# FOREST TYPE CLASSIFICATION USING MORPHOLOGICAL OPERATORS AND FOREST PA METHOD

M. Ustuner[1] *, F. B. Sanli[1], S. Abdikan[2]

[1] Dept. of Geomatic Engineering, Yildiz Technical University, Istanbul, Turkey - (mustuner, fbalik)@yildiz.edu.tr
[2] Dept. of Geomatic Engineering, Bulent Ecevit University, Zonguldak, Turkey – sabdikan@beun.edu.tr

**KEY WORDS:** Morphological Operators, Forest PA, Support Vector Machines, Random Forests

**ABSTRACT:**

Morphological operators have obtained great attention in the fields of image processing and pattern recognition due to the providing relevant spatial information for the classification. This study evaluates the impacts of morphological operators from dual-polarized, Advanced Land-Observing Satellite (ALOS) and Phased Array L-band Synthetic Aperture Radar (PALSAR) data for the classification of forested areas. To this aim, the opening and closing operators with different size of structuring elements were exploited as morphological features in our study. For the classification of the forested areas, three different classification models (Support Vector Machines, Random Forests and Forest PA) were performed. The experimental results show that morphological features have increased the classification accuracy by 6.82% and 10.36% for Forest PA (Forest by Penalizing Attributes) and Random Forests (RF), respectively, while decreased the accuracy by 1.43% for Support Vector Machines (SVM). Furthermore, Forest PA yields the highest accuracy with morphological features followed by RF.

## 1. INTRODUCTION

The temporal monitoring of forest distribution and accurate mapping of forest types are essential for decision makers in climate regulation, environmental planning and management for forest biodiversity. Due to the high coverage of clouds in the tropical to sub-tropical regions, SAR (Synthetic Aperture Radar) sensors are only optimal way to monitor the large forestry areas in those regions with the advantages of penetration through clouds. L-band SAR data have been found more applicable for forestry application rather than X and C bands since longer wavelength and deeper penetration capability. Deeper penetration into the canopy provides greater characterisation of the multiple scattering mechanism. (Baltzer et al., 2003). Microwave signals are more sensitive to forest canopy size and structure and therefore SAR sensors have been found more useful and advantageous for forestry applications (forest type classification, detection of deforestation, forest volume estimation etc.) compared to optical sensors.

Many studies have been reported the use of L-band SAR data for forest cover/type mapping. Thiel et al. (2006) used the L-band SAR data (JERS-1) for forest cover mapping in five different test sites. Liesenberg and Gloaguen (2013) investigated the polarimetric, interferometric and textural features from the ALOS PALSAR data for forest classification. Mitchell et al. (2014) tested the C and L-band SAR images for forest cover mapping and also investigated their interoperability. Middinti et al. (2017) explored the L-band SAR data and textural features for the classification of forest types by using support vector machines. Abdikan (2018) assessed the performance of optical and L-band SAR data for the classification of forested areas.

To able to benefit from the spatial characteristics of the objects for the classification process, several methods have been used so far like image segmentation and mathematical morphology which is a non-linear image processing technique. The morphological operators (opening, closing, erosion and dilation) try to model the spatial characteristics of the object (shape, orientation etc.) given in a structuring element. (Bioucas-Dias et al., 2013). Many research have been carried out for the spatial-spectral classification of high-resolution multi/hyperspectral by using the morphological operators however, seldom performed for SAR (dual or quad pol) data, especially for forest classification.

Only few studies explored the benefits of morphological operators for the classification purposes of SAR/PolSAR data. Du et al. (2015) combined the polarimetric and spatial features for the land use classification of PolSAR data by using RF, rotation forest and SVM. Wang et al. (2017) proposed a new method called “composite kernel” that benefits from both spatial and polarimetric information for the PolSAR data classification. Wurm et al. (2017) investigated the spatial features (textural, morphological and polarimetric) for slum mapping using dual-polarized SAR data. Recently, Samat et al. (2018) explored the potential use of polarimetric signatures and morphological features for mapping the halophyte plants in wetland environments by using their newly proposed classifiers. Samat et al. (2018) also reported that higher classification accuracies were obtained when using morphological features. All these above studies illustrated the spatial features by exploiting of mathematical morphology for the classification purposes however none of them addressed the forest type classification.

The main objective of this research is to test the impacts of the morphological operators on the classification accuracy for the classification of forested areas (including forest types). For this purpose, we implemented three different model (SVM, RF and Forest PA). To the best of our knowledge, this is the first study exploiting Forest PA method for SAR data classification in the field of remote sensing and pattern recognition.

The major contributions of our experimental study can be shortly summarized as follows.

* Corresponding author

(1) We tested Forest PA for the first time in remote sensing for the classification of SAR imagery.
(2) The performance of Forest PA in comparison to RF and SVM was evaluated for classification of forested areas using morphological features

## 2. MATERIALS AND METHODS

### 2.1 Study Area and Dataset

The study area is located in Zonguldak city of Turkey. There are several hard coal mining areas in the region and city is surrounded by dense forests.

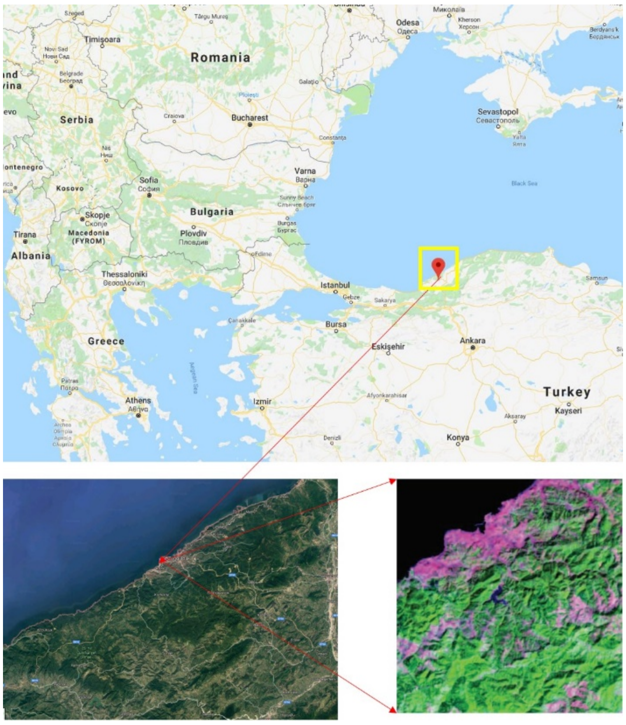


Figure 1. Study area (ALOS PALSAR composite image)

Since L-band SAR data can penetrate deeply due to its long wavelength (23 cm), it was preferred for this experimental study. The data has dual polarization (HH and HV, H: Horizontal, V: Vertical). The specifications of the data are presented in Table 1.

| Specifications | Description |
|---|---|
| Date | 10.09.2009 |
| Sensor | PALSAR/FBD |
| Spatial Resolution | 15 m |
| Flight Direction | Ascending |
| Polarization | HH+HV |
| Wavelength | 23 cm (L-Band) |

Table 1. Data Specifications

The data was acquired in Single-Look-Complex (SLC) format that needs to be processed before the classification. The pre-processing of the data includes the following steps: multi-looking, speckle filtering (3x3 Gamma Map), terrain correction (30 m ASTER GDEM) and scaling the data into decibel. As a final step of pre-processing, the data was exported as GeoTIFF format with 15 meter spatial resolution. Two different dataset were exploited to test the impact of morphological features (Table 2). Dataset I has two bands and Dataset II has ten bands. Morphological features were defined by using the opening profile and closing profile with different size of structuring elements. The size of the structuring elements are 3×3 and 5×5.

| Dataset | Description |
|---|---|
| I | Original Bands (HH+HV) |
| II | Original Bands + Morphological Features |

Table 2. Dataset description

Our study area consists of five classes, including bare land, deciduous forest, deciduous mixed forest, mixed forest and urban areas. Table 3 demonstrates the corresponding number of training and test samples for the classes.

| Class Name | Number of Samples | |
|---|---|---|
| | Training | Test |
| Bare land | 400 | 392 |
| Deciduous forest | 1467 | 1451 |
| Deciduous mixed forest | 1741 | 1452 |
| Mixed forest | 1366 | 846 |
| Urban Areas | 737 | 492 |

Table 3. The number of training and testing samples

The accuracy of classified images was assessed using error matrix that is used to calculate overall accuracy. Overall accuracy which is calculated by summing correctly classified values and dividing it by the total number of values is formulated as follows.

$$Overall\ Accuracy = \frac{Correctly\ classified\ values}{Total\ number\ of\ values}$$

### 2.2 Morphological Operators

For the extraction of relevant spatial information prior to image classification, there are several approaches employed so far and mathematical morphology (morphological operators, morphological profiles, extended morphological profiles etc.) is one of these approaches. The main purpose of the mathematical morphology is to understand the spatial relationship between the pixels within a structuring element that is also known as a kernel (e.g. a 5×5 square). The fundamental operators of mathematical morphology are erosion and dilation. The erosion basically enlarges the particular (selected) objects which is darker than their surroundings, however dilation makes them smaller. Through the combination of erosion and dilation, both opening and closing operators are derived. The opening operator is defined as the erosion following by dilation and the closing operator is defined as first dilation and then the erosion for a particular pixel of an image (Fauvel et al., 2013; Bioucas-Dias et al., 2013; Samat et al., 2018). In this research, we used 3×3 and 5×5 size of structuring elements for both opening and closing operators.

### 2.3 Image Classification

In this study, we implemented three different classification models (SVM, RF and Forest PA). SVM and RF, which are two of the popular machine learning algorithms, have been extensively used for the classification and regression problems in the fields of pattern recognition and remote sensing. Forest PA which is a new decision forest algorithm has been proposed by Adnan and Islam (2017) and was tested for the classification of the freely available data sets from the UCI Machine Learning Repository. In the following paragraph, the short overview of the classification models are presented. More details for SVM and RF can be found in Melgani and Bruzzone (2004) and Pal (2005), respectively.

Support vector machine which is one of the kernel-based learning algorithms tries to separate the two classes by defining the optimal hyperplane. SVM uses the kernel tricks for non-linear classification and needs user-defined parameters to be set for kernels (Melgani and Bruzzone, 2004). We implemented radial basis function (RBF) kernel for SVM classification. Optimum parameters for RBF kernel (gamma and cost parameters) were determined by using grid search algorithm. Random Forest which is one of the decision tree-based ensemble classifiers produce multiple decision trees and utilize the majority voting strategy to assign a class label to the unknown instance. RF uses a random subset of training samples and variables for producing the multiple decision trees. RF is popular for the classification and regression problems because of its high classification accuracy and low computational cost (Pal, 2005). Forest PA (Forest by Penalizing Attributes) is a novel decision forest algorithm, recently introduced to the community by Adnan and Islam (2017). Forest PA benefits from the power of all the non-class attributes in a dataset to create the high accurate decision trees and has novel features like weight assignment strategy and bootstrap sampling. More technical details for Forest PA can be found in Adnan and Islam (2017) and Siers and Islam (2018). SVM and RF classifications were performed using the open-source Scikit-learn module in Python v3.6.4 (Pedregosa et al., 2011) and Forest PA classification was performed using the open-source data mining software Weka 3.8.2 (Frank et al. 2016).

## 3. RESULTS AND DISCUSSION

Each classification model has different sensitivity and attributes in terms of classification accuracy for the same dataset due to the main differences at learning phase of the model. In this section, we will present the classification results for the dataset-I and dataset-II and discuss them with underlying reasons. The overall accuracies varied from 39.33% to 50.66%. The overall accuracies for dataset-I were presented in Figure 3. The highest accuracy was obtained by SVM while the lowest one was received by RF for the dataset-I. The highest accuracy for dataset-I was reached up to 44.53% which is still less than 50% and may not be acceptable for the practical applications.

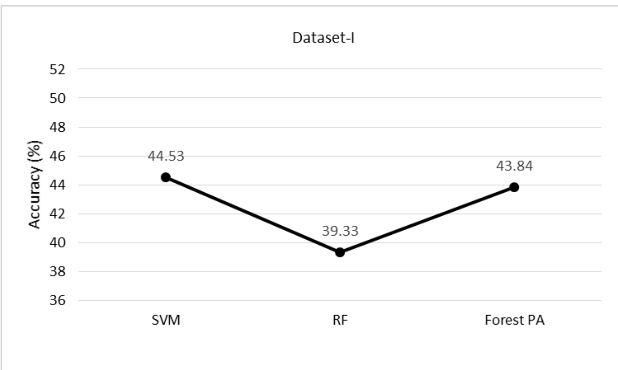


Figure 2. Classification accuracies for Dataset-I

The possible reason of this low accuracy for dataset-I is the spectral confusion between the forest types in the training set and insufficient spectral information to discriminate the classes. When morphological features were added to the original bands, the classification accuracies were changes. Figure 3 shows the classification results for dataset-II and obviously the overall accuracy of Forest PA is the highest among the other methods. SVM obtained the lowest accuracy for dataset-II while it performed better than the other methods for dataset-I. The possible reason of this decline might be the overfitting problem of SVM based upon the choice of cost parameters. In principle, SVM is a robust method against the overfitting problem however it’s really based upon the choice of kernel parameters in practical applications.

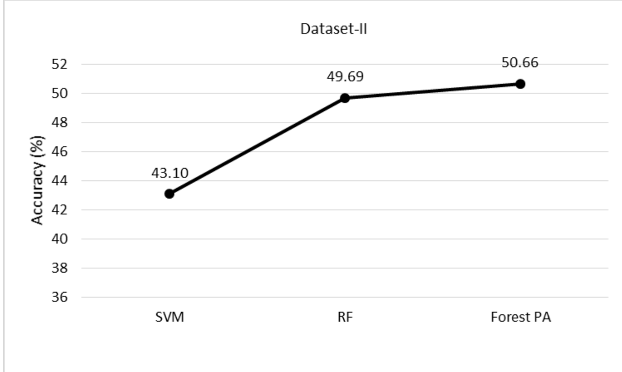


Figure 3. Classification accuracies for Dataset-II

When Figure 4 is examined, it’s clear that morphological operators have increased the classification accuracy for Forest PA and RF. However the accuracy of SVM is decreased by 1.43%. The best improvement (sharp increase in graph) in terms of classification accuracy was observed by RF as 10.36%, followed by Forest PA. The classification accuracy for dataset-II was exceed the 50% by Forest PA, which is still insufficient rate for practical applications in remote sensing.

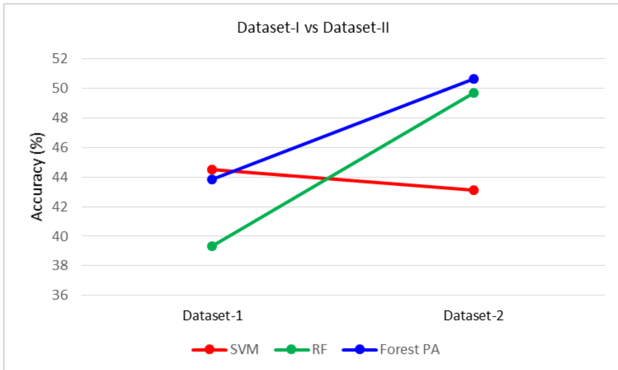


Figure 4. Comparison of classification accuracies

## 4. CONCLUSION

This study assessed the morphological operators from dual-polarized ALOS-PALSAR data for the classification of forested areas. To investigate the impacts of morphological operators on classification accuracy, two different dataset and three machine learning models were utilized. Our experimental results show that morphological operators increased the accuracy by 6.82% and 10.36% for Forest PA and RF, respectively, while decreased the accuracy by 1.43% for SVM. Any method could not reach up to the 50% accuracy for Dataset-I. The spectral confusion between the forest classes caused to low classification accuracy. Sharp increase was observed on RF when morphological features were incorporated. Only Forest PA exceed the 50% accuracy however such level of accuracy still cannot be acceptable for most practical applications in remote sensing. Our results concluded that dual-polarized ALOS-PALSAR, either by itself or with morphological features is insufficient for the classification of forested areas. We believe that multispectral optical data should be incorporated to able to reach up to the desirable accuracy level. Our future work will focus on the multiple/integrated use of vegetation indices obtained from multispectral data and SAR data for the classification of forested areas with advanced ensemble classifiers.

## ACKNOWLEDGEMENTS

The first author of the paper was granted by TUBITAK (The Scientific and Technological Research Council of Turkey) 2214/A International Doctoral Research Fellowship Programme and experiments were held while he visited the Department of Geography at Friedrich-Schiller University Jena, Germany as a guest researcher. He would like to thank TUBITAK (Grant No #1059B141700579) for the support.

## REFERENCES

Abdikan, S., 2018. Exploring image fusion of ALOS/PALSAR data and LANDSAT data to differentiate forest area. *Geocarto International*, 33, pp. 21–37.

Adnan, M.N., Islam, M.Z., 2017. Forest PA: Constructing a decision forest by penalizing attributes used in previous trees. *Expert Systems with Applications*, 89, pp. 389–403.

Balzter, H., Skinner, L., Luckman, A., Brooke, R., 2003. Estimation of tree growth in a conifer plantation over 19 years from multi-satellite L-band SAR. *Remote Sensing of Environment,* 84, pp. 184–191.

Bioucas-Dias J. M., Plaza A., Camps-Valls G., Scheunders P., Nasrabadi N., Chanussot J., 2013. Hyperspectral Remote Sensing Data Analysis and Future Challenges. *IEEE Geoscience and Remote Sensing Magazine,* 1, pp. 6–36.

Du, P., Samat, A., Waske, B., Liu, S., Li, Z., 2015. Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, pp. 38–53.

Fauvel M., Tarabalka Y., Benediktsson J. A., Chanussot J., Tilton J. C., 2013. Advances in Spectral-Spatial Classification of Hyperspectral Images. *Proceedings of the IEEE*, 101, pp. 652–675.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11, 10–18.

Liesenberg, V., Gloaguen, R., 2013. Evaluating SAR polarization modes at L-band for forest classification purposes in Eastern Amazon, Brazil. *International Journal of Applied Earth Observation and Geoinformation,* 21, pp. 122–135.

Melgani F., Bruzzone L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42, pp. 1778–1790.

Middinti, S., Jha, C.S., Reddy, T.B., 2017. Forest type classification with combination of advanced polarimetric decompositions and textures of L-band synthetic aperture radar data. *Journal of Applied Remote Sensing,* 11, 016035.

Mitchell, A.L., Tapley, I., Milne, A.K., Williams, M.L., Zhou, Z.-S., Lehmann, E., Caccetta, P., Lowell, K., Held, A., 2014. C- and L-band SAR interoperability: Filling the gaps in continuous forest cover mapping in Tasmania. *Remote Sensing of Environment*, 155, pp. 58–68.

Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing,* pp. 26, 217–222.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Samat, A., Gamba, P., Liu, S., Miao, Z., Li, E., Abuduwaili, J., 2018. Quad-PolSAR data classification using modified random forest algorithms to map halophytic plants in arid areas. *International Journal of Applied Earth Observation and Geoinformation* 73, pp. 503–521.

Samat A., Persello C., Liu S., Li E., Miao Z., Abuduwaili J., 2018. Classification of VHR Multispectral Images Using ExtraTrees and Maximally Stable Extremal Region-Guided Morphological Profile. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11, pp. 3179–3195.

Siers, M.J., Islam, M.Z., 2018. Novel algorithms for cost-sensitive classification and knowledge discovery in class imbalanced datasets with an application to NASA software defects. *Information Sciences* 459, pp. 53–70.

Thiel, C., Drezet, P., Weise, C., Quegan, S., Schmullius, C., 2006. Radar remote sensing for the delineation of forest cover maps and the detection of deforestation. *Forestry: An International Journal of Forest Research,* 79, pp. 589–597.

Wang, X., Cao, Z., Ding, Y., Feng, J., 2017. Composite Kernel Method for PolSAR Image Classification Based on Polarimetric-Spatial Information. *Applied Sciences,* 7.

Wurm, M., Taubenböck, H., Weigand, M., Schmitt, A., 2017. Slum mapping in polarimetric SAR data using spatial features. *Remote Sensing of Environment,* 194, pp. 190–204.