Very-High-Resolution Object-Based Land Use-Land Cover Urban Classification using Extreme Gradient Boosting

Stefanos Georganos, Tais Grippa, Sabine Vanhuysse, Moritz Lennert,
Michal Shimoni, *Member*, *IEEE* and Eléonore Wolff

***Abstract*— In this letter the recently developed Extreme Gradient Boosting (Xgboost) classifier is implemented in a very-high-resolution (VHR) object-based urban Land Use-Land Cover application. In detail, we investigated the sensitivity of Xgboost to various sample sizes, as well as to feature selection (FS) by applying a standard technique, Correlation Based Feature Selection. We compared Xgboost with benchmark classifiers such as Random Forest (RF) and Support Vector Machines (SVM). The methods are applied to VHR imagery of two Sub-Saharan cities of Dakar and Ouagadougou and the village of Vaihingen, Germany. The results demonstrate that, Xgboost parametrized with a Bayesian procedure, systematically outperformed RF and SVM, mainly in larger sample sizes.**

***Index Terms*— Extreme gradient boosting (Xgboost), very-high-resolution classification, random forest (RF), feature selection (FS), support vector machine (SVM)**

I. Introduction

Non-parametric supervised machine learning (ML) classifiers such as Random Forest (RF) and Support Vector Machines (SVM) have been widely utilized in Geographic Object Based Image Analysis (GEOBIA) Land Use-Land Cover (LULC) mapping due to their effectiveness and ease to use [1]. For the past decade, their use has constantly been expanding and numerous studies have demonstrated their applicability [2], [3]. Recent studies have shown that more advanced variants of some of the aforementioned algorithms such as Rotation Forest ensembles, Logistic Model Trees and Canonical Correlation Forests, have exhibited superior results in several recent test cases [4]–[6]. On the other hand, advances in computer vision and the advent of Big Data have shifted the research to Artificial Neural Networks (ANN) and specifically to Convolutional Neural Networks (CNN) techniques, as an alternative to traditional supervised GEOBIA methods. Although it is generally established that most types of CNNs can be impressive with large volumes of training data at the cost of computational resources and model complexity [7], [8], most studies compare their results with other CNN techniques and not with GEOBIA [9], [10]. In fact, only a few studies have tackled the issue from a comparative perspective [11] but they focused on object detection and not on a multiclass LULC scheme. Moreover, they usually benchmarked their results with RF and SVM and not with more advanced ML implementations. Finally, limitations in training data have not taken under considerations as huge volumes of sample sizes are the exception rather than the rule. Consequently, in order to more realistically perform comparisons between deep learning methods and GEOBIA in the future, rigorous and new state-of-the art machine learning classifiers need to be investigated.

Another supervised classification technique, that belongs to the Classification and Regression Trees (CART) family is Gradient Boosting Machines (GBM) [8]. In the past, GBM and their variants have successfully been applied for several remote sensing (RS) applications such as species prediction [12], above ground biomass estimation [13] and scene classification [14]. Even though they have been shown to perform similarly to the rest of the state-of-the art classifiers [15], the effort and expertise needed to implement them overshadowed their predictive prowess for RS applications. The main reason is that traditional boosting algorithms are more prone to overfitting and have a larger number of parameters to optimize than machine learning algorithms such as RF and SVM.

Recently, Chen and Guestrin [16], published a new, regularized implementation of GBM, called Extreme Gradient Boosting (Xgboost). Since then, it has made a very strong impact in the machine learning community, being the winning solution of most machine-learning competitions [17]. A number of studies in other scientific fields have already demonstrated its superior performance compared to popular algorithms [18], [19]. Although preliminary results of its efficacy for VHR LULC were shown in [3], to our knowledge, this is the first systematic implementation of Xgboost in urban classification.

In this letter, we introduce Xgboost and evaluate its efficiency considering its sensitivity to sample size and feature selection (FS) using VHR datasets over the cities of Dakar and Ouagadougou along with the village of Vaihingen. A Bayesian-based parameter optimization method is followed to

Manuscript submitted November 03, 2017. This work was supported by BELSPO (Belgian Science Policy Office) in the frame of the STEREO III program – project REACT (SR/00/337).

S. Georganos, T. Grippa, S. Vanhuysse, M. Lennert and E. Wolff are with the Department of Geosciences, Environment and Society (DGES), Institute of Environmental Management and Spatial Planning (IGEAT), Universite Libre de Bruxelles, 1050 Bruxelles, Belgium (e-mails: sgeorgan /tgippa /svanhuysse/mlennert/ewolff@ulb.ac.be).

M. Shimoni is with the Signal and Image Center, Royal Military Academy, 1000 Bruxelles, Belgium (email: mshimoni@elec.rma.ac.be).

TABLE I
CLASSIFICATION SCHEME AND VALIDATION DATA FOR DAKAR, OUAGADOUGOU AND VAIHINGEN

Dakar (Senegal)		**Ouagadougou (Burkina Faso)**		**Vaihingen (Germany)**	
Class name	Validation set	Class name	Validation set	Class name	Validation set
Asphalt	107	Asphalt	80	Impervious surface	1000
Building type A	73	Building	351	Building	1000
Building type B	433	Bare Soil	312	Low vegetation	1000
Bare soil / light and dusty concrete	428	Tree	134	Tree	1000
Tree	117	Mixed bare soil/ vegetation	319		
Grass	142	Other vegetation	254		
Bush	130	Inland water	129		
Inland water	40	Swimming pool	107		
Shadow	113	Shadow	117		

maximize the performance of the algorithm. The results of the Xgboost classification, both in terms of accuracy and computational cost, are compared with traditional benchmark classifiers, namely RF and SVM.

MATERIALS AND METHODS

A. Ouagadougou, Burkina Faso

The first classification scheme is implemented in Ouagadougou, the capital of Burkina Faso and a major Sahelian city which has been facing an extensive urban growth since the last two decades. A pansharpened stereo WorldView-3 image (VNIR, 0.5m) collected in October 2015, as well as a normalized Digital Surface Model (nDSM) produced by stereo-photogrammetry on the same dataset, are used for this study. The city mostly consists of planned and unplanned residential buildings, commercial structures and natural materials.

B. Dakar, Senegal

Pleiades tri-stereo imagery (VNIR, 0.5m) of Dakar, Senegal, collected in July 2015, was also used for LULC classification along with the respective nDSM. The urban fabric is more diverse and complex than that of Ouagadougou, especially in the city centre.

C. Vaihingen, Germany

Finally, the openly accessible ISPRS Vaihingen (GE) dataset is employed. The dataset is characterized by various Germanic architectural styles and is consisted of orthophotos at 9 cm spatial resolution (NIR, R, G). The orthophotos are split into several referenced numbered tiles. We trained and tested the various classifiers in tile nº 13.

D. Image Segmentation and Training Samples

For segmentation, we use an open source, semi-automatic processing chain, recently developed in a Python environment, exploiting GRASS GIS [20]. The processing chain uses region growing segmentation, with Unsupervised Segmentation Parameter Optimization (USPO) as implemented in the i.segment.uspo module of GRASS [21] (Fig. 1). The classifications are trained using different number of classes for each dataset as described in Table I. The collection of training objects for Dakar and Ouagadougou is performed using random and stratified random sampling with strata defined using OpenStreetMap while the objects are validated and labeled by detailed visual interpretation. For Vaihingen, we created the training and validation samples by labeling each object in the image with the reference layer that is provided. Afterwards, we randomly extracted objects for training and validation. To reduce the bias of disproportional class sizes we used three fixed training sets of 20, 40 and 60 objects per class for Dakar and Ouagadougou. Due to the abundance of training data in the Vaihingen dataset, we investigated the sensitivity to sample size in more depth (between 50 and 400 objects per class) while we also used ten randomly drawn distributions of training objects for each sample size category, further increasing the inference power of the results.

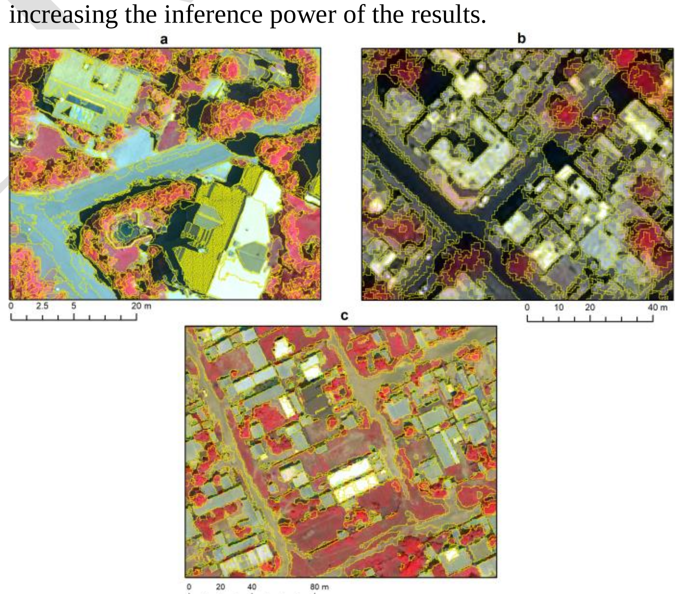

Fig. 1. Examples of different segmentations produced by the i.segment.uspo module of GRASS. Infrared false color composite image subsets of a) Vaihingen, b) a dense built-up area in Dakar, c) a sparsely built-up area in Ouagadougou.

E. Classification Algorithms

Along with Xgboost, which is presented in detail here after, two popular benchmarked classifiers, RF and SVM are selected. Comprehensive reviews of their operational framework and their applications can be found in [22] and [23]. All three classifiers are utilized through their R software implementations.

Xgboost is a regularized extension of traditional boosting ensemble techniques that belong to the CART family. As an ensemble tree-boosting method, its fundamental function predicts a new classification membership after each iteration. This is done in an additive way, meaning that the predictions are made from weak classifiers that constantly improve over the previous classifiers error. Incorrectly classified samples receive higher weights at the next step, forcing the classifier to focus on their performance in the following iterations. The final classification is the most vigorous as it includes the combined improvement of all the previous modelled trees. The learning of these classifiers is based on defining an objective function [24]. This function represents the *training loss* and the *regularization*. The former describes the predictive accuracy of the model while the latter the complexity. So far, GBM lacked a robust regularization factor, which made them susceptible to overfitting [25]. Xgboost overcomes this shortcoming by providing a stronger regularization framework, that constrains overfitting.

F. *Feature Selection*

We consider several features as initial input to the classification of each image. In the sub-Saharan regions of Dakar and Ouagadougou these include object descriptive statistics (mean, median, min, max, range, standard deviation, sum, first and third quartiles) for each spectral band, nDSM and NDVI, as well as geometrical covariates (object compactness, perimeter, fractal dimension and area) which total up to 60 features. With respect to the Vaihingen dataset, the same object statistics and covariates were computed for the spectral bands, NDVI, the wetness index (NDWI), as well as several textures of the Gray-level-co-occurrence matrix (GLCM) summed to 102 variables. These covariates were used as input to a typical supervised multiclass model for each classifier.

Recent studies have shown that in OBIA VHR classifications, machine learning classifiers are affected in different ways from FS techniques [26]. Consequently, in this study we employ a widely used standard FS method named Correlation Based Feature Selection (CFS) to assess the selected classifiers sensitivity to simple dimensionality reduction. CFS is a filter method which examines the relationships between covariates and the dependent variable through statistical metrics such as correlation or mutual entropy. It is often preferred due to its high computational efficiency albeit having the drawback of not capturing feature interaction, a phenomenon documented in CART based classifiers [27]. The CFS method creates an objective function which performs a greedy search by maximizing the dependency of a feature subset with the dependent variable, while minimizing the collinearity among the features of the subset [28] as shown in (1):

$$g(F)=FCi/IFC \quad (1)$$

where, *F is* a candidate feature subset, *g(F)* is the objective function, *FCi* is the average correlation between subset *F* and the dependent variable and *IFCi* is the average correlation within the subset *F*.

G. *Parameter Optimization*

The Xgboost parameters are optimized with a Bayesian optimization (BO) procedure through the *'rBayesianOptimization'* package in *R* as it has been demonstrated to be the most appropriate method for boosting classifiers [29]. As input, the algorithm requires the minimum and maximum boundaries for a given parameter. Afterwards, 10 model runs are constructed based on random combinations of a 10-fold cross-validated subset and the internal classification error is reported. From this point onwards, a Bayesian model is fit aiming to predict parameter combinations that minimize classification error using posterior probabilities. The procedure ends when a predefined number of iterations is reached (50 runs in our case). We predefined a total of 600 trees for the boosting to make sure that the optimization will be conservative and reduce the chances of overfitting. For the rest of the parameters such as the learning rate, tree depth and subsampling of training data and features, we used a wider range of values as input to the BO. For the SVM optimization, we construct a sequential grid search (GS) with a radial kernel function by applying exponentially growing sequences to identify the optimal cost and gamma parameters. Regarding RF parameters, an appropriate number of trees and number of features selected at each tree node by minimizing the Out of Bag Error (OOB). Finally, parameter optimization is performed both before and after FS. The training and testing of the classifiers is performed using an Intel Xeon E5-2690 16 core processor at 2.9 GHz with 96 GB of RAM.

II. RESULTS

The CFS feature selection identified 26, 22 and 16 covariates for Dakar, Ouagadougou and Vaihingen, respectively. To evaluate the efficiency of each classifier we computed the overall classification accuracy (OA) prior and after FS in each examined sample size.

A. *Dakar Results*

The results of the classification for Dakar are described in Table II and Fig.2. It can be observed that Xgboost's efficiency is more evident with larger sample sizes. For example, with the maximum sample size (N_{60}) it outperformed both RF (~ 1.5%) and SVM (~ 4%), regardless of feature selection. The highest recorded accuracy (OA=82.10%) is obtained when Xgboost trained with the full set of features. Nonetheless, Xgboost appears to be slightly negatively influenced by a reduced CFS feature subset, except for very small sample sizes (N_{20}). In contrast, RF and to a smaller extent SVM, appear to be benefited from CFS selection in all cases.

TABLE II

OVERALL ACCURACIES (%) FOR DAKAR USING DIFFERENT NUMBERS OF TRAINING OBJECTS. THE TRAINING TIME REFERS TO THE LARGEST SAMPLE SIZE (N_{60})

Training sample	SVM	SVM_{CFS}	RF	RF_{CFS}	Xgb	Xgb_{CFS}
N_{20}	74.65	75.85	77.12	**77.67**	74.81	75.56
N_{40}	77.12	77.31	78.33	79.4	**80.09**	79.25

N_{60}	77.88	78.07	80.33	80.88	**82.1**	81.78
Time (s)	0.25	**0.18**	1.51	1.01	3.62	2.21

TABLE III
OVERALL ACCURACIES (%) FOR OUAGADOUGOU USING DIFFERENT NUMBERS OF TRAINING OBJECTS. THE TRAINING TIME REFERS TO THE LARGEST SAMPLE SIZE (N_{60})

Training sample	SVM	SVM_{CFS}	RF	RF_{CFS}	Xgb	Xgb_{CFS}
N_{20}	**84.53**	83.53	82.57	82.55	83.42	84.23
N_{40}	86.74	85.64	85.83	86.19	87.21	**87.30**
N_{60}	88.39	88.02	86.78	86.69	**88.44**	88.10
Time (s)	0.29	0.21	1.75	0.82	4.15	1.93

B. Ouagadougou Results

In a similar fashion for Ouagadougou, Xgboost systematically outperformed for all but the smallest sample size (Table III). Xgboost before and after CFS, obtained the highest accuracy using medium and large sample sizes. SVM performed very good having similar accuracies with Xgboost. On the contrary, RF under performed in all the cases. In most cases, the classifiers were relatively unaffected by FS, with the exception of the first category, where for SVM the FS led to a drop of 1%.

C. Vaihingen Results

The results of the average and standard deviation of OA for 10 randomly sampled training distributions for each class size are reported in Fig 3. Xgboost displays higher accuracy in all different cases, with or without FS. SVM and RF largely underperformed in comparison to Xgboost with quasi 5% and 2% differences in OA. Generally, the standard deviation of the mean OA is decreasing as sample size increase for all classifiers. However, models with FS appear smoother in their trends potentially because noisy features that might affect stability are discarded. Regarding FS, it is clear that with a very large number of covariates (102), FS can be beneficial in small to moderate sample sizes. Critically, from 200 samples per class, Xgboost appears to increase in accuracy when the whole data are used, while SVM and RF remain the same.

III. DISCUSSION

Although far from extensive, the results demonstrate that Xgboost is a promising alternative to other state-of-the-art classifiers. Optimized through a Bayesian procedure, its performance for VHR LULC classification is highly encouraging in comparison to RF and SVM. Even at cases were typical CART classifiers such as RF underperformed (Ouagadougou), Xgboost managed to achieve the highest accuracy. On the negative side, the high number of trees, made the algorithm computationally inefficient in comparison to RF and SVM (Tables II – III). With respect to the training time, SVM was by far the fastest with RF and Xgboost following. It is suggested that Xgboost performs just as good if a lower amount of iterations is selected [19]. This indicates that computation cost may be reduced by optimizing the parameters around a smaller number of trees.

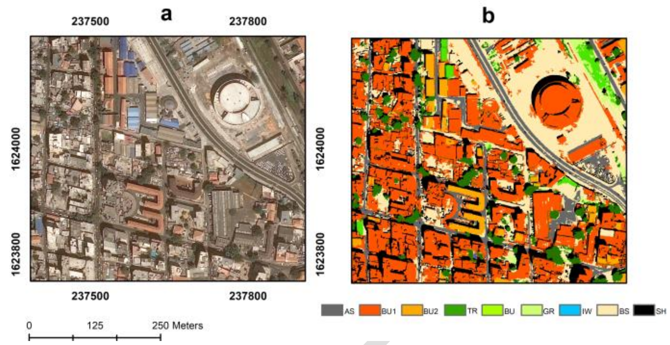

Fig. 2. a) Pleiades RGB composite of a central urban region in Dakar and b) LULC map of the same region using Xgboost with the largest sample size (N_{60}).

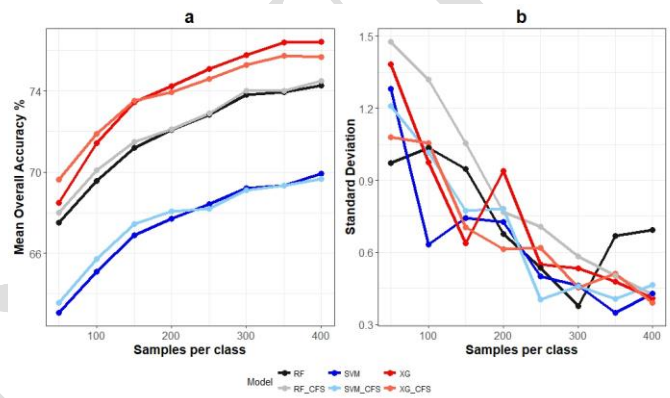

Fig. 3. Mean overall accuracy and standard deviation of 10 random distributions of training data for each sample size category

Another important facet of the applied methodology is that parameter optimization is performed for each classifier before and after FS. In most studies, optimization is performed only once. The effects of re-optimizing parameters when the number of input features is changed, is largely unexplored and subject of further research. Moreover, FS using CFS appeared to be systematically beneficial for relatively small amount of training objects in comparison to the amount of input covariates. This can be explained by the curse of dimensionality where the accuracy of prediction can be affected by the number of predictors [3]. Given that we used more than 100 covariates in the Vaihingen dataset, it is not surprising that even with 150 objects per class there was merit for performing FS. In larger sample sizes RF and SVM classifiers were largely unaffected but Xgboost exhibited increase in OA using the whole set of features. These results imply that the merit of using CFS as an FS method on Xgboost is beneficial for small to moderate sizes but can be harmful with large volumes of training data. As such, alternative selection methods should be investigated. In [26] it was demonstrated that instead of focusing on FS methods that produce single feature subsets, it is more appropriate to perform selection with methods that produce sequential feature rankings [26], [30]. Detailed methodologies regarding the GEOBIA processing along with documented code can be found in [20] while a simple example of the classification, parameter optimization and feature selection procedures can be found in the following github repository (https://github.com/ANAGEO/Xgboost_FeatureSelection_Optimization). Finally, it should be noted that Xgboost has recently been included in the list of classifiers of the v.class.mlR module in GRASS GIS.

IV. CONCLUSIONS

This study evaluates the implementation of Xgboost for VHR LULC urban classification. The results demonstrate that optimized Xgboost with a Bayesian model, consistently outperforms RF and SVM in different VHR datasets and classification schemes but at the cost of increased computational time. The improvement of Xgboost offers expands as the amount of training data increase. The benefits of CFS as FS method for reducing dimensionality are more beneficial for small sample sizes.

With respect to supervised machine learning algorithms, further research should compare Xgboost with more refined classifiers such as Rotation Forest [5] and other CART ensembles [4] that have been shown to outperform several traditional techniques. Moreover, comparative studies between GEOBIA methods employing Xgboost and deep learning techniques should be investigated. It is recommended the studies take into account computational burden and complexity, and training data limitations. Finally, more advanced FS methods, such as recursive feature elimination [28] or genetic algorithms should be examined, particularly for Xgboost which was shown to be sensitive to the results of dimensionality reduction.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their comments and suggestions which greatly improved the quality of the manuscript.

REFERENCES

[1] M. Li, L. Ma, T. Blaschke, L. Cheng, and D. Tiede, “A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 49, pp. 87–98, 2016.

[2] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, “A review of supervised object-based land-cover image classification,” *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 277–293, 2017.

[3] S. Georganos *et al.*, “Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application,” *GIScience Remote Sens.*, pp. 1–22, Nov. 2017.

[4] I. Colkesen and T. Kavzoglu, “Ensemble-based canonical correlation forest (CCF) for land use and land cover classification using sentinel-2 and Landsat OLI imagery,” *Remote Sens. Lett.*, vol. 8, no. 11, pp. 1082–1091, 2017.

[5] T. Kavzoglu, I. Colkesen, and T. Yomralioglu, “Object-based classification with rotation forest ensemble learning algorithm using very-high-resolution WorldView-2 image,” *Remote Sens. Lett.*, vol. 6, no. 11, pp. 834–843, 2015.

[6] I. Colkesen and T. Kavzoglu, “The use of logistic model tree (LMT) for pixel-and object-based classifications using high-resolution WorldView-2 imagery,” *Geocarto Int.*, vol. 32, no. 1, pp. 71–86, 2017.

[7] N. Mboga, C. Persello, J. R. Bergado, and A. Stein, “Detection of Informal Settlements from VHR Images Using Convolutional Neural Networks,” *Remote Sens.*, vol. 9, no. 11, p. 1106, 2017.

[8] B. Bischke, P. Helber, J. Folz, D. Borth, and A. Dengel, “Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks,” *arXiv Prepr. arXiv1709.05932*, 2017.

[9] M. Kampffmeyer, A.-B. Salberg, and R. Jenssen, “Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 1–9.

[10] V. Mnih and G. E. Hinton, “Learning to detect roads in high-resolution aerial images,” in *European Conference on Computer Vision*, 2010, pp. 210–223.

[11] E. Guirado, S. Tabik, D. Alcaraz-Segura, J. Cabello, and F. Herrera, “Deep-learning Versus OBIA for Scattered Shrub Detection with Google Earth Imagery: Ziziphus Lotus as Case Study,” *Remote Sens.*, vol. 9, no. 12, p. 1220, 2017.

[12] G. G. Moisen, E. A. Freeman, J. A. Blackard, T. S. Frescino, N. E. Zimmermann, and T. C. Edwards, “Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods,” *Ecol. Modell.*, vol. 199, no. 2, pp. 176–187, 2006.

[13] A. M. Filippi, İ. Güneralp, and J. Randall, “Hyperspectral remote sensing of aboveground biomass on a river meander bend using multivariate adaptive regression splines and stochastic gradient boosting,” *Remote Sens. Lett.*, vol. 5, no. 5, pp. 432–441, 2014.

[14] F. Zhang, B. Du, and L. Zhang, “Scene classification via a gradient boosting random convolutional network framework,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1793–1802, 2016.

[15] J. C. W. Chan and D. Paelinckx, “Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery,” *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2999–3011, 2008.

[16] T. Chen and C. Guestrin, “XGBoost : Reliable Large-scale Tree Boosting System,” *arXiv*, pp. 1–6, 2016.

[17] M. Graña and J. M. López-guede, *Advances in Intelligent Systems and Computing 527 International Joint Conference*. 2016.

[18] I. B. Mustapha and F. Saeed, “Bioactive molecule prediction using extreme gradient boosting,” *Molecules*, vol. 21, no. 8, pp. 1–12, 2016.

[19] R. P. Sheridan, W. M. Wang, A. Liaw, J. Ma, and E. Gifford, “Extreme Gradient Boosting as a Method for Quantitative Structure-Activity Relationships,” *J. Chem. Inf. Model.*, p. acs.jcim.6b00591, 2016.

[20] T. Grippa, M. Lennert, B. Beaumont, S. Vanhuysse, N. Stephenne, and E. Wolff, “An Open-Source Semi-Automated Processing Chain for Urban Object-Based Classification,” *Remote Sens.*, vol. 9, no. 4, p. 358, 2017.

[21] M. . Lennert and G. D. Team, “Addon i.segment.uspo,” *Geogr. Resour. Anal. Support Syst. SoftwareVersion 7.3*, 2016.

[22] G. Mountrakis, J. Im, and C. Ogole, “Support vector machines in remote sensing: A review,” *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011.

[23] M. Belgiu and L. Drăgut, “Random forest in remote sensing: A review of applications and future directions,” *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.

[24] T. Chen and T. He, “xgboost : eXtreme Gradient Boosting,” pp. 0–3, 2015.

[25] M. Sewell, “Ensemble Learning,” *IEEEACM Trans. Comput. Biol. Bioinforma. IEEE ACM*, vol. 2007, no. April 2007, pp. 1–16, 2008.

[26] S. Georganos, T. Grippa, S. G. Vanhuysse, M. Lennert, and E. Wolff, “Optimizing classification performance in an object-based very-high-resolution land use-land cover urban application,” in *Remote Sensing Technologies and Applications in Urban Environments II*, vol. 10431.

[27] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *J. Mach. Learn. Res.*, vol. 3, no. Mar, pp. 1157–1182, 2003.

[28] L. Ma *et al.*, “Evaluation of Feature Selection Methods for Object-Based Land Cover Mapping of Unmanned Aerial Vehicle Imagery Using Random Forest and Support Vector Machine Classifiers,” *ISPRS Int. J. Geo-Information*, vol. 6, no. 2, p. 51, 2017.

[29] Y. Xia, C. Liu, Y. Li, and N. Liu, “A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring,” *Expert Syst. Appl.*, vol. 78, pp. 225–241, 2017.

[30] M. Pal and G. M. Foody, “Feature selection for classification of hyperspectral data by SVM,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, 2010.