Using NASA's Night Light Data Analyzing Poverty Index Bangladesh

by

M A Moontasir Abtahee
19301150

Maisha Jarin
20101125

Jannatul Ferdoush
20301291

Md Farhad Mahamud Azad
20301378

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
October 2024

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Abtahee

M A Moontasir Abtahee
19301150

Maisha

Maisha Jarin
20101125

Jannatul

Jannatul Ferdoush
20301291

Md Farhad Mahamud.

Md Farhad Mahamud Azad
20301378

Approval

The thesis/project titled "Using NASA's Night Light Data Analyzing Poverty Index Bangladesh" submitted by

1. M A Moontasir Abtahee (19301150)
2. Maisha Jarin (20101125)
3. Jannatul Ferdoush (20301291)
4. Md Farhad Mahamud Azad (20301378)

Of Summer, 2024 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science October 17,2024.

Examining Committee:

Supervisor:
(Member)

____

Abdulla Hil Kafi
Lecturer
Department of Electrical and Electronic Engineering
Brac University

Co Supervisor:
(Member)

____

Raihana Shams Islam Antara
Lecturer
Department of Electrical and Electronic Engineering
Brac University

Thesis Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

In the never-ending quest to reach Sustainable Development Goal 1 (SDG 1), which is to end poverty, new and different ways of doing things are needed, especially in places where collecting data the old way is hard. "Analysis of Poverty Using NASA Black Marble NTL Data with the DHS Wealth Index of Bangladesh," the title of this study, describes a new way to figure out how poor a country is by using NASA's VNP46A4 dataset, which is an important part of the Black Marble suite. The VNP46A4 dataset gives annual averages from 2017 to 2018 based on NTL radiance corrected for atmospheric and lunar factors, containing 28 layers of useful information. Initial analysis using VNP46A4 and DHS data revealed inefficiencies and accuracy issues. Therefore, we propose a new approach: utilizing datasets for the year 2022, including OSM, Google Static Maps, and NASA's VNP46A3, which provides monthly data that we have merged into yearly aggregates, to achieve higher accuracy. This method focuses on specific data layers, h26v06 and h27v06, within the VNP46A4 dataset covering Bangladesh. By integrating the unique nighttime brightness data of Bangladesh through merging and clipping, and combining it with other significant datasets, this study aims to present a comprehensive overview of the country's socioeconomic status. Using NASA's nighttime light data intelligently, this approach seeks to provide policymakers, development agencies, and academics with a transformative tool, leveraging advanced technology and extensive datasets to enable informed decisions and targeted assistance for struggling communities.

Keywords: Deep Learning, Machine Learning, NTL data, Poverty Index, lunar factors, SDG 1, VNP46A4 and VNP46A3 database, Nighttime Luminosity Data, Bangladesh, DHS cluster, sinusoidal grid, OSM, Google Static Maps

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.

Secondly, to our co-advisor Abdulla Hil Kafi sir for his kind support and advice in our work. He helped us whenever we needed help.

Thirdly, to our co-supervisor Raihana Shams Islam Antara mam for his kind support and advice in our work.

And finally to our parents without their throughout sup-port it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Bangladesh being a geographically diverse country with massive socio economic issues. Bangladesh has faced the issues of the lack of timely and credible data on the household economic situation, which hampers the efficiency of policy measures directed at reducing poverty[30].Prediction of poverty is an important stage toward the reduction of world inequalities for sustainable development. It enables the formulation of policies and decisions for appropriate funding and intervention, and humanitarian assistance plans, the implementation of developing strategies and projects. However, other techniques like Demographic and Health Surveys (DHS) which give poverty estimates have some drawbacks[34]. The commonly used approaches are slow, costly, and sometimes descriptive only and this is particularly more real in developing nations like Bangladesh. Due to the quick dynamics of the socioeconomic environment, the requirement for fresh, accurate, and easily available data on poverty has become more significant.

In recent years, enhanced technology in data access has transformed poverty prediction as a field. This study opens new horizons in estimating poverty through the availability of satellite data for capturing geographical features as well as Open-source geographical data from OSM, and using or implementing tools such as GSM[37]. For example, nighttime luminescence or NTL, as it is also referred to, has evolved as an ideal proxy for economic activity capturing infrastructure improvements, electricity consumption and urbanization. Likewise, satellite imagery taken during the day shows features of geography that one can use to improve estimates of the degree of wealth and destitution[33]. In our study, we concentrate on poverty prediction. As a result of the above challenges, the current paper puts forward the following two-stage strategy to enhance the accuracy of poverty estimation.
At the first level, we use annual NTL data from the Black Marble dataset in conjunction with DHS data to model poverty. We use K-Nearest Neighbors (K-NN), Random Forest, and Decision Tree as a subset of machine learning (ML) models. These models assist in selecting features of the satellite data relevant to a household's wealth and we can attain accuracy of between 60% and 70%. For instance, applying the K-NN model generates an accuracy of 61% and the Decision Tree model generates an accuracy of 63%. While these results are not efficient, we aimed at increasing the accuracy and, for this purpose, extended the dataset and used more profound methods.

In the second stage, we include monthly satellite data from NASA's VNP46A3 data set (2022), OpenStreetMap data (2022), Google Static Map data (2022) and latest DHS data (2022). Using both these datasets, we are able to obtain more detailed and timely fluctuations in the economic environment in various districts of Bangladesh. Here, we expand the methods such as VGG16 and SDG to analyze the GSM data by population density and density intensity which is divided into three classes. This means that the GSM data considered here retains the rural/urban breakdown which helps to provide a better assessment of economic processes. From these datasets, our deep learning models learn features that greatly enhance impoverished predictions. At this stage, our models get an astounding 80% accuracy in poverty prediction all over the country making the integration of satellite imagery along with open-source and survey data a significant propulsion.

In addition, our research is not only interested in methodological questions about estimating poverty but also shows how development policy implicitly relies on certain methodologies and their assumptions in Bangladesh[4]. Since poverty maps are already used to distribute financial assistance, development programs and other support structures, it is important that these maps are as accurate as possible to reach the correct populations. Besides, it is free from the limitations of political/household surveys and is more cost-friendly than approaching households directly for surveys.This research uses machine learning and deep learning techniques to create a strong model fit for predicting poverty in Bangladesh. Thus, using open-access big data and advanced ML algorithms, we would like to join the global initiative to tackle poverty and promote more effective and equitable development in the future years.

1.1 Aims and Objectives

This study attempts to transform Bangladeshi poverty estimation through the use of novel techniques and satellite data. Household surveys were the traditional technique of assessing poverty, yet they consist of flaws such as biases and insufficient coverage. We suggest a multidimensional method that is based on two studies. Firstly, we use the Demographic and Health Survey (DHS) wealth measure along with NASA's Black Marble Nighttime Lights (NTL) dataset to find the poorest areas of Bangladesh[28]. When social data is integrated with satellite data, the picture reflects economic disparity completely. Following that, we employ various advanced approaches, including machine vision and deep learning using the OSM dataset, Google Static Maps[32] and NTL dataset[35] with these tools, we could perfect poverty mapping, eliminate prejudice and study how wealth is distributed besides income. Our broad approach impacts the decision which is fact-based, concerned with special groups and cases, and projects designed for sustainable development.

1.2 Problem Statement

Due to the limitations of traditional survey methods, it is harder to get an accurate assessment of how much poverty there is, especially in countries like Bangladesh.

These methods often fail to take into account how quickly the social and economic landscape changes, causing a scarcity of real-time information. This paper shows an innovative approach to solve this problem using NASA's VNP46A4 dataset, the Demographic and Health Survey (DHS) data, OpenStreetMap (OSM) data, Google Static Maps, and NASA's VNP46A3 dataset[32]. Utilizing the availability of satellite data makes poverty predictions far more accurate and up-to-date, which makes it a vital resource for dealing with social and economic problems. Nevertheless, the utilization of extensive and intricate datasets for poverty mapping entails a distinct array of obstacles, which this research endeavors to surmount.

1.3 Background

Bangladesh, one of the most densely populated countries in the world, has undergone many changes in its current form. Bangladesh suffered from liberation war in 1971 and then it faced several political instabilities , natural calamities and widespread poverty. Such problems delayed it for a long time. But this was to change after the severing of the chains of democracy in 1991. Bangladesh has demonstrated remarkable economic growth in the period from 2018 and 2019 with an 8% economic growth in 2018 and 7.9% in 2019. It was able to assist millions of its population to come out of poverty while offering support in education, life expectancy and farming that boosted the country's growth. Bangladesh was also a developing country on its way out of the United Nations Least Developed Countries (LDC) list by 2024[14]. Bangladesh faces new challenges that may jeopardize its poverty-reduction achievements. The "Quota Reform Movement," while addressing fundamental social issues, resulted in political and economic turmoil. The country's growing external debt—USD 100.6 billion as of December 2023—has prompted questions about its long-term economic viability[24].These factors underscore the nation's persistent vulnerability to poverty, despite its endeavors to extricate itself from the UN's designation as a Least Developed Country. Given these challenges, the necessity for accurate poverty forecasting and targeted responses becomes increasingly critical. This project aims to provide a comprehensive picture of poverty in Bangladesh through the examination of many variables, including daylight imagery, OpenStreetMap data, and nocturnal illumination, thereby informing effective policy interventions for its reduction.

However, there are areas that contain many challenges to this day. A very serious problem which continues to exist is the inequality, which exists between the rich and the poor, and between the people living in the rural and the urban areas. Although there has been a decline in poverty in the rural areas, many people in urban areas are still living in poverty, and the increase has apparently favored large cities such as Dhaka[12]. It has led to emigration from the rural areas in the quest for such opportunities thereby exerting more stresses on the urban structure. Further, it has also slowed down the poverty eradication rate making it very crucial to discuss how we are going to sustain the rate achieved.

The story of Bangladesh is both a tale of success and struggle. The country has really come a long way and yet it has experienced and still experiences some difficulties of inequality, poor urban areas, and the pressure caused by loans and social movements[12]. In the future, the government needs to guarantee that all people will also be a stakeholder in development and that those who managed to avoid poverty again fall into it.

Chapter 2

Literature Review

The article titled "Remote Sensing of Night Lights: A Review and an Outlook for the Future" offers a thorough examination of the progress and uses of remote sensing technology for capturing and analyzing nocturnal illumination. The paper examines the historical context and development of space-based sensors for quantifying nocturnal illumination. It explores the diverse applications of nocturnal illumination, including urbanization mapping, environmental disaster monitoring, and armed conflict analysis. Additionally, it investigates the influence of holiday and decorative lighting on nocturnal illumination levels. The study also covers the challenges and opportunities of the subject. The work discusses the historical background of remote sensing of night lights, noting that the use of nighttime Earth images was limited by the RS community before the 1990s for a lack of digital formats until 1992. It also provides useful information on the development of space-born electromagnetic emitters for monitoring night light even in urban regions.Such data incorporates OLS from the Defence Meteorological Satellite Programme (DMSP), and the Visible Infrared Imaging Radiometer Suite (VIIRS).
Furthermore, the research presents the application of REMS(remote sensing) of NI(nocturnal illumination) such as a tool to identify the pattern of urban growth, events related to natural disasters and wars. It analyzes the application of night-only lights in evaluating impacts on electrical networks, observing armed hostilities, and evaluating the impact of wars on IDPs. It also explores the impact of holiday and decorative lights on night lights that are valuable data on fluctuations in the energy consumption of countries across the globe. The study also analyzes different spaceborne instruments for measuring nightlight, including DMSP/OLS, VIIRS, Landsat, and other exploratory cubesats. The text provides a systematic review of space borne sensors for mapping nightlights. The sensors are compared regarding the spatial resolution with emphasis on the advantages and limitations of each sensor as mentioned in the text.
In the paper, the author comprehensively reviewed the development and application of remote sensing technology for night light research. This makes it an important reference book for scholars and professionals working in the fields of remote sensing and geographical information systems[15].
The paper, "Utilizing NASA's Black Marble to Measure the Change in Electrification and Regional Inequality in Dhaka, Bangladesh ," is a master's thesis work of Kh Shakibul Islam. This paper was completed as a partial fulfillment of the course work for the award of a Masters of Arts degree in Geography from Binghamton

University, State University of New York in the year 2023. Discussed in this report is a refined Nighttime Development Index-NLDI that uses advanced Black Marble Nighttime data. It is needed to overcome the limitations of previous research that used VIIRs or DMP OLS satellite images for analysis. The literature review part of the study focuses on the method and approach of using remote sensing data and GIS technologies to map and compare land use and land cover (LULC) in Dhaka – a city characterized by rapid urban growth and population expansion. The research also employs the computation of GDP per capita using nighttime remote sensing data for Dhaka. Research has further revealed that the Nighttime Light Development Index(NLDI) levels have a positive response to electrical intensity and hence GDP per capita.

The research employed various data types including daytime and nighttime remote sensing data, global population counts, and land use and land cover classification systems. Initially, an assessment was made about the accuracy of the data and during the causehand investigation it was found out that there are some changes in land use, land cover, electricity and NLDI during the year of 2010–2020 in Dhaka of Bangladesh. In this study, the authors affirm that the application of the Nighttime Development Index is a step forward in city planning.This method provides a detailed and sophisticated view of the nighttime environment and its possibility for business, social, and cultural development[20].
The article under consideration is entitled "Using Earth Observations to Help Developing Countries Improve Access to Reliable, Sustainable, and Modern Energy" which was authored by K. M. de Beurs, B. A. et al. The authors conducted an electronic search using a vocabulary list of keywords in Google Scholar and Google Search. Applying this search resulted in 9,856 potential titles that were related to the application of Earth Observation (EO) data into energy management. The authors filtered the volume to 2,763 papers for the abstract/summary review and then prioritized 384 documents of them for the full-text analysis according to certain categories and criterion. The final and sources of literature comprise the peer-reviewed journals, Industry/consortia reports, workshop/conference report and important original papers in government organization and other agencies. Towards the end of the paper, it stresses on decency of data or model uncertainty and other technical constraints particularly in the context of climate information or projections. It also raises awareness of the need for lessons in different areas of study such as data handling using statistical or geographical information system software, and other statistical or artificial intelligence approaches. According to the authors, researchers and DataBase providers should ensure the accessibility and flexibility of training or capacity development programmes of a global nature. Moreover, the investigation highlights the importance of clearly describing, for example, analyses, modeling, or other activities that might be conducted while completing a project and storing such documentation over time[10].
The paper titled "On the Use of Satellite Nightlights for Power Outages Prediction", seeks to explore the applicability of substituting satellites VIIRS night light data products for power supply in order to predict power outages resulting from hurricanes where there is dearth of adequate historic information. The study uses machine learning algorithms that are learned from historical data of power outages, the climate conditions, and other relevant environmental data to predict power outages

during future hurricanes with special interest in hurricane events in the Caribbean region including hurricane Maria. The paper undertakes the challenge of generating prediction models for utility location that have little or no historical power loss data aggregated through satellite-based night light observations. The study also explores the use of various weather explanatory variables as well as non-weather variables and the creation of the meteorological variables using the WRF(Weather Research and Forecasting)model.The authors of the study are Juan P. Montoya-Rincon, Shams Azad, Rabindra Pokhrel, Masoud Ghandehari, Michael P. Jensen, and Jorge E. Gonzalez. It specifically was published in the year 2022. The study was sponsored by the Brookhaven National Laboratory and the U.S. National Science Foundation. In the paper under review, the authors provide a comprehensive review of employing satellite nightlights in predicting power outages.It addresses this challenge of developing prediction models for areas that lack the power outage data and looks at the possibilities of using this study on better preparedness for severe weather disturbances[38].
The article titled 'Spatial heterogeneity of uncertainties in daily satellite nighttime light time series' analyzes the daily satellite nighttime lights (NTL) patterns of change in spatial heterogeneity and temporal dynamics. It also measures the general effect of the environment as well as observational conditions on the difference in NTL intensity. Thus, the study proposes a systematic procedure for evaluating the fluctuation of nighttime lights (NTL) at multiple spatial and temporal resolutions. It uses generalized linear models (GLM) to identify the relationship between daily changes in NTL and environmental parameters and observational conditions at a grid cell level. This is done in the study and also isolates uncertainty in the daily NTL time series which arises from seasonal effects, angular effects as well as daily varying variables. In models used in the research, there are harmonic terms, linear terms, quadratic functions, and interaction terms. These are used to depict to a better probability how environmental factors and observational conditions are unlikely to exhibit simple relationships with the variation of nighttime light (NTL).To compare the models, we apply the coefficient of determination (R2) and to check how each specific variable influences the daily NTL, we employ the average marginal effect (AME). The paper gives a detailed description of the methods used and results obtained which can help in understanding of regional characteristics of uncertainty involved in daily satellite nighttime light time series[22].

The research paper "Multi-scale estimation of poverty rate using night-time light imagery" is an attempt at finding a new method of evaluating poverty amongst different scales of a country with the help of Night-Time Light Imagery (NTLI) and other sources of open data.The research applies various models and methods including, Random Forest Regression (RFR), Principal Component Analysis (PCA), the Alkire-Foster (A-F) approach, multi-angle NTL data, and land cover data. The article examines the case study of Mozambique in which the suggested approach is applied in order to estimate poverty at different spatial levels. The accuracy of the findings concerning the estimates is ascertained through the deployment of relative error and R-square. This shows that the adopted model is highly effective in predicting the poverty rates with a level of accuracy of 85.21% together with R-square of 0.94. The research focuses on this idea that NTL imaging and the application of machine learning can help predict poverty rates on various levels of spatial disaggregation. This approach could be taken to other countries to predict poverty levels and fetch useful information for focused interventions towards poverty.The paper provides a clear analysis of the technique and data employed in developing the poverty rate estimate model. This includes the use of Household survey data, Remote sensing, data, Statistical poverty rate data[21].

The study conducted by Ditmer et al. (2021), "Estimating the loss and fragmentation of dark environments in mammal ranges from light pollution," evaluates the impact of light pollution on mammalian species by estimating the loss and fragmentation of dark habitats within their ranges. The researchers used data from NASA's Visible Infrared Imaging Radiometer Suite to quantify the impact of full moonlight brightness in order to define a scientifically-informed threshold of They used the ranges of 351 mammals in the contiguous United States to figure out how much of each species' range was affected by light pollution above the levels associated with direct emissions and skyglow, how much of the range was not affected ("dark environment"), and how many pieces of dark environments there were.extent of "dark environments" where light pollution was absent, as well as the degree of fragmentation in these dark environments. The researchers discovered that, on average, the habitat of mammals covered 2.6% of the region, where the brightness from point-source emissions continuously surpassed that of a full moon. However, there was a wide variation across species, ranging from 0% to 47.4%. The phenomenon of the skyglow had a significant impact on a far larger proportion of the areas, namely 24.3%. Diurnal species exhibited a somewhat lower level of exposure compared to nocturnal species. Several families, such as the Molossidae, which includes free-tailed bats, were particularly affected and consisted of species that are of conservation concern. Due to the isolation of animal habitats brought about by light pollution, they are now more susceptible to long-term exposure to artificial light. According to the authors, it is crucial to identify the species that are most affected by light pollution and habitat fragmentation in order to focus conservation efforts on protecting the remaining dark areas that serve as refuges for light-sensitive species [16].

The study "Using radiant intensity to characterize the anisotropy of satellite-derived city light at night" by Xi Li et al. (2022) introduces a novel method to evaluate the anisotropy of city light at night, acquired from satellite data, by using radiant intensity. The authors suggest using a linear regression model to depict the correlation between the viewing zenith angle (VZA) and radiant intensity. This approach yields an average regression R2 ranging from 0.26 to 0.73 across fourteen worldwide cities.Further, they recommend the cosine correction linear model that describes the relationship between the view's zenith angle (VZA) and radiance. This model is obtained mathematically from the linear model that exists between VZA and intensity. The paper relies on NASA Black Marble product suite for the imagery of nighttime lights and uses LiDAR data to analyze the relationship between artificial night-time light anisotropy and urban geometry. As proclaimed by the authors themselves, radiant intensity is perhaps the most suitable physical quantity for quantifying artificial light at night in terms of its directionality. Perhaps this measure can help to explain the anisotropy in artificial light at night.

The paper gives a good overview of the directional dependence of CLDN as observed from satellites and proposes a new method that estimates CLDN through the radiant intensity. The authors did a great job to identify ways on how artificial light is distributed at night time. To determine how the viewing zenith angle influenced the radiant intensity, they specified linear regression models as well as cosine-corrected models.By using NASA's Black Marble product suite and LiDAR data, the study makes data more credible by considering high-quality, openly available data. In general, the work offers some valuable information about the dependency of artificial light at night and presents a new approach to explore it[19].

A 2023 academic paper by Qiming Zheng et al. called "Nighttime light remote sensing for urban applications: Progress, challenges and prospects" reviews a lot of the work that has already been done in the field of nighttime light remote sensing (NLRS) and what may or may not be possible in cities. Some of the areas of the NLRS that the authors investigate include; the socioeconomic and environmental indices; calculation of NTL data for function zones; and analysis of human behaviors. Moreover, they stress the necessity of analyzing the cause and consequence of such variances with respect to NTL data as this is a key part among the creation of new applications and progress of the field. The authors explore the possibilities to estimate the socioeconomic and environmental parameters like GDP, population, and energy utilization by analyzing the nocturnal light (NTL) brightness.The writers discuss the various methods applied in these studies including deep learning strategies, machine learning algorithms and regression models. The authors explore how urban function zones, including local climate zones and urban infrastructures, can be classified using NTL data.
The authors investigate how changes in the intensity of human activity brought about by urbanization, public holidays, regional wars, natural disasters, and public events like the COVID-19 pandemic may be tracked using NTL data.
updated study guidelines. Future studies in NLRS-based urban applications are recommended to focus on four interrelated topics by the authors. These include deciphering NTL changes, handling scale effects and sources of variation, comprehending NTL data at a deeper level, refining techniques for handling missing data and taking atmospheric circumstances into account. In their assessment of the literature, the authors highlight the potential and challenges of this technology for understanding and managing urban environments by citing several studies that have made use of NLRS data for various urban applications[8].

A survey of the literature on the many applications of light pollution maps and their social ramifications can be found in the article "Maps of light pollution in odd places: Are night time satellite pictures making us to forget natural darkness?" by J. Lyytimäki. The author contends that light pollution maps—which are derived from satellite data collected at night—are used in a variety of different communication contexts in addition to the context of light pollution discussions. According to the study, there may be a loss of natural darkness due to the widespread usage of these maps, which might reinforce the perception of an artificially lit nighttime world as the standard for human experience. To highlight the possible unintended consequences of map-based visual information, the article offers a few instances of how light pollution maps have been used in various settings, including energy reporting, environmental advocacy, and fiction. In order to achieve fact-based knowledge and well-informed decision-making, the author also emphasizes the need for critically analyzing the framing provided by the maps. The research comes to the conclusion that, particularly in a culture that is becoming more and more information-intensive and digitally mediated, it is important to be mindful of the possible unintentional signals and interpretations that the maps may produce. 2020 saw the publication of the work in the International Journal of Sustainable Lighting[9].

The paper "Development of a nighttime shortwave radiative transfer model for remote sensing of nocturnal aerosols and fires from VIIRS" models the transport of shortwave radiation at night for remote sensing purposes. The multiple-scattering linearized vector discrete ordinary VLIDORT radiative transfer code is integrated and improved in the UNL-VRTM model.This is the lighting employed by moonlight, artificial lighting, fires and sources of light on the surface.This paper presents studies on the UNL-VRTM and VLIDORT models together with the Planck function, Stokes vector, and the utilization of moonlight as a collimated light source.

The study by Jun Wang et al. examines the methods used to standardize the images obtained with a DSLR camera in ISS. This calibration is underlined as crucial for understanding the effects of artificial light at night on the Earth's atmosphere and ecosystems. Housed under this goal is the primary aim of this project, that is, to design an effective and reliable method to solve calibration problems in DSLR photography. The authors give a detailed description of how DSLR pictures need to be adjusted to enhance their quality for space-related analyses. In order to correct an uneven illumination due to either the lens of the camera or the curvature of the Earth it is required to consider flat field, illumination, and vignetting problems. Astrometric calibration is that image phase in which the pixel of an image is aligned with the coordinates of the Earth. An orientation in space is accurate by reference to star fields at each stage of the calibration. If star field calibration is impossible, the paper describes the method of ground based photometric calibration. Geo-referencing is a procedural component which relates the degree of aerial photo's pixel to a physical georeferenced coordinate system and improves the spatial referencing of calibrated pictures. The authors also support their calibration method with actual star brightness data from the Ducati II/237 catalog, shown against Nikon D3S calibration pictures. This validation approach assesses the accuracy and reliability of the proposed calibration methods. The paper also examines the process of shutter closing and suggests a universal adjustment factor for f/number. The calibration technique is comprehensive since it takes into account several elements to resolve potential flaws and ensure the correctness of the image. Ultimately, the study offers a meticulous and comprehensive methodology for calibrating DSLR photographs taken from the International Space Station (ISS), uncovering the impact of artificial light on Earth at night. The research's authors demonstrate a high level of technical expertise and the capacity to apply their findings in a practical manner. This makes the study a significant contribution to the fields of remote sensing and Earth Ub observation[7].

In the paper, "Combining satellite imagery and machine learning to predict poverty", Neal Jean and others conducted a literature review and pointed out that there is a lack of data in developing countries especially in Africa which limits the prospects of economic analysis and planning. The authors stress that most African countries performed less than two nationally representative surveys in 2000-2010 which hinders understanding and combating poverty. Other approaches, including nighttime lights (NTL) using the satellite data obtained at night, have been effective at the country level but fail to differentiate economic status within low luminosity regions/areas in the LMICs. The authors develop a new machine learning method based on high-resolution daytime satellite imagery and a CNN trained using transfer learning to extract socioeconomic information. This method can accurately estimate poverty and wealth with nearly as much detail as possible when only using publicly available data, and has predictive ability that accounts for 60–75% of variation in local economic conditions across five African countries: Nigeria, Tanzania, Uganda, Malawi, Rwanda.The study makes an important contribution into the directions of poverty estimation and points to the broader possible use of the machine learning techniques for various scientific fields where data is scarce.Moreover, the authors detail how their model can sometimes yield better results than the nightlight model and is a more scalable solution for low-income clusters for policymakers who want to deliver their interventions in the right locations[1].

The "D4N Poverty Estimates Report" published by the Bangladesh Bureau of Statistics (BBS) uses a new approach to estimate poverty by combining satellite imagery and machine learning that is intended to generate timely and geographical breakdowns of poverty rates that are critical for policy interventions in Bangladesh. It was undertaken within the framework of the Data4Now project and with the support of the UN Statistics Division and the University of Southampton. The model employs open-source satellite imagery along with Nighttime Light (NTL) data that has an 84% accuracy with the 2016 Household Income and Expenditure Survey (HIES). The approach uses a grid cell size of 3890 x 3890 meters, enabling poverty estimates at the union level and showing a national average poverty rate of 19.0 % for 2022 with regional disparities, the lowest poverty at Khulna at 18.3 % and the highest at Barisal at 20%. The report also points out that more data from other layers that include POI, land covers and roads must be integrated into the model for improved accuracy of the final results. Among the capacity building was the training of more than 25 of its officials from different governmental sectors on how to harness this innovative method. Through the use of a Random Forest based CNN, the BBS is also designed to significantly increase the efficiency of poverty monitoring and help policymakers to make insurance informed decisions for poverty reduction as articulated in the SDGs as well as improved disaster occurrence analysis.plans. The initiative also highlights the need for methodology development within NSOs to harness advanced analytics for further poverty tracking [2].

The paper "Deep Learning Approach Analysis Model for Prediction and Classification of Poverty Status" by Musli Yanto et al. addresses the urgent issue of poverty in Indonesia, particularly exacerbated by the COVID-19 pandemic. The deep learning (DL) framework developed for this study combines K-means clustering, artificial neural networks (ANN) and support vector machines (SVM) for efficient prediction and classification of poverty levels. Taking into account the accuracy of the model obtained by using the main variables such as population density, the income levels, and poverty rates, the model is highly accurate, with an accuracy of 99.8%. The importance of advanced analytical techniques integration, data processing, and actionable insights for policymakers is underscored by the authors. This novel approach seeks to serve as a source of guidance to governments in crafting their economic management and poverty fighting strategies, leveraging the potential of artificial intelligence in solving complex social economic problems.

The authors review existing literature, which shows that traditional poverty classification methods are often unable to characterize the many dimensions of poverty dynamics. Previous studies show varying accuracy of 71.93 obtained using SVM alone. This field is then advanced by Yanto et al.'s model that optimizes feature selection using Pearson correlation analysis and K-means clustering of data for efficient data grouping. The research promotes the importance of AI in improving predictive capabilities as well as systematic data processing for government interventions. This study makes a constructive contribution to deep learning in social science research by advocating further exploration of additional data layers and machine learning algorithms. The study has shown to be a promising avenue for policy making interventions in Indonesia, and more broadly, by exploiting data to alleviate poverty [23].

In a paper titled "Measuring Poverty in India with Machine Learning and Remote Sensing," Daoud et al. proposed a novel way of measuring human development in India by the fusion of machine learning (ML) and remote sensing (RS). However, standard welfare measures — such as the census data or household surveys — are often incapable of delivering the temporal or spatial resolution needed to watch local progress. The authors point out that while these traditional methods are exhaustive, rarely implemented and expensive, they leave big gaps in localized data needed for targeted public policies. This study seeks to overcome these limitations by using high resolution satellite imagery and deep learning techniques to understand living conditions across Indian villages in a more dynamic and granular manner.

The research specifically addresses two key gaps in existing literature: the overwhelming prevalence of studies focusing on Africa and the scantness of outcomes measured. The authors attempt to provide a more comprehensive assessment of human development, by applying their EO-ML methodology to a broad array of more than 90 human development indicators, across different income and asset regimes. For this reason, the study emphasizes a need to employ an asset index based on census data instead of lamp luminosity, which is the more commonly used proxy in other research. In addition to improving the accuracy of poverty assessments this innovative approach also enables continuous monitoring of developmental trajectories over time. The findings ultimately suggest that EO-ML methods have the potential to revolutionize how we measure the phenomenon of human development globally, and especially in very densely populated regions like India, where lack of data can impede the formulation of effective public policy and attainment of Sustainable Development Goals [18].

The paper titled "Predicting City Poverty Using Satellite Imagery" by Piaggesi et al. looks at whether machine learning and satellite imagery can create a good socio economic conditions forecast of urban space. Gathering socio-economic data using traditional methods of surveys or censuses is expensive and labor intensive, with many gaps in reliable information. To predict household income in five metropolitan areas in North and South America, including Santiago, Chile, and major U.S. cities, the authors propose a novel approach that combines high resolution satellite images with pre trained CNNs. The study seeks to apply transfer learning techniques to adapt models that were already used for poverty mapping in resource poor settings to the more heterogeneous populations of developed countries: that is, with differing population density and resource concentration.The paper gives a literature review of the applications of deep learning techniques to remote sensing data for socio-economic analysis; the literature is presented to show how significant developments have been recently in this regard. Previous research that has attempted to establish relationships between the poor and nightlight intensity has focused mostly on developing countries in rural areas. In contrast, this work examines whether these methodologies can be applied to urban contexts and whether existing models trained over lower resolutions can predict poverty at finer scales of cities. The results show that pre-trained CNNs can robustly predict household income without requiring fine tuning or reliance on proxy variables; they even predict households income at unknown cities in the test sample. This work is part of a growing body of literature aiming to develop new poverty measurement techniques by combining advanced technological methods with more conventional socio-economic indicators while delivering a scalable solution for real time socio-economic analysis in urban contexts[11].

Hall et al. (2015) in their paper 'A Review of Machine learning and Satellite Imagery for Poverty Prediction' have undertaken an exhaustive evaluation of the combination of machine learning (ML) techniques with satellite imagery (SI) for supporting improved poverty measurement and analysis. Yet traditional approaches, including household surveys, are expensive, infrequent and are spatially constrained, which is something the authors point out. However, while these traditional approaches furnish detailed data, the temporal granularity of their outcome does not suffice in providing enough detail to support effective monitoring in the changing events. The review demonstrates that the integration of ML and SI presents an attractive alternative whereby welfare indicators can be estimated from satellite images with the same or better accuracy than traditional survey methods. Statistical analyses are presented showing that the use of asset based indicators results in a 17 point performance advantage over the use of income or consumption measures. They also say that integrating deep learning techniques into ML can boost predictive power by as much as 15 percentage points.

The literature review additionally discusses the complications with measuring poverty, including its necessity to incorporate both 'hard' measures (i.e., income and nutrition, as well as 'hard' measures like physical assets in the predictive models. Specifically, the authors claim that soft measures like insecurities are perfectly amenable to imagery, whereas physical indicators such as housing quality and infrastructure are particularly well suited for detection using high resolution images. They discuss how diverse methodologies were employed in these recent studies, including the use of daytime satellite data instead night time light data (which is shown to be more successful in impoverished areas where night time light data might not correctly represent economic activity). The paper also showcases how ML algorithms can be successfully applied on various datasets and how combining different data sources such as mobile phone metadata, environmental indicators and satellite imagery can be used to significantly improve the accuracy of poverty estimation. In addition to synthesizing existing research, this comprehensive review acts as a foundational resource for scholars and practitioners wishing to use advanced technologies for poverty analysis and informed policy making[5].

Chapter 3

Description of the Data

3.1 NASA'S BLACK MARBLE

NASA's Black Marble facilitates numerous applications for various data users. Primarily helping short-term in the weather forecasting and disaster response sectors, it also provides novel data to track wildfires, chemical flares, and illumination. The data holds several economic programs, such as serving as a proxy for GDP, monitoring changes in urban energy infrastructure, and assisting humanitarian organizations in turmoil areas. Numerous variables influence the nighttime environment, including urban illumination, lightning, navigation lights from fishermen vehicles, gas flares, flows of lava, and northern lights. The moon's illumination of the Earth enables the study of oceanic and terrestrial features through reflections from ice, snow, and other reflective surfaces. The Visible Infrared Imaging Radiometer Suite (VIIRS), integrated within the Suomi National Polar-orbiting Partnership (NPP) satellite, contains 22 spectral bands spanning from ultraviolet to mid-infrared wavelengths. The Day Night Band (DNB) detects nighttime illumination due to its receptivity to visible and near-infrared wavelengths. This extensive instrument enables comprehensive analysis of several environmental and social-economic factors[6].

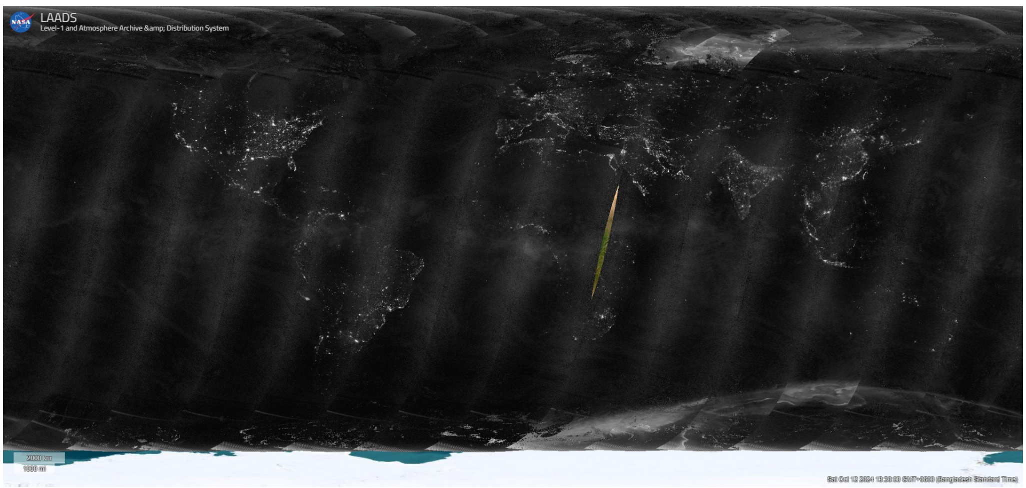

Figure 3.1: Earth At Night 2022 . Courtesy of NASA/NOAA Suomi NPP

3.1.1 VNP46A4

The fourth NTL product in the Black Marble suite, which was introduced by the Suomi National Polar-orbiting Partnership (SNPP) Visible Infrared Imaging Radiometer Suite (VIIRS), is abbreviated as VNP46A4. This dataset contains annual composites generated by BRDF-correcting the daily NTL radiance of the atmosphere and moon in order to eradicate the influence of unrelated artifacts and biases. Furthermore, the dataset comprises 28 layers that furnish details on the NTL composite, quantity, quality, and standard deviation for various categories of multi-view zenith angles (including near-nadir, off-nadir, and all angles), alongside their snow-covered and snow-free conditions, in addition to the land-water mask, latitude, and longitude coordinate data[41].

3.1.2 VNP46A3

The third nighttime lights (NTL) product in the Black Marble suite, VNP46A3, generates monthly composites with daily atmospherically and lunarly-BRDF-corrected NTL radiance, effectively eliminating biases and artifacts. The VIIRS/NPP Lunar BRDF-Adjusted Nighttime Lights Monthly L3 Global 15 arc-second Linear Lat Lon Grid comprises a total of 28 levels. They report land-water mask, latitude and longitude coordinates, snow-covered and snow-free statuses, the NTL composite, quantity of observations, quality, and standard deviation for multi-view zenith angle categories (near-nadir, off-nadir, and all angles). They also go into great length on quality flags. January 1, 2012, is the start date of Version-ID 001 in the file name of this SNPP VIIRS collection-1 product record[40].

3.2 DHS Survey Data

The 2022 Bangladesh Demographic and Health Survey (2022 BDHS) is the ninth national survey to report on the demographic and health conditions of women and their families in Bangladesh. The survey was conducted under the authority of the National Institute of Population Research and Training (NIPORT), Medical Education and Family Welfare Division, Ministry of Health and Family Welfare (MOHFW), Government of Bangladesh. The primary goal of the 2022 BDHS is to provide current data for basic demographic and health factors. The BDHS specifically collected information on reproductive and juvenile mortality rates, demographic preferences, awareness, acknowledgment, and usage of family planning possibilities, mother and child health (including breastfeeding habits), nutritional status, and newborn care. The data gathered by the 2022 BDHS aims to aid policymakers and program administrators in formulating and assessing initiatives and plans to enhance the health of the Bangladeshi population. The survey additionally offers statistics relevant to the Sustainable Development Goals (SDGs) for Bangladesh[28].

3.2.1 DHS Household Data (BDHR81FL)

The 2022 Bangladesh Demographic and Health Survey (BDHS), a part of Phase 8 of the DHS series, gathered a lot of data on Bangladeshi households and the country's economy. This survey, which is essential for making modifications to policies, asked regarding family size, structure, income, expenditure, assets, and members. The four main surveys used were the household, the woman's, the biomarker, and the verbal analysis. Fieldworkers completed a questionnaire on their own. It had been feasible to talk to 30,018 families and 30,078 women; 19,987 women completed the whole questionnaire. This a lot of information helps policymakers make decisions and evaluate health programs and strategies that will improve the health of the citizens of Bangladesh or help the country reach its Sustainable Development Goals (SDGs).

3.2.2 DHS Geographic Data (BDGE81FL)

The Bangladesh Demographic and Health study (BDHS) 2022 is the ninth national study that NIPORT and the Ministry of Health and Family Welfare have been in charge of. The research is meant to give up-to-date estimates for significant medical and demographic indicators. It includes a geographic dataset (BDGE81FL) with details regarding where families reside and where the boundaries of the administration are. The dataset includes 675 cluster sites extend out in both rural and urban areas. This helps with mapping and analyzing how demographic and health indicators are distributed in an area. This location information helps us learn more about differences and trends in various regions of Bangladesh. The BDHS gathered a lot of information about fertility and infant mortality rates, reproductive preferences, understanding and utilization of family planning tools, health of mothers and children (including breastfeeding), nutrition levels, and care for newborns. This information is essential for policymakers and program managers to plan and assess health programs. It also helps Bangladesh achieve its goals for sustainable development (SDGs). We have gathered data regarding places in country and urban areas, including their latitude and longitude, and turned it into point geometric and cluster numbers[28].

3.3 Google Static Maps

Google Static Map Imagery
The daylight satellite imagery is obtained from the Google Static Map API. For a land mass area of approximately 130,170 km^2, a total of 414,757 images have been amassed to create a comprehensive representation of the country, featuring a zoom level of 16, a pixel resolution of 2.5 m, an image size of 400x400 pixels, with each image corresponding to a single pixel of the nighttime luminosity data, encompassing approximately 0.25 km^2. The daytime satellite photos are arranged to align with their nocturnal counterparts regarding light intensity. This would expedite the training and extraction of geographical characteristics during unsupervised training. These images are classified into 3 different classes based on intensity.

3.4 OpenStreetMap (OSM)

Although Google Static Map is very accessible, it remains a proprietary enterprise offering, resulting in costs for picture acquisition. Although this price is justifiable relative to a costly survey aimed at representing a national population, it remains an impediment to information access and introduces a temporal delay in analysis (Google imposes a daily limit of 25,000 Static Map image requests, which for a country of considerable land area, would result in a postponement of poverty identification analysis). Consequently, an alternative data source for geographical feature mapping from Open Street Map for Bangladesh is also utilized. Open Street Map (OSM), as a crowd-sourced community, offers a complimentary option to Google Static Image, facilitating geo-mapping of both natural and artificial features within a selected country's terrain. The features derived from OSM data may signify the wealth distribution levels throughout the country, either as an independent feature set or in conjunction with nightlight luminosity data to enhance the predictive efficacy of the models developed from these resources.

- **Road Networks:** The OSM dataset includes detailed mapping of road networks in Bangladesh, covering national highways, regional highways, zilla or district roads, Upazila Roads, Union Roads, and Village Roads. This comprehensive mapping helps in understanding the connectivity and infrastructure of the country[36][3].

- **Land Use:** The data set provides us about the way different parts of land are used, like fields, pastures, cities, dwellings, industrial operations, agriculture, educational institutions, and amusement areas. The data contained in this file is very important when investigating how land spreads and planning future strategies of managing land[27].

- **Buildings:** Man-made buildings with close to stable roofs are identified via geo-mapping in the OSM dataset. Urban growth and planning require an understanding of the kinds and locations of buildings throughout the country, as is given by this data[26].

- **Points of Interest (POI):** Points of interest, which identify significant locations like hospitals, schools, parks, and shopping centers are included in the dataset. Knowing the distribution of basic services and facilities is essential to pinpoint areas that may be neglected or in need of service improvements [29].

3.5 Shapefile

In GIS software, a shapefile is a popular geographically vector data format utilized to store the precise location, form, and other details of geographic features. Operational boundaries are provided for nations at several levels by the Global Administrative Areas (GADM) dataset; GADM Level 3 offers unique boundaries for communes, municipalities, and comparable units. Administrative levels 0 through 3 are included in the subnational administrative boundaries dataset for Bangladesh from OCHA. We have utilized Level 2 and Level 3 shapefiles from OCHA and GADM, respectively, in our research. In order to correlate population data with the geographic distribution of health and demographic factors, these datasets are crucial for knowing national trends and regional inequalities. Whenever paired with point geometry-converted longitude and latitude data, this data offers an entire tool for geographic investigation[13].

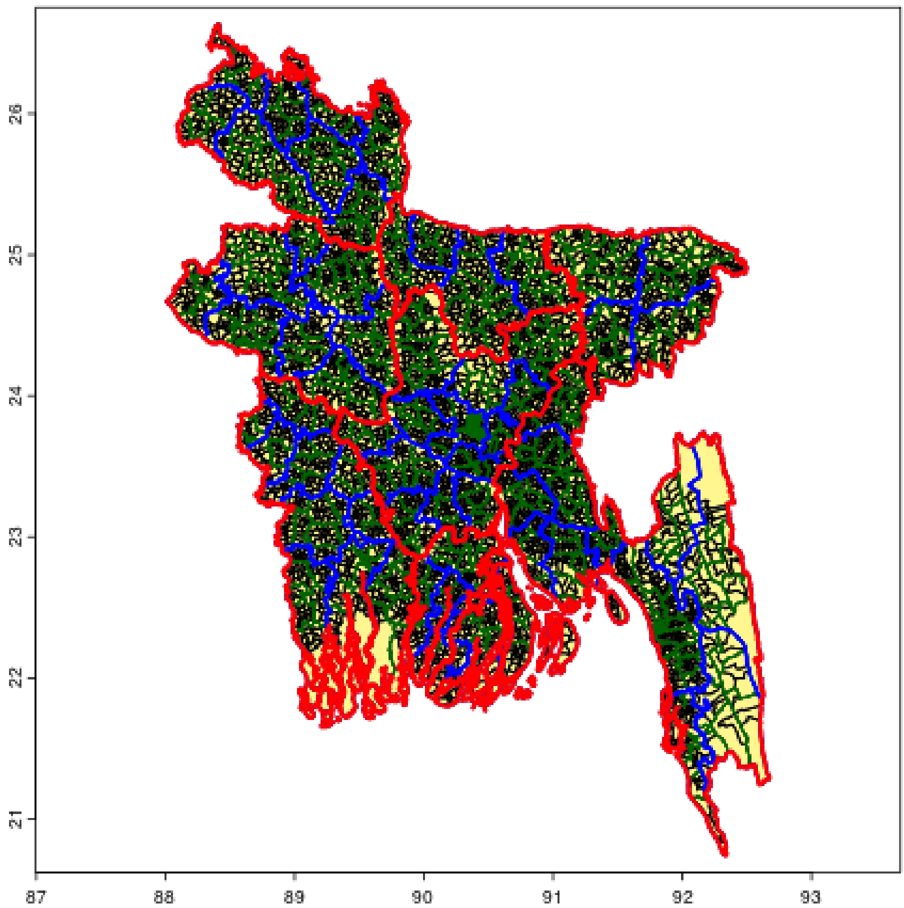

Figure 3.2: Shapefile of Bangladesh from GIS Software

Chapter 4

Methodology

4.1 Data Preprocessing

The data preparation for this research was carried out in to multiple phases, each corresponding to distinct data sources: NASA's Night Time Luminosity (VNP46A4,VNP46A3), DHS Survey Data (BDGE81FL,BDHR81FL), DHS Cluster Location, OSM, and Google Street Maps

4.2 DHS Survey Data

In order to determine the cluster location for each cluster number provided by the DHS dataset, we first gathered the DHS geographic data (BDGE81FL) and extracted specific columns for cluster number, urban or rural area, longitude, latitude, and point geometry. Next, we combined the cluster location with the wealth index derived from the DHS household data (BDHR81FL) based on cluster number. As a result, the DHS data was preprocessed; although there were no null or missing values, the DHS data was frequently inaccurate, and the DHS wealth index does not contain the actual wealth of that area but rather the wealth factor of that entire area.

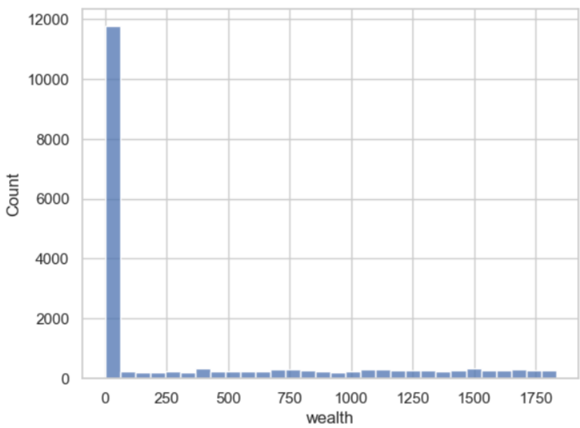

Figure 4.1: The DHS Wealth Index

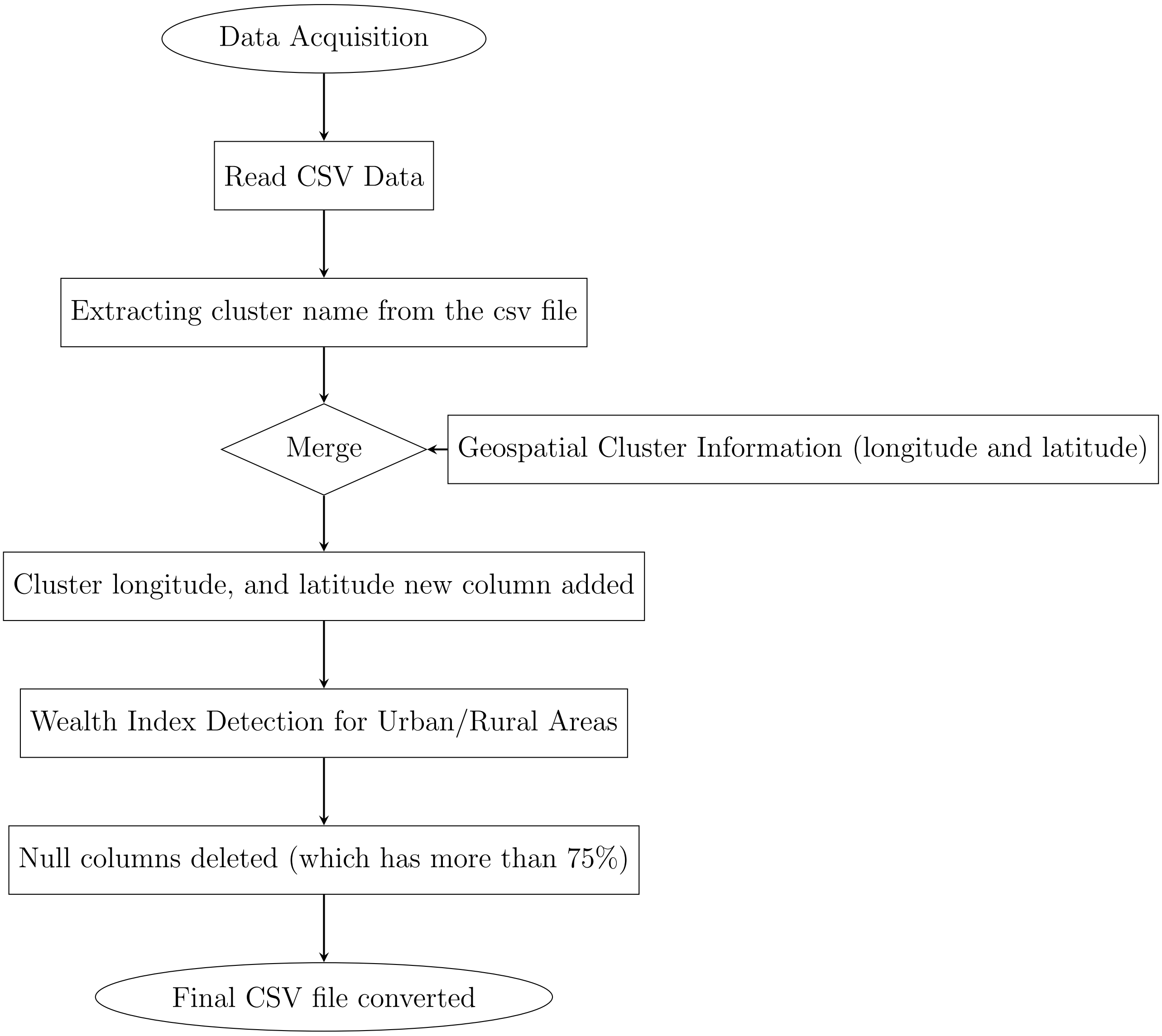

Figure 4.2: DHS Survey data and Geospatial DHS data

4.3 VNP46A4 and VNP46A3

We are using The nighttime lights (NTL) data from both the VNP46A4 and VNP46A3 datasets. These datasets, provided in HDF5 (h5) format, contain multiple layers of information about nighttime lights captured by the Suomi National Polar-orbiting Partnership (NPP) satellite.

The VNP46A4 data for Bangladesh is divided into two files, namely h26v06 and h27v06, for each year due to the spatial coverage of the satellite imagery. We focused on extracting specific layers from these files: layers 4, 14, and 22, corresponding to the near-nadir, off-nadir, and all-angle snow-free composites, respectively, along with the Land Water Mask layer.

The VNP46A3 dataset provides monthly composites of nighttime lights, and we followed a similar process of extracting the relevant layers. This dataset offers a different perspective on nighttime lights with its monthly aggregation, complementing the VNP46A4 data[40].

After extracting the necessary layers from both VNP46A4 and VNP46A3, we converted them into GeoTIFF format for easier handling and processing within a geographic information system (GIS). We then merged the individual GeoTIFFs into single, multi-band images, combining all relevant information for each geographic area. These merged images were then clipped to the boundaries of Bangladesh using a shapefile to remove any extraneous data.

Finally, we combined the processed data from both VNP46A4 (h26v06 and h27v06) and VNP46A3 to create a comprehensive and multi-temporal nighttime lights dataset for Bangladesh. Given the relatively low levels of nighttime illumination in some areas, we chose the maximum pixel value from the available data to ensure the use of the most reliable and informative data points for our poverty prediction model.

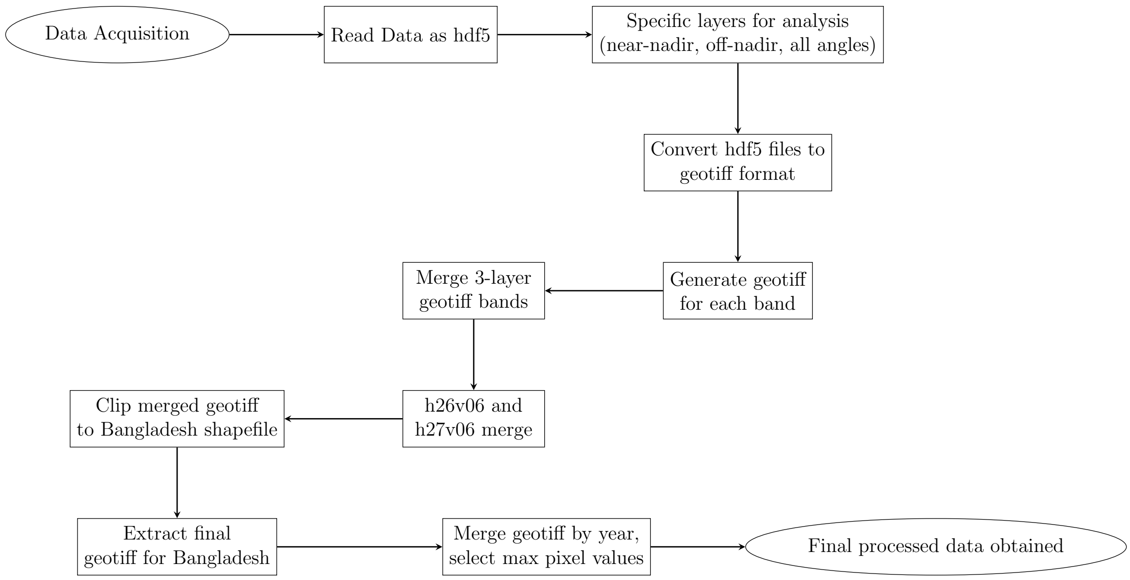

Figure 4.3: Nasa Night Light Yearly and Monthly data preprocessing Flowchart.

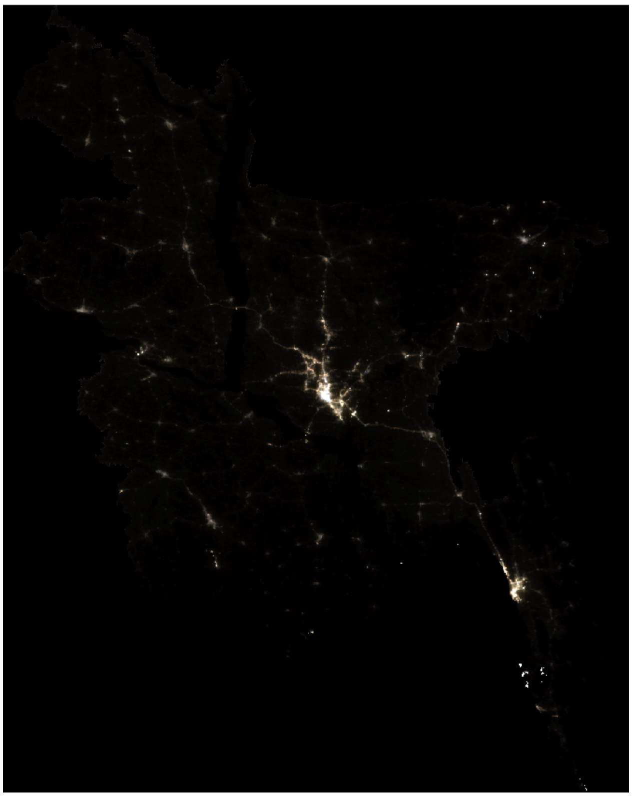

Figure 4.4: Nighttime light intensity for VNP46A4

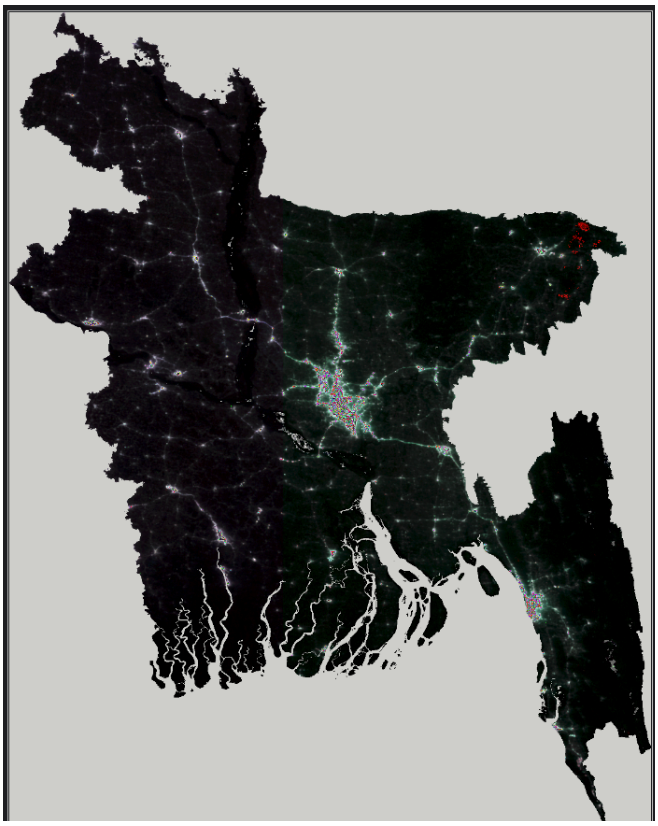

Figure 4.5: Nighttime light intensity for VNP46A3

4.4 Merge Nightlight and DHS data

Nighttime lights data is being merged with survey data on household wealth to explore the relationship between the two. The goal is to determine if areas with more nighttime lights also tend to have higher levels of wealth. The process starts by reading in the nighttime lights data, which is like a map showing the brightness of different areas at night. Then, survey data containing information on household wealth and location is read in.
The key step is combining these two datasets. For each location in the survey data, the nighttime lights in the surrounding area are examined. Evaluations are made for the maximum, minimum, average, and variability in brightness of various light intensity parameters.A new dataset combining these readings of light intensity with wealth data is the final result. This makes it possible to determine if places having brighter illumination at night are often more wealthy. To make this relationship easier to understand, a visualization is offered as well. To put it simply, it's like seeing if the wealthier areas correlate with the brighter areas on a nighttime image of Bangladesh. This information can be very useful for comparing economic activity as well as pinpointing regions that require additional assistance.

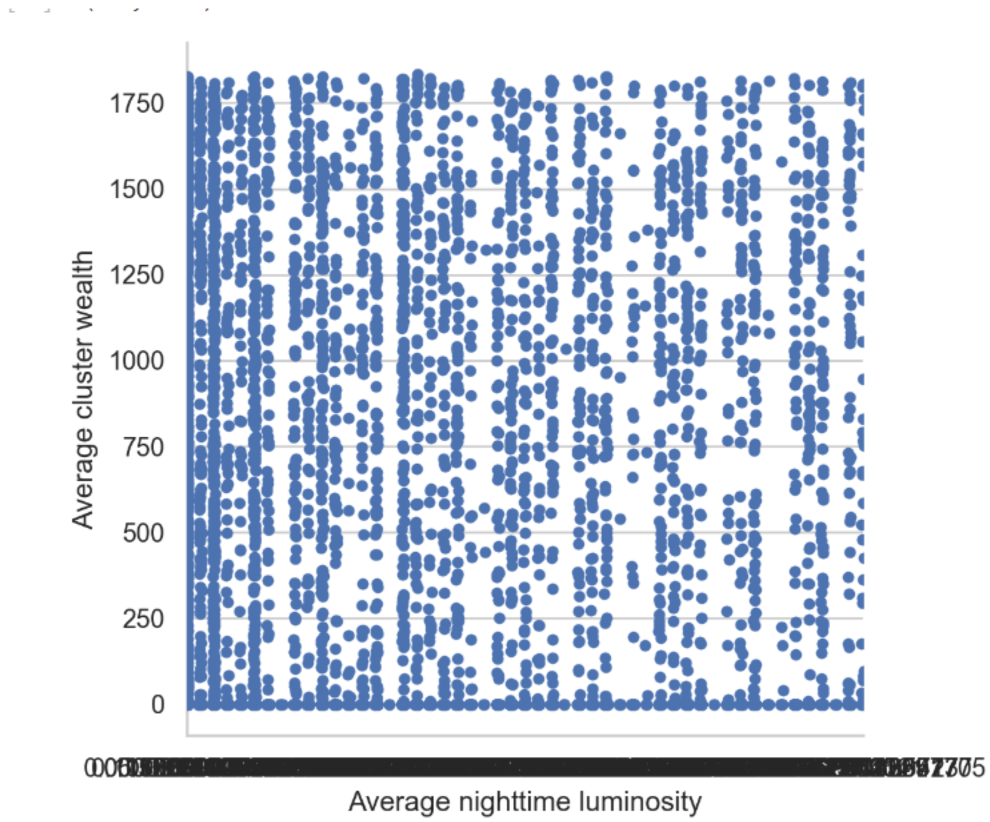

Figure 4.6: The Distribution of Nighttime Light intensity with wealth data

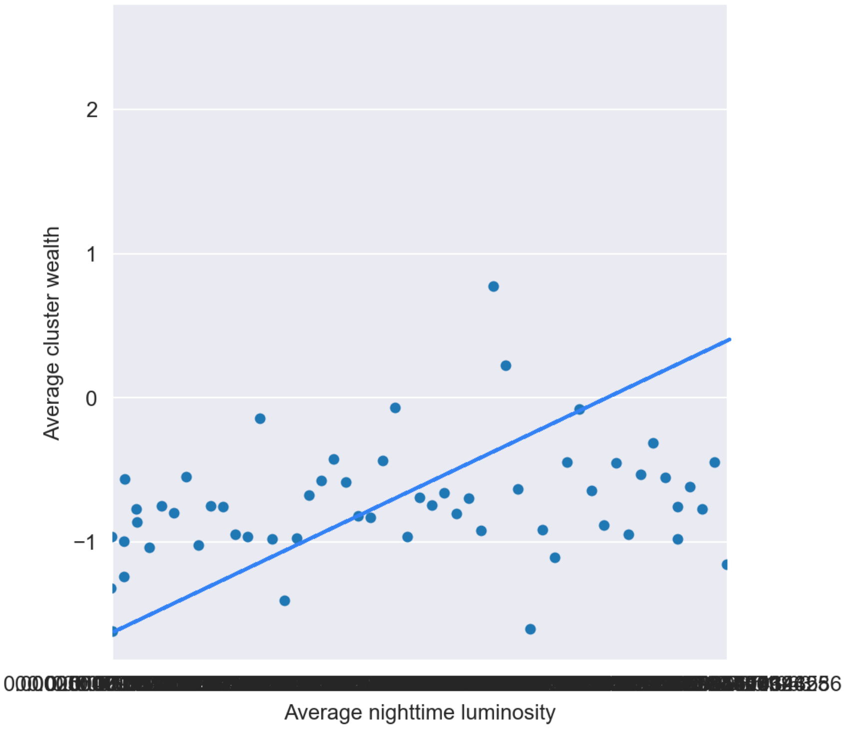

Figure 4.7: Average Nighttime Luminosity Correlation with Wealth Distribution in Household Clusters for testing data

4.5 Pre-processing of OSM Dataset

In term OSM data we put a focus on only 4 main features and they are buildings, Point of interest , landuse and roads. The features are extracted and preprocessed to get substantial information about the build environment and architectural infrastructure for all the cluster locations. Thus , we get the density and typological structure of buildings in both urban and rural settings . We measured a variety of building types to examine the form of the build environment. All the services and amenities are listed under the POI which includes school, health facilities and markets at a predefined distance from each cluster. It helps to understand the degree of reachability of basic features and dynamics of the economic activities across different regions . In the same way the land use features show how a land is being used in an economic and social structure according to our research settings. We assessed the total area of various land use categories, including homes, businesses, and agricultural, encircling each cluster. This analysis explicates the prevailing economic processes and how to allocate land for different reasons.

Road networks are essential for accessibility and mobility. We collected data from the road system and calculated the total road length in a certain radius surrounding every cluster. With each cluster, we additionally calculated its distance from the closest road, which gives us an idea of how accessible and interconnected each cluster is. This data aids in our comprehension of how simple it is to go around and move about in many different places.

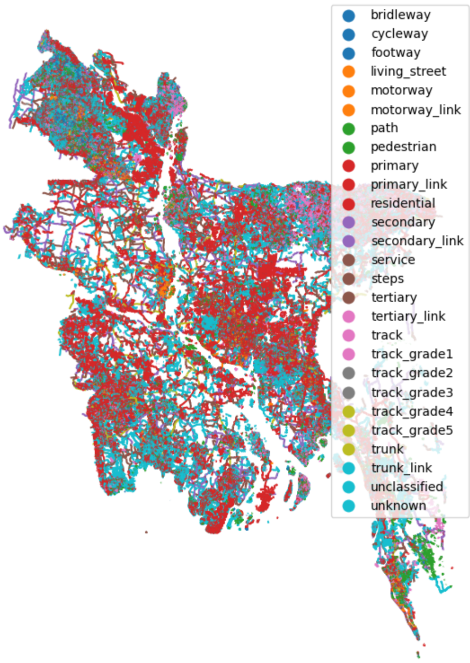

Figure 4.8: All Mapped roads in Bangladesh OSM Data

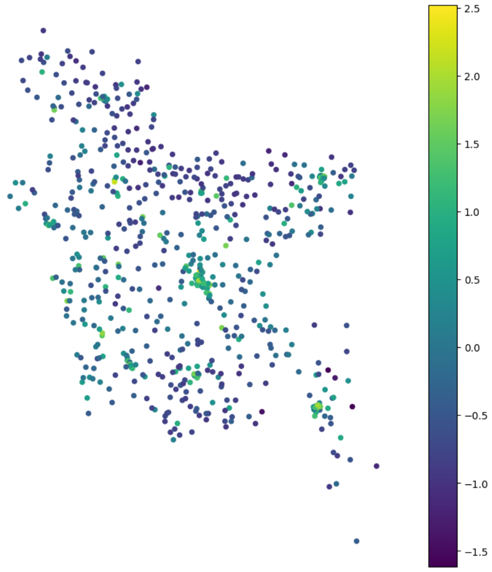

Figure 4.9: Nightlight Based Poverty Mapping

4.6 Pre-processing Google Static Map dataset

The satellite image was initially segmented into 64 levels before downloading, and for the aim of maximizing model training efficiency, we further divided it into 3 levels based on intensity. Basic features were removed from the downloaded images by processing. Among these characteristics included the RGB color values of intensity for each pixel in the images. Statistical measures such as the maximum, minimum, average, and variation of color intensities were calculated to capture the visual characteristics of the images.

This preprocessing of GSM data aimed to create a set of features that could be used in conjunction with nighttime lights and survey data for poverty prediction. The extracted features provide insights into the daytime environment, complementing the nighttime lights data and potentially enhancing the accuracy of poverty prediction models.

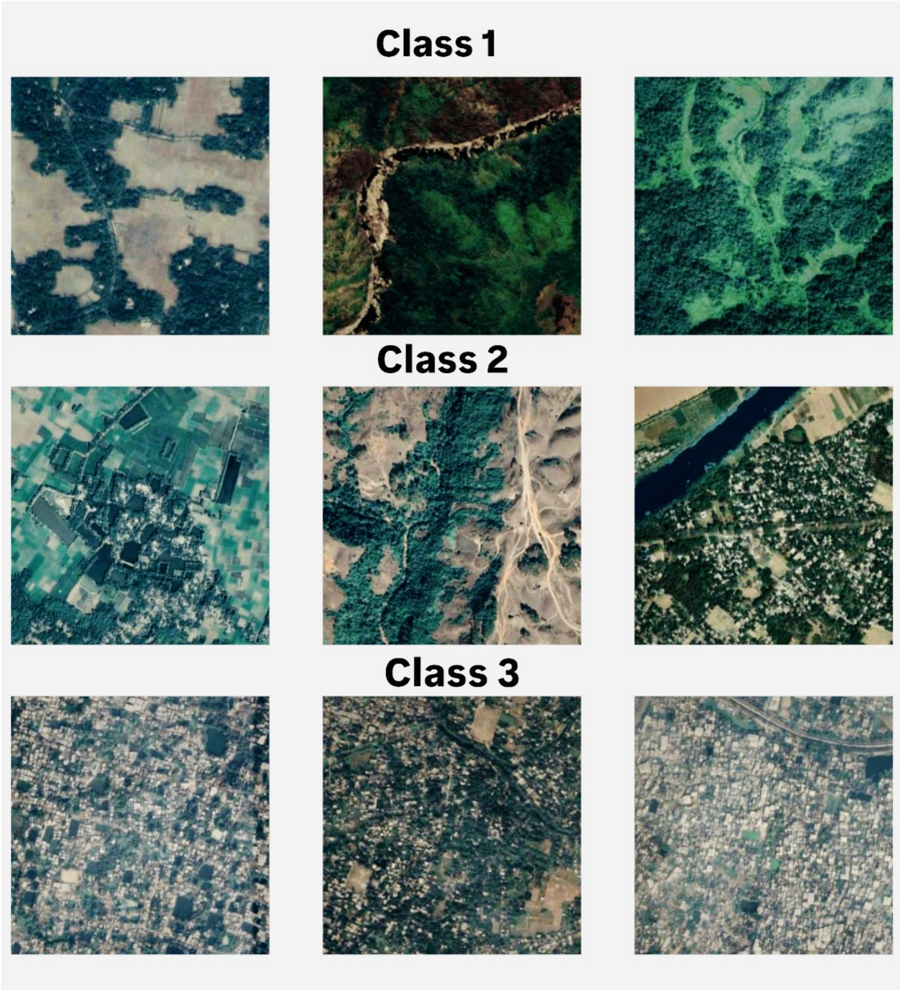

Figure 4.10: Day Light Satelite images after cluster them into three classes

4.7 Target identification

The investigation aims to uncover solutions to the subsequent inquiries:
1. Given the unavailability of survey data, can other publically available data sources effectively quantify and map poverty and wealth distribution?
2. Which data sources or combinations thereof are most effective as predictors of variations in economic outcomes?
The Wealth Index from the DHS survey was utilized as the target variable to describe and quantify poverty as well as welfare levels in response to these problems. This metric has been employed as a common indicator of household wealth in national surveys and is derived by decreasing the high dimensionality of household asset data through principal component analysis.
According to DHS, this approach to constructing the wealth index is capable of adapting and achieving precision. Articulate the socioeconomic conditions of both urban and rural regions. The households surveyed in the DHS were selected to represent the whole population of Bangladesh, so their economic statistics provide a comprehensive overview of the country's wealth distribution.

4.8 Feature Space Engineering

The primary challenge in utilizing real-world data, such as nighttime luminosity images (.tif format), daytime satellite images (.jpeg format), and OpenStreetMap geospatial data (.shapefile format), lies in its unstructured and noisy nature. These diverse data types require careful transformation into a suitable format for machine learning models. This method entails turning these diverse sources of data into numbers that the models can understand. The method of feature engineering, which is an example of data preparation, is vital for assuring the accuracy and suitability of the data to estimate poverty.

4.9 Nighttime Luminosity as a stand-alone predictor

This research explores the idea of using nighttime illumination as an analogy for wealth, predicting that higher luminosity levels are linked with greater electricity use, construction expenditures, and economic activity. Luminosity features, which include maximum, minimum, mean, median, and standard deviation, are taken from nighttime imaging and compared to the wealth levels of 600 Bangladeshi household clusters so as to investigate this association. The investigation thoroughly assesses the degree to which changes in richness can be attributed to nighttime luminosity utilizing machine learning models with K-fold cross-validation. The models are trained to correspond to the extracted nightlight attributes as a measure of average wealth. Using this method allows a thorough assessment of the prediction ability of at night lighting to estimate the wealth degrees of the household clusters.
This relationship can be shown visually in the scatter plot that displays a household's location and asset-based wealth index overlaid over emission of light at night. Greater light intensity locations tend to correspond with wealthier household clusters, particularly in Dhaka, the country's capital city. On the contrary hand, lower-income families are more likely to live in areas with less nighttime light.

4.10 Daytime Satellite Images as a stand-alone predictor

The application of daytime satellite imagery to forecast social and economic prosperity is examined in this study. The process starts by taking basic data out of the photos, like the compressed RGB color channel' maximum, minimum, mean, median, and standard deviation. After that, a model that predicts wealth based on these basic image metrics is developed using these features. This first model creates a basis for assessing daylight imagery's predictive capacity.
A Convolutional Neural Network (CNN) that has been pre-trained on the ImageNet dataset is used to derive significant characteristics from the daylight images in order to improve the analysis even more. The pre-trained model is already capable of identifying unique characteristics such as curves, forms, and lines. The model can be modified to identify elements indicating socioeconomic level, such as roads,

rivers, land usage, or roofing materials, by freezing the last stage and retraining it on Google daytime images. The wealth levels inside the clusters are then predicted using these tabulated deep traits that were extracted. The aim of this study is to ascertain if the integration of deep features with fundamental image metrics may enhance the precision of wealth level prediction in contrast to the use of basic features solely. This method makes use of CNNs' capacity to extract complex information from photos and evaluate how well they predict social and economic prosperity.

4.11 Deep Learning Techniques Discussed in this Study

A Convolutional Neural Network architecture (CNN) is used to extract the deep characteristics from the satellite photos. CNNs are a particular kind of neural network that are essentially composed of two parts: a front-end part that performs the feature learning that preprocesses the image data, and a back-end part that performs the classification task, i.e. classifying the wealth level of the household clusters for our use case using the information from the front-end input (Ferlitsch, 2020).
In the part that follows, the application of neural networks will be covered in more detail. For the sake of image processing, we will first concentrate on the convolutional layers, which are the front end of the feature extraction stage.

4.11.1 Processing Satellite Images for Computer Vision

The satellite images must be transformed into numerical representation in order to be processed by the computer model as inputs. The first step in implementing image processing is realizing that a typical RGB image consists of three distinct channels, each of which corresponds to a different intensity of red, green, and blue light that is compatible with human color receptors. The pixel values in these photos correspond to each layer and range from 0 (no color visible) to 255.
(The pixel has the highest level of color). As a result, the pixel values from these satellite photos can be analyzed and used as input for the models.
But the magnitude of the input is the primary barrier to picture processing. For instance, each satellite image in our use case includes 400x400 pixels spread across three RGB channels. With a single image that is 400x400x3, or 480,000 pixels, we need an input vector with 480,000 elements of a comparable length. Furthermore, we need to handle 414,757 photos in total, which will yield an input vector with over 160 billion components overall. The amount of weights we would need to update at just the input layer of a typical Deep Neural Network for this task, assuming that it has 4096 input nodes to learn from the data, would be in the trillions, making it impossible for our standard computer hardware to handle.

4.11.2 Convolutional Layer

Prior to the implementation of convolutional layers, training and processing real-world visual data using neural networks proved impractical. As previously mentioned, it serves as a preprocessing front-end for a neural network, facilitating the reduction of the high dimensionality of a pixel-based image to a more manageable lower dimensionality feature-based representation[25].
The latter can thereafter be utilized as the input vector for Deep Neural Networks for additional processing and learning.

Translational Variance : In conventional neural networks, an object's or feature's recognition performance is largely reliant on where it is located within the image. A small change can cause identification problems. The capacity of a model to identify an object or feature regardless of where it is in the image is known as translational variance.CNNs' convolutional layers deal with translational variation. These layers can identify characteristics wherever they may be found using filters that scan the whole image. By using these filters methodically across the image, the network has the ability to recognize the existence of a feature in any area. The primary advantage it has over traditional neural networks, that are sensitive to the exact location of pixels, is this[17].

Feature Detection: Feature detection is a process of determining the location of significant structures or patterns in a picture, such as edges, corners, or textures. These features are the raw elements which are used to interpret the content of an image. We have done this feature detection via the convolutional layers of the VGG16 model. These layers apply learned filters to match a set of various features present in the input images. Application of a large number of filters each only designed to detect one type of feature.

Pooling: Pooling is a downsampling technique where we make the feature map smaller which is helpful for making the model less expensive and generalization of features by lessening the complexity of the dataset. We have used average pooling which can downsample to reduce dimensionality where the overall feature is intact. It also removes unnecessary details of the model but keeps contacting the necessary ones.

Flattening: In flattening , a multidimensional array of feature maps is taken and converted into a single dimensional vector. To provide a feature to feed into a neural network we need to convert the feature into a vector otherwise data can't be fed to the input layer of fully connected CNN. As its an essential step , where the previous layers data are preserved in a vectorized form.

4.11.3 Deep Neural Network

A neural network with several hidden layers is called a deep neural network (DNN). The network may discover intricate, non-linear relationships in the data because of these layers. As higher-level features are gradually extracted by each layer, a hierarchical representation of the input data is created. A DNN with two hidden layers and 4096 nodes per layer is defined by your code. These layers learn to forecast wealth levels by processing the convolutional layers' flattened features. The network can capture complex patterns and dependencies between the features and the target variable because of its various levels.

Transfer Learning: Transfer learning is the process of starting a new task using a pre-trained model. In order to enhance performance and shorten training time on a smaller dataset, this makes use of the knowledge gained from a larger dataset. When the new task is similar to the task the model was trained on initially, it is especially helpful. The VGG16 model, which has been pre-trained on the ImageNet dataset, is used in your code. The model's later layers are retrained to precisely forecast income levels from your satellite photos, but the early layers—which identify generic traits like edges and shapes—remain in place. This method saves a lot of time and money by not having to train the network from scratch.

VGG16: The deep structure and strong performance of the VGG16 CNN architecture on image recognition tasks make it stand out. It was created by Oxford University's Visual Geometry Group and produced cutting-edge outcomes on the ImageNet challenge. The foundation for transfer learning in your code is the VGG16 model. The VGG16 pretrained weights are loaded and utilized to initialize the model, giving feature extraction a solid foundation. The model can gain from the experience gained on a large and varied dataset of photos thanks to this initialization.

Convolutional Neural Network Architecture: A CNN's general architecture establishes the placement and connections between its many layers. The network's communication flow as well as the methods for feature extraction and processing are determined by its design. The pre-trained VGG16 model is combined with extra convolutional, pooling, and dense layers in a CNN architecture defined by your code. This architecture is intended to anticipate wealth levels by extracting pertinent information from satellite photos. The performance of the network depends on the precise configuration of the layers and their hyperparameters (filter size, stride, and number of nodes).

4.12 Description of the model:

4.12.1 Machine Learning Techniques Discussed in this Study:

In this research, we have examined several machine learning algorithms, including Decision Trees, Random Forests, Naive Bayes, k-Nearest Neighbours (K-NN), Gradient Boosting Classification, and Adaptive Boosting (AdaBoost). We have selected these algorithms as it shows a spectrum of algorithms ranging from elementary ones like Decision Trees and K-NN, to more complex ensemble methods like Random Forests, Gradient Boosting, and AdaBoost. Each step has its own advantages and disadvantages The Naive Bayes classifier is very simple, but efficient in text categorization tasks. Nevertheless, all of these methods need calibrations to avoid overfitting and properly manage the dataset which has a large number of dimensions.

4.12.2 Implementation of Models:

In this study, we implemented several machine learning models to testify our prediction for poverty. Let's talk about the implemented models.

Decision Trees: A basic machine learning model that is used to split data into branches based on feature values. We have used a Decision Tree model with a maximum depth of 10, and then the split criterion is set to entropy so that we can ensure balanced information gain, preventing overfitting.

Random Forests: An advanced machine learning method where we combine multiple decision trees together to improve accuracy and robustness and its more effective than decision trees. Our Random Forest model has 1000 decision trees to predict poverty from features and each with a maximum depth of 10 which creates a diverse and robust random forest algorithm which goes with the research.

Naive Bayes: This is a probabilistic machine learning model based on Bayes' theorem which often shows good results for text classification. The Naive Bayes model assumes a Gaussian distribution of features and works well for continuous data.

k-Nearest Neighbors (K-NN): The KNN model is used for classification and regression by finding the most similar data points which is also known as clustering. In this model we used 5 nearest points from the dataset to find the poverty range.

Gradient Boosting Classification: An machine learning algorithm that builds models sequentially to minimize errors as low as possible. We have built the Gradient Boosting model to optimize a differentiable loss function. This algorithm is known for its high predictive accuracy along with its ability to manage complex data and data structure as well.

Adaptive Boosting (AdaBoost): An ensemble method that combines weak learners to create a strong classifier. Our AdaBoost model combined 100 weak learners, typically decision trees, into a strong model.
Each of these models was trained on the training dataset and evaluated on the test dataset based on accuracy, precision, recall, and F1 score. The detailed results are discussed in the following sections.

Ridge Regression : Among all the linear regression methods, ridge regression is one in which the penalty term is directly added to the OLS objective function. That penalty is just the sum of squared coefficients multiplied by a regularization parameter, alpha. That is L2 regularization. Adding this penalty to the objective function shrinks the coefficients toward zero, hence decreasing their variance at the expense of a slight bias. Because of this, it has been particularly useful when facing multicollinearity in the data.

Lasso : LASSO means Least Absolute Shrinkage and Selection Operator and is yet another linear regression-one with an applied penalty term. The L1 Regularization: In L1 regularization, one uses the sum of absolute values of coefficients multiplied by a regularization parameter. Because of this penalty, some of the coefficients are forced to be precisely zero. Therefore, the feature selection is done inherently. That becomes really useful when the number of features is huge and you want to find the most predictive ones.

ElasticNet : ElasticNet in its very nature is a combination of penalties of Ridge Regression and Lasso. It introduces L1 and L2 regularizations in a linear way by using two different parameters: alpha and l1_ratio. Thus, ElasticNet inherited all the advantages of both handled multicollinearity as well as Ridge and performed feature selection like Lasso. It is an elastic all-around method.

XGBoost : Another powerful ensemble method is that of XGBoost; the trees are built one after another in a sequence, with each round trying to correct some error made in the previous one. The XGBoost algorithm implements a gradient-boosting algorithm that progressively and additively fits toward an optimum loss function. The XGBoost algorithm makes use of various regularization techniques to avoid overfitting and is known to be quite fast and accurate.

4.12.3 Preliminary Analysis

We present in this section, a comparison of the performance of six distinct machine learning models in predicting poverty from Nasa's Night Light where the target variable is wealth index . These learning models are Decision Tree, Random Forests, Naive Bayes, K-NN (K-Nearest Neighbors), Gradient Boosting Classification, and AdaBoost Classifier. We have used these four measures to evaluate the models: Accuracy, r2_Score, mse (Mean Squared Error), and rmse (Root Mean Squared Error).Following table shows us the concise result for the preliminary analysis.

From the analysis Decision Tree, Random Forests, and Gradient Boosting Classification models have the best accuracy of 0.637 but the negative r-square values suggest the poor fitting of the data in the model. The Naive Bayes and K-NN models shows the same poorer accuracy and worse r2_scores, showing their unsuitability for this application.The mean squared error (MSE) and root mean squared error (RMSE) results also projects little variation across all models, with MSE ranging from 6.333 to 6.664 and RMSE ranging from 2.516 to 2.581.

Model Name	Accuracy	r2_Score	mse	rmse
Decision Tree	0.637	-0.4453	6.333	2.516
Random Forests	0.637	-0.446	6.338	2.517
Naive Bayes	0.586	-0.505	6.596	2.568
K-NN	0.613	-0.521	6.664	2.581
Gradient Boosting Classification	0.637	-0.445	6.334	2.516
AdaBoost Classifier	0.636	-0.448	6.334	2.518

Table 4.1: The Performance of Six Distinct Machine Learning models for Nighttime Luminosity Predictor

Based on this first step of research, it can be inferred that none of these machine learning models demonstrate significant efficacy and accuracy in forecasting the wealth index of the DHS clusters using the NTL data. We must find out other methodologies or characteristics to enhance the Precision, accuracy and r2_score of the models. We must also take into account the constraints and presuppositions of each of these models, and how they can impactfully contribute to understanding of the outcomes.

Chapter 5

Result and Discussion

5.1 Early Stage of Development

In the Preliminary analysis we have focused on predicting poverty with the help of Nasa Black Marble yearly Night Light Data and DHS household data . Through this we tried to evaluate the status of this research and find out any gaps that our research may address in future steps.In this step we collected Night light data and DHS data , cleaned it with exploratory data analysis to find out any outliers or inconvenience in our dataset. This EDA helps us to understand the data structure and connection of the dataset. We must then create my research approach. My preliminary data analysis and literature study will serve as the foundation for this. Our technique will specify the precise actions I'll take to fulfill my research goals. It will be comprehensible, specific, and repeatable. While our technique is being developed, we also need to choose the machine learning models that I will use to predict wealth. Decision Tree, KNN, Naive Bayes, Random Forest, and AdaBoost Classifier are a few examples of this. These models must be trained using a portion of my data, and their effectiveness must be assessed. We will be able to improve our methods and get an indication of how well my models are expected to function thanks to this preliminary model training and assessment. We may need to make revisions to my study strategy in light of the outcomes of our first model training and assessment. This can include choosing other models, adjusting my process, or gathering more information. This iterative procedure is essential to guaranteeing the accuracy and dependability of our findings and is a typical aspect of the research method. Lastly, we need to draft an outline for our essay. All the parts that will be included in my final paper, including the introduction, methods, findings, and conclusion, should be included in this. A well-defined framework will facilitate our writing process and guarantee that my paper is cohesive and well-structured. Keep in mind that exploration and comprehension are the main goals of the early stages of inquiry. As we gain more knowledge about our study subjects and data, we shouldn't be scared to make changes to our strategy.
Lackings: DHS data are utilized to survey the poverty area. Bangladesh has a high data scarcity, an accuracy rate, and almost zero nighttime light intensity. This data distribution is thus not optimal for further research. Thus, in order to get additional data about the poor region, we must work on a future plan.

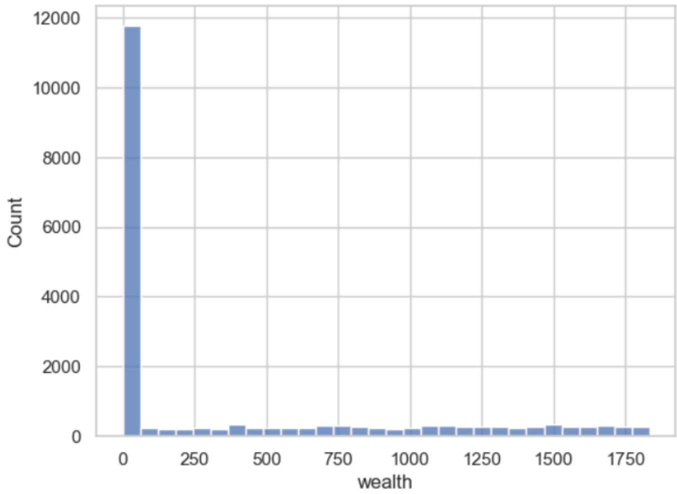

Figure 5.1: Wealth Distribution in Household Clusters

In this figure, wealth is divided into 7 parts, where the 0-150 wealth rate is the poorest and 1500¡ is considered the richest. The histogram illustrates the 'wealth' column from a dataset, with the x-axis denoting various wealth levels and the y-axis indicating the frequency of occurrences. Significantly, the majority of data points reside under the lowest wealth category, signifying a concentration of diminished wealth values. This visualization clarifies wealth differences within the dataset.

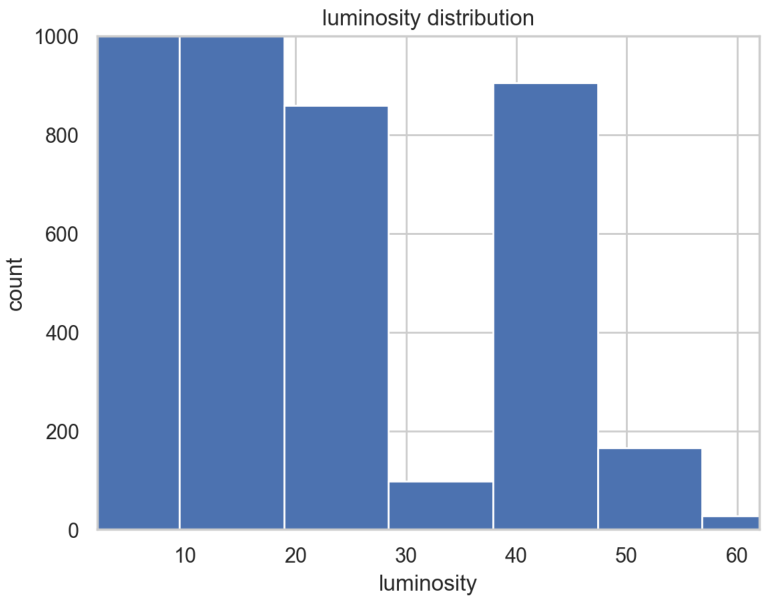

Figure 5.2: Luminosity-Count (Luminosity Distribution Level)

The histogram illustrates the distribution of luminosity levels. The x-axis denotes luminance, spanning from 2 to 62, and the y-axis illustrates the frequency of occurrences. The histogram consists of many bars, each denoting a range of brightness levels. The highest bar probably represents the most common brightness range. This image clarifies the distribution of brightness values within the dataset.

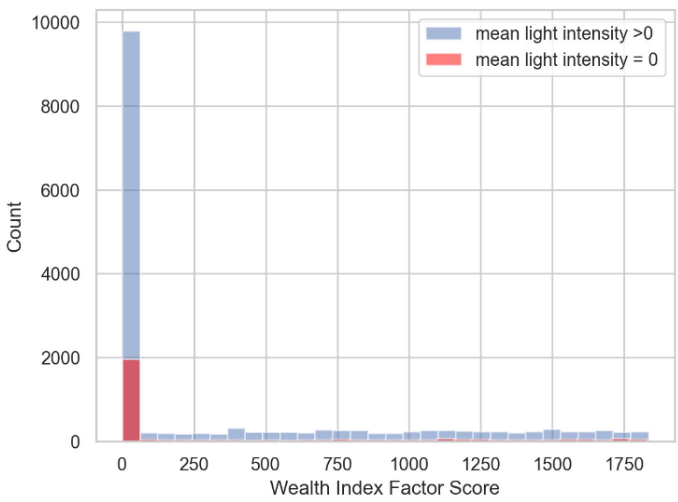

Figure 5.3: Wealth Index Factor Score and mean light intensity

From this figure we can see, Wealth Index Factor on the basis of mean light intensity. The histogram displays two overlay datasets: one for mean light intensity exceeding 0 (shown by blue bars) and another for mean light intensity equal to 0 (indicated by red bars). The x-axis represents the "Wealth Index Factor Score," whilst the y-axis indicates the frequency of occurrences. This image facilitates the comparison of income distribution patterns according to varying mean light intensity levels, offering insights into their interrelation. Here, mean light intensity = 0 is considered as the poorest area with almost no light area. On the other hand, mean light intensity ¿0 coincides as a rich area. In this term Zero light intensity is 15.92%

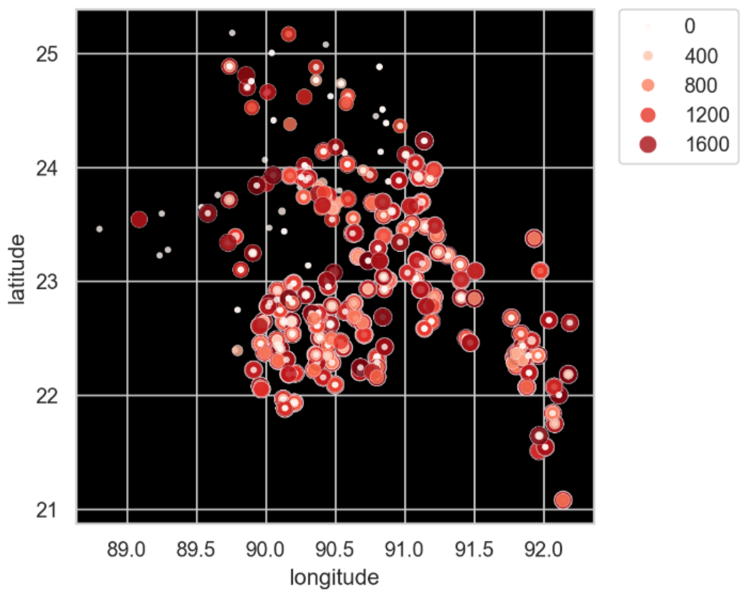

Figure 5.4: Longitude - Latitude VS Wealth Rate

In this figure we have shown the wealth index on the basis of Longitude - Latitude area. The scatter plot features circles of diverse sizes and colors of red distributed on a coordinate grid. These circles denote varying levels of affluence according to latitude and longitude coordinates. The larger circles represent elevated wealth values, whereas regions with denser data point concentrations indicate wealth clusters. This image clarifies geographical variations in wealth distribution, highlighting the correlation between location and economic prosperity.

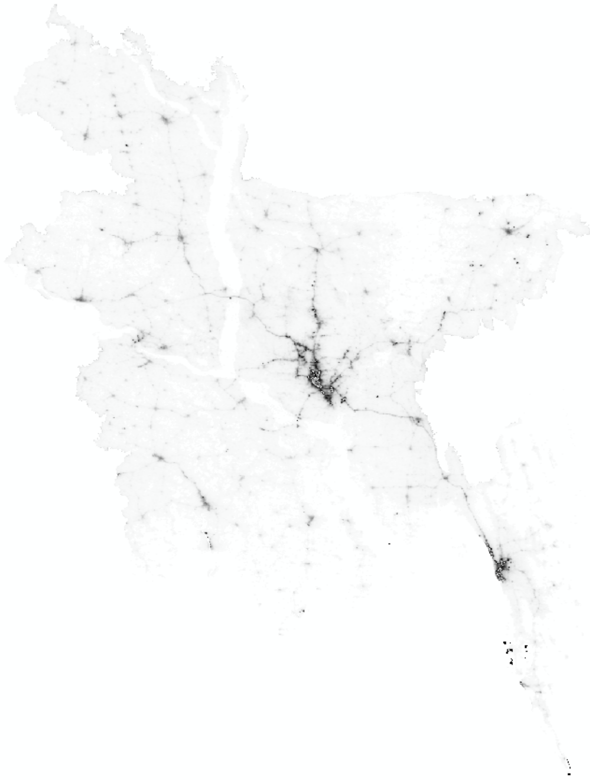

Figure 5.5: Map of the Bangladesh Inverted Night Light Data (link)

We have displayed an inverted map of NASA night light image data in this image.Areas depicted in the image that appear darker at night in reality are actually brighter at night.

5.2 Post Stage of Development

5.2.1 Nighttime Luminosity Modeling

After implementing significant variations in light illumination from the nighttime satellite view data, there seems to be an important connection between the features and the aim of wealth index level in household clusters. This encourages the concept that nighttime light intensity can be used as an alternative for calculating wealth and socio-economic health. As expected, the data points are placed jointly at the lower end of the range for illumination as well as poverty level.

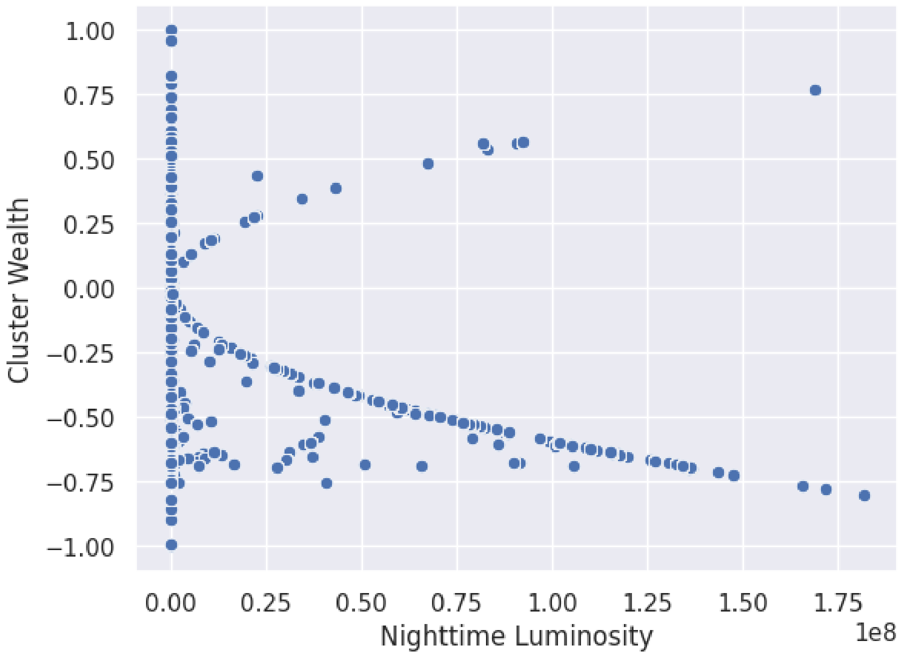

Figure 5.6: Average Nighttime Luminosity Correlation with Wealth Distribution in Household Clusters

The wealth index of household clusters exhibits significant correlation with the right nightlight characteristics that have been determined from the satellite picture. This helps for the argument for using the quantity of illumination at night as a stand-in for assessing wealth and social and economic well-being. In this research, five distinct models—Ridge Regression, Lasso Regression, ElasticNet, Random Forest Regressor, and XGBoost—were created and tested using a ten-fold cross-validation approach. The results show that all of the models worked regarding the same. The XGBoost had the most predictive power, explaining 98.69% of the disparity in social status (R-squared score).

Method	**R-squared**
Ridge Regression	17.78%
Lasso	0.58%
ElasticNet	7.55%
Random Forest	43.54%
XGBoos	98.69%

Table 5.1: R-squared Model for Nighttime Luminosity Predictor

Below is the comparison of performances for five different machine learning models on a continuous target variable. According to the closer $R\hat{2}$ scores to 1, the best fits to the data are Decision Tree and Gradient Boosting. Also, quite well was the SVR, while K-Nearest Neighbors and AdaBoost were less accurate. Again, these models may be improved by the usage of other techniques such as hyperparameter tuning and cross-validation.

Method	R-squared	Best Parameters
KNN	91.74%	{'n_neighbors': 3, 'weights': 'distance'}
SVR	97.70%	{'C': 0.1, 'kernel': 'linear'}
Decision Tree	99.93%	{'max_depth': None}
Gradient Boosting	99.93%	{'learning_rate': 0.1, 'n_estimators': 200}
AdaBoost	99.30%	{'learning_rate': 1.0, 'n_estimators': 200}

Table 5.2: R-squared values and best parameters for the nighttime luminosity predictor after fine-tuning.

5.2.2 Daytime Satellite Image Modeling

First, basic traits are compiled from the Google daytime satellite images. Those include the RGB pixel color data max, min, mean, median, and deviations from the mean. The wealth level of the different residence clusters can be determined by applying these features to machine learning models. The outcome was expected as daylight satellite images' basic luminance can't show the amount of electricity used by people, making it an unreliable indicator for estimating socioeconomic well-being.

Basically a confusion matrix is a tool which can be used to assess the effectiveness of the classification model by comparing the true class and the predicted class of any model. In this particular confusion matrix, we are working with three classes based on the amount of building and tree area. Class 0 stands for the areas with only trees with no buildings,Class 1 is areas partly occupied by trees and smaller extent of buildings while class 2 is areas fully occupied by buildings with no trees. This table shows how the model has predicted these classes according to the true labels Save Table Image File The matrix can be explained in terms of the following: From this matrix, we see that the model got an accuracy of about 80% that implies that the model gave correct prediction 80% of the time.

Based on the confusion matrix, the model predicts the best in Class 0 (only trees), with 6534 true positives which gives correct predictions. However, it wrongly labeled 859 images as Class 1 (mixed area) when they were actually True Negative, and labeled 4 images as Class 2 (building-only area) when actually they are False Negative. In Class 1(mixed areas), there were 1329 True Positive (correct Prediction) but 972 tree-only areas and 75 building-only areas were also considered as wrong (True Negative).Lastly, for Class 2 (building-only areas), the model predicted 170 samples correctly but assigned a classification of building-only to 23 tree-only regions and 75 mixed areas. These numbers demonstrate how the classification went for each of the classes and reveal a trend of the confusion between mixed and building-only regions with other classes especially Class 1.As it can be seen, Class 0 produces the most accurate outcomes owing to the high number of correctly identified true positives, while the results for Class 2 are the lowest. The True Negatives (TN) are instances for which the model could correctly predict that a region did not belong to any particular class improving the general accuracy, especially for Class 0.

This research shows that the model worked at 80% efficiency; therefore, 80% of the time the model was accurate in its predictions. As you can see, this is a good result, however, it should be mentioned that in most of the machine learning cases, it's rather hard to achieve 100% accuracy of the given data due to some limitations. Another limitation that was encountered in this work was the computational cost especially when training large data sets. For this model, 4 epochs have been performed, and fine-tuning was applied to enhance the model's effectiveness. But it is possible to get more accuracy by tuning hyperparameters and by performing more epochs, though this comes with a huge cost. Due to the large size of the data set, the future study can try to enhance the process to bring enhancement into accuracy while at the same time ensuring high computational complexity.

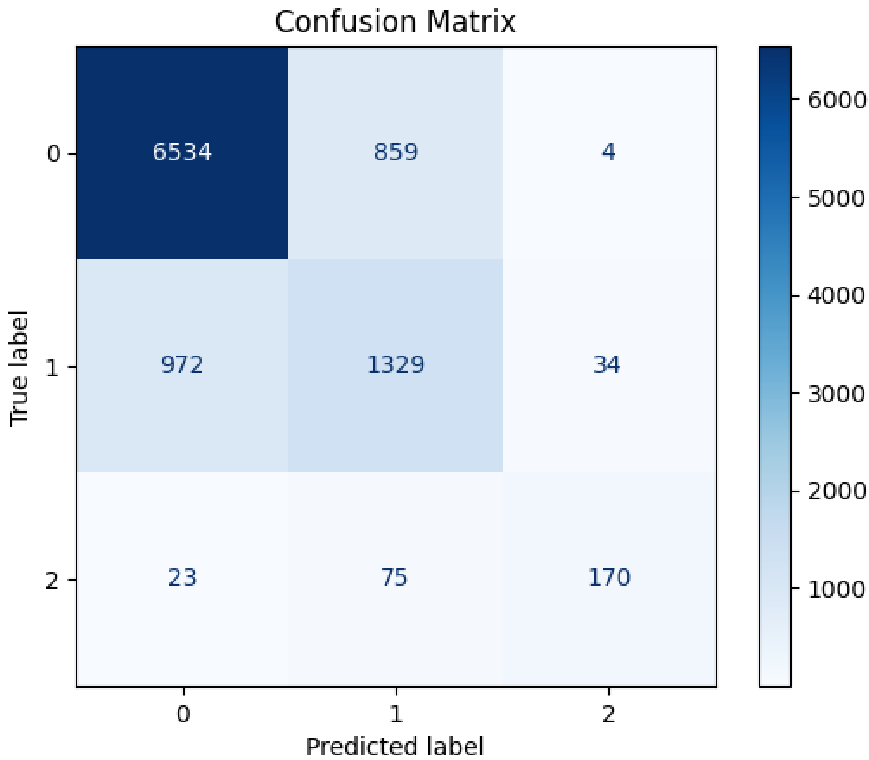

Figure 5.7: Confusion Matrix for 3 classes

Convolutional Neural Network (CNN) architecture with pre-trained weights from the VGG16 model was used to extract deep features from daylight satellite images in order to enhance the performance of the model. By collecting specific geographic features surrounding residence clusters, this approach generates tabular data that may be utilized for fitting models. PCA, or principal component analysis, was implemented to reduce the dimension of these features as a sanity check. PCA1, that explained most of the variance, was contrasted to the average wealth index. Since data on poor housing is common, it is not unexpected that the vast majority of data points clustered at lower wealth levels. The model utilizing nighttime luminosity as a stand-alone predictor produced a similar result, showing that both approaches effectively reflect the link between each of their input features and the target variable.

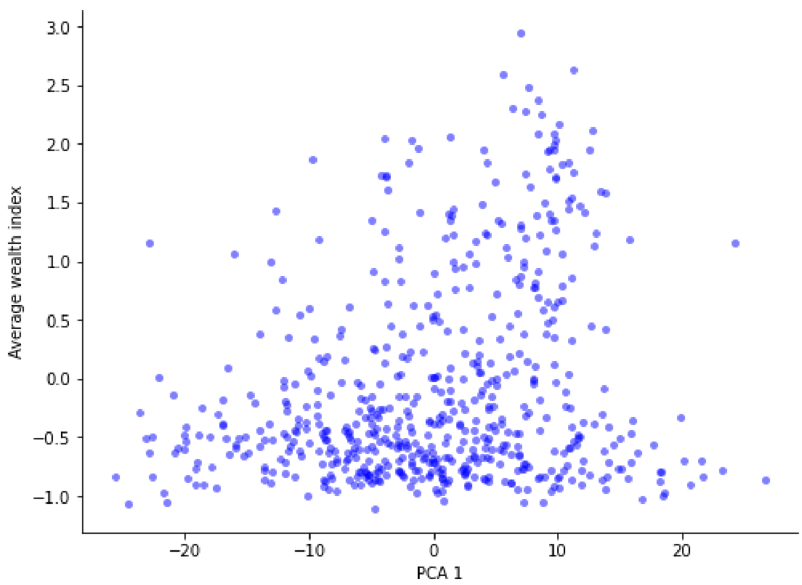

Figure 5.8: Average Wealth Index against PCA

5.2.3 OSM Model

The relationship between geographic features and wealth levels was assessed in Bangladesh using OpenStreetMap (OSM) data. There were over 260 specific features that we featured in four categories.

RoadsFeatures: Type of roads walking service primary secondary etc., length MOI type road, distance to the nearest road from center cluster
Structure Feature: Buildings, e.g. residential or commercial are a potential representation of area development and wealth;
Land Use: What land is used for (parks, forests or residential) and its connection to socio-economic well-being.
POI (Point of Interest) Features: Such as hospitals, schools or markets which reflect the investment on facilities by an area and affluence.

It was conjectured that these human features have a high correlation with the socioeconomic status of neighboring populations. Spearman and Pearson's correlation coefficients were computed, showing that certain building features regarding low-cost structures agriculture transport in general to have the strongest association with average wealth levels. As a standalone feature space, the OSM data explained 43%-56.99% of variance in socio- economic well-being (in agreement with Pearson correlations results). This results in the possibility of using OSM data to improve predictions for wealth and, if plugged into existing predictors such as nighttime luminosity, can potentially significantly boost predictive power.

Method	R-squared
Ridge Regression	12.58%
Lasso	97.12%
ElasticNet	98.51%
Random Forest	99.921%
XGBoos	89.99%

Table 5.3: R-squared Model for OSM Predictor

Poverty map for Bangladeshi household clusters
The model predicted the poverty map for Bangladesh based on nighttime light luminosity and OSM data. This type of map makes it possible to trace levels of poverty across the nation, since in between survey data points average wealth for any location with known coordinates can be estimated with a reasonable accuracy. This feature can be very useful in helping targeted poverty reduction programs and financial inclusion programmes even where survey data is missing. The figures below exhibit the poverty map which is derived from model predictions.

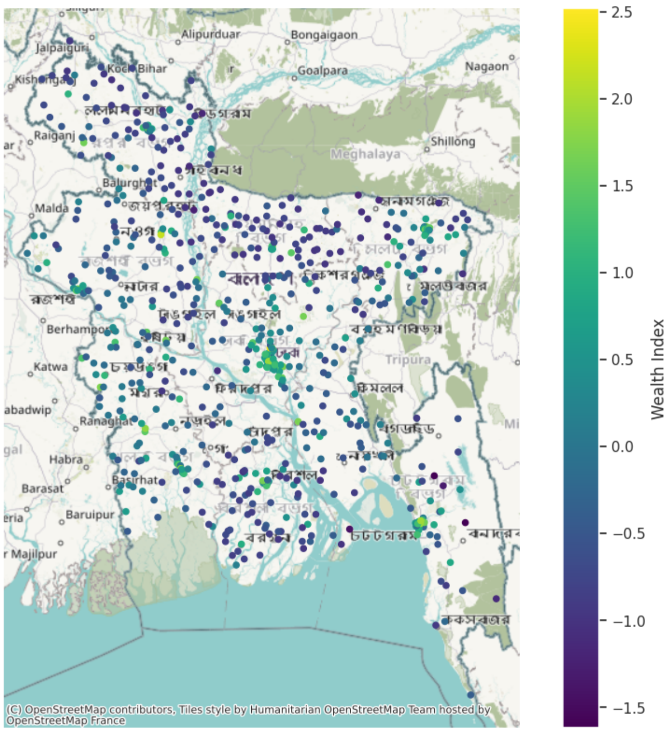

Figure 5.9: Distribution of clusters based on wealth in Bangladesh From DHS Data

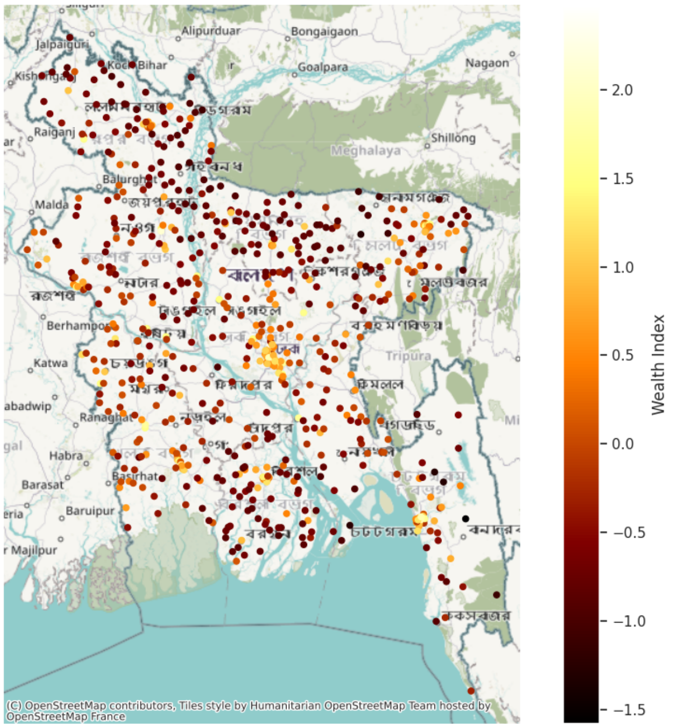

Figure 5.10: Predicted poverty In the provided cluster location

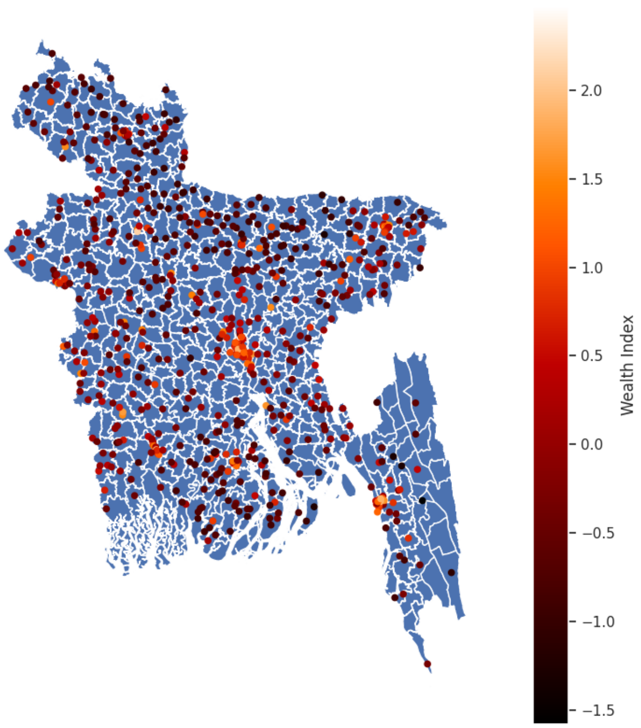

Figure 5.11: Bangladesh's Map of average wealth index prediction for household clusters layered with nightlight emission

Chapter 6

Future plans

In our research we have enhanced the poverty prediction models in Bangladesh with multiple data sources like the DHS surveys and OSM(OpenStreetMap), GSM (Google Static Maps) imagery, shapefiles for Bangladesh, and NTL(NASA's Night-time Lights) dataset. In the early stages, each dataset was distinctly analyzed for prediction, but if we incorporate these data sources with the DHS survey data simultaneously, the results will be further enhanced and the model will be more efficient. But in this case the costing will be much more expensive. Hence for future work we will train the model by incorporating the dataset and by reducing the cost of training by some following methods.

Future poverty analysis research could also be done with other data sources, such as demographic surveys, by considering satellite imagery, cell phone data, and economic indicators to reach even better predictions. Such datasets allow for rich data preprocessing, feature engineering, and machine learning techniques to be performed for high-value information extraction from them[14].

While it is true that the refinement of the DHS data is still on course, in deep learning, multi-dimensional integrations with Night Light and Day Light imagery may find computational problems and yield even better results.

Other measures of poverty involve multidimensional poverty indices, responsive sub-national estimates of the multidimensionality of poverty ; such measures would likely be considered. Longitudinal analysis and integration with time-series data could also be of immense help in this respect.

The overall objective, therefore, should be to devise a multi-source and complex methods-of-analyses integrated robust framework for poverty analysis in Bangladesh.

Chapter 7

Limitations

1. The accuracy of DHS data is very low cause connected and maintain poorly way
2. Computation is expensive and can limit scalability in this high resolution satellite image dataset
3. Any Kind of data loss or presence of noise in the data can affect the accuracy in the model.
4. The nighttime lights may not be useful to determine economic activity especially during the nighttime in the low lighted rural regions.
5. Some of the shortcomings of the model are that low spatial resolution may not capture detailed small-scale poverty trends. Low social, political or cultural factors in poverty are not considered in the model.
6. Another weakness is that the model used in the study may not be generalized when applied to other geographical area

Chapter 8

Conclusion

This study was conducted to improve poverty estimates in Bangladesh through unorthodox data sources and advanced analytical techniques. To overcome the limitations of traditional survey methods, and to compile a more complete and up-to-date picture of poverty over time, researchers leveraged nighttime lights data along with daytime satellite imagery, as well as OpenStreetMap (OSM) infrastructure[32].

In this paper, we provide evidence for the usefulness of alternative data sources in predicting poverty through using a two-stage approach. First pass analysis using NASA Black Marble Nighttime Lights dataset and DHS data suffered from inefficiencies and inaccuracy[34]. The study nevertheless used other 2022 datasets such as OSM (Open Street Map), Google Static Maps and NASA's VNP46A3 to provide more accurate poverty estimation[6].

We have achieved 80% accuracy by splitting Bangladesh into 3 different cluster based on intensity of buildings, land-uses and roads.We have used VGG16 for supervised learning as well as Image-net pretrained model[31].

These results provide important insights both for policymakers, development institutions and academia. Here we claim the contribution of this research as part of an initiative addressing world poverty and promoting development more effectively and fair, due to its use (and reuse) of data source open access large leagues in previous work machine learning. By leveraging increasingly available, timely and low-cost satellite technology alongside cutting-edge analytic tools (like machine learning), we have a cost-effective means to track poverty for quicker real-time decisions — helping us help in more targeted ways towards communities under struggle.

By exploring new data sources and refining analytical techniques, future research may broaden on this foundation and improve the accuracy and degree of data in poverty estimation. In the end, this research can help accomplish Sustainable Development Goal 1 by providing decision-makers with the information required to develop and implement more efficient strategies for eradicating poverty[39].

Bibliography

[1] I. Bhandari, E. Colet, J. Parker, Z. Pines, R. Pratap, and K. Ramanujam, "Advanced scout: Data mining and knowledge discovery in nba data," *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 121–125, Mar. 1997, ISSN: 1573-756X. DOI: 10.1023/A:1009782106822.

[2] D. R. Choudhury, P. Bhargava, Reena, and S. Kain, "Use of artificial neural networks for predicting the outcome of cricket tournaments," *International Journal of Sports Science and Engineering*, vol. 1, no. 2, pp. 87–96, 2007, ISSN: 1750-9823.

[3] *List of roads in bangladesh*, 2011. [Online]. Available: https://en.wikipedia.org/wiki/List_of_roads_in_Bangladesh.

[4] M. Schreiner, "Estimating expenditure-based poverty from the bangladesh demographic and health survey," *BIDS*, vol. XXXIV, no. 4, 2011, Accessed: https://bids.org.bd/uploads/publication/BDS/34/34-4/04Schreiner.pdf.

[5] S. Akhtar and P. Scarf, "Forecasting test cricket match outcomes in play," *International Journal of Forecasting*, vol. 28, Jul. 2012. DOI: 10.1016/j.ijforecast.2011.08.005.

[6] NASA, *Nasa's black marble*, Accessed: https://blackmarble.gsfc.nasa.gov/, 2012.

[7] G. Gartheeban and J. Guttag, "A data-driven method for in-game decision making in mlb: When to pull a starting pitcher," Aug. 2013, pp. 973–979. DOI: 10.1145/2487575.2487660.

[8] T. Tulabandhula and C. Rudin, "Tire changes, fresh air, and yellow flags: Challenges in predictive analytics for professional racing," *Big Data*, vol. 2, no. 2, pp. 97–112, 2014, PMID: 27442303. DOI: 10.1089/big.2014.0018.

[9] V. Veppur Sankaranarayanan, J. Sattar, and L. Lakshmanan, "Auto-play: A data mining approach to odi cricket simulation and prediction," Apr. 2014. DOI: 10.1137/1.9781611973440.121.

[10] A. Leibrand, N. Sadoff, *et al.*, "Using earth observations to help developing countries improve access to reliable, sustainable, and modern energy," *Frontiers in Environmental Science*, vol. 7, 2019. DOI: 10.3389/fenvs.2019.00123. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fenvs.2019.00123/full.

[11] S. Piaggesi, L. Gauvin, *et al.*, "Predicting city poverty using satellite imagery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 90–96.

[12] H. Rahman and R. Hill, "Poverty in urban bangladesh: Trends, profiles and spatial differences," *BIDS*, vol. XLII, no. 2-3, 2019.

[13] Humanitarian Data Exchange, *Bangladesh - subnational administrative boundaries-hdx*, 2020.

[14] D. Huy, "Mapping poverty in bangladesh with satellite images and deep learning," *GitHub*, 2020. [Online]. Available: https://github.com/huydang90/Mapping-Poverty-With-Satellite-Images/blob/master/Final%20Paper/Mapping_Poverty_In_Bangladesh.pdf.

[15] N. Levin, C. Kayba, *et al.*, "Remote sensing of night lights: A review and an outlook for the future," *Remote Sensing of Environment*, vol. 237, 2020. DOI: 10.1016/j.rse.2019.111443.

[16] A. Ditner, C. Stoner, and H. Carter, "Estimating the loss and fragmentation of dark environments in mammal ranges from light pollution," *Biological Conservation*, vol. 257, 2021. DOI: 10.1016/j.biocon.2021.109135.

[17] M. Celik, "What is translation invariance? what is the main difference between invariance and equivariance in cnn? why do convolutional neural networks have translation invariance?," 2022. [Online]. Available: https://celik-muhammed.medium.com/what-is-translation-invariance-3ec5fc4f3cd4.

[18] A. Daoud, F. Jordan, *et al.*, "Measuring poverty in india with machine learning and remote sensing," *Cornell University*, 2022. DOI: 10.48550/arXiv.2202.00109. [Online]. Available: https://arxiv.org/abs/2202.00109.

[19] X. Li, X. Shang, *et al.*, "Using radiant intensity to characterize the anisotropy of satellite-derived city light at night," *Remote Sensing of Environment*, vol. 271, 2022. DOI: 10.1016/j.rse.2022.112920. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425722000347.

[20] S. Islam, "Utilizing nasa's black marble to measure the change in electrification and regional inequality of dhaka, bangladesh," *ProQuest*, 2023. [Online]. Available: https://search.proquest.com/openview/a6923b6c3fa1d45399a6160df9de4ae9/1?pq-origsite=gscholar&cbl=18750&diss=y.

[21] Z. Shao and X. Li, "Multi-scale estimation of poverty rate using night-time light imagery," *,nternational Journal of Applied Earth Observation and Geoinformation*, vol. 121, 2023. DOI: 10.1016/j.jag.2023.103375. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569843223001991.

[22] X. Tan, R. Chen, *et al.*, "Spatial heterogeneity of uncertainties in daily satellite nighttime light time series," *International Journal of Applied Earth Observation and Geoinformation*, vol. 123, 2023. DOI: 10.1016/j.jag.2023.103484. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569843223003084.

[23] M. Yanto, Y. Wiyandra, *et al.*, "Deep learning approach analysis model prediction and classification poverty status," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 1, pp. 459–468, 2023. [Online]. Available: https://d1wqtxts1xzle7.cloudfront.net/94993773/13567-libre.pdf?1669683695=&response-content-disposition=inline%3B+filename%3DDeep_learning_approach_analysis_model_pr.pdf&Expires=1728946021&Signature.

[24] D. A. Borthakur, "Quota reform movement in bangladesh: A deep dive into its origin and impact," *Vivekananda International Foundation*, 2024. [Online]. Available: https://www.vifindia.org/article/2024/august/2/Quota-Reform-Movement-in-Bangladesh#:~:text=The%20curfew%20and%20internet%20disruption,which%20accounts%20for%20about%2085.

[25] "Convolutional neural network (cnn) in machine learning," 2024. [Online]. Available: https://www.geeksforgeeks.org/convolutional-neural-network-cnn-in-machine-learning/.

[26] "Key:building," 2024. [Online]. Available: https://wiki.openstreetmap.org/wiki/Key:building.

[27] "Land use," 2024. [Online]. Available: https://wiki.openstreetmap.org/wiki/Land_use.

[28] n.d., "Demographic and health survey 2022," vol. 1, p. 16 374, 2024. [Online]. Available: https://microdata.worldbank.org/index.php/catalog/6290/study-description#study_desc1684331110642.

[29] "Points of interest," 2024. [Online]. Available: https://wiki.openstreetmap.org/wiki/Points_of_interest.

[30] "The world bank in bangladesh," 2024. [Online]. Available: https://www.worldbank.org/en/country/bangladesh/overview.

[31] "Vgg-16 — cnn model," 2024. [Online]. Available: https://www.geeksforgeeks.org/vgg-16-cnn-model/.

[32] "About google static maps api," [Online]. Available: http://www.birdtheme.org/useful/staticmaps.html.

[33] "Find out how bangladesh is using satellite imagery to better address national poverty levels," [Online]. Available: https://unstats.un.org/UNSDWebsite/capacity-development/data-for-now/story-details/bangladesh-use-satellite-image-poverty-estimation.

[34] n.d., "The dhs program," [Online]. Available: https://dhsprogram.com/what-we-do/survey-Types/dHs.cfm.

[35] "Openstreetmap," [Online]. Available: https://www.openstreetmap.org/#map=6/23.72/90.35.

[36] *Openstreetmap*. [Online]. Available: https://www.openstreetmap.org/.

[37] "Register to wekeo services to enjoy the full potential of sentinel mission and copernicus environmental data," [Online]. Available: https://www.wekeo.eu/subscription-benefits.

[38] M. Ricon, S. Azad, *et al.*, "On the use of satellite nightlights for power outages prediction," *Power Outages*, vol. 10, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9706181.

[39] "Sustainable development goals," [Online]. Available: https://www.un.org/sustainabledevelopment/poverty/.

[40] "Vnp46a3 - viirs/npp lunar brdf-adjusted nighttime lights monthly l3 global 15 arc second linear lat lon grid," DOI: 10.5067/VIIRS/VNP46A3.001. [Online]. Available: https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/VNP46A3/.

[41] "Vnp46a4 - viirs/npp lunar brdf-adjusted nighttime lights yearly l3 global 15 arc second linear lat lon grid," Accessed: https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/VNP46A4/.