Predicting Levels of Household Electricity Consumption in Low-Access Settings

Simone Fobi [*1], Joel Mugyenyi[2], Nathaniel J. Williams[3], Vijay Modi[1], and Jay Taneja[2]

[1]Department of Mechanical Engineering, Columbia University
[2]Department of Electrical & Computer Engineering, University of Massachusetts Amherst
[3]Golisano Institute for Sustainability, Rochester Institute of Technology
{sf2786,modi}@columbia.edu, {jmugyenyi,jtaneja}@umass, njwgis@rit.edu

Abstract

In low-income settings, the most critical piece of information for electric utilities is the anticipated consumption of a customer. Electricity consumption assessment is difficult to do in settings where a significant fraction of households do not yet have an electricity connection. In such settings the absolute levels of anticipated consumption can range from 5-100 kWh/month, leading to high variability amongst these customers. Precious resources are at stake if a significant fraction of low consumers are connected over those with higher consumption. This is the first study of it's kind in low-income settings that attempts to predict a building's consumption and not that of an aggregate administrative area. We train a Convolutional Neural Network (CNN) over pre-electrification daytime satellite imagery with a sample of utility bills from 20,000 geo-referenced electricity customers in Kenya (0.01% of Kenya's residential customers). This is made possible with a two-stage approach that uses a novel building segmentation approach to leverage much larger volumes of no-cost satellite imagery to make the most of scarce and expensive customer data. Our method shows that competitive accuracies can be achieved at the building level, addressing the challenge of consumption variability. This work shows that the building's characteristics and it's surrounding context are both important in predicting consumption levels. We also evaluate the addition of lower resolution geospatial datasets into the training process, including nighttime lights and census-derived data. The results are already helping inform site selection and distribution-level planning, through granular predictions at the level of individual structures in Kenya and there is no reason this cannot be extended to other countries.

1. Introduction

Improved engineering and new business models for electrification have contributed to increasing access to electricity around the world. However, 840 million people still lack access to electricity services [43], many of them residing in places that are difficult to reach and, as a result, expensive to serve [45, 31]. Energy providers, constrained by limited investment budgets, face a perpetual trade-off between expanding electricity access and cost recovery. When consumption levels are low, as can occur in low-income settings, utilities struggle to recover the cost of servicing a grid connection, and the government subsidies[12] for initial capital are poorly utilized. Alternatives to grid extension such as Solar Home Systems (SHS) can support smaller loads without the large wire investments, while in some cases clustered homes (with clusters far from each other) can make mini-grids viable[13]. In practice, identifying those likely to become high consumers is critical to the energy provider, as these are critical to revenue generation and system cost recovery. Given the diversity of electrification technologies, planners rely on energy access planning tools (e.g., the *Open Source Spatial Electrification Tool (OnSSET)*[27]) that utilize electricity consumption tiers, to match potential customers with technologies that can cost-effectively meet consumption. Consumption predictions can assist matching areas with cost-effective energy technologies, enabling a country to provide electricity access to a larger population given the same investment.

We make four unique contributions. *First*, we introduce a data-driven method to predict levels of future electricity consumption for individual households, using information prior to the household being electrified. Our approach trains a CNN to predict levels of household consumption using pre-electrificaion daytime satellite images. Although accurate individual household electricity consumption predictions are difficult to achieve [37], we show that

high-resolution daytime satellite imagery (0.5 m/pixel) performs better (preserving performance at different levels of consumption) than other approaches (historical consumption, census indicators, and Nighttime Lights) that result in heavily-skewed prediction performance. *Secondly*, our proposed method shows that learning about buildings through a building segmentation task and over a large volume of images improves the downstream task of electricity consumption prediction. *Thirdly*, we demonstrate a method for model interpretation that quantifies the importance of building characteristics relative to the surrounding context. Specifically, we show that building roof sizes and color are relevant to predicting consumption levels. Our approach also shows that learning about the household's surrounding context improves prediction performance between 2-5% depending on the consumption tier. *Finally*, we present additional validation of our results using the World Bank's Multi-Tier Framework survey of electricity consumption among households. Sample weighted Pearson correlation scores between the survey and our predictions for 5.3 million residential buildings were 0.82 when excluding the over-sampled and already-electrified capital city of Nairobi, and 0.64 otherwise. Outputs from our model can be used in planning tools such as OnSSET, which utilize electricity demand tiers as an input parameter for spatial electrification planning. Given the potential dependence of consumption on tariffs and policies for recovery of installation costs, the specific results of this data-driven approach apply to Kenya. However, our methods can be extended to other countries, thereby offering insights to electricity planners.

2. Related Work

Predicting Electricity consumption: [38, 3] present a comprehensive survey on residential energy consumption prediction. First, we review load forecasting in time for individual households, highlighting how the problem at hand is different. One approach [15] forecasts individual household level electricity loads 24 hours ahead using sequence mining and smart meter data. Another [41] clusters customer smart meter data into behavioral groups and later use supervised techniques such as Random Forest to predict customer clusters given unseen smart meter data. Other residential load forecasting studies [32, 20, 24] also predict the short-term future consumption of houses given their historical data or appliance usage, and deep learning techniques. Past consumption data is essential for such studies and hence not suitable for future consumption prediction where no prior data exists. Some studies have attempted to address the problem of predicting the future consumption of a currently unelectrified household. In [39] the average consumption of previously connected customers (by municipality) is used to estimate consumption for unelectrified households. This would not capture the variations amongst households. [34] use support vector regression to study the relevance of 48 household survey variables (demographics, appliance ownership, household personality traits) in predicting household consumption. [30] use an energy end-use model to estimate demand for off-grid communities in Myanmar, Indonesia and Laos through household surveys that measure appliance ownership and usage. [29] take a similar approach to estimating residential electricity consumption in Nigeria by collecting survey responses on appliance ownership and usage. [1] use machine learning to predict daily electricity consumption tiers upon connecting to a microgrid, using features obtained from customer application surveys pre-electrification. All of these studies use data that would be difficult and/or expensive to obtain at scale for a country. Our approach provides a scalable and faster approach to estimating consumption from proxy household features which are available in satellite imagery.

Satellite Imagery and Machine Learning: Recently, there has been a surge of studies applying CNNs to satellite imagery to assess building damage [35], measure road quality [4], estimate population density[8], detect solar farms [21], segment roads and buildings[7], estimate rooftop density by type [42] and measure poverty. One approach [22] predicts wealth for multiple African countries by combining overhead daytime images with CNNs. The authors use high resolution daytime images 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. Their results suggest that predictions about economic development can be made from satellite image derived features; this insight provides additional motivation for developing methods that extract information from imagery for electricity consumption prediction. Building on [22], multiple works [19, 44, 5, 40, 25, 2, 17] have assessed wealth, poverty and development using satellite images. [10] use VIIRS nighttime lights, gridded population data and land cover to estimate binary electricity access rates and electricity consumption tiers at 1 x 1 km grids. These studies demonstrate the value of satellite imagery in serving as a proxy measure for varying features such as poverty, electricity access and consumption. However, all these studies are carried out at a larger spatial scale to preclude evaluations of poverty levels or electricity consumption tiers of individual households[1]. [37] is the only study to the best of our knowledge that predicts individual building energy consumption using overhead imagery in Gainesville, Florida, and San Diego, California. While performance improves at a spatially aggregated level, at the individual building level, they report low correlation (r^2=0) between predictions and the training data in Gainesville. The context of this study is quite different given the consumption levels, the size and formal construc-

---

[1]Household consumption tiers are relevant for electrification planning

tion practices in the U.S.. Also, the problem they address is that of estimation for electrified households rather than consumption prediction for unelectrified households.

To the best of our knowledge, ours is the first study of its kind that predicts electricity consumption at an individual household level using overhead imagery. We formulate our task as a classification rather than a regression problem, and provide model interpretation around learned features.

3. Models

3.1. Problem definition

Given a set of households found in buildings $\mathbf{B} = \{b_1, b_2, b_3...b_n\}$, where each building has a corresponding satellite image prior to the household being electrified $\mathbf{X} = \{x_1, x_2, x_3...x_n\}$, the objective of our proposed model $\mathcal{F}(x_i)$ is to use each building's corresponding satellite image pre-electrification, to predict its consumption class ($\hat{y}$) after the building has been electrified (i.e. $\hat{y}_i = \mathcal{F}(x_i)$). Data from electric meters after electrification serve as ground truth values. Binary labels y_i are obtained by applying a threshold *thres* (e.g. $<=$ *thres* kWh/month) to the average monthly consumption values of each individual household given its electric meter readings.

We propose a two-phase supervised method to predict binary consumption classes ($\hat{y}_i$). First, we prioritize learning about building features through the help of a building segmentation task. Next, the building segmentation model is used to initialize a supervised model to classify consumption levels. Figure 1 illustrates the steps used for training, and their corresponding losses. Our method is compared to other models that match commonly used approaches to predict electricity consumption levels after electrification.

3.2. Electricity Prediction Models

Here we present 4 different approaches to predicting electricity consumption. The first two are widely used approaches, the third is our proposed method, and the fourth is an approach to support the interpretation of our method.

3.2.1 Model A: Average Historical Consumption

Model A represents the simplest model and serves as a baseline. This method mimics approaches commonly used by energy planners to estimate consumption for the unelectrified. Energy providers have historical consumption data for already-connected households. Thus, this model assumes that the average historical consumption of households in the same administrative unit is sufficient to approximate the consumption of customers that will be electrified in the future. In the case of Kenya, energy planning is done at administrative levels (following Kenya's policy of decentralized energy planning). For each connection year t_k, house-

holds electrified prior to t_k are used to calculate the average monthly consumption ($\bar{y}(c_j, t_k)$) of each constituency c_j. A constituency is an administrative unit and there are 290 constituencies across the country. Households who share the same constituency and electrification year are assigned the same $\bar{y}(c_j, t_k)$ value. A threshold value *thres* (e.g. $<=$ *thres* kWh/month) is applied to the assigned consumption of every household to determine the expected consumption class ($\hat{y}_i$). $\hat{y}_i$ is compared to the true consumption class y_i.

3.2.2 Model B: MLP with Non-Visual Data

Varying lower-resolution datasets are widely available and can serve as proxies for electricity access. In fact [10, 17] use non-visual data to evaluate electricity access and economic development. We present *model B*, based on publicly available non-visual datasets to evaluate their performance in predicting consumption levels upon electrification. *Model B* offers more complexity than an average historical consumption strategy. Non-visual features are inputted into a Multi-Layer Perceptron (MLP) containing 3 dense layers with 64, 32, and 16 filters respectively (Supplementary Material: A.1.). The model is trained by minimizing the Negative Log-Likelihood loss as shown in Equation 1.

$$\mathcal{L}_{task} = \sum_y -\log(p(y_i; \theta)), \qquad (1)$$

θ and y_i are the model weights and consumption labels.

3.2.3 Model C: Building Characteristics and Context

Model C combines both information about building characteristics with information about the surrounding context. In this model both the building of interest (b_i) and its surrounding context (in the form of a 128 x 128 image patch pre-electrification) are used to predict consumption levels post-electrification. Electricity consumption levels post-electrification are learnt in two phases. First we train an encoder-decoder building segmentation model to learn relevant building features. Next, we extract the trained encoder, add a classifier head and use the learnt building weights to initialize the consumption prediction task. Below we present a description of each phase as shown in Figure 1.

Phase 1: Learning about buildings: Deep learning has been shown to thrive in the presence of large amounts of labels. Although our electricity 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. We hypothesize that learning a proxy task (such as building segmentation) could provide relevant image encodings for predicting levels of electricity consumption, especially when small numbers of labels are available. We employ a much larger dataset of 6,928,078
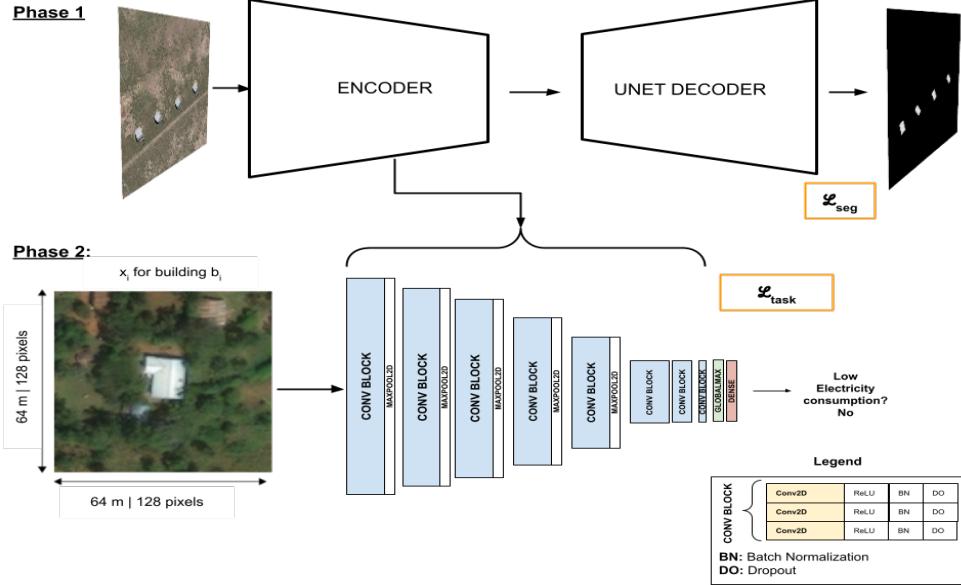
Figure 1. Phase 1: A building segmentation model is trained using the encoder in Phase 2 and a UNET decoder. The segmentation model is trained with a dissimilarity loss ($\mathcal{L}_{seg}$). Skip connections are omitted to maximize information funneling. Phase 2: The pretrained encoder is used in phase 2 to learn the electricity prediction task. An image (x_i) containing a household's building is input into the pretrained encoder. This encoder is trained with the negative log-likelihood ($\mathcal{L}_{task}$) loss to predict electricity consumption levels upon electrification.

building footprint polygon geometries in Uganda released by Microsoft[26] for building segmentation. Building polygons from Uganda are used because there is no large high quality building footprint data in Kenya, and Uganda is the closest geographic country to Kenya with building polygons.[2] Noisy (misaligned or missing) building polygons were observed within the Microsoft data in some parts of Uganda. Nevertheless, RGB patches of 128 x 128 pixels were used to train the building segmentation model in Uganda. We combine a custom encoder with a UNET-decoder to perform building segmentation (Figure 1). This encoder architecture is used both as an encoder for building segmentation and as an encoder for the classifier in phase 2. This architecture was inspired by the DeepSense architecture[6] and has been shown to be helpful in remote sensing applications such as building segmentation. Skip connections between the encoder-decoder are excluded to maximize information funnelling through the encoder during phase 2. 64 filters were used in each layer of the encoder. The building segmentation model was trained with a dissimilarity loss ($\mathcal{L}_{seg}$) as shown in Equation 2, which builds on the Jaccard index $\mathcal{J}(U,\hat{U})$.

$$\mathcal{L}_{seg} = 1 - \mathcal{J}(U, \hat{U}) = 1 - \frac{(U \cdot \hat{U}) + \epsilon}{(U + \hat{U} - U \cdot \hat{U}) + \epsilon} \quad (2)$$

where U represents the true footprints, $\hat{U}$ represents the predicted footprints and ϵ is used for numerical stability.

---

[2]This work was done prior to the release of Google Footprints [16]

The learnt encodings are later used in the downstream consumption level prediction task to bootstrap the classifier.

Phase 2: Predicting electricity consumption levels: After training the building segmentation model using Uganda data, the encoder-decoder network is initialized with the best building segmentation weights. The encoder is extracted and merged with a classification head (consisting of a global max-pooling and a dense layer) to predict consumption levels. The image patch is fed into the encoder with the classifier head, which outputs the predicted class ($\hat{y}_i$), and is trained with $\mathcal{L}_{task}$. Data augmentations (e.g vertical and horizontal image flips, 90 degree random rotations and 15% zooms) are performed during training.

3.2.4 Model D: Building Characteristics Only

The goal of *Model D* is to provide additional interpretation around the black box CNN in *Model C*. Rather than evaluate the whole image, this model aims to evaluate the importance of **only** roof characteristics of the building of interest, while ignoring the surrounding context of the household. To achieve this, *Model D* utilizes only building roof characteristics (area and type) as predictors of consumption levels. Specifically, building roof area and the RGB 3-channel intensities are extracted and used as features for prediction. Building roof area and color are inputted into the previously defined MLP to predict consumption levels (Supplementary Materials: A.1.). The MLP is also trained with $\mathcal{L}_{task}$.

Roof-top area extraction: The point indicator approach proposed by [11] is chosen over conventional segmenta-

tion because the building polygons available for segmentation (Microsoft Building Footprints in Uganda [26]) suffer from misaligned and omitted labels when compared to our satellite images. First we select only polygons that are well aligned with structures in satellite imagery. The well-aligned polygons together with the Pointer Segmentation Network[11] are used to train a segmentation model that learns when some of the instances within the images are omitted. This segmentation model was also trained with the dissimilarity loss ($\mathcal{L}_{seg}$). After training on Uganda, we also generate 1000 hand-labelled footprints in Kenya and use the small sample from Kenya to tune the Pointer Segmentation Network. Once the model is tuned to Kenya, the GPS locations of the buildings (b_i) in our dataset, are used to obtain a point within each image (x_i). This point when combined with the tuned Pointer Segmentation Network is used to extract the footprint for building b_i. The extracted footprint is then used to crop out the pixel intensities of the roof. We assume building roofs have a uniform color, thus the roof pixel mean for each channel is used in addition to the roof area (obtained from images pre-electrification) as input features to predict consumption levels after electrification.

4. Data

The dataset used in this work has 3 components: 1) Monthly post-paid electricity bills, 2) Overhead daytime satellite imagery, and 3) Public data sources. We unify these 3 data sources by matching the billing dataset to images or public data sources using customer locations within the billing dataset. Following are some details about each.

4.1. Ground truth electricity data

Previously[12], we conducted a longitudinal study of 100k+ randomly sampled electrified households, observing that median customers in Kenya typically reach a consistent level of electricity consumption roughly 12 months after receiving an electricity connection. Given this observation, we define the average monthly consumption of a household after 12 months of a connection as the *expected stable electricity consumption*. For each 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 are obtained by placing a threshold (*thres*) at 30 kWh/month. Figure 2 illustrates our levels of consumption relative to the MTF tiers. We select a 30 kWh/month boundary because it aligns with MTF break points and energy access practitioners rely on the MTF tiers to support spatial electrification planning. Rather than defining the bi-
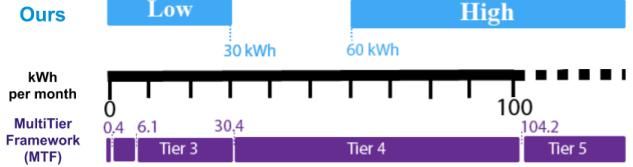

Figure 2. Illustration of World Bank Multi-Tier Framework Consumption Tiers relative to our levels of consumption .

nary class with low being $<= 30$ kWh and high being > 30 kWh, we select a discontinuous boundary where stable monthly consumptions (kWh) $<=30$ kWh are considered low while $>= 60$ kWh are considered high [3]. To develop a matched dataset of bills and images, customers are grouped by location to obtain electrified buildings. We select residential buildings with only one customer account and these buildings are matched to contemporaneous daytime satellite imagery. From 135,702 Kenya Power customers, 52,083 single customer buildings are used to calculate the monthly stable consumption of each customer. Single customer buildings are chosen over multi-customer buildings because our billing dataset does not contain all the customers in each multi-customer building. Keeping in mind our goal of predicting expected levels of consumption upon electrification ($\hat{y}_i$) using images pre-electrification (x_i), satellite image acquisition dates are used to select buildings with satellite imagery acquired prior to the stable consumption phase. Our selection assumes that the socioeconomic benefits of electrification do not become apparent within a daytime satellite image immediately (< 1 year) after the household is electrified. Labels are obtained by applying the discontinuous threshold, to obtain binary consumption levels.

4.2. Satellite Imagery

Satellite imagery used in this work consists of 3-band (RGB) 50 cm daytime DigitalGlobe Satellite Imagery obtained between 2002 and 2020. The DigitalGlobe imagery while providing country-wide coverage only contains a single image per location (there are no temporal images for the same location). To train the building segmentation task, images with corresponding building polygons were used irrespective of the image acquisition date. To predict electricity consumption levels, buildings whose images (x_i) occurred pre-electrification are selected as part of the training, validation and test datasets.

---

[3]Customers between 30-60 kWh represent a harder set to study given that we use proxy measures (building characteristics) from satellite imagery. The disjoint boundaries enable electricity planners to still identify customers with low electricity consumption ($<=30$ kWh) to target lower-cost electricity technologies and customers, while also enabling planners to target customers likely to have high electricity consumption ($>=60$ kWh) for more traditional grid-based connections. This design choice, made with significant input from electricity system planning practitioners, supports the twin goals of enhancing the financial sustainability of electricity providers and preserving model performance for the relevant task.

Table 1. Non-visual data used for electricity prediction.

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)
Intensity
VIIRS Nighttime lights

4.3. Public data sources

Census Information: The 2009 Kenya census [28] provides low-resolution demographic information on households at the ward administrative level, for which there are 1450 wards in Kenya. The 2009 census is selected over the more recent 2019 census because the recent census data are not yet publicly available and also occur significantly after our electricity consumption data. In addition, the 2009 census better aligns with our formulation for latent electricity prediction using data the occurs prior to when the household was electrified. 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 used as additional data. Customers in the same ward are assigned the ward census value. *Intensity*: 15 arcseconds/pixel (450m at the equator) VIIRS Satellite Nighttime Light data [14] is often used to study economic development and electricity. Average monthly nighttime light intensities for every year (2012 - 2015) were calculated using monthly VIIRS composites. The nighttime light intensity for the year prior to when the building was electrified is retrieved, for the grid cell in which the building is located. If the building was electrified before 2012, the 2012 intensity is used, as VIIRS composites are only available after 2011.

5. Experiments & Results

5.1. Experimental Setup

After matching satellite images pre-electrification to the mean electricity consumption of households in the stable phase, the datasets consist of 20,000 individual households. 75 % was used for training, 15 % for validation and 10 % were held-out as the test set. The distribution of the overall electricity data is preserved within each sub-group of train, val, and test. Results are reported for the 10 % in the held-out test set. All models were trained with an Adam optimizer with a learning rate of $1e^{-5}$. This learning rate was chosen over others ($1e^{-3}$, $1e^{-4}$ and $1e^{-6}$) as it offered the best overall performance and training convergence. The building segmentation model in Phase 1 was

trained for 30 epochs (as both the train and validation curve had converged). The MLP models are trained for 20 epochs and the CNN model is trained for 100 epochs. A batch size of 64 was used and 25% dropout was applied on all models to prevent overfitting. Feature standardization and normalization was performed. We used an input patch size of 128 x 128 pixels to provide enough field of view that captures the building in the centre and some context around it.

5.2. Performance Evaluation

Table 2 shows the performance of each of the four models presented in Section 3. Our evaluation metrics include: 1) Class Accuracies shown as True Negative (TN - low consumers) and True Positive (TP - high consumers) 2) Equally weighted F1-score, and 3) Area-Under-Curve (AUC).

5.2.1 (A) Average Historical Consumption

Using average historical consumption levels as predictors for yet-to-be connected customers results in a highly-skewed prediction (0.35 F1-Score), with 99% of high consumers correctly predicted while only 2% of low consumers are correctly predicted. The strong performance skew is because of the electrification bias, where high consumers (who are often wealthier) are electrified first while lower consumers are added over time. The average historical consumption will always over estimate the consumption levels of the newly connected (often lower consuming) customers. An energy planner using administrative level averages will spend large investments to connect low consumers via grid when cost-effective alternatives might be more suitable.

5.2.2 (B) Non-Visual Data

Census indicators offer a range in F1-scores (0.57 - 0.65), with the highest obtained from rooftop materials. This suggests that building characteristics are important proxy features for predicting consumption levels upon electrification. Census parameters while performing better than *Model A*, also show a performance skew towards the lower consumption class. Nighttime lights only achieved a 0.51 F1-score in predicting individual consumption levels of future electricity connections. Overall F1-scores and AUCs with non-visual is still below that obtained with images.

5.2.3 (C) Building Characteristics and Context

Using only daytime satellite images as the basis for prediction, our approach (*Model C*) achieves an balanced F1-score of 0.68 with an AUC of 0.75. This image-based model ensures good performance in both classes (70 % and 66 % correctly predicted as low and high respectively). Good performance in both classes is crucial for energy planners (especially in highly heterogeneous regions) and suggests that

Table 2. Comparison of electricity prediction models in Kenya. Area-Under-Curve (AUC) & Balanced F1-score metrics are presented. True Negative (TN) shows the fraction of low consumers that were correctly predicted while True Positive (TP) shows the fraction of high consumers that were correctly predicted.

Model	Method	Data Input	AUC	F1-score	TN	TP
A	Historical Consumption	Average kWh	NA	0.35	0.02	0.99
B	Census	i) Water Src.	0.69	0.63	0.82	0.47
		ii) Sanitation	0.62	0.57	0.63	0.51
		iii) Lighting Fuel	0.68	0.63	0.82	0.47
		iv) Floor Mat.	0.67	0.61	0.84	0.41
		v) Cooking Fuel	0.67	0.60	0.86	0.39
		vi) Wall Mat.	0.66	0.63	0.69	0.57
		vii) Rooftop Mat.	0.69	0.65	0.66	0.64
B	Nighttime Lights	VIIRS	0.52	0.51	0.77	0.30
C (Ours) Building Seg. Weights without Contrastive Loss	Building Characteristics & Context	Images	**0.75**	**0.68**	**0.70**	**0.66**
C (Ours) Building Seg. Weights with Contrastive Loss	Building Characteristics & Context	Images	0.73	0.67	0.66	0.67
D	Building Characteristics	Roof Area	0.65	0.61	0.66	0.56
		Roof Color	0.66	0.62	0.56	0.68
		Roof Area & Color	0.69	0.64	0.65	0.64
B & C	Building Characteristics & Context, Census, Nighttime Lights	Images, i-vii, VIIRS	0.77	0.70	0.76	0.65

images better support local class differentiation, compared to the other lower resolution data sources. Our CNN architecture performs comparable to well-known architectures such as VGG16[36] and ResNet50[18], even though our custom encoder only has 728k trainable parameters compared to millions in VGG-16 and ResNet-50 (Table 3).

The classifier encoder was pretrained on a building segmentation task. The value of this pretraining step is validated by 2 approaches. First, we compare classification performance with and without pretraining and noticed that pretraining the encoder on building segmentation improves the electrification classifier accuracy from 0.63 to 0.68 when all the electricity training data is used. When the training data is reduced, performance is preserved for the building segmentation-pretrained model when compared to a model trained from scratch (Supplementary Materials: A.2.). Second, we also compare the performance with and without a supervised constrastive loss[23] when initialized with building segmentation weights. Here, we hypothesize that if relevant embeddings are obtained through building segmentation, further optimizations of the embeddings (through a contrastive loss) would provide no additional performance gains. The classifier (initialized with building segmentation weights) was trained with a supervised contrastive loss (temperatures: 0.08 and 0.1) prior to finetuning the final layer for classification. Adding a contrastive loss did not further improve performance (0.67 F1-score), suggesting that the building segmentation task learnt relevant embeddings needed for the classification task.

Table 3. Performance comparison of well-known architectures compared to our encoder

	Weights	F1-score	# Parameters
VGG16	Random	0.62	14,714,688
Resnet-50	Random	0.62	23,587,712
Our Encoder	Random	0.63	728,0065

Figure 3. Sample segmentation outputs using an indicator point to specify which building(s) to segment[11]. White dots show input points given to the model to specify which buildings to segment. Green shows predictions and blue ground-truth.

Combining visual and non-visual features (*Model B & C*) through a multi-modal architecture (Supplementary Materials: A.3.), increased the F1-score to 0.70. Using multi-modal data can be helpful to improve electricity predictions.

The image model was evaluated on households with monthly consumption between 31-59 kWh. We observed a 4% decline in F1-score when a threshold of $<=30$ kWh and > 30 kWh is used for low and high, respectively. [4]

5.2.4 (D): Building Characteristics Only

Complementary to the CNN, *Model D* isolates and quantifies the relevance of building characteristics (**only rooftop area and type**) when learning to predict consumption. Prior to discussing *Model D*'s performance, we first discuss the performance of the pointer segmentation model used to obtain building footprints. **Performance of rooftop segmentation:** Hand-labelled polygons in Kenya showed a validation Intersection-Over-Union (IOU) of **0.54**. Figure 3 shows sample segmented footprints in Kenya given indicator points (white dots) specifying buildings. This segmentation model was applied on the electricity training, validation and test set to extract building footprints (roof area) and average roof pixel intensities for each channel (rooftop type).

Model D Performance: Roof area and roof color features respectively predicted 66% and 56% of the low class correctly while respectively predicting 56% and 68% of the high class correctly. However, combining both roof area and color reduced the skew in performance while encouraging better predictions for both low and high levels of consumption. This suggests that individual roof sizes may be more indicative of low consumers while roof materials (from mean pixel intensities) are helpful for better identifying high consumers. Direct use of images, which includes

---

[4]This set make up 28 % of single household customers within our data.

both the building characteristics and the surrounding context of the building improves the F1-score (relative to using only building characteristics) by 4% . This added benefit is likely a combined effect of bypassing segmentation error and the additional information found within the surrounding context of the building.

5.3. GRAD-CAM (Qualitative Interpretation):

Class activation maps are used to qualitatively explain the relevant features that the image-based classifier in *Model C* uses to predict levels of electricity consumption. GRAD-CAM [33] is used to visualize portions of the image that have high neural activations when making predictions. Some GRAD-CAM visualizations are shown in Figure 4. Strong activations (red) on buildings are observed when predicting high-consuming buildings while the activation is more distributed between the building and its surrounding context (blue) when predicting low-consuming buildings. The image-based model utilizes both building size and surrounding land as indicators of consumption levels.

5.4. Validation with independent survey data

We present an extra validation of our approach against an independently collected and nationally representative baseline household survey (4473 households) of both electrified and unelectrified households[9]. The Kenya Multi-Tier Framework (MTF) Survey conducted between 2016 - 2018, asks grid-connected households how much electric-
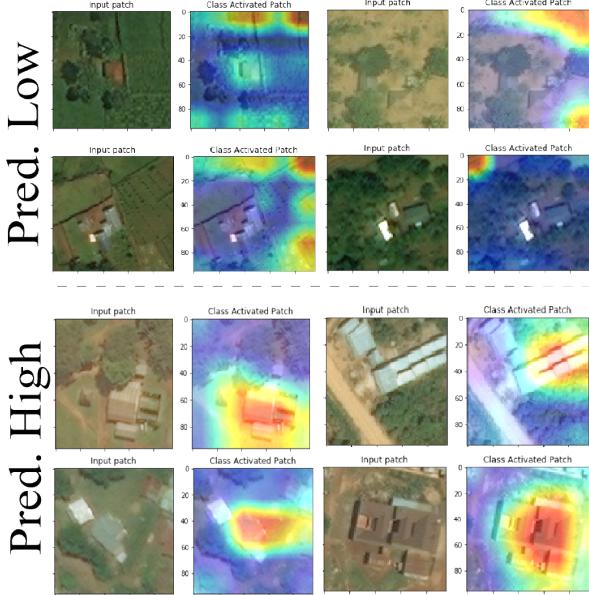

Figure 4. Gradient-based class activation maps for sample in test set. Stronger neural activations are in Red while weaker neural activations are in Blue. Buildings are strongly activated when predicting high levels of consumption while the activation is more distributed between the building and its surrounding context when predicting low levels of consumption.

Table 4. County-level consistency between independently collected Multi-Tier Framework (MTF) Survey and predictions for 5.3 million Kenya Power residential customers. Results (p-value <0.0005) for counties with at least 15 MTF samples.

	29 Counties	28 Counties (Excluding Nairobi)
Pearson correlation	0.64	0.82

ity they consumed in the most recent month. The reported and binned onetime consumptions are compared with our predictions for Kenya Power residential grid-connected customers. Images alone are used to predict consumption levels of 5.3 million Kenya Power residential customers connected by the start of 2016. MTF samples are binned at $<=30$ kWh as low and $>=60$ as high. Sample weighted Pearson correlations (Table 4) between MTF and predicted consumption levels are reported for 29 counties with at least 15 MTF survey samples of grid-connected customers. Using all 29 counties, a correlation of 0.64 is observed. Excluding Nairobi county increases, the correlation to 0.82[5]. These correlations show strong agreement (p <0.0005) given an independent source of national data.

6. Conclusion & Future work

This paper proposes a method to estimate levels of electricity consumption for unconnected households using pre-electrification images. Our results show that our novel methodology of using satellite images for electricity prediction outperforms existing approaches currently used by energy planners. We also present a multi-modal approach that combines satellite images with other data sources to further improve the overall prediction performance. The predictions of our model (currently deployed in Kenya) provide a birds-eye view of relative levels of consumption upon electrification throughout the country and equip 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 investments in electrification to connect more people to modern energy sources.

Predicting electricity usage from satellite images remains a difficult task, mainly because elements in satellite images (rooftops, roads, fields) are only proxy measures for electricity. Utilities could improve their own predictions with additional much larger and comprehensive data (e.g. bills and locations of all existing customers) that they possess. We are keen to co-develop such methodologies with partners. We also plan to evaluate the transferability of our method by extending our approach to other countries and sectors (e.g., commercial and industrial).

---

[5]Nairobi (the largest city in Kenya), is excluded because the survey over-samples recently-electrified informal settlements.

References

[1] Andrew Allee, Nathaniel J Williams, Alexander Davis, and Paulina Jaramillo. Predicting initial electricity demand in off-grid tanzanian communities using customer survey data and machine learning models. *Energy for Sustainable Development*, 62:56–66, 2021.

[2] Kumar Ayush, Burak Uzkent, Kumar Tanmay3 Marshall Burke2 David Lobell, and Stefano Ermon. Efficient poverty mapping from high resolution remote sensing images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12–20, 2021.

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

[4] Gabriel Cadamuro, Aggrey Muhebwa, and Jay Taneja. 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*, 2019.

[5] Partha Sarathi Das, Harsh Chhabra, and Sanjay Kumar Dubey. Socio Economic Analysis of India with High Resolution Satellite Imagery to Predict Poverty.

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

[7] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. 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*, June 2018.

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

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

[10] Giacomo Falchetta, Shonali Pachauri, Simon Parkinson, and Edward Byers. A high-resolution gridded dataset to assess electrification in sub-saharan africa. *Scientific data*, 6(1):1–9, 2019.

[11] Simone Fobi, Terence Conlon, Jayant Taneja, and Vijay Modi. Learning to segment from misaligned and partial labels. In *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*, pages 286–290, 2020.

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

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

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

[15] Krzysztof Gajowniczek and Tomasz Zabkowski. Electricity forecasting on the individual household level enhanced based on activity patterns. *PloS one*, 12(4):e0174098, 2017.

[16] Google. Mapping Africa's Buildings with Satellite Imagery. https://ai.googleblog.com/2021/07/mapping-africas-buildings-with.html?m=1, Accessed 2021-08-11.

[17] Sungwon Han, Donghyun Ahn, Sungwon Park, Jeasurk Yang, Susang Lee, Jihee Kim, Hyunjoo Yang, Sangyoon Park, and Meeyoung Cha. Learning to score economic development from satellite imagery. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2970–2979, 2020.

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

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

[20] Ye Hong, Yingjie Zhou, Qibin Li, Wenzheng Xu, and Xiujuan Zheng. A deep learning method for short-term residential load forecasting in smart grid. *IEEE Access*, 8:55785–55797, 2020.

[21] Xin Hou, Biao Wang, Wanqi Hu, Lei Yin, and Haishan Wu. 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. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

[23] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *arXiv preprint arXiv:2004.11362*, 2020.

[24] Weicong Kong, Zhao Yang Dong, Youwei Jia, David J Hill, Yan Xu, and Yuan Zhang. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1):841–851, 2017.

[25] Chiara Ledesma, Oshean Lee Garonita, Lorenzo Jaime Flores, Isabelle Tingzon, and Danielle Dalisay. Interpretable poverty mapping using social media data, satellite images, and geospatial information. *arXiv preprint arXiv:2011.13563*, 2020.

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

[27] Division of Energy Systems KTH Royal Institute of Technology. OnSSET - Open Source Spatial Electrification Tool,

2021.

, Accessed 2021-06-04.

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

[29] Kayode Olaniyan, Benjamin C McLellan, Seiichi Ogata, and Tetsuo Tezuka. Estimating residential electricity consumption in nigeria to support energy transitions. *Sustainability*, 10(5):1440, 2018.

[30] Andante Hadi Pandyaswargo, Mengyi Ruan, Eiei Htwe, Motoshi Hiratsuka, Alan Dwi Wibowo, Yuji Nagai, and Hiroshi Onoda. Estimating the energy demand and growth in off-grid villages: Case studies from myanmar, indonesia, and laos. *Energies*, 13(20):5313, 2020.

[31] Lily Parshall, Dana Pillai, Shashank Mohan, Aly Sanoh, and Vijay Modi. 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):2395–2410, 2009.

[32] Muhammad Sajjad, Zulfiqar Ahmad Khan, Amin Ullah, Tanveer Hussain, Waseem Ullah, Mi Young Lee, and Sung Wook Baik. A novel cnn-gru-based hybrid approach for short-term residential load forecasting. *IEEE Access*, 8:143759–143768, 2020.

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

[34] Meng Shen, Huiyao Sun, and Yujie Lu. Household electricity consumption prediction under multiple behavioural intervention strategies using support vector regression. *Energy Procedia*, 142:2734–2739, 2017.

[35] Yu Shen, Sijie Zhu, Taojiannan Yang, Chen Chen, Delu Pan, Jianyu Chen, Liang Xiao, and Qian Du. Bdanet: Multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

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

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

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

[39] Usman Hassan Syed. Estimation of un-electrified households & electricity demand for planning electrification of un-electrified areas: Using south africa as case, 2013.

[40] I. Tingzon, A. Orden, K. T. Go, S. Sy, V. Sekara, I. Weber, M. Fatehkia, M. García-Herranz, and D. Kim. MAPPING POVERTY IN THE PHILIPPINES USING MACHINE LEARNING, SATELLITE IMAGERY, AND CROWD-SOURCED GEOSPATIAL INFORMATION.

[41] Anastasia Ushakova and Slava Jankin Mikhaylov. Big data to the rescue? challenges in analysing granular household electricity consumption in the united kingdom. *Energy Research & Social Science*, 64:101428, 2020.

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

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

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

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

A. Supplementary Material

A.1. Multi-Layer Perception (MLP) Architecture

Figure 5 shows the MLP architecture used to train *Model B* (Non-visual Data) and *Model D* (Building Characteristics Only). The MLP consists of 3 dense layers with 64, 32, and 16 filters respectively, all with ReLU activations. The last dense layer consists of a softmax activation. 25 % dropout was applied to minimize overfitting.

A.2. Performance of building segmentation

Additional evaluation of the building segmentation task is done by observing how the classifier performs at varying training data sample sizes. Figure 6 shows the F1-score at different training data sample sizes when random subsets of the data are selected and either random weights or building segmentation weights are used to initialize model training. 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. Initializing with building segmentation weights offers performance gains especially at smaller sample sizes. The improved performances with building segmentation weights suggests that underlying characteristics about buildings (rooftop type, color, size) provides relevant features for consumption prediction. This is inline with our initial findings that building characteristics are relevant in predicting consumption levels. In addition to improved model performance, building segmentation weights make the classifier less susceptible to label quality. Specifically when random weights are used for initialization, it is observed that the randomly selected sub-sample at 60 % of the full dataset, performed the best and performance dropped as more samples were added. This suggests that the ease | difficulty of the sub-sample significantly affects performance. Building segmentation weights initializes the model in a suitable learning space and has a regularizing effect even as harder labels may be introduced, allowing only additional useful information to be extracted. Obtaining large amounts of useful samples to appropriately predict consumption of yet to be connected customers can be challenging. For energy practitioner looking to apply our
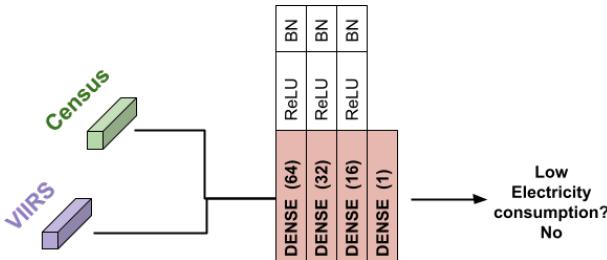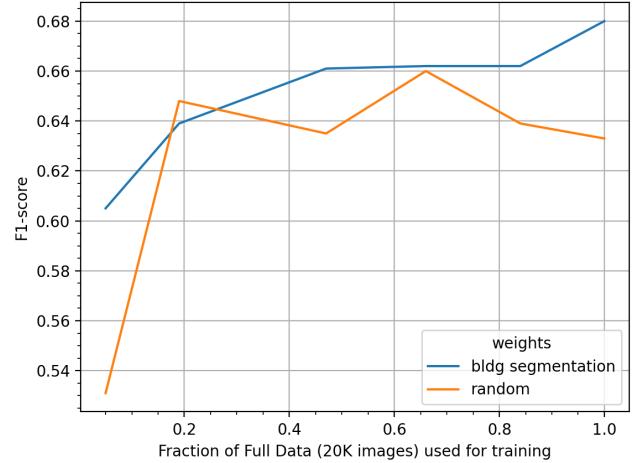

Figure 6. Comparison of prediction performance when the classifier is initialized with random weights versus building segmentation weights. Learning about building segmentation improves performance in low-data regimes and makes performance less susceptible to harder labels thereby offering a regularizing effect.

approach, we show that learning about buildings from using a segmentation task, provides useful weight tuning needed for appropriate prediction of consumption tiers.

A.3. Multi-modal architecture: Encoder and MLP

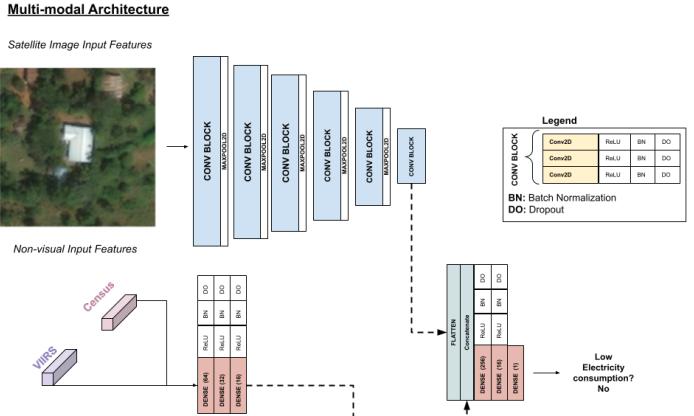Figure 7 shows the multimodal architecture used to combine satellite images with public data sources.

Figure 7. Multi-modal architecture combining the CNN image-based encoder with an MLP to predict consumption levels using visual images and non-visual public data sources.

Figure 5. MLP architecture used to train *Model B* and *Model D*