Power to the People: Predicting Levels of Household Electricity Consumption in Low-Access Settings

Simone Fobi
Columbia University
NYC, USA
sf2786@columbia.edu

Joel Mugyenyi
University of Massachusetts
Amherst, USA
jmugyenyi@umass.edu

Jay Taneja
University of Massachusetts
Amherst, USA
jtaneja@umass.edu

Vijay Modi
Columbia University
NYC, USA
modi@columbia.edu

ABSTRACT

Today, 840 million people remain without electricity access at home. Still, many governments struggle to fund system expansion despite goals for universal electricity access. Aiming to balance traditional approaches like centralized grid extension with nascent distributed electricity technologies while maximizing the reach of scarce electrification budgets, the field of integrated electricity system planning has emerged. A crucial input into electricity system planning is accurate and detailed consumption predictions, but accuracy is difficult in hard-to-reach settings comprised of households who have little to no prior experience with electricity services and are likely to have highly heterogeneous consumption levels once connected. This paper presents a novel data-driven approach to the challenge of predicting future electricity consumption for individual buildings using daytime satellite imagery, a widely-available data source.

Our approach trains a Convolutional Neural Network (CNN) over daytime satellite imagery with a sample of bills from 27,585 geo-referenced electricity customers in Kenya (0.01% of Kenya's residential customers). We achieve an accuracy that is competitive with other approaches while far better capturing the challenging variability of consumption that underpins electricity consumption distributions. We also evaluate the incorporation of other geospatial datasets into the training process, including nighttime lights and census-derived data. Additionally, we compare our model's performance to a best-in-class survey independently collected by the World Bank, showing strong agreement throughout the country (sample weighted r^2 of 0.84 when excluding the over-sampled and already-electrified capital city of Nairobi, r^2 of 0.69 otherwise). Our technique can help to inform not only site selection (as previous approaches to consumption prediction have) but also distribution-level planning, as we provide highly granular predictions at the level of individual structures. We have publicly released a dataset of granular residential consumption predictions over the entire country of Kenya and have an active set of electricity access practitioners incorporating our predictions into their operations. This adoption validates that our approach is solving a real-world challenge and can also be applied in other countries striving for universal electricity access.

ACM Reference Format:
Simone Fobi, Joel Mugyenyi, Jay Taneja, and Vijay Modi. 2018. Power to the People: Predicting Levels of Household Electricity Consumption in Low-Access Settings. In *Singapore '21: ACM SIGKDD Conference on Knowledge Discovery and Data Mining, August 14–18, 2021, Singapore.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Over the past decade, improved engineering, reduced costs, and innovative business models for electrification technologies have contributed to increasing access to electricity around the world. However, 840 million people still lack access to electricity services [35], most of them residing in places that are difficult to reach and, as a result, expensive to serve [28, 37]. While achieving universal electricity access is a priority, many energy providers struggle to meet this mandate while remaining financially viable. At low consumption levels, which are common among new customers in low-income settings, energy providers may struggle to recover the cost of servicing a grid connection even with government subsidies[15]. Alternatives to grid extension such as Solar Home Systems (SHS) can support smaller loads without the wire required for a grid connection, while in some cases clustered homes (with clusters far from each other) can make mini-grids viable[16]. If electricity consumption predictions can assist in identifying appropriate alternatives, then a country can provide electricity access to a larger population with the same investment.

We define electricity consumption prediction as follows: given current information about a set of unelectrified households, what is the latent or expected future consumption levels after these households received a grid connection? In our formulation, we use information about households at time $t_{\mathbf{un}electrified}$ to predict their consumption levels at $t_{electrified}$. It is worth noting that this formulation differs significantly from consumption estimation, where information about households at $t_{electrified}$ is used to estimate consumption levels at $t_{electrified}$. While the latter makes for an

Singapore '21, August 14–18, 2021, Singapore

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

easier research question, the former underpins the questions that an electrification planning decision-maker faces.

In this paper, we introduce a data-driven method to predict levels of latent future electricity consumption for residential buildings upon receiving an electricity connection. Our approach trains a Convolutional Neural Network (CNN) to predict levels of consumption from overhead daytime images of homes. Although accurate individual electricity consumption predictions are difficult to achieve [31], we show that high-resolution daytime satellite imagery (0.5 m/pixel) contains a larger number of relevant features than approaches using lower-resolution widely available geospatial data such as Nighttime Lights and census indicators.

An approach based on daytime imagery also captures more of the spatial heterogeneity inherent in consumption, an important characteristic for distribution network planning[1]. Previous work [24] shows that taking into account heterogeneity of individual consumption produces electrification plans that are less costly than plans that assume homogeneous consumption. Thus, we adopt evaluation metrics that maximize overall performance – such as accuracy and F1-score – as well as a domain-specific metric – the spatial variability in the underlying electricity consumption data – in order to better support distribution network planning.

In this work, we show the value of combining satellite images with non-visual datasets to jointly learn about levels of consumption, and we also demonstrate that learning on a proxy building segmentation task improves prediction performances especially in low-data regimes. Finally, we present our novel dataset of predicted levels of consumption for individual buildings, aggregated at 250m resolution (publicly available through a custom API) and discuss its value to our active and growing user base. Our predictions for 11.9 million buildings show strong agreement throughout the country (sample weighted r^2 of 0.69 at the county-level) with the World Bank's Multi-Tier Framework survey of electricity consumption among households. Our dataset can be used to support large-scale site selection of electrification projects and as an aid in distribution network planning. Given the potential dependence of consumption on tariffs and policies for recovery of first costs, the specific results of the data-driven approach here apply to Kenya. However, we believe that our methods can be extended to other countries, thereby offering insights to decision makers for planning.

Beyond energy planning, electricity consumption levels are often correlated with economic activity and levels of development [4]. However, surveys that measure access to public services such as electricity are typically available for a small subset of the population [36]. In addition, surveys record a one-time instance of economic activity or electricity consumption, which may not reflect the overall consumption trend. Higher resolution country-wide electricity consumption predictions present an opportunity for multiple estimates of economic development, with higher spatial granularity.

2 RELATED WORK

Satellite Imagery and Machine Learning: More recently, there has been a surge of work applying CNNs to satellite imagery, measuring wealth and urbanization. Jean et al. combine overhead daytime images (aided by Nighttime Lights) to predict wealth for multiple African countries [22]. High resolution daytime images were used in training a CNN to predict nighttime lights; features extracted from the trained model were then used to estimate household expenditure and wealth at a 10 x 10 km resolution. Results from this paper suggest that predictions about economic development can be made from remote sensed imagery using features derived from that imagery; this insight provides additional motivation for developing methods that extract information for electricity consumption prediction. [20] replicate the work by Jean et al and extend the analysis to other survey indicators. They observe that while some indicators like wealth can be predicted relatively well using satellite imagery, the technique might not lend itself so easily to other socioeconomic indicators. Specifically, they show that the methodology excels when predicting access to electricity in Rwanda and Haiti, but it does not in Nigeria and Nepal. This suggests that while satellite imagery provides a powerful method for country-level coverage, it is important to quantify its predictive power and to propose complementary methodologies for specific indicators such as electricity consumption. [33] combined machine learning and satellite imagery to measure the density of rooftop types in villages. Outputs were used to prioritize villages for direct cash transfers. The methodology was also used to support microgrid site selection in India. Other applications of satellite imagery are measuring road quality [5], population density[13], detecting solar farms [21], segmenting roads and buildings[12].

Electricity consumption: Electricity consumption data is a vital input to electrification planning tools. [3, 32] present a comprehensive literature review of residential energy consumption modeling in developing regions. They found two approaches to modeling residential energy consumption: i) top-down (aggregate estimates) and ii) bottom-up (individual) techniques. Top-down approaches treat groups of buildings as a single unit of energy usage and correlate macroeconomic indicators (population, wealth, income) at national or regional levels with the aggregated consumption. Top-down approaches are appealing because they do not require individual building billing data and can rely on widely available lower resolution data. However, top-down approaches ignore the underlying variability in consumption, which is necessary for distribution planning. Bottom-up approaches combine consumption from a sample of households with a variety of techniques to predict electricity consumption. This paper falls in the latter category, with the aim of preserving heterogeneity of consumption. [34] use machine learning to predict daily electricity consumption tiers upon connecting to a microgrid, using features obtained from customer application surveys. Customer spending on electricity (pre-microgrid connection), electricity tariffs, spending on airtime and number of existing lights were some of the important predictors. [31] use overhead imagery to estimate residential building energy consumption in Gainesville, Florida, and San Diego, California. At the individual building level, they report low correlation (r^2=0) between predictions and the training data in Gainesville. However, improved performances are observed after spatially aggregating buildings (r^2 = 0.81) to 1 x 1 km. This work suggests that predicting kWh consumption at the individual level using overhead imagery is challenging, thus we formulate our task as a classification rather than a regression problem. While they show that spatial aggregation improves the r^2, it is dependent on having consumption labels for all buildings. Our dataset is a spatially representative sample

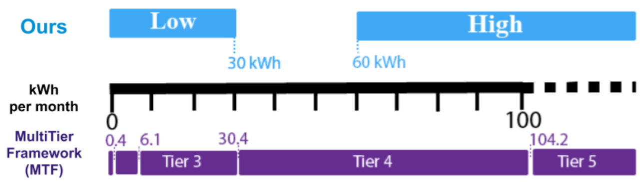

Figure 1: Illustration of World Bank Multi-Tier Framework Consumption Tiers relative to the defined levels of stable electricity consumption in our work.

of consumption labels rather than a census. Within the context of increasing electricity access, we evaluate the value of overhead daytime imagery (especially in the absence of customer surveys) in predicting levels of electricity consumption upon electrification. To the best of our knowledge, this work is the first of its kind.

3 METHODOLOGY

3.1 Electricity Consumption Dataset

Previously, we conducted a longitudinal study of 100k+ randomly sampled electrified households [15], observing that customers in Kenya typically reach a consistent level of electricity consumption roughly 12 months after receiving an electricity connection; we define this level as the *expected stable electricity consumption*. For every household, all bills after one year of connection are averaged to obtained a single stable estimate of electricity consumption. The World Bank's Multi-Tier Framework (MTF) divides electricity consumption into a series of Tiers, based on levels of electricity services. We consider low levels of consumption as corresponding to Tiers 0 - 2 of the framework while high consumption levels correspond to >= Tiers 3. Our levels of consumption (low | high) are determined by thresholding the expected stable consumption values (kWh) at <= 30 kWh and >= 60 kWh. Figure 1 illustrates our definition of levels of consumption relative to the MTF tiers, as many energy access practitioners rely on the MTF tiers to classify levels of consumption. Our disjoint boundary still enables us to include customers most suited for off-grid services (low consumption levels) and the customers that electric utilities cannot afford to miss (high consumption levels), while ensuring sufficient model accuracy. We discuss the implications of this decision in Section 4.

To obtain the dataset used in this work, we begin with monthly post-paid electricity bills from 2010 to 2015 for 135,702 randomly sampled Kenya Power customers. The Kenya Power data contains GPS locations and electricity connections dates. Please refer to our previous work [15] for a further characterization of the dataset. To develop a matched dataset of bills and images, we group customers by location to obtain electrified buildings. Next we select residential buildings that have only one customer account and these buildings are matched to contemporaneous daytime satellite imagery. Keeping in mind our goal of predicting expected levels of stable residential consumption upon electrification, satellite image acquisition dates are used to select buildings with satellite imagery acquired prior to the stable consumption phase, resulting in 42,852 buildings. We train and validate our model using a disjoint boundary, where low consumption corresponds to <= 30 kWh and high consumption corresponds to >= 60 kWh average monthly stable consumption, resulting in 27,585 building/customer pair labels.

3.2 Consumption levels from satellite images

Three band 50 cm daytime DigitalGlobe satellite imagery acquired between 2002-2016 is paired with the electricity consumption data. 74% of the 27,585 buildings had satellite images captured after 2010. Given building GPS locations, 48m x 48m image patches are retrieved, with the building of interest centered in the patch. The image patch is inputted into a CNN that outputs the probability $P(x_i)$ of a building (x_i) having a low or high level of consumption. Our custom lightweight architecture is used for electricity predictions as shown in Figure 2a), which is a modification of the DeepSense architecture[11]. With the exception of the last 2 convolutional blocks, 64 filters were used. 32 and 1 filter(s) were used in the second-to-last and last convolutional block, respectively. Parameter tuning (e.g., image size, learning rate, batch size) was performed to obtain suitable model weights. Using standard machine learning data splitting approaches, a train, validation, and held-out test data ratio of 0.75, 0.15, 0.10 was used. All results are reported on the held-out test set. Given the data split, a 60:40 class imbalance of low-consuming samples to high-consuming samples was observed. To address the imbalance, class oversampling, with image flips, 90 degree random rotations and 20% zooms were used to augment samples in the high-consuming class, such that at each training epoch the model saw an equal number of samples in each class. Equally-weighted balanced accuracies and F1-scores are reported. While accuracy and F1-score capture the overall prediction performance, they do not capture the spatial diversity of predictions made. This is of relevance when making predictions for individual buildings because varying consumption profiles often exist within the same geography. Therefore, it is desirable to have good diversity of predictions for the same spatial geography while preserving performance. A binomial standard deviation of levels of consumption for each ward is computed and the average of all wards is reported, in addition to well-known metrics. Wards represent an administrative unit in Kenya for which there are approximately 1500, with each having an average of about 30k total residents.

3.3 Combining daytime satellite images with non-visual data for multi-modal learning.

Varying lower-resolution datasets are widely available and can serve as proxies for wealth and electricity access. In this section we introduce the geospatial datasets considered and present our multi-modal approach that evaluates the value of these non-visual datasets when combined with satellite imagery. Feature standardization and normalization was applied to all 38 non-visual data features.

3.3.1 ***Geospatial data****.* *Census Information*: The 2009 Kenya census [27] (available for each ward) provides low-resolution demographic information on households at the ward administrative level. The 2009 census is selected over more recent 2019 census because the recent census data are not yet publicly available and also occur significantly after the electricity consumption data. In addition, the 2009 census better aligns with our formation for latent electricity prediction. Table 1 shows a summary of parameters obtained from the 2009 Kenya census, grouped by semantic meaning. The census reports the % of households in a ward for every category. Seventeen census indicators were extracted and used as additional

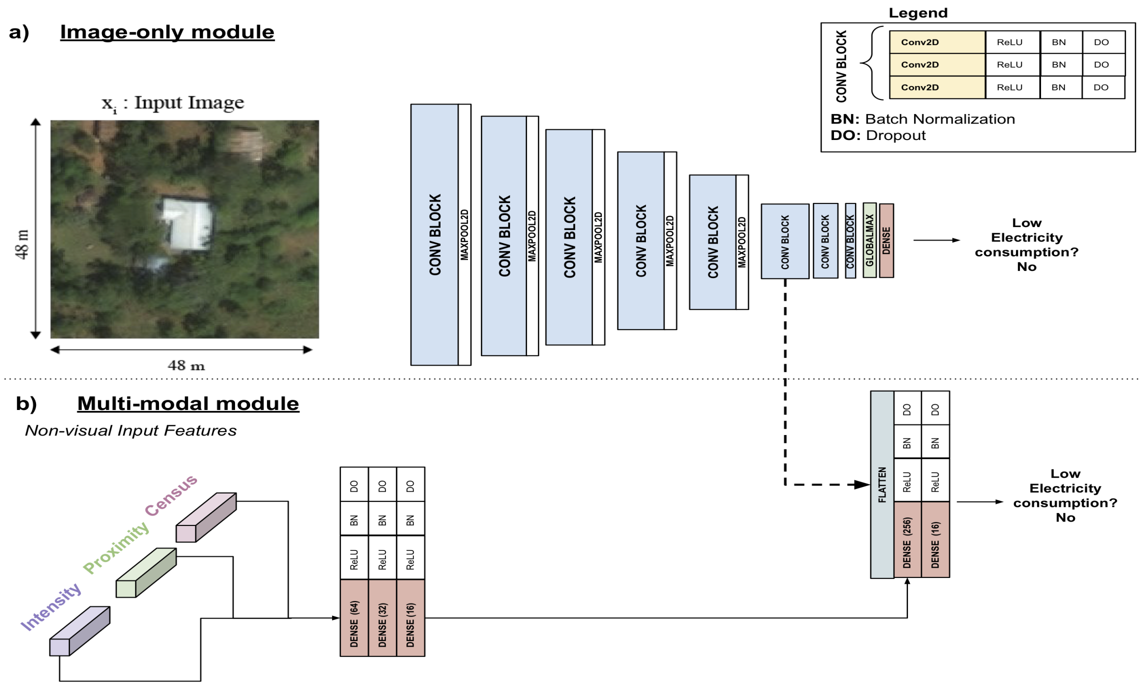

Figure 2: a) Satellite images are inputted into the Image-only module that predicts levels of electricity consumption. 64 filters are used in all but the last 2 convolutional blocks. 32 filters and 1 filter are used in the last 2 convolutional blocks, respectively. b) Multi-modal module combines features extracted from non-visual inputs with extracted features from the image-only module. During multi-modal learning, the image encoder is initialized with pretrained weights from the image-only module.

data. *Proximity metrics*: Proximity to key infrastructure such as roads, schools, markets, and cell towers may provide information on wealth and urbanization. Although some of these indicators (e.g., roads) may be present in images, proximity to key infrastructure provides human-understandable features which could be helpful for predicting levels of latent electricity consumption. Humanitarian Data Exchange (HDX) provides diverse datasets on key infrastructure in Kenya. As a result, we calculate the distance of the buildings within our dataset to the closest roads[18], schools[26], markets[10], cell towers[9], health facilities[8], and financial service points [7]. Eighteen proximity features are used (see Table 1), 11 obtained from roads by type, 3 obtained from cell towers by type, and the rest from health facilities, markets, schools, and financial service points. *Intensity*: Two intensity datasets were considered: i) 15 arcseconds/pixel (450m at the equator) VIIRS Nighttime Light data [17] and ii) 30 m High Resolution Settlement Layer (HRSL)[23] containing both population density and binary labels for presence | absence of structures for 2015. Average monthly nighttime light intensities for every year (2012 - 2015) were calculated using monthly VIIRS composites, clipped to Kenya. We retrieved the nighttime light intensity of the year the building was electrified for the grid cell in which the building is located. If the building was electrified prior to 2012, we use the 2012 intensity, as VIIRS composites are only available starting from 2012. For the HRSL layers, in addition to the intensity of the grid cell containing the building of interest, neighboring pixels (2x2) were also retrieved. We observed better performances when neighboring cells were included as features.

3.3.2 *Multi-modal learning with satellite images and non-visual data.*

We present our multi-modal module in Figure 2b) that combines images with non-visual data to predict consumption levels. Non-visual features are passed through a series a dense layers to encourage a relevant transformation of the data. Outputs from the last dense layer are concatenated with the latent image representation. The concatenated features are passed through additional dense layers prior to predicting levels of electricity consumption. Batch normalization and dropout is applied to discourage overfitting. During joint learning, the encoder is initialized with the best weights from the image-only module. Using the best image-only weights initializes the model in a suitable space. To establish baselines for comparison, predictions are made by combining the simple neural net with non-visual features.

3.4 Learning an embedding for sparse labels

Deep learning has been shown to thrive in the presence of large amounts of labels. Although our billing dataset is the largest of its kind (i.e., in a similar context) ever studied, its size remains small relative to the amounts frequently used to train data-hungry CNNs. In reality, many energy planners rely on expensive surveys to estimate electricity consumption for site selection. In this section, we investigate the value of learning on a proxy task to improve prediction performances. We hypothesize that learning a proxy task (such as building segmentation) could provide relevant image embeddings for predicting levels of electricity consumption, especially when small numbers of labels are available. We employ a dataset of

Table 1: Non-visual data used for multi-modal learning.

Census (% of ward)
Water Source (Surface | Improved | Unimproved)
Sanitation (Improved | Unimproved)
Lightfuel (Finished | Rudimentary)
Floor material (Finished | Rudimentary)
Cook fuel (Finished | Rudimentary)
Wall material (Finished | Rudimentary | Natural)
Rooftop material (Finished | Rudimentary | Natural)
Proximity
Health Facilities
Markets
Schools
Financial Services Points
Roads (Primary, Secondary, Tertiary etc)
Cell towers (GSM, UMTS, LTE)
Intensity
HRSL (Population & Settlement)
VIIRS Nighttime lights

6,928,078 building footprint polygon geometries in Uganda released by Microsoft[25]. Offsets between building polygons and our imagery dataset were observed for some parts of Uganda. As a result, we calculate the intersection between Microsoft polygons and a structures dataset containing GPS point locations for every building in Uganda to find polygons with the highest agreement. The ground-truth structures data was obtained from the Uganda Rural Electrification Agency and includes GPS information for structures identified through satellite imagery. 642,468 RGB image patches of size 128 x 128 pixels were used to train the building segmentation model for Uganda. We combine our custom encoder with a decoder for building segmentation. Skip connections between the encoder-decoder are excluded to maximize information funnelling through the bottleneck. The learnt embedding is later used in downstream consumption level prediction, to bootstrap the predictions. After training, the encoder-decoder network is initialized with the best building segmentation weights, the encoder is then extracted and merged with a classification head for predicting consumption levels.

4 RESULTS & DISCUSSION

4.1 Predictions from Satellite Imagery Alone

We discuss our balanced model performances (equally weighting each class) for the held-out test set of 2919 samples. Using daytime satellite images only as the basis for prediction, our approach achieves an accuracy of 0.64 and an F1-score of 0.64. Our custom model (trained with fewer parameters) from scratch performs comparably with well-known architectures such as VGG16[30], ResNet50[19], and Xception[6] (initialized with ImageNET weights). In addition to the accuracy and F1-score, we compared the true average ward binomial standard deviation for levels of consumption to that of the predicted levels; this metric allows observation of the heterogeneity of predictions within a ward, which is crucial for capturing the variability in electricity consumption. An average ward standard deviation of 0.43 was recorded for the true labels in the test set, while an average ward standard deviation of 0.34 was recorded for predictions made with images only. Images support diversity in predictions within the same geographies. The value of diversity in predictions becomes more apparent in Table 2, where other datasets have comparable accuracies and F1-scores but do not return predictions that reflect the spatial variability in consumption levels within a ward. Images maximize both the machine learning and domain-specific metrics of performance and heterogeneity that support distribution planning. We investigate prediction performance behavior of the image-only model by looking at the confidence scores of predictions. Figure 3 shows a barplot of predictions for the low and high class made at different prediction confidences. From the figure, it can be observed that the image-only model is better at identifying low-consuming buildings as more samples in this class are correctly predicted. This suggests that the image characteristics of lower-consuming buildings may be more similar than those of higher-consuming buildings, making them easier to identify. For the low-consuming class, no predictions in our test set were made at confidence scores greater than 0.8. At higher confidence scores, a larger % of samples are correctly predicted. This notion of uncertainty is also returned to our users to serve as an input for their decision making, giving them a sense of our model's uncertainty.

The GRAD-CAM approach by [29] is used to visualize portions of the image that are activated when making predictions. Some GRAD-CAM visualizations are shown in Figure 4. Strong activations (red) on buildings are observed when predicting high-consuming buildings. Both the building and the surrounding context are weakly considered when predicting low-consuming buildings. The image-based predictive model leverages both building size and surrounding available land as indicators of consumption levels.

We evaluate the impact of disjoint classes when defining levels of consumption and observe that disjoint classes improve model performance, especially amongst the extremes of consumption. For electricity planning, there is interest in identifying lower-consuming

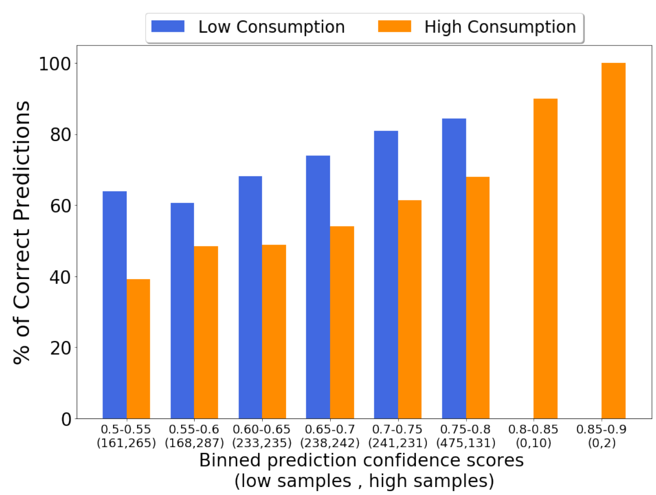

Figure 3: Prediction confidence scores showing the model uncertainty in the test data. A larger % of correct predictions are made for low-consuming buildings. At higher confidences, a larger % of samples are correctly predicted.

and higher-consuming customers as these groups represent the set suitable for offgrid systems and those the utility cannot afford to leave unconnected, respectively. In addition, disjoint classes encourage better differentiation between both levels of consumption. Note that of the 11,229 samples with stable consumption between 31 - 59 kWh, samples with stable consumption < 45 kWh tend to be classified as low, while the rest have a random chance of being assigned to either consumption level.

4.2 Predictions from Multi-modal Learning

To effectively understand the value of multi-modal learning, we present prediction results for each of the non-visual features by semantic group. Similar to the previous section, accuracy, F1-score and average ward prediction standard deviations are reported. From the 2009 census features, we observe that housing characteristics (rooftop, wall, floor material) and the energy fuel source are most indicative of levels of latent electricity consumption upon electrification. These results align with the intuition that household characteristics and wealth are proxies for latent electricity consumption. Despite the census holding relevant information, the low-resolution nature of the dataset prohibits it from providing differentiated predictions within the ward. Proximity metrics provide more unique inputs relative to their census counterparts. Specifically, proximity to roads proves to be the best performing feature, performing marginally better in accuracy and F1-score than the image-only model. However, predictions with proximity to roads features do not capture the diversity of latent consumption levels within the wards. On this front images still outperform other features. Ideally,

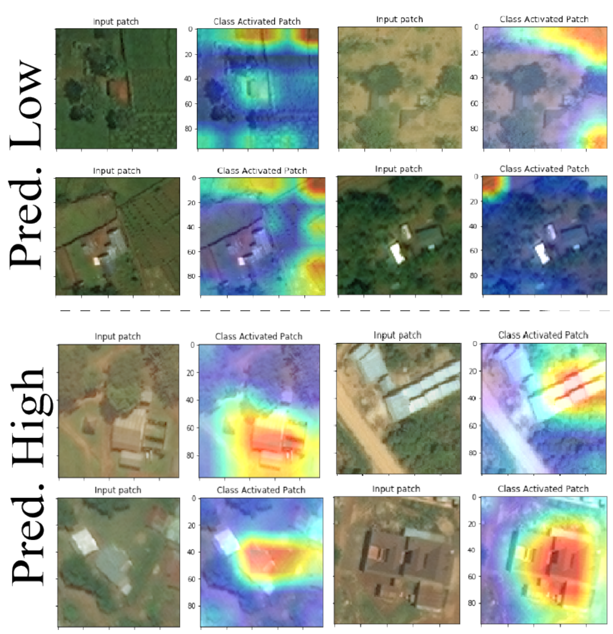

Figure 4: Gradient-based class activation maps for sample in test set. Strongest activations in Red, weakest activations in Blue. Buildings are strongly activated when predicting high levels of consumption while both the building and context are considered when predicting low levels of consumption.

Table 2: Prediction performances across varying datasets.

Type: Name (# features)	Acc.	F1-Score	Avg. ward std
C: Water Source (3)	0.60	0.60	0.00
C: Sanitation (2)	0.58	0.57	0.00
C: Light Fuel (2)	0.61	0.61	0.00
C: Floor material (2)	0.59	0.59	0.00
C: Cook Fuel (2)	0.61	0.61	0.00
C: Wall material (3)	0.61	0.60	0.00
C: Rooftop material (3)	0.62	0.61	0.00
P: Roads (11)	0.65	0.65	0.09
P: Schools (1)	0.57	0.57	0.45
P: Markets (1)	0.53	0.52	0.34
P: Cell towers (3)	0.62	0.63	0.16
P: Health Centers (1)	0.60	0.60	0.08
P: Financial Inst. (1)	0.59	0.59	0.01
I: VIIRS DNB (1)	0.55	0.55	0.19
I: HRSL (2)	0.60	0.58	0.37
All Non-visual features	*0.66*	*0.66*	*0.13*
Images	*0.64*	*0.64*	*0.34*
All	**0.66**	**0.65**	**0.31**

features that both maximize overall performance and the spatial diversity in predictions are desirable. Combining image and all non-visual features provides a small boost in performance while preserving the spatial diversity of predictions. While multi-modal learning provides a path to merge imagery with non-visual geospatial data, for the challenging task of individual building predictions, the resolution of datasets still limits the richness of the input signal needed to significantly improve prediction performance.

4.3 Performance of Learnt Embeddings

We evaluate the value of pretraining on a proxy task, especially in small data regimes. The pretrained weights learned from building segmentation in Uganda are used to predict consumption levels in Kenya. Figure 5 shows the F1-score at different data sample sizes when random and building segmentation weights are used to initialize model training. Random sample sizes were selected representing 5%, 20%, 50%, 70% and 90% of our full dataset. At each sample size increment, samples from the previous sample size are included. E.g. the 20 % dataset contains all the samples from the 5 % dataset. The same samples are used to compare the performance under random (blue) and building segmentation (orange) weights. We observe that initialization with building segmentation weights offers gains in performance especially at smaller sample sizes. The improved performances with building segmentation weights also suggests that underlying characteristics about buildings (rooftop type, color, size) provides relevant insights to levels of consumption. In addition to improved model performance, building segmentation weights decreased over-fitting to small sample sizes, thereby improving transferability to unseen samples at inference. Faster training times are also observed when the model is initialized with building segmentation weights. A mean-Intersection-Over-Union (mIOU) of 0.31 was observed, when the trained segmentation model was used to segment 1000 images of buildings from Central and Western

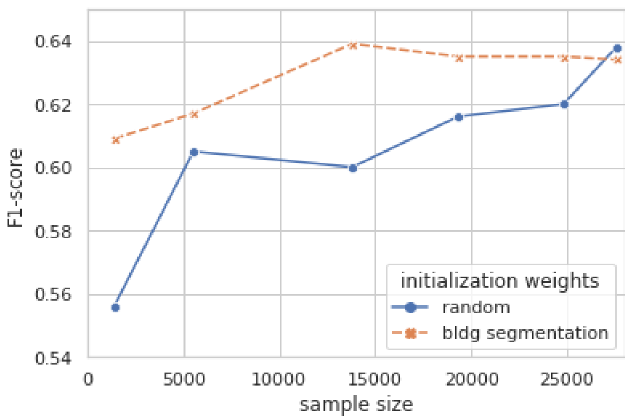

Figure 5: Comparison of prediction performance when the image-only module is initialized with random weights versus building segmentation weights. Learning about building segmentation improves performance in low-data regimes.

Kenya. This is due to the absence of skip connections and the observed offset between the Microsoft polygons and our RGB images. We hypothesize that improvements in mIOU might provide additional gains, especially in rural areas where structure segmentation is harder. Nevertheless, learning about buildings creates relevant embeddings for predicting levels of electricity consumption, especially in low-data regimes. From the figure, it remains unclear whether the maximum performance has been achieved or whether larger electricity consumption samples sizes will offer more gains.

4.4 Verification with MTF Survey

Few datasets other than those produced by the utility measure electricity consumption in Kenya. The Kenya Multi-Tier Framework (MTF) Survey – administered by the World Bank and collected independently from the utility – is a nationally representative baseline household survey of both electrified and unelectrified households[14]. The survey, conducted between 2016 - 2018, records responses for 4473 households. Specifically, the survey asks grid-connected households how much electricity they consumed in the most recent month. We compare this reported onetime consumption with our predictions for Kenya Power residential grid-connected customers. The image-only model is used to predict consumption levels of 5.3 million Kenya Power residential customers connected by the start of 2016. MTF samples are binned at 45 kWh, where < 45 kWh is considered low and the rest considered high. The 45 kWh threshold is selected to ensure proper comparison with our predictions. Figure 6 shows the correlation between MTF levels of consumption and predicted levels of consumption for Kenya Power residential customers. Correlations are reported at the county-level for counties with at least 15 MTF survey samples of grid-connected customers in order to ensure a sufficient MTF sample. Between the two measurements, we observe a sample weighted r^2 of 0.69. This correlation indicates that predictions from the image-only model are in agreement with MTF survey-reported consumption. For the county of Nairobi (the largest city in Kenya), we observe strong disagreement between the MTF survey and our predictions,

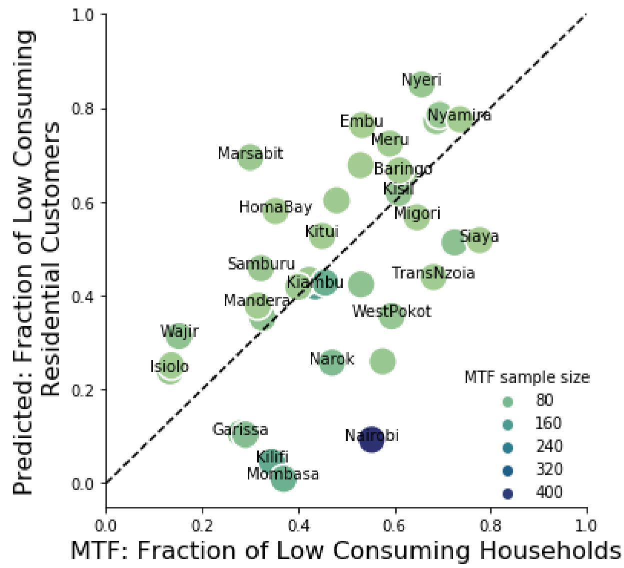

Figure 6: Correlation (r^2=0.69) between World Bank Multi-Tier Framework (MTF) survey levels of household consumption and our predicted levels for 5.3 million Kenya Power residential customer buildings, aggregated at the county level. Counties with >= 15 MTF survey samples are shown.

where MTF shows around 55 % of grid-connected survey respondents are low-consuming. However, survey collection documents indicate that oversampling was performed in Nairobi (specifically in recently-connected informal settlements), which may explain the high number of reported low-consuming respondents. When Nairobi County is excluded from the set, the r^2 increases to 0.84. This strong agreement with an independent source of national data signals the accuracy of our technique.

4.5 Country-wide predictions

We inferred consumption levels for 11.9 million buildings in Kenya using building GPS locations collected as part of the Kenya National Electrification Strategy - Structures Survey. The trained model and image patches of corresponding GPS locations are used to obtain consumption level predictions for all buildings. Statistics for each predicted level of consumption are reported in a 6-band TIF at resolutions of 250m, 500m, 1000m and 10,000m. Band 1 shows the predicted number of buildings with low levels of consumption, band 2, the mean predicted probabilities for band 1, and band 3 the standard deviation of prediction probabilities. Bands 4-6 capture similar information as the first three but are for high levels of consumption. The Kenya map in Figure 7 shows the proportion of low-consuming building predictions (aggregated at 250m) for the 11.9 million buildings. This is obtained by dividing band 1 in our generated TIF by the sum of band 1 and 4. Blue shows regions where more buildings have low levels of expected stable consumption, while red shows regions where more buildings have high levels of expected stable consumption. Our training data is from a random sample of customers nationally – there are no areas where we have exhaustive coverage. Thus, we are unable to obtain performance

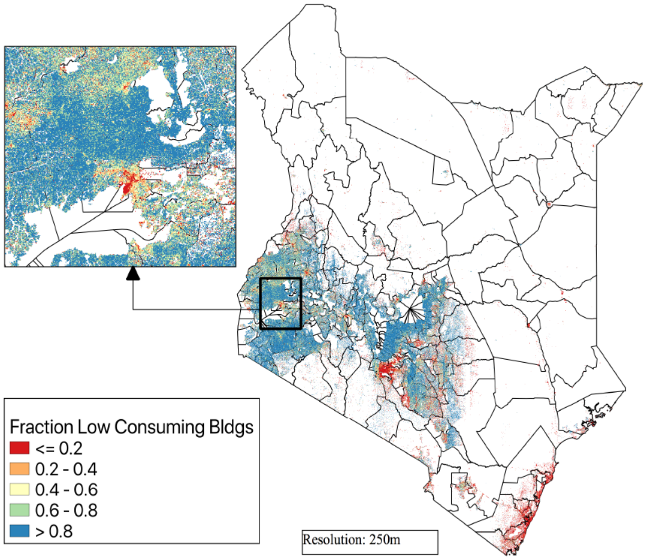

Figure 7: Novel predictions of electricity consumption levels for Kenya, aggregated at 250m. Blue shows regions with a large fraction of low-consuming buildings while Red shows regions with a large fraction of high-consuming buildings.

metrics for our aggregations (though we would expect substantial improvement in accuracies over aggregated regions, as seen in other similar work [31]).

4.6 Our API and Users

Working with a small beta user group of electricity access planning professionals, we co-developed an API to share building consumption predictions from Kenya freely with the general public. Users can access consumption predictions of a single cell or a collection of cells using point or polygon queries. Figure 8 shows a three-step process to make a request from our API. First, point or polygon geometries are input into the request. Next, we determine the grid cells in our prediction TIFs that intersect with the requested geometries and retrieve the predictions for the intersecting grid cells. Finally, a JSON response is returned containing building counts in each class and prediction confidences for every cell intersecting the requested geometries. Our API supports summarization of all intersecting cells as shown in Figure 8, where the *summary_only* attribute is **true**. Currently, only residential consumption predictions from Kenya are available with the goal of including multiple other countries and consumption categories. Since its launch in November 2020, the API has seen significant engagement from a variety of energy practitioners (NGOs, private companies, institutions, and individuals). The API has so far registered 7 active users who combined have made nearly 15,000 requests. We fully expect the engagement to grow as we make more countries and consumption categories available. A recent whitepaper [2] demonstrates an actual user's application of our data to support off-grid site selection for rural villages in Kenya. For more information on how to access and use the API, please see: https://eguide.io/#api.

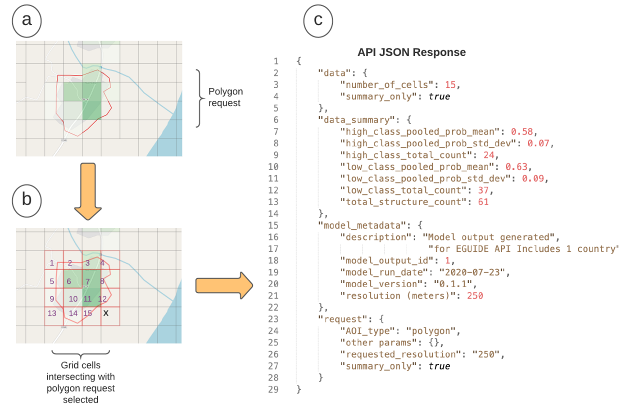

Figure 8: Sample consumption prediction API request workflow. a) a polygon or point query is made b) grid cells in our prediction TIFs that intersect with the query geometry are selected c) a JSON response is returned containing the number of buildings and prediction confidences of each class.

4.7 A note on proper applications of our work

This paper proposes a methodology to estimate anticipated levels of electricity consumption. Such an exercise – given the premise itself of estimating how much electricity a specific household will consume – is fraught with many uncertainties in the prediction, the embedded assumptions, and (im)proper applications of the results. First, this paper focuses on residential customers only with an estimate of whether they are expected to be in a low or high level of consumption if they are grid-connected. Secondly, an electrification bias may be present, as the analysis cannot and does not evaluate customers that are currently electrified with off-grid systems. Thirdly, given that our validation results show that the odds of correct predictions are roughly two out of three, electrification planners risk classifying individual households or groups of households (perhaps in some geographies/landscapes or perhaps based on roof materials/footprints) with otherwise high consumption as low – potentially leading to biased outcomes. Hence we believe that there is no substitute for individual and community agency and representation, and no substitute for utility/planner surveys. On the flip side, utilities could uniformly end up simply estimating that all new consumers are low-consuming by simply extrapolating from their recent observations. Analysis such as that presented here could be one additional input in decision-making. Utilities could improve their own predictions with the much larger and comprehensive data (e.g. bills and locations of all existing customers) that they possess. Our novel results are aimed at providing a new methodology and high-level guidance – making them suitable for site prioritization across larger landscapes, where a *human-in-the-loop* approach can be taken, to validate the true consumption (through appliance ownership and other indicators) after initial sites have been determined. We are keen to co-develop such methodologies with partners.

4.8 Future work

We plan to extend our approach to other countries to evaluate the general transferability of our method. Also, we aim to extend such work to non-residential sectors (e.g., commercial and industrial). With increasing applications of machine learning to developing policy, it is important to evaluate fairness and minimize biased outcomes in prediction models. As a result, we plan to incorporate a fairness loss that encourages geo-spatial accuracy in predictions especially in underrepresented or otherwise marginalized regions.

5 CONCLUSION

Predicting electricity usage from satellite images remains a difficult task, mainly because elements in satellite images (rooftops, roads, fields) are only proxy measures for consumption. As shown in previous work [31], while overhead imagery provides a new data-driven approach for predictions, explicit features in images are hard to correlate with implicit indicators such as electricity consumption, especially at the individual level. Still, despite undertaking a challenging problem at a far higher resolution, we achieve r^2 correlations comparable to seminal work in the machine learning-based measurement for economic development literature [22]. Overall our technique demonstrates a method to obtain levels of electricity consumption from satellite images. Our predictions provide a birds-eye view of relative levels of consumption upon electrification throughout the country and equips decision-makers with a direct measure of expected energy usage as well as a novel proxy for economic activity. This can enable better system planning and stretch ongoing investments in electrification to connect more people to modern energy sources.

REFERENCES

[1] Jason Edwin Adkins, Vijay Modi, Shaky Sherpa, Roy Han, Ayse Selin Kocaman, Naichen Zhao, Chris Natali, and Jonathan Carbajal. 2017. A geospatial framework for electrification planning in developing countries. In *2017 IEEE Global Humanitarian Technology Conference (GHTC)*. IEEE, 1–10.

[2] Village Data Analytics. 2020. Electricity Demand Estimation and Viability Analysis for Off-grid Villages in Kenya. https://www.rockefellerfoundation.org/report/electricity-demand-estimation-and-viability-analysis-for-off-grid-villages-in-kenya/, Accessed 2021-02-05.

[3] Subhes C. Bhattacharyya and Govinda R. Timilsina. 2010. Modelling energy demand of developing countries: Are the specific features adequately captured? *Energy Policy* 38, 4 (2010), 1979 – 1990. https://doi.org/10.1016/j.enpol.2009.11.079 Energy Security - Concepts and Indicators with regular papers.

[4] Paul J. Burke, David I. Stern, and Stephan B. Bruns. 2018. The Impact of Electricity on Economic Development: A Macroeconomic Perspective. *International Review of Environmental and Resource Economics* 12, 1 (2018), 85–127.

[5] Gabriel Cadamuro, Aggrey Muhebwa, and Jay Taneja. 2019. Street smarts: measuring intercity road quality using deep learning on satellite imagery. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*.

[6] François Chollet. 2017. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1251–1258.

[7] OpenStreetMap Contributors. 2021. HOTOSM Kenya Financial Services (OpenStreetMap Export). https://data.humdata.org/dataset/hotosm_ken_financial_services, Accessed 2021-01-05.

[8] OpenStreetMap Contributors. 2021. Kenya-healthsites. https://data.humdata.org/dataset/kenya-healthsites, Accessed 2021-01-05.

[9] American Red Cross. 2021. Kenya Cell Tower Locations. https://data.humdata.org/dataset/kenya-cell-tower-locations, Accessed 2021-01-05.

[10] American Red Cross. 2021. Kenya Markets. https://data.humdata.org/dataset/kenya-markets, Accessed 2021-01-05.

[11] deepsense.ai. 2020. Deep learning for satellite imagery via image segmentation. Retrieved March 6, 2020 from https://deepsense.ai/deep-learning-for-satellite-imagery-via-image-segmentation/.

[12] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. 2018. DeepGlobe 2018: A Challenge to Parse the Earth Through Satellite Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[13] Patrick Doupe, Emilie Bruzelius, James Faghmous, and Samuel G. Ruchman. [n.d.]. Equitable development through deep learning: The case of sub-national population density estimation. ([n. d.]).

[14] World Bank Energy Sector Management Assistance Program. 2018. Multi-Tier Framework for Measuring Energy Access - Kenya Household Survey (2018). https://datacatalog.worldbank.org/dataset/multi-tier-energy-access-tracking-framework-global-survey-2016-2018, Accessed 2020-08-15.

[15] Simone Fobi, Varun Deshpande, Samson Ondiek, Vijay Modi, and Jay Taneja. 2018. A longitudinal study of electricity consumption growth in Kenya. *Energy Policy* 123 (2018), 569–578.

[16] Simone Fobi, Ayse Selin Kocaman, Jay Taneja, and Vijay Modi. 2021. A scalable framework to measure the impact of spatial heterogeneity on electrification. *Energy for Sustainable Development* 60 (2021), 67–81.

[17] Earth Observation Group Payne Institute for Public Policy. [n.d.]. Version 1 VIIRS Day/Night Band Nighttime Lights. https://eogdata.mines.edu/download_dnb_composites.html, Accessed 2020-07-01.

[18] WFP Government of Kenya. [n.d.]. Kenya - Roads Network. https://data.humdata.org/dataset/kenya-roads, Accessed 2021-01-05.

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

[20] Andrew Head, Mélanie Manguin, Nhat Tran, and Joshua E Blumenstock. 2017. Can human development be measured with satellite imagery?. In *Ictd*. 8–1.

[21] Xin Hou, Biao Wang, Wanqi Hu, Lei Yin, and Haishan Wu. 2019. SolarNet: a deep learning framework to map solar power plants in China from satellite imagery. *arXiv preprint arXiv:1912.03685* (2019).

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

[23] Facebook Connectivity Lab and Center for International Earth Science Information Network CIESIN Columbia University. 2016. High Resolution Settlement Layer (HRSL). https://ciesin.columbia.edu/data/hrsl/, Accessed 2020-01-20.

[24] Stephen J Lee, Eduardo Sánchez Jacob, Andrés González García, Pedro Ciller Cutillas, Pablo Dueñas Martínez, Jay Taneja, Fernando Cuadra García, Julio Lumbreras Martín, Hannah Daly, Robert J Stoner, et al. 2019. Investigating the necessity of demand characterization and stimulation for geospatial electrification planning in developing countries. (2019).

[25] Microsoft. 2020. Uganda-Tanzania-Building-Footprints. https://github.com/microsoft/Uganda-Tanzania-Building-Footprints, Accessed 2020-12-27.

[26] Kenya Ministry of Education. 2021. Kenya Schools. https://data.humdata.org/dataset/kenya-schools, Accessed 2021-01-05.

[27] Kenya National Bureau of Statistics. 2021. The 2009 Kenya Population and Housing Census. http://statistics.knbs.or.ke/nada/index.php/catalog/55, Accessed 2021-01-21.

[28] Lily Parshall, Dana Pillai, Shashank Mohan, Aly Sanoh, and Vijay Modi. 2009. National electricity planning in settings with low pre-existing grid coverage: Development of a spatial model and case study of Kenya. *Energy Policy* 37, 6 (2009), 2395–2410.

[29] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.

[30] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[31] Artem Streltsov, Jordan M Malof, Bohao Huang, and Kyle Bradbury. 2020. Estimating residential building energy consumption using overhead imagery. *Applied Energy* 280 (2020), 116018.

[32] Lukas G. Swan and V. Ismet Ugursal. 2009. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews* 13, 8 (2009), 1819 – 1835. https://doi.org/10.1016/j.rser.2008.09.033

[33] Kush R Varshney, George H Chen, Brian Abelson, Kendall Nowocin, Vivek Sakhrani, Ling Xu, and Brian L Spatocco. 2015. Targeting villages for rural development using satellite image analysis. *Big Data* 3, 1 (2015), 41–53.

[34] Nathaniel J Williams. 2017. *Microgrid Utilities for Rural Electrification in East Africa: Challenges and Opportunities*. Ph.D. Dissertation. Carnegie Mellon University.

[35] WorldBank. 2019. More People Have Access to Electricity Than Ever Before, but World Is Falling Short of Sustainable Energy Goals. https://www.worldbank.org/en/news/press-release/2019/05/22/tracking-sdg7-the-energy-progress-report-2019, Accessed 2019-12-22.

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

[37] Alex Zvoleff, Ayse Selin Kocaman, Woonghee Tim Huh, and Vijay Modi. 2009. The impact of geography on energy infrastructure costs. *Energy Policy* 37, 10 (2009), 4066–4078.