Automating Electricity Access Prediction with Satellite Imagery

Fangge Deng, Shamikh Hossain, Prithvir Jhaveri, Ashley Meuser, Harshvardhan Sanghi, Joe Squillace, Anuj Thakkar, Brian Wong, Xiaolan You

Faculty Advisors: Dr. Kyle Bradbury (Energy Initiative), Dr. Leslie Collins (Pratt), Dr. T. Robert Fetter (Nicholas Institute), Dr. Marc Jeuland (Sanford), Dr. Timothy Johnson (Nicholas)

Duke UNIVERSITY | BASS CONNECTIONS

Duke ENERGY INITIATIVE

Introduction & Overview

Energy access is a key resource in improving the wellbeing, economic prosperity, and gender equality of a region. Particularly, it is linked to an increase in the number of students enrolled in school, time students spend studying, business hours, agricultural productivity and labor supply, and a reduction of the poverty rate (Khandker, S.R. et al., 2012).

Despite these benefits, an estimated 1.2 billion people do not have electricity access, and even more have either too unreliable or insufficient supply to reap the aforementioned welfare gains (World Energy Outlook, 2017).

This study aims to fill current data gaps on global energy access, particular in resolving finer-scale geographic access metrics. The study aims to overcome inaccurate or biased data, to provide a method for continuously measuring progress in electricity access over time, and to get more refined electricity access data on a village-to-village basis.

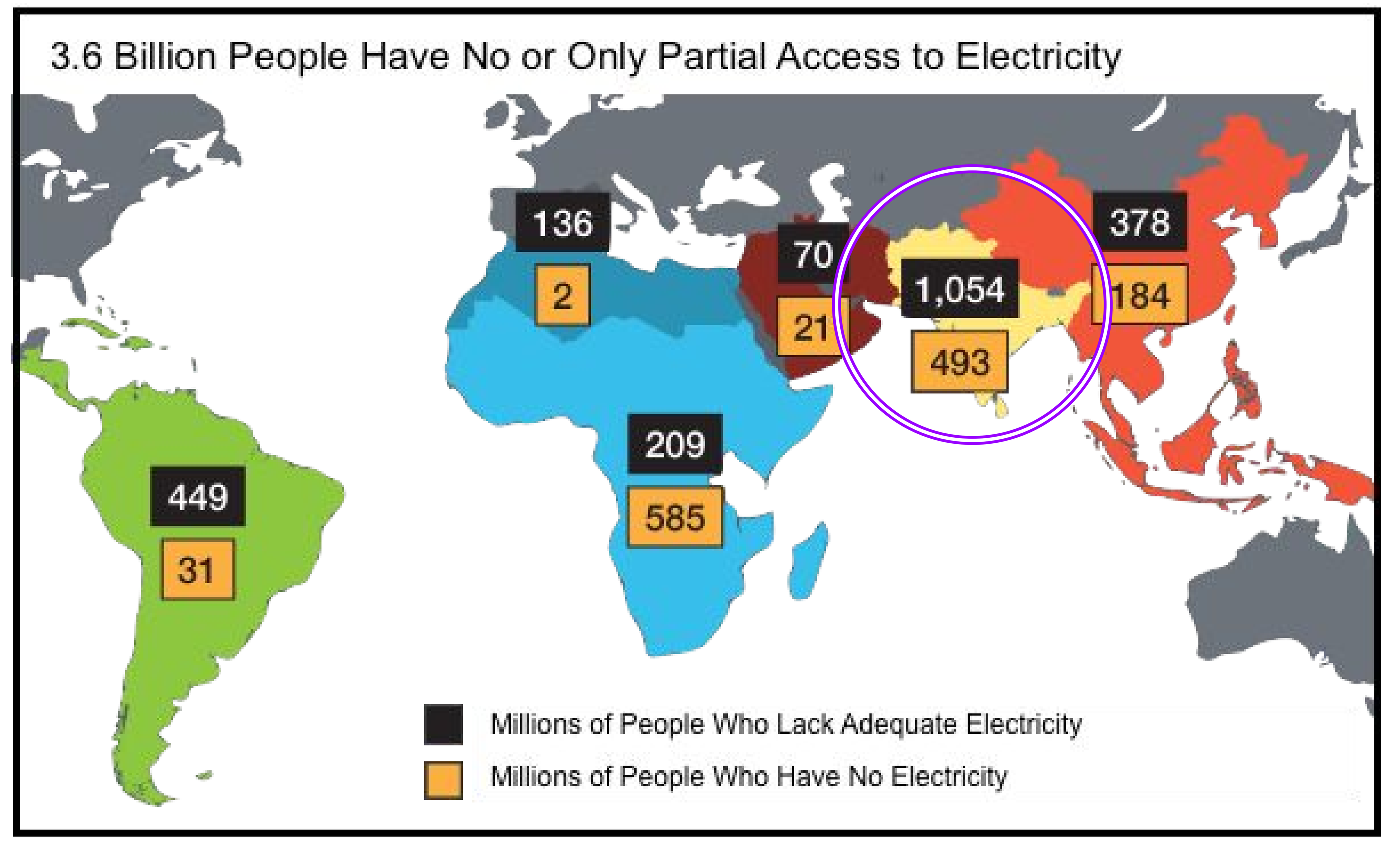


Figure 1: Global population with access to no or inadequate electricity

Source: International Energy Agency World Energy Outlook 2011 and The World Bank World Development Indicators 2011

The primary deliverable of this preliminary study is to produce a functional collaborative machine learning infrastructure with VIIRS Lights at Night data capable of predicting electrification rates at the village level in Bihar, India. In doing so, the study aims to advance energy access mapping at finer resolutions (Shi et al., 2014).

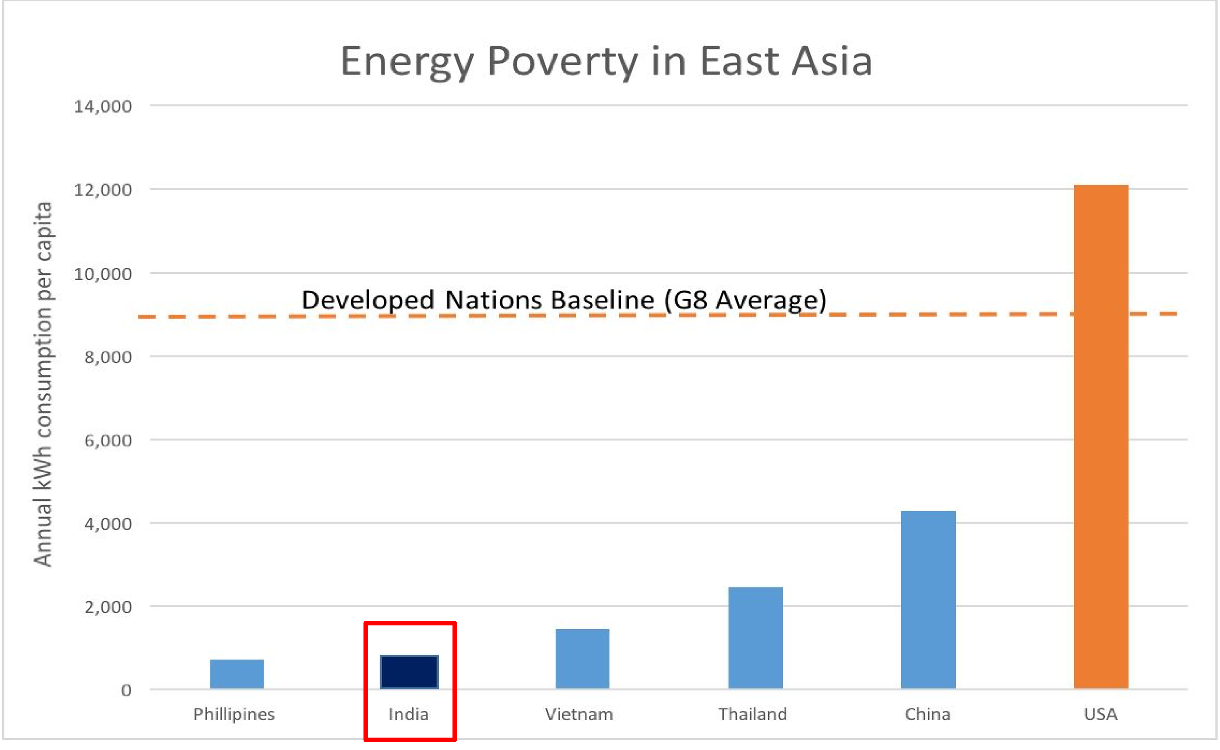


Figure 2. Energy poverty in East Asia India relative to developed nations and the G8 average.

Process Summary

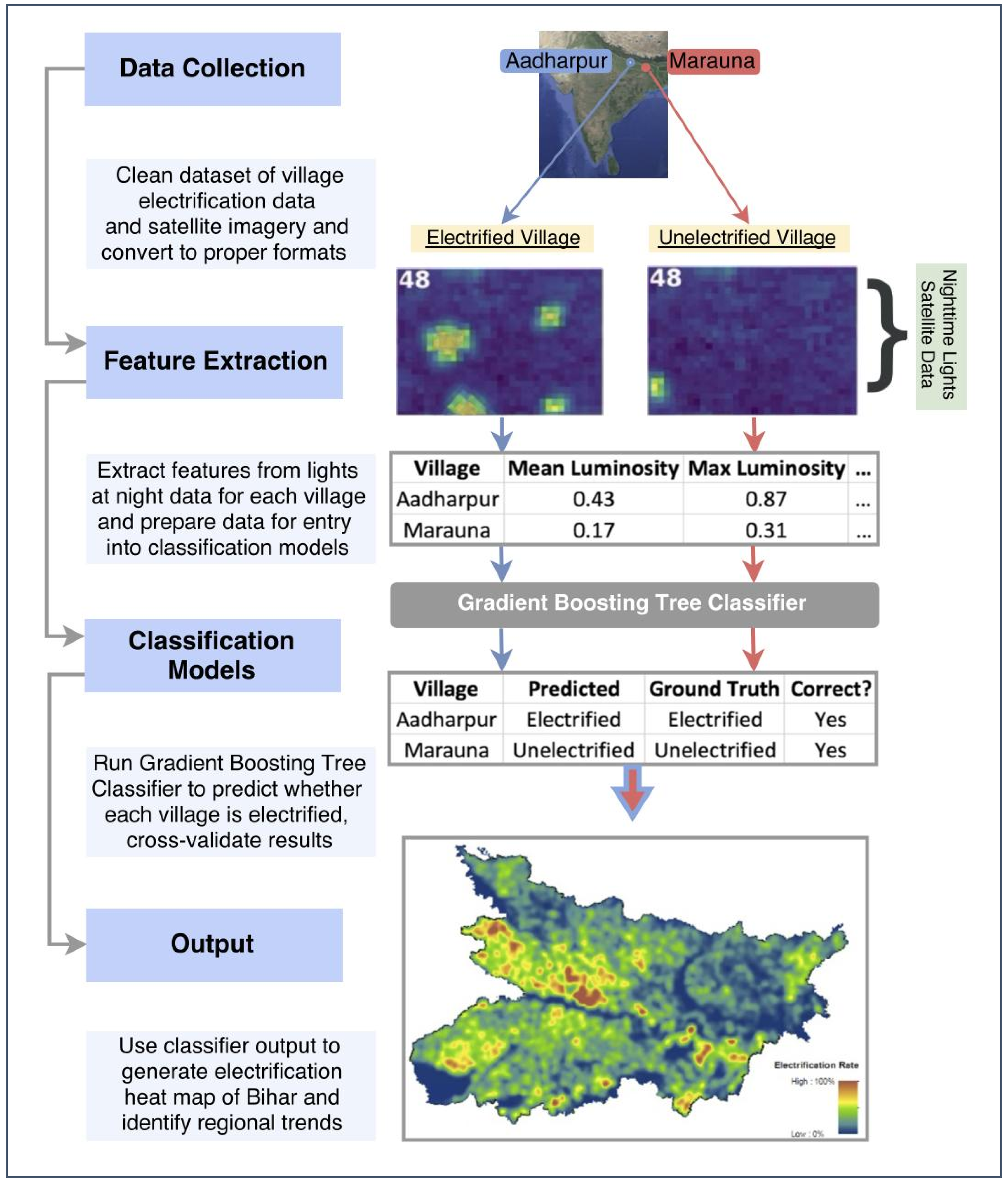


Figure 3. Process of data collection, feature extraction, village electrification classification and output validation.

Our team performed operations on VIIRs band arrays that correspond to villages in Bihar. We first masked each array to include only pixels that are within the boundaries of their corresponding village. From these arrays we calculated the mean, max, and sum radiance values as well as the 10th, 25th, 50th, 75th, and 90th radiance percentiles. We input these values as features to train our classifier to predict for each village whether it belongs to the "electrified" class or the "unelectrified" class, for which the threshold for being considered electrified was if 10% of the households have electricity access. Finally, we produced an electrification heat map of Bihar from the classifier's predictions for all villages.

Results

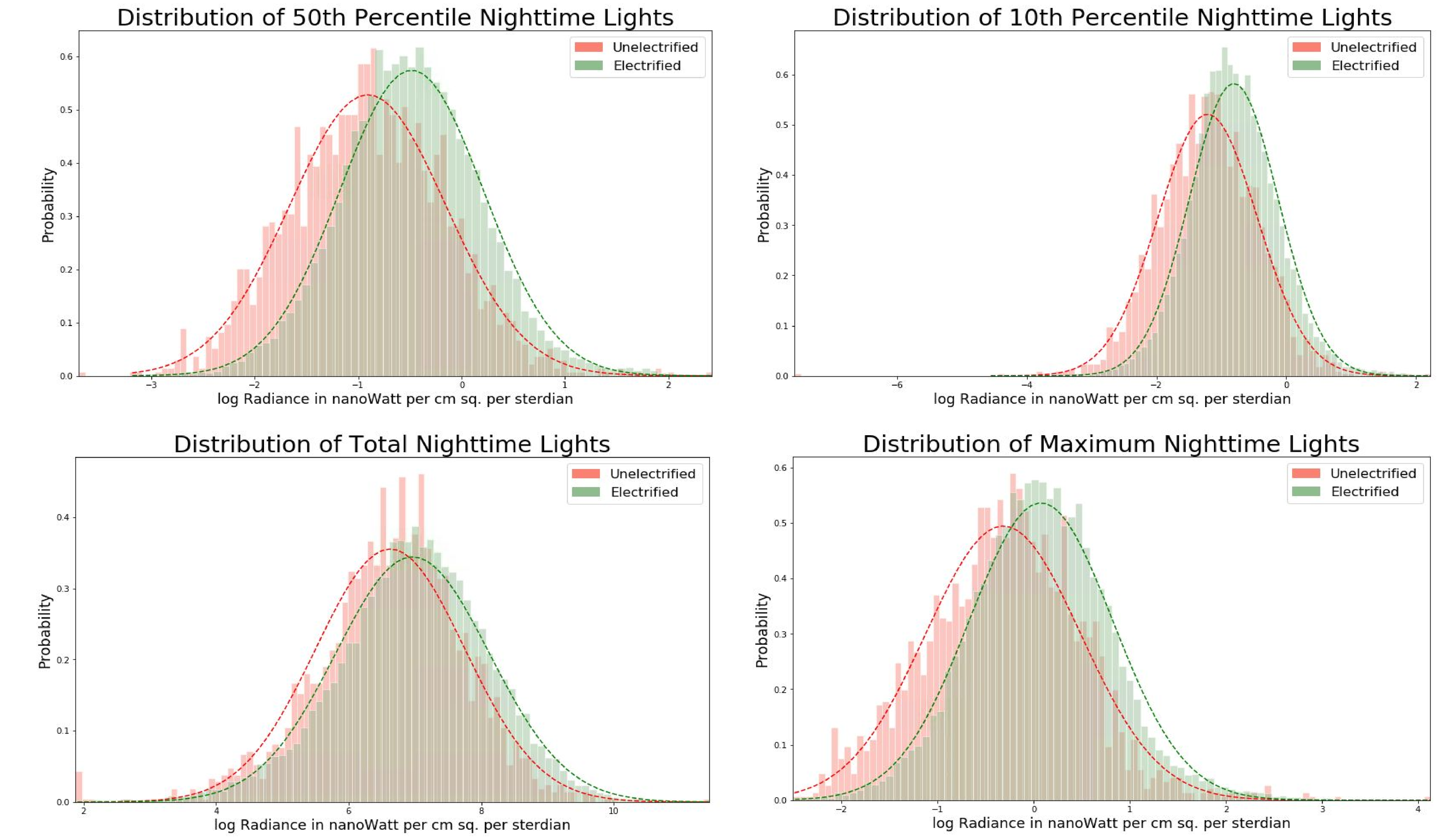


Figure 4. Distribution of top four features of importance separated by Electrified and Unelectrified classes.

Machine learning classifiers were used to predict village data into two categories: un-electrified and electrified. The classifier identified villages with less than 10% access as unelectrified and villages with larger values as electrified, thresholds that were based on the literature (Min and Gaba, 2014). The results of the Gradient Boosting Classifiers are shown to right in the form of Receiver Operating Characteristic (ROC) curves in Figure 5. Min and Gaba 2014 identified difficulties for nighttime lights data to detect small villages, even if electrified, thus three thresholds (minimum number of households in a village) were selected to pre-filter the data.

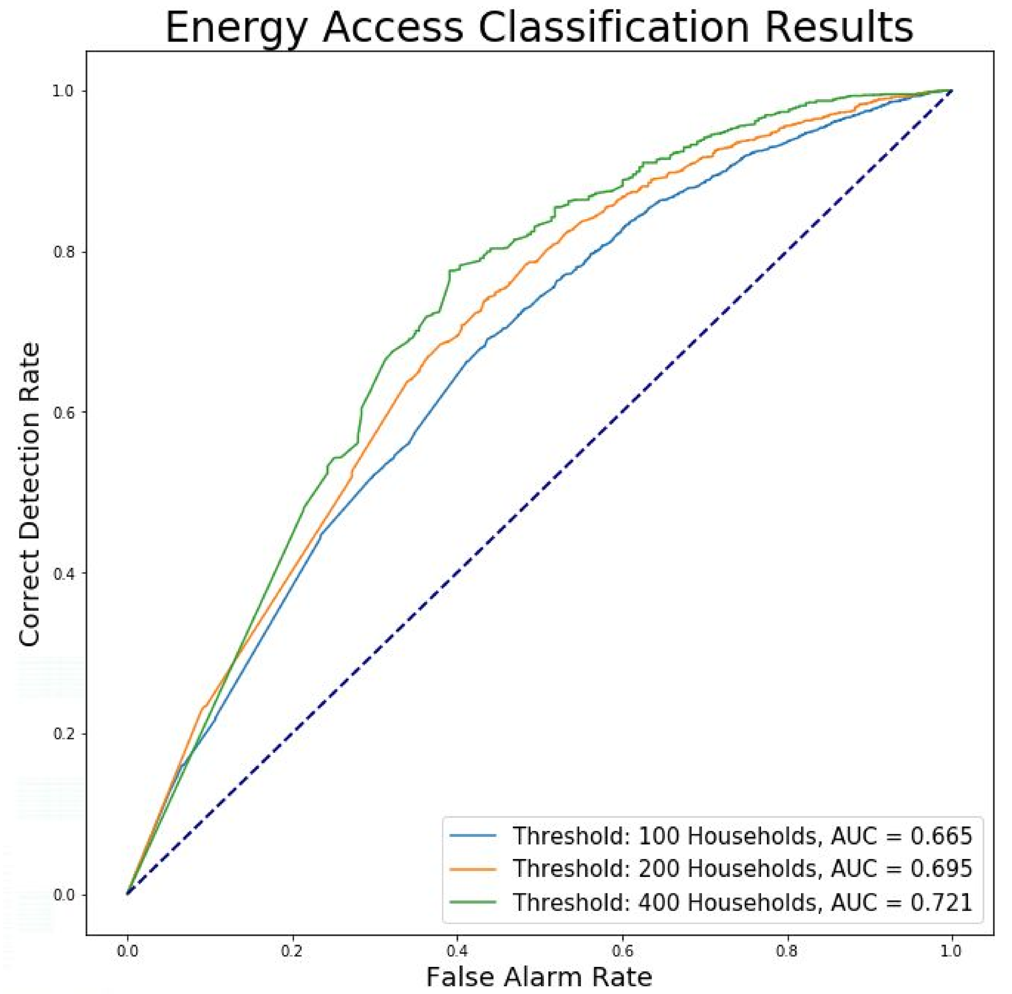


Figure 5. ROC Curve demonstrating results of energy access projections separated by three models, each by selecting a minimum number of households threshold.

Conclusion & Future Steps

This preliminary study demonstrated that the data are nonlinear, and thus a nonlinear classifier will be more successful in differentiating between village electrification status. It also yielded a development environment that is flexible to expanded feature extraction and testing for quicker iterations of modeling in the future. These descriptors may include other features detected from VIIRS or additional satellite imagery, such as the presence of buildings, irrigation, and other energy access indicators. Given the constraints of using only nighttime lights to predict energy access, this preliminary study indicates significant potential for model improvements with the incorporation of additional predictors.

Sources

- Energy access database. (n.d.). Retrieved October 03, 2017, from http://www.worldenergyoutlook.org/resources/energydevelopment/energyaccessdatabase/.
- Khandker, S.R., Samad, H.A., Ali, R., & Barnes, D.F. (2012). Who Benefits Most from Rural Electrification? Evidence in India. *Policy Research Working Papers*. doi:10.1596/1813-9450-6095
- Min, B., & Gaba, K. M. (2014). Tracking Electrification in Vietnam Using Nighttime Lights. Remote Sensing, 6(10), 9511–9529.
- Shi, K., Yu, B., Huang, Y., Hu, Y., Yin, B., Chen, Z., Wu, J. (2014). Evaluating the Ability of NPP-VIIRS Nighttime Light Data to Estimate the Gross Domestic Product and the Electric Power Consumption of China at Multiple Scales: A Comparison with DMSP-OLS Data. Remote Sensing, 6(2), 1705–1724.

Estimating Electricity Access using Satellite Imagery

Duke UNIVERSITY | BASS CONNECTIONS

Fangge Deng, Shamikh Hossain, Prithvir Jhaveri, Ashley Meuser, Harshvardhan Sanghi, Joe Squillace, Anuj Thakkar, Brian Wong, Xiaolan You

Faculty Advisors: Dr. Kyle Bradbury (Energy Initiative), Dr. Leslie Collins (Pratt), Dr. Marc Jeuland (Sanford), Dr. Timothy Johnson (Nicholas)

Duke ENERGY INITIATIVE

Introduction & Overview

Energy access is a key resource in improving the wellbeing, economic prosperity, and gender equality of a region. Particularly, it is linked to an increase in the number of students enrolled in school, time students spend studying, business hours, agricultural productivity and labor supply, and a reduction of the poverty rate (Khandker, S.R. et al., 2012).

Despite these benefits, an estimated 1.2 billion people do not have electricity access, and even more have either too unreliable or insufficient supply to reap the aforementioned welfare gains (World Energy Outlook, 2017).

This study aims to fill current data gaps on global energy access, particular in resolving finer-scale geographic access metrics. The study aims to overcome inaccurate or biased data, to provide a method for continuously measuring progress in electricity access over time, and to get more refined electricity access data on a village-to-village basis.

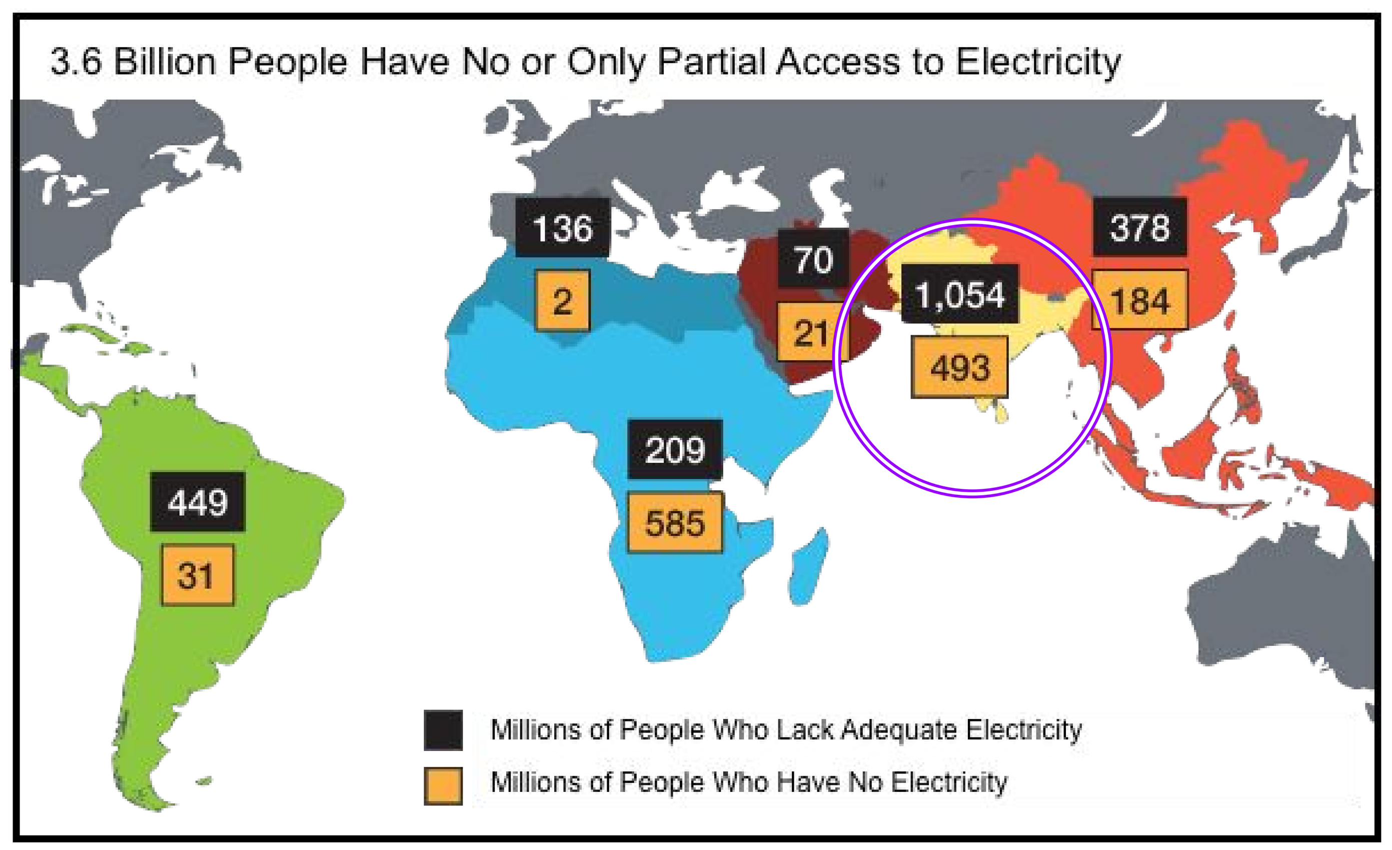
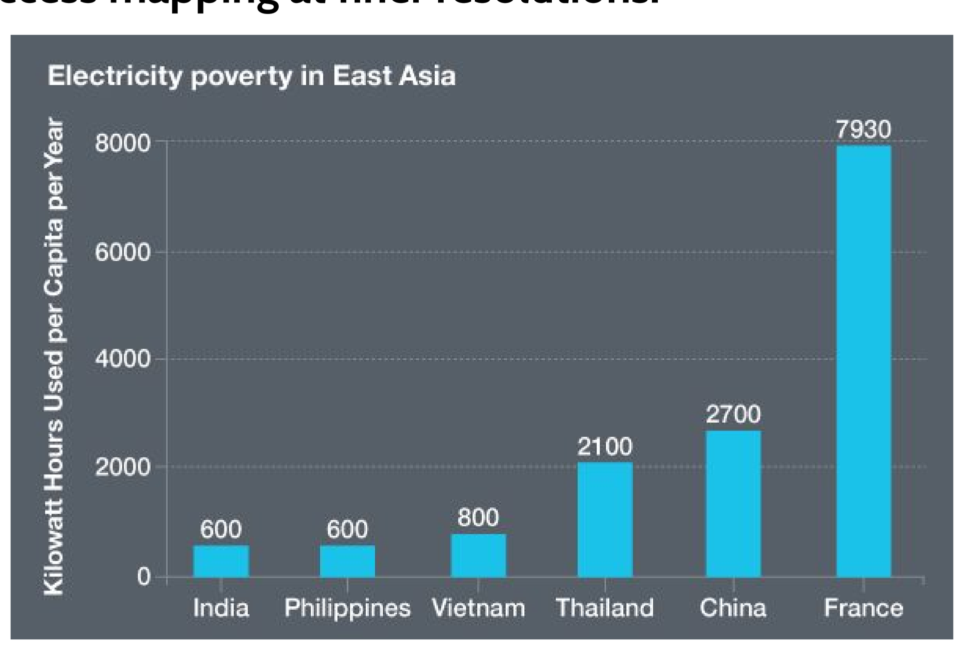


Figure 1: Global population with access to no or inadequate electricity

Source: International Energy Agency World Energy Outlook 2011 and The World Bank World Development Indicators 2011

The primary deliverable of this preliminary study is to produce a functional collaborative machine learning infrastructure with VIIRS Lights at Night data capable of predicting electrification rates at the village level in Bihar, India. In doing so, the study aims to advance energy access mapping at finer resolutions.



Sources

- Energy access database. (n.d.). Retrieved October 03, 2017, from http://www.worldenergyoutlook.org/resources/energydevelopment/energyaccessdatabase/.
- Khandker, S.R., Samad, H.A., Ali, R., & Barnes, D.F. (2012). Who Benefits Most from Rural Electrification? Evidence in India. *Policy Research Working Papers*. doi:10.1596/1813-9450-6095

Process Summary

Input Data | Extracted/Output Data

Data Collection

Download dataset of village household electrification data in Bihar and satellite imagery, convert to proper formats.

Satellite Data

Village	% Electrified
Aadharpur	22.80%
Marauna	UNKNOWN
...	...

Indian Government Electrification Data

Feature Extraction

Extract features from lights at night data for each village and prepare data for entry into classification models.

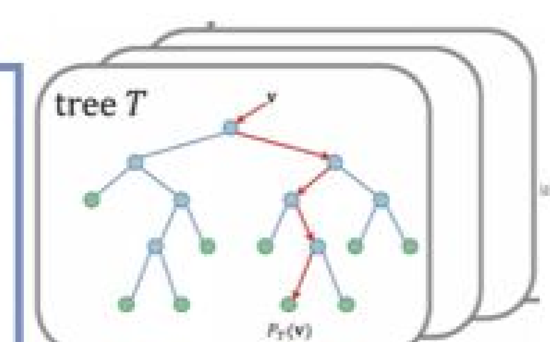
Lights at Night

Village	Mean	Max	...
Aadharpur	0.40	0.86	...
Marauna	0.32	0.71	...
...	...	...	...

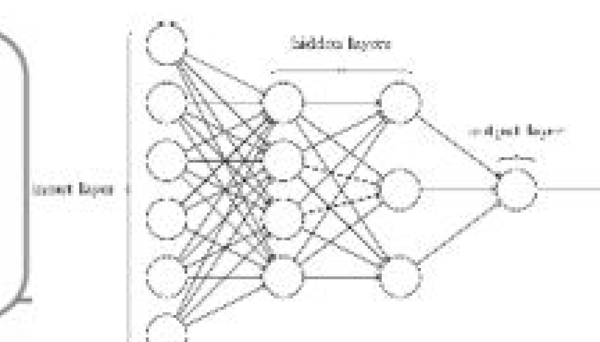
Radiance values

Classification Models

Run Random Forest and MLP binary classification models to estimate electricity access from lights at night values.



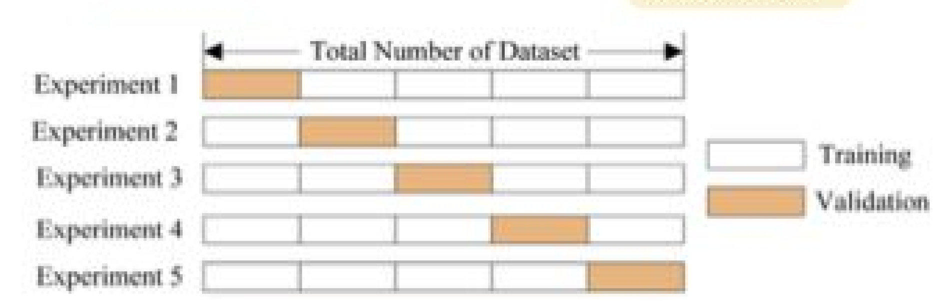
Random Forest



Multilayer Perceptron

Validation

Use cross validation to test classifiers and evaluate performance using ROC curve.



K-folds cross validation

Output

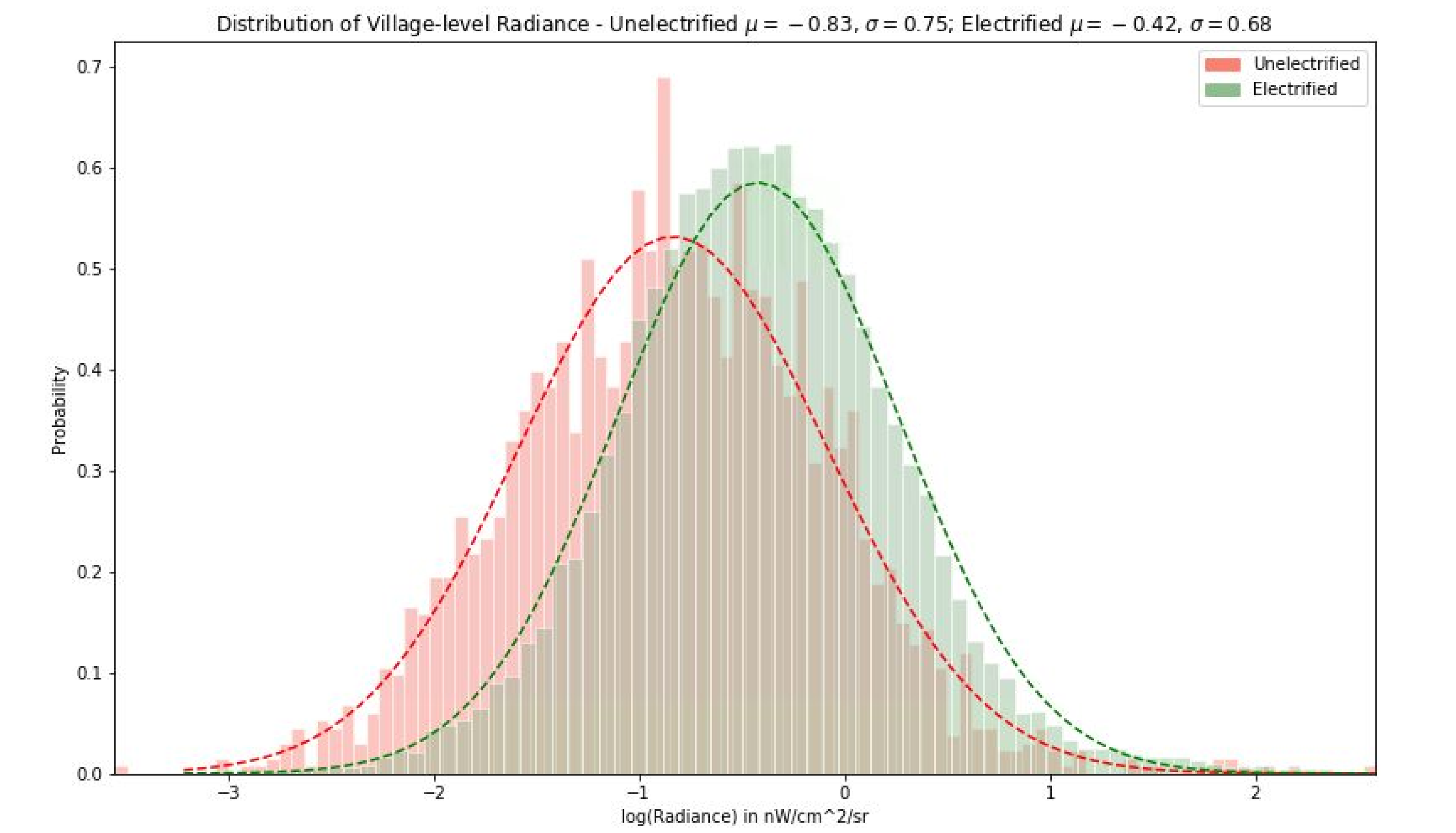
Data Processing

Operations were performed on VIIRs band arrays that corresponded to villages in Bihar. Each array was first masked to include only pixels that corresponded to the specific village. From these arrays the mean and max radiance values and the 10%, 25%, 50%, 75%, and 90% radiance percentiles were calculated.

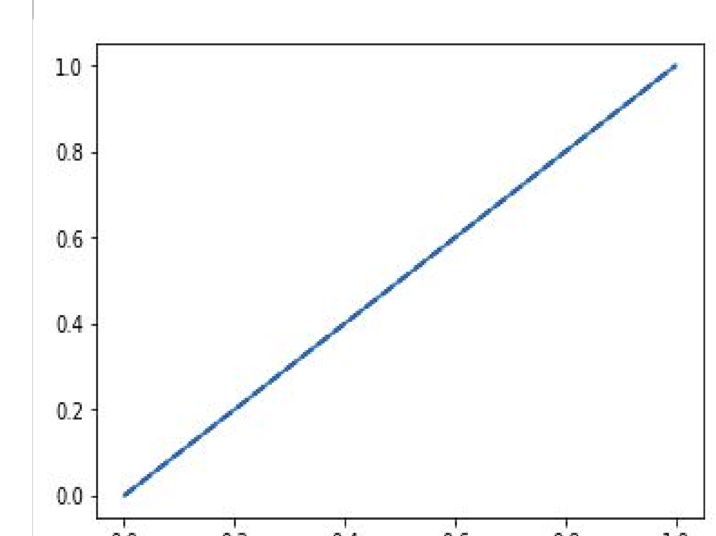
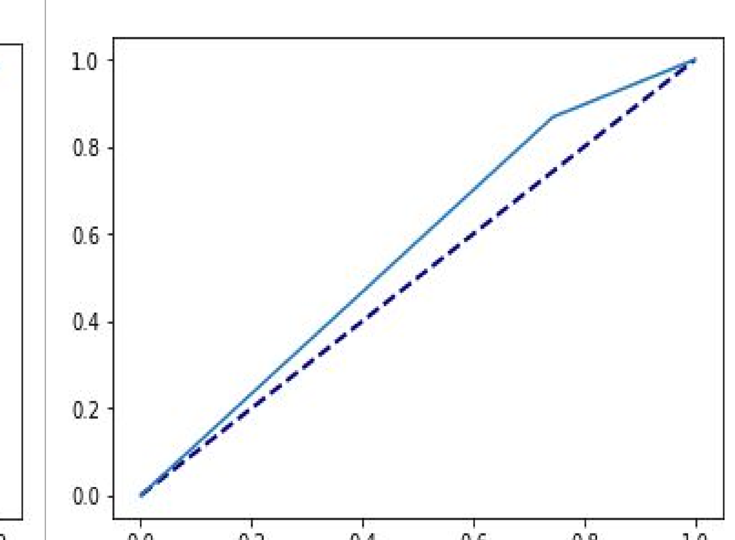
Our team consolidated the following software infrastructure for adding new features:

1. load_features.py to add new features for each village
2. viirs_feature_extraction.ipynb to merge with other village data and produce cleaned data ready for experimentation with different modeling techniques.

Machine Learning Model

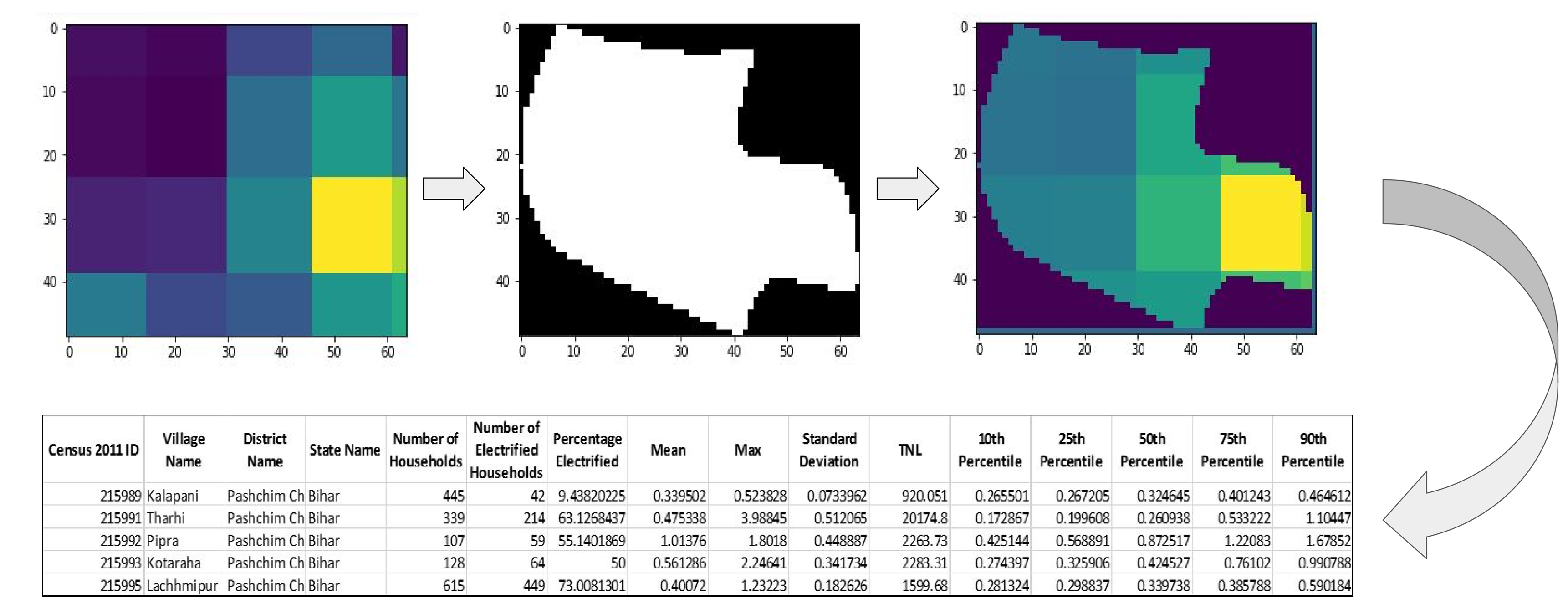


Machine learning classifiers were run to classify the village data into two categories: un-electrified and electrified, ere represented with labels 1 and 2. The first classifier assumed less than 10% as not electrified, the second classifier assumed less than 50% as not electrified. These thresholds were chosen based on visual analysis of data clustering. The results of our classifiers are shown below in the form of Receiver operating characteristic (ROC) curves.

Classifier	Neural Network MLP 10%	50%
Accuracy	91.75%	53.83%
Area under ROC	0.50	0.5624
ROC curve		

Conclusion & Future Steps

This exploratory study demonstrated that the data are nonlinear, and thus a nonlinear classifier will be more successful in differentiating between village electrification rates. It also yielded Python scripts that are flexible to new feature extraction and testing for quicker iterations of modeling in the future. These features may include other objects detected from the VIIRS imagery, such as buildings or other features of the built environment. However, tests are limited by the size of the village imagery dataset, so future tests must continue to limit the number of features that are tested. Once more features are extracted, a CNN may prove effective at determining village electrification rates.



Census 2011 ID	Village Name	District Name	State Name	Number of Households	Number of Electrified Households	Percentage Electrified	Mean	Max	Standard Deviation	TNL	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
215989	Kalapani	Pashchim Ch	Bihar	445	42	9.43820225	0.339502	0.523828	0.0733962	920.051	0.265501	0.267205	0.324645	0.401243	0.464612
215991	Tharhi	Pashchim Ch	Bihar	339	214	63.1268437	0.475338	3.98845	0.512065	20174.8	0.172867	0.199608	0.260938	0.533222	1.10447
215992	Pipra	Pashchim Ch	Bihar	107	59	55.1401869	1.01376	1.8018	0.448887	2263.73	0.425144	0.568891	0.872517	1.22083	1.67852
215993	Kotaraha	Pashchim Ch	Bihar	128	64	50	0.561286	2.24641	0.341734	2283.31	0.274397	0.325906	0.424527	0.76102	0.990788
215995	Lachhmipur	Pashchim Ch	Bihar	615	449	73.0081301	0.40072	1.23223	0.182626	1599.68	0.281324	0.298837	0.339738	0.385788	0.590184