A Booster Analysis of Extreme Gradient Boosting for Crop Classification using PolSAR Imagery

Mustafa Ustuner
Department of Geomatic Engineering
Yildiz Technical University
Istanbul, Turkey
mustuner@yildiz.edu.tr

Fusun Balik Sanli
Department of Geomatic Engineering
Yildiz Technical University
Istanbul, Turkey
fbalik@yildiz.edu.tr

Saygin Abdikan
Department of Geomatics Engineering
Zonguldak Bulent Ecevit University
Zonguldak, Turkey
sabdikan@beun.edu.tr

Gokhan Bilgin
Department of Computer Engineering
Yildiz Technical University
Istanbul, Turkey
gbilgin@yildiz.edu.tr

Cigdem Goksel
Department of Geomatics Engineering
Istanbul Technical University
Istanbul, Turkey
goksel@itu.edu.tr

Abstract*—This study evaluates the impacts of three booster types (two tree-based and one linear model) in extreme gradient boosting (XGBoost) for crop classification using multi-temporal PolSAR (Polarimetric Synthetic Aperture Radar) images. Ensemble learning algorithms have received great attention in remote sensing for classification due to their greater performance compared to single classifiers in terms of accuracy. Extreme gradient boosting is the regularized extension of traditional boosting techniques and could overcome the overfitting constrain of gradient boosting (a.k.a gradient boosting machine). Three types of booster which are linear booster, tree booster and DART (Dropouts meet Multiple Additive Regression Trees) booster were tested on XGBoost for crop classification. From the multi-temporal PolSAR data, two types of polarimetric dataset (linear backscatter coefficients and Cloude–Pottier decomposed parameters) were extracted and incorporated into the classification step. The impacts of polarimetric features for crop classification were also analyzed in detailed besides exploring the boosting types of XGBoost. Our experimental results demonstrated that tree booster and DART booster were found to be superior compared the linear booster in terms of overall classification accuracy for both polarimetric dataset. The highest classification accuracy (87.97%) was achieved by tree booster with linear backscatter coefficients. Furthermore, linear backscatter coefficients achieved higher performance with respect to Cloude–Pottier decomposition in terms of classification accuracy.

Keywords*— *extreme gradient boosting, polarimetry, PolSAR, crop classification, agriculture

I. INTRODUCTION

Crops, by their nature, are significantly affected by climate change and accordingly have very dynamic and heterogeneous pattern in terms of spatio-temporal domain. These changes requires the mapping of the crop distribution and changes over time for the sustainable management and development of croplands [1,2]. For this matter, remote sensing technology provides unique advantages with its cost-effective solutions and synoptic viewing capability in temporal domain. In particular, Synthetic Aperture Radar (SAR) images could be of service for timely monitoring of croplands when cloud-free optical sensors are limited to be operational [3]. SAR signals are sensitive to crop structural characteristics and dielectric properties, and hence SAR images could provide unique textural and geometric information for crops. PolSAR images, with respect to single or dual polarized images, provides richer information content in terms of different types of scattering since crops are polarization-dependent [4,5].

PolSAR images are complex-valued data where each pixel has complex-valued vectors, required to be projected (or converted) to the real domain [6]. Such projection makes the interpretation and analysis of PolSAR data more comprehensible and transferable. For this purpose, polarimetric target decompositions or PolSAR-spesific feature extraction methods (radar vegetation index, total scattering power, pedestral height, polarimetric discriminators etc.) are commonly used [7,8]. In this study, we chose the polarimetric target decomposition for this analysis.

Polarimetric target decompositions are grouped into two subcategories: (i) coherent and (ii) incoherent decompositions. Coherent decompositions based on Sinclair scattering matrix while incoherent decomposition based on coherency/covariance matrix. For distibuted targets (i.e. natural targets), incoherent target decompositions are more suitable to characterize the complex scattering mechanism. Incoherent decompositions are divided into two subcategories: (i) model-based (e.g., Freeman–Durden or Yamaguchi) and (ii) eigenvalue-eigenvector-based decompositions (Cloude–Pottier or Touzi). In this study, we implemented Cloude–Pottier (a.k.a H/A/α where H: entropy, A: anisotropy and α: alpha angle) decomposition that provides unique mathematical outputs to be interpreted for understading the complex scattering mechanism. Cloude–Pottier decomposition do not provide the direct features representing the different scattering types for targets however extracted polarimetric features have to be analysed and interpreted. The entropy measures randomness degree of the scattering and varies between 0-1. The alpha angle indicates the scattering types for the target and values from 0 to 90 degrees. The anisotropy describes the relative importance between second and third eigenvalues [7,9].

Ensemble learning algorithms (a.k.a multiple classifiers) have been of great interests by data scientists because of their superiority compared to single classifiers in terms of classification performance and computational time [10,11]. In the last few year, recently developed ensemble learning algorithms have been tested in remote sensing image classification, such as canonical correlation forest (CCF), extreme gradient boosting (XgBoost), Light Gradient Boosting Machine (LightGBM), ForestPA and deep forest [12-17].

978-1-7281-2116-1/19/$31.00 ©2019 IEEE

In this study, we preferred the XgBoost for classification since XgBoost gained the great popularity in data science and found to be the winning solutions of many machine learning competitions [18].

This study investigates the impacts of three booster types (two tree-based and one linear model) in extreme gradient boosting (XGBoost) for crop classification using multi-temporal PolSAR images. Furthermore, extracted two polarimetric features (linear backscatter and polarimetric decomposed features) were compared in terms of overall classification accuracy.

The remainder of this article is organized as follows. Section 2 describes the study area, data characteristics and classification algorithm. Section 3 presents and discusses the experimental results. Conclusions and future research topics are provided in Section 4.

II. METHODOLOGY

A. *Study area and Data*

The study area is located in Konya, Turkey and approximately situated approximately 65 km north of Konya city center (see Figure 1). The study area is flat and favorable for precision agriculture, that is covered from only with croplands including maize, potato, wheat, sunflower and alfalfa . RGB composite image of the study area can be seen in Figure 1.

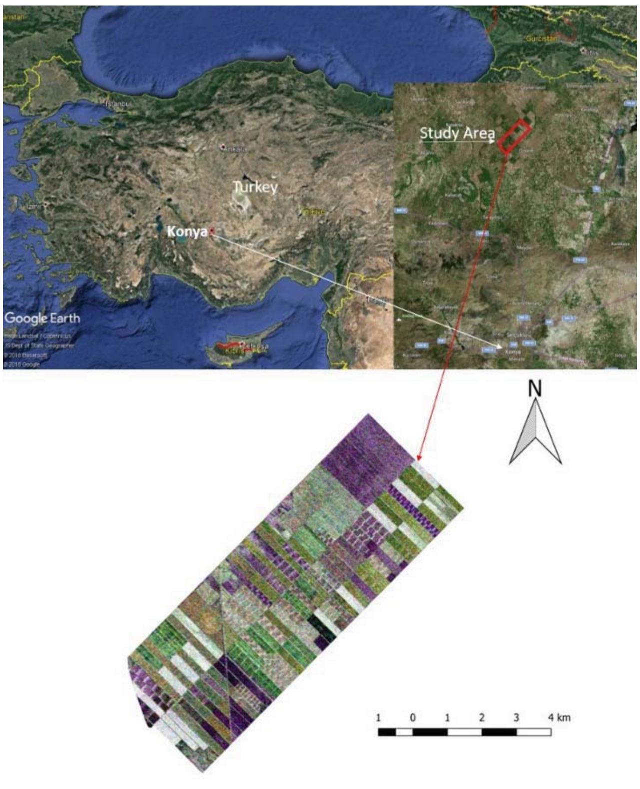

Figure 1: Study area

A total of four fine-quad polarimetric mode (FQW) RADARSAT-2 PolSAR images were acquired from June 13 to August 24 in 2016 for the study site. The data were acquired in single look complex data format to able to generate polarimetric features using decomposition methods. The resolution of the imagery is 4.7 m × 5.1 m in the range and azimuth direction, respectively. The specifications of the RADARSAT-2 images are presented in Table 1.

TABLE I. IMAGE SPECIFICATIONS

Image Specifications	
Sensor Type	RADARSAT-2
Wavelength	C band- 5.6 cm
Resolution	4.7 m × 5.1 m (range × azimuth)
Incidence Angle	40^0
Pass Direction	Descending
Acquisition Type	Fine Quad Pol
Polarization	Full Polarimetric
Nominal Scene Size	25 km × 25 km
Acquisition Times	June 13, July 7, July 31 and August 24 in 2016
Product Type	Single Look Complex

For each data acquisition time, field work was carried out and ground truth information was collected. Table II shows the ground truth information (incl. training and testing data) and crop type information for the study area.

TABLE II. GROUND TRUTH INFORMATION

Class	Training Set (in pixel)	Testing Set (in pixel)
Alfalfa	1918	3542
Maize	5581	14217
Potato	2274	10604
Sunflower	3524	6338
Wheat	3729	8915

RGB composite images of the study area for each acquisition time could be seen in Figure 2. These images are the final product of pre-processed data from the linear backscatter.

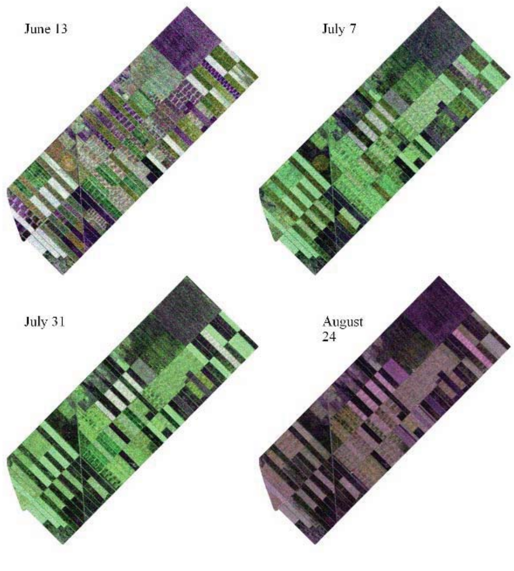

Figure 2: Study area

B. *Extreme Gradient Boosting*

XgBoost received the great popularity and attention recently in data science since it was the first choice of teams winning the many machine learning competitions at Kaggle[18]. It is an optimized distributed gradient-boosting machine learning library that was developed by Tianqi Chen and Carlos Guestrin from University of Washington [19]. XgBoost is also called as a scalable end-to-end tree boosting system where developers proposed a novel sparsity-aware algorithm and weighted quantile sketch for the learning process[19]. The popularity of XgBoost comes from its faster processing time and higher performance compared to existing boosting frameworks where the underlying factor is XGBoost's scalability in all scenarios. The parameters of XGBoost used in our experiment are listed in Table III.

TABLE III. XGBOOST PARAMETERS

Parameter	Value
nrounds	100
learning_rate	0.05
gamma	0
max_depth	5
min_child_weight	1
subsample	0.8
colsample_bytree	0.8

XgBoost has been tested in only a few studies in remote sensing for the classification of SAR data. Dong et al. [20] tested the Chinese Gaofen-3 PolSAR images for land cover classification using polarimetric spatial information and XGBoost. In terms of polarimetric features, they also used the polarimetric features from Cloude–Pottier. They reached up to 89% classification accuracy. Jiang et al. [21] investigated the Sentinel-1 and Sentinel-2 time series data for the early season mapping of sugarcane using random forest classification and XGBoost. Their experimental results demonstrated XGBoost showed greater performance than random forest classification in terms of computation speed.

III. RESULTS AND DISCUSSION

In this section, overall classification, kappa coefficient as well as class-based accuracies (based on F1-scores) are presented.

TABLE IV. OVERALL CLASSIFICATION ACCURACIES

Classification Accuracy (%)		
Booster Type	**Linear Backscatter**	**Cloude–Pottier decomposition**
Tree booster	87.97	78.72
DART booster	87.89	78.78
Linear booster	80.12	61.73
Kappa coefficient		
Tree booster	0.8411	0.7178
DART booster	0.8400	0.7187
Linear booster	0.7345	0.4657

The highest classification accuracy (87.97%) was achieved by tree booster with linear backscatter coefficients, followed DART booster (87.89%) linear booster (80.12%) for both with linear backscatter coefficients. For all booster types, linear backscatter coefficients achieved higher performance with respect to Cloude–Pottier decomposition in terms of classification accuracy. The classified images are depicted in Figure 3.

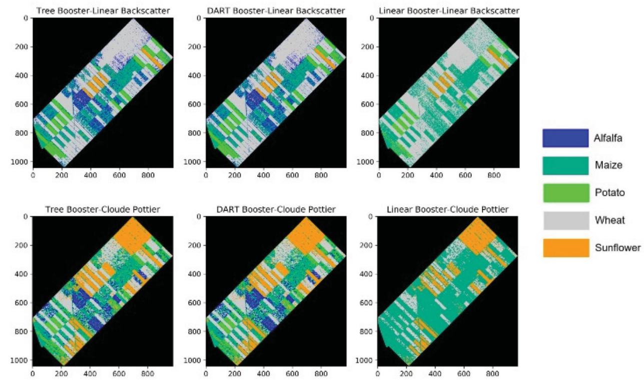

Figure 3: Classified Images

The class-based accuracies (based on F1-scores) for linear backscatter and Cloude-Pottier decomposition are listed in Table V and VI, respectively.

TABLE V. F1-SCORES FOR LINEAR BACKSCATTER

Class	Tree Booster	DART Booster	Linear Booster
Alfalfa	0.37	0.36	0.00
Maize	0.86	0.86	0.81
Potato	0.94	0.93	0.80
Sunflower	0.95	0.95	0.93
Wheat	0.97	0.97	0.82

TABLE VI. F1-SCORES FOR CLOUDE–POTTIER DECOMPOSITION

Class	Tree Booster	DART Booster	Linear Booster
Alfalfa	0.27	0.27	0.00
Maize	0.76	0.76	0.63
Potato	0.70	0.70	0.05
Sunflower	0.99	0.99	0.93
Wheat	0.95	0.95	0.85

The linear backscatter outperformed the Cloude-Pottier decomposition in terms of the F1-score values for alfalfa, maize, potato and wheat whereas the Cloude-Pottier decomposition yielded higher F1-score value than for linear backscatter for sunflower. This contradiction shows that crops have various scattering characteristics and hence the impact in the classification accuracy. It was also noted that alfalfa could not be classified through linear booster for both linear backscatter and Cloude-Pottier decomposition. The reason of this failure might be the less number of training data for alfalfa compared to other crop types in our experimental study.

IV. CONCLUSION

This paper investigated the comparative performance of booster types for XGBoost classification for crop classification. Also, class-based comparative analysis was conducted to investigate how crops behaves for each booster types.

Our experimental results demonstrated that tree-based booster types provided similar accuracies to each other and outperformed linear booster. The highest classification accuracy (87.97%) was achieved by tree booster with linear backscatter coefficients When F1-score values were investigated, it was concluded that linear backscatter outperformed the Cloude-Pottier decomposition except sunflower. The main drawback of the XGBoost algorithm is the high number of parameters and their optimization. Authors addresses the need of in-depth parameter analysis of XGBoost to determine the impact of each parameter on classification accuracy. Our future research will focus on testing the model based decomposition with XGBoost for crop classification

ACKNOWLEDGMENT

This research has been funded by Yildiz Technical University, Scientific Research Projects Office [Project No: FBA-2017-3062]. The authors gratefully acknowledge the data support from TAGEM (Project No: TAGEM/TSKAD/14/A13/P05/03)

REFERENCES

[1] H. McNairn and J. Shang, "A Review of Multitemporal Synthetic Aperture Radar (SAR) for Crop Monitoring," in *Multitemporal Remote Sensing: Methods and Applications*, Y. Ban, Ed., ed Cham: Springer International Publishing, 2016, pp. 317-340.

[2] A. Larrañaga and J. Álvarez-Mozos, "On the Added Value of Quad-Pol Data in a Multi-Temporal Crop Classification Framework Based on RADARSAT-2 Imagery," *Remote Sensing,* vol. 8, p. 335, 2016.

[3] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geoscience and Remote Sensing Magazine,* vol. 1, pp. 6-43, 2013.

[4] J. Soria-Ruiz, Y. Fernandez-Ordonez, H. McNairm, and J. Bugden-Storie, "Corn monitoring and crop yield using optical and RADARSAT-2 images," in *2007 IEEE International Geoscience and Remote Sensing Symposium*, 2007, pp. 3655-3658.

[5] H. Tamiminia, S. Homayouni, H. McNairn, and A. Safari, "A particle swarm optimized kernel-based clustering method for crop mapping from multi-temporal polarimetric L-band SAR observations," *International Journal of Applied Earth Observation and Geoinformation,* vol. 58, pp. 201-212, 2017/06/01/ 2017.

[6] R. Hänsch and O. Hellwich, "Skipping the real world: Classification of PolSAR images without explicit feature extraction," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 140, pp. 122-132, 2018/06/01/ 2018.

[7] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 34, pp. 498-518, 1996.

[8] Y. X. Lim, M. S. Burgin, and J. J. v. Zyl, "An Optimal Nonnegative Eigenvalue Decomposition for the Freeman and Durden Three-Component Scattering Model," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, pp. 2167-2176, 2017.

[9] X. Jiao, J. M. Kovacs, J. Shang, H. McNairn, D. Walters, B. Ma, and X. Geng, "Object-oriented crop mapping and monitoring using multi-temporal polarimetric RADARSAT-2 data," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 96, pp. 38-46, 2014/10/01/ 2014.

[10] T. Kavzoglu and I. Colkesen, "An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping," *International Journal of Remote Sensing,* vol. 34, pp. 4224-4241, 2013/06/20 2013.

[11] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 114, pp. 24-31, 2016/04/01/ 2016.

[12] J. Xia, N. Yokoya, and A. Iwasaki, "Hyperspectral Image Classification With Canonical Correlation Forests," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 55, pp. 421-431, 2017.

[13] S. Georganos, T. Grippa, S. Vanhuysse, M. Lennert, M. Shimoni, S. Kalogirou, and E. Wolff, "Less is more: optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application," *GIScience & Remote Sensing,* vol. 55, pp. 221-242, 2018/03/04 2018.

[14] S. Georganos, T. Grippa, S. Vanhuysse, M. Lennert, M. Shimoni, and E. Wolff, "Very High Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting," *IEEE Geoscience and Remote Sensing Letters,* vol. 15, pp. 607-611, 2018.

[15] M. Ustuner and F. Balik Sanli, "Polarimetric Target Decompositions and Light Gradient Boosting Machine for Crop Classification: A Comparative Evaluation," *ISPRS International Journal of Geo-Information,* vol. 8, p. 97, 2019.

[16] A. Samat, S. Liu, C. Persello, E. Li, Z. Miao, and J. Abuduwaili, "Evaluation of ForestPA for VHR RS image classification using spectral and superpixel-guided morphological profiles," *European Journal of Remote Sensing,* vol. 52, pp. 107-121, 2019/01/01 2019.

[17] Y. Boualleg, M. Farah, and I. R. Farah, "Remote Sensing Scene Classification Using Convolutional Features and Deep Forest Classifier," *IEEE Geoscience and Remote Sensing Letters,* pp. 1-5, 2019.

[18] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A Data-Driven Design for Fault Detection of Wind Turbines Using Random Forests and XGboost," *IEEE Access,* vol. 6, pp. 21020-21031, 2018.

[19] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794..

[20] H. Dong, X. Xu, L. Wang, and F. Pu, "Gaofen-3 PolSAR Image Classification via XGBoost and Polarimetric Spatial Information," *Sensors (Basel, Switzerland),* vol. 18, p. 611, 2018.

[21] H. Jiang, D. Li, W. Jing, J. Xu, J. Huang, J. Yang, and S. Chen, "Early Season Mapping of Sugarcane by Applying Machine Learning Algorithms to Sentinel-1A/2 Time Series Data: A Case Study in Zhanjiang City, China," *Remote Sensing,* vol. 11, p. 861, 2019.