Required data sources

- Demographic and Health Survey (DHS)
  The household wealth for each of the clusters is averaged and held as the principal indicator for socioeconomic well-being as it represents the possession of common assets for these households
- Nightime Luminosity Data
- Google Static Map Images
- Open Street Map Data

The Target Variable is the Wealth Index

Trained Model to find Accuracy with

- Nighttime Luminosity as a stand-alone predictor
  The results are relatively close to one another with the XGBoost model being able to explain the most variation in socio-economic status at 66.28 percent R-square score.
- Daytime Satellite Images as a stand-alone predictor
  Basic Features
  daytime satellite images such as the max, min, mean, median and standard deviation of the .attened RGB color channels of the images for locations close to the household clusters.
  The results demonstrate that the rudimentary information of the satellite images is only able to explain 36-47.87 percent of the variation in wealth levels, depending on the algorithm used.
  Deep Features
  The next step in this procedure is to extract deep features and insights from the daytime images using Convolutional Neural Networks (CNN), taking advantage of the existing model pretrained on the ImageNet dataset. capture the features indicative of socioeconomic status such as roads, rivers, land usage, or roofng materials.
  These features will subsequently be tabulated and utilized as predictors for wealth level within the clusters to discern whether they can improve upon the baseline of basic features computed previously.
  The models fitted with the deep features extracted from the daytime satellite images improve considerably from those of basic features, explaining up to 60.33 percent of the overall variance in average wealth index level of the households by employing the XGBoost method.
- OSM(Open Street Map) data as a stand-alone predictor
  Road Features
  Buildings Features:
  Land Use Features:
  Point-of-interest Features: speci%c point locations that people may %nd useful such as hospitals, schools, supermarkets, public attractions, etc.
  building features appear to hold the most correlation with average wealth index

the OSM data as a stand-alone feature space is able to explain about 43-56.99 percent of the total variance in socio-economic wellbeing.

Transfer Learning approach

Nighttime Luminosity & Daytime Satellite Images Model

The hybrid model using both nighttime luminosity and daytime satellite images' deep features achieves the best performance out of all permutations of features combination at 72 percent for wealth index R-squared score.

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

Nighttime Luminosity & OSM Model

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

household clusters

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