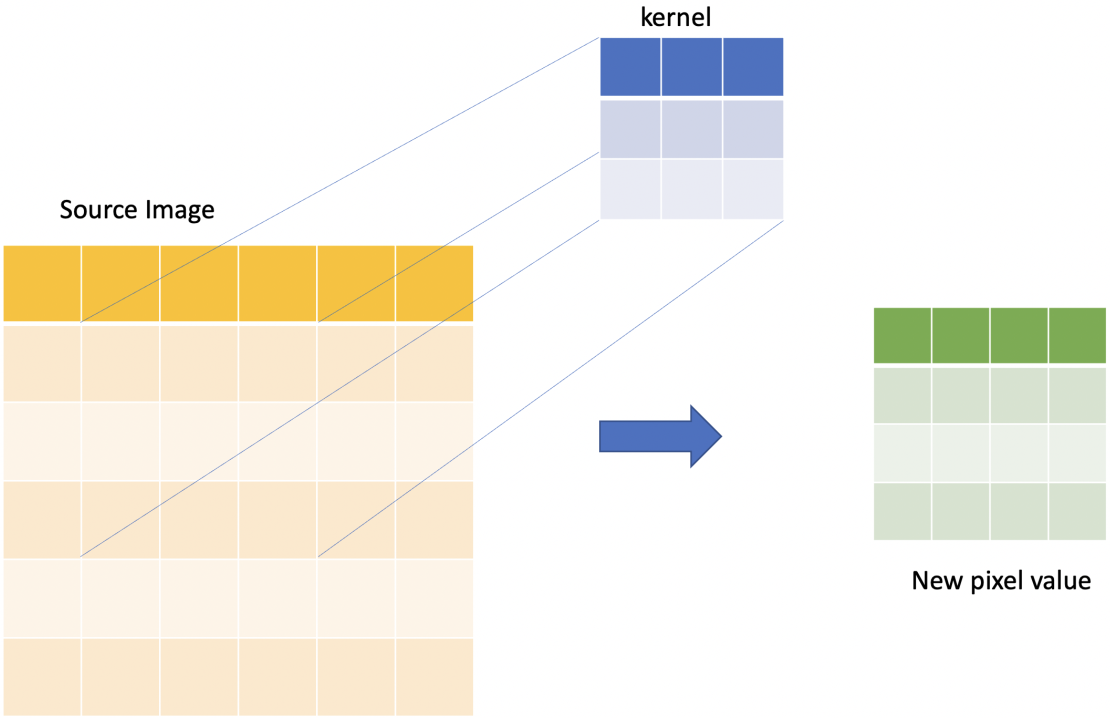


Figure 3.2: Filter applied to an input image

3.1.2 Back propagation algorithm

The back propagation algorithm is significantly important in training neural networks [28]. Neural networks have three main layers namely the input layer, hidden layer and output layer [10]. The back propagation algorithm works its way backwards from the output layer to the input layer, calculating error gradients along the way [28]. After computing the gradients of the cost function with respect to each parameter (weights and biases) in the neural network, the algorithm uses these gradients to update the value of each parameter in the network using a gradient descent step towards the minimum. The Gradient is nothing but the weighted derivative of the loss function [25]. It is used to update the weights in neural networks to minimize the loss of function during back propagation. When we move backwards with each layer during back propagation, the derivative or slope gets less and smaller until it vanishes. During back propagation, an exploding gradient happens when the derivatives or slope grow larger and larger as we go backwards with each layer. This is the absolute opposite of the vanishing gradients' problem.

3.1.3 Residual Networks (ResNets)

ResNet is one of the most powerful deep neural networks [33]. An increase in the number of layers in a deep neural network leads to a vanishing/exploding gradient, which will rapidly degrade the accuracy of the model [25]. This rapid degradation of accuracy is not because of overfitting. Instead, it is caused when the gradient becomes zero or too large. Thus, training deep neural networks is difficult. To solve the problem of the vanishing/exploding gradient, residual networks are used. In

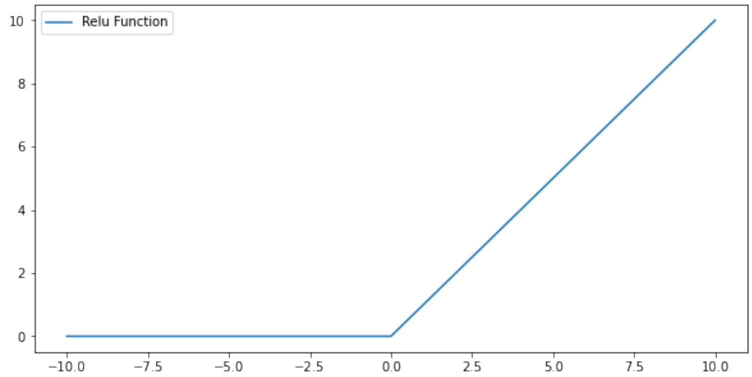
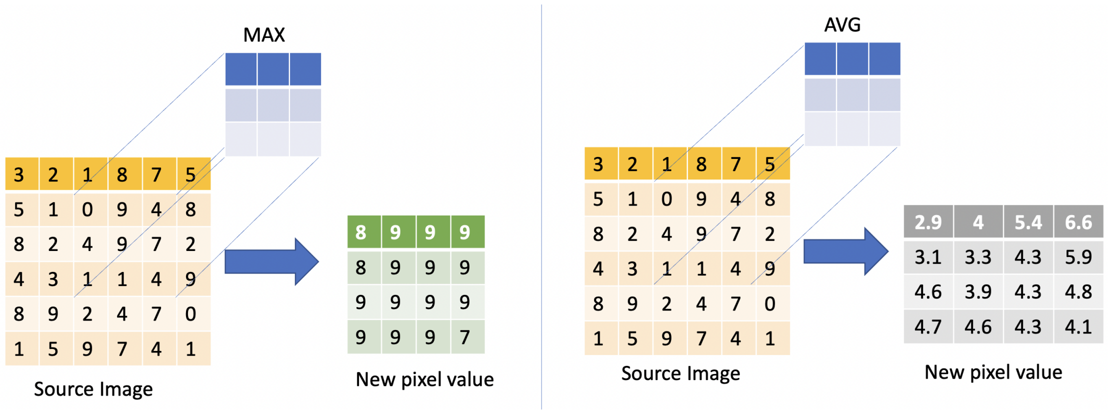


Figure 3.3: Value of ReLU function



3	2	1	8	7	5
5	1	0	9	4	8
8	2	4	9	7	2
4	3	1	1	4	9
8	9	2	4	7	0
1	5	9	7	4	1

8	9	9	9
8	9	9	9
9	9	9	9
9	9	9	7

3	2	1	8	7	5
5	1	0	9	4	8
8	2	4	9	7	2
4	3	1	1	4	9
8	9	2	4	7	0
1	5	9	7	4	1

2.9	4	5.4	6.6
3.1	3.3	4.3	5.9
4.6	3.9	4.3	4.8
4.7	4.6	4.3	4.1

Figure 3.4: Max and average pooling

Residual Network, a technique called skip connections is used. In skip connection, a few layers are skipped and connected directly to the output[26]. Shortcut connections are used by ResNet to significantly minimize the difficulty of training, resulting in considerable improvements in both training and generalization error. The Deep Learning community started to build deeper networks as they were able to achieve high accuracy values since 2013. Deeper networks can also represent more complicated properties, which improves the model's robustness and performance. Adding more layers, on the other hand, did not work for the researchers. The problem of accuracy decline was discovered while training deeper networks. In other words, adding more layers to the network either saturated the accuracy value or caused it to drop suddenly. The vanishing gradient effect, which can only be seen in deeper networks, was the cause of the accuracy decline.

The error is calculated, and gradient values are determined during the back propagation stage. After the gradients are transmitted back to hidden layers, the weights are modified. The gradient determination process is repeated until the input layer is reached, after which it is sent back to the next concealed layer. As we get closer to the bottom of the network, the gradient gets less and smaller. As a result, the weights of the first layers will either update slowly or remain unchanged. In other words, the network's initial layers will not be able to learn successfully. As a result, deep network training will not converge, and accuracy will begin to deteriorate or

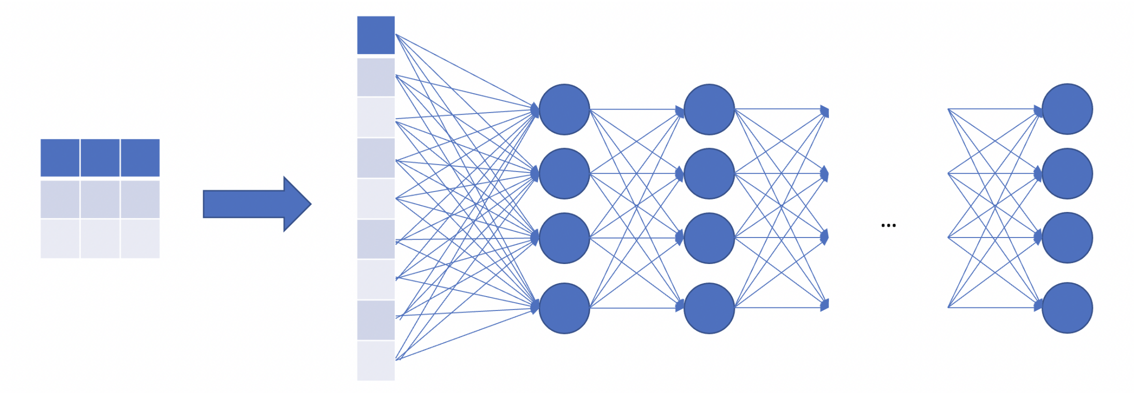


Figure 3.5: Fully Connected layer

saturate at a specific value. Although the vanishing gradient problem was solved by employing normalized weight initialization, deeper network accuracy did not improve.

Deep Residual Networks resemble networks that include convolution, pooling, activation, and fully-connected layers piled one on top of the other. The identity connection between the layers is the only construction that transforms the simple network into a residual network. Fig 3.6 shows the identity connection as the curved arrow originating from the input and sinking to the end of the residual block.

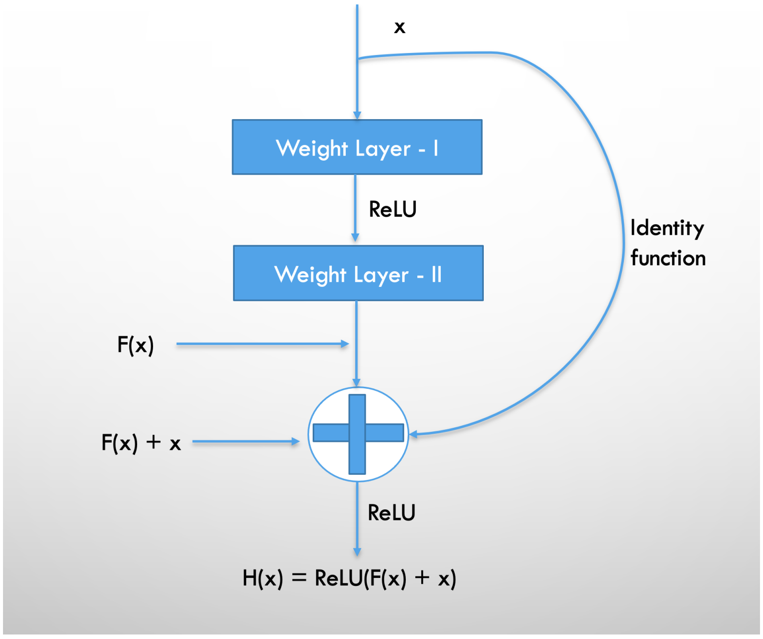


Figure 3.6: Residual Block

3.1.4 Ridge regression

The covariates (the columns of X) are super-collinear when the design matrix is high-dimensional. In regression analysis, recall collinearity occurs when two (or more) factors are highly linearly connected. As a result, the parameter space's space spanned by super-collinear variables is a lower-dimensional subspace. It is (nearly) difficult to disentangle the contributions of the various variables if the design matrix X, which contains the collinear covariates as columns, is (close to) rank deficient. The uncertainty about the covariate responsible for the variation explained in Y is frequently reflected in the fit of the linear regression model to data by a large error in the estimates of the regression parameters corresponding to the collinear covariates, and, as a result, by large values of the estimates [35]. The first step in ridge regression is to normalize the variables (both dependent and independent) by dividing by their standard deviations and removing their means. This creates a notation problem because we need to declare whether the variables in a formula are standardized or not. All ridge regression computations are based on standardized variables in terms of standardization. The final regression coefficients are rescaled to their original scale when they are displayed. The ridge trace, on the other hand, is on a standardized scale. The base of any regression machine learning model is the standard regression equation, which is stated as:

$$Y = XB + e$$

3.1.5 Performance Evaluation Metrics

Performance Evaluation of a Deep Learning model is extremely important as it provides a more realistic measure of how the model will perform when deployed in a production environment, which helps to avoid overfitting and keep the model simple. A correlation is the quantitative measure of the association between observed and predicted values, we use this metric for evaluating the performance of DL models. A measure of an association between variables is called correlation in its broadest definition. In correlated data, a change in one variable's magnitude is linked to a change in another variable's magnitude, either in the same (positive correlation) or opposite (negative correlation) direction. In this study, we use the coefficient of determination and squared Pearson correlation coefficient metrics. The squared Pearson correlation coefficient (r^2) is used to identify patterns and is used to calculate the effect of change in one variable when the other variable changes, whereas the coefficient of determination (R^2) is used to identify the strength of a model. Therefore, the coefficient of determination (R^2) is used to evaluate the performance of the models in this study.

3.1.5.1 Coefficient of Determination

The coefficient of determination is a metric used to explain the amount of variability of one factor that is caused by its relationship to another related factor. In other words, the coefficient of determination measures the goodness-of-fit based on explained variance [18][45]. The coefficient of determination, denoted by, R^2 is typically used to evaluate regression models. The value of R^2 can range from 0 to 1. R^2 with value 0 indicates the regression line does not fit the set of data points, and R^2 with value 1 indicates a perfect fit for the set of data points.

Let the n values of a dataset marked by y_1,..., y_n each associated with a fitted (or modeled, or predicted) value f_1,...,f_n.
Then the residuals will be $e_i = y_i - f_i$ (forming a vector e).
If $\bar{y}$ is the mean of the observed data:

$$\bar{y} = \tfrac{1}{n}\sum_{i=1}^{n} y_i$$

then the variability of the data set can be measured with two sums of squares formulas:

- The sum of squares of residuals, $SS_{res} = \sum_i (y_i - f_i)^2 = \sum_i (e_i)^2$
- The total sum of squares, $SS_{tot} = \sum_i (y_i - \bar{y})^2$

Therefore,

$$R^2 = 1 - \tfrac{SS_res}{SS_tot}$$

3.1.5.2 Squared Pearson Correlation Coefficient

Pearson's correlation coefficient, abbreviated as r, is a measure of the strength of a linear relationship between two variables [13] [45]. A Pearson's correlation is an attempt to build a line of best fit through the data of two variables. The Pearson correlation value, r, shows how far apart from all of these data points are from the best-fit line. Pearson's r can take values between -1 and 1. Squared Pearson Correlation Coefficient is the squared value of r. The squared Pearson correlation coefficient is usually not equal to the coefficient of determination (or $r^2 \neq R^2$).
We can obtain a formula for r_{xy} by substituting estimates of the covariances and variances. Given paired data ((x_1,y_1), ..., (x_n,y_n) consisting of n pairs, r_{xy} is defined as:

$$r_{xy} = \tfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})^2}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Squared Pearson Correlation Coefficient is obtained by squaring the r_{xy}

$$r_{xy}^2 = \left(\tfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})^2}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}\right)^2$$

4

Methods

This chapter provides a detailed description of the methods used in developing the deep learning models relevant to this study. It starts with an introduction to the wealth index and DHS survey, followed by an exploration of the DHS survey and preprocessing of satellite images. The model architecture and implementation of various deep learning models are described in this chapter.

4.1 Deep Learning workflow

Figure 4.1 illustrates the nine stages of the deep learning workflow. Some stages are data-oriented (e.g., data collection, data labelling, data cleaning) and others are model-oriented (e.g., model training, model evaluation). There are many feedback loops in the workflow. Typically, the model evaluation may loop back to any of the previous stages.

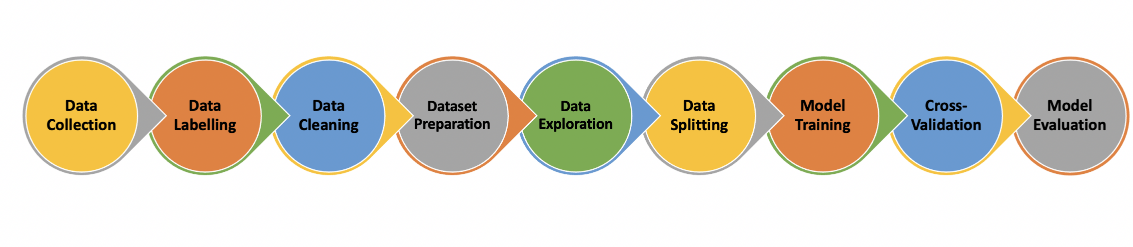


Figure 4.1: Deep learning workflow

4.1.1 Data Collection

We have collected data from two different data sources for conducting the study, namely DHS Program and Google Earth Engine. DHS survey data is tabular data downloaded from the DHS Program website and satellite images are obtained from Google Earth Engine.

4.1.1.1 Demographic and Health Surveys (DHS)

The DHS Program collects and shares various information about people such as infant and child mortality, fertility, maternal health, child immunization, malnutrition levels, HIV prevalence etc. [49]. The data thus collected is freely available and can be used for analysis purposes [51]. The program is funded by the United States Agency for International Development since 1984 [49].

DHS surveys use nationally representative samples of childbearing-age women and, more subsequently, males. The survey is conducted, and the results are reviewed by national government agencies [49]. Each country's specific demands were addressed by adopting a core questionnaire. A list of the 44 nations that conduct DHSs was supplied, together with details on the most recent year(s) of the survey and the number of respondents. Questions on fertility and mortality, anthropometry, family planning, maternity care, infant nutrition, vaccination, child illness, and AIDS were included in the core questionnaire [3]. The surveys were beneficial in that they provided a wide range of health and healthcare indicators. Continuous data quality checks were performed to improve instruments, ensure qualified field people, use contemporaneous data entering and editing, and provide feedback to interviewers while the instrument was being administered in the field. After the fieldwork was completed, the results were publicized rapidly, and tabulations were ready within 2-3 months [3]. Reporting and recall bias were among the flaws, especially for age or other retrospective statistics based on the memory of a prior occurrence. Omissions were not regarded as a severe issue [49].

The data collected from households can be used for analyzing the trend, planning and monitoring development programs. In the late 1980s, the DHS project began georeferencing cluster coordinate data, and in 2003, it began making georeferenced GPS datasets available to the public [4]. To anonymize the data collected with georeferencing, households in a cluster are changed to the same latitude/longitude. Clusters in Urban areas are randomly displaced to two kilometres and clusters in rural areas are displaced to five kilometres, and one per cent of randomly selected clusters are displaced to ten kilometres [50]. International Wealth Index (IWI) is one of the most extensively used wealth indices in surveys like DHS [47].

4.1.1.2 International Wealth Index (IWI)

The first comparable asset-based index of households is called as International Wealth Index. This can be used to compare the economic status of households in low and middle-income countries [47]. A questionnaire is used to collect information about the IWI in a household. A questionnaire is prepared with data which are easy to collect such as television, bicycle, type of water access and sanitation facilities and materials used for housing construction [43]. Other reasons are also considered while collecting the information on each of these items. For example, diarrhoea among children is directly associated with floor type, water supply and sanitation facilities of their households. Mass media health messages are received through television and radio, hence this is considered in the questionnaire. Possession of a vehicle is related to emergency medical transportation. Multiple persons sleeping in the same area and non-electrical source of lighting is linked to higher transmission of respiratory illness[5]. If two or more households have the same IWI value doesn't mean that they have the same assets. It means that these households have reached the same level of material satisfaction [47].

DHS surveys for India with a Geographic coordinate system (GPS) from 2015 to 2016 are available for download and analysis [50]. IWI score in this report stands for IWI mean value. Because, to anonymize the data collected with georeference, households in a cluster are changed to the same latitude/longitude. Clusters in Urban areas are randomly displaced to two kilometres and clusters in rural areas are displaced to five kilometres, and one per cent of randomly selected clusters are displaced to ten kilometres. Therefore, an IWI score at a specific place is certainly not obtained. Fig 4.2 illustrates IWI scale, which ranges from 0 to 100 [9]. The wealth asset index value of a household with all durables and the highest quality housing and services is 100 and if its value is 0 if it lacks all the durables and has the poorest housing and services.

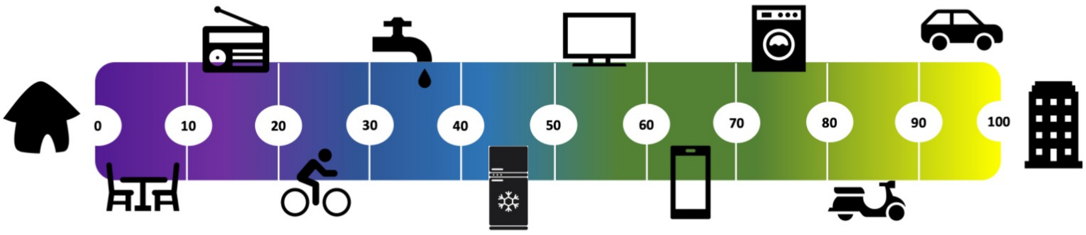


Figure 4.2: International Wealth Index scale

4.1.1.3 DHS data

Since 1984, the Demographic Health Survey (DHS) program has been working with governments to collect and share key information about people, their health, and their health systems for national representation. Data is collected nationally through various surveys on malaria, HIV/AIDS, child health, nutrition, fertility, etc. Wealth indexes are also included in some of these surveys. The wealth index used in the DHS survey is known as IWI (International Wealth Index). IWI can be used to measure the economic well-being of a household. The IWI score is taken from the DHS survey, where the score is calculated from the responses to a set of questions prepared by DHS. Questions made by DHS are based on common assets of ownership, such as if you have a television, a bicycle, the type of material used for the construction of the house, the quality of water used in the house, the number of rooms in the house, if the house is powered by electricity or not, and so on [55]. IWI ranges from 0 to 100, where 0 indicates households with no assets and the poorest housing quality, and 100 represents households with all assets and the best housing quality. DHS surveys with georeferencing coordinates for the years 2015-2016 are available for India, which can be downloaded directly from the website [50]. The data from the DHS survey was downloaded for conducting this study, and we observed that there are 30,798 household cluster points spread across India. Fig 4.3 illustrates the DHS cluster points plotted on the map of India.

4.1.1.4 Administrative divisions of India: States and Union Territories

Names of administrative divisions vary greatly between continents. India, the seventh-largest country in the world(in terms of area) consists of 28 states and

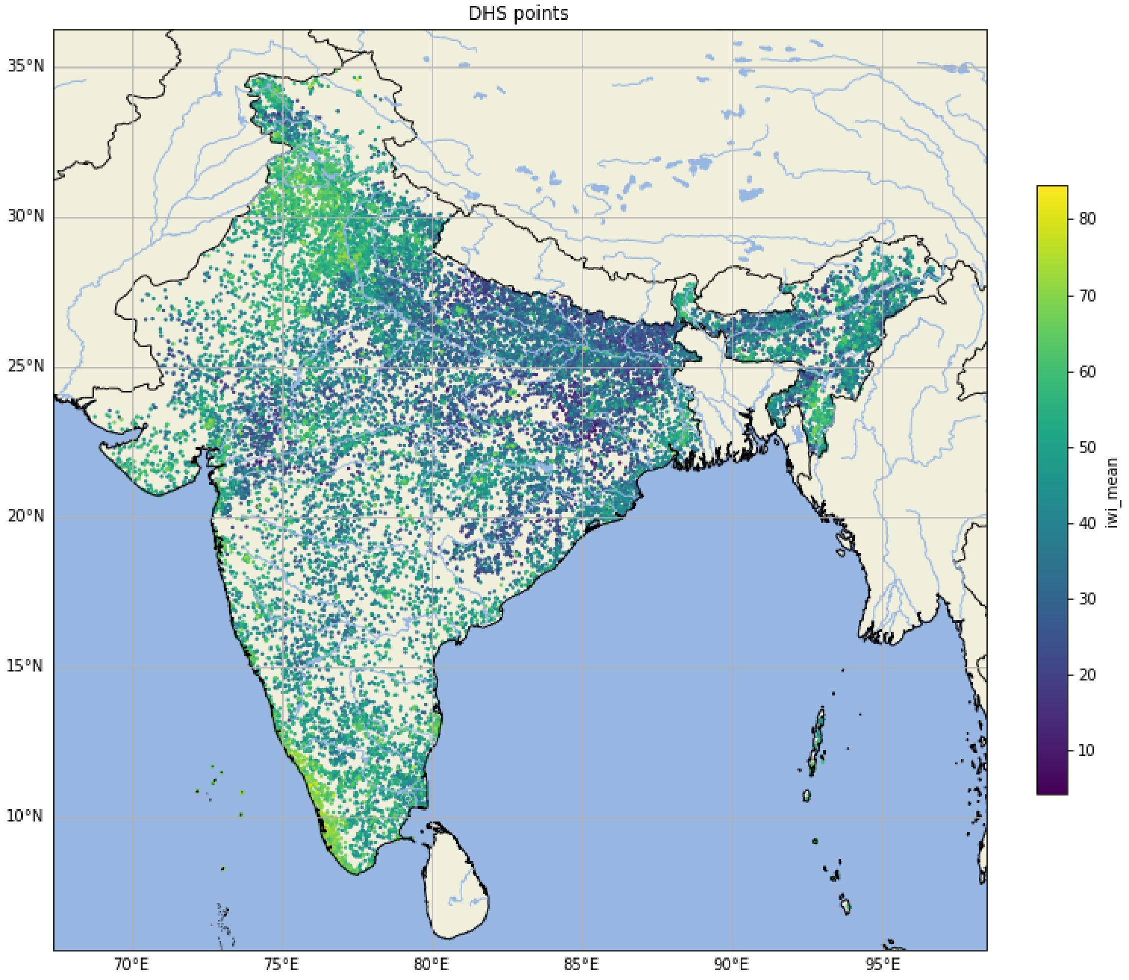


Figure 4.3: 30,798 DHS points plotted on Indian's Map.

8 union territories, constituting a total of 36 entities. These states and union territories are further subdivided into districts and smaller administrative units. States are formed based on linguistic lines [8]. DHS dataset of India has states which are further divided into districts, sampling clusters and households. Typically, countries in the DHS dataset are divided into clusters based on census reports. Clusters are a group into which the population is divided. For example, for a rural area, a cluster can be an entire village, part of a village or a group of small villages whereas for an urban area, a cluster can be a building block, city block etc. Each cluster consists of 20-30 households. Each household has a latitude and longitude co-ordinate associated with it. To anonymize the data collected with georeference, households in a cluster are changed to the same latitude/longitude. Clusters in Urban areas are randomly displaced to two kilometres and clusters in rural areas are displaced to five kilometres, and one per cent of randomly selected clusters are displaced to ten kilometres.

4.1.1.5 Satellite Imagery

Google Earth Engine (GEE) is a cloud computing platform that may be used to process satellite images as well as other geographical and observational data. It gives access to a big library of satellite imagery as well as the computing capacity [23]. The entire remote sensing images from Landsat images and Sentinel-1 and Sentinel-2 are used in GEE [23]. The project uses a DHS survey from 2015 to 2016 hence images from Landsat satellites 7, and 8 are used. The images collected from these satellite has multispectral bands and has a spatial resolution of 30 m/pixel. Nightlight Images (NL) are also used for the thesis. Figure 4.4 shows the bands in a satellite imagery.

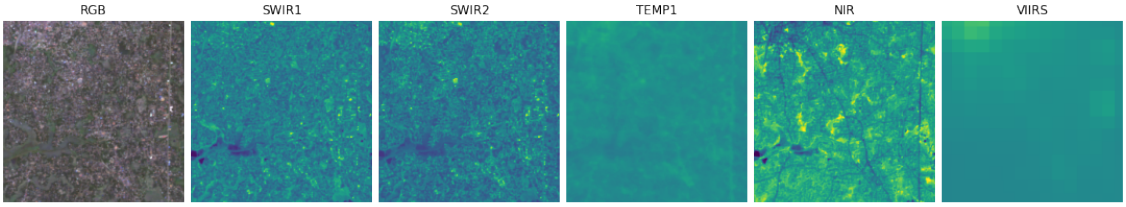


Figure 4.4: Satellite Image

Multi-spectral Images: Satellite images are stored as rasters [29]. These raster images contain seven bands. These bands are called multispectral (MS) bands. Multispectral bands consist of 7 bands which are RED, GREEN, BLUE, NIR (Near Infrared), SWIR1 (Short wave Infrared 1), SWIR2 (Short wave Infrared 2), and TEMP1 (Thermal). Multispectral imaging refers to spectral imaging methods that provide images that correspond to at least a few spectral channels. All these bands have a spatial resolution of 30 m/pixel. While some multispectral imaging devices (also called multispectral cameras) are used on space satellites and aeroplanes, there are also hand-held devices as well as imaging devices installed in industrial settings. Multispectral cameras are frequently tailored to individual applications, especially in terms of the spectral bands utilized.

Nightlight Images:

Nightlight images are captured through Visible Infrared Imaging Radiometer Suite (VIIRS) from 2015 to 2017 [55]. The use of remote sensing of nocturnal light emissions provides a unique perspective on some of these human activities. The Visible Infrared Imaging Radiometer Suite (VIIRS) instruments aboard NASA/Suomi NOAA's National Polar-orbiting Partnership (Suomi NPP) and NOAA-20 satellites provide global daily measurements of nocturnal visible and near-infrared (NIR) light suitable for Earth system science and applications research. Data from the VIIRS Day/Night Band (DNB) is used to estimate population, analyze electrification of rural areas, monitor disasters and conflict, and better understand the biological effects of growing light pollution [37].

Using the Landsat archives available on Google Earth Engine, we obtained Landsat surface reflectance and nighttime lights (nightlights) photos centred on each cluster region [23]. We used 3-year median composite Landsat surface reflectance photos of India collected by the Landsat 7 and Landsat 8 satellites. For compositing, we used three years: 2015–17. Each composite is made by taking the average of all cloud-free pixels available during three years. The use of three-year composites was motivated by two factors. First, multi-year median compositing is a successful strategy for gathering clean satellite data in similar applications 26; nevertheless, even with 1-year compositing, we continued to see the significant influence of clouds in some places due to defects in the cloud mask. For the images of our nightlights, we also developed 3-year median composites for comparison. For the 2015–17 composites, VIIRS pictures are used as nightlight images.

Both MS and NL pictures were processed in Google Earth Engine and exported as 255 x 255 tiles, which were then centre-cropped to 224 x 224, the input size of our CNN architecture, covering 6.72 km on each side (30 m Landsat pixel size = 6.72 km). Fig 4.5 shows the normalized multispectral and nightlight bands of a satellite image.

4.1.2 Data Labelling

We used a python API to export Landsat satellite image composites from Google Earth Engine to Google cloud. The images are saved in gzipped TFRecord format (*.tfrecord.gz). DHS surveys with latitude, longitude and IWI index are used to download the satellite images. Satellite images for corresponding latitude and longitudes are downloaded. The downloaded images contain metadata information such as latitude, longitude and IWI. IWI is used as a label for each image, as our objective is to fill the data gap by identifying the missing wealth indices for different regions in the survey. The downloaded satellite images consume a significant amount of storage space, and it is difficult to store them on a personal computer. Therefore, these images are stored in a Google bucket, which is then transferred to SNIC for further data preprocessing and analysis. In this thesis, we haven't done explicit data labelling, as the metadata of the image already had the labels embedded in it.

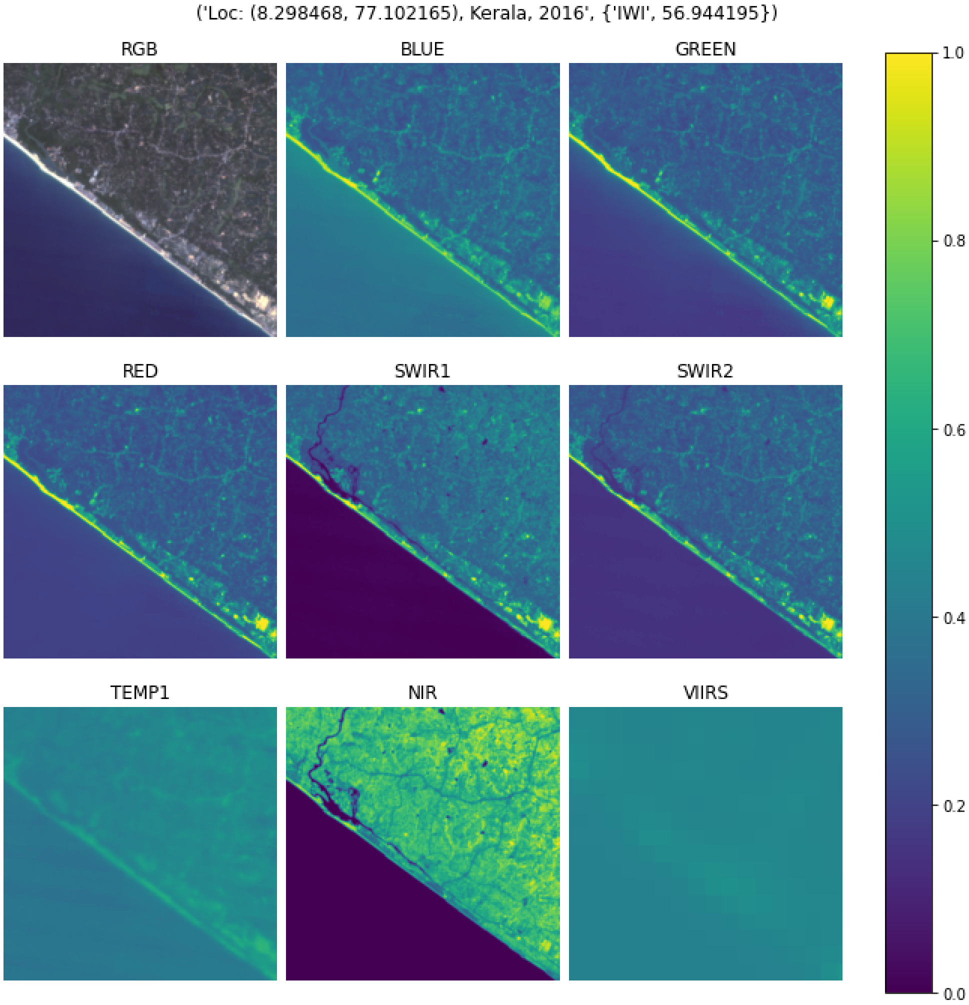

('Loc: (8.298468, 77.102165), Kerala, 2016', {'IWI', 56.944195})
RGB
BLUE
GREEN
RED
SWIR1
SWIR2
TEMP1
NIR
VIIRS
1.0
0.8
0.6
0.4
0.2
0.0

Figure 4.5: Satellite Image with multispectral bands and nightlight band

4.1.3 Data Cleaning

Our dataset consists of, 30789 cluster points from the DHS survey. We then verified downloaded satellite images based on the GPS locations of samples in the DHS survey. This verification is performed to check if the fields in the TFRecords match with the original CSV files generated through the DHS survey. As a part of data cleaning, images with incomplete bands are removed for maintaining the input data quality. After this, our dataset will have only the images will all seven bands required for a multispectral image and one band required for a nightlight image. Consequently, the total number of points is reduced from 30789 to 30787.

4.1.4 Dataset Preparation

To prepare the dataset, we split each monolithic TFRecord file exported from Google Earth Engine into one file per record using a python script so that we can easily shuffle the data. With the TFrecords in one monolithic file, it is impossible to shuffle the order. However, with splitting, we can shuffle the order of the files, which allows us to approximate shuffling the data as we have access to the individual TFRecords. With the increased number of splits, the better is the approximation of shuffling the data. In this study, we have thousands of training examples saved, and we want to repeatedly run them through a training process. Furthermore, for each repetition of the training data (i.e. each epoch), we have to load the data in completely random order. Splitting a monolithic TFRecords file into multiple files has essentially 2 more advantages. The first advantage is that files can be spread across multiple servers, processing several files from different servers in parallel will optimize bandwidth usage (rather than processing one file from a single server). This can improve performance significantly compared to processing the data from a single server. The second advantage is that huge files can be difficult to manage: in particular, transfers are much more likely to fail. Moreover, it's harder to manipulate subsets of the data when it's all in a single large file. Finally, as we were using the GCS bucket (Google Cloud bucket) the amount of throughput can be tremendously increased by having multiple files and thus multiple streams and the TPU sit less in the ideal state. Therefore, the splitting was performed and the single files per record were used for further processing and analysis.

4.1.5 Data Exploration

In the data exploration stage, the DHS survey data, which is downloaded from the DHS program website, is examined to detect the dirty data problems such as ill-formatted data, missing data, duplicate data and erroneously parsed data. After data examination, clusters without georeferencing coordinates are removed. Other columns in the CSV file of the DHS survey are checked in detail to understand the features relevant in the context of this study, and irrelevant columns are removed. For instance, country code and continent are the same for all the entries as we have taken the data for a single country, India. The month of the interview was also removed, as it has no relevance in predicting IWI value. IWI_kurtosis, IWI_variance and IWI_skewness were also removed as IWI_mean was a more meaningful label in the context of poverty prediction. After analyzing the data, we found that the ClusterID column was delivering the information contained in the RegionID column. Therefore, we retained the ClusterID and removed RegionID. Thus, the final CSV file obtained has columns such as ClusterID, Latitude, Longitude, Country_year, Year and IWI_mean. This IWI_mean value is considered as the label IWI. Total points in the DHS survey for India are plotted as a scatter-plot on the map of India. Fig 4.3 illustrates the DHS cluster points plotted on the map of India. From the figure, it can be observed that the minimum IWI value for India is around 4 and the maximum is 81.

The total number of clusters per state is taken from the DHS survey. It can be noted that Uttar Pradesh has got the highest count of clusters, 3950 and Lakshadweep have got the lowest number of clusters, 32. Fig 4.6 shows the number of counts per state.

Fig 4.7 shows a box-plot of IWI distribution in each state and the union territories. From the box plot, it can be observed that Lakshadweep and Kerala have a mean

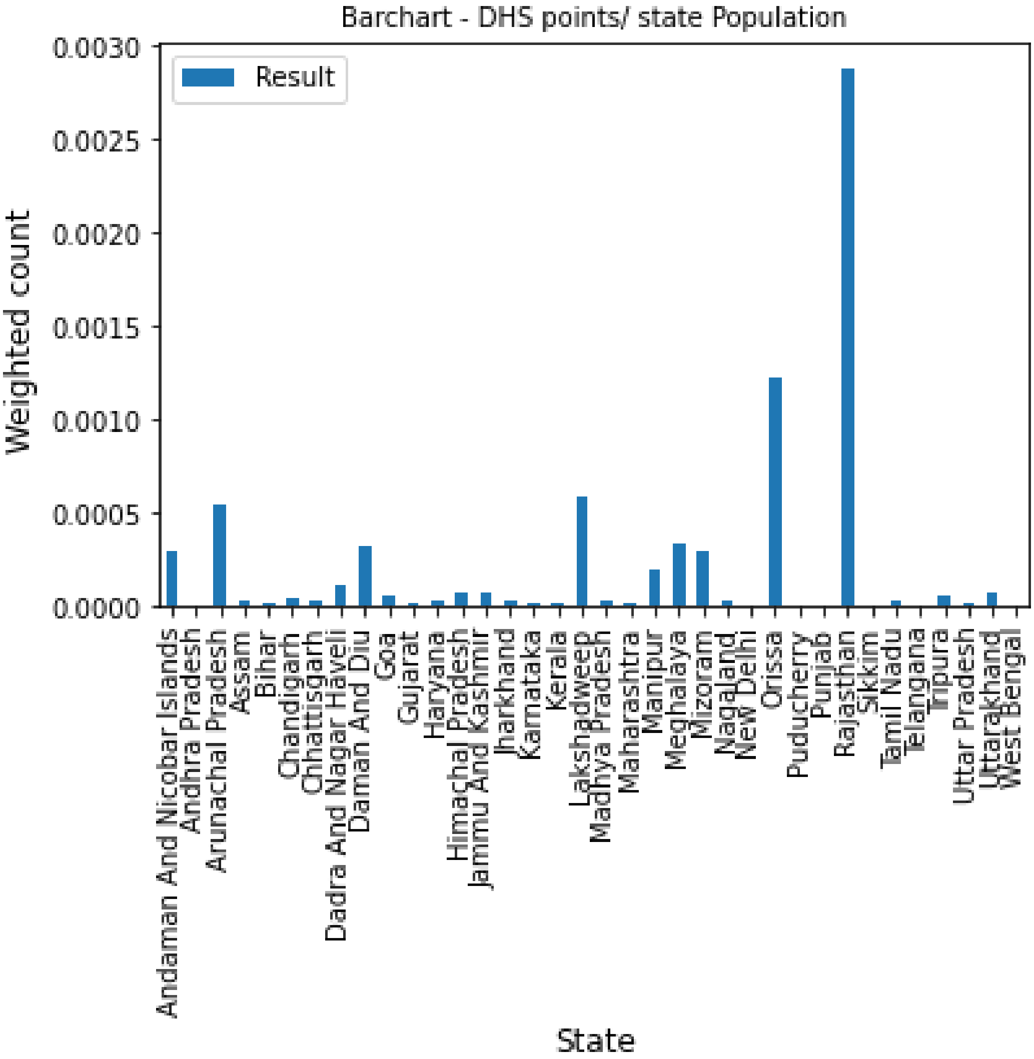


Figure 4.6: Number of DHS survey points per state population.

IWI score of 63. However, Kerala is a state and Lakshadweep is a union territory. When we sort the IWI score state-wise, the state with the highest IWI score is Kerala (67.3), followed by Goa (61.5) and Punjab (60.6). Union territory with the highest IWI score is Lakshadweep (67.8) followed by Chandigarh (62.8) and New Delhi (62.7). The lowest IWI score is for Bihar (32.08), followed by Jharkhand (35) and Assam (38.2).

The total number of cluster points in the urban area is less than the clusters in the rural area. The total number of cluster points in the urban area is 9185 and in the rural area is 21602. Fig 4.8 shows the fraction of urban cluster points among the various states of India. It should be noted that for most of the states' the fraction of urban cluster points is below 0.4. However, union territories like Chandigarh, Daman and Diu, Lakshadweep, New Delhi and Puducherry has more urban area compared to the states of India. One possible reason for this difference is that the size of union territories is comparatively smaller than states in India.

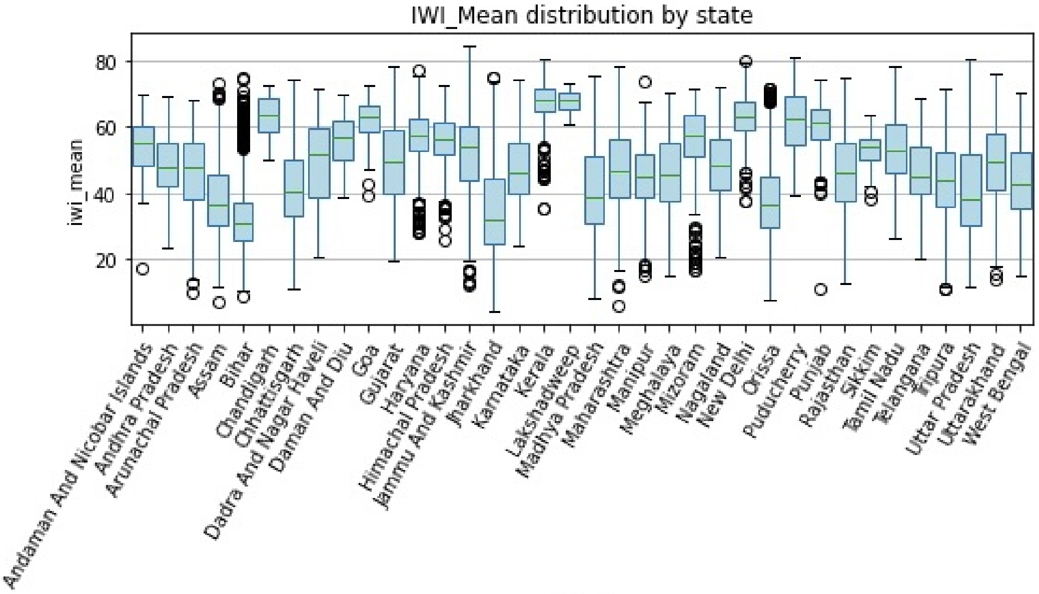


Figure 4.7: IWI distribution state wise

Fig 4.9 shows the satellite image for the location with a high IWI value. Fig 4.10 shows the satellite image for the location with a low IWI value.

4.1.6 Data Splitting

Households surveyed by DHS is grouped into clusters. Our goal is to train models to predict the average cluster wealth-index with maximum accuracy given a satellite image of the cluster. To train our models, we assign the clusters into training (train), validation (val), and test (test) splits. However, we do not arbitrarily assign clusters to splits because many clusters are located very close to each other such that their satellite images overlap. If one cluster was put in train and a nearby cluster was put into test, this may constitute "peaking" at the test set. This is not what we want. Instead, we want our model to be generalizable, able to estimate the cluster wealth-index in geographic regions that the model has not necessarily seen before. Thus, we have to take special care that the satellite images between splits do not overlap. We do this through 2 separate approaches: "out-of-state" and "in-state". For "out-of-state" split, we assign an entire state to a split, so naturally there is no overlap between splits. For "in-state" split, we allow different clusters within the same state to be assigned to different splits, taking care that their satellite images have no overlap.

Out-of-state Split:- In out of state split, the entire state is assigned to train, val or test splits. Thus, all the points in a state will be included in one and only one split, namely train, val or test split.

In-state Split:- For "in-state" splits, we allow different clusters within the same state to be assigned to different splits, taking care that their satellite images have no overlap. DBSCAN algorithm was used to group villages with overlapping satellite images, then ordered the groups in decreasing order by the number of villages per group, then greedily assigned each group to the fold with the fewest villages.

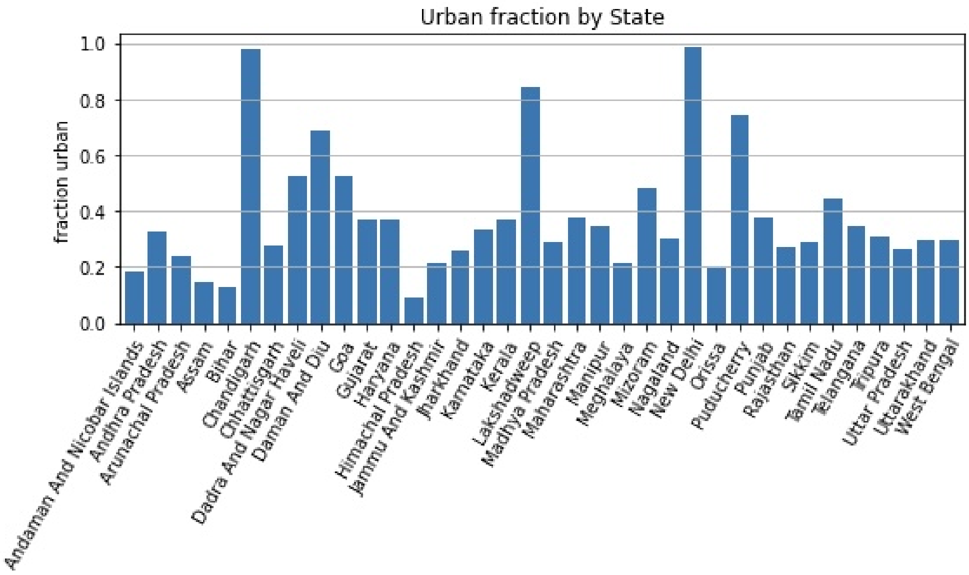


Figure 4.8: Urban Fraction in Indian administrative units

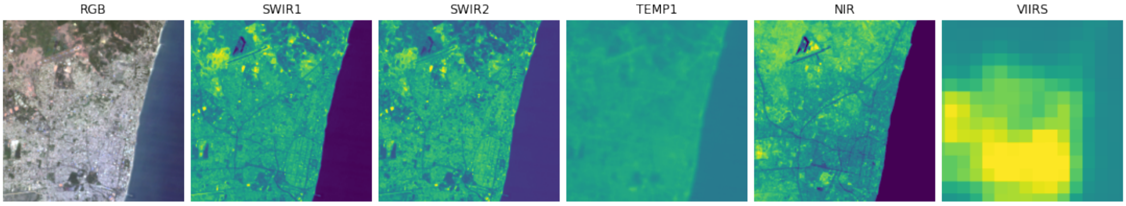


Figure 4.9: Satellite Image with high IWI value. The location is Puducherry (11.950013, 79.822952) and the IWI score is 81

DBSCAN Algorithm In unsupervised learning, there won't be any label associated with data points. Clustering is an unsupervised learning method to group data with similar data points. DBSCAN algorithm is a clustering algorithm. DBSCAN stands for density-based spatial clustering of applications with noise. Epsilon and pinpoints are the two main parameters used in DBSCAN algorithm.

- **Epsilon (ϵ)** :- ϵ is used to specify the neighbourhood. That is if the distance between two points is less than or equal to ϵ, these points are considered to be in the same neighbourhood, otherwise considered as a different neighbourhood.
- **minPoints(n)** :- The smallest number of data points required to define a cluster.

Steps of DBSCAN Algorithm

1. Classify the points.
2. Discard noise.
3. Assign cluster to a core point.
4. Color all the density connected points of a core point.

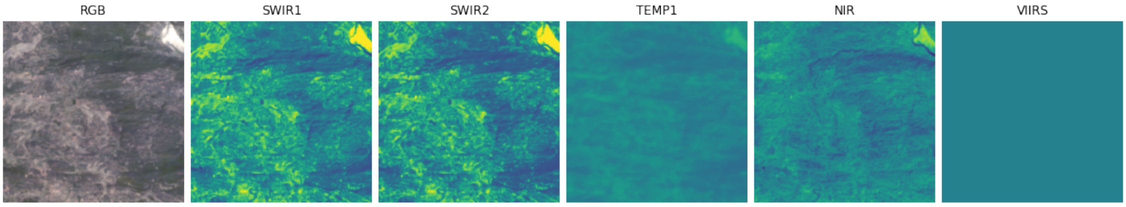


Figure 4.10: Satellite Image with Low IWI value. The location is Jharkhand (22.879873, 85.438171) and the IWI score is 4.2

5. Color boundary points according to the nearest core point.

The algorithm begins by randomly selecting a point (x) (one record) from the dataset and assigning it to cluster 1. In the next step, it counts the number of points located with the epsilon (ϵ) distance from the point x. If the number of points counted is greater than or equal to n i.e. minPoints, this point will be considered a core point, and all of these neighbours will be pulled into the same cluster 1. In the next step, it will look at each member in cluster 1 and will find the ϵ neighbours to the cluster respectively. In some cases, some members of cluster one will have n or more ϵ neighbours, in that case, it will enlarge cluster 1 by adding those neighbours to cluster 1. Cluster 1 will continue to grow until there are no more examples to add. In the latter scenario, it will select a point from the dataset that does not belong to any cluster and place it in cluster 2. This will continue until all cases are assigned to a cluster or designated as outliers.

For DHS survey data, we split the data into 5 folds for cross-validation. For the DHS out-of-state split, we manually split the 36 states into 5 folds namely A, B, C, D, and E. As described below, models were trained using cross-validation to select optimal hyperparameters. Each model was trained on 3-folds, validated on the 4th fold, and tested on the 5th fold. For DHS in-state split, we split the 30,787 points into 5 folds such that there is no overlap in satellite images of the villages between any fold, where the overlap is defined as any area (however small) that is present in both images. For instance, both MS and NL images are 255 x 255 tiles, which are then centre-cropped to 224 x 224, the input size of CNN architecture, covering 6.72 km on each side (30 m Landsat pixel size = 6.72 km). Thus, two or more images can have a set of pixels common that comes at the intersection. If one of these images is used for training and the others are used for testing, the model will see this common set of pixels and therefore results in peaking. We used the DBSCAN algorithm to group together villages with overlapping satellite images. We first sorted the groups by the number of villages per group in decreasing order, and then greedily assigned each group to the fold with the fewest villages.

4.1.7 Model Training

We have trained three models and all three of them make use of the CNN model with ResNet-18 architecture. From the literature, we understood that most existing CNN models are designed to work with 3-channel RGB images and thus are not directly compatible with multi-band satellite images. Thus, we adapted the existing architecture of ResNet to work on multi-band satellite images. We selected ResNet-18 architecture v2, with preactivation [27], because of its balance of compactness and high accuracy on the ImageNet image classification challenge [42]. We used SNIC (Swedish National Infrastructure for Computing) which is a national research infrastructure that provides large-scale high-performance computing resources, storage capacity, and better user assistance to Swedish researchers for training and testing all the models.

4.1.7.1 Baseline model

Baseline models are used to establish a meaningful point of reference, which is usually a model that is there in the state-of-the-art literature [52]. For this thesis, we train a CNN model with ResNet-18 architecture having 18 layers on nightlight images and evaluate the performance. This ResNet-18 trained on only nightlight images is used as a baseline model. Thus, the performance from this model is considered the baseline performance for the study. Fig 4.11 shows the architecture of the baseline model used in the thesis.

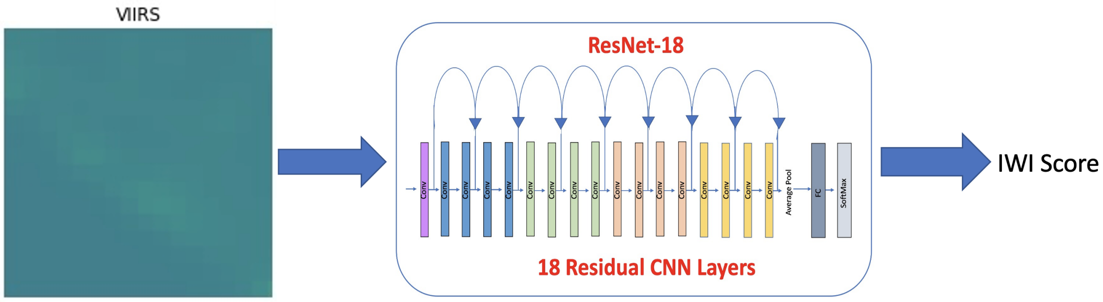


Figure 4.11: Baseline Model - ResNet-18 trained only on nightlight images

4.1.7.2 Deep CNN trained on Nightlight and Multispectral images

We train two deep CNN models apart from the baseline model. The first model is a deep CNN model with ResNet-18 architecture trained on multispectral images as shown in fig 4.12. As a second model, we trained a ResNet-18 architecture the predict the IWI score. Unlike the baseline model and the first model, we use both nightlight images and multispectral images as input to the ResNet-18 architecture. The first convolution layer of our model is an alteration of the CNN model to accommodate multispectral images and nightlight images. The final layer of the model is also changed to get a single final scalar output. We have trained the model with the Adam optimizer and with the mean squared error loss function. The batch size is fixed as 64 for each epoch. Learning rate decay is set as 0.96 [55] and the model has trained 150 epochs for in-state and 200 epochs for out-of-state splits. We have performed a grid search with different learning rates such as

1e-2, 1e-3, 1e-4, 1e-5 and L2 weight regularization such as 1e-0, 1e-1, 1e-2, 1e-3 for obtaining the best combination. Furthermore, there are no bias terms in the ResNet-18 architecture because each convolutional layer is preceded by a batch-normalization layer. As a regularization technique, the model with the highest r^2 (Pearson's correlation coefficient) on the validation set across all epochs is used as the final model for comparison. After the training, the best model is selected based on the model's performance on the validation fold. Fig 4.13 shows the ResNet-18 architecture trained on both multispectral and nightlight images. The baseline model (deep CNN trained on NL bands) and first model (deep CNN trained on MS bands) are used for selecting optimal hyperparameters for the single-frame model.

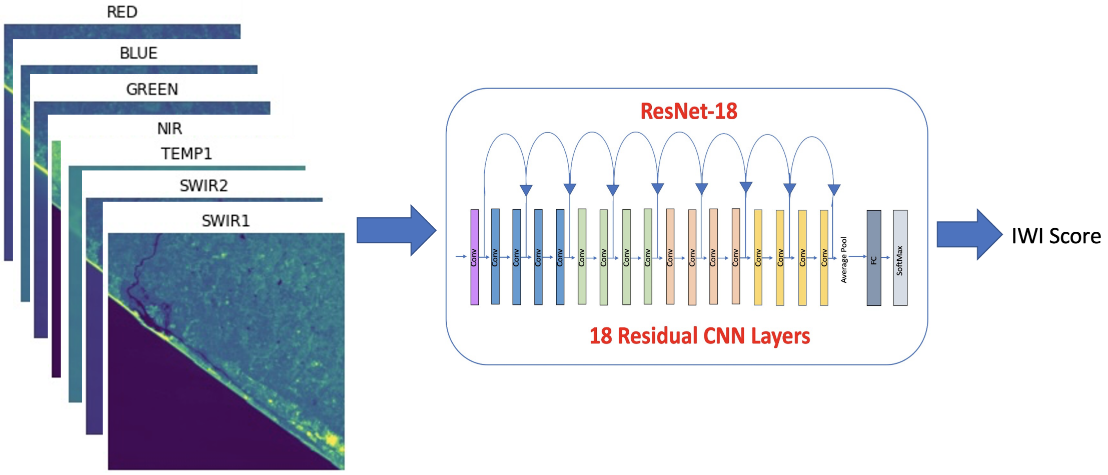


Figure 4.12: ResNet-18 with only multispectral images as input

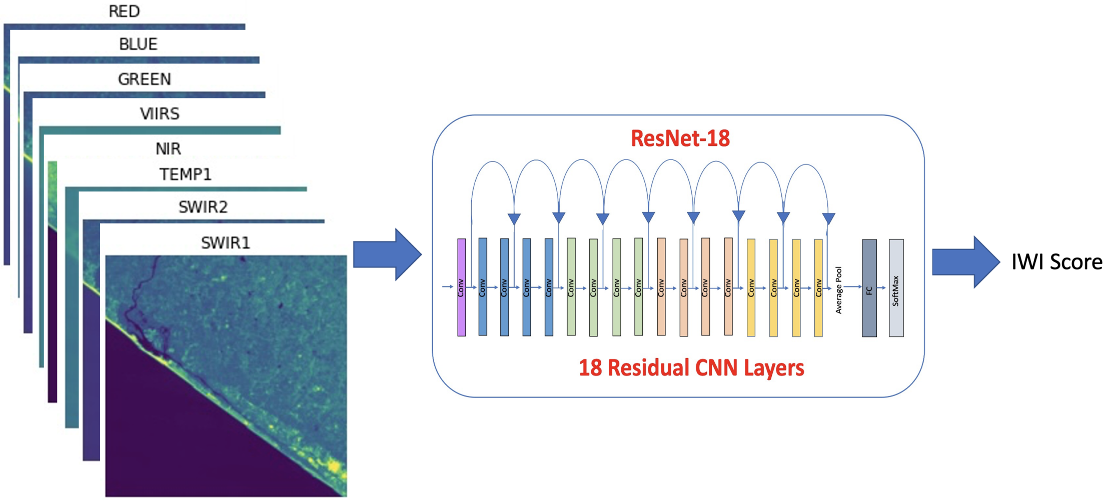


Figure 4.13: ResNet-18 with both multispectral images and Nightlight images as input

4.1.7.3 Single frame model

Single frame model [34] is a combination of two deep CNNs with ResNet-18 architectures. We use a single frame model as the final DL model for predicting IWI in our study. Here, one CNN with ResNet-18 architecture is used for training multispectral images and the other CNN with the same ResNet-18 architecture is used for training nightlight images. i.e, One ResNet-18 architecture is trained on nightlight images and the other ResNet-18 on daylight images. In our case, the single-frame model is a combination of our baseline model and ResNet-18 trained on MS bands. This is different from the second model, deep CNN trained on nightlight and multispectral images. Because we want to analyse and confirm whether these two approaches result in the same performance or not. Since the number of bands for the satellite images (eight bands) is different from normal images (three bands), the first convolution layer of the CNN is modified, making it suitable to fit all the eight bands of satellite images. Further, we also modify the final convolutional layer to output a scalar for regression. Fig 4.14 illustrates the structure of the Single Frame model. The images are augmented by random horizontal and vertical flips to prevent overfitting. The brightness and contrast of the multispectral bands are also subjected to random adjustments using the python library pillow.

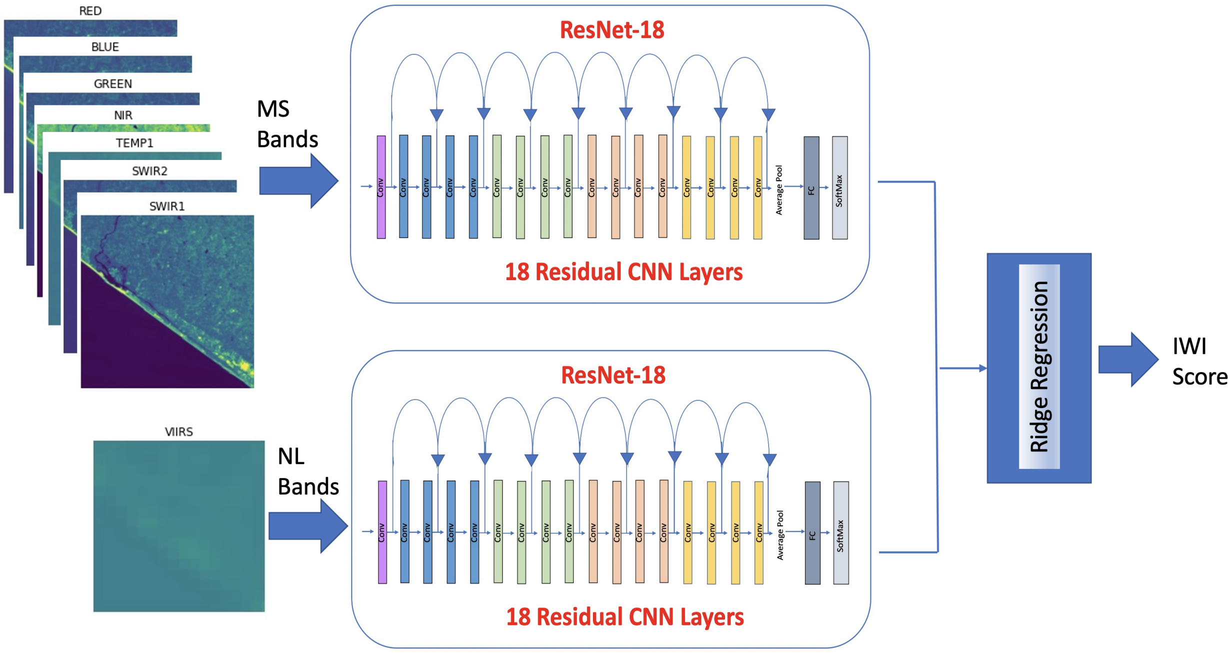


Figure 4.14: Single Frame Model

4.1.8 Cross-Validation and Hyper Parameter Tuning

We trained five distinct models, each with a different test fold, for each of the input band combinations MS, MS+NL, and NL. We split the entire dataset into five folds, A, B, C, D and E. Three of the folds were used to train the models, with the fourth serving as a validation set for early stopping and tweaking additional hyperparameters. We have performed a grid search with different learning rates such as 1e-2, 1e-3, 1e-4, 1e-5 and L2 weight regularization such as 1e-0, 1e-1, 1e-2, 1e-3 for obtaining the best combination. We used Adam optimizer with mean squared-error loss function as it has faster computation time, and require fewer parameters for tuning. We chose a mini batch size, 64 less than the total number of samples, to reduce memory usage and for faster training. Learning rate decay is set as 0.96 [55] and the model is trained 150 epochs for in-state and 200 epochs for out-of-state splits. The number of epochs were selected randomly, as there are no hard and fast rule for selecting the batch size and number of epochs. Since the ResNet-18 architecture has a batch-normalization layer following each convolutional layer, there are no bias terms. As a regularization technique, the model with the highest r^2 (Pearson's correlation coefficient) on the validation set across all epochs is used as the final model for comparison. i.e. we did the following steps, not necessarily in the same order.

- Train on A, B and C, validate on D for selecting hyperparameters and test on the remaining fold E.
- train on B, C and D, validate on E for selecting hyperparameters and test on the remaining fold A.
- train on C, D and E, validate on A for selecting hyperparameters and test on the remaining fold B.
- train on D, E and A, validate on B for selecting hyperparameters and test on the remaining fold C.
- train on E, A, and B validate on C for selecting hyperparameters and test on the remaining fold D.

Thus we have test results of all five folds and consequently test results for the entire dataset. After the CNNs were trained, we used ridge regression with hold-one-group-out cross-validation to fine-tune the final fully connected layer. We fine-tuned the final layer individually for each test fold under the in-state split scenario, utilizing data from all other folds. As a result, the CNNs' convolutional layers essentially saw data from four of the five folds, whereas the final layer saw data from every fold except the test fold.

For optimal generalization performance on unseen data, cross-validation should be used to tweak hyperparameters for machine learning models. Hold-one-group-out cross-validation, on the other hand, is unreasonably time-consuming due to the significant computer resources required to train deep neural networks (where in our setting, each group is a state). Therefore, we utilized hold-one-fold-out cross-validation to tune the regularization parameter for training the weights in the final fully connected layer, and simply used hold-one-group-out cross-validation to tune the hyperparameters for the body of the CNN.

5
Results

Model Evaluation which is the final step in the deep learning workflow is explained in this chapter. Model Evaluation is organized as the performance of the three different models the baseline model, deep CNN trained on MS and NL and the Single frame model on inputs such as NL, MS, and MS+NL.

5.1 Performance of baseline model

For this study, we use deep CNN with ResNet-18 architecture having 18 layers trained on nightlight images as a baseline model. The baseline model was chosen in such a combination as most of the existing literature was using nightlight bands for predicting economic well-being or poverty. We chose ResNet-18 architecture, as the Single Frame model is a combination of 2 ResNet-18 architectures. Moreover, most of the existing studies use nightlight bands for predicting poverty. Therefore, we wanted to compare the results with one ResNet-18 trained with nightlight images.
The model is trained with different learning rates (1e-2, 1e-3, 1e-4, 1e-5). Five-fold cross validation is performed on both in-state and out-of-state split and evaluated using the metric r^2, which is the squared Pearson's coefficient of correlation. Table 5.1 shows the r^2, R^2, values of baseline model in various folds for out-of-state split, and Table 5.2 shows the r^2, R^2, values of baseline model in various folds for in-state split.
From the table 5.1, it can be seen that, the baseline model with out-of-state split has $r^2 = 0.42$ and $R^2 = 0.32$. This means that the correlation between actual IWI values and baseline predicted IWI values are 42% correlated for fold A. Further, the strength of the model or goodness of fit of the baseline model is 0.32. i.e. the quality of fit of baseline model on fold A is 0.32.

Table 5.1: Performance of baseline model with out-of-state split

Folds	r^2	R^2
A	0.42	0.32
B	0.37	0.28
C	0.36	0.33
D	0.41	0.41
E	0.16	0.1
Mean	0.34	0.29

From the table 5.2, it can be seen that, the baseline model with in-state split has $r^2 = 0.41$ and $R^2 = 0.39$. This means that the correlation between actual IWI values and baseline predicted IWI values are 41% correlated for fold A. Further, the strength of the model or goodness of fit of the baseline model is 0.39. i.e. the quality of fit of the baseline model on fold A is 0.39. When we consider the mean performance value for both types of splits, it can be seen that with an in-state split, we get a stronger correlation between actual and predicted IWI values. Furthermore, the quality of fit of the baseline model is better on in-state spit compared to the out-of-state split.

Table 5.2: Performance of baseline model with in-state split

Folds	$\boldsymbol{r^2}$	$\boldsymbol{R^2}$
A	0.41	0.39
B	0.44	0.38
C	0.44	0.44
D	0.28	0.24
E	0.41	0.37
Mean	0.40	0.36

5.2 Performance of Deep CNN model trained on MS and MS+NL bands

A Deep CNN model trained with only multispectral bands is the first model we developed for performance comparison. Table 5.3 shows r^2, R^2, values of deep CNN model in various folds for out-of-state split, and Table 5.4 shows the r^2, R^2, values of deep CNN model in various folds for the in-state split.

From the table 5.3, it can be seen that, the model with out-of-state split has $r^2 = 0.50$ and $R^2 = 0.48$. This means that the correlation between actual IWI values and model predicted IWI values are 50% correlated for fold A. Further, the strength of the model or goodness of fit of the model is 0.48. i.e. the quality of fit of the model on fold A is 0.48. The performance of this model is better than the baseline model.

From the table 5.4, it can be seen that, the model with in-state split has $r^2 = 0.60$ and $R^2 = 0.60$. This means that the correlation between actual IWI values and predicted IWI values is 60% correlated for fold A. Further, the strength of the model or goodness of fit of the model is 0.60. i.e. the quality of fit of the model on fold A is 0.60. When we consider the mean performance value for both types of splits, it can be seen that with an in-state split, we get a stronger correlation between actual and predicted IWI values. Furthermore, the quality of fit of the model is better on in-state spit compared to out-of-state split.

The second model is different from the baseline model in terms of the input given to it. The input to the second model is both multispectral and nightlight images. It can be observed that this model that has input MS and NL images has a better performance compared to the baseline model on all the five folds for both in-state and

Table 5.3: Performance of deep CNN model trained on MS bands with out-of-state split

Folds	r^2	R^2
A	0.50	0.48
B	0.48	0.39
C	0.48	0.48
D	0.39	0.34
E	0.37	0.18
Mean	0.44	0.37

Table 5.4: Performance of deep CNN model trained on MS bands with in-state split

Folds	r^2	R^2
A	0.60	0.60
B	0.59	0.60
C	0.56	0.56
D	0.56	0.56
E	0.55	0.49
Mean	0.57	0.56

out-of-state splits. However, it can be also observed that the model's performance on folds C and D are less compared to folds A, B and E. Table 5.5 shows the r^2, R^2, values for five-folds with out-of-state split.

From the table 5.5, it can be seen that, the model with out-of-state split has $r^2 = 0.57$ and $R^2 = 0.56$. This means that the correlation between actual IWI values and model predicted IWI values are 57% correlated for fold A. Further, the strength of the model or goodness of fit of the model is 0.56. i.e. the quality of fit of the model on fold A is 0.56. The performance of this model is also better than the baseline model.

From the table 5.5, it can be observed that folds C and D have lower r^2 values compared to the other three folds. As the split was done manually without considering the variance of data which could have resulted in the comparatively lower r^2 values for folds C and D. To overcome the discrepancy, we did an in-state split. For in-state split, data is not split according to the states. Instead, data is split in such a way that no neighbouring data points have an overlap. Further, we allowed different clusters within the same state to be assigned to different splits, and extreme care was taken to avoid overlap in the satellite images. To perform an in-state split, we used the DBSCAN algorithm. Table 5.6 shows the r^2, R^2, values for five-folds with in-state split.

From the table 5.6, it can be seen that, the model with in-state split has $r^2 = 0.62$ and $R^2 = 0.62$. This means that the correlation between actual IWI values and

Table 5.5: Performance of Deep CNN model trained on MS+NL bands with out-of-state split

Folds	r^2	R^2
A	0.57	0.56
B	0.54	0.54
C	0.34	0.31
D	0.35	0.27
E	0.50	0.50
Mean	0.46	0.44

baseline predicted IWI values are 62% correlated for fold A. Further, the strength of the model or goodness of fit of the model is 0.62. i.e. the quality of fit of the baseline model on fold A is 0.62. When we consider the mean performance value for both types of splits, it can be seen that with an in-state split, we get a stronger correlation between actual and predicted IWI values. Furthermore, the quality of fit of the model is better on in-state spit compared to out-of-state split.

Table 5.6: Performance of Deep CNN model trained on MS+NL bands with in-state split

Folds	r^2	R^2
A	0.62	0.62
B	0.59	0.59
C	0.57	0.57
D	0.57	0.57
E	0.55	0.50
Mean	0.58	0.57

Table 5.7 and 5.8 shows the mean values of r^2 and R^2 for baseline model and deep CNN model trained on MS bands, MS+NL bands with out-of-state split and in-state split respectively.

Table 5.7: Mean of performance metrics of models with out-of-state split

Model	**Mean** r^2	**Mean** R^2
Baseline model	0.34	0.29
Deep CNN trained on MS Bands	0.44	0.37
Deep CNN trained on MS and NL Bands	0.46	0.44

From the tables 5.7 and 5.8, it can be observed that the performance of the model on the out-of-state split is extremely poor compared to the performance of the model

Table 5.8: Mean of performance metrics of models with in-state split

Model	Mean r^2	Mean R^2
Baseline model	0.40	0.36
Deep CNN trained on MS Bands	0.57	0.56
Deep CNN trained on MS and NL Bands	0.58	0.57

on the in-state split. The reason for poorer performance may be due to the reason that the out-of-state split was done manually without considering the geographical and demographical distribution of states. To eliminate the problem of overfitting, we have augmented the images with random horizontal and vertical flips. The non-nightlights bands are also subject to random adjustments to brightness (up to 0.5 standard deviation change) and contrast (up to 25% change). Moreover, we did an evaluation using held-out locations that the model did not use in training, an approach that limits overfitting as well as replicates the real-world setting of making predictions where ground data do not exist. Consequently, overfitting couldn't be a reason for the better performance of the in-state split. Therefore, the out-of-state split is not considered from this point onwards. i.e. for the single-frame model, we have used only the in-state split. Further, the deep CNN model trained on MS bands (referred to as the first model) together with the baseline model (deep CNN model trained on NL bands) is used for hyperparameter tuning.

5.3 Performance of Single Frame Model

The single-frame model is a combination of two deep CNN models. Two ResNet-18 models trained separately on the Landsat bands and nightlights bands, respectively, and joined the models in their final fully connected layer. The first convolution layer of ResNet-18 architecture is modified separately, to fit all the bands in the multi-spectral image, as well as nightlight bands. The final convolutional layer is also modified to output a scalar for regression, which is the IWI value. The results from cross-validation and hyperparameter tuning are used as input to the Single Frame model. i.e. we select the best combination of the fold, seed, keep and hyperparameters such as learning rate and L2 weight regularization based on the performance of the baseline (ResNet-18 with NL bands) and first deep CNN model (ResNet-18 with MS bands) on the validation fold. The following results are obtained for the Single Frame model after evaluating the performance on the test set.

- r^2, (weighted) squared Pearson correlation coefficient = 0.59
- R^2, (weighted) coefficient of determination = 0.59

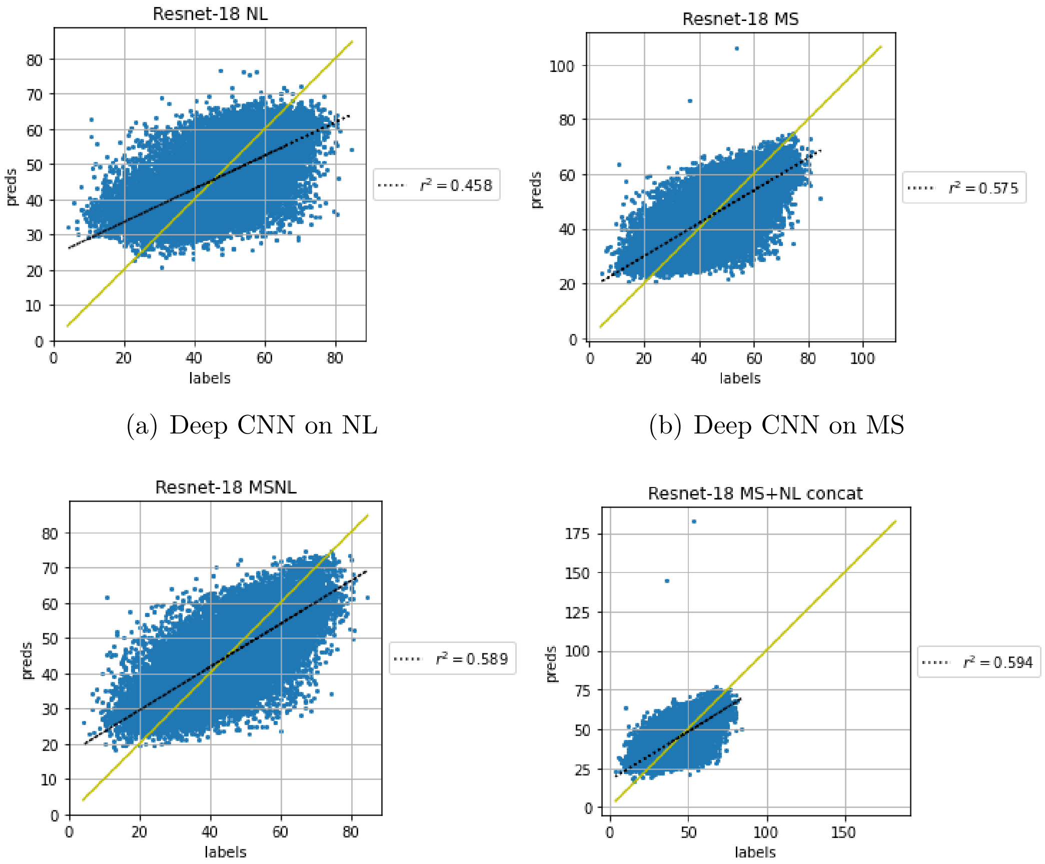
These results show that the quality of fit (R^2) of a single frame model on the data is 0.59, which is higher compared to all the remaining models discussed in this thesis. Further, the single-frame model explains 59% of IWI variance.

Fig 5.9 shows the consolidated performance of all the models discussed in this thesis, which can be used to compare the performances of different models on the in-state split. Although all models are predictive of average cluster level IWI, the best performance is delivered by the single-frame model, which explains 59% of IWI variance.

Table 5.9: Consolidated Results from all four CNN models

Model	$\boldsymbol{R^2}$	$\boldsymbol{r^2}$
Single Frame Model	0.594	0.592
ResNet -18 on MS & NL	0.588	0.587
ResNet -18 on MS	0.574	0.573
Baseline Model	0.457	0.456

For better visualization of predicted values and ground truth, model predictions are plotted against ground truth asset wealth indices (IWI). The explained variance, r^2 of the models, is marked with a black line.



(a) Deep CNN on NL

(b) Deep CNN on MS

(c) Deep CNN on MS and NL

(d) Single frame Model

Figure 5.1: Model predictions plotted against ground truth asset wealth indices

5.4 Performance Comparison with related works

From table 5.10, it can be seen that all the models in the related work are used for predicting different indicators of poverty/health conditions. Moreover, the datasets used are also different across the studies. Two studies are using only few states in India for training the model. The study by Pandey et al. describes a classification model rather than a regression model. Further, the study by Chi et al. [16] has trained a model on data from 56 low and middle income countries to predict micro estimates of wealth. However, they have not tested their model on data from India. Instead, they have built a linear regression model to predict the model error. Therefore, comparing the performance of these models with our model is a bit difficult. Nevertheless, if we simply consider the performance of a Single frame model described in this thesis, we can see that the performance is not bad/poor compared to any of the related works. As a result, the reported results should be regarded as indicative of the proposed solution's capabilities rather than the definitive maximum performance of the models.

Table 5.10: Performance Comparison with related works

Study	**Dataset used**	**States included in dataset**	**Performance**
Subash et al.	Nightlight data and GDP	All states	Predict rural poverty with $R^2 = 0.56$
Pandey et al.	Census data and satellite imagery	One state - Uttar Pradesh	Predict income category with avg. classification accuracy 0.91
Daoud et al.	Census data and satellite imagery	Six states - Uttar Pradesh, West Bengal, Bihar, Jharkhand, Punjab and Haryana	Predict 93 health outcomes with R^2 as 0.92 to 0.60 for 21 outcomes; 0.59 to 0.30 for 25 outcomes, 0.29 to 0.00 for 28 outcomes, and 19 outcomes had negative R^2
Chi et al.	DHS Survey data, satellite imagery, mobile phone networks, topographic maps, deidentified connectivity data from Facebook	All states	Predict residual from wealth prediction model with MSE = 0.50
Lee et al.	DHS data and street level imagery	All states	$r^2 = 0.56$
Thesis	DHS data and satellite imagery	All states	$r^2 = 0.59$ and $R^2 = 0.59$

5.5 Hyper Parameter Tuning

All ResNet-18 models except the single-frame model discussed in this thesis are trained with the Adam optimizer and a mean squared error loss function. The batch size is 64, and the learning rate is decayed by a factor of 0.96 after each epoch. The models are trained for 150 epochs (200 epochs for DHS out-of-country). The final model for comparison is the one with the highest r^2 on the validation set across all epochs. This is a regularization approach that is similar to early stopping. We performed a grid search over the learning rate (1e-2, 1e-3, 1e-4, 1e-5) and L2 weight regularization (1e-0, 1e-1, 1e-2, 1e-3) hyperparameters to find the model that performs the best on the validation fold. To prevent overfitting, the images are augmented by random horizontal and vertical flips. The non-nightlight bands are also subject to random adjustments to brightness (up to 0.5 standard deviation change) and contrast (up to 25% change). On the other hand, our pipeline didn't include a post-hoc linear regressor for post-hoc calibration.

Many hyperparameters can influence the performance of a neural network during training. Due to time restrictions, no systematic effort was undertaken to tweak these parameters to improve the performance of the single-frame model discussed in this thesis. As a result, the reported results should be regarded as indicative of the proposed solution's capabilities rather than the definitive maximum performance of the models.

6 Discussion

Our work demonstrates that satellite imagery combined with DHS data can be utilized to produce fairly accurate predictions about the international wealth index (IWI) across India. Despite inaccuracies in the timing of satellite imagery and the location of clusters in the training data, our model works well, and more precise data in either of these dimensions is expected to improve model performance further. For instance, to anonymize the DHS survey, clusters are randomly displaced, which results in the shifting of cluster points. This shifting causes the inexactness of location for the satellite images.

Notably, we show that our model's predictive power declines when a model is trained based on the out-of-state split, in which a model trained on a set of states is used to estimate consumption or assets in another state. One of the possible reasons for this declining performance is that the out-of-state split is done manually without considering the geographical and demographical distribution of states. Differences in economic and political institutions across states could be indirect determinants of livelihoods across settings. For instance, larger states in India with more data points typically have a lower IWI. Consequently, a model trained on larger states cannot accurately predict the IWI of smaller states. In contrast, the in-state split always yields good performance for all the models we tried in this study, suggesting that our approach could be used to fill in the large data gaps resulting from poor survey coverage in many Indian states. Further, our method uses only publicly available data and so is straightforward and nearly costless.

Although our approach is nearly similar to the algorithm applied to the dataset from Africa by Yeh et al., poverty prediction for the dataset from India based on satellite imagery and DHS data has not been done before. To reiterate, all the models in the related work are used to predict different indicators of poverty/health conditions. Moreover, the datasets used are also different across the studies. Two studies are using only a few states in India for training the model. Consequently, we don't have many performance benchmarks for comparison. Through our literature survey, we found five papers that used data from India. However, two of them used census data and satellite imagery, the third one by Subash et al. [48] was using nightlight data and GDP to predict poverty, and the fourth by Lee et al. [31] was using DHS data and street-level imageries for predicting poverty. The first study by Pandey et al. [38] uses both census data and satellite imagery, but is not comparable with our study. Because, they have only taken one state, whereas our study has considered the data from the entire country. Besides, they have developed a classification model and our study aims to predict the IWI index of missing places across India. The second paper by Daoud et al. [17] implements a DL model with the data from six Indian states (constituting 30% of the Indian landmass) to predict 93 health outcomes out of which they obtained R^2 of 0.92 to 0.60 for 21 development outcomes; 0.59 to 0.30 for 25 outcomes; and 0.29 to 0.00 for 28 outcomes, and 19 outcomes had negative R^2. Although this model has a better R^2 value for some health predictors, it cannot be directly compared with our study, as this thesis has considered DHS data from entire India and satellite images from Landsat-7 to predict IWI. The fourth paper by Lee et al. [31] implements a DL model and obtains r^2 of 0.56, which is less than the performance obtained in this thesis. The final paper by Chi et al. [16] has trained a model on data from 56 low and middle-income countries to predict micro estimates of wealth. However, they have not tested their model on data from India. Instead, they have built a linear regression model for predicting the model error when tested on countries other than Togo, Kenya, and Nigeria.

Several studies have been conducted on the dataset from Africa. However, Africa is a continent and India is a country. The DMSP band is not present in satellite imagery of India, while it is there in the dataset from Africa. Further, the closest study by Yeh et al. used the DHS survey as well as the LSMS survey, while in our case, we had only the DHS survey. Furthermore, the dataset from Africa has data from 2009-2017, whereas our study used data for the years 2016-17. New sources of ground truth data, whether from more disaggregated surveys or novel crowdsourced channels, could enable a better evaluation of our model.

Due to time constraints, we were not able to use the transfer learning approach to predict the International wealth index (IWI). We assume that transfer learning models might have yielded better performance compared to our approach. Further, as training deep neural networks demand substantial computational resources, hold-one-out cross-validation is extremely time-intensive, which made us move it to future works. Finally, the results of our approach are not directly comparable to findings from other small area estimate approaches like the results from the study by Pandey et al.

We strongly believe that our approach could have broad application across many scientific domains and may be immediately useful for inexpensively producing granular data on other socioeconomic outcomes of interest to the international community, such as the large set of indicators proposed for the United Nations Sustainable Development Goals (5). Further, our model can also be used to predict the child mortality rate, maternal health, malnutrition, etc.

7
Conclusion

The objective of this study was to train models to predict the average cluster wealth index given a satellite image of the cluster. In the immediate future, increasing amounts of high-resolution satellite imagery become available, and therefore IWI predictions based on satellite imagery would be very helpful to both researchers and policy-makers. Along with the increased popularity of DL models, the relevance of the approach discussed in this study will also increase. Policy-makers and other experts working with poverty prediction can make use of the predicted IWI values for inexpensively producing granular data on other socioeconomic outcomes. We have used two types of data splits, namely out-of-state split and in-state split, in this study. Dataset splitting is performed with extreme care to avoid peaking. We selected optimal hyperparameters after 5-fold cross-validation. Each model was trained on 3-folds, validated on a 4th, and tested on a 5th fold. Hyperparameters for DL models were tuned by cross-validation for optimal generalization performance on unseen data. The single-frame model has the best performance compared to the other two models. However, the single-frame model is only tested with an in-state split, as the out-of-state split yielded poor performance on deep CNN trained on MS and NL bands. Our results are not directly comparable to findings from other small area estimate approaches. In future work, we plan to do an out-of-state split, considering the geographical and demographical distribution of states. Further, we would like to perform a hold-one-out cross-validation, which we couldn't perform due to time and computational resource constraints.

Bibliography

[1] 2014wesp_country_classification.pdf. https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf. (Accessed on 05/20/2022).

[2] Cs 230 - convolutional neural networks cheatsheet. https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks. (Accessed on 12/20/2021).

[3] Demographic and health surveys (dhs): contributions and limitations - pubmed. https://pubmed.ncbi.nlm.nih.gov/8017081/. (Accessed on 12/22/2021).

[4] The dhs program - gps data collection. https://dhsprogram.com/methodology/GPS-Data-Collection.cfm. (Accessed on 05/20/2022).

[5] The dhs program - research topics - wealth index. https://dhsprogram.com/topics/wealth-index/. (Accessed on 12/19/2021).

[6] A gentle introduction to the rectified linear unit (relu). https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/. (Accessed on 05/21/2022).

[7] Least developed countries (ldcs) | department of economic and social affairs. https://www.un.org/development/desa/dpad/least-developed-country-category.html. (Accessed on 05/20/2022).

[8] States uts - know india: National portal of india. https://knowindia.india.gov.in/states-uts/. (Accessed on 12/22/2021).

[9] Wealth indices - global data lab. https://globaldatalab.org/iwi/#:~:text=The%20IWI%20scale%20is%20additive,its%20IWI%20value%20is%200. (Accessed on 05/21/2022).

[10] Oludare Isaac Abiodun, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad. State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11):e00938, 2018.

[11] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6, 2017.

[12] Yahaya Alhassan and Uzoechi Nwagbara. Microfinance and sustainable development in africa, 2021.

[13] Agustin Garcia Asuero, Ana Sayago, and AG Gonzalez. The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, 36(1):41–59, 2006.

[14] Joshua Evan Blumenstock. Fighting poverty with data. *Science*, 353(6301):753–754, 2016.

[15] Hollis B Chenery. Foreign assistance and economic development. In *Capital movements and economic development*, pages 268–292. Springer, 1967.

[16] Guanghua Chi, Han Fang, Sourav Chatterjee, and Joshua E Blumenstock. Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences*, 119(3), 2022.

[17] Adel Daoud, Felipe Jordan, Makkunda Sharma, Fredrik Johansson, Devdatt Dubhashi, Sourabh Paul, and Subhashis Banerjee. Using satellites and artificial intelligence to measure health and material-living standards in india. *arXiv preprint arXiv:2202.00109*, 2021.

[18] Alessandro Di Bucchianico. Coefficient of determination (r 2). *Encyclopedia of Statistics in Quality and Reliability*, 1, 2008.

[19] Gary W Evans. The environment of childhood poverty. *American psychologist*, 59(2):77, 2004.

[20] Bruce Fuller. *Raising School Quality in Developing Countries: What Investments Boost Learning? World Bank Discussion Papers 2.* ERIC, 1986.

[21] Dhaneshwar Ghura, Carlos A Leite, and Charalambos G Tsangarides. Is growth enough? macroeconomic policy and poverty reduction. *IMF Working Papers*, 2002(118), 2002.

[22] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning.* MIT press, 2016.

[23] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202:18–27, 2017.

[24] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48, 2016.

[25] Boris Hanin. Which neural net architectures give rise to exploding and vanishing gradients? *Advances in neural information processing systems*, 31, 2018.

[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[27] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.

[28] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural networks for perception*, pages 65–93. Elsevier, 1992.

[29] GM Jacquez, JR Meliker, RR Rommel, and PE Goovaerts. Exposure reconstruction using space-time information technology. *Encyclopedia of Environmental Health*, pages 636–644, 2019.

[30] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

[31] Jihyeon Lee, Dylan Grosz, Burak Uzkent, Sicheng Zeng, Marshall Burke, David Lobell, and Stefano Ermon. Predicting livelihood indicators from community-generated street-level imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 268–276, 2021.

[32] Jihyeon Janel Lee, Dylan Grosz, Sicheng Zeng, Burak Uzkent, Marshall Burke, David B. Lobell, and Stefano Ermon. Predicting livelihood indicators from crowdsourced street level images. *CoRR*, abs/2006.08661, 2020.

[33] Yiping Lu, Aoxiao Zhong, Quanzheng Li, and Bin Dong. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In *International Conference on Machine Learning*, pages 3276–3285. PMLR, 2018.

[34] Juila Ortheden Markus Pettersson. *Predicting Economic Well being in Africa using temporal satellite imagery and deep learning.* PhD thesis, Chalmers University of Technology, 2021. unpublished thesis.

[35] Gary C McDonald. Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1):93–100, 2009.

[36] Ambar Narayan and Rinku Murgai. Looking back on two decades of poverty and well-being in india. *World Bank Policy Research Working Paper*, (7626), 2016.

[37] Kenji Obayashi, Keigo Saeki, Junko Iwamoto, Nozomi Okamoto, Kimiko Tomioka, Satoko Nezu, Yoshito Ikada, and Norio Kurumatani. Positive effect of daylight exposure on nocturnal urinary melatonin excretion in the elderly: a cross-sectional analysis of the heijo-kyo study. *The Journal of Clinical Endocrinology & Metabolism*, 97(11):4166–4173, 2012.

[38] Shailesh M Pandey, Tushar Agarwal, and Narayanan C Krishnan. Multi-task deep learning for predicting poverty from satellite images. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[39] Anthony Perez, Christopher Yeh, George Azzari, Marshall Burke, David Lobell, and Stefano Ermon. Poverty prediction with public landsat 7 satellite imagery and machine learning. *arXiv preprint arXiv:1711.03654*, 2017.

[40] Thomas W Pogge. Eradicating systemic poverty: brief for a global resources dividend. *Journal of Human Development*, 2(1):59–77, 2001.

[41] Aiswarya Raj, Jan Bosch, H Holmström Olsson, Anders Arpteg, and Björn Brinne. Data management challenges for deep learning. In *45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pages 1–8, 2019.

[42] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[43] Shea O Rutstein. Steps to constructing the new dhs wealth index. *Rockville, MD: ICF International*, 2015.

[44] Juan Sapena, Vicent Almenar, Andreea Apetrei, María Escrivá, and María Gil. Some reflections on poverty eradication, true development and sustainability within cst. *Journal of Innovation & Knowledge*, 3(2):90–92, 2018.

[45] Patrick Schober, Christa Boer, and Lothar A Schwarte. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5):1763–1768, 2018.

[46] Vickie L Shavers. Measurement of socioeconomic status in health disparities research. *Journal of the national medical association*, 99(9):1013, 2007.

[47] Jeroen Smits and Roel Steendijk. The international wealth index (iwi). *Social indicators research*, 122(1):65–85, 2015.

[48] Surendran Padmaja Subash, Rajeev Ranjan Kumar, and Korekallu Srinivasa Aditya. Satellite data and machine learning tools for predicting poverty in rural india. *Agricultural economics research review*, 31(2):231–240, 2018.

[49] J Ties Boerma and A Elisabeth Sommerfelt. Demographic and health surveys (dhs: contributions and limitations. *World health statistics quarterly 1993; 46 (4): 222-226*, 1993.

[50] usaid. The dhs program - india: Standard dhs, 2019-20. `https://dhsprogram.com/methodology/survey/survey-display-541.cfm`. (Accessed on 09/07/2021).

[51] USAID. The demographic and health surveys program, Jul 2021.

[52] Peter A Whigham, Caitlin A Owen, and Stephen G Macdonell. A baseline model for software effort estimation. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 24(3):1–11, 2015.

[53] Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. Transfer learning from deep features for remote sensing and poverty mapping. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[54] Christopher Yeh, Chenlin Meng, Sherrie Wang, Anne Driscoll, Erik Rozi, Patrick Liu, Jihyeon Lee, Marshall Burke, David B. Lobell, and Stefano Ermon. Sustainbench: Benchmarks for monitoring the sustainable development goals with machine learning, 2021.

[55] Christopher Yeh, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature communications*, 11(1):1–11, 2020.