PowerScour: Tracking Electrified Settlements Using Satellite Data

Santiago Correa
UMass Amherst, USA
scorreacardo@umass.edu

Zeal Shah
UMass Amherst, USA
zshah@umass.edu

Yuezi Wu
Columbia University, USA
yw3054@columbia.edu

Simon Kohlhase
Village Data Analytics, Germany
simon@villagedata.io

Philippe Raisin
Village Data Analytics, Germany
philippe@villagedata.io

Nabin Raj Gaihre
Village Data Analytics, Germany
nabin.gaihre@villagedata.io

Vijay Modi
Columbia University, USA
modi@columbia.edu

Jay Taneja
UMass Amherst, USA
jtaneja@umass.edu

ABSTRACT

Access to electricity is crucial for poverty reduction and economic growth. However, almost 759 million people still do not have access to electricity. More than 90% are located in the global South, where low-income countries struggle to provide clean, reliable, and affordable energy sources to ameliorate the basic living standards. Even though there are many opportunities to provide basic electrification in these settings, the lack of reliable and updated information about electrification has become one of the main challenges for policymakers and developers to better plan grid extensions and prioritize communities with higher needs. The increasing availability of remote sensing data has created opportunities to obtain information about electricity access at a larger and quicker scale. Using ground truth data of 57k distribution transformer locations from Kenya, we present a processing pipeline to validate and compare state-of-the-art techniques that use very high resolution (VHR) daytime imagery (50cm Digital Globe) or low-resolution nightlight (NTL) imagery (450m VIIRS-DNB) to identify electricity access. Further, we propose a supervised-learning approach called `PowerScour` that outperforms three techniques from the commercial, scientific and public fields. By assessing the trade-offs between temporal and spatial resolution and comparing population and settlement patterns, we find that `PowerScour` improves the F1-score of existing techniques by up to ≈27% in deep rural areas. In Kenya, our model correctly identified ≈73.4% of places with and without access to electricity between 2013 and 2017. All data processing and modeling scripts are available at https://github.com/santiagocorrea/PowerScour.

CCS CONCEPTS

• **Computing methodologies** → **Supervised learning by classification**.

BuildSys '22, November 9–10, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9890-9/22/11…$15.00
https://doi.org/10.1145/3563357.3564069

ACM Reference Format:
Santiago Correa, Zeal Shah, Yuezi Wu, Simon Kohlhase, Philippe Raisin, Nabin Raj Gaihre, Vijay Modi, and Jay Taneja. 2022. PowerScour: Tracking Electrified Settlements Using Satellite Data. In *The 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '22), November 9–10, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3563357.3564069

1 INTRODUCTION

According to the United Nation's Sustainable Development Goal 7 (SDG7), providing universal access to affordable, reliable, and sustainable electricity is the cornerstone to addressing major challenges in sustainable and social development and reducing poverty in developing countries. Today, approximately 759 million people lack access to electricity. In Sub-Saharan Africa alone, 570 million people still lack access, accounting for three-quarters of people without electricity [20]. Moreover, the Covid-19 crisis has exacerbated the problem by delaying or preventing many electrification projects. Particularly in poorly electrified countries, the lack of access stems from weak regulatory frameworks and insufficient electrification plans [1]. Achieving SDG7 by 2030 requires significant efforts to track progress in this field. It is a crucial challenge to enable policymakers, national and international organizations, and commercial entities to facilitate the adoption of renewable technologies, expand and upgrade existing infrastructure, and prioritize resources to regions with higher needs.

Traditionally, electricity access information is compiled from national/regional household surveys/censuses. These methods are expensive and inefficient for frequent data collection from zones with difficult terrains and remote areas, which limits the spatio-temporal resolution of surveys/census and ultimately hinders the assessment of progress on electrification. To overcome these issues, researchers are leveraging the growing corpus of remote sensing data by developing side-channel measurement techniques such as Unmanned Aerial Vehicles [26, 27, 31, 38], high-resolution daytime satellite images, and night-time light satellite data. Remote sensing data is facilitating electricity planning and hence supporting the tracking of SDG7. For example, daytime satellite images are commonly used to detect the presence of solar panels using data-driven models [4, 6, 21, 24]. Crowdsourcing mechanisms through smartphones and open data have been used to infer distribution grid

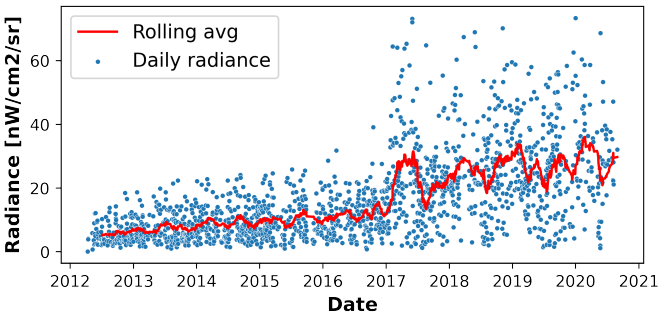

Figure 1: Radiance profile over time for a single pixel in Kenya from daily measurement of the VIIRS-DNB sensor. The red curve is a monthly rolling average.

topologies [30]. However, crowdsourcing requires significant computing power for extensive coverage and relies on locally-available labeled training data which may be scarce in developing regions.

Night-time lights (NTL) data is a widely used open-source dataset due to its global coverage, timely availability, and consistency [11]. It has been used as a proxy for multiple applications such as urban area mapping [37], population and economic activity estimations[17], well-being and conflict assessment [16, 23, 36], and disaster monitoring [25, 32]. Furthermore, researchers have leveraged this capability of detecting artificial lighting at night to study electrification and grid reliability [10, 28, 32, 33]. Currently, the daily NTL dataset provided by the Suomi NPP Visible Infrared Imaging Radiometer Suite (VIIRS) is the most popular NTL dataset. Daily NTL measurements are noisy as shown in Figure 1 and so the research community has preferred using monthly/annual composites instead. However, in this work we show that daily NTL can help detect electrification in settlements with dim and irregular lighting levels.

In this study, we address these challenges using ground-truth data from Kenya – a dataset of 57,000+ geo-located transformers provided by the local utility company. We develop a supervised learning model called `PowerScour` and evaluate its performance relative to current state-of-the-art techniques for the task of identifying electricity access. The compared systems include `gridfinder`[2], `gridlight`, the High-Resolution Energy Access dataset (`HREA`) [29] and the Gridded Dataset for Electrification in Sub-Saharan Africa (GDESSA) [12]. We discuss the trade-offs between existing techniques and our model, and compare the detection across time and in areas with a wide-range of population dynamics. Validation results for the Kenya dataset show that the model exceeds the performance of state-of-the-art techniques by up to 27%. As an exploration of a potential future avenue for improvement, we provide a performance comparison of a technique that employs high-resolution daytime imagery for estimating electricity access, and discuss the trade-offs of using low-resolution (NTL) alongwith high-resolution daytime imagery. Through this work, we aim to characterize the limits of measuring electrification using remote sensing data to support universal electrification.

2 RELATED WORK

In recent years, researchers and commercial entities have identified several approaches to predict electricity access using remote sensing data. In the following we will outline the most successful approaches thus far.

Gridfinder and Gridlight: The most notable model thus far has been `gridfinder`. Based on work by Rohrer at Facebook [15], Arderne et. al. (2020) developed `gridfinder`, an open source tool to predict the extent of the powergrid globally. Relying on NTL imagery and OpenStreetMap´s high-voltage grid and roads data, the model is able to predict the global power grid with a reported accuracy of 75%. The technique uses monthly NTL composites to identify regions with consistent illumination. Gridfinder applies a 2D convolutional filter to the temporally stacked NTL composites to extract pixels that are brighter than their surroundings, enabling it to pick up dimly lit areas. A threshold is then applied to obtain a binary raster of electrification targets. Further, the areas with zero population are filtered out from the electrification targets. Assuming that all identified electrification targets are grid-connected, a medium-voltage network is generated using Dijkstra's shortest path algorithm to connect all of the points [2]. `Gridfinder` has been applied in various energy access planning projects, but has a crucial shortcoming in that it is a one-time estimation rather than a timeseries; this limits its utility for tracking changes in electricity access. One variation that addresses this shortcoming is produced by *Village Data Analytics (VIDA)* which has developed `gridlight`, a modified version of `gridfinder` that uses time-series NTL data. The targets and medium-voltage grid estimates generated by `gridlight` serve to identify unelectrified settlements (typically defined as being at least 2.5-5km away from the grid). `Gridlight` has reported a balanced average accuracy of 72% in three regions (Nassarawa (82%), Ondo state (72%) in Nigeria, and Mozambique(61%)). Additionally, `gridlight` uses a tunable threshold (0.1) during the convolution to filter out noise background that can be adjusted to increase sensitivity. A lower threshold increases accuracy by detecting dim lights from small settlements, but increases the number of false positive targets. `Gridlight` has been used in more than 14 countries, mainly in Sub-Saharan Africa. The data is part of the VIDA analysis to support investors and policymakers in developing national electrification strategies and selecting mini-grid deployment sites. For this work, we used `gridlight` electrification targets from 2014 to 2017 for Kenya, which are discussed in Section 4.

High Resolution Energy Access (`HREA`): Another approach is the `HREA` method by Brian Min and Zachary O'Keeffe at the University of Michigan [29]. `HREA` uses daily NTL imagery to generate likelihood estimates for electricity access over all populated areas within a country. First, luminosity detected over areas without any settlements trains a statistical model for background noise and exogenous factors, such as lunar illumination or land cover to yield an estimate for baseline illumination for each day and land cover. Next, the detected illumination is compared to the baseline illumination on every pixel. Settlement pixels with significantly higher illumination than the baseline are assumed to have electricity access on that night. By aggregating over all nights throughout a year, an artificial light score is computed to determine the overall likelihood of an area to be electrified. With this method, it is possible to identify

	Gridfinder	HREA	GDESSA
Source	[2]	[29]	[12]
Processing Steps	1. threshold 2. transient filter 3. cost modeling	1. statistical modeling of background noise	1. noise correction 2. urban & rural areas identification
Input Data	- monthly NTL - GHS population data - ESA Land Cover - NASA Elevation Model - OSM roads & electricity - GHS settlement data	- daily NTL - Facebook population density maps - GADM country boundaries	- monthly NTL - Landscan gridded population - MODIS Land Cover
Output Data	- LV network (vector data) - MV network (vector data)	- likelihood estimates for electrification (450m x 450m raster)	- access rates & - consumption tiers (1km x 1km raster)
Characteristics	- high precision - distribution network	- daily results	- non-binary results - access quality

Table 1: Comparison of existing satellite-based technologies to predict electricity access.

regions where electricity access is uncertain. In contrast to traditional binary classifiers, HREA provides an indicator regarding the quality and reliability of power supply for regions with middling scores [29]. The results of HREA are openly available and hosted by the World Bank in the *Light Every Night AWS Data Archive* [3].

Gridded Dataset for Electrification in Sub-Saharan Africa (GDESSA): A third approach to measure electricity access with remote sensing data was presented by Falchetta et. al. of the International Institute for Applied Systems Analysis (IIASA). The GDESSA model combines yearly NTL data, MODIS land cover classification, and LandScan population data to estimate electricity access rates and consumption levels in populated regions within Sub-Saharan Africa. To identify electricity access, median radiance for each pixel and each year is computed and compared to a lower-bound noise floor, determined from by-definition zero-radiance pixels - for example from large water bodies. Pixels above the noise floor are considered electrified, pixels below are unelectrified. Further, each pixel is classified as rural or urban based on land cover type and population density. Non-zero radiance pixels are then assigned to one of eight consumption tiers (four rural and four urban) according to the World Bank Multi-Tier framework based on the distribution of radiance quartile values [12].

Table 1 provides an overview of all three models and compares them with regards to input, output, processing steps and distinct characteristics. A detailed, quantitative validation of these models against ground-truth data in Kenya is presented in Section 4.

3 METHODOLOGY

To investigate the limits of remote sensing data to measure electricity access, we developed a machine learning-based model called PowerScour and compare its performance with current state-of-the-art techniques. We aim to understand the trade-offs between different spatial characteristics, performance over time, and opportunities that arise when other sources of remote sensing data are used. In the following subsection we explain the datasets used and the processing pipeline of our implementation.

3.1 Datasets

To explore the performance of existing NTL-based methods, we use the following datasets as inputs for our framework:

Daily NTL data: The first initiative to collect low-light imagery of the earth goes back to the mid-1960s with sensors onboard the Defense Meteorological Satellite Program (DMSP) platforms [5]. Since then, two main products have been developed: the DMSP Operational Linescan System (DMSP-OLS) and the Visible Infrared Imaging Radiometer Suite (VIIRS) day/night band (DNB) on the Suomi National Polar-orbiting Partnership (S-NPP) satellite mission. Data from the DMSP-OLS sensor were used to create annual composites of stable light produced by different sources of human activity such as agricultural fires, city lights, fishing boats, and gas flares from 1994 to 2017. However, its large spatial resolution (30 arc-seconds or ≈1km at the equator), low radiometric resolution (6-bit data, values range from 0-63), and saturation on urban cores pose a significant challenge to analyze electrification at a usable temporal and spatial granularity [7]. These challenges limit the ability to draw conclusions from time-series variations in low-density urban areas since it may overlook small or dimly-lit areas.

Currently, the best available source of global NTL data is collected by the VIIRS-DNB sensor [11]. The DNB data are collected each day (during the night) with monthly and annual composites also published from 2012 to the present-day. These data provide radiance from surface lighting in $nW/cm^2/sr$ units with a spatial resolution of 15 arc-seconds (≈450 meters at the equator). This sensor overcomes the limitations of the OLS sensor and are typically used as cloud-free composites on a monthly basis. Recently, the World Bank in collaboration with the National Oceanic and Atmospheric Administration (NOAA) and the University of Michigan released Light Every Night, a publicly-available data repository of raw daily data collected from the two aforementioned sensors over the last three decades [3]. Given the advantages of the DNB data over the OLS sensor, in this work we use daily VIIRS-DNB data and perform our analysis in Kenya, where we have significant ground-truth data to validate existing techniques and evaluate our proposed method. Daily NTL offers consistent measurements across administrative borders and, in comparison to aggregated annual and monthly, mitigates the under representation of NTL signals in deep rural areas. However, it is extremely noisy, which requires additional pre-processing steps before being used. We discuss these shortcomings in Section 3.2.

Distribution transformers and minigrids in Kenya: Our ground-truth data of electrification is composed of the geographic location of distribution transformers and minigrids in Kenya. Transformer locations were provided from the national power utility. This dataset includes latitude and longitude, date of commissioning, and power capacity in kVA units for more than $57k$ transformers. Dates of commissioning span from 1966 to 2017, which facilitates the analysis of electrification between 2014 and 2017 using NTL data. We use this time period since it overlaps with the daily VIIRS-DNB data that are used to train our model and validate the performance of existing techniques.

Minigrid locations were obtained from a national report in [22]. It includes information from 21 minigrids such as installed capacity (kW), number of connections by June 2016, and date of commissioning. Unfortunately, this report does not contain geographic coordinates so we use the geo-location of the settlement where the minigrids belong to.

Population counts: Gridded population data provides a consistent and comparable data format that is useful in scenarios where the aggregation of population estimates is required for different spatial or administrative units. This capability facilitates the analysis and integration of diverse spatial datasets and enables evaluation

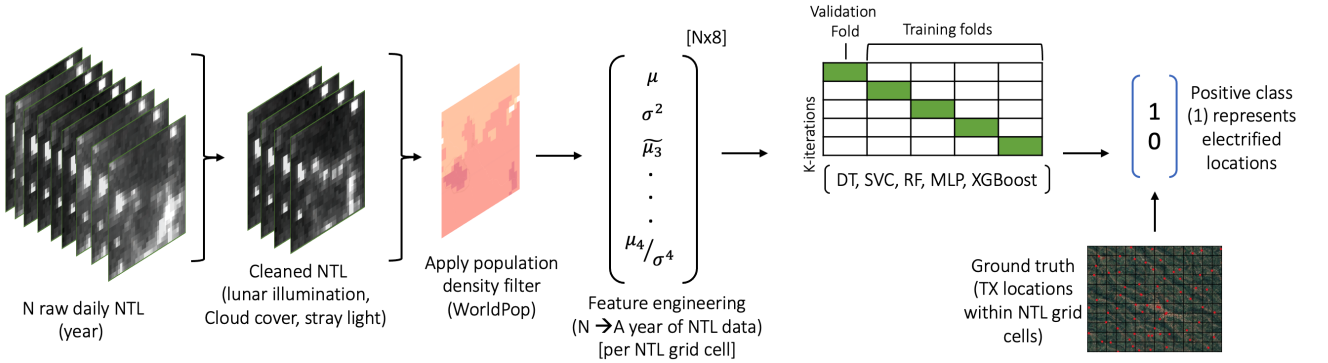

Figure 2: PowerScour: Data pipeline to estimate access to electricity using nighttime light data (NTL) and supervised learning algorithms. The input data are raw daily NTL composites following a preprocessing step to clean and prepare the data to avoid examples affected by cloud cover, stray light and lunar illumination effects. Ground-truth data was obtained from existing locations of distribution transformers (TX). Summary statistics of each pixel constitute the feature space used for the binary classification task. Hyperparameter tuning was implemented using k-fold cross-validation.

of the impact of population densities at a sub-national level. In this work, we use gridded population estimates for two reasons: as a feature for our electricity access estimation model, and as a mechanism to filter out uninhabited regions where electricity access is not needed, hence reducing the computational complexity of our assessment. A widely used and publicly available source of gridded population estimates is WorldPop [35], which provides population counts from 2000 to 2020 at a resolution of 3 and 30 arc-seconds (100m and 1km at the equator, respectively) globally. WorldPop estimates population using a variety of models that leverage census data and a stack of covariates.

3.2 Processing Framework

We propose a supervised learning model and designed a data processing pipeline as illustrated in Figure 2. Our problem is defined as a binary classification task where places that are electrified and non-electrified are labeled as 1 and 0 respectively. We construct these labels for training our model based on the presence or absence of distribution transformers and minigrids within NTL pixels. Using our ground truth in Kenya, If there is at least one transformer that intersects with the area of an NTL pixel ($450x450m$), it is classified as 1; if the pixel does not contain any electricity infrastructure, then it is labeled as 0. Based on this problem definition, we estimate electrification at an NTL pixel level so the input for our model is a feature vector built on summary statistics of the radiance and the population in each pixel, and the output indicates if that area ($450x450m$) is electrified.

Given our problem definition, we complete the following steps according to our data processing pipeline in figure 2.

Clean Data: Raw NTL data is noisy as we have shown in figure 1 and it cannot be directly employed in our approach. As opposed to monthly and annual composites, daily data suffers from background noise, solar and lunar contamination, data degradation due to stray light, cloud cover, and events unrelated to electric lighting [10].

Background noise refers to the residual radiance in areas where surface lighting is extremely low and are part of areas without settlements such as water bodies and dense forests. Solar and lunar illumination affects NTL data when their lighting reflects on Earth's surface so measurements taken when the sun is well in the horizon (solar zenith $> 101°$) and outside the full moon phase are preferred. The DNB sensor provides solar zenith angles for each pixel which facilitates the filtering of the desired signals. Stray light degradation occurs when the sensor is collecting data from the Earth's surface while is hit by sunlight as the sun is under the horizon. Cloud cover impacts the radiance of lighting by obscuring and scattering the measurements and events such as wildfires, biomass burning and gas flares add noise to electricity-based lighting [9].

The Light Every Night dataset provides quality bitflags that indicate if some of the aforementioned issues were corrected in each pixel. For example, cloud cover index varies between 1–5, where 1 indicates very cloudy and 5 cloudless. For our analysis we only use the pixels with $index > 4$. In [8], the authors present the detailed steps and algorithms necessary to produce clean and global VIIRS NTL, which were incorporated in our dataset.

In addition to these corrections, we filtered out all the uninhabited regions where estimating electrification would be futile. Filtering out those regions as a preprocessing step reduces the computational complexity for building our training set and accelerates the inference time. We select inhabited regions in Kenya using the WorldPop dataset presented in Section 3.1. For each NTL pixel, we compute the population count by aggregating the WorldPop geospatial raster dataset based on the vector geometry of each pixel. Note that the aggregation is required since the spatial resolution provided in the WorldPop dataset is $100m$.

Feature Engineering: An important advantage of working with full-resolution temporal profiles is the opportunity to work with additional indices besides the mean [10]. After filtering and merging with population data, we leverage the variable nature of daily profiles by creating feature vectors using 4 statistical moments and a score of the daily profiles: mean, variance, skew, kurtosis and $95th$ percentile. Kurtosis provides information about the presence of outlier and the tail heaviness of a distribution. High kurtosis indicates that the radiance is more stable since the values are more concentrated in the central peak of the probability distribution. Low

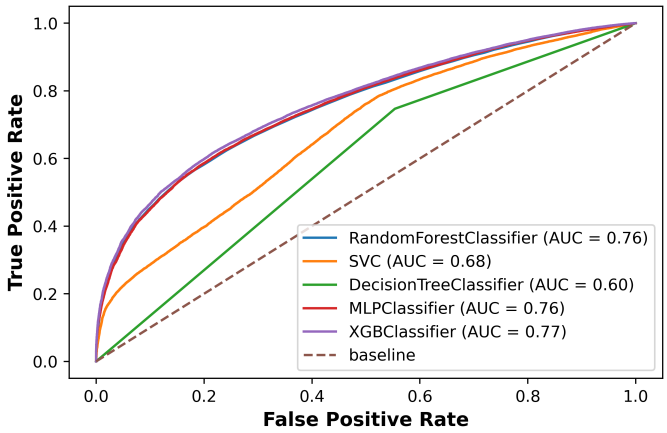

Figure 3: ROC curve for different learning algorithms. XGBoosting shows slightly better performance than random forest and multi-layer perceptron classifiers.

kurtosis indicates more fluctuation within the grid cell since the distribution have heavy tails and the pixels can exhibit low and high radiance levels at different nights. Skew measures the asymmetry from the mean of the data distribution. Positive skew shows that most of radiance levels are in the lower end of the brightness spectrum so the mean value is larger than the median, and a negative skew indicates the opposite. These indices are obtained for each NTL pixel which are the examples in the training and held-out sets. These statistical characterization of daily data captures changes in brightness from different types of human settlements and mitigate the impact of noisy examples, which allow PowerScour to identify places that are likely to be electrified.

Training and Held-out Sets: Our dataset is composed of more than $225k$ examples that represent NTL grid cells excluding uninhabited areas. We split our dataset into training (70%) and test (30%) sets, and use 5-fold cross-validation on the training set to simulate a validation set and perform hyperparameter tuning as we illustrate in algorithm 1.

Algorithm 1 5-Fold Cross-validation for Hyperparameter Tuning

Require: Dataset D, Model M with set of hyperparameters H
  Randomly split D into a training set T_r and test set T_e
  Randomly split T_r into a set of 5 folds $F_1, ..., F_5$
  for Each cross-validation fold $k = 1, ..., 5$ **do**
    Let $V = F_k$ and $L = T_r - F_k$
    Learn M_{ik} on L for choice of H_i
    Computer performance metric P_{ik} on V
  end for
  Select H_* $s.t.$ max $\frac{1}{5}\sum_{k=1}^{5} P_{ik}$ on V
  Evaluate performance of M_* on T_e

Hyperparameter Tuning: Hyperparameters are a key component during the model training process since they control bias-variance and precision-recall tradeoffs, and in some cases, the speed of training. These hyperparameters are not optimized or "learnt" by the classifier during training but given as part of the algorithm setup. Each supervised learning algorithm has a unique set of hyperparameters that are usually tuned using different techniques such as grid search, random search, or coarse-to-fine search (a combination of grid and random search). In this work, we use grid search in combination with 5-fold cross-validation as described in algorithm 1. Table 2 presents the commonly used hyperparameters for each learning technique and their respective optimal value. Random Grid search is the simplest technique that is used when the set of hyperparameters is relatively small. Next, we explain how we selected which traditional learning algorithm to use.

Learning Algorithm Selection: For our supervised binary classification task, we aim to find a method that given a set of example pairs $D = \{(x_i, y_i), i = 1 : N\}$ where $x_i \in \mathbb{R}^D$ is a feature vector and $y_i \in \mathcal{Y}$ is a binary class label, is able to learn a function $f : \mathbb{R}^D \rightarrow \mathcal{Y}$ that accurately predicts the class label y for any feature vector x. There is a wide variety of learning classifiers each with different properties such as interpretability, training and prediction speed, decision boundary, and so on. We short-listed traditional classifiers based on different principles such as kernel-based (Support Vector Classifier (SVC)), shallow learning (Decision Trees (DT), and Random Forest (RF)), deep learning (Multi-layer Perceptron Classifier (MLP)) and ensembles (XGradient Boosting (XGB)). Table 2 summarizes the main characteristics of each learning algorithm. SVC is particularly robust when different classes are linearly separable (decision boundary is linear); however, sparse feature vectors can significantly degrade its performance.

Recently, interpretability has become a key component in ML to verify predictions, identify flaws and biases, learn about the problem and ensure compliance to legislation. Tree-based methods offer an abstraction of feature importance that provide a global interpretation of the model behavior which make them very useful to comply with the aforementioned characteristics. Moreover, these types of algorithms are generally fast during the training and inference steps, which make them attractive for production environments that handle real-time predictions. But, ensemble algorithms are known to provide more predictive power for complex tasks. Artificial Neural Networks (ANNs) are widely used due to their great predictive power and ability to represent non-linear relations between the target and the input examples. However, ANNs tend to overfit when trained using limited (small) data and their interpretability is considerably more challenging.

After training and tuning each type of learning algorithm we assess the performance of each classifier using the receiver operating characteristic (ROC) curve. The output of our binary classification task belongs to one of four classes: true positive (TP), places that were predicted electrified are actually electrified; true negative (TN), an example is correctly predicted as non-electrified; false positive (FP), a pixel is predicted as electrified but is actually non-electrified; and false negative (FN), a pixel that is incorrectly predicted as non-electrified. ROC curves combine true positive rate (TPR) and false positive rate (FPR) of the classifiers which are defined as $TP/TP + FN$ and $FP/FP + TN$ respectively. Learning models with higher area under the ROC curve (AUC) are typically better classifiers. A perfect model has an AUC of 1 and if a classifier obtains an AUC below 0.5, it means that the model is not better than a random classifier. Figure 3 illustrates the performance of five traditional learning models. The Decision Trees classifier performs poorly with an AUC of only 0.6. In contrast, MLP, random

Model	Principle	Properties	Hyperparameters	AUC
SVC	kernel-based	↓ capacity for large feature vectors ↓ interpretability ↓ train and prediction speed	kernel type (RBF) C (1), gamma (10)	0.68
DT	shallow learning	↑ interpretability ↑ train and prediction speed ↓ predictive power	criterion (entropy) max. depth (3), max. features	0.60
RF	ensemble	↑ capacity for large feature vectors ↑ train and prediction speed ↓ interpretability	num. estimators (100) max. depth (3), max. features (4)	0.76
XGB	ensemble	↑ capacity for large feature vectors ↑ train and prediction speed ↓ interpretability	num. estimators (1000) max. depth (10), subsample (0.7) learning rate (0.01)	0.77
MLP	deep learning	↑ capacity for large feature vectors ↓ train and prediction speed ↓ interpretability	num. hidden layers (3) optimizer(Adam) learning rate (0.001)	0.76

Table 2: Comparing principles, properties, hyperparameters, and AUC for conventional supervised learning methods.

forest and XGBoost show the best performance with an AUC of 0.76-0.77. Since XGB classifier shows the best AUC, we choose it as the learning algorithm to conduct our study.

4 EVALUATION

In this section, we present a performance analysis of existing techniques and PowerScour based on three different conditions: detection of electrified sites over time, performance based on different population densities and settlement patterns, and comparison of detection using very high resolution satellite imagery. As it was presented in Section 2, HREA, GDESSAA, and `gridlight` do not explicitly provide their estimations as a binary class format nor use the same spatial resolution. We transformed these approaches to enable a fair comparison between all of them. For each method, we aimed to obtain an electrification estimate for each NTL grid cell that overlaps with inhabited regions as explained in Section 3.2.

Even though the HREA dataset uses NTL pixels for the estimation, the outcome of their model has a higher spatial resolution linked to a settlement layer ($\approx 30m$) and the outcome is given as a likelihood electrified estimate. To match our binary measurements of access, we apply a threshold of 0.5 to each HREA cell. Pixels with a likelihood greater than the threshold are classified as electrified (1), otherwise they are classified as unelectrified (0). The output of this step is a binary raster file that is used to match the spatial resolution of the VIIRS data. For each NTL pixel, we compute the number of HREA pixels that are located within each NTL cell. If there is at least one HREA electrified pixel in the NTL polygon, then the NTL pixel is set to have electricity access.

`Gridlight` and GDESSA provide a binary access layer as an intermediate step. `Gridlight` exposes rasters of electrification targets that refer to places with consistent illumination. Similarly, GDESSA first computes the electrified areas and then adds a population layer to estimate the number of people electrified at different tier levels. We query these binary rasters to obtain their value at the centroid location of each NTL cell. Given that both use NTL data with consistent spatial resolution, no further processing is required.

To evaluate the performance of existing techniques, we use the following standard metrics to assess the performance of the binary classifier: accuracy $((TP + TN)/(TP + TN + FP + FN))$, precision$(TP/(TP + FP))$, recall $(TP/(TP + FN))$ and F1-score $(2 \cdot TP/(2 \cdot TP + FP + FN))$. Accuracy represents the ratio between the correctly classified and total number of classified examples. It is commonly used when the errors in each class are equally important. However, this metric is susceptible to class imbalance and has to be used with care. Precision is the ratio of correct electrified predictions to the total number of examples estimated as electrified. Recall, also known as true positive rate, quantifies the ratio of correct electrified predictions to the overall number of electrified examples in the dataset. F1-score is the harmonic mean between precision and recall. It decreases with an increase of false positives or false negatives examples. This metric provides better interpretability and a balance between precision and recall. Each of these metrics above gain value based on the prediction goal and ultimate use. If the goal is effectively detecting locations for electrification, then the model needs to reduce the number of false negatives. If it is assigning electricity grid auditors, one needs to avoid false positives.

4.1 Performance over time

One important feature of remote sensing techniques is the ability to measure changes over time. For the application of tracking the progress of electricity access, this is especially valuable as it allows continuous performance monitoring of investments and the reallocation of resources as necessary. Remote sensing data can also serve as an independent "check" on stakeholders who may be politically incentivized to over-report or under-report electrification. All the techniques presented in this study provide annual electrification estimates that can be linked to our ground-truth data from 2013 to 2017, except for GDESSA which is available from 2014 onward. For each year, we built a ground-truth dataset as explained in Section 3.2 using the commissioning date for the transformers and minigrid dataset in Kenya. For our learning model, we train only with examples from 2017 (70% for training and 30% as a held-out set for testing in the year) and evaluate with the examples in previous years. Since each year has independent NTL features and population densities, data leakage is prevented and the class proportions are maintained using stratified sampling.

Table 3 summarizes the performance of each approach across different years and illustrates the class proportion for each year which shows the electrification increase year by year in this NTL sample. Across the years of analysis, `gridlight` outperforms all remaining techniques in terms of precision; however, it performs poorly for recall. As we discussed in Section 2, `gridlight` finds electrified areas based on their consistent illumination. Since only monthly composites are used in this technique, places with high variability or dim light such as small settlements or villages are mostly undetected as electrified. High precision indicates that the number of false positives is very low. It means that `gridlight` is very certain when a place is identified as electrified, specifically in the case of large settlements of cities, but classifies as non-electrified places that are actually electrified due to the aforementioned shortcoming. This characteristic of being cautious about labelling settlements as electrified and aiming for high precision is a key requirement for `gridlight` when being used to support policymakers. Similarly, GDESSA shows a low recall and a competitively high precision. It applies additional land cover layers that facilitate the identification of settlements regardless of their size. However, this technique uses yearly composites of NTL which may underestimate places with radiance close to the noise floor, which could be the reason why the recall is low as in `gridlight`.

	2013				2014				2015				2016				2017			
Metric	**GL**	**HREA**	**GDESSA**	**PS**	**GL**	**HREA**	**GDESSA**	**PS**	**GL**	**HREA**	**GDESSA**	**PS**	**GL**	**HREA**	**GDESSA**	**PS**	**GL**	**HREA**	**GDESSA**	**PS**
Accuracy	0.566	**0.704**	n.a.	0.647	0.605	**0.693**	0.543	0.659	0.530	**0.676**	0.555	0.663	0.497	**0.683**	0.496	0.681	0.632	0.695	0.624	**0.735**
Precision	**0.970**	0.865	n.a.	0.620	**0.970**	0.867	0.889	0.647	**0.974**	0.872	0.879	0.658	**0.974**	0.870	0.897	0.740	**0.973**	0.854	0.858	0.748
Recall	0.268	0.515	n.a.	**0.840**	0.231	0.520	0.190	**0.830**	0.210	0.517	0.270	**0.869**	0.220	0.585	0.227	**0.761**	0.354	0.626	0.486	**0.876**
F1-score	0.321	0.645	n.a.	**0.714**	0.374	0.650	0.313	**0.727**	0.346	0.649	0.413	**0.749**	0.359	0.699	0.362	**0.751**	0.519	0.722	0.621	**0.807**
Class prop.	**elect.(52.3%), non-elect.(47.8%)**				**elect.(55%), non-elect.(45%)**				**elect.(58%), non-elect.(42%)**				**elect.(63%), non-elect.(37%)**				**elect.(63%), non-elect.(37%)**			

Table 3: Comparison of performance over time for existing NTL-based techniques. Best performance across different techniques is highlighted in bold. `gridlight` shows better precision results for all the years. `PowerScour(PS)` outperforms `gridlight`(GL) and `HREA` in recall and F1-score.

Both HREA and PowerScour consistently show the best balanced performance among the techniques. HREA generally shows significantly better accuracy than `gridlight` and GDESSA, and marginally better than PowerScour approach except in 2017 where our learning algorithm outperforms HREA by ≈ 4%. For recall, however, PowerScour notably reduces the number of false-negative examples across all the years and shows an increase of ≈ 35% in comparison to HREA in 2015. The recall is an important metric since we are mainly interested in reducing the number of false negatives; in other words, we primarily aim to find unelectrified places with high confidence due to its importance for measuring and reaching universal access to electricity. On the other hand, PowerScour shows low precision in comparison to the other approach indicating that our model is not as confident detecting true electrified places as the other models are. However, the precision is not substantially low to indicate degradation in the model performance. Moreover, F1-score, which provides a balance between precision and recall, is measurably higher for our model for each year of the analysis. This represents an improvement on the state-of-the-art techniques available for this problem, enabling better tracking of electricity access over time.

4.2 Population and Settlement Pattern

Another dimension of our analysis is observing performance based on population densities and settlement characteristics which can affect the ability of sensors to detect nighttime light signals. Moreover, settlement categories such as rural, peri-urban, and urban areas can implicitly demonstrate the challenges to identify access to electrification in a given region using NTL-based methods.

We obtained the settlement classification from previous work [13] which employs population density, land use classification, and NTL data to compute clusters of the aforementioned categories in a raster format. We merge this classification with our NTL grid cells adding an attribute of urban class. Then we subdivide each settlement category into thirds (Q1, Q2, Q3) based on population counts of each NTL pixel and observe the performance. In this setting, the dominant settlement class is rural which accounts for ≈ 96% of the total NTL grid cells. Urban and peri-urban areas represent the remaining ≈ 4% which refers to major cities and settlements in the country. Figure 4 illustrates the F1-score for each model at different population and urban organization. We observe improvements in performance for all the techniques as we move from rural places with low population densities in Q1 to highly populated urban areas. Also, most of the techniques show an F1-score close to one when transitioning from rural to peri-urban, with the exception of `gridlight` which reaches its maximum score at the urban category. On the other hand, the lowest population third in rural areas shows the lowest performance across all the models with less than 60% in F1-score. This behavior clearly demonstrates the limitations to understand electricity access in places with lower population density using NTL data. Rural areas with little or scarce street lighting do not emit enough light to be easily detectable by the existing models. However, it is worth noting that HREA and our model, which are the only two methods that use daily measurements of NTL, significantly outperform `gridlight` and GDESSA in this category. Moreover, our model shows better performance than HREA across the different population characteristics, likely as a result of our statistically-informed feature engineering approach.

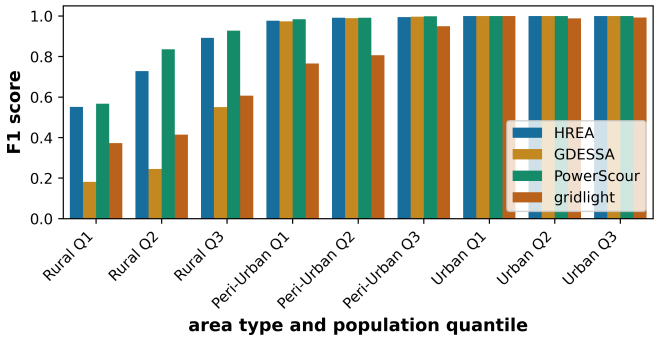

Figure 4: Comparing performance across different areas and levels of population density. Each type of area is subdivided into thirds (Q1,Q2,Q3) based on population count. Population increases from left to right.

However, the aforementioned settlement categories are not uniform when we evaluate different settings. A typical scenario that highlights the shortcomings is: assuming a rural settlement pattern of households largely dependent on agriculture, where each household may have a regular grid farm evenly spread out over an administrative area. This type of settlement needs long wires with multiple electric infrastructures in order to electrify the whole area through grid extension. However, another settlement pattern with the same number of households located in a nucleated settlement surrounded by farmland may need much less wires and a fraction of infrastructure. Both settlements have the same overall population density measured at the administrative area scale, and might both be classified as rural administratively.

Define the former kind of settlement pattern as non-nucleated rural and the latter as nucleated rural. Even if both settlements have full access to electricity by grid extensions, the latter pattern will most certainly have a distribution infrastructure located within the tight settlement area. Moreover, from the NTL perspective, the nucleated rural is more likely to be visible, as opposed to the more

diffuse non-nucleated rural. These different settlement patterns would become one of the reasons that the model performs worse in the vast and diverse rural areas. This highlights the importance of considering the settlement pattern as an additional metric in the validation of the model.

In previous work [14], the authors developed a settlement pattern metric used to better estimate the power infrastructure for distribution system planning. Given the building structure locations in Kenya, this model outputs the deployment strategy of medium and low voltage lines and transformers to reach electrification for each ward in Kenya with the least infrastructure costs. Because a single transformer has a maximum length limit to connect structures, more structures connected to a transformer means a higher nucleation level. In other words, every hypothetical transformer represents a cluster of structures with different settlement patterns.

We apply the hypothetical transformer as a metric of settlement pattern, by merging to each NTL grid cell according to the location, and investigate if this metric impacts our learning model performance. For each transformer that intersects with an NTL pixel, we average its number of structures per transformer (str/TX) as a new feature, so that each NTL not only contains summary statistics across daily NTL measurements but also a new predictor that represents the settlement pattern. We re-train our model taking into account the new attribute and only evaluated in areas classified as rural. We illustrate the confusion matrices of the two models (with and without str/TX) in Figure 5. Not every NTL pixel intersects with hypothetical transformer locations, so the feature is set to zero for those particular cases. We can observe that including the settlement pattern indicator slightly improves the accuracy of the model reducing the number of false positive examples (top right corner). The reason might be that the new feature helps the model recognize that sparse rural wards are prone to lacking access. However, the improvement is marginal and computing the new attribute is computationally intensive, which likely renders it superfluous.

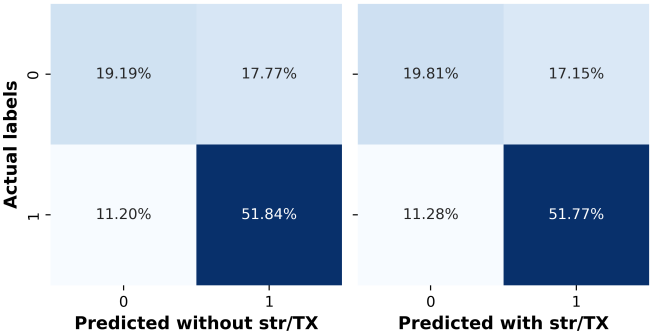

Figure 5: Confusion matrices for rural areas when the average number of structures per hypothetical transformer (str/TX) is added as a feature in our learning model.

4.3 Low Versus High Resolution Satellite Data

Even though NTL data has been widely used to identify economic activity and electricity access due to its global coverage and temporal resolution, recent advancements in imaging technology and remote sensing have made it possible to also acquire daytime images at high resolution (30-50cm) which facilitates the analysis of infrastructure development at a higher granularity. Unfortunately, these images are often only available from commercial providers at significant expense and with infrequent temporal frequency, particularly in rural developing regions. In this section, we aim to evaluate if the use of high resolution visible (red, green, and blue bands) imagery would improve the electrification measurements relative to the low resolution of NTL data (450m). This analysis can give an indication of how much room for improvement there can be in accurately estimating electricity access.

For this study, we compare the performance of PowerScour model trained using NTL data with a CNN model trained on DigitalGlobe's (DG) high resolution daytime satellite images [18] in Kenya, for the task of detecting electrification. This DG dataset consists of approximately 7000 images of size $10km \times 10km$, with each image having a spatial resolution of $50cm$. For ease of training a CNN, large DG images were divided into smaller image tiles of size $250 \times 250m$ (or 500×500 pixels). Even though these images were captured ranges from 2010 to 2017, it is important to note that each region of Kenya was only captured once and therefore the daytime image tiles do not have a temporal component since they do not overlap. Similarly as in our NTL data preprocessing step, image tiles were labelled as "electrified" or "non-electrified" based on presence or absence of electrified structures in each image tile during the corresponding year of capture. the geo-location of electrified structures across Kenya was recorded as a part of the Kenya National Electrification Strategy (KNES) [19].

The entire dataset was grouped using a standard 70-20-10 split for training, validation and testing, ensuring similar distributions of buildings per image. Splitting was done such that model would get evaluated on images from different counties and with varied electrified structure densities. Furthermore, we created an additional test dataset called our comparison test set. We manually appended the comparison test set with all the daytime image tiles whose centroids belonged to the held-out NTL images. All the image tiles added to the comparison test set were removed from training, validation, and testing sets to ensure there was no leakage of information. The comparison test dataset was completely hidden from the model during training and evaluation process. This new test set was created to allow for consistent performance comparison of the two models – the NTL-based PowerScour model and the daytime image-based CNN – in the same held-out regions of Kenya.

A VGG11 CNN model [34] pre-trained on the ImageNet dataset was used for the task of classifying daytime image tiles into electrified or not electrified. The VGG11 network was modified to handle image tiles of size $500\times 500px$ and to output binary results. The entire network was trained end-to-end using a batch size of 16 images, learning rate of 1e-6 and training time equal to 50 epochs. After training and evaluation of the CNN model, we made the model predict electrification status of each image tile in the comparison test dataset. Since multiple daytime image tiles in the comparison set belong to one larger held-out NTL image, we labelled the held-out NTL image as "electrified" if at least one corresponding image tile was predicted as "electrified". Finally, for each held-out NTL image, we also obtained labels predicted using a CNN trained on high resolution daytime satellite images.

	2014		2015		2016		2017	
Metric	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Accuracy	**0.849**	0.816	0.675	**0.705**	0.624	**0.678**	0.693	**0.709**
Precision	**0.691**	0.642	0.664	**0.735**	0.608	**0.692**	0.711	**0.770**
Recall	**0.898**	0.882	**0.885**	0.767	**0.917**	0.765	**0.873**	0.8
F1-score	**0.781**	0.743	**0.759**	0.751	**0.731**	0.726	0.784	**0.785**

Table 4: Comparing performance over time for PowerScour with $450m$ resolution NTL data (Low) and a CNN using $50cm$ resolution RGB imagery (High).

Table 4 summarizes the comparison between the CNN-based model and PowerScour approach across the 2014-2017 timespan. For the test set in 2014 we can observe that the high resolution tool outperforms the NTL-based approach; however, as we move towards more recent years the differences become increasingly marginal and in 2017 we see a better performance of the NTL model across most of the metrics but recall. The NTL-based model shows better accuracy and precision in three of the four years of our analysis with an average difference in performance of $\approx 0.3\%$ and $\approx 6\%$ respectively. In contrast, the CNN-model outperforms the NTL model for recall during the same years. The balance between precision and recall can be observed in the F1-score which has an absolute difference of only $\approx 0.4\%$ between the two models. Even though the CNN-based model is able to provide higher granularity, the performance difference with the NTL-based model is marginal which may indicate that we are hitting an upper bound in performance to detect access to electricity using the current remote sensing tools. Thankfully, this strong performance can be achieved with a free, globally-available, and frequently-collected dataset, rather than an expensive and infrequently-updated dataset, positioning the NTL-based PowerScour method as a strong candidate for continuous tracking of electricity access.

5 DISCUSSION AND FUTURE WORK

We have shown the dynamics of electricity access detection across time, population, settlement patterns, and types of remote sensing data. Traditionally, measuring access to electricity in high-income economies is unnecessary due to universal electricity access. By contrast, emerging economies are dynamic, requiring additional data sources such as NTL data that are also changing and correlate with human activity. However, detection of electrified settlements in developing regions using the current state-of-the-art techniques still struggles to track electrification in deep rural areas as we have observed in Figures 4. Even though we have provided an analysis that combines spatial dynamics from two different perspectives (population and settlement patterns), we consistently observe that sparse rural areas are the most difficult to identify electrification.

Besides the inherent difficulty detecting low night light from DNB sensors, one possible reason why some techniques underperform in these settings is the underlying NTL data source used. For example, `gridlight` and `GDESSA` use monthly and annual composites which are heavily aggregated and filtered, discarding possible insights in the variability of the signals as illustrated in Figure 1. By contrast, we have seen that `HREA` and our PowerScour learning model, which use daily nightlight measurements, improve over the aforementioned techniques in rural areas by at least $\approx 20\%$. This difference in performance highlights the additional information present in the noisy daily data and how the results for `gridlight` and `GDESSA` could be improved if daily NTL data were incorporated to control their respective filtering processes.

We have also explored the performance of using high-resolution daytime satellite imagery versus our NTL-based PowerScour model. As shown in Table 4, the improvement in performance in detecting electricity access from high-cost, high-resolution images is marginal, indicating that our NTL-based technique may be approaching a regime of diminishing returns. Nonetheless, daytime images of $50cm$ resolution can potentially identify more characteristics of the environment and critical infrastructures such as transmission towers, power plants, and solar PV arrays. This type of imagery also offers the possibility to assess household conditions such as roof type and size, distance to primary roads, and more, which can be correlated to access to electricity. However, this type of approach requires more computing power, and it is more difficult to scale globally since few satellite imagery providers offer affordable global coverage. These shortcomings are reflected in the high cost of computing infrastructure and data acquisition, posing a challenge for developing countries and policymakers with budget constraints.

As more substantial ground-truth data become available in developing settings, we plan to investigate the implications for transferability of our learning model and what volume and distribution of local data are needed to train a generalizable model. For this study, we concentrate our analysis in Kenya where we have access to detailed ground truth. However, to deploy this model widely, it is required to analyze the levels of domain shift across regions to assess what data are needed to effectively fine-tune `PowerScour`.

Furthermore, we have seen that our model can do better than current state-of-the-art NTL-based techniques; however, we believe that adding more sources of publicly-available data such as geographic information about land cover, roads, and building footprints should be further investigated. These additional datasets can potentially be predictors of electrification and are extremely useful in rural areas where NTL data alone are not sufficient.

6 CONCLUSION

Tracking access to electricity in developing regions is an essential step in the quest to achieve universal access to clean, affordable, and reliable electricity for all. Fortunately, technological advances in remote sensing and side-channel techniques and the ubiquity of open-source data have enabled the development of data-driven tools to monitor progress towards achieving universal access more frequently and at a fraction of the cost of traditional census.

In this study, we analyzed three of the most prominent techniques to detect access to electricity using NTL data and proposed a learning model called PowerScour that leverages daily NTL data. Using substantial ground truth from the national utility company in Kenya, we compare the performance of each approach across time, population, and settlement patterns. As expected, we found that identifying electricity access from NTL data in rural areas with scarce population densities poses a difficult task due to the low levels of brightness that characterize these settings. However, we observe that techniques using more granular daily NTL data are able to improve the detection over those which only use monthly

and annual composites by at least ≈ 20%. Moreover, our PowerScour learning model outperforms the existing methods, especially in areas with low population density.

Further, we also find that the performance of our PowerScour approach compares favorably with that of a CNN with high-resolution daytime satellite imagery, an approach that employs expensive imagery and substantially more compute-intensive processing. This indicates that estimating electricity access using remote sensing techniques with only low-resolution data, may have limited improvement potential. We are keen to contribute our technique towards the governments, investors, and other organizations striving for universal electricity access.

REFERENCES

[1] International Energy Agency(IEA). 2021. *World Energy Outlook 2021.* OECD. https://doi.org/10.1787/14fcb638-en

[2] C. Arderne, C. Zorn, C. Nicolas, and E. E. Koks. 2020. Predictive mapping of the global power system using open data. *Scientific Data* 7, 1 (jan 2020). https://doi.org/10.1038/s41597-019-0347-4

[3] World Bank. 2020. *Light Every Night.* "https://worldbank.github.io/OpenNightLights/wb-light-every-night-readme.html"

[4] Kyle Bradbury, Raghav Saboo, Timothy L. Johnson, Jordan M. Malof, Arjun Devarajan, Wuming Zhang, Leslie M. Collins, and Richard G. Newell. 2016. Distributed solar photovoltaic array location and extent dataset for remote sensing object identification. *Scientific Data* 3, 1 (2016). https://doi.org/ScientificData

[5] Thomas A. Croft. [n. d.]. The brightness of lights on Earth at night, digitally recorded by DMSP satellite. , 66 pages. https://doi.org/10.3133/ofr80167

[6] Julian de Hoog, Stefan Maetschke, Peter Ilfrich, and Ramachandra Rao Kolluri. 2020. Using Satellite and Aerial Imagery for Identification of Solar PV: State of the Art and Research Opportunities. In *Proceedings of the Eleventh ACM International Conference on Future Energy Systems* (Virtual Event, Australia) *(e-Energy '20)*. Association for Computing Machinery, New York, NY, USA, 308–313. https://doi.org/10.1145/3396851.3397681

[7] Christopher Elvidge, Kimberly Baugh, Mikhail Zhizhin, and Feng-Chi Hsu. [n. d.]. Why VIIRS data are superior to DMSP for mapping nighttime lights. 35 ([n. d.]), 62–69. https://doi.org/10.7125/APAN.35.7

[8] Christopher D Elvidge, Kimberly Baugh, Mikhail Zhizhin, Feng Chi Hsu, and Tilottama Ghosh. 2017. VIIRS night-time lights. *International Journal of Remote Sensing* 38, 21 (2017), 5860–5879. https://doi.org/10.1080/01431161.2017.1342050 arXiv:https://doi.org/10.1080/01431161.2017.1342050

[9] Christopher D. Elvidge, Kimberly E. Baugh, Paul C. Sutton, Budhendra Bhaduri, Benjamin T. Tuttle, Tilotamma Ghosh, Daniel Ziskin, and Edward H. Erwin. [n. d.]. Who's in the Dark-Satellite Based Estimates of Electrification Rates. In *Urban Remote Sensing*, Xiaojun Yang (Ed.). John Wiley & Sons, Ltd, 211–224. https://doi.org/10.1002/9780470979563.ch15

[10] Christopher D. Elvidge, Feng-Chi Hsu, Mikhail Zhizhin, Tilottama Ghosh, Jay Taneja, and Morgan Bazilian. 2020. Indicators of Electric Power Instability from Satellite Observed Nighttime Lights. *Remote Sensing* 12, 19 (2020). https://doi.org/10.3390/rs12193194

[11] Christopher D. Elvidge, Mikhail Zhizhin, Tilottama Ghosh, Feng-Chi Hsu, and Jay Taneja. 2021. Annual Time Series of Global VIIRS Nighttime Lights Derived from Monthly Averages: 2012 to 2019. *Remote Sensing* 13, 5 (2021). https://doi.org/10.3390/rs13050922

[12] Giacomo Falchetta, Shonali Pachauri, Simon Parkinson, and Edward Byers. 2019. A high-resolution gridded dataset to assess electrification in sub-Saharan Africa. *Scientific Data* 6, 1 (jul 2019). https://doi.org/10.1038/s41597-019-0122-6

[13] Simone Fobi, Varun Deshpande, Samson Ondiek, Vijay Modi, and Jay Taneja. [n. d.]. A longitudinal study of electricity consumption growth in Kenya. 123 ([n. d.]), 569–578. https://doi.org/10.1016/j.enpol.2018.08.065

[14] Simone Fobi, Ayse Selin Kocaman, Jay Taneja, and Vijay Modi. 2021. A scalable framework to measure the impact of spatial heterogeneity on electrification. (2021), 67–81. https://doi.org/10.1016/j.esd.2020.12.005

[15] Dimitry Gershenson, Brandon Rohrer, and Anna Lerner. 2019. *A new predictive model for more accurate electrical grid mapping.* https://engineering.fb.com/2019/01/25/connectivity/electrical-grid-mapping/

[16] Tilottama Ghosh, Sharolyn J Anderson, Christopher D Elvidge, and Paul C Sutton. 2013. Using nighttime satellite imagery as a proxy measure of human well-being. *sustainability* 5, 12 (2013), 4988–5019.

[17] Tilottama Ghosh, Rebecca L Powell, Christopher D Elvidge, Kimberly E Baugh, Paul C Sutton, and Sharolyn Anderson. 2010. Shedding light on the global distribution of economic activity. *The Open Geography Journal* 3, 1 (2010).

[18] Digital Globe. [n. d.]. *Basemap + Vivid Product.* http://dg-cms-uploads-production.s3.amazonaws.com/uploads/document/file/2/DG_Basemap_Vivid_DS_1.pdf

[19] The World Bank Group. 2018. Kenya National Electrification Strategy: Key Highlights. (2018). https://pubdocs.worldbank.org/en/413001554284496731/Kenya-National-Electrification-Strategy-KNES-Key-Highlights-2018.pdf

[20] IEA, IRENA, UNSD, World Bank, and WHO. 2021. *Tracking SDG 7: The Energy Progress.* Retrieved January 14, 2022 from https://www.worldpop.org/project/categories?id=18

[21] Nevrez Imamoglu, Motoki Kimura, Hiroki Miyamoto, Aito Fujita, and Ryosuke Nakamura. 2017. Solar Power Plant Detection on Multi-Spectral Satellite Imagery using Weakly-Supervised CNN with Feedback Features and m-PCNN Fusion. In *Proc. of the 28th British Machine Vision Conference.*

[22] New Climate Institute. 2018. *The role of renewable energy mini-grids in Kenya's electricity sector.* Retrieved February 12, 2021 from https://ambitiontoaction.net/wp-content/uploads/2019/11/A2A-Kenya_Mini-grids-study_201911.pdf

[23] Wei Jiang, Guojin He, Tengfei Long, and Huichan Liu. 2017. Ongoing conflict makes Yemen dark: From the perspective of nighttime light. *Remote Sensing* 9, 8 (2017), 798.

[24] Stephen Lee, Srinivasan Iyengar, Menghong Feng, Prashant Shenoy, and Subhransu Maji. 2019. *DeepRoof: A Data-Driven Approach For Solar Potential Estimation Using Rooftop Imagery.* Association for Computing Machinery, New York, NY, USA, 2105–2113. https://doi.org/10.1145/3292500.3330741

[25] Noam Levin, Christopher C.M. Kyba, Qingling Zhang, Alejandro Sánchez de Miguel, Miguel O. Román, Xi Li, Boris A. Portnov, Andrew L. Molthan, Andreas Jechow, Steven D. Miller, Zhuosen Wang, Ranjay M. Shrestha, and Christopher D. Elvidge. 2020. Remote sensing of night lights: A review and an outlook for the future. *Remote Sensing of Environment* 237 (feb 2020), 111443. https://doi.org/10.1016/j.rse.2019.111443

[26] L. F. Luque-Vega, B. Castillo-Toledo, A. Loukianov, and L. E. Gonzalez-Jimenez. 2014. Power line inspection via an unmanned aerial system based on the quadrotor helicopter. In *MELECON 2014 - 2014 17th IEEE Mediterranean Electrotechnical Conference.* 393–397. https://doi.org/10.1109/MELCON.2014.6820566

[27] C. Martinez, C. Sampedro, A. Chauhan, and P. Campoy. 2014. Towards autonomous detection and tracking of electric towers for aerial power line inspection. In *2014 International Conference on Unmanned Aircraft Systems (ICUAS).* 284–295. https://doi.org/10.1109/ICUAS.2014.6842267

[28] Brian Min. 2015. Measuring Electricity from Space. In *Power and the Vote.* Cambridge University Press, 51–73. https://doi.org/10.1017/cbo9781316272121.004

[29] Brian Min and Zachary O'Keeffe. 2020. High Resolution Energy Access Indicators. https://github.com/zachokeeffe/nightlight_electrification

[30] Pezhman Nasirifard, Jose Rivera, Qunjie Zhou, Klaus Bernd Schreiber, and Hans-Arno Jacobsen. 2018. A Crowdsourcing Approach for the Inference of Distribution Grids. In *Proceedings of the Ninth International Conference on Future Energy Systems* (Karlsruhe, Germany) *(e-Energy '18)*. Association for Computing Machinery, New York, NY, USA, 187–199. https://doi.org/10.1145/3208903.3208927

[31] Jakub Ondracek, Ondrej Vanek, and Michal Pechoucek. 2015. Solving Infrastructure Monitoring Problems with Multiple Heterogeneous Unmanned Aerial Vehicles. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (Istanbul, Turkey) *(AAMAS '15)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1597–1605.

[32] Zeal Shah, Feng-Chi Hsu, Christopher D. Elvidge, and Jay Taneja. 2020. Mapping Disasters amp; Tracking Recovery in Conflict Zones Using Nighttime Lights. In *2020 IEEE Global Humanitarian Technology Conference (GHTC).* 1–8. https://doi.org/10.1109/GHTC46280.2020.9342937

[33] Zeal Shah, Noah Klugman, Gabriel Cadamuro, Feng-Chi Hsu, Christopher D. Elvidge, and Jay Taneja. 2022. The electricity scene from above: Exploring power grid inconsistencies using satellite data in Accra, Ghana. *Applied Energy* 319 (2022), 119237. https://doi.org/10.1016/j.apenergy.2022.119237

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

[35] WorldPop. 2020. *Population Density.* Retrieved January 14, 2022 from https://www.worldpop.org/project/categories?id=18

[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] Yuyu Zhou, Steven J. Smith, Christopher D. Elvidge, Kaiguang Zhao, Allison Thomson, and Marc Imhoff. 2014. A cluster-based method to map urban area from DMSP/OLS nightlights. *Remote Sensing of Environment* 147 (2014), 173–185. https://doi.org/10.1016/j.rse.2014.03.004

[38] Lingli Zhu and Juha Hyyppä. 2014. Fully-Automated Power Line Extraction from Airborne Laser Scanning Point Clouds in Forest Areas. *Remote Sensing* 6, 11 (2014), 11267–11282. https://doi.org/10.3390/rs61111267