Towards Equitable Access to Information and Opportunity for all: Mapping Schools with High-resolution Satellite Imagery and Machine Learning

Zhuang-Fang Yi[1], Naroa Zurutuza[2], Drew Bollinger[1], Manuel Garcia-Herranz[2], Dohyung Kim[2]

[1] Development Seed, [2] UNICEF

Abstract

Having accurate data about schools is key for organizations to provide quality education and promote lifelong learning, listed as UN sustainable development goal 4 (SDG4), ensure equal access to opportunity (SDG10) and eventually, reduce poverty (SDG1). However, this is a challenging task since educational facilities' records are often inaccurate, incomplete or non-existent. By leveraging machine learning and high-resolution imagery, we are able to determine school detection at the national scale.

Despite their varied structure, many schools have identifiable overhead signatures that might make possible to detect them with modern deep learning techniques applied to high-resolution satellite imagery. This paper presents the results of a Convolution Neural Networks school classifier for high-resolution satellite imagery. This classifier is based on the pre-trained models, Xception and MobileNetV2, from ImageNet, with a new cleaned school dataset in Colombia.

We selected the best-performed model with 0.94 area under the ROC curve and 9% of the false positive rate from our nearly 200 training iterations. A large scale model inference across Colombia and the Eastern Caribbean islands was implemented over 52 million DigitalGlobe *Vivid imagery tiles. A group of five expert mappers was able to validate 73,000 predicted school tiles from our machine learning inference within eight working days. We added about 11,000 schools to the map in Colombia and the Caribbean islands, and around 7,000 of them were unmapped schools.*

Our study showed that current deep learning and inexpensive cloud computing can assist humans to detect schools at scale in a rapid, rigorous manner. This provides the first object-based detection model for schools. A complete and accurate school facility map can further reduce the digital divide in education and improve children's access to information, digital goods, and opportunities, and make the best use of limited educational resources.

1. Introduction

Equal access to quality education is one of the Sustainable Development Goals[1] (SDGs). An accurate, comprehensive map of schools - where no school is left behind - is critical to measuring and improving the quality of learning. The map, combined with connectivity data collected by UNICEF's Project Connect initiative, can be used to reduce the digital divide in education and improve access to information, digital goods and opportunities for entire communities. In addition, understanding the location schools, can help governments and international organizations gain critical insights around the needs of vulnerable populations, and better prepare and respond to exogenous shocks such as disease outbreaks or natural disasters.

However, accurate information on school locations, especially in developing countries, is often unavailable. Traditionally, school gelocations are gathered and mapped by local governments - sometimes at the regional level, other times at national level[2]. Unfortunately, some national governments still don't know where all the schools in their country are. Crowdsourcing can be used as an alternative for these cases: e.g. through OpenStreetMap[3], volunteers can map schools at their crowdsourcing platform. This kind of maps is hard to verify and keep up-to-date, leading into inaccurate, outdated information.

Underlying patterns of the school building(s) can be spotted from space. Even in developing countries, schools normally have a distinguishable shape: bigger size building(s) and bigger bare ground(s) surrounding compared to other residential buildings in rural areas. Moreover, schools in wealthier urban areas have additional facilities such as basketball courts, swimming pools, playfields, parking lots and large building complexes(Figure 3). Hence, all these attributes form a matrix that makes a school observable from fine spatial resolution, submeter, satellite images.

The aim of this paper is to explore the use of contemporary deep learning methods for assisted automated mapping of schools to locate accurate school geo-locations in a timely manner. This methodology will allow us to rapidly, frequently and rigorously detect schools in different terrains and regions. This work forms the foundations towards an approach that could be used to allow point-to-point mapping of schools - and eventually other facilities like health centers - at national scale on a repeatable basis.

Recent advances in the fields of earth observation and computer science offer tremendous potential for mapping schools with machine learning at scale. First, fine spatial and temporal resolution satellite sensor data is widely available. In this study, we used DigitalGlobe Vivid Imagery [4] (DG Vivid) at zoom level 18 [5], which is 0.596m spatial resolution.

DigitalGlobe has supported this research by providing access to its 31cm resolution Vivid Basemap. This project was initially selected as a recipient of a GBDX Research Award [6], and continued to be supported by DigitalGlobe after the life of the award. DG Vivid is a snapshot of the earth, mosaiced, sharpened and color-enhanced. This high-resolution imagery product has a 50cm spatial resolution and global coverage. At this submeter spatial resolution imagery, objects like buildings, cars, and trees can be observed clearly from space.

The second key technological advance that enables this work is a GPU-powered image classification based on Convolutional Neural Networks (CNN). At the emerging spotlight of machine learning, CNNs allow partial remote sensing applications, especially object-based and pixel-based remote-sensed image analysis, moving towards high performing GPU-powered parallel computed machine learning. In fact, multiple studies indicate that CNN based machine learning methods can yield into more accurate classifications[7, 8] than conventional statistical classifiers.

Last but not least, deploying the school detection model at a large spatial scale was only possible because of the advances in cloud computing. Cloud computing allows for on-demand, cheap access to computing power, data storage, applications, and other IT resources [9].

The above three technological developments offer the possibility to train a school classification to national or continent level in days that would have not been possible five years ago. To our knowledge, this study is the first case of using deep learning to detect schools from satellite imagery at the national scale. We hope that in the near future it will inspire further research that contributes to SDGs. We hope our work can shape a methodology for routine, inexpensive detection of schools globally, and accelerating the work towards the 2030 SDGs.

2. Methodology

Our workflow of school detection with DG Vivid imagery and machine learning can be simplified as a diagram as follows (Figure 1).

2.1. Preparation of training dataset

A high-quality training dataset is essential for the machine learning models to learn the right object features. Our first step is to validate and clean the schools provided by UNICEF. Development Seed's Data Team, a group of five expert mappers, validated the dataset that was used for training. The goal of the validation exercise is to split the original data into 1) a set of image tiles that contain schools, 2) a set of image tiles that clearly do not contain a school, 3) a set of image tiles that it's uncertain if it contains a school or not. The first two sets of clear school tiles and clear non-school tiles were used by the neural network to learn how to distinguish between the two. We use red, blue and green (RGB) spectral bands from DG Vivid to create the training image tiles.

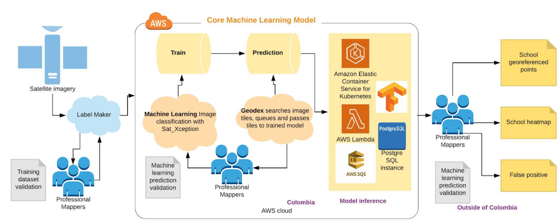

Figure 1. The diagram of our school detection system with high-resolution satellite imagery and machine learning. The Core Machine Learning Model and model inference sit at the center of the diagram. To its left is the pipeline for cleaning, validating and generating the training dataset, and to its right is the machine learning prediction and validation.

2.2. School data cleaning and training data creation

A team of expert mappers tagged the school geolocations within the original dataset of 44,665 schools with the attributes of "confirmed", "unrecognized" and "not-school". **Confirmed schools** are these observed from the high-resolution satellite imagery and have very clear school features, e.g. building size, shape, and facilities. For instance, below are some of the school features that were used as criteria for schools and that can be used to label the tiles as “confirmed” schools.

Figure 2. School feature examples that our expert mappers use to clean and validate schools across Colombia.

Figure 3. A few examples of “not-school” covers various look of buildings and not buildings in Colombia.

The **unrecognized school** referred to geolocations that were part of the original 44,665 school dataset but that had no clear school features, especially in urban areas with high building density or, in rural areas that can’t be distinguished from residential buildings. Another case of unrecognized schools is school building(s) that can not be seen on DG Vivid because of cloud/tree cover.

The **not-school** refers to locations from the original 44,665 school dataset where the expert mappers could not find any school looking buildings at the provided school geolocations. As an example, some of the schools were mislocated in the middle of the ocean. This can be caused the school geolocation was recorded incorrectly or because the DG Vivid imagery has been updated in particular areas of Colombia after schools were built.

Two categories of datasets, ‘school’ and ‘not-school’, were generated as the training dataset for the machine learning model. We randomly sampled half of the school geolocations from the validated “confirmed schools”, and generated 5,904 tiles as “school” training dataset (Table 1.). The not-school category is not as trivial. This category can contain forest, grassland or agricultural fields without any buildings, among others. It can also be a building complex or facility that looks very similar to “school” from space, e.g. hospitals, market places, courthouses etc.

To enrich the “not-school” category, and allow our machine learning algorithm to detect real schools more accurately, we queried available hospital, farmlands, parks, courthouses and marketplaces from OpenStreetMap and added them to the “not-schools” category. We ended up with 9,092 ‘not-school’ tiles (Table 1.). Through the machine learning model training iterations, we learned that the model was over-confident in certain areas, and therefore we adjusted the training dataset accordingly. For instance, we randomly included more schools as well as purposely added more regular building tiles that are not schools. See the detailed information(Table 1.) for the training dataset for the different sessions.

Tasks	Confirmed	Unrecognized	not-school	Total
Data Cleaning	10,951	26,638	7,066	**44,655 schools**
Training dataset for the 1st session	5,904	-	9,092	**14,996 tiles**
Training dataset for the 2nd session	8,716	-	11,192	**19,908 tiles**

Table 1. School data cleaning and training dataset creation for machine learning.

2.3. Sat-Xception and the school classifier

To quickly train the school classifier, we create a deep learning python package called Sat-Xception. It is a deep learning package that utilizes pre-trained models from ImageNet [10]. It is currently private but will soon be open sourced. The package is designed to quickly install, transfer-learn and fine-tune image classifiers with the built-in pre-trained models, which can be used to train other classifiers rather than the school classifier. Xception [11] and MobileNetV2 [12] were two pre-trained models built-in in the package. They were written by Keras [13], a high-level python package that can allow users to quickly reconstruct neural networks Google's Tensorflow [14] was used as backend.

Xception is one of current state-of-the-art CNN architectures and pre-trained models on top of ImageNet [15]. It's a high performing and efficient network compared to other pre-trained networks [16]. MobileNetV2 [17] on the other hand, is a model that is slightly less accurate compared to Xception. However, it's a very light-weight, fast, and easy to tune when limited resources available (appropriate for scenarios with resource vs. accuracy tradeoff). Both Xception and MobileNetV2 have fewer hyper-parameter available to tune compared to other pre-trained models, e.g. VGG and Inception, but both are high-performed models.

We broke the training sessions into two sessions. The first session was designed to test the feasibility of using Sat-Xception to train a well-performed school classifier in Colombia. The model was over-confident in rural Colombia in the first session, leading into too many false predictions in the area. To overcome the issue, we created a new training dataset that was slightly different from the training dataset in the first session. In the second training session, 2,048 'not-school' buildings were added. In addition, for the "school" category, we only kept rural schools that have very clear school features (Figure 2.). We also randomly selected another 2500 confirmed school tiles to add to the category (Table1.).

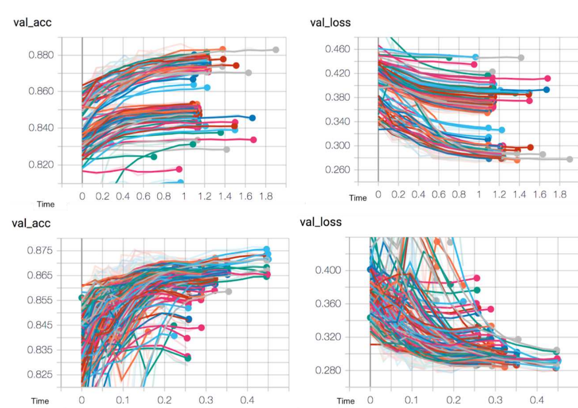

Figure 4. The validation accuracy and loss from model training iterations. Model training iteration with Xception at the first row and MobileNetV2 at the second row. The two models were trained on the same dataset, and the following graph only reflected the training with the 2nd session of the training dataset (Table 1).

We trained about 200 model iterations on two separate AWS EC2 [18] P3.2xlarge [19]. They are AWS's deep learning AMI [20] machines that have deep learning virtual environment setup, e.g. python3 with Tensorflow GPU version pre-installed in our case, and ready-to-use. We found the best-performing model from MobileNetV2 with a validation accuracy of 0.88. However, Xception reached a validation accuracy of 0.89, and therefore, we picked the model trained with Xception (see Figure 4.). We packaged the best-trained Xception model with Tensorflow Serving [21]. Tensorflow Serving helps to package the Keras and Tensorflow model as a Docker image [22]. The image can serve as an endpoint for large spatial scale model inference, which allows us to run model inference on tens of millions of image tiles per hour without manually watching the inference.

2.4. Large scale model inference

To run the inference at a country-wide scale on high-resolution imagery, we developed a library called "chip-n-scale-queue-arranger". It is currently private but will soon be open sourced. It is composed of a configuration code for Amazon Web Services (AWS CloudFormation) as well as small scripts for automating the necessary tasks associated with the satellite imagery and machine learning predictions. A high-level overview of the process looks like this:

A. A user sends `x/y/z` tile indices to an AWS SQS queue in the form of `'{ "x": 1, "y": 2, "z": 3}'`. These indicate which geographic region to run our model over.
B. Each SQS message triggers the Lambda function `DownloadAndPredict` which downloads images, posts to a prediction cluster (via a Load Balancer endpoint), and saves the result to an RDS database.
C. The prediction cluster on ECS runs the TensorFlow Serving image to predict each tile. All instances are behind an Application Load Balancer which will dynamically register new instances that appear on the cluster and allocate jobs to them evenly.
D. A user manually downloads predictions from RDS after the full inference process is complete.

The prediction from RDS was converted to a shapefile so that our expert mappers can upload a map editor for validation. In the editor, the mappers can overlay the predicted school tiles only and focus their attention on confident predictions, avoiding the tedious task of reviewing the entire area for Colombia and the eastern Caribbean islands.

3. Results and discussion

Searching for Points of Interest (POIs) (e.g. schools and hospitals) manually is time-consuming, especially at the national level. Our approach provided a way where humans are assisted by machine learning to narrow down the search space of the POIs from high-resolution satellite imagery. However, schools can look very different in different regions, for instance, schools in rural and urban areas often don't have the same characteristics. Furthermore, the looks will vary a lot from country to country. To overcome this challenge, the parameterization of the school classifier can be further explored for optimal classification to specific geo-locations.

3.1. The overall model performance

The machine learning model took individual tiles as input and provided output in the form of a probability between zero and one for each. By running the model prediction over the test set (about 2000 tiles) we plotted a ROC curve. The ROC curve gives us an overall model performance and guidance on threshold cutoff.

The ROC curve indicates the area under the curve is up to 0.94 from the test set with the given threshed of 0.44 (Figure 5) - we got a false positive rate of 9%. The ROC curve (Figure 5, A) shows that the model we selected was a high-performed model, where only 9% of schools have ring false alarms.

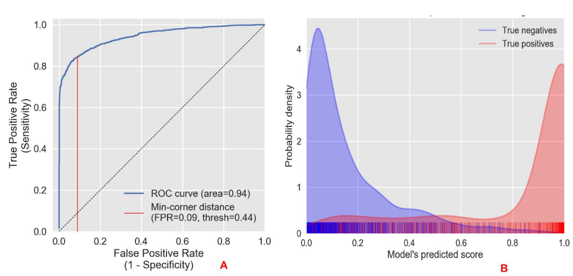

Figure 5. The ROC curve and false positive rate for the school classifier in Colombia.

We found that the Xception model was more accurate but that MobileNetV2 was much faster for each model training iteration. Specifically, MobileNetV2 only used a quarter of time per training iteration on exactly the same training set. In the case of our study, we had the AWS infrastructure set up and available and we were able to take more computing resources for increasing the accuracy. However, for whom does not have access to rich cloud computing resources, we recommend using MobileNetV2 to train the model instead.

During the validation process, the expert mappers validate each predicted school tile and tag it as "yes", "unrecognized" and "no" based on the school features part of the selection criteria defined during the initial data cleaning process (Figure 2). With the increase of the threshold (e.g. from 0.44 to 0.99) we would limit the false predictions but in the process, we will also lose an increasing proportion of correct predictions (Figure 5).

3.2. AI/ML-assisted school detection

A threshold score of 0.92 is, and the model inference is run on Colombia and the eastern Caribbean islands. 73,717 tiles were predicted as school tiles and passed to the team of expert mappers to validate. It's almost impossible to search for schools at the national level, however, the machine learning model was able to narrow down the searching space to less than 0.15% of 52 million tiles for our expert mappers to validate. With a validation speed of 10,000 tiles per day, the mappers identified 10,998 school geolocations, where 6,954 of them are unmapped schools (schools that were not part of the initial dataset of 44,665 schools) (Figure 7). 60,568 predicted school tiles were tagged as "unrecognized" by our expert mappers, which turned into a heatmap (Figure 6). These tiles don't have clear school features (Figure 2). During the machine learning prediction validation, we found that schools in rural areas are hard to verify as schools because a "school" can't be distinguished from residential houses.

The machine learning model generalizability is an active research area [23, 24], and in our study, the school classifier we trained in Colombia generalized well in the Eastern Caribbean islands. We added 262 schools to the islands that had not been mapped before. If users are looking at different countries or regions that have different terrains than Colombia, we recommend adding few but representative schools to fine-tune our current trained school classifier. Our school classifier, the Tensorflow Serving image(GPU version), lives on DockerHub now. It's open-source and free to run as an end-points to users who want to send DG Vivid zoom 18 images tiles to classify schools in their area of interest.

Total tiles reviewed	**73, 717**
Confirmed ml_school tiles	12,250
Confirmed and added ml_school geolocations	10,998
ml_school unrecognized	60,568
ml_school no	899

Table 2. The final stats from our machine learning school detection in Colombia and the eastern Caribbean island.

The contemporary machine learning shed light on narrowing down the search space for POIs. The trained model can be run anytime to update the predictions frequently. An accurate school map, produced by a human-in-the-loop type of approach (combining machine learning and the expert mappers together), will still need field agents on the ground to further validate and confirm 'unrecognized' schools with local authorities.

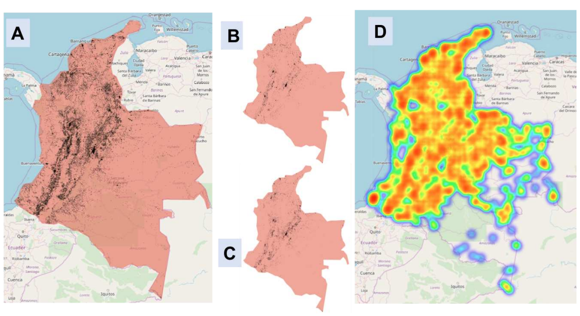

Figure 6. Schools are represented as black dots from A to C in the figure. We had 44,655 schools provided by UNICEF(A.) and among them, 10,951 schools (B.) have been confirmed as schools by their clear school features shown in Figure 2. We use 70% of schools (10,951) in (B.) to train the machine learning model, and after the validation of predictions, 10,988 schools were added to the map (C.), around 7000 of them are unmapped schools that we newly added after the ML validation. The school heatmap was created from the machine learning predicted and validated as "unrecognized" by our expert mappers (D). The heatmap is an interactive map that the field agents/validators on the ground can room in and see much more detail of the possible geolocation of the school within a 152 by 152 m searching space.

Another AI-assisted high-voltage grid mapping model developed by Development Seed has been proved working. In this case, a similar model and workflow were used to detect high-voltage towers in Pakistan, Nigeria, and Zambia that the expert mappers afterward used to map the high-voltage grid at the country-wide scale [25]. The grid map now lives in OpenStreetMap that the platform users who have local knowledge about the electric grid can update and correct the grid. Machine learning has provided a means to study the SDGs related issues [26] the school mapping case we present in this paper address it's another example of machine learning or AI assisted infrastructure mapping from satellite imagery.

Figure 7. A few snapshots of newly discovered schools by the machine learning model that were validated by the expert mappers and they were not part of the initial dataset of 44,665 schools.

4. Conclusion

Satellite imagery enables the generation of unique insights into population dynamics, urbanization, and economic development at a scale that cannot be achieved using conventional survey techniques in isolation. Machine learning is capable of detecting many types of objects in satellite imagery. By leveraging machine learning and high-resolution imagery, we are able to determine school locations at a national scale, proofing that machine learning can assist humans to narrow down the search space for points of interests. This can speed up the process of mapping schools or other POIs considerably.

For this study, we fine-tuned pre-trained models (Xception and MobileNetV2), with cleaned school data in Colombia. Our school classifier model achieved 0.89 validation accuracy with Xception, and

0.88 with MobileNetV2 on high-resolution satellite imagery. Specifically, the satellite image tiles with 59.6cm spatial resolution were queried from DG Vivid Maps API. The school classifier is a high-performed model on current training dataset, with 0.94 AUC and 9% of false positive rate. A large scale model inference across Colombia and the Eastern Caribbean islands was implemented over 52 million tiles with our best-performed trained model. The model inference returned more than 73,000 predicted school tiles that were further validated by our expert mappers. As a result, about 11,000 schools were added to school maps for Colombia and the Caribbean islands, among which around 7,000 of schools are new (Figure 6, 7), previously unmapped schools.

Our school classifier, including model training and model inference, can be fully automated and scaled up to much larger geospatial areas in the future. These results suggest considerable potential for mapping schools at a scale, quickly with human mappers' in the loop for validation. This will support the improvement of education information management systems, reduce gaps in access to information and opportunity, improve the quality of education, and further disaster response to vulnerable populations. Thereby aid the achievement of the relevant UN Sustainable Development Goals of equal access to quality education.

References

[1] UN Sustainable Development Goals. https://sustainabledevelopment.un.org/?menu=1300

[2] School Mapping | UNICEF Office of Innovation. https://www.unicef.org/innovation/school-mapping

[3] OpenStreetMap. https://www.openstreetmap.org/

[4] Imagery and Basemaps - DigitalGlobe Maps API Documentation. https://mapsapidocs.digitalglobe.com/docs/imagery-and-basemaps

[5] Zoom levels - OpenStreetMap Wiki. https://wiki.openstreetmap.org/wiki/Zoom_levels

[6] "GBDX for Sustainability Challenge: Mapping every school in the world ...." May. 2018, http://blog.digitalglobe.com/industry/gbdx-for-sustainability-challenge-mapping-every-school-in-the-world-and-reducing-the-digital-divide-in-education/.

[7] X. Chen, S. Xiang, C. Liu and C. Pan, "Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks," in *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1797-1801, Oct. 2014.doi: 10.1109/LGRS.2014.2309695

[8] E. Maggiori, Y. Tarabalka, G. Charpiat and P. Alliez, "Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 645-657, Feb. 2017. doi: 10.1109/TGRS.2016.2612821

[9] What is Cloud Computing? - Amazon Web Services https://aws.amazon.com/what-is-cloud-computing/

[10] "ImageNet - Wikipedia." https://en.wikipedia.org/wiki/ImageNet.

[11] F. Chollet. Xception: Deep Learning with Depth Wise Separable Convolutions.Oct. 2016. https://arxiv.org/abs/1610.02357.

[12] "Applications - Keras Documentation." http://keras.io/applications/.

[13] "Keras Documentation: Home." http://keras.io/.

[14] "TensorFlow." https://www.tensorflow.org/.

[15] ImageNet. http://www.image-net.org/

[16] Applications - Keras Documentation. http://keras.io/applications/

[17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks." 13 Jan. 2018, https://arxiv.org/abs/1801.04381.

[18] "Amazon EC2 - AWS - Amazon.com." https://aws.amazon.com/ec2/.

[19] "Amazon EC2 P3 – Ideal for Machine Learning and HPC - AWS." https://aws.amazon.com/ec2/instance-types/p3/.

[20] "AWS Marketplace: Deep Learning AMI (Ubuntu)." https://aws.amazon.com/marketplace/pp/B077GCH38C.

[21]"Serving Models | TFX | TensorFlow." https://www.tensorflow.org/tfx/guide/serving.

[22] "About images, containers, and storage drivers | Docker Documentation." https://docs.docker.com/v17.09/engine/userguide/storagedriver/imagesandcontainers/.

[23] "Understanding deep learning requires rethinking generalization." 10 Nov. 2016, https://arxiv.org/abs/1611.03530.

[24] "Generalization | Machine Learning Crash Course | Google Developers." https://developers.google.com/machine-learning/crash-course/generalization/video-lecture.

[25] "Mapping the electric grid - Development Seed." http://devseed.com/ml-grid-docs/.

[26] "Earth Observation and Machine Learning to Meet Sustainable ... - MDPI." 29 Jan. 2019, https://www.mdpi.com/2072-4292/11/3/266.