Poverty Mapping In Bangladesh Using Geospatial Data Distribution And Machine Learning Techniques

MD Ohidur Rahman

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	MD Ohidur Rahman
Date of Birth	:	21st January 2002
Title	:	Poverty Mapping In Bangladesh Using Geospatial Data Distribution And Machine Learning Techniques
Academic Session	:	Spring 2022

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☐	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

_____
(Student's Signature)

221-35-902
_____
Student ID
Date:

_____
(Supervisor's Signature)

Mr. Kazi Rifat Ahmed
_____
Name of Supervisor
Date:

NOTE : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

THESIS DECLARATION LETTER (OPTIONAL)

Librarian,
Daffodil International University,
Daffodil Smart City,
Ashulia.Dhaka,Bangladesh

Dear Sir,

CLASSIFICATION OF THESIS AS RESTRICTED

Please be informed that the following thesis is classified as RESTRICTED for a period of three (3) years from the date of this letter. The reasons for this classification are as listed below.

Author's Name :	MD Ohidur Rahman
Thesis Title	Poverty Mapping In Bangladesh Using Geospatial Data Distribution And Machine Learning Techniques
Reasons	(i)
	(ii)
	(iii)

Thank you.

Yours faithfully,

_____
(Supervisor's Signature)

Date: 24/12/2025

Stamp:

Note: This letter should be written by the supervisor and addressed to the Librarian, *Daffodil International University* with its copy attached to the thesis.

APPROVAL FORM

APPROVAL

This thesis titled on **Poverty Mapping In Bangladesh Using Geospatial Data Distribution And Machine Learning Techniques**, submitted by **MD Ohidur Rahman (ID: 221-35-902)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. S. M. Hasan Mahmud
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

13.12.23

Internal Examiner 1

A.H.M Shahariar Parvez
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Tapushe Rabaya Toma
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 3

Khalid Been md. Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Dr. Md Sazzadur Rahman
Professor
Institute of Information technology
Jahangirnagar University, Bangladesh

SUPERVISOR's DECLARATION

I/We* hereby declare that I/We* have checked this thesis/project* and in my/our* opinion, this thesis/project* is adequate in terms of scope and quality for the award of the degree of *Bachelor of Science/ Master of Science.

(Supervisor's Signature)

Full Name : Mr. Kazi Rifat Ahmed

Position : Lecturer

Date :

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

(Student's Signature)

Full Name : MD Ohidur Rahman

ID Number : 221-35-902

Date :

Poverty Mapping In Bangladesh Using Geospatial Data Distribution And Machine Learning Techniques

MD Ohidur Rahman

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

November 2025

ACKNOWLEDGEMENTS

All praise and gratitude are due to Almighty Allah, the Most Gracious and the Most Merciful, for granting me the strength, patience, and opportunity to complete this thesis. Without His blessings and guidance, this work would not have been possible.

I would like to express my sincere appreciation to my respected supervisor for his continuous support, valuable guidance, constructive feedback, and patience throughout the entire research process. His insightful suggestions and encouragement played a crucial role in shaping this work and helping me overcome every challenge along the way.

Finally, I extend my heartfelt gratitude to my family for their unwavering love, prayers, and constant encouragement. Their support has been my greatest source of strength and motivation throughout my academic journey. This achievement is as much theirs as it is mine.

DEDICATION

This work is dedicated to the countless individuals and families whose lives are shaped by poverty and inequality. May this research contribute, even in a small way, to the understanding, awareness, and efforts needed to build a more just and equitable world for them

ABSTRACT

In order to make proper policies, measurement of poverty needs to be accurate, but conventional survey-based instruments like the Demographic and Health Survey (DHS) are not only expensive, but also scarce and sparse spatial-resolution instruments. This paper will look at the question of whether the geospatial features of poverty in Bangladesh can be estimated using freely available features as an alternative. Based on the ground truth (DHS 2018 household recode data), several geospatial indicators, including Night-Time Lights (NTL) and Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), rainfall, land cover distribution, and Point-of-Interest (POI) density, were obtained including 672 survey clusters on Google Earth Engine and other open sources. Two measures of poverty were tubed, such as the Make-go of Wealth Index and Wealth Quintile, which were predicted by a series of machine learning models, such as the Generalized Least Squares (GLS), Random Forest Regressor (RFR), Support Vector regression (SVR), the XGBoost, the KNN, Decision Tree, Gradient Boosting, and Multi-layer perceptron (MLP) models. The R2, RMSE, MAE evaluation of the Wealth Index and Accuracy, Precision, and Recall evaluation of the Wealth Quintile prove that geo deterministic based models can be used to have a significant approximation of the official estimates of DHS. Random forest performed better among all the other models with the lowest RMSE of 365.220439 and the highest R2 of 0.759106. Compared to the other predictive models which had an accuracy of 69 and 53 percent, the predictive model Random forest was the most appropriate to be used in mapping poverty with a 86 percent accuracy. The feature importance analysis revealed that NTL, land cover, NDVI and POI accessibility have the highest role in predicting poverty.

The results verify multi-source geospatial features are a cost-efficient and timely technique of mapping poverty.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

y — Actual dependent variable (wealth index or mean wealth quintile)

$\hat{y}$— Predicted value

X — Feature matrix (geospatial predictors)

x — Single input sample

$\hat{\beta}$ — Regression coefficients

ε — Error term

n — Number of samples

p — Number of features

Ω — Covariance matrix of error terms

αi — Lagrange multipliers

γ — RBF kernel width parameter

ϵ — SVR margin parameter

λ — Regularization strength

w_j — Weight of leaf j

LIST OF ABBREVIATIONS

ACC — Accuracy

AI — Artificial Intelligence

API — Application Programming Interface

BBS — Bangladesh Bureau of Statistics

BUI — Built-Up Index

CNN — Convolutional Neural Network

CO — Carbon Monoxide

CPU — Central Processing Unit

CSV — Comma-Separated Values

DEM — Digital Elevation Model

DHS — Demographic and Health Survey

DNN — Deep Neural Network

DSM — Digital Surface Model

DT — Decision Tree

EO — Earth Observation

F1 — F1-Score

GBR — Gradient Boosting Regressor

GEE — Google Earth Engine

GIS — Geographic Information System

GLS — Generalized Least Squares

GPS — Global Positioning System

GPU — Graphics Processing Unit

HCI — Human Capital Index

HCR — Head Count Ratio

HI — Household Income

HRI — High-Resolution Imagery

JSON — JavaScript Object Notation

KNN — K-Nearest Neighbors

LR — Linear Regression

LST — Land Surface Temperature

MAE — Mean Absolute Error

ML — Machine Learning

MLP — Multi-Layer Perceptron

MPI — Multidimensional Poverty Index

MSE — Mean Squared Error

NDVI — Normalized Difference Vegetation Index

NDWI — Normalized Difference Water Index

NGO — Non-Governmental Organization

NO_2 — Nitrogen Dioxide

NTL — Night-Time Light

OSM — OpenStreetMap

POI — Point of Interest

PR — Precision

RAM — Random Access Memory

REC — Recall

RF / RFR — Random Forest / Random Forest Regressor

RMSE — Root Mean Squared Error

SDG — Sustainable Development Goal

SES — Socioeconomic Status

SO_2 — Sulfur Dioxide

SR — Surface Reflectance

SVR — Support Vector Regression

UI — User Interface

UNDP — United Nations Development Programme

VIIRS — Visible Infrared Imaging Radiometer Suite

WI — Wealth Index

WQ — Wealth Quintile

XGB / XGBoost — Extreme Gradient Boosting

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background:

Poverty has remained one of the most perennial issues around the globe affecting billions of individuals and hampering sustainable growth across the globe. In its effort to minimize poverty, the United Nations (UN) has come up with 17 Sustainable Development Goals (SDGs) to be achieved within 2015 and 2030 constantly including poverty dealing in all forms in the entire world. This is the reason why proper policymaking and specific interventions require timely and accurate mapping of poverty, especially in developing countries, such as Bangladesh. Conventionally collected data is however in most cases, time consuming, costly and space limited, especially using conventional-survey-based data collection. In addition, the survey data can have years before it becomes unavailable in very poor countries (or war-torn countries) (Zhao et al., 2019). It prompts the suggestion of the alternative solution of poverty mapping with the help of freely available geospatial data and satellite image objects (Zheng et al., 2024, Newhouse, 2023, Jean et al., 2016) set of features. With the rising access to high resolution geospatial data, such as satellite imagery, Nighttime Lights (NTL), Normalized Difference Vegetation Index (NDVI), Land Cover, Land Surface Temperature (LST), Rainfalls, Points of Interest (POIs), and various other geospatial distributed information, it is now possible to estimate poverty and wealth at finer spatial levels rather than usual methods, which are often expensive and time intensive. Machine learning and deep learning models have facilitated the precision and promptness of poverty estimates while also increasing the ability to derive significant social economic characteristics of the multiple sources of data (Hu et al., 2022, Zhao et al., 2019, Browne et al., 2021).

1.2 Motivation:

The problem of poverty reduction in Bangladesh stems from poor planning and forecasting of poverty, and more often than not, it does not take the steps necessary that would result in socio-

economic inequality, poor living conditions, an unsanitary ecosystem, malnutrition and preventable diseases (Titumir and R. A. M.,2021, Rahman et al., 2021). The method of poverty mapping is essential to address these problems successfully, whereas traditional census method, which is Bangladesh Demographic and Health Survey (BDHS) of the Demographic Health Survey program and the Household Income and Expenditure Survey (HIES) by the Bangladesh Bureau of Statistics (BBS) is carried out every 3-5 years. This creates a massive vacuity to know the present poverty situation of the state of the country. The updated geospatial data that is given out on a weekly or daily basis using satellite platforms is a potentially good alternative to coming up with the right estimates of the prevalence of poverty at the right time. Spatial resolution of poverty at high resolution will, consequently, be invaluable in empowering government agencies, non-governmental organizations, development planners and policy makers to discover the vulnerable locations and/or environments thereafter formulating focused interventions. Once more, the modified approach for managing executive coaching has secured benefits within strategic human resources management. The adjusted method of executive coaching management has again attained gains in strategic human resources management.

1.3 Problem Statements:

Although the world and nations have continued to put efforts in alleviating poverty, the most effective way of identifying and tracking the spatial dispersion of poverty has continued to be a challenge to most developing nations, including Bangladesh. Traditional poverty measurement systems, including consumption metrics and multiple measures of poverty, are highly dependent on population survey to measure economic well-being and poverty (Alkire et al., 2011). Although they are both statistically powerful and policy relevant, they both have inherent operational limitations, especially in low and medium-income nations where the costs of data collection are high, labor-intensive and logistically complicated. In Bangladesh, the number of individuals living in poverty is still estimated, mainly based on household surveys that are conducted on a periodic basis, including the Demographic and Health Survey (DHS) and other nationally representative surveys. These surveys are normally done after every period of many years leading to huge gaps in time between consecutive estimates of poverty. Such updates become infrequent and, therefore, restrict the capacity of the policymakers to perform timely monitoring, respond to shocks, or

design the adaptive poverty-reducing interventions (as pointed out) (Newhouse 2023). Thus, small areas estimation techniques based on surveys do not tend to capture quickly changing socioeconomic conditions especially in areas where the climate varies, urbanize, migrate or suffer due to economic shocks. The other significant weakness of conventional methods of poverty mapping is that they have a coarse spatial resolution. Estimates collected by surveys tend to be valid only at administrative scales (i.e., districts or divisions) and hide much heterogeneity on smaller scales (i.e., communities, villages, urban neighborhoods, etc.) (De Nicolo et al., 2023). This spatial aggregation issue is commonly recorded in the developing territories and leads to the problem of failing to identify pockets of localized deprivation, which diminish the efficiency of localized policy interventions (Engstrom et al., 2017; Steele et al., 2017). New opportunities have been created by the recent developments in geospatial technologies and machine learning to overcome these limitations. Recent scientific sources also indicate that remotely sensed data (including nighttime lights, vegetation indices, land surface temperature, rainfall, land cover, and points of interest) have a high correlation with economic activity, infrastructure development, and living conditions (Jean et al., 2016; Browne et al., 2021). These are liberally accessible, timely updated, and spatially explicit data sources and hence are appealing substitutes or supplements to conventional survey data on poverty estimation. Empirical research in a wide range of geographic settings demonstrated encouraging findings in terms of forecasting the level of poverty based on geospatial data and machine learning algorithms at more specific scales. As an example, Zhao et al. (2019) established that in Bangladesh, wealth indices could be accurately predicted with a reasonable error by the random forest models based on multi-source satellite data. Other countries have also reported similar success (Zheng et al., 2024; Putri et al., 2022; Ramadhan et al., 2023; Puttanapong et al., 2020; Tang et al., 2024; Agyemang et al., 2023; and many other African and Southeast Asian countries). Nevertheless, in the current literature, there are still a number of critical gaps and unsolved issues in spite of these advances. To begin with, extensive research is based on mixed methods where household survey data remains a dominant model input and calibration objective, restricting the scalability and autonomy of geospatial-only poverty modeling methods (Newhouse, 2023; Engstrom et al., 2017). Second, most of the available literature concentrates on single or small groups of geospatial data, mostly nighttime lights intensity, which is unlikely to be a sufficient multidimensional measure of poverty- particularly in the rural or agricultural setting where electrification is low (Yong et al., 2022; Yu et al., 2023). Moreover, on

a case-by-case basis deep learning and advanced machine learning architectures have demonstrated higher predictive power (Ni et al., 2021; Ayush et al., 2021; Hu et al., 2022), but it is unclear about how they will perform across diverse regions. Most models are trained and tested in geographically or socioeconomically different settings, so their applicability to other countries, such as Bangladesh, needs to be questioned, as the particular factors of poverty, such as high population density, susceptibility to climate, and informal economy, are unique to it (Hall et al., 2022). Notably, the empirical studies on the possibility of estimating poverty indicators based on the only data on multi-source geospatial data in Bangladesh, without the use of standard socioeconomic variables are rather scarce. Whereas it has been established in the recent past that relationships between features defined by satellites and indices of wealth can be observed in Bangladesh (Zhao et al., 2019), it is yet unknown whether a completely geospatial, machine-learning-intensive strategy results in the accuracy and reliability of these variables being high enough to consider a valid alternative to the classical surveys used to map poverty levels. Consequently, there is a methodological and empirical gap between analyzing the possible study of only geospatial covariates and machine learning models that can be used to produce high-resolution, timely and scalable poverty estimates in Bangladesh. The necessity to fill this gap will help to improve the poverty monitoring systems, better spatial targeting of social programs, and evidence-based policymaking in resource limited environments. In this work, I aim to discuss how these challenges can be overcome through systematic analysis of whether the geospatial data based on multiple sources, as used to signify the signs of poverty in Bangladesh and as used to predict poverty in that nation, in case of using the same, including the methods of machine learning can alleviate or at least in part reduce, the sense of poverty in Bangladesh.

1.4 Research Question:

1. Is the geospatial features the alternative of the old census and survey method of counting poverty?
2. What is the best predictive performance of a machine learning model of poverty in Bangladesh based on geospatial data versus the official data based on R2, MAE and RMS of Wealth index and Wealth Quintile based on Accuracy, Precision, and Recall?
3. Which geospatial characteristics are the most effective in predicting poverty in Bangladesh?

1.5 Objectives:

1. To suggest a scalable geospatial ML model to estimate poverty in low-data third world nations.
2. To construct and test a machine learning model to predict poverty, comparison of prediction and official data in terms of R2, MAE, and RMSE regarding wealth index. Wealth quintile Accuracy, Precision, Recall, and F1-score.
3. To produce a high resolution poverty map of the distribution of poverty in Bangladesh regions.

1.6 Thesis Organization

This thesis follows as follows:

Chapter 1 gives an introduction to the research, motivation behind it and problem statements with clear objectives

Chapter 2 mentions literature reviews in terms of recent geospatial poverty mapping with machine learning methods and gaps (2021-2025).

Chapter 3 gives the methodology, which contains data collection, pre-processing, model selection and development and metrics of evaluation.

Chapter 4 contains the results, visualizations and comparative analyses. The last chapter of **Chapter 5** comprises the conclusion of the key findings, restrictions, the response of the research questions, and future research improvements.

CHAPTER 2

LITERATURE REVIEW

2.1 Related works

As the world targets to eliminate poverty and upheld the 17 Sustainability Goals, many studies have been conducted to reduce poverty and tackle the problem poverty brings. Efficient policy making and timely intervention is the most crucial part in order to reach this goal. Therefore the need of mapping poverty takes the first priority. The advancement to Machine Learning (ML), Deep Learning (DL) and remote sensing data has made this goal within our grasp without intensive work or expensive costs. In this chapter, we demonstrated some of the works that have been previously done related to our work with methodologies varying from traditional machine learning to CNN, to Deep learning, and their strength and limitation in mapping poverty with geospatial or remote sensing data.

Introduced by Putri et al. (2022), the novel approach to mapping poverty on a grid scale allows making the process of mapping faster and cheaper (1.5 km grid). In their initial model, they employed different satellite data which comprised of the following: Nighttime light intensity (NTL) to denote the economic activity, NDVI to denote rural regions, Built-Up Index (BUI) to denote urban regions, NDWI to denote land cover and Land Surface Temperature (LST) to denote urban heat. They also included latest pollution data such as CO, NO2 and SO2 to monitor the industrial activity and conducted point of Interest (POI) data to verify the access to significant infrastructure. They used deep learning tools like ResNet-34 in this model to transfer learning and MLP, daytime multispectral, and NTL. The authors discovered that the CNN-1D model of the former approach and the ResNet-34 + MLP model of the latter approach provided the most reasonable results. The CNN-1D model had a measured RMSE, calculated as 1.95, and a calculated adjusted R2, which was 0.84 in predicting poverty (Putri et al., 2022).

Tingzon et al. examined one example of poverty mapping in developing nations by providing a case of the Philippines (2019). They brought into relief the concerns of boring and sometimes costly traditional modes of collecting socioeconomic information. They were able to predict significant aspects, based on their study based on the data provided by the OpenStreetMap, and nighttime satellite images, including wealth, education and access to utilities, such as electricity and water. The regional point of view indicates that on average 63 per cent of asset-based wealth could be attributed to those models. Besides, the researchers have also concluded that models constructed on the volunteered public data were more effective than the models constructed on the costly private satellite images. This study suggests that open-source statistics may be rather useful to make estimates on poverty.

Zheng et al. (2024) approached the issue of providing a different method of estimating poverty in developing countries by suggesting a Random Forest Regression (RFR) model that integrates various sources of data to estimate poverty more effectively. The research objective was to have a rough measure of poverty using a measure of wealth index (WI) at a 10 km x 10 km spatial scale. The model used many environmental and geographical data such as, NTL data available on the NPP-VIIRS, Day/Night Band (DNB), Google satellites, land cover, road map with and positions of heads of division. It was found that when applied to Bangladesh and Nepal data, the model would predict well with an R2 of 0.70 and 0.61 in Bangladesh and Nepal, respectively, and complete external validity. A notable result was the high contribution of the proximity to urban areas to the explanation of the poverty with a contribution of 37.9 per cent to the explanatory power of the model. The current research identifies the idea of using various data in estimating poverty as more valid and presents a methodology that may be implemented worldwide, even to other parts of the world.

A study by Puttanapong et al. (2022) aimed to utilize the easy-to-access geospatial data to estimate the distribution of poverty in Thailand with the background on such variables as Night-Time Light Intensity (NTL), land cover, vegetation index, land surface temperature, built-up areas, and point of interest. The comparison included several machine learning tools such as the Random Forest (RF), the Support Vector Regression (SVR), the Generalized Least Squares (GLS), and the Neural Networks (NN). The findings revealed that poverty was closely related to the NTL and population density proxies. The most accurate results were obtained with the Random Forest model, since the

values of Root Mean Square Error (RMSE) of 2015 and 2017 were 0.067 and 0.084 respectively, which was more accurate than SVR(RMSE 0.129 and 0.161) and GLS (RMSE 0.133 and 0.170). The Neural Network model has the poorest results where the RMSE values of the two years stand at 0.419 and 0.549 respectively. The article emphasizes the impact and efficiency of the geospatial and machine learning utilization to predict poverty precisely (Puttanapong et al., 2022).

Hu et al. (2022) offered a system to detect the village level poverty in Yunyang County, Hubei Province, China, based on high-resolution imagery (HRI), Point-of-Interest (POI) data, OpenStreetMap (OSM) data, and Digital Surface Model (DSM) data. They studied 338 villages, and pulled out variables of access to facilities, agricultural condition, village building and the geographical distribution of settlements. Regression outcomes obtained by applying the Random Forest regression revealed that the proportion of built-up land and time cost to facilities are the most crucial variables in predicting poverty whereas agricultural conditions of production were not conclusive. The model had an overall accuracy of 54% and highest accuracy of 72% since it predicted poverty in poor villages. This paper showed that the benefits of satellite images and geospatial data might be efficient in recognizing high-poverty areas especially in regions where the statistical facilities are scarce, which underscores the significance of public services and natural situations to predict poverty (Hu et al., 2022).

Gulecha et al. (2024) have touched the issue of outdated ways of collecting poverty data prevalent in India that tend to be expensive, time-intensive, and scalp-less. To deny these drawbacks, the research combined satellite imagery plus geospatial statistics with socio-economic surveys and the Point of Interest (POI) statistics to forecast poverty levels in villages and towns. The algorithms used in the study were a variety of machine learning and deep learning algorithms such as Decision Tree Regressor, Random Forest Regressor (RFR), Convolutional Neural Networks (CNNs) and Multi-layer Perceptron (MLP). The Random Forest Regressor model showed the highest results with R2 of 0.778. The model also provided encouraging outcomes where the RMSE = 10.178 and MAE = 7.077 were obtained in the forecasting of poverty. This experiment shows that machine learning models, especially RFR, are feasible in estimating poverty on a grainy level that policy makers may utilize to direct their efforts of addressing poverty (Gulecha et al., 2024).

Agyemang et al. (2023) achieved the goal of discussing the problem of high-resolution poverty mapping in low- and middle-income countries (LMICs), where no (or insufficient) traditional

survey data is usually available. The research paper presented the transfer learning idea based on a Convolutional neural network (CNN) on satellite images to forecast chronic poverty on a scale of 1 km 2 of rural Sindh, Pakistan. The models were trained on a big, spatially randomized georeferenced house-hold survey with poverty scores of 1.67 million households, as well as publicly accessible information, daytime and nighttime satellite images and accessibility data. Based on hold-out and k-fold validation exercises, the study adopted a collection of three CNN models, which performed better than prior studies on major metrics of accuracy. The superior accuracy of the ensemble model was once again validated by a third test which was a ground-truthing test using survey data of 7,000 households. The findings indicate that simply a scalable and affordable approach would be a substantive method of targeting poverty in Sindh and other LMICs (Agyemang et al., 2023).

To create estimates of poverty at a chosen spatial resolution 10 x 10 km in size, Tang et al. (2024) suggested a model to combine characteristics of several sources of data to estimate poverty: nighttime light remote sensing data, normalized difference vegetation index (NDVI), surface reflectance, land cover type, and slope data. The model was implemented on Nigeria where the wealth index was taken as the measure of poverty based on Demographic and Health Survey (DHS) program. The analysis also included time-series elements obtained over convolutional long short-term memory (LSTM) networks, as it was observed that environmental trends can serve as useful factors to reveal poverty. The model performed well, with regression being higher by 0.73 with 2018 data of Nigeria and 0.69 with the 2021 forecasting of poverty, a sign of good generalization. The accuracy of the model was also checked by validating them by comparing them with high-resolution satellite images, and a multidimensional poverty index at the state-level. The analysis revealed that the use of time-series characteristics enhanced the estimation of poverty, the R 2 rose by 0.64 to 0.73. It provides a relatively inexpensive way of estimating poverty in areas that do not have the historic data (Tang et al., 2024).

Jean et al. (2016) suggested a new and scorable technique to forecast economic activities, including consumption expenditure and asset wealth, with high-resolution satellite imagery. This paper utilized the survey and satellite data of five African nations namely Nigeria, Tanzania, Uganda, Malawi and Rwanda and used a Convolutional Neural Network (CNN) to detect image qualities that describe up to 75 per cent of the variation in local-level economic performance. The research

proved that the technique that uses publicly available data is a good and cost-effective way of estimating economic livelihood in the developing countries. With the help of satellite imagery and machine learning, Jean et al. demonstrated the opportunity to enhance tracking and targeting of the poverty issue in areas where access to the classic data is scarce. The effectiveness of this method despite a lack of training data indicates that it may be applicable to many other areas of science, as a strong instrument to deal with poverty in the conditions of a lack of data (Jean et al., 2016).

Ayush et al. (2020) addressed the challenge of using high-resolution satellite imagery for poverty prediction by proposing a reinforcement learning approach to reduce the cost of acquiring such images. High-resolution imagery, while accurate, is expensive, limiting its scalability for widespread use in tasks like poverty prediction. The authors' innovative approach uses low-resolution imagery to dynamically identify locations where high-resolution images are most needed, minimizing the number of high-resolution images required. Applied to poverty prediction in Uganda, the method outperformed previous benchmarks by using 80% fewer high-resolution images. This approach builds on earlier work that utilized object detection to count objects, which were then used to predict poverty. The results demonstrated that reinforcement learning could make high-resolution-based approaches more cost-effective and scalable, with potential applications across various domains requiring such imagery (Ayush et al., 2020).

2.2 Gaps and proposed solutions

After completing the literature review of related work, their findings, results, identifiable gaps and Proposed solutions are shown at Table 1

Table 1: Gaps and Proposed solution of related works

Research Works	Identifiable Gaps	Proposed Solution	Model	Model Evaluation
Putri et al., 2022	Poor granularity of poverty maps plus using	Multi-source satellite data and POI data based on	CNN, Random Forest, SVR, MLP	CNN-1D achieves better results compared to others; its RMSE = 1.95, and $R^2 = 0.84$.

	conventional sources of data.	high-resolution poverty map.		
Tingzon et al., 2019	Poor quality of data on poverty prediction in rural regions	Poverty mapping based on satellite images and crowd-sourced POI data.	Random Forest, Support Vector Machines (SVM)	Metrics not explicitly specified but results were explained to be Very high precision in the cities but poor rural information processing
Zheng et al., 2024	The models of poverty estimation based on standalone data.	Combination of various data (NTL, Satellite imagery (Google), road maps.).	Random Forest Regression (RFR)	R^2 in both Bangladesh and Nepal is 0.70 and 0.61 respectively; this is good predictive power.
Puttanapong et al., 2020	Lack of predictive models using a combination of geospatial data for poverty	Integration of NTL, land cover, vegetation index, and POI data for poverty mapping	Random Forest, Support Vector Regression (SVR)	Random Forest showed the best performance, with the lowest RMSE values (0.067 and 0.084)
Hu et al., 2022	Unavailability of data of high detail in estimating poverty in areas that have a low level of statistical systems.	Combining HRI, POI, OSM, and DSM to predict poverty in the village level	Random Forest	54 percent overall, 72 percent poor villages.

Gulecha et al., 2024	Expensive traditional ways of collecting poverty data.	Satellite imagery, POI data and socio-economic surveys	Random Forest Regressor (RFR), CNN, MLP	R^2 = 0.778, RMSE = 10.178; best model performance RFR.
Agyemang et al., 2023	Predictions of high-resolution spatial poverty that is not available in rural areas	Transfer learning with CNNs to forecast poverty at 1 km^2	CNN Ensemble using transfer learning	Ensemble model was the most accurate compared to other models
Tang et al., 2024	Poor resolution to poverty.	A combination of time-series satellite feature and convolutional LSTM was used.	Multi-source convolutional LSTM	R^2 0.73 in 2018 and 0.69 in 2021; accuracy significantly increased when time-series features are added
Jean et al., 2016	Unaffordable and scalable ways of poverty estimation in developing countries.	Consumer expenditure and asset wealth are estimated using CNN involving satellite pictures.	Convolutional Neural Network (CNN)	75-percent of the economic outcomes explained.
Ayush et al., 2020	High-resolution imagery is expensive to obtain at large scale.	High-resolution image acquisition optimization via reinforcement learning.	Deep Learning (Object Detection, CNN) with reinforcement learning	80 percent less high-resolution pictures consumed, achieved higher benchmarks than before.

CHAPTER 3

METHODOLOGY

Our main objective of this study is to identify and map areas around Bangladesh to demonstrate and accurately predict poverty. The methodology of our study composes of 5 key modules. Data collection, Data preprocessing, Satellite feature extraction, Model training and model evaluation. This system architecture has been clearly demonstrated in Figure 1.

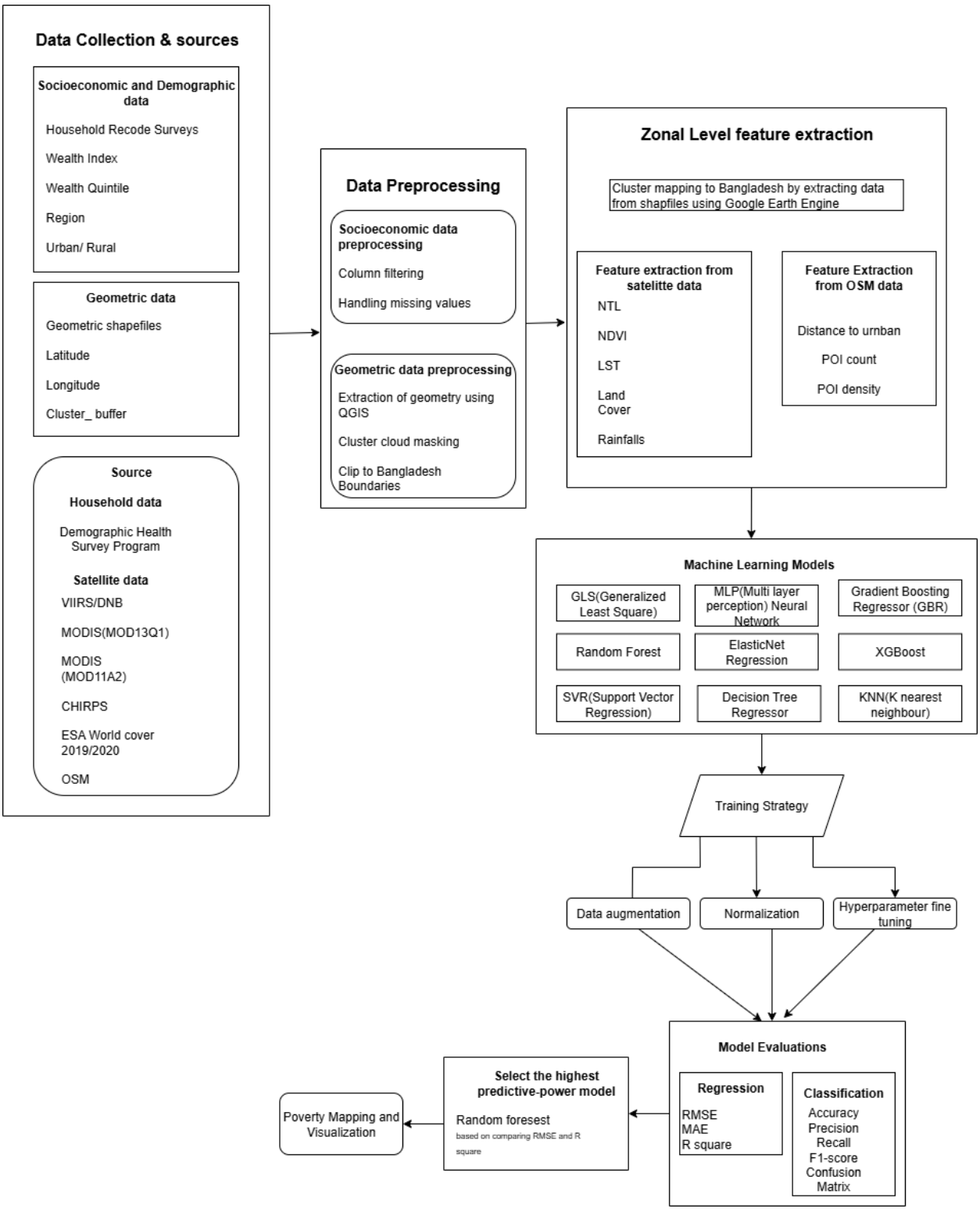

Fig 1: Methodology Architecture

3.1 Data Collection

Two types of datasets were used in this study to prepare our modeling dataset for the models. The study analysis relies on two main categories of data: (1) Target variables (wealth index and wealth quintile) obtained from the Demographic Health Survey (DHS) 2022 standard for Bangladesh and predictor variables extracted from remote sensing geospatial datasets processed via Google Earth engine.

Demographic Health Survey is a reliable source for our ground truth dataset as it is a global research initiative program funded primarily by the United States Agency for International Development (USUAID) "Citations" that has been conducting nationally representative household surveys over 90 poor to developing countries since 1980. DHS constructs the wealth index using Principal Component Analysis on asset ownership and housing qualities making the household recode dataset of DHS a strong indication of poverty levels.

DHS household recode dataset as shown in figure 2, consists data from 19457 households with 3707 features. This dataset contains detailed socio-economic details for each surveyed household across 672 clusters for Bangladesh. Each cluster contains about 28 households on average. The raw dataset contains features such as household size, age and sex of household member, living conditions, assets, literacy, region, wealth index, wealth quintile, fertility rate, Urban Rural and many other predictors that make this an ideal dataset for mapping poverty. For our study we only need the two target variables, regions and Urban Rural features

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	hv001	hv002	hv003	hv004	hv005	hv006	hv007	hv008	hv008a	hv009	hv010	hv011	hv012	hv013	hv014	hv015	hv016	hv017	hv018	hv019	hv020	hv021	hv022	hv023	hv024
2	1	4	2	1	671011	11	2017	1415	43044	5	2		5	5	0	complete	5	1	129	4	ever marr	1	barisal - ru	barisal - ru	barisal
3	1	8	3	1	671011	11	2017	1415	43046	6	1		6	4	0	complete	7	1	129	4	ever marr	1	barisal - ru	barisal - ru	barisal
4	1	13	4	1	671011	11	2017	1415	43044	5	1		5	2	0	complete	5	1	129	4	ever marr	1	barisal - ru	barisal - ru	barisal
5	1	17	4	1	671011	11	2017	1415	43045	6	1		6	6	1	complete	6	1	129	4	ever marr	1	barisal - ru	barisal - ru	barisal
6	1	21	1	1	671011	11	2017	1415	43044	4	2		3	4	0	complete	5	1	129	4	ever marr	1	barisal - ru	barisal - ru	barisal
7	1	30	2	1	671011	11	2017	1415	43044	5	1		4	5	0	complete	5	1	170	4	ever marr	1	barisal - ru	barisal - ru	barisal
8	1	35	2	1	671011	11	2017	1415	43044	5	1		5	5	0	complete	5	1	170	4	ever marr	1	barisal - ru	barisal - ru	barisal
9	1	39	2	1	671011	11	2017	1415	43044	4	0		4	4	0	complete	5	1	170	4	ever marr	1	barisal - ru	barisal - ru	barisal
10	1	43	2	1	671011	11	2017	1415	43045	4	1		4	4	0	complete	6	1	170	4	ever marr	1	barisal - ru	barisal - ru	barisal

Fig 2: Raw DHS Household Recode Dataset

Additionally with the socio-economic data we need the GPS coordinate data to pinpoint which cluster belongs to which location on the Bangladesh geographical map. DHS also surveys GPS

coordinate for each cluster point with a buffer of 5000 meter for rural regions and 2000 meter for urban regions, protecting privacy of those household by not giving exact coordinates but masking it in a radius of 5 kilometers for rural region clusters and 2 kilometers for urban region clusters. As shown in figure 3 The GPS dataset contains cluster id, latitude and longitude with which we can point geometry to extract geospatial features needed for our predictor variables

DHSCLUST	CCFIPS	ADM1FIPS	ADM1FIPSNA	ADM1SALBNA	ADM1SALBCO	ADM1DHS	...	DHSREGCO	DHSREGNA	SOURCE	URBAN_RURA	LATNUM	LONGNUM
1.0	BG	NULL	NULL	NULL	NULL	1.0	...	1.0	BARISAL	GPS	R	21.907454	90.106474
2.0	BG	NULL	NULL	NULL	NULL	1.0	...	1.0	BARISAL	GPS	R	22.171946	90.298709
3.0	BG	NULL	NULL	NULL	NULL	1.0	...	1.0	BARISAL	GPS	R	22.167150	90.187479
4.0	BG	NULL	NULL	NULL	NULL	1.0	...	1.0	BARISAL	GPS	R	22.195625	90.122065
5.0	BG	NULL	NULL	NULL	NULL	1.0	...	1.0	BARISAL	GPS	R	22.348901	90.179272
6.0	BG	NULL	NULL	NULL	NULL	1.0	...	1.0	BARISAL	GPS	R	22.151993	89.927550
7.0	BG	NULL	NULL	NULL	NULL	1.0	...	1.0	BARISAL	GPS	R	22.941232	90.183298
8.0	BG	NULL	NULL	NULL	NULL	1.0	...	1.0	BARISAL	GPS	R	22.869642	90.327852
9.0	BG	NULL	NULL	NULL	NULL	1.0	...	1.0	BARISAL	GPS	R	22.585755	90.459976
10.0	BG	NULL	NULL	NULL	NULL	1.0	...	1.0	BARISAL	GPS	R	22.481506	90.225066

Fig 3: GPS coordinate dataset using geopandas to visualize shape file

3.2 Data preprocessing

3.2.1 Preprocessing the DHS Household Recode Dataset

Originally the Household Recode Dataset contained 3307 features, not all of them are useful for socioeconomic analysis and spatial matching. Column filtering was applied to keep only 5 relevant features among 3307 and renamed them accordingly. hv001 Cluster ID , hv270 Wealth Quintile, hv271 Wealth Index (continuous target variable), hv024 region, hv025 URBAN_RURA. Two target variables were used for the model predictions. Wealth Index is the continuous value good for measuring poverty at regression metrics and wealth quintile being poorest to richest is good for visualization of poverty at classification level for policy makers and ease of understanding.

Cluster-level aggregation was done to convert household-level poverty data into cluster-level ground truth, the wealth index (hv271) was averaged for each cluster. For wealth quintile, the values were ranged from 1 to 5 replacing poorest to richest by averaging the mean wealth quintile. Later it was converted to back to textual value from the cluster's majority poverty influence.

3.2.2 Cleaning and Preparing the DHS GPS Dataset

Similarly to Household Recode preprocessing, for the GPS dataset column filtering was applied to keep only relevant features and renamed them accordingly. LATNUM latitude, LONGNUM longitude, URBAN_RURA urban/rural, DHSCLUST cluster ID, ADM1NAME region. In figure 4 we can see both Household Recode dataset and GPS dataset were merged by cluster id to keep the GPS coordinates and the wealth metrics needed for geospatial feature extraction from Google Earth Engine. Ensuring GPS coordinates CRS were kept in WGS 84 (EPSG:4326) provided compatibility for Google Earth Engine and other geospatial dataset.

1	cluster_id	wealth_in	mean_we	region	latitude	longitude	URBAN_R	buffer_m
2	1	-76583.3	1.655172	barisal	21.90745	90.10647	R	5000
3	2	-69557	1.733333	barisal	22.17195	90.29871	R	5000
4	3	-33966.8	2.551724	barisal	22.16715	90.18748	R	5000
5	4	-51788.6	2.4	barisal	22.19563	90.12207	R	5000
6	5	-46230.5	2.275862	barisal	22.3489	90.17927	R	5000
7	6	-88162.4	1.344828	barisal	22.15199	89.92755	R	5000
8	7	15935.45	3.344828	barisal	22.94123	90.1833	R	5000
9	8	-23780.1	3.071429	barisal	22.86964	90.32785	R	5000
10	9	-51606.7	2.4	barisal	22.58576	90.45998	R	5000
11	10	-44631	2.5	barisal	22.48151	90.22507	R	5000
12	11	-36759.9	2.633333	barisal	22.75154	90.09002	R	5000
13	12	-27561.9	2.9	barisal	23.03155	90.1585	R	5000
14	13	-27562	2.814815	barisal	22.95821	90.46295	R	5000
15	14	-43865.1	2.357143	barisal	22.72245	90.43908	R	5000

Fig 4: Merged Household Recode dataset + GPS coordinate dataset for Google Earth Engine geospatial feature extractions

3.2.3 Preprocessing Geospatial Covariates (Satellite-derived Features)

The preprocessed merged dataset combining from household recode and GPS coordinate was used to extract 6 geospatial covariates.

Table 2: Satellite-derived geospatial features

Data Name	Satellite / Sensor Source	Features Extracted	Frequency
Night-Time Lights (NTL)	VIIRS/DNB – Visible Infrared Imaging Radiometer Suite (NOAA/NASA)	Mean NTL	Monthly
NDVI (Normalized Difference Vegetation Index)	MODIS (MOD13Q1) or Landsat 8/9 Surface Reflectance	Mean NDVI	Monthly
Land Surface Temperature (LST	MODIS (MOD11A2 or MYD11A2)	LST_mean, LST_min, LST_max, LST_stdDev	Monthly
Rainfall (Precipitation)	CHIRPS (Climate Hazards Group InfraRed Precipitation with Stations)	Rain_mean, Rain_min, Rain_max, Rain_stdDev	Monthly
Land Cover Fractions	ESA WorldCover 2019/2020 (Sentinel-1 & Sentinel-2 based)	frac_bare, frac_built, frac_cropland, frac_shrub, frac_tree,	Monthly

		frac_water, frac_wetland	
POI Density	OpenStreetMap (OSM)	POI_count, POI_density	

This table represents the zonal level geospatial features that were extracted through Google Earth Engine as our predictor variables

NTL is the intensity of artificial light occurring at night and is a powerful proxy of human activity, infrastructure construction, and economy. Regions with increased NTL are usually those with greater access to electricity power, urbanization and increased income as compared to low-lit regions that are usually poor and rural. When it comes to the problem of poverty prediction NTL assists the model by differentiating the high threshold and low threshold clusters of people, however, in this case luminosity is interrelated with economic activity. The continuous values of NTL contribute to the model throughout the training period by giving the model clear variation that helps to enhance the regression model and minimize the error of prediction.

The NDVI is a measurement of vegetation health and density based on satellite effect and near infrared and red light. An increase in NDVI value is associated with a healthy vegetation, mostly associated with the rural and agricultural land, and the same decrease in the value signifies barren, urban or industrial land. Rural agricultural areas in developing countries are usually more prone to poverty so NDVI can be used to distinguish socioeconomic states between two dissimilar kinds of land. NDVI gives significant variance to the model learning, as it will assist the model to identify pattern variables connected to rural livelihoods, and environmental factors where which are associated with poverty.

LST reveals surface temperature on the earth and it is determined by urban heat island, vegetation cover and land use. High LST value tends to be in large urbanization or bare land

whereasoured ones tend to be in the vegetated or the water areas region. LST assists in estimation of poverty by dividing dense populated developed regions and low density in the rural region. When training a model with LST, an environmental dimension is incorporated that helps to increase the diversity of features and augment the strength of prediction particularly when NTL and NDVI are used together.

Surface features that land cover data recognizes include water bodies, farmlands, forests, grasslands and the urban built ups. Given that the socioeconomic situation is different in such types of land (e.g., urban = higher income, agricultural = lower income), land cover assists the model in comprehending the spatial patterns associated with the living standards. Categorical or fractional values of land cover add more context to the other features such as NDVI, and LST, the model is more likely to classify the environmental contexts in relation to the level of poverty.

Rainfall is the quantity of precipitation per time and it is significant to the agricultural output, water security and rural livelihoods. Poor rainfall patterns have usually been associated with food insecurity and poverty in the agricultural regions. The rainfall is also included, which is useful to the model to capture the climatic stress factors and make predictions stronger. Being a continuous environmental characteristic, the variability of rainfall enhances learning in the model, provides time-related significance, and helps in recognizing the poor regions vulnerable to climate change factors.

POI density indicates the accessibility and availability of basic services that include schools, hospitals, markets, and the public facilities. The high POI concentration is a typical sign of superior infrastructure and high economic activity, whereas low POI in availability is typical in underserved or poor areas. The POI density enhances the prediction of poverty with the immediate association of service accessibility to household well being. It also supports the scrutinization of the models by offering socioeconomic information of high significance that will support the satellite-based environmental indicators.After extraction, all datasets were merged using cluster id to create the final modeling dataset that were used to train and test models. Figure 5 demonstrates the final modeling data that we used to train our models

A	B	C	D	E	F	G	H	I	J	K	L	M	N
cluster_id	wealth_in	mean_we	latitude	longitude	buffer_m	ntl_mean	ndvi_mea	frac_bare	frac_built	frac_cropl	frac_grass	frac_shrul	frac_tree
1	-76583.3	1.65517	21.9075	90.1065	5000	0.28161	0.42737	0.00819	0.00185	0.46442	0.02012	0	0.18888
2	-69557	1.73333	22.1719	90.2987	5000	0.31078	0.52842	0.00108	0.0022	0.64	0.00426	1.38E-06	0.34707
3	-33966.8	2.55172	22.1672	90.1875	5000	0.33186	0.52032	0.00111	0.00157	0.481	0.00447	0	0.39043
4	-51788.6	2.4	22.1956	90.1221	5000	0.41887	0.58576	0.0007	0.01769	0.50924	0.00392	0	0.46472
5	-46230.5	2.27586	22.3489	90.1793	5000	0.32713	0.53366	0.00052	0.00463	0.47557	0.00174	0	0.42866
6	-88162.4	1.34483	22.152	89.9276	5000	0.26872	0.43759	0.00198	0.0017	0.46333	0.0043	1.38E-06	0.30699
7	15935.4	3.34483	22.9412	90.1833	5000	0.40953	0.61152	0.00033	0.00238	0.36565	0.00086	0	0.62796
8	-23780.1	3.07143	22.8696	90.3279	5000	0.32893	0.50825	0.00055	0.00064	0.46408	0.00744	0.00118	0.39049
9	-51606.7	2.4	22.5858	90.46	5000	0.32306	0.46592	0.00093	0.00139	0.42505	0.00326	0	0.42454
10	-44631	2.5	22.4815	90.2251	5000	0.30476	0.55029	0.0009	0.00182	0.44718	0.00188	6.90E-06	0.47419
11	-36759.9	2.63333	22.7515	90.09	5000	0.45304	0.58472	0.00142	0.00827	0.09606	0.00452	0.00036	0.78421
12	-27561.9	2.9	23.0316	90.1585	5000	0.34255	0.58124	0.0001	0.00126	0.60357	0.00359	0.002	0.37545
13	-27562	2.81481	22.9582	90.463	5000	0.33212	0.46816	0.00502	0.00104	0.41912	0.01058	3.05E-05	0.38178
14	-43865.1	2.35714	22.7224	90.4391	5000	0.45987	0.45083	0.0028	0.0039	0.32095	0.00781	0	0.45998
15	-91431.7	1.33333	22.7735	90.5101	5000	0.26984	0.31538	0.00521	0.00033	0.28644	0.01693	4.56E-05	0.25859

Fig 5: Final dataset with target variables and predictor variables

3.2 Model Description

To quantify the amount of wealth and poverty at the household level based on geospatial predictors, the several machine learning models were applied. Both models were chosen because they can represent various socio-economic outcomes by remote sensing variables. In this sub-section, each model is described in detail and how it is appropriate in poverty mapping through the use of geospatial data. Our modeling data were used in 9 models to establish the models that will give the most precise estimates of poverty.

3.3.1. Generalized Least Squares (GLS)

GLS is a linear regression approach that considers both heteroscedasticity and dependency among the error terms. It was included because it was decided to have a statistical base of the prediction of wealth indices. GLS offers understandable linear coefficients that draw attention to the direct impact of such variables as nighttime lights, NDVI, land surface temperature, and proximity-based characteristics. GLS gave the study the possibility to determine the linear relationships among geospatial indicators and outcomes of poverty as a benchmark to compare more complex non-linear models.

Model Equation

$$y=X\beta+\varepsilon \ldots\ldots (1)$$

GLS Estimator

$$\hat{\beta}GLS = (X^T \Omega^{-1})^{-1} X^T \Omega^{-1} y \ \ldots\ldots(2)$$

Where, Ω= covariance matrix of errors y = Xβ

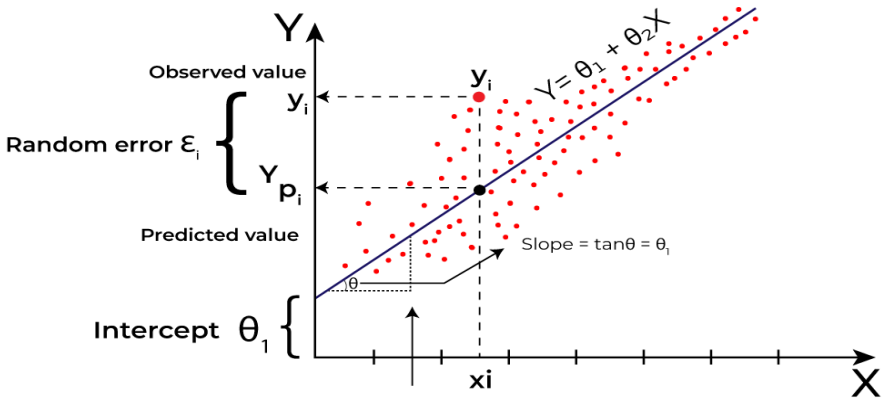

Fig 6: Linear Regression Diagram

3.3.2. Random Forest Regressor (RF)

Random Forest refers to an ensemble tree-based algorithm that has the ability to model complex non-linear relationships. It was chosen because it is resistant to noise, has the strong capacity of generalization and it has a high-dimensional geospatial data capacity. RF has been found to be effective especially in poverty mapping as it is capable of automatically discovering interactions among features, e.g. vegetation health, density of built-up, heterogeneity of land cover, and spatial accessibility. Its feature importance scores also give information regarding the strongest variables of the geospatial variable which significantly impact on wealth differences across DHS clusters. The Support Vector Regression (SVR) method is used to identify the presence or absence of an element within a dataset, as well as to predict and counteract the impact of elements on one another

Model Equation

$$\hat{\gamma} = \frac{1}{T} \sum_{t=1}^{T} h\, t(x) \ \ldots..(3)$$

Where, h(x) = prediction from the t^{th} tree T = number of trees

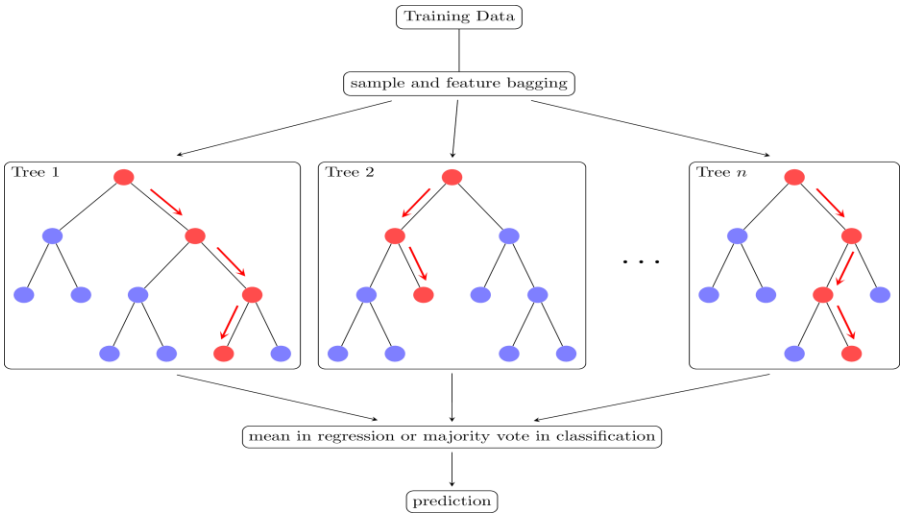

Fig 7: Random forest Diagram

3.3.3 Support Vector Regression (SVR)

The Support Vector Regression (SVR) is employed to detect an element or its absence in a dataset, and to forecast and neutralize the effect of each element on the others. The RBF kernel of SVR is very good at the task of taking non-linear marginal effects in medium sized data sets like DHS cluster level samples. It can generalize well when there are outliers and non-linear relationships, as its learning can be learned based on margins. The SVR was selected to establish the existence of smooth non-linear relationships between indicators of remote sensing and economic well-being, in particular, those that are not well established with tree-based models. Its capability to plot interactions among feature high dimensions render it useful towards detecting subtle geospatial-poverty processes.

Model Equation

$$f(x) = \sum_{i=1}^{n} (\alpha_i \alpha_i^*) K(x_i, x) + b \quad \text{.....(4)}$$

RBF Kernel

$$K(x_i, x) = exp(-\gamma \|x_i - x\|^2) \quad \text{.....(5)}$$

Where, $\propto_i, \propto_i^*$= dual coefficients K = kernel function b = bias

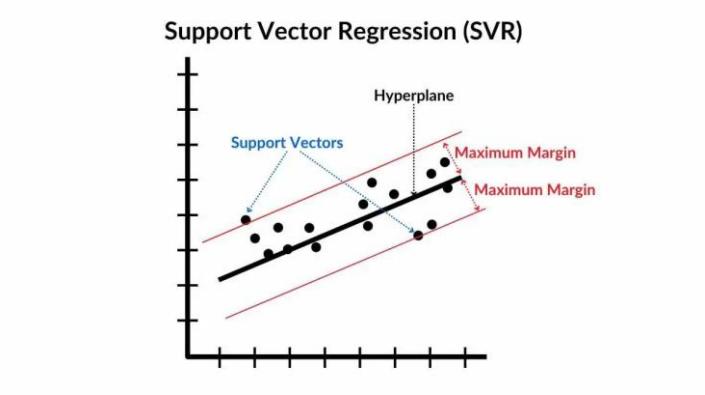

Fig 8: Support vector Regression Diagram

3.3.4. Multi Layer Perception (MLP) Neural Network.

The MLP is a feedforward neural net, which is able to learn the sophisticated functional associations using multi-layer hidden units. It was added to test whether deep non-linear representations enhance the accuracy of poverty estimation, at least when there are interactions between geospatial variables that are highly non-linear. Spatial complexity is common in features like night time luminosity, vegetation indexes, climatic variables and POI densities and might not be well represented by simpler models. The flexibility of the MLP renders it to be applicable in the extraction of underlying non-linear geospatial patterns in relation to the levels of wealth within DHS cluster level.

Model Equation:

$$a^{(l)} = \sigma\left(W^{(l)}a^{(l-1)} + b^{(l)}\right) \ldots\ldots\ldots (6)$$

Final Prediction

$$\hat{y} = a^{(L)}$$

Where, $W^{(l)}$ = weight matrix, $b^{(l)}$ = bias vector, σ = activation (ReLU), L = number of layers

ReLU Activision

$\sigma(z)=\max(0,z)$

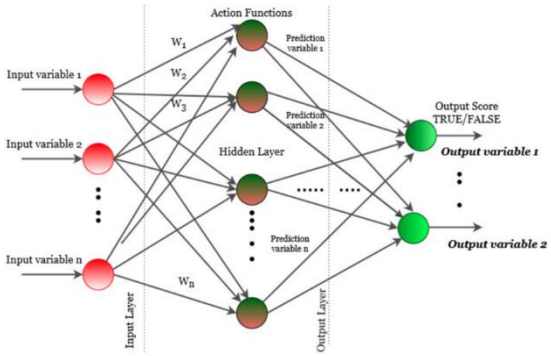

Fig 9: MLP architecture diagram

3.3.5. ElasticNet Regression

ElasticNet is a synthesis of the L1(Lasso)-L2(Ridge) regularization and identical to Lasso, ElasticNet is perfect regarding multicollinearity, as well as determining the most important geospatial predictors. To avoid overfitting it was selected to generate a sparse and interpretable model that can determine the major geospatial determinants of poverty. Since Earth Engine-based features tend to be correlated (e.g. NDVI and land cover, rainfall and vegetation), ElasticNet is a reasonable compromise between feature reduction and dimension stability.

Loss Function

$$\mathcal{L}(\beta) = \|y - x\beta\|_2^2 + \lambda[\alpha\|\beta\|_1 + (1-\alpha)\|\beta\|_2^2] \ldots\ldots\ldots(7)$$

Final prediction:

$$\hat{y} = X\hat{\beta}$$

3.3.6. Decision Tree Regressor

Decision Trees have an intuitive hierarchical format whereby the feature space is split into homogenous parts by threshold divisions. They were chosen to analyze the level of poverty partition by geospatial predictors on an individual basis. Despite its simplicity, Decision Trees provide interpretable decision rules describing the input of land cover types, temperature patterns or urban accessibility to wealth status. Their openness is useful in interpreting the structure of the information.

Prediction Equation:

$$\hat{y}(x) = \frac{1}{N_{leaf}} \sum_{i \in leaf_{(x)}} y_i \text{.....(8)}$$

Where the model takes the **mean value** of all training samples reaching the same leaf

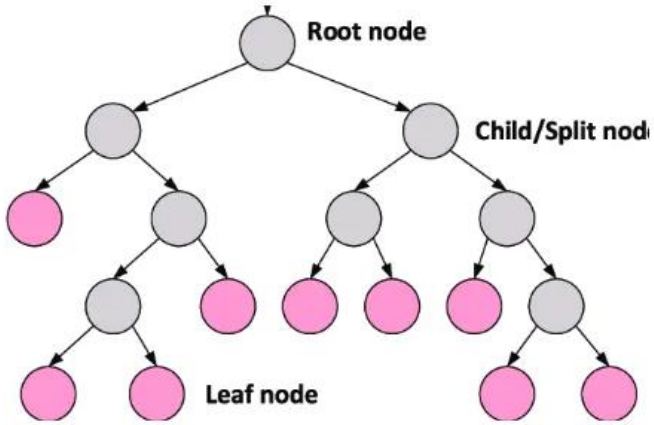

Fig 10: Decision tree diagram

3.3.7. Gradient Boosting Regressor (GBR)

GBR constructs trees in a sequential manner with a tree fixing the errors made in the previous tree. It was selected because it has a better capability of representing intricate non-linear interaction involving geospatial variables. In the case of poverty mapping, GBR is the best tool in the understanding of the subtle changes in the wealth index due to images of nighttime lights variations, vegetation gradients, and the extent of built-up situations. It is a good candidate of high-resolution estimation of poverty since it focuses on minimizing bias.

Stage-wise Update

$$F_m(x) = F_{m-1}(x) + v.h_m(x) \text{.....(9)}$$

Where, F_m = prediction after m-th tree ,ν = learning rate, $h_m(x)$h = new tree trained on residuals

Residual Computation

$$r_{im} = y_i - F_{m-1}(x_i) \text{.....(10)}$$

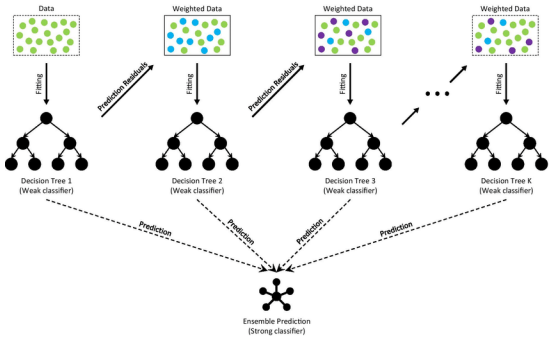

Fig 11: Gradient Boosting Regressor diagram

3.3.8. XGBoost Regressor

XGBoost is a high-performance and efficient gesture of gradient boosting that is associated with the high predictive capability. It was chosen due to its success in the socio-economic prediction activity and its capacity to deal with heterogeneity of geospatial characteristics. XGBoost may capture complex interactions of remotely sensed indicators, which makes the model give stabilized predictions even in cases when environmental variable-poverty indicator interaction is highly non-linear. The regularization it contains also minimizes the chances of overfitting.

Objective Function

$$\mathcal{L} = \sum_i l(y_i, \hat{y}) + \sum_k \Omega(f_k) \ldots\ldots\ldots\ldots (11)$$

Regularization

$$\Omega(f_k) = yT + \frac{1}{2}\lambda \sum_j w_j^2 \ldots\ldots\ldots\ldots\ldots (12)$$

Tree Output

$$\hat{y} = \sum_{k=}^{k} 1^{f_k(x)} \ldots\ldots\ldots\ldots\ldots\ldots (13)$$

3.3.9. K-nearest Neighbors Regression

KNN is a non-parametric distance based model which makes predictions based upon the value of the nearest neighbors in the features domain. It was added to experiment on a simple, instance based form of learning in which the poverty level of clusters is supposed to correlate with alike

geospatial features. KNN assists in assessing the possibility of similar clusters (based on environmental similarities (e.g., vegetation, climatic conditions, accessibility to urban amenities), having similar wealth distributions. This model generates a background on how spatial similarity can be observed in the dataset.

Prediction Equation

$\hat{y}(x) = \frac{1}{K}\sum_{i \in N_K(x)} y_i$.....(14)

Where, $N_K(x)$ = the set of K nearest neighbors based on Euclidean distance:

$$d(x, x_i) = \sqrt{\sum_{j=1}^{p} (x_i - x_{ij})^2}$$

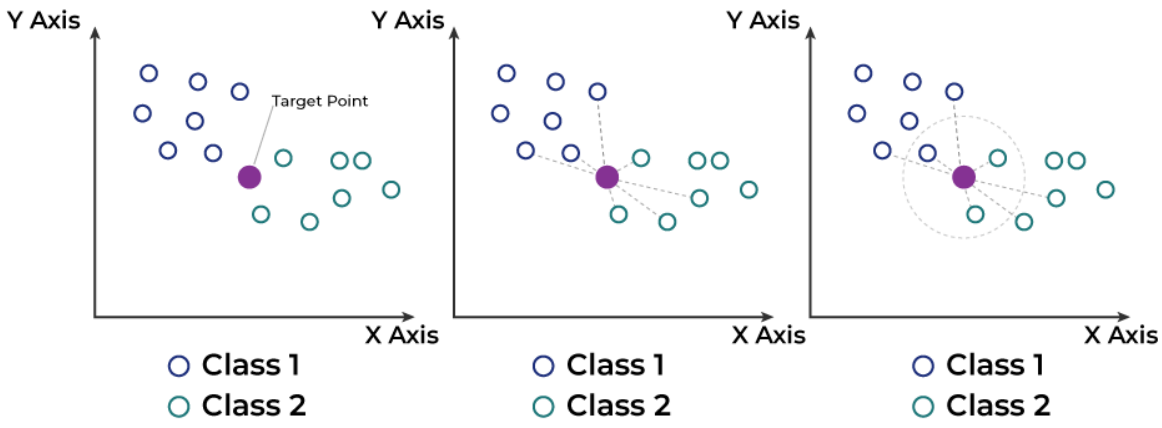

Fig 12: KNN diagram

3.3 Experimental Setup

3.3.1 Hardware configuration

The experiments carried out in this paper were all carried out in a personal computer with standard mid-range specifications to make sure that the modelling structure could be replicated without the need to use some special equipment. The evaluation and training machine was based on an Intel based 6 core processor with a speed of about 3.0 GHz

CPU: Intel(r) Coretm i5-10400 CPU @ 2.90 GHz

RAM: 16 GB DDR4

GPU: No dedicated GPU was used

There was no special graphics acceleration, and all calculations were made with the help of the central processing unit of the system. This hardware evaluated example shows that the proposed method of poverty mapping is able to be conducted on widely available computing resources and this aspect indicates that the methodology is available to researchers with comparable resources or with relatively limited resources.

3.3.2Software Environment

The used software environment was a Python based ecosystem installed on Windows 11 (64-bit) operating system. Modelling, analysis, and visualisations were done by using Python 3.10 with Jupyter Notebook.

Operating System Windows 11 (64-bit)

Programming Language Python 3.10

Development System: Jupyter Notebook.

Package Manager: pip 3.4.3

Machine Learning Libraries There are a number of machine learning and data processing libraries that were necessary during the research.

NumPy as a numerical computation library.

Pandas to manipulate data in spreadsheets and tabular operations.

Scikit-learn (sklearn) of GLS, ElasticNet, of Random Forest, SVR, Decision Tree, Gradient Boosting, KNN, and MLP model. Gradient-boosted regression with XGBoost.

Visualization using matplotlib and seaborn. statisticsite: Bayesian statistics using Python.

Joblib to save and load model.

The application of Google Earth engine API to extract geospatial features (NDVI, NTL, LST, etc.).

The reproducibility was guaranteed by maintaining high environmental consistency and controlled computing environments. Every single script was executed in the identical Python environment during the entire study, and the version of important libraries did not change during the experimentation. Each model employed deterministic operations, which could proceed using the same seeds, and the preprocessing pipeline was run across models the same way so it would not introduce differences due to difference in data handling. A combination of these steps guaranteed the reliable repetition of the experiments that were provided in this study by other scientists with the same amount of hardware and software resources.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Result Analysis

The chapter reports experiment findings of applying different machine learning models in order to predict accurately the poverty of 672 clusters of each and every cluster according to the wealth index and the wealth quintile (poorest to richest). In order to make a complete assessment, our research question was simulated by 9 machine learning models. The least RMSE as well as the highest R2 value were used to select the best of these 9 models. Random rainforest model performed better as compared to other 8 models. With RMSE of 365.220439 and R2 of 0.759106. The ensemble model (random forest) had a 86% accuracy on the prediction of the result. It actually outscored the correct forecasts of the wealth quintile of the cluster of the poorest, poorer, middle, rich, and richest by 86 percent. The finding was satisfactory as per other studies findings concerning the mapping of poverty.

4.1.1 GLS(Generalized Least Square)

The works of Generalized Least Squares (GLS) performed in the form of Linear Regression was the original baseline model in the present research. GLM showed surprisingly good results in this study, and the R2 values that were generated by it were very high in comparison to some other complex models. This was an indication that, a significant part of the variation in poverty in Bangladesh in terms of DHS clusters is well described in terms of linear relationships. Though simple, GLS gave consistent forecasts and was a strong benchmark that has indicated what non-linear models actually offered an advantage over a linear organization. The cause of GLS doing good is due to the influence of the light of night, the land surface temperature, or the urbanized areas-often has the tendencies of linearity.

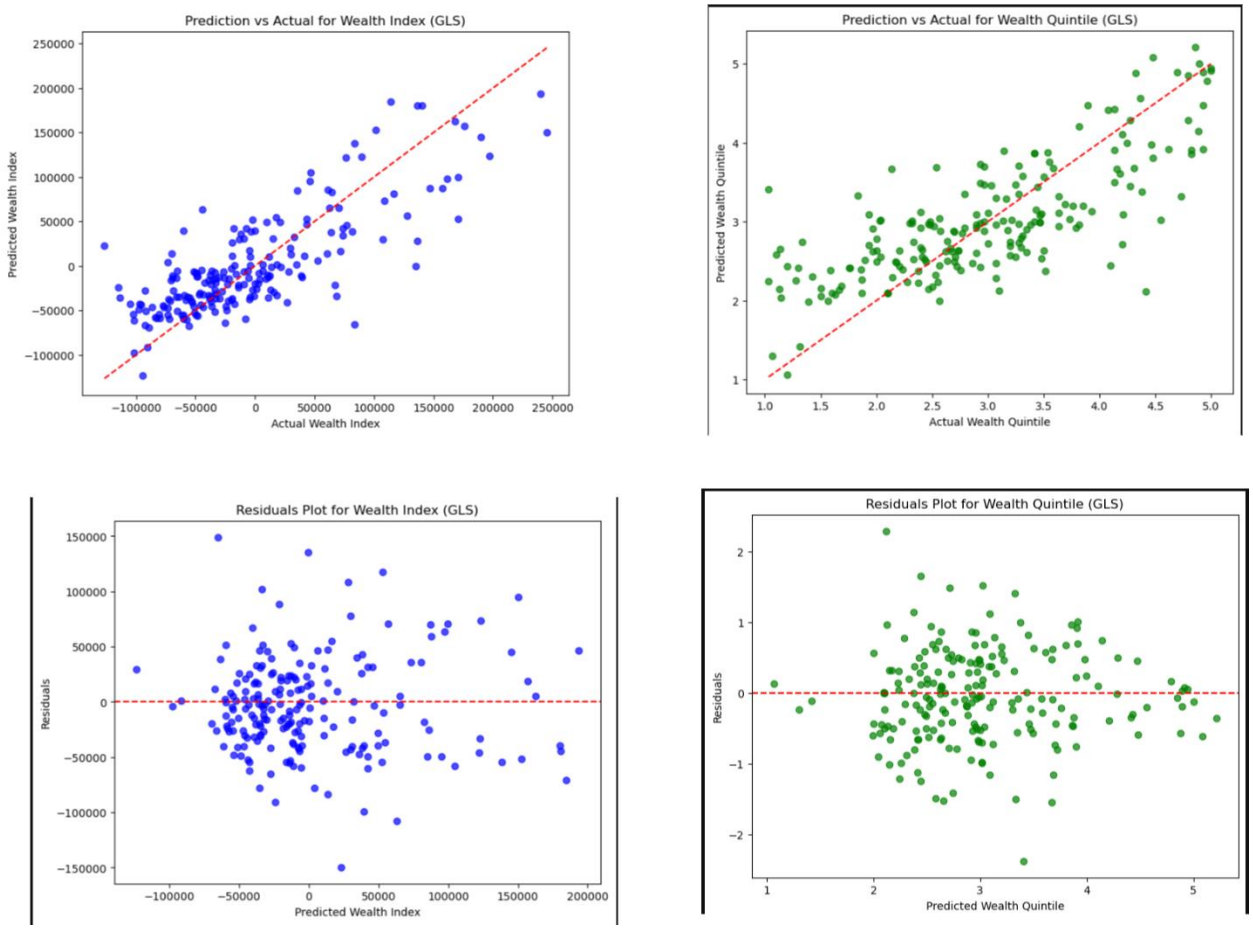

Fig 13: Model evaluation for GLS model

4.1.2 Random Forest

Random Forest Regression was selected because it has been proven to be useful in predicting multi-layered and non-linear relationships and has strong resistance to overfitting. RF is able to work with data that is of high dimension and it can resist noise, and is thus suitable in working with multi-source features as in this study. The model was effective in both target variables and when compared with any other algorithms it did very well making it able to represent real heterogeneity in the DHS clusters. The performance and the interpretability of RF are highly sensitive to well-tuned hyperparameters (depth, amount of trees and bagging of features, etc). Various features which had the greatest impact on our prediction could be recognized and their importance measured by the use of random forest.

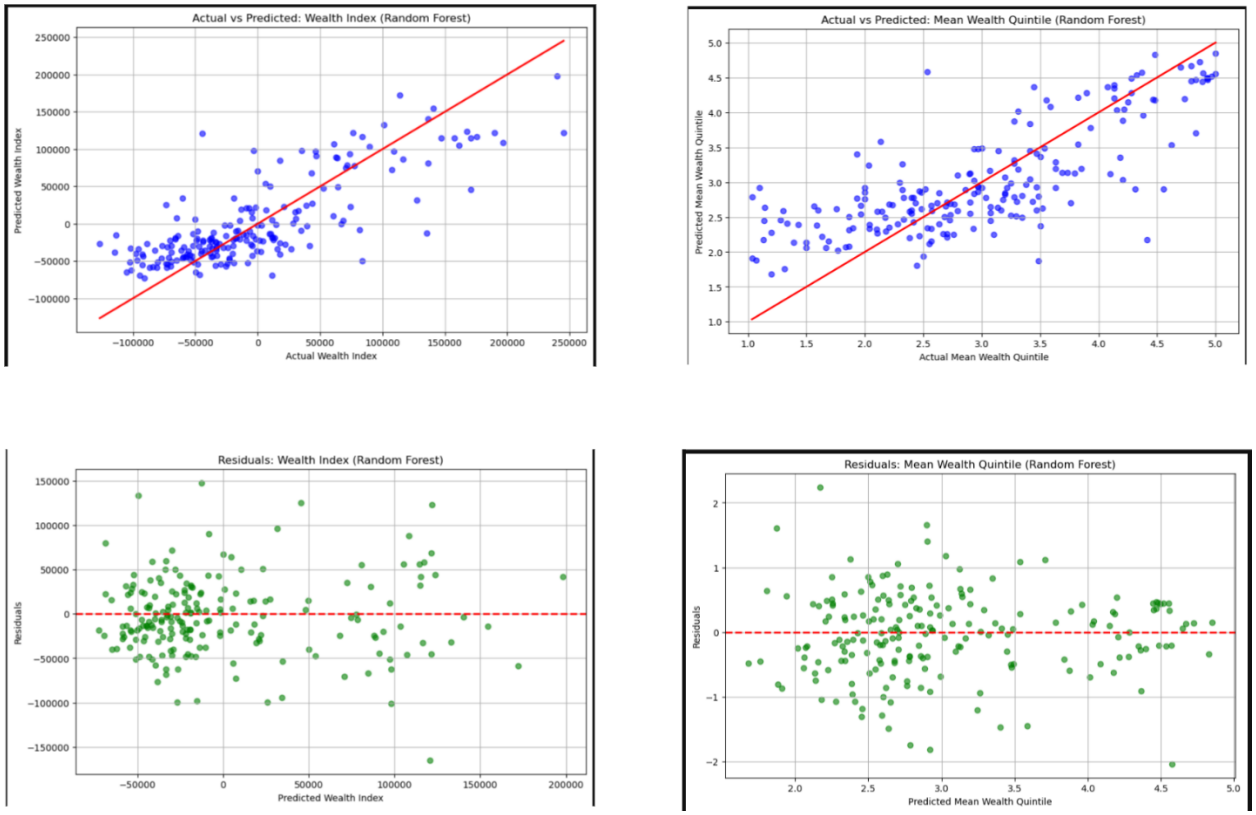

Fig 14: Model Evaluation for Random Forest

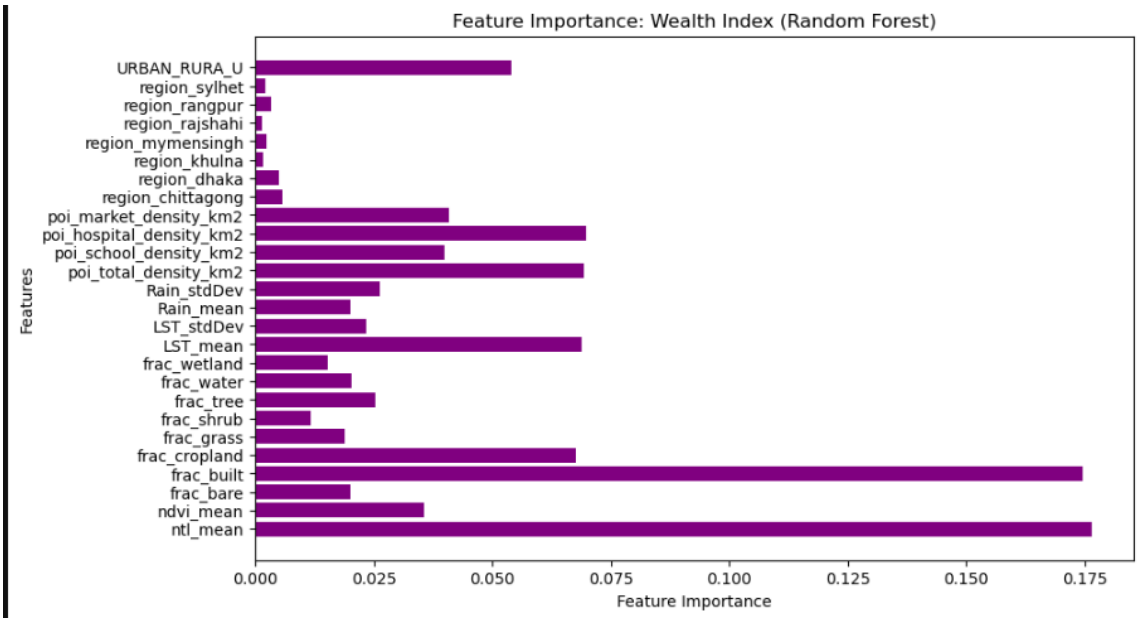

Fig 15: Result of variable importance analysis on poverty mapping

4.1.3 Support Vector Regression (SVR)

The appearance of SVR was predetermined to determine the efficiency of the margin-based optimization methods in the process of poverty prediction. In this experiment, SVR gave comparatively the weakest performance with the highest RMSE and negative R2. The poor levels of R2 and the large error statistics imply that SVR was not able to generalize to the different clusters of DHS. This result indicates observed difficulties of SVR with relatively large tabular data of mixed-feature characteristics. Moreover, the amount of noise entertained on cluster-level indicators of poverty would have posed a challenge to the model in making a stable margin of error. But even SVR was still useful as far as the Residual Error Distribution.

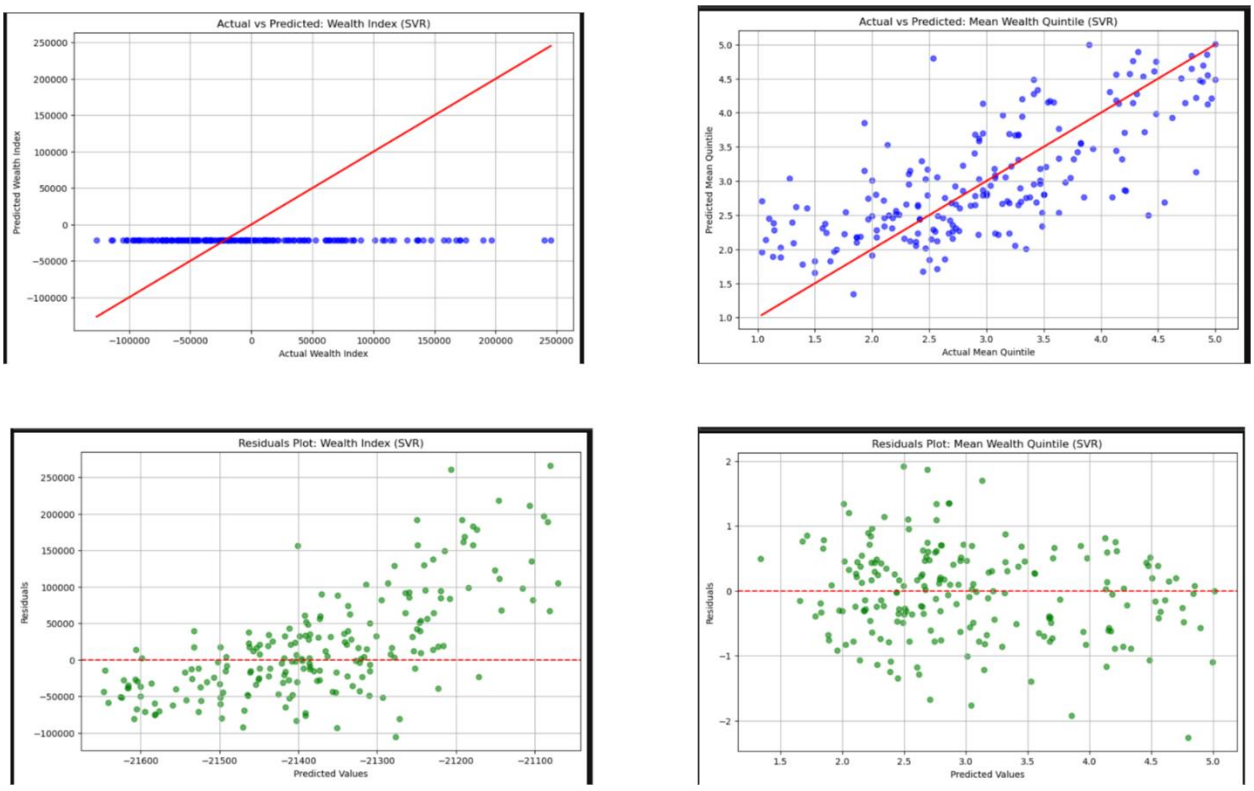

Fig 16: Model Evaluation for SVR

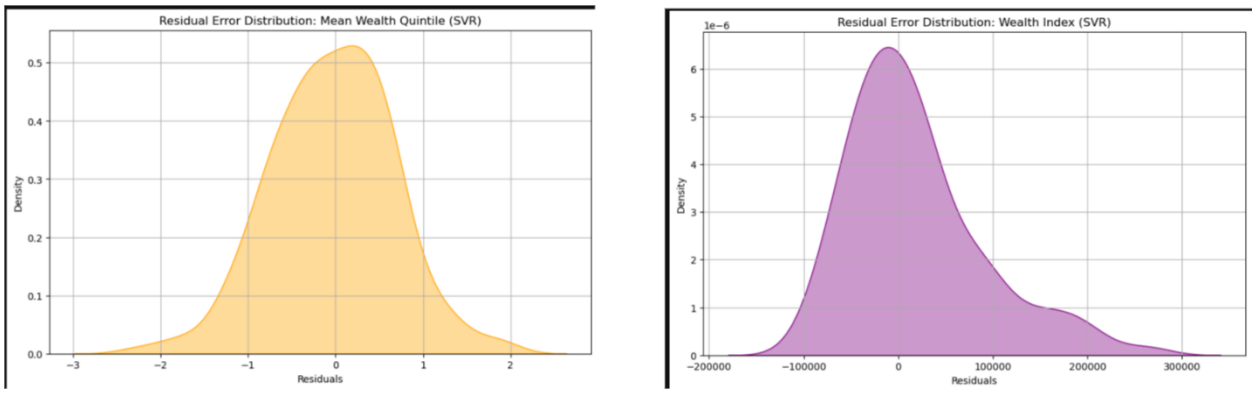

Fig 17: Result of Residuakl Error distribution using SVR

4.1.4 Multilayer Perceptron (MLP)

The feed-forward neural network architecture Multilayer Perceptron (MLP) was chosen to estimate whether or not artificial neural networks would discover more profound non-linearities in the geospatial and environmental predictors. A moderate architecture (64 and 32 neurons) that was used in the current study is a shallow architecture as compared to deep learning architecture, but it fits well with structured tabular data. Nevertheless, the performance of MLP was fluctuating and below tree based in a number of circumstances. This is due to a number of known weaknesses: neural networks usually need much larger datasets to avoid underfitting, the model is sensitive to size and hyperparameters, and neural networks do not deal well with tabular data that has limited number of rows and a large number of engineered features. Moreover, data on poverty have irregular distributions of spaces intrinsically, and there is noise and irregularity between clusters which brings up issues on the optimization of MLP.

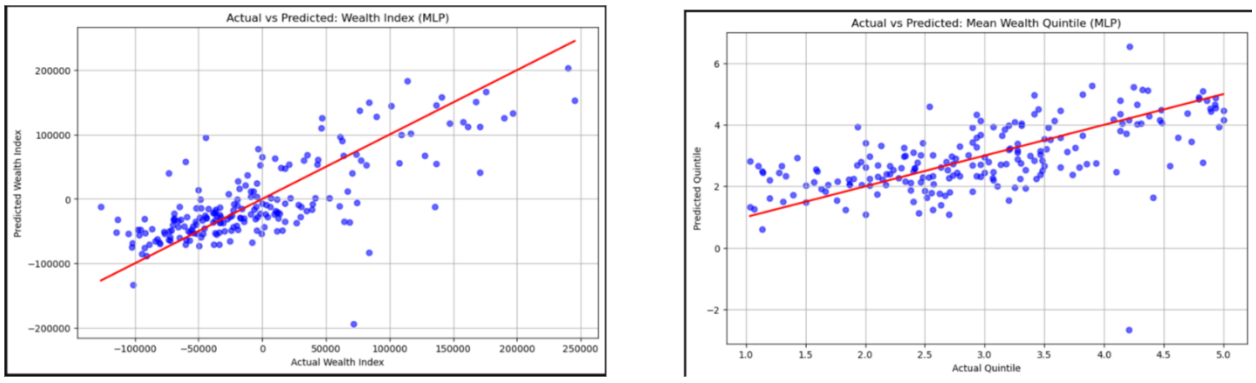

Fig 18: Model Evaluation of Predicted vs Actual scatter plot for MLP

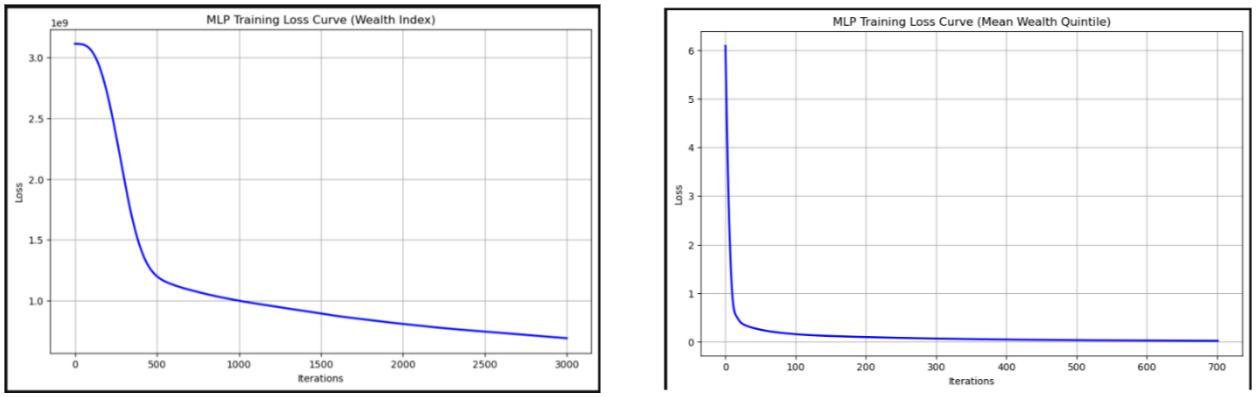

Fig 19: Result of training loss curve using MLP

4.1.5 ElasticNet Regression

ElasticNet Regression was used to discuss the effectiveness of the regularized linear models in predicting poverty. ElasticNet is a hybrid of L1 (Lasso) and L2 (Ridge) penalties and is used to train the model to resolve multicollinearity, which is widely observed in geospatial data sets where the various features of the environment e.g. NDVI, rainfall and land cover fractions are found to co-relate with one another. Despite the fact that ElasticNet worked better than some non-linear model, its performance was hampered by the fact that it was linear. The model was moderately accurate, which implies that although regularization contributed to stabilizing the estimation of coefficients, the reduction in poverty in Bangladesh cannot be explained by penalized linear models only..

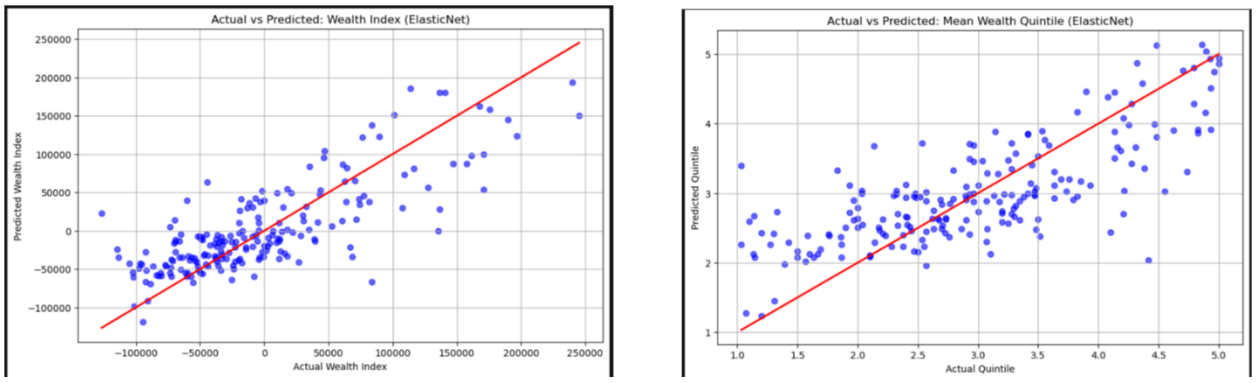

Fig 20: Model Evaluation of Predicted vs Actual scatter plot for ElasticNet

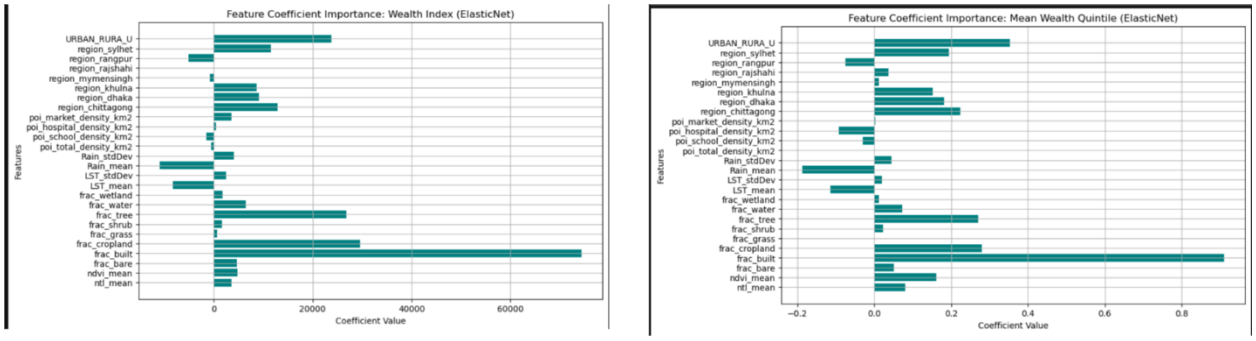

Fig 21: Result of feature coefficient importance using ElasticNet

4.1.6 Decision Tree Regressor

A single Decision Tree Regressor was also added to study a simple, non-linear learner that subdivides the feature space into hierarchical decision-rules. Single decision trees are however prone to over-fitting and are normally less predictively stable than their ensemble-based counterparts. The study itself had this limitation since the Decision Tree Regressor achieved some of the lowest R2 scores of the other models. The model was not very successful with generalizing to data that could not be seen and also had large error variation.

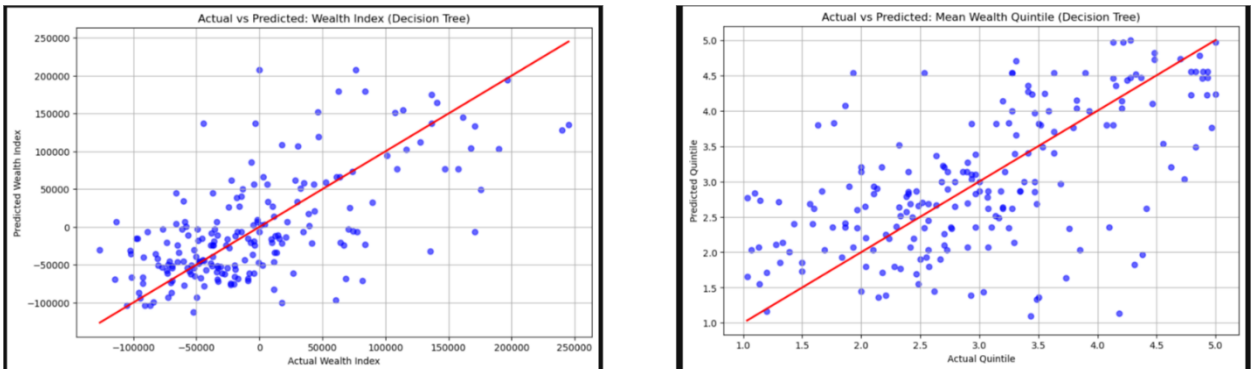

Fig 22: Model Evaluation of Predicted vs Actual scatter plot for Decision Tree Regressor

4.1.7 Gradient Boosting Regressor (GBR)

Gradient Boosting was chosen because it has a good reputation in the prediction operations involving structured data. Such a model constructs trees in series, each of which corrects the errors of the parents, and the ensemble is able to learn in detail the complicated relationships and subtle interactions of the data. Gradient Boosting was found to be very reliable and worked well

in this study mostly outperforming more basic linear and distance based model models such as KNN. This character along with its capacity to bundle shallow trees into an effective ensemble was especially useful in modeling poverty variations that entail non-line patterns of environmental, socio-economic, and point-of-interest characteristics.

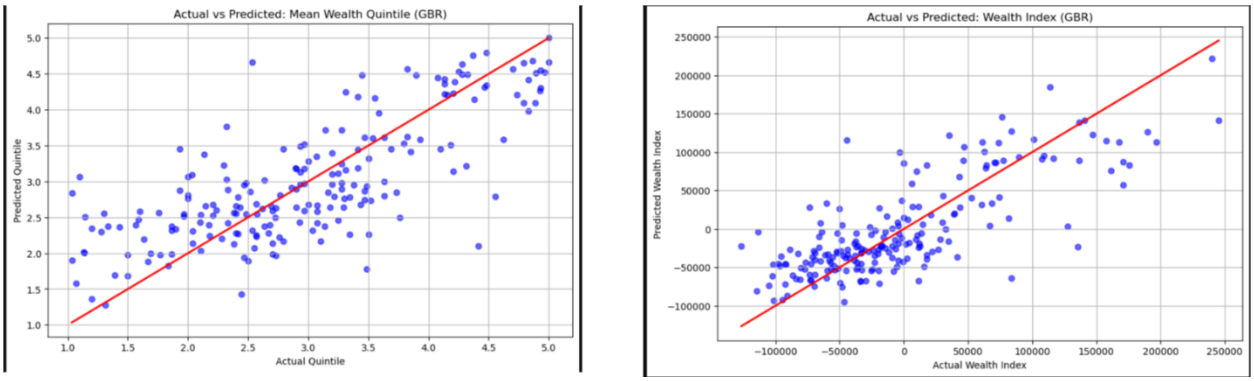

Fig 23: Model Evaluation of Predicted vs Actual scatter plot for Gradient Boosting Regressor

4.1.8 XGBoost Regressor

XGBoost which is a recent version of gradient boosting that has been optimized to be fast and efficient was chosen due to its good performance in machine learning contests and the capacity to fit complex non-linear interactions in structured data. XGBoost performed averagely in the selected work: in some instances where it performed on par with Gradient Boosting it failed to outperform the default GLS model. This indicates that XGBoost might have been sensitive to the considerably small amount of data used and parameter hypertrying. Cluster-based poverty mapping datasets which take only cluster-level characteristics do not always have the volume required to enable XGBoost to take advantage of all its boosting functions

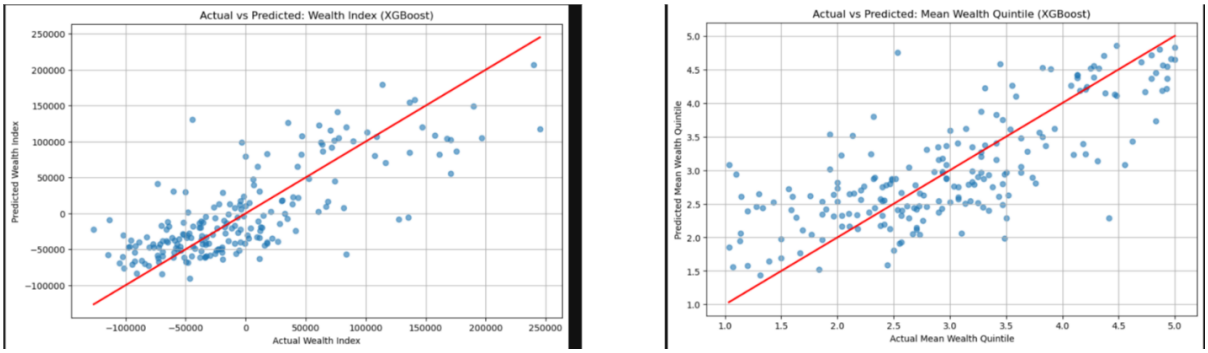

Fig 24: Model Evaluation of Predicted vs Actual scatter plot for XGboosting Regressor

4.1.9 K-Nearest Neighbors (KNN)

KNN has been added as a distance based baseline to determine the performance of the instance based learning with regard to predicting cluster levels of poverty. KNN had a poor ability to predict in this research. Valuable features represent a high-dimensional data space coupled with noisy geospatial measures, making KNN prone to the curse of dimensionality, in which distance metrics are no longer discriminatory. Moreover, the difference in poverty between clusters is not always smoothly distributed in feature space, and this decreases the ability of KNN to extrapolate.

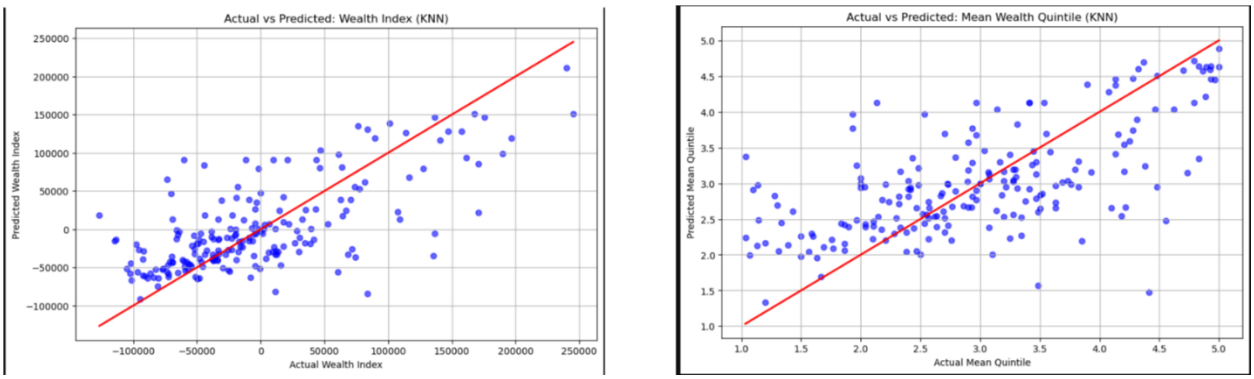

Fig 25: Model Evaluation of Predicted vs Actual scatter plot for KNN

4.2 Model comparison:

Comparison between all 9 of these models showed that Random Forest was able to capture mixed-feature types well and generalize it to get significantly more accurate results. Linear methods such as GlS and ElasticNet didn't have the best error evaluation but relatively high r-squared score meaning poverty is positively linear to some features such as NTL, Built up area. Boosting methods performed generally well but lack of significant amount of data resulted to have marginal errors which means the dataset couldn't leverage from it's boosting capabilities. Table 3 contains all models evaluation metrics for ease of comparison in performance

Table 3: Model Evaluation Metrics

	RMSE	MAE	R^2
GLS	435.001785	336.321763	0.634608
Random Forest	365.220439	277.589122	0.759106
SVR	1106.069473	802.878750	-0.060453
MLP	520.428176	378.52220	0.557336
ElasticNet	473.179010	365.672928	0.634065
Decision Tree Regressor	776.063627	574.161416	0.328264
Gradient Boosting regressor	487.042010	364.564932	0.603835
XGBoosting Regressor	542.211147	418.553943	0.589842
KNN	737.704542	542.126378	0.485082

This table shows the evaluation metrics and compares each models result with other ones

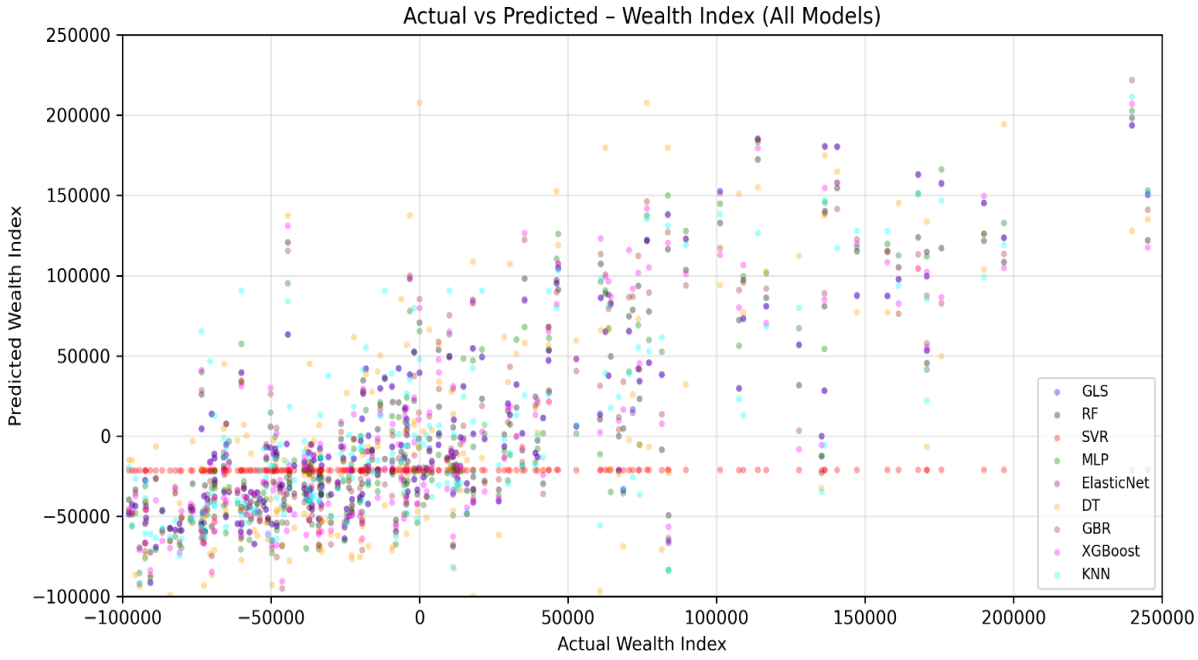

Fig 26: Scatter Plot comparing actual vs predicted for wealth index by all models

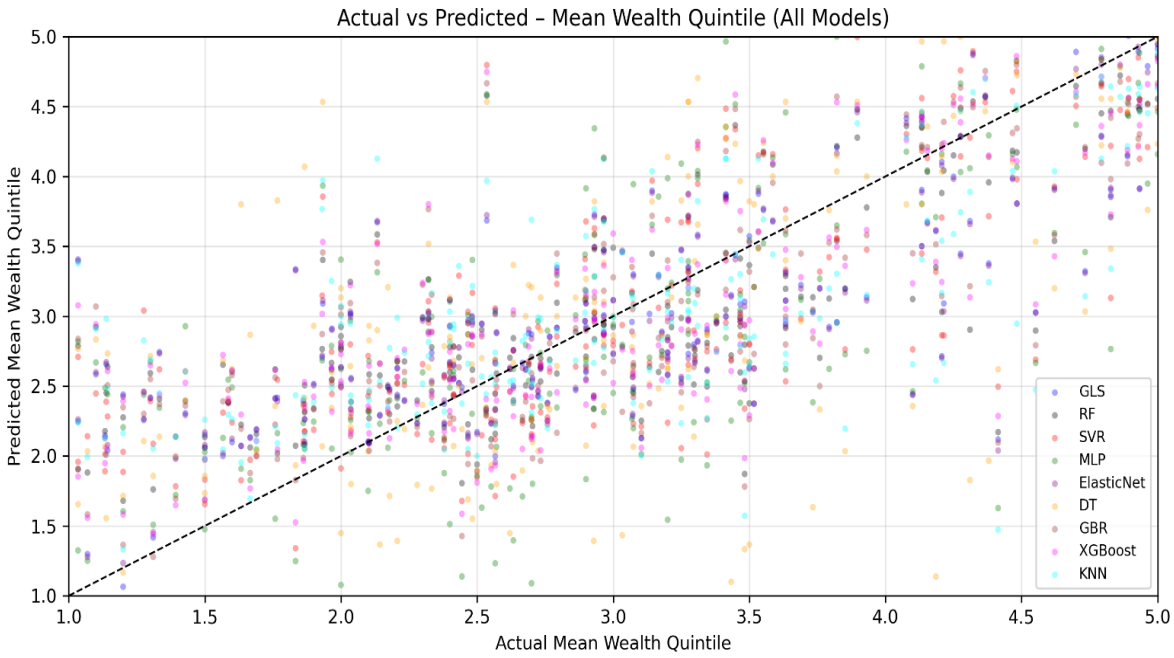

Fig 27: Scatter Plot comparing actual vs predicted for wealth quintile by all models

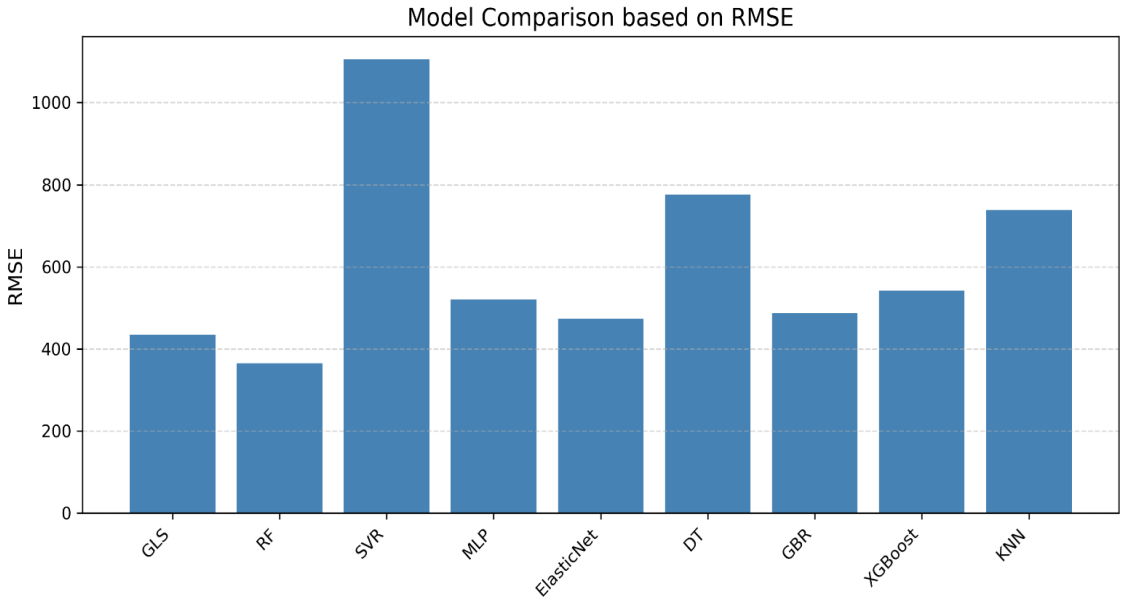

Fig 28: Comparison of average RMSE obtained from nine machine learning models

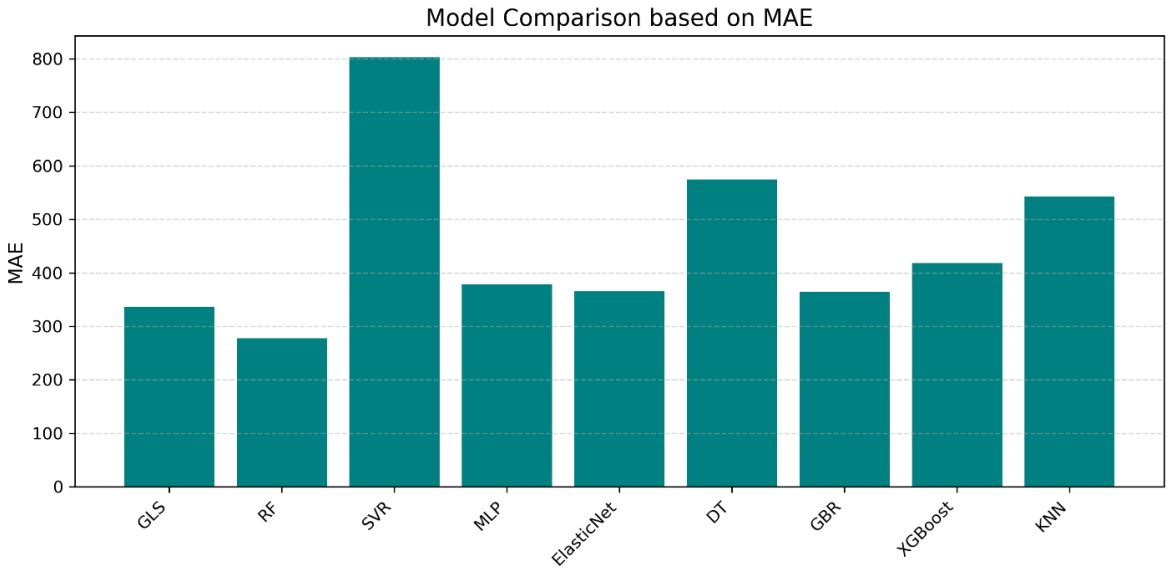

Fig 29: Comparison of average MAE obtained from nine machine learning models

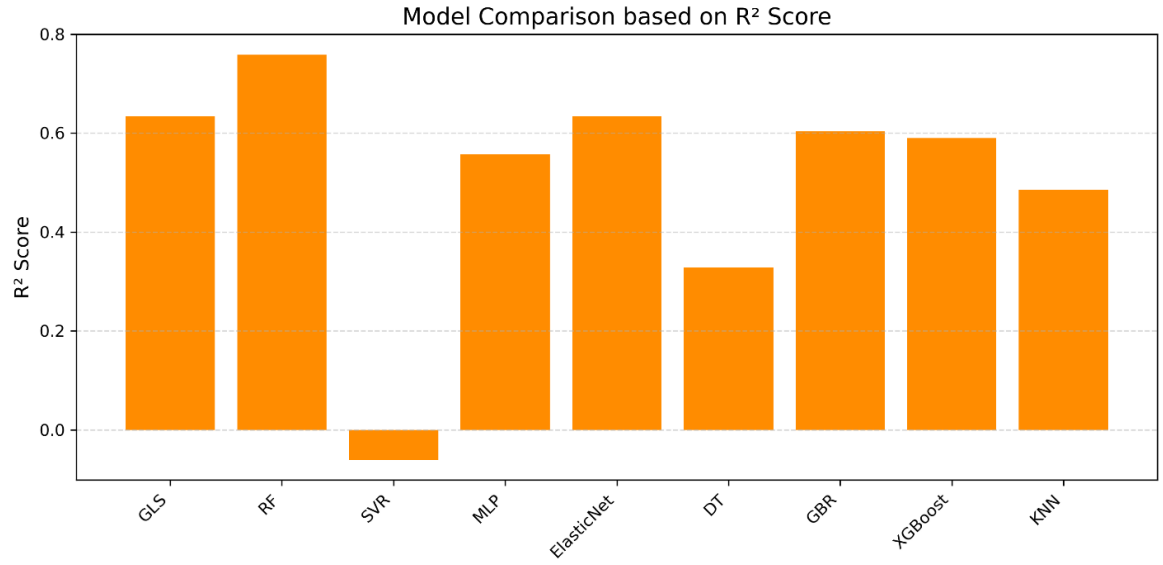

Fig 30: Comparison of average R^2 obtained from nine machine learning models

In summary, Random Forest was selected as the final model for poverty mapping because it consistently delivered the most accurate and stable predictions among all tested algorithms. Unlike linear or distance-based models, Random Forest effectively captured the complex, non-linear relationships present in the geospatial predictors and handled noise, multicollinearity, and feature interactions with minimal tuning.

As shown in table x, the R^2 score value of random forest being the highest makes it suitable for the predictive result of mapping poverty. We used the mean wealth quintile value as our classification target for the predictive model, converting back to 1-5 by the most influential wealth index of that cluster. One being poorest and five being richest. The classification metrics used for evaluations were Accuracy, Precision, Recall, F1-score and confusion matrix. Random forest was able to reach an accuracy of 86%, Precision of 88%, recall of 77% and f1-score of 84%

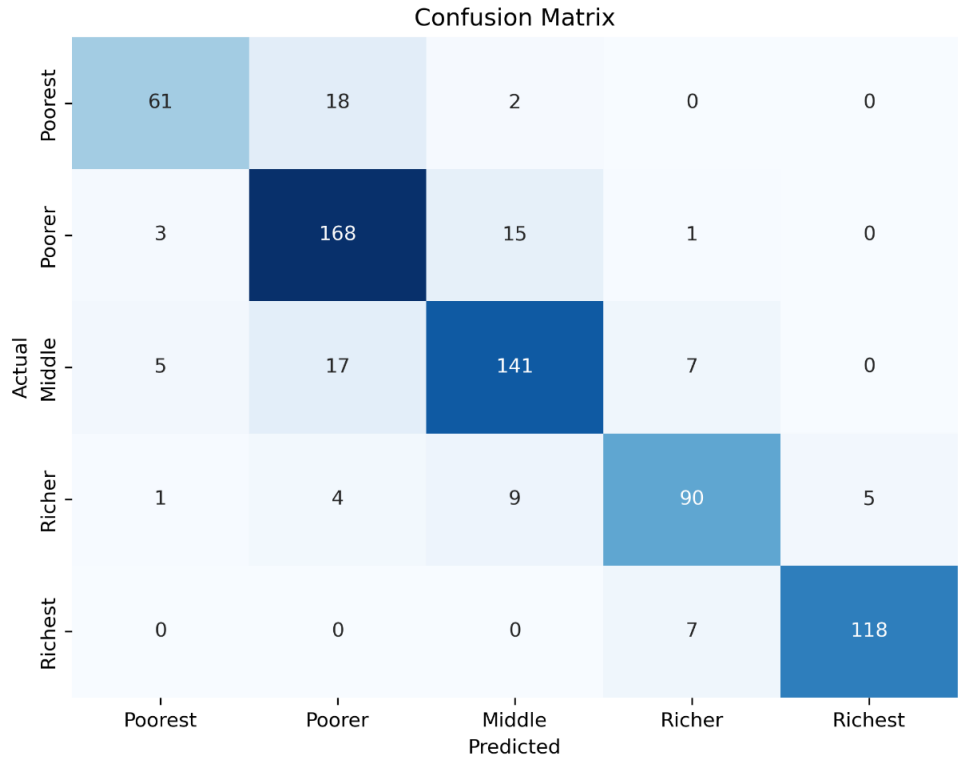

Fig 31: Confusion matrix using Random Forest

We used the actual wealth quintile to generate the cluster and visualize the original poverty in figure x and used the predicted wealth quintile from the random forest model to generate the cluster and visualize the predicted poverty.

4.3 Mapping

Initially data collection is done from multiple satellites using the GPS coordinated clusters for Bangladesh. Google Earth Engine was used to calculate features and generate maps for each extracted data. figure 32, 33, 34, 35, 36 x are the visual representation of the calculated NTL, NDVI, LST, Land Cover and Rainfalls

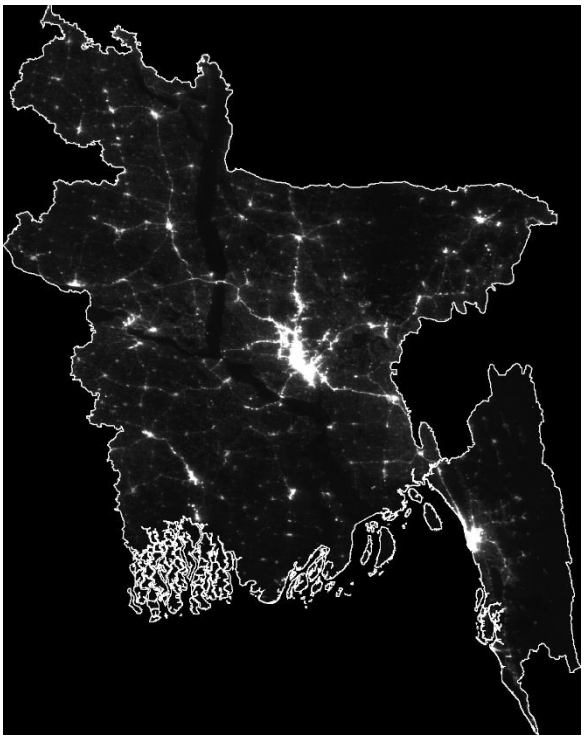

Fig 32: NTL intensity for Bangladesh in 2022. Image source was generated by Google Earth Engine by the author

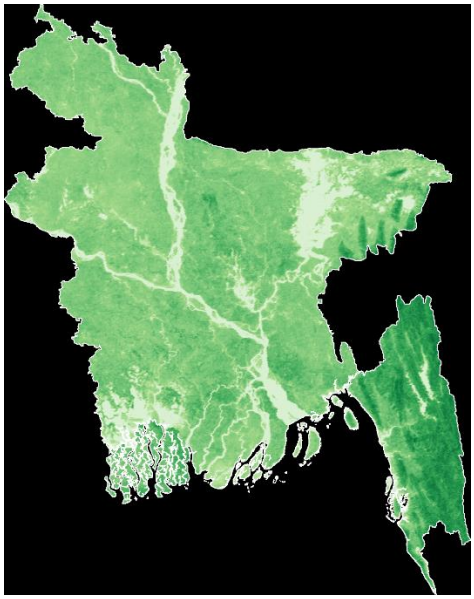

Fig 33: NDVI spatial distribution for Bangladesh in 2022. The image source was generated by Google Earth Engine by the author

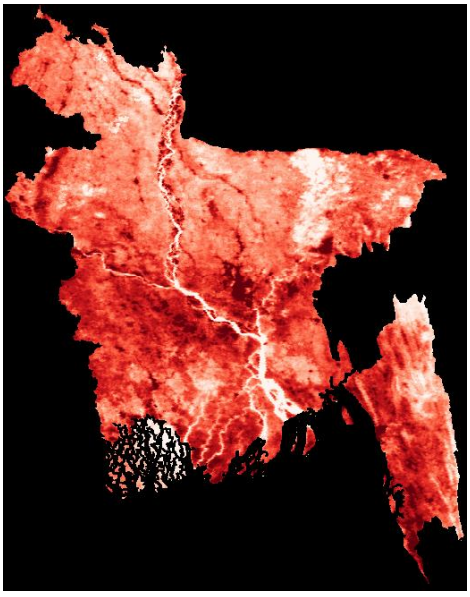

Fig 34: Spatial distribution of Land Surface Temperature for Bangladesh in 2022. Image source was generated by Google Earth Engine by the author

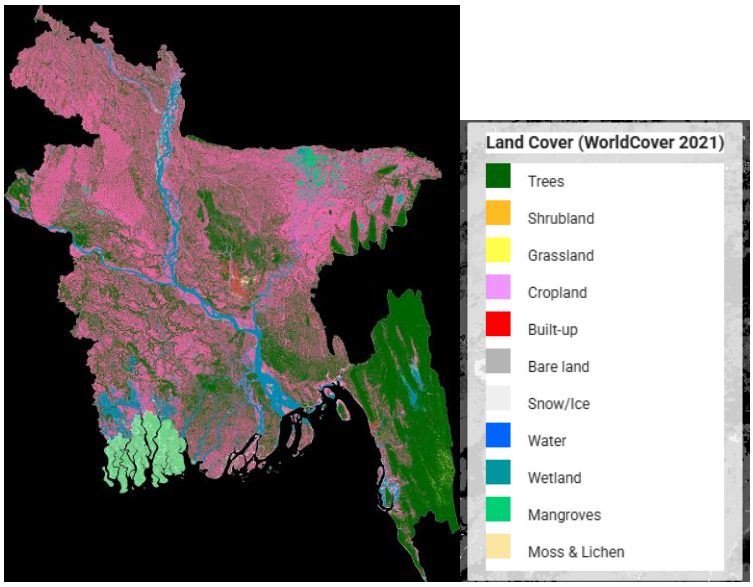

Fig 35: Spatial Distribution of Land cover for Bangladesh in 2021. Image source was generated by Google Earth Engine by the author

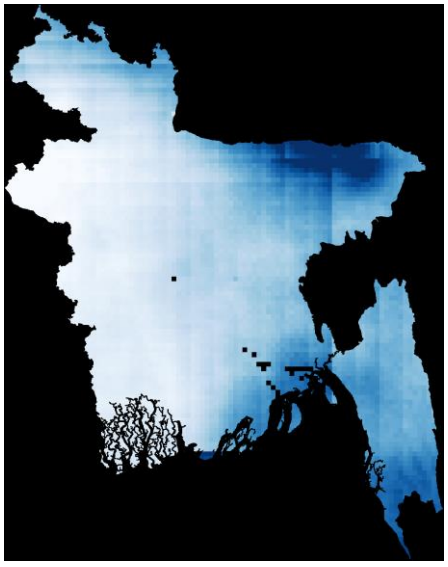

Fig 36: Spatial Distribution of cumulative rainfall in 2022. Image source was generated by Google Earth Engine by the author

The actual and predicted poverty levels were both mapped at the DHS cluster level as shown in figure 37,so that the spaces where poverty had been distributed across Bangladesh could be viewed visually to determine how effective the model-predictive was. The maps allow comparing the observed conditions in wealth with the forecasted conditions using the Random Forest model and allow providing an intuitive picture of the way the poverty is distributed across the territory. The visualization offers important details concerning the merits and shortcomings of the model and potential areas of spatial error since the visualization shows those spaces where the model is most similar to the real-life trends and those where the variances exist. Spatial performance of the model can be more cognitive and the entire evaluation of the model in terms of being suitable in poverty mapping can be facilitated when the actual and predicted poverty maps are presented side by side.

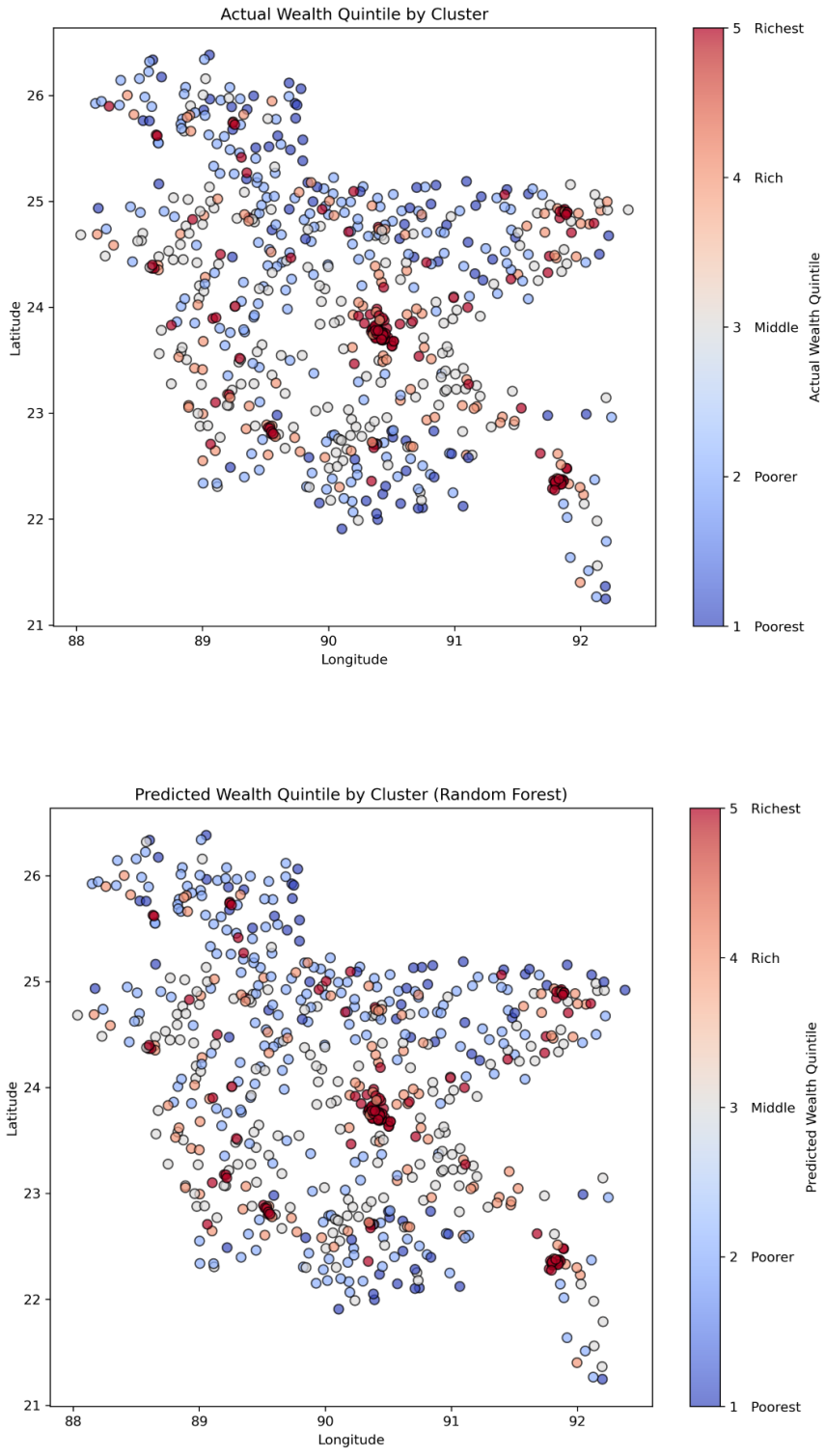

Fig 37: Actual vs Predicted Poverty mapping of Bangladesh at cluster level

CHAPTER 5

CONCLUSION

5.1 Findings and contributions

This paper aimed to examine the question of whether freely available geospatial data used with machine learning algorithm can be utilized as a credible and low-cost alternative to conventional household surveys to estimate poverty in Bangladesh. The research combines the latest satellite-derived features, including lights operating at night, vegetation indices, land surface temperature, rainfall patterns, land cover features, and point-of-interest information to display promising evidence under the support of this approach.

The initial research question addressed the question on whether geospatial characteristics could substitute, or supplement, traditional ways of counting poverty. The findings show that indicators based on geography generate a significant amount of socioeconomic diversities within DHS clusters in Bangladesh. The random forest model allowed the models to forecast the continuous wealth index and categorical wealth quintile more than 80 percent of the time, indicating that geospatial data has the potential to serve as a useful proxy when census data are costly, old-fashioned, or unavailable. Geospatial-based models, though not a total replacement of elaborate household level surveys, are an answer to national and sub-national poverty monitoring, that are both scaled as well as can be integrated muster an update at a fast rate.

The second research question considered the validity of machine learning models in prediction of poverty when comparing with official DHS values. Through various algorithms: GLS, Random Forest, SVR, XGBoost, KNN and MLP, the predictive performance was good based on common evaluation measures. In the case of the wealth index, both the obtained R2, MAE, and RMSE values indicate that the models had repeatedly learned the relationships within the geospatial features and, therefore, could do a good task with unobservable clusters. To have the wealth quintile classification, the Accuracy, Precision and Recall have been used to further assure that the

models could distinguish between the household groups to form sensible socioeconomic groupings. The levels of performance indicate that machine learning could predict the official poverty indicators with a degree of fidelity in a situation where it is being trained using only environmental and infrastructural proxies.

The third research question was the one where it was aimed to determine the geospatial characteristics that have the most significant impact on poverty levels in Bangladesh. The analysis of feature importance showed that the variables included night lights during the night, land cover, vegetation indices, and the proximity to POI were highly important in explaining the socioeconomic disparity in the regions. The results appear to be similar to those of other world literature and the concept that neighbors the environmental patterns and urban fabric and service availability are closely related to household well-being is supported.

5.2 Limitations

Nevertheless, the study has many limitations although they promise good results. DHS data of the wealth are just available at the cluster level and therefore finer-scale validation is not possible. Minor quantity of house hold information implied that the models lacked sufficient information to train on resulting in overfitting in certain instances and underfitting in certain instances. The radius that DHS used was the 16 km square area which is quite a large margin. The area of 3 km square cluster radius would have led to improved augmentation and accurate results. As much as adequate and significant Geospatial features were obtained and utilized in training our data, there were features that hardly survived to carry any significance. Our modeling dataset could have been enhanced with additional geospatial features, e.g. population density and carbon emissions. 3 major POIs were employed in our modeling dataset. Less POI might not be necessarily the most effective. Absence of other POIs like bank, atm, shop, etc led to a low coefficient rate. Also POI data may be either incomplete or irregularly covered by the country. Also, geospatial characteristics on their own are incapable of defining some of the social or behavioral components of poverty, restricting the explanation of particular inequality at the regional level. They may also have seasonal variations in performance due to the performance of imagery in different seasons, cloud cover and the frequency of satellite revisiting.

5.3 Future improvements

Moving forward, it could be improved by the future study, which could entail deep learning networks like CNNs or hybrid spatio-temporal networks in extrapolating a more detailed information of the raw images. Time-series data could also be a good choice in promoting temporal stability and strengthening against environmental change. The predictive accuracy can be further honed by adding mobile phone metadata or transportation networks or crowdsourced social economic information. Lastly, a real-time or annual poverty-monitoring dashboard should be created to assist policymakers monitor the vulnerability of households in real-time; also, the policy makers can work towards targeting interventions more effectively. Overall, this study reveals that geospatial characteristics, using machine learning models, can provide a promising and scalable way of estimating poverty in Bangladesh. Although making these surveys a full substitute of a household survey is impossible, these approaches offer helpful, timely, and granular data that could substantially improve poverty-monitoring systems and make informed policy judgments.

REFERENCES

1. Aiken, E., Rolf, E., & Blumenstock, J. (2023). *Fairness and representation in satellite-based poverty maps: Evidence of urban-rural disparities and their impacts on downstream policy*. *arXiv*. https://arxiv.org/abs/2305.01783 arXiv
2. Alkire, S., Roche, J., Santos, M., & Seth, S. (2011). Multidimensional poverty index 2011: brief methodological note.
3. Zhao, X., Yu, B., Liu, Y., Chen, Z., Li, Q., Wang, C., & Wu, J. (2019). Estimation of Poverty Using Random Forest Regression with Multi-Source Data: A Case Study in Bangladesh. *Remote. Sens.*, 11, 375. https://doi.org/10.3390/rs11040375.
4. Newhouse, D. (2023). Small Area Estimation of Poverty and Wealth Using Geospatial Data: What have We Learned So Far?. *Calcutta Statistical Association Bulletin*, 76, 7 - 32. https://doi.org/10.1177/00080683231198591.
5. Zheng, X., Zhang, W., Deng, H., & Zhang, H. (2024). County-Level Poverty Evaluation Using Machine Learning, Nighttime Light, and Geospatial Data. *Remote. Sens.*, 16, 962. https://doi.org/10.3390/rs16060962.
6. Hu, S., Ge, Y., Liu, M., Ren, Z., & Zhang, X. (2022). Village-level poverty identification using machine learning, high-resolution images, and geospatial data. *Int. J. Appl. Earth Obs. Geoinformation*, 107, 102694. https://doi.org/10.1016/j.jag.2022.102694.
7. Putri, S., Wijayanto, A., & Pramana, S. (2022). Multi-source satellite imagery and point of interest data for poverty mapping in East Java, Indonesia: Machine learning and deep learning approaches. *Remote Sensing Applications: Society and Environment*. https://doi.org/10.1016/j.rsase.2022.100889.
8. Puttanapong, N., Martinez,, A., Addawe, M., Bulan, J., Durante, R., & Martillan, M. (2020). Predicting Poverty Using Geospatial Data in Thailand. *Asian Law eJournal*. https://doi.org/10.2139/ssrn.3785116.
9. Steele, J., Sundsøy, P., Pezzulo, C., Alegana, V., Bird, T., Blumenstock, J., Bjelland, J., Engø-Monsen, K., De Montjoye, Y., Iqbal, A., Hadiuzzaman, K., Lu, X., Wetter, E., Tatem, A., & Bengtsson, L. (2017). Mapping poverty using mobile phone and satellite data. *Journal of the Royal Society Interface*, 14. https://doi.org/10.1098/rsif.2016.0690.

10. Tang, J., Zhao, X., Zhang, F., Qiu, A., & Tao, K. (2024). Poverty Estimation Using a ConvLSTM-Based Model With Multisource Remote Sensing Data: A Case Study in Nigeria. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 3516-3529. https://doi.org/10.1109/jstars.2024.3353754.
11. De Nicolò, S., Fabrizi, E., & Gardini, A. (2023). Mapping poverty at multiple geographical scales. .
12. Ramadhan, R., Wijayanto, A., & Pramana, S. (2023). Geospatial Big Data Approaches to Estimate Granular Level Poverty Distribution in East Java, Indonesia using Machine Learning and Deep Learning Regressions. *Proceedings of The International Conference on Data Science and Official Statistics*. https://doi.org/10.34123/icdsos.v2023i1.359.
13. S, N., Kumar, G., M, K., & S, J. (2025). Poverty Mapping with Satellite Vision for Social Good Using R-CNN Algorithm. *2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*, 1306-1311. https://doi.org/10.1109/icmlas64557.2025.10968622.
14. Jean, N., Burke, M., Xie, S., Davis, W., Lobell, D., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353, 790 - 794. https://doi.org/10.1126/science.aaf7894.
15. Browne, C., Matteson, D., Mcbride, L., Hu, L., Liu, Y., Sun, Y., Wen, J., & Barrett, C. (2021). Multivariate random forest prediction of poverty and malnutrition prevalence. *PLoS ONE*, 16. https://doi.org/10.1371/journal.pone.0255519.
16. Yu, B., Chen, F., Ye, C., Li, Z., Dong, Y., Wang, N., & Wang, L. (2023). Temporal expansion of the nighttime light images of SDGSAT-1 satellite in illuminating ground object extraction by joint observation of NPP-VIIRS and sentinel-2A images. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2023.113691.
17. Engstrom, R., Hersh, J., & Newhouse, D. (2017). Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being. *Econometrics: Data Collection & Data Estimation Methodology eJournal*. https://doi.org/10.1596/1813-9450-8284.
18. Ni, Y., Li, X., Ye, Y., Li, Y., Li, C., & Chu, D. (2021). An Investigation on Deep Learning Approaches to Combining Nighttime and Daytime Satellite Imagery for Poverty

Prediction. *IEEE Geoscience and Remote Sensing Letters*, 18, 1545-1549. https://doi.org/10.1109/lgrs.2020.3006019.

19. Tingzon, I., Orden, A., Go, K., Sy, S., Sekara, V., Weber, I., Fatehkia, M., García-Herranz, M., & Kim, D. (2019). MAPPING POVERTY IN THE PHILIPPINES USING MACHINE LEARNING, SATELLITE IMAGERY, AND CROWD-SOURCED GEOSPATIAL INFORMATION. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 425-431. https://doi.org/10.5194/isprs-archives-xlii-4-w19-425-2019.
20. Yong, Z., Li, K., Xiong, J., Cheng, W., Wang, Z., Sun, H., & Ye, C. (2022). Integrating DMSP-OLS and NPP-VIIRS Nighttime Light Data to Evaluate Poverty in Southwestern China. *Remote. Sens.*, 14, 600. https://doi.org/10.3390/rs14030600.
21. Yu, B., Chen, F., Ye, C., Li, Z., Dong, Y., Wang, N., & Wang, L. (2023). Temporal expansion of the nighttime light images of SDGSAT-1 satellite in illuminating ground object extraction by joint observation of NPP-VIIRS and sentinel-2A images. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2023.113691.
22. Owusu, M., Engstrom, R., Thomson, D., Kuffer, M., & Mann, M. (2023). Mapping Deprived Urban Areas Using Open Geospatial Data and Machine Learning in Africa. *Urban Science*. https://doi.org/10.3390/urbansci7040116.
23. Hall, O., Ohlsson, M., & Rögnvaldsson, T. (2022). A review of explainable AI in the satellite data, deep machine learning, and human poverty domain. *Patterns*, 3. https://doi.org/10.1016/j.patter.2022.100600.
24. Srishti Gulecha, R., Muthu Reshmi, K., Rishitha, N., & Vani, K. (2024). Poverty Mapping in India using Machine Learning and Deep Learning Techniques. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *10*, 319-326
25. Agyemang, F. S. K., Memon, R., Wolf, L., & Fox, S. (2023). High-resolution rural poverty mapping in Pakistan with ensemble deep learning. *PLOS ONE*, 18(4), e0283962. https://doi.org/10.1371/journal.pone.0283962
26. Ayush, K., Uzkent, B., Tanmay, K., Burke, M., Lobell, D., & Ermon, S. (2021, May). Efficient poverty mapping from high resolution remote sensing images. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 1, pp. 12-20).

27. Aiken, E., Rolf, E., & Blumenstock, J. (2023). Fairness and representation in satellite-based poverty maps: Evidence of urban-rural disparities and their impacts on downstream policy. *arXiv preprint arXiv:2305.01783*.
28. Das, S., Deepawansa, D., & Lahiri, P. (2025). Multidimensional Poverty Mapping for Small Areas. *arXiv preprint arXiv:2510.08898*.
29. Yin, J., Qiu, Y., & Zhang, B. (2020). Identification of Poverty Areas by Remote Sensing and Machine Learning: A Case Study in Guizhou, Southwest China. *ISPRS Int. J. Geo Inf.*, 10, 11. https://doi.org/10.3390/ijgi10010011.
30. Corral, P., Henderson, H., & Segovia, S. (2023). Poverty Mapping in the Age of Machine Learning. *Journal of Development Economics*. https://doi.org/10.1596/1813-9450-10429.
31. Hersh, J., Engstrom, R., & Mann, M. (2021). Open data for algorithms: Mapping poverty in Belize using open satellite derived features and machine learning. *27*(2), 362–382. https://doi.org/10.1080/02681102.2020.1827088
32. Tingzon, I., Orden, A., Go, K., Sy, S., Sekara, V., Weber, I., Fatehkia, M., García-Herranz, M., & Kim, D. (2019). Mapping poverty in the Philippines using machine learning, satellite imagery, and crowd-sourced geospatial information. *ISPRS Archives, 42(4/W19)*, 425–431. https://doi.org/10.5194/isprs-archives-XLII-4-W19-425-2019
33. Pokhriyal, N., & Jacques, D. (2017). Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E9783 - E9792. https://doi.org/10.1073/pnas.1700319114.
34. Smythe, I., & Blumenstock, J. (2022). Geographic microtargeting of social assistance with high-resolution poverty maps. *Proceedings of the National Academy of Sciences of the United States of America*, 119. https://doi.org/10.1073/pnas.2120025119.
35. Hall, O., Dompae, F., Wahab, I., & Dzanku, F. (2023). A review of machine learning and satellite imagery for poverty prediction: Implications for development research and applications. *Journal of International Development*. https://doi.org/10.1002/jid.3751.
36. Elvidge, C., Sutton, P., Ghosh, T., Tuttle, B., Baugh, K., Bhaduri, B., & Bright, E. (2009). A global poverty map derived from satellite data. *Comput. Geosci.*, 35, 1652-1660. https://doi.org/10.1016/j.cageo.2009.01.009.

APPENDICES

Appendix A: Dataset Availability

Dataset Link : https://dhsprogram.com/data/dataset/Bangladesh_Standard-DHS_2022.cfm?flag=0

LIBRARY CLEARANCE

PLAGARISM REPORT

ACCOUNT CLEARANCE