A Comparative Study of Gradient Boosting Algorithms for Cropland Delineation

Dr. Anand Sahadevan[1], Aneri Patel[2], and Rudra Patel[2]

[1]Space Applications Center, Indian Space Research Organization
[2]Dhirubhai Ambani Institute of Information & Communication Technology, Gandhinagar, Gujarat, India

***Abstract*—In this study, a dual class cropland classification for the area of study is produced using pixel based machine learning methods on Sentinel-2 satellite multi-spectral temporal images using the red, green, blue, near infrared and shortwave infrared bands and derived indices - Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI) and Norm. The classification output of the various algorithms implemented is validated using ground truth values provided by the Ministry of Forest and Agriculture of Slovenia. [1]. Finally, the accuracy of 3 gradient boosting [2] machine learning models - LightGBM [3], Catboost [4] and Random Forests for this classification problem are compared in order to conclude that the Catboost algorithm with it's default parameters is optimum for the problem of cropland delineation, resulting in a classification accuracy of 91.2%.**

I. Introduction

Land cover mapping has many uses in the area of sustainable agriculture and monitoring variation in distribution of cultivated land. Timely and accurate information on the global cropland extent is critical for such applications. Remote sensing can be used for such purposes either through direct use of geospatial data itself or after integrating statistical data with it for better interpretation [5]. Remote sensing significantly lowers costs and improve reliability when compared to ground visits for the purpose of cropland delineation. The increasing spatial and temporal resolution of globally available satellite images, such as those provided by Sentinel-2, creates a better input to generate accurate cropland maps. Traditional works in this field use machine learning algorithms like random forests [6] or segmentation for an object based approach [7]. This paper studies the efficiency and accuracy of recently developed gradient boosting decision tree [8] based algorithms that achieve a greater accuracy.

II. Study Area

The area of interest for this study is a square box bounded region in Slovenia. Projected longitude and latitude values of the upper left and bottom right corners of this bounding box into meters are ([357,737.03, 632593.82], [5016135.99, 5204957.80]). Slovenia has a total area of $2.0273e+7m^2$, out of which a total area of approximately $9e+4m^2$ (1000x1000 pixels) is studied. Slovenia being a small, predominantly hilly and mountainous country located in the middle of Europe, finds its principal model of agriculture to be family farming. Thus, the cultivated lands in this area are fragmented.

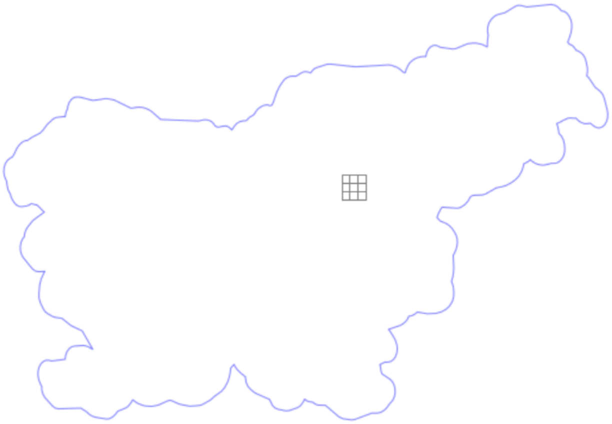


Fig. 1. Location of Study Area

III. Data

A. Data

The multi-spectral time series data used is free and open access data obtained using the eo-learn [9] package in python that downloads data from the European Space Agency's (ESA) website [10]. Sentinel-2 satellite images corresponding to the study area were utilized for this.

B. Data Preparation

Because of the internal malfunction of satellite sensors and poor atmospheric conditions such as thick cloud, the acquired remote sensing data often suffer from missing information, i.e., the data usability is greatly reduced [11]. Due to this missing information, the data we have to work with is irregularly sampled. We handle this issue by using linear interpolation to have a regularly sampled dataset.
A cloud that discards images with greater than 20% cloud cover in the images is also applied, leaving behind 69 images to use as input to the algorithms.

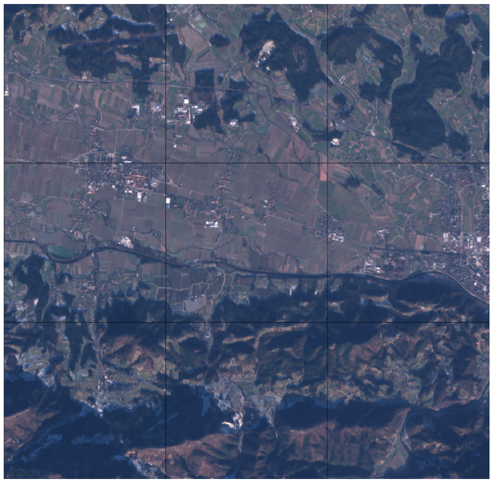


Fig. 2. True Color Image of AOI

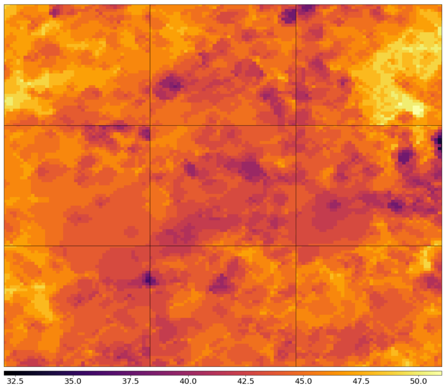


Fig. 3. A heatmap showing a count of the number of times a valid image of the area represented by a pixel is captured.

C. Data Organization

The study area is divided into 9 square tiles in order to split the training and testing data in a tile-wise manner. This is done in order to produce a scalable classification model that requires only a part of the complete image in order to classify the rest of it.

The data is split into training and testing datasets by randomly choosing n_{test} number of testing tiles and $n_{train} = b - n_{test}$ number of training tiles. In this case, as there are only 9 total tiles, a single testing tile seems to suffice. This tile is chosen manually by taking into account the land cover elements present in the region that each tile covers, but ideally, it should be done manually for cases where there are a larger number of tiles.

The time span for which data was retrieved ranges over a period of 12 months, from 1^{st} January 2017 to 1^{st} January 2018.

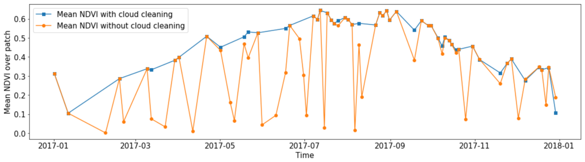


Fig. 4. A time-series plot indicating the spatial mean of NDVI values of each pixel in an image before and after cloud filtering at each instant

Typical features used for the classification process involving Earth Observation data (EO-data) are values in a particular band, NDVI or other indices.

Band Name	Band ID	Resolution (m)
Blue	2	10
Green	3	10
Red	4	10
NIR	8	10
SWIR	11	20
SWIR	12	20

In addition to these bands, a few other related indices are calculated in order to augment the feature space and improve classification accuracy. These indices that act as bands themselves are derived from the values of other bands on which their values are dependent. They are calculated for each pixel, for each point in the time series data.

D. Derived Spectral Indices

- NDVI : Normalized Difference Vegetation Index - $\frac{NIR-RED}{NIR+RED}$
- NDWI : Normalized Difference Water Index - $\frac{NIR-SWIR}{NIR+SWIR}$
- Euclidean Norm : Norm = $\sqrt{(\sum_i B_i^2)}$ - where B_i represents the value of the i^{th} individual band.

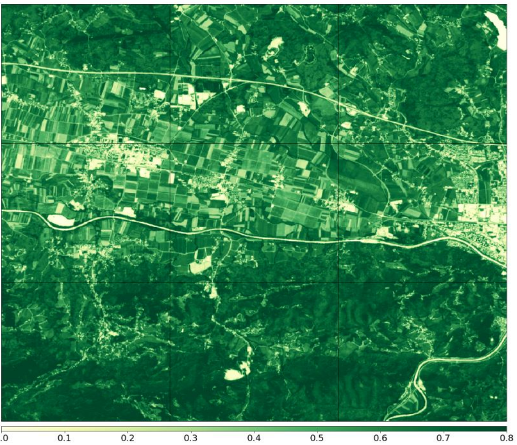


Fig. 5. A graphical representation of the temporal mean of NDVI feature values for each pixel covering the study area.

E. Ground Truth

The output of the pixel classification algorithms is validated using a reference dataset [1] that has ground truth values indicating the type of land cover over each point of the

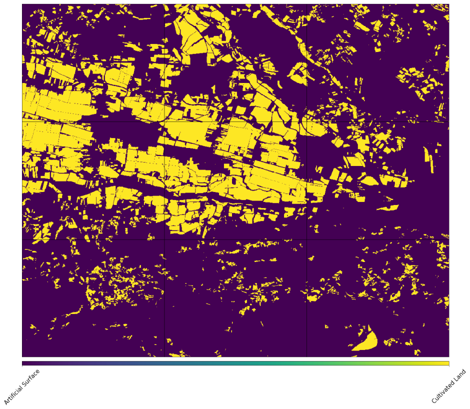


Fig. 7. Ground truth values for 2 classes

study area in the year 2017. It distinguishes land cover into 11 different classes (No data, Cultivated land, Forest, Grassland, Shrubland, Water, Wetlands, Tundra, Artificial surface, Bareland, Snow and ice) and is provided in Scalable Vector Graphics (SVG) [12] format by the Ministrstvo Za Kmetijstvo, Gozdarstvo in Prehrano of Slovenia. For the purpose of cropland delineation, all land cover classes other than cultivated land in the reference data were grouped under a super-class 'non-cultivated land'.

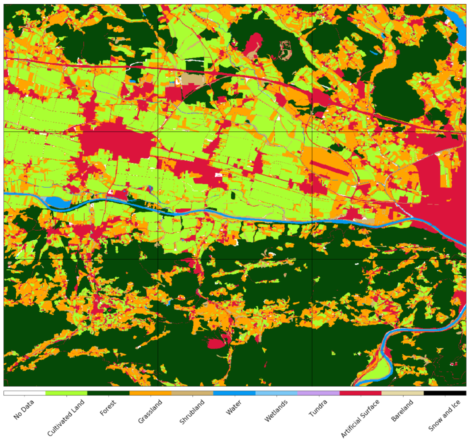


Fig. 6. Ground truth values for 11 classes

IV. Methodology

A. *Conceptual description of classification methods*

When designing a model in domain-specific areas, one strategy is to build a model from theory and adjust its parameters based on the observed data. [13] The lack of such models for cropland delineation can be circumvented if one applies non-parametric machine learning techniques like neural networks, support vector machines, or any other algorithm at one's own discretion, to build a model directly from the data. [13]. These are supervised learning methods that required labelled data.

The most prominent examples of such machine learning ensemble techniques are random forests (Breiman, 2001) and neural network ensembles (Hansen and Salamon, 1990), which have found many successful applications in different domains (Fanelli et al., 2012; Qi, 2012).

Previous research has shown that an ensemble is often more accurate than any of the single classifiers in the ensemble (Opitz et. al.,1999c). Bagging (Breiman, 1996c) and Boosting (Freund Shapire, 1996; Shapire, 1990) are two relatively new but popular methods for producing ensembles.

Boosting methods are based on a different strategy involving ensemble formation where new models are added to the ensemble sequentially, training a new weak base-learner model in every iteration with respect to the error of the whole ensemble learnt so far.(Natekin et.al., 2013) In Gradient boosting machines (GBM's) the learning procedure consecutively fits new models to provide a more accurate estimate of the response variable (Natekin et.al., 2013).

1) *Random Forest:* Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [6]

2) *LightGBM:* Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) are two novel features implemented in LightGBM in order to tackle the problem of unsatisfactory efficiency and scalability in prior implementations, where the feature dimension is high and data size is large [3]

3) *Catboost:* CatBoost is a variant of of GBM's with the implementation of ordered boosting, a permutation driven alternative to the classic algorithm. This was implemented to fight a prediction shift caused by a special kind of target leakage present in all currently existing implementations of gradient boosting algorithms [4].

B. *Accuracy measures*

- Precision - Precision and recall are well-suited to evaluating problems where the goal is to find a set of items from a larger set of items. Precision describes the proportion of entities – e.g. mentions of people, events, or any given target phenomenon – which a system returns that are correct
$precision = \frac{tp}{tp+fp}$
- Recall - Recall describes the proportion of all entities that potentially should be found, that a given system actually returns

$recall = sensitivity = \frac{tp}{tp+fn}$

- F1 - F score is derived from two summary measures: precision and recall.

$F - measure = \frac{(\beta^2+1)*precision*recall}{(\beta^2*precision*recall)}$.

When one system achieves higher recall or precision than another, it does not imply that the better-scoring system accurately reproduces the results of the other and then exceeds them; rather, just the overall selection of entities is more precise, or more comprehensive, in some way.

All three measures distinguish the correct classification of labels within different classes. Recall is a function of its correctly classified examples (true positives) and its misclassified examples (false negatives). Precision is a function of true positives and examples misclassified as positives (false positives). The F-score is evenly balanced when $\beta = 1$. It favours precision when$\beta 1$, and recall otherwise.

V. Results & Discussion

A. CatBooster

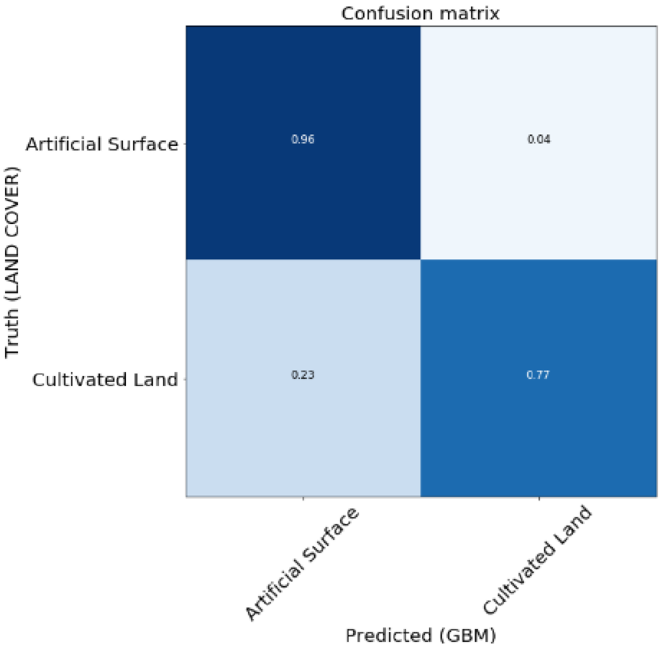
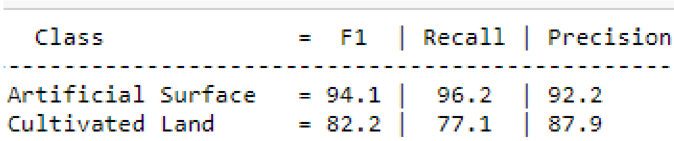


Fig. 8. Confusion Matrix of Catbooster Classifier



```
 Class                =  F1  | Recall | Precision
--------------------------------------------------
Artificial Surface    = 94.1 |  96.2  | 92.2
Cultivated Land       = 82.2 |  77.1  | 87.9
```

Fig. 9. F1, Recall and Precision scores of Catbooster Classifier

B. LightGBM

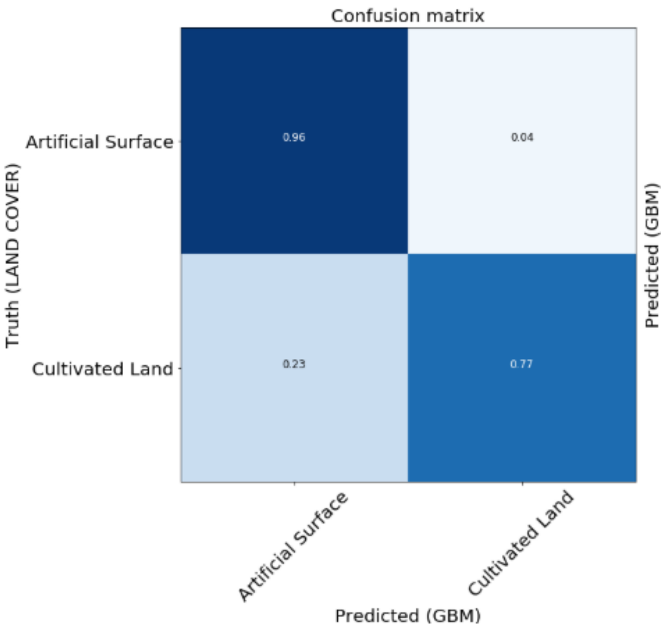
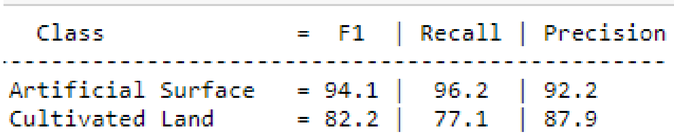


Fig. 10. Confusion Matrix of LightGBM Classifier



```
 Class                =  F1  | Recall | Precision
--------------------------------------------------
Artificial Surface    = 94.1 |  96.2  | 92.2
Cultivated Land       = 82.2 |  77.1  | 87.9
```

Fig. 11. F1, Recall and Precision scores of LightGBM Classifier

C. Random Forest

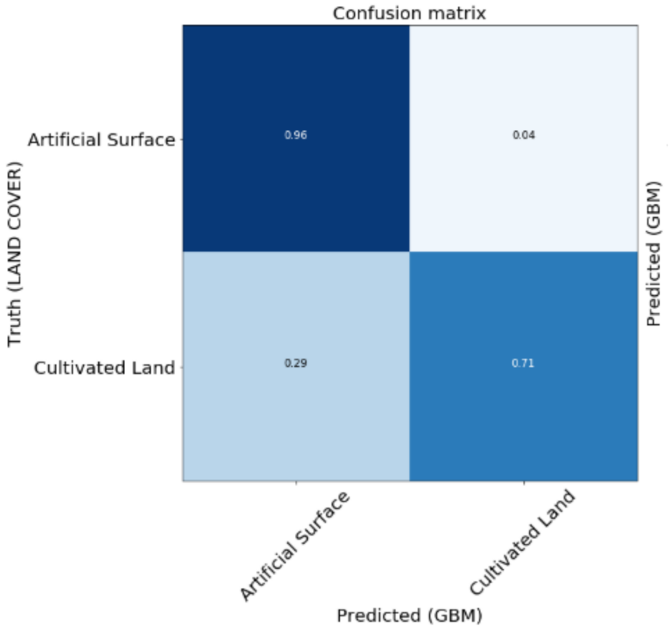
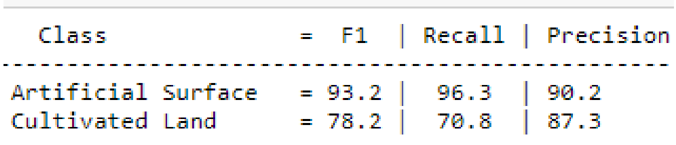


Fig. 12. Confusion Matrix of Random Forest Classifier



```
  Class                = F1   | Recall | Precision
--------------------------------------------------
 Artificial Surface    = 93.2 |  96.3  | 90.2
 Cultivated Land       = 78.2 |  70.8  | 87.3
```

Fig. 13. F1, Recall and Precision scores of Random Forest Classifier

D. Comparison of models

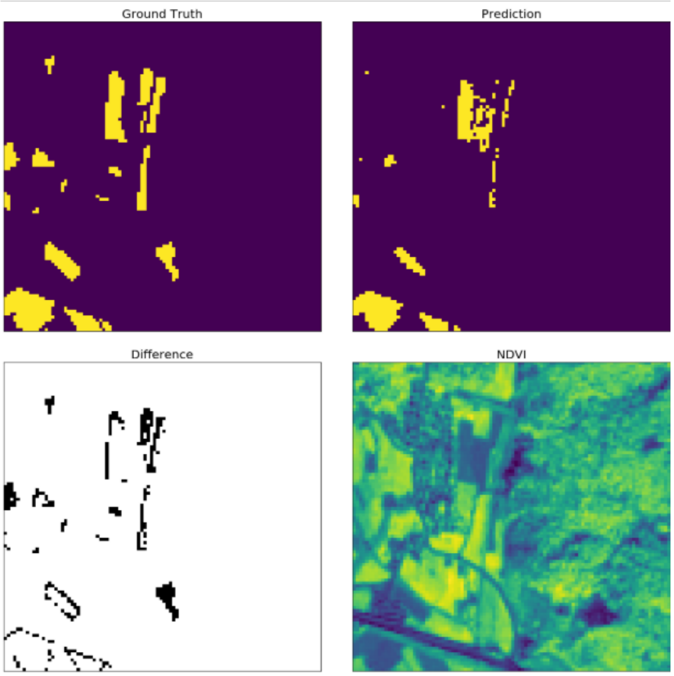
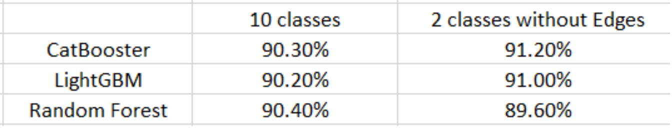


Fig. 14. Comparison of true and predicted images using Catbooster Classifier



	10 classes	2 classes without Edges
CatBooster	90.30%	91.20%
LightGBM	90.20%	91.00%
Random Forest	90.40%	89.60%

Fig. 15. Classification Accuracy of different models

From these observations we can see that all three classifiers gives almost same scores with Catboost giving slightly higher score. Along with scores, Catboost and LightGBM being Gradient Boosting Methods converges very fast than simple Random Forest. The false rate of all classifiers is very less which is a good sign for any model.

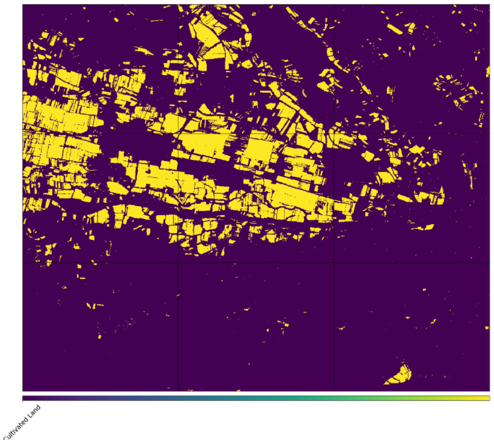


Fig. 16. Mask of cultivated and uncultivated land

References

[1] “Graficini podatki RABA za celo Slovenijo, howpublished = Available at http://rkg.gov.si/GERK(2019/09/27).”

[2] G. Ridgeway, “Generalized boosted models: A guide to the gbm package,” *Update*, vol. 1, no. 1, p. 2007, 2007.

[3] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “Lightgbm: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems*, 2017, pp. 3146–3154.

[4] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “Catboost: unbiased boosting with categorical features,” in *Advances in Neural Information Processing Systems*, 2018, pp. 6638–6648.

[5] F. Waldner, S. Fritz, A. Di Gregorio, and P. Defourny, “Mapping priorities to focus cropland mapping activities: Fitness assessment of existing global, regional and national cropland maps,” *Remote Sensing*, vol. 7, no. 6, pp. 7959–7986, 2015.

[6] M. Pal, “Random forest classifier for remote sensing classification,” *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.

[7] A. Darwish, K. Leukert, and W. Reinhardt, “Image segmentation for the purpose of object-based classification,” in *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477)*, vol. 3. Ieee, 2003, pp. 2039–2041.

[8] J. H. Friedman, “Stochastic gradient boosting,” *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[9] “eo-learn,” Available at https://github.com/sentinel-hub/eo-learn (2019/09/20).

[10] “Sentinelhub,” Available at https://www.sentinel-hub.com/ (2019/09/20).

[11] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, “Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4274–4288, 2018.

[12] D. Jackson, “Scalable vector graphics (svg): the world wide web consortium’s recommendation for high quality web graphics,” pp. 319–319, 2002.

[13] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” *Frontiers in neurorobotics*, vol. 7, p. 21, 2013.