This Little Light of Mine: Electricity Access Mapping Using Night-time Light Data

Santiago Correa, Zeal Shah, Jay Taneja
University of Massachusetts Amherst - {scorreacardo, zshah, jtaneja}@umass.edu

ABSTRACT

The tracking of electricity infrastructure locations is crucial to making informed decisions on grid expansion and energy supply alternatives. However, in developing settings, these tasks are limited by technical and budget capacity constraints where the most recent data on the locations of low- and medium-voltage grids is outdated or even unknown. Currently, utilities in high-income economies monitor these lines using sophisticated sensing devices, airborne laser scanning, and field surveys which are unaffordable in emerging economies. In this work we aim to improve upon an existing open-source electricity mapping tool that uses night-time light data as the main proxy of electrification. Using ground-truth data from Kenya, we validate the performance of the existing tool and proposed a learning model to improve the detection of electrified sites. Our results show that our learning model is able to correctly identify ≈78% of those places which had electricity but were not identified before and improve the detection accuracy by up to ≈7%. Moreover, we show that using daily composites of nighttime data combined with other open-source data sources significantly helps the generation of accurate electricity access maps.

CCS CONCEPTS

• Computing methodologies → Supervised learning by classification; **• Information systems → Data analytics**;

ACM Reference Format:
Santiago Correa, Zeal Shah, Jay Taneja. 2021. This Little Light of Mine: Electricity Access Mapping Using Night-time Light Data. In *The Twelfth ACM International Conference on Future Energy Systems (e-Energy '21), June 28-July 2, 2021, Virtual Event, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3447555.3464871

1 INTRODUCTION

Accurately tracking the state of electricity access is a crucial capability to assess progress and prioritize investments towards universal electrification [14]. Throughout the years, high-income economies have collected such information using sophisticated techniques such as airborne laser scanning [32], field and unmanned aerial system surveys [20, 21, 25], and conventional simulation methods to decide when and how to extend the grid. However, in emerging economies, these traditional methods are expensive and even unfeasible to implement since information about low- and medium-voltage power infrastructure is outdated or difficult to access [27].

Fortunately, the growing availability of open-source data and the ubiquity of remote sensing techniques are shifting the paradigm of how electricity planning and monitoring is accomplished. For example, the World Resources Institute has released the Global Power Plant Database [23] containing data on approximately 30k geo-located power plants across 164 countries with information on fuel type, capacity, and generation. OpenStreetMaps (OSM) [12] provides crowd-sourced ground-truth data about multiple features such as roads, power infrastructure, buildings, forests, rivers, and so on. This data availability is not limited to energy infrastructure but also includes population density [31], natural resources data such as wind speed [24], photovoltaic power potential [30], and land cover [3], all of which combined with energy-related data may offer significant insights into how to meet the United Nations Sustainable Development Goal (SDG) 7: to ensure access to affordable, reliable, sustainable and modern energy for all.

The single most-employed proxy for estimating electrification is night-time light data (NTL), collected by satellites over the entire Earth and made available on an open-source basis [11]. The particular strengths of this dataset are its global coverage, frequent collection (daily), consistency of measurement across administrative borders, and long historical record (the VIIRS satellite has reported data since 2012, and its predecessor DMSP-OLS satellite has reported data since 1992). However, the data exhibit substantial noise, tend to overrepresent streetlights, suffer from stray light effects, can be difficult to distinguish from lunar illuminance, and are powerless in the face of clouds. Further, deep rural areas, even when electricity access is present, may not register any signature because of extremely low levels of external lighting; still, despite these shortcomings, this valuable dataset remains the best available for assessing electricity access broadly.

Recently, a promising technique emerged for the challenge of locating existing power infrastructure. `gridfinder` takes a composite approach – using monthly and annual NTL data in concert with other widely available infrastructure, land cover, and population data – to produce an estimate of the world's electricity grids [4]. Even though this framework offers an impressive accuracy of 75% across a validation set of 14 countries and 88% specifically for Kenya, the presence of low and noisy NTL data in rural areas prevents `gridfinder` from identifying some settings as electrified. Improving on this performance requires new approaches and detailed ground-truth data for validation.

In this work, we aim to validate the estimations from `gridfinder`, specifically in Kenya where we have substantial ground-truth data containing more than 57k geo-located transformers provided by the local utility company. We perform a detailed analysis of locations

e-Energy '21, June 28-July 2, 2021, Virtual Event, Italy

ACM ISBN 978-1-4503-8333-2/21/06…$15.00
https://doi.org/10.1145/3447555.3464871

where `gridfinder` fails to predict the presence of power infrastructure and propose a learning model to improve the estimations of electrified sites using daily NTL data, which is beyond the temporal resolution generally used elsewhere. Improvements on areas where `gridfinder` performs poorly can especially enhance the prospects of accurate, global-scale energy access tracking.

2 RELATED WORK

Remote sensing data are valuable for a number of applications in low-income countries, where other data sources tend to be absent. Among these are crop yield prediction [26] and measuring road quality [6]. An interesting application that has raised interest is the use of NTL satellite data and machine learning to predict poverty. In [18], the authors state that NTL data are a rough proxy for economic wealth. A deep learning model was implemented using a multi-step transfer learning approach due to the lack of ground truth. The model is used to estimate either average household expenditures or average household wealth at a cluster of geographic aggregation. To estimate these outcomes, their transfer learning pipeline involves three main steps. Using a pre-trained CNN on ImageNet, the model learns how to identify low-level features such as edges and corners. Then, the model is fine-tuned to predict the NTL intensity corresponding to input daytime satellite imagery. Finally using survey data along with the features extracted, the authors train a ridge regression model that estimates cluster-level expenditures and assets. These models can explain up to 75% of the variations in local-level economic outcomes, emphasizing the heterogeneity captured in NTL data.

There are numerous examples employing remote sensing data for enhancing energy systems measurement, primarily in high-income countries. These include detection of solar PV arrays and power plants [5, 9, 16, 19]. Another application aims to estimate generation capacity using weather forecasting and solar irradiation data [28]. For low-income regions specifically, there is work aiming to measure electric power stability using NTL data [10]. The authors show evidence that long-term assessment of power supply growth and stability can be accomplished looking at indices such as the mean, variance, and lift of NTL irradiance.

In terms of electricity mapping we found that the World Bank Group and Development Seed built a pipeline to map high-voltage (HV) lines at country-scale[29]. Firstly, their model finds HV towers in satellite imagery using XCeption Neural Network [8]. The model was pretrained on ImageNet and generated a probability score on the interval [0,1] that a high-voltage (HV) tower was present. For training data, they used the Digital Globe Vivid layer which is a privately-owned high-resolution imagery repository. Then, the results were compiled and provided as a map overlay on top of satellite imagery. However, this technique still relies on manual tagging since professional mappers perform manual tracing from tower to tower to identify sections of HV power lines. While the end result does accurately trace out power networks, the limited scope to HV lines limits the value of this approach.

Tackling a similar problem, in [22] the authors propose a crowd-sourced framework to map electricity infrastructure using mobile devices and an algorithm that aims to collect open imagery data and the geographical location of grid infrastructure. However, this approach is bounded by the willingness of the crowd to participate and the availability of smartphone devices, both of which may be lacking in rural areas with recent electricity access.

The most robust approach found so far is `gridfinder` [4]. Their approach begins with a custom image filtering process of the monthly NTL satellite imagery to identify regions with consistent illumination (electrification targets). Then, assuming that all these regions are grid-connected, they are used as proxy for the existence of grid electricity. The implementation is based on a modified version of Dijkstra's shortest path algorithm [13] seeking to make connections of night lights in a specific region the most efficient way possible. The algorithm infers grid paths based on the likelihood to follow roads and avoid water. The ability for this system to estimate medium-voltage (MV) lines makes it the state-of-the-art for electricity grid estimation and, therefore, electricity access tracking.

3 DATA AND METHODOLOGY

3.1 NTL and ground-truth data

To explore and validate the performance of `gridfinder`, we conduct an analysis using the following datasets:

NTL and population counts. We leverage the NTL data recorded by the Day/Night Band of Visible Infrared Imaging Radiometer Suite (VIIRS) on the NPP-Suomi satellite. This is captured by VIIRS on a daily basis at a spatial resolution of 750 meters. These data are then processed and converted into an NTL grid of 15 arc-seconds (approximately 450 meters at the equator) which is publicly available in the form of daily, monthly, and annual composites [1]. Daily NL data has been made available for open access by the World Bank under the Light Every Night dataset [2]. For this study, we use the NTL grid of Kenya, which contains daily radiance profiles for every pixel in Kenya's NL grid from April 2012 to the present day. Radiance of pixels is given in units of $nW/sr/cm^2$. We computed the population counts per NL-grid pixel by aggregating the 2018 population count reported by WorldPop inside those pixels [31].

Transformer locations and minigrids of Kenya. We use data of distribution transformers and minigrids in Kenya. The transformer dataset contains geo-locations of more than 57k transformers, information about the date of commissioning (from 1966 to 2017), county name, and capacity (in kVA). Since we are interested in identifying the presence of access, we only make use of geo-locations and the date of commissioning because we need to match with the temporal dimension of the daily NTL. In terms of minigrids, we use the town location of the existing 21 minigrids in Kenya [17]. This dataset also contains commissioning date, number of connections (June 2016), and installed capacity (kW).

3.2 Methodology

Data Preprocessing. To evaluate the performance of `gridfinder`, we merge the three different data sources that comprise our data. According to `gridfinder`, 97% of the world's population lives within 10km of an MV line [4]; because of this, we create a generous buffer of 5 km around `gridfinder` predictions and intersect this buffer with transformer locations buffered by 600 meters. We choose 600 meters since it is the radius in which the utility charges customers a flat fee to connect in Kenya. All the transformer locations that

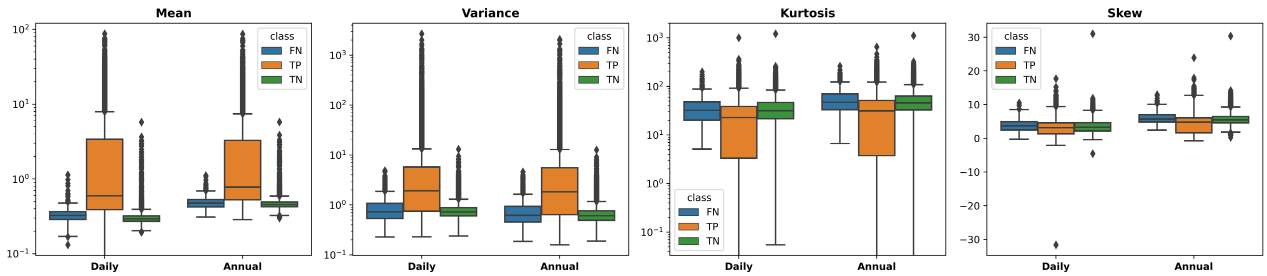

Figure 1: Comparison between different statistics of NTL data performed on examples classified as False Negatives (FN), True Positives (TP), and True Negatives (TN) from `gridfinder` and validated with ground-truth data. Each index is calculated using daily and annual NTL. We expect to see large radiance metrics in the FN group due to the presence of power infrastructure; however, the variations between FN and TN are not significant, which may explain why `gridfinder` fails to estimate the presence of electricity infrastructure in those regions.

Class	Group	Num. examples	Class Prop.
Electricity Access (1)	TP	48,011	75.9%
	FN	608	
No Access (0)	TN	15,431	24.1%

Table 1: Summary of number of examples per class and group obtained from NTL and transformer data.

overlap with `gridfinder` are labeled as True Positives (TP); otherwise, they are classified as False Negatives (FN). The goal of this paper will be to identify FN examples that could have been correctly identified as sites with electricity access.

The NTL data are composed of a spatial resolution of 15 arc-seconds ($\approx$ 450m). By the time of writing this paper, the daily composites available for Kenya covers $\approx$ 9.46% of the country so we need to match the location of each transformer classified in the previous step to the closest NTL cell available that covers the transformer – within the spatial resolution of the NTL – based on Haversine distance. The goal is to be able to map the transformer location to the most accurate NTL radiance at that specific point. To do so, we use k-nearest neighbor with $k = 1$ and filter out NTL data examples that have distances above the spatial resolution. NTL locations that do not intersect either with `gridfinder` prediction or the transformer locations are classified as True Negative (TN) examples. Table 1 summarizes the ground-truth dataset with the number of examples for each category.

We also performed some summary statistics within each group and look at different indices such as kurtosis, skew, and variance index between the daily and annual data. Kurtosis indicates how peaked a distribution is. Higher kurtosis represents that radiance is relatively stable and the distribution has shorter tails. Lower the kurtosis means the distribution of radiance for a pixel is flat and there is an equal chance of a pixel exhibiting high and low radiance (more fluctuations). A positive or negative skew value indicates that a pixelś radiance distribution is right- or left-tailed, i.e., a pixel experiences radiance in the lower or higher end of the brightness spectrum, respectively. Figure 1 illustrates a comparison between each index within each group and daily and annual radiance values. Mean and variance are higher for TP examples than for FN and TN. Also, across these indices, the difference between FN and TN is minor and varies slightly between daily and annual data which is likely the reason why `gridfinder` fails to estimate grid presence for the FN group. Kurtosis is higher for TN and FN which can be seen as less stable radiance levels in comparison to the TN group.

By looking at the radiance distributions among the groups it is difficult to observe substantial differences between TN and FN examples. However, there are minor differences that might be detected by ML-based models and help to reduce the FN rate.

ML-Based models. To assess the aforementioned observation, we define the problem as a binary classification task in which places without and with electricity access are labeled as [0,1] respectively. We build a dataset with daily and annual radiance indices, summary statistics, and population counts as predictors. We use traditional supervised learning classifiers such as decision trees, random forest, gradient boosting, support vector machine, and artificial neural networks (Multilayer Perceptron classifier). Since our data are imbalanced, we make sure that we stratify our datasets so train and test sets have proportionally the same number of examples for each class. Also, we separate the FN examples for testing since we aim to evaluate the reduction of the misclassification in this group.

Feature Selection. It is possible to reduce the complexity of a learning method if we perform feature selection on our set of predictors. Even though the curse of dimensionality is not an issue in this problem due to the number of examples in our dataset, this technique is useful to identify important predictors and their influence on performance. Tree-based models such as random forest and gradient boosting perform internal feature selection as part of the procedure so they are immune to the inclusion of many irrelevant predictor variables [15]. Table 2 shows the top 10 rankings of predictors for three different feature selection methods: feature importance applied to random forest, gradient boosting, and statistical filtering. We highlight predictors that are common for all three techniques and ranked within the top 5: Population count, daily mean, and 95th percentile NL values, all of which seem reasonable to be relevant to identify places with electricity access.

Hyperparameter optimization. We use exhaustive search over a discrete grid of values and evaluate the performance using 5-fold cross-validation. We tune the hyperparameters using different score metrics and refit the model using the best-found hyperparameters on the whole dataset for four different score metrics: Accuracy, Precision, Recall, and F1 score.

4 EVALUATION

We evaluate the performance of different models and the impact of daily and aggregated NTL data that is being used as predictors. We use the following metrics: *Accuracy* quantifies the number of

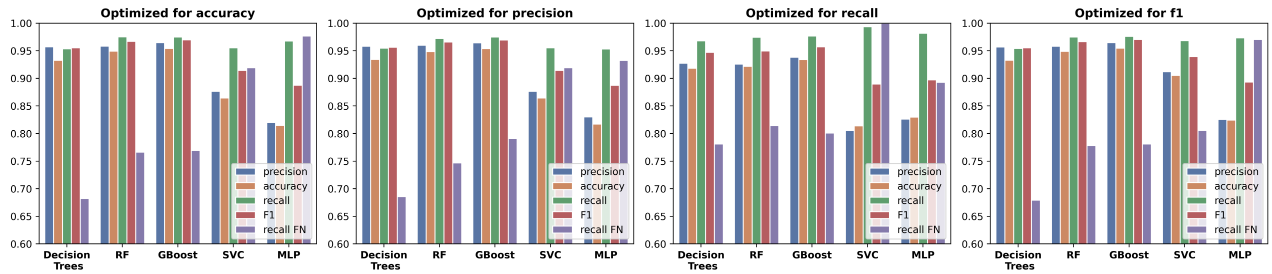

Figure 2: Performance metrics for the learning models. Each figure shows the performance for models tuned to maximize the different score functions. The legend indicates the performance metrics in the test set and recall FN point out the recall obtained only in the hold-out FN set.

Rank	Random Forest	Gradient Boosting	Statistical Filtering
1	Population	Population	Daily 95th percentile
2	Daily Mean	Daily Mean	Annual Skew
3	Annual Mean	Daily 95th percentile	Daily Mean
4	Daily Dispersion Index	Annual Variance	Annual Mean
5	Daily 95th percentile	Dispersion Index	Population
6	Daily Variance	Daily Kurtosis	Annual Kurtosis
7	Annual Variance	Annual Skew	Daily Kurtosis
8	Annual Skew	Annual Mean	Annual Kurtosis
9	Daily Kurtosis	Annual Kurtosis	Annual Variance
10	Daily Skew	Daily Skew	Daily Variance

Table 2: Top 10 predictors for different feature selection techniques. Common predictors within the top 5 are highlighted.

Metric	Daily	Annual Only
Accuracy	0.955	0.934
Precision	0.965	0.947
Recall	0.976	0.964
F1-score	0.970	0.956
Recall FN	0.781	0.731

Table 3: Impact of daily NTL in electricity access estimation using gradient boosting. Features obtained from daily NTL indices were removed in the annual model. Removing predictors from daily data reduces the performance of the model by on average $\approx 2\%$ for the test set and $\approx 5\%$ for the FN set obtained from `gridfinder`.

examples that are correctly classified. *Precision* is defined as the ratio $TP/(TP+FP)$ and looks at the impact of FP. *Recall* is the ratio $TP/(TP+FN)$ which is commonly used to evaluate models that associate a high cost to FN. *F1 score* is a function of precision and recall and it is important when we want to observe balance between those two metrics.

Performance of different learning models optimized for different classification metrics. We tune the parameters of each of our models using an exhaustive search over a discrete set of hyperparameters for each model. Using 5-fold cross-validation we can measure how well a given model generalizes. Furthermore, it is possible to optimize the selection of these hyperparameters that maximize the score of a specific metric. Figure 2 illustrates the performance of each model tuned for different scores. Since our dataset is imbalanced, the true performance improvement of our model needs to be compared to the zero rule benchmark or major class classifier [7]. For tree-based models we observe increments in accuracy from 4 to 7%. On the other hand, support vector (SVC) and multilayer perceptron (MLP) classifiers are extremely good at reducing the FN rate (high recall) but increase the number of false positives affecting precision and accuracy.

Tree-based models are usually more consistent than SVC and MLP classifiers, which show high variability when they are forced to optimize for recall and F1-score. For our problem domain, we aim to reduce the number of FN examples which is reflected in the recall and F1-score. However, our gradient boosting approach shows significant improvement in comparison to the zero rule for the test set and the recall FN. Recall FN is the recall obtained if we estimate electrifications only in the places that `gridfinder` did not recognized as electrified (FN examples). As we can observe in figure 2 and table 3, our model identifies ≈78% of those places which had electricity but `gridfinder` did not identify.

Performance using daily vs annual predictors. Even though the variations illustrated in figure 1 are minor, our learning models can still detect interactions that improve the classification performance. In Figure 2 we observe that the most consistent model across the optimization metrics was gradient boosting which generally has good recall performance. In order to measure the impact of daily data for our model, we retrain our gradient boosting model using only predictors from annual data and compare its performance with the model that includes daily predictors. Note that the original `gridfinder` implementation does not use daily NTL data.

Figure 3 summarizes the result of using predictors from daily versus only annual data. Daily data improves overall performance, by 2% on average for precision, recall, accuracy, and F1-score in the test set. We pay special attention to the performance for recall in the FN set since our goal is to improve the detection where `gridfinder` is failing in terms of reducing the number of false negatives. We can identify 78.1% of those cases and using daily features improve 5% of the recall obtained only with annual data.

5 CONCLUSION AND FUTURE WORK

Improving data on power infrastructure in developing settings is essential to reach the goal of universal electrification, increase efficiency in grid expansion investments, and mitigate the impact of climate change. In this work, we evaluate the performance of the existing state-of-the-art electricity access mapping technique using nighttime lights radiance data and evaluate the estimations with ground-truth data from Kenya. Using more granular composites of NTL data, we show that we can improve the detection accuracy by up to 7% and identify $\approx 78\%$ of electrified sites that were previously missed. Furthermore, even though daily nighttime lights data contains significant noise, the combination of such data with additional publicly-available datasets can provide better means of measuring electrification in developing countries.

REFERENCES

[1] [n.d.]. *Earth Observation Group, the Payne Institute for Public Policy, Colorado School of Mines.* https://payneinstitute.mines.edu/eog
[2] [n.d.]. *World Bank - Light Every Night.* https://worldbank.github.io/OpenNightLights/wb-light-every-night-readme.html
[3] European Space Agency. 2018. Climate Change Initiative. http://maps.elie.ucl.ac.be/CCI/viewer/.
[4] 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
[5] 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
[6] Gabriel Cadamuro, Aggrey Muhebwa, and Jay Taneja. 2018. Assigning a Grade: Accurate Measurement of Road Quality Using Satellite Imagery. In *NIPS 2018 Workshop on Machine Learning for the Developing World. Workshop paper.*
[7] Machine Learning Catalog. [n.d.]. *The role of renewable energy mini-grids in Kenya's electricity sector.* Retrieved February 18, 2021 from https://machinelearningcatalogue.com/algorithm/alg_zero-rule.html
[8] François Chollet. 2016. Xception: Deep Learning with Depthwise Separable Convolutions. arXiv:1610.02357 [cs.CV]
[9] 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
[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] The Earth Observations Group (EOG). 2021. *Nighttime Lights.* Retrieved February 12, 2021 from https://payneinstitute.mines.edu/eog/nighttime-lights/
[12] Geofabrik. 2018. OpenStreetMap Data Extracts. http://download.geofabrik.de/.
[13] Dimitry Gershenson, Brandon Rohrer, and Anna Lerner. 2019. Pathfinder: A many-to-many variant of the Dijkstra's shortest distance algorithm. Facebook Research. https://github.com/facebookresearch/many-to-many-dijkstra.
[14] Joel Gurin, Laura Manley, and Audrey Ariss. 2015. *Sustainable Development Goals and Open Data.* Retrieved February 6, 2021 from https://blogs.worldbank.org/digital-development/sustainable-development-goals-and-open-data
[15] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2001. *The Elements of Statistical Learning.* Springer New York Inc., New York, NY, USA.
[16] 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.*
[17] NewClimate 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
[18] Neal Jean, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 6301 (2016), 790–794. https://doi.org/10.1126/science.aaf7894 arXiv:https://science.sciencemag.org/content/353/6301/790.full.pdf
[19] 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
[20] 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
[21] 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
[22] 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
[23] Global Energy Observatory, Google, KTH Royal Institute of Technology in Stockholm, Enipedia, and World Resources Institute. 2018. *Global Power Plant Database.* Retrieved February 10, 2021 from https://datasets.wri.org/dataset/globalpowerplantdatabase/
[24] Technical University of Denmark (DTU Wind Energy) and the World Bank Group. 2021. *Global Wind Atlas.* Retrieved February 12, 2021 from https://globalwindatlas.info
[25] 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.
[26] Reid Pryzant, Stefano Ermon, and David Lobell. 2017. Monitoring Ethiopian Wheat Fungus with Satellite Imagery and Deep Feature Learning. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Workshop paper.*
[27] UN Global Pulse. 2012. *Big Data for Development: Challenges and Opportunities.* Retrieved February 10, 2021 from https://www.unglobalpulse.org/document/big-data-for-development-opportunities-and-challenges-white-paper/
[28] Fermin Rodríguez, Alice Fleetwood, Ainhoa Galarza, and Luis Fontán. 2017. Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renewable Energy* (2017).
[29] Development Seed and World Bank. 2019. Mapping the electric grid. Using ML to Augment Human Tracing Of HV Infrastructure. http://devseed.com/ml-grid-docs/.
[30] Solargis. 2020. *Global Solar Atlas v2.4.* Retrieved February 12, 2021 from https://globalsolaratlas.info/
[31] WorldPop. 2020. *Population Density.* Retrieved February 14, 2021 from https://www.worldpop.org/project/categories?id=18
[32] 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