https://doi.org/10.7494/geom.2023.17.1.57

Rashmi Saini[1]

Integrating Vegetation Indices and Spectral Features for Vegetation Mapping from Multispectral Satellite Imagery Using AdaBoost and Random Forest Machine Learning Classifiers

Abstract: Vegetation mapping is an active research area in the domain of remote sensing. This study proposes a methodology for the mapping of vegetation by integrating several vegetation indices along with original spectral bands. The Land Use Land Cover classification was performed by two powerful Machine Learning techniques, namely Random Forest and AdaBoost. The Random Forest algorithm works on the concept of building multiple decision trees for the final prediction. The other Machine Learning technique selected for the classification is AdaBoost (adaptive boosting), converts a set of weak learners into strong learners. Here, multispectral satellite data of Dehradun, India, was utilised. The results demonstrate an increase of 3.87% and 4.32% after inclusion of selected vegetation indices by Random Forest and AdaBoost respectively. An Overall Accuracy (OA) of 91.23% (kappa value of 0.89) and 88.59% (kappa value of 0.86) was obtained by means of the Random Forest and AdaBoost classifiers respectively. Although Random Forest achieved greater OA as compared to AdaBoost, interestingly AdaBoost provided better class-specific accuracy for the Shrubland class compared to Random Forest. Furthermore, this study also evaluated the importance of each individual feature used in the classification. Results demonstrated that the NDRE, GNDVI, and RTVIcore vegetation indices, and spectral bands (NIR, and Red-Edge), obtained higher importance scores.

Keywords: ensemble classifiers, Machine Learning, Random Forest, AdaBoost, vegetation mapping, vegetation indices

Received: 18 May 2022; accepted: 12 October 2022

[1] Department of CSE, G. B. Pant Institute of Engineering and Technology, Pauri Garhwal, India, email: 2rashmisaini@gmail.com, https://orcid.org/0000-0002-9802-7089

1. Introduction

In recent years, with the availability and accessibility of remote sensing data, a huge variety of applications like crop type classification, vegetation mapping, forestry, precision agriculture, landslide susceptibility mapping, built up extraction, etc. have attracted the attention of multidisciplinary researchers [1–8]. Vegetation mapping is one of the essential application needs which has to be addressed effectively for overall environmental monitoring [8]. The utilization of remotely sensed data is the optimal way for vegetation mapping because of the free availability of medium and coarser spatial resolution data, having different spatial and spectral properties, cost effective and less time consuming in comparison to traditional field survey methods [7, 8]. Quantifying vegetation provides a valuable information for socio-economic applications. It is essential to obtain the accurate information about vegetation cover for various vegetation restoration and protection schemes.

In the past decade, various Machine Learning techniques such as Artificial Neural Network (ANN), k-Nearest Neighbor (k-NN), Random Forest, Support Vector Machine (SVM), Rotation Forest, various boosting techniques (AdaBoost, Schotostic Gradient Boosting etc.) have attracted considerable interest from multidisciplinary researchers as well as within the remote sensing community because these classifiers provide high classification accuracy and are robust to noise [2–8]. The ensemble classifier is a kind of Machine Learning method which consists of a number of base classifiers and combines their prediction by means of some voting scheme or mechanism. The basic concept behind an ensemble classifier makes the assumption that a combination of multiple base classifiers provides better prediction results as compared to a single classifier [10–12]. Random Forest is one popular ensemble method used for a variety of remote sensing applications like landslide susceptibility mapping [5], vegetation mapping [2, 4, 7, 12], ecotope mapping [1], land cover classification [3], etc. AdaBoost is a popular boosting technique proposed by Freund and Schapire [13], and this algorithm enhances the performance of base learner or weak models by converting them into the strong learner or models [8, 14].

Optical remote sensing data are the most widely used data types for a variety of agricultural mapping and monitoring applications [4, 8, 12, 14–18]. Remote sensing data have been used successfully for other challenging applications such as drought identification and analysis [19, 20]. Rotjanakusol and Laosuwan [19] carried out a study to analyze droughts using Terra/MODIS data for the time duration of one decade. Another study utilized Landsat and NDVI for drought detection [20]. However, some studies in the domain of agriculture mapping have been carried out using Synthetic Aperture Radar (SAR) and hyperspectral data. Chan and Paelinckx [1] carried out a study for classification using hyperspectral data (Airborne HyMap) at the study site of Belgium. The results of this study indicates that the AdaBoost classifier and Random Forest achieved nearly similar classification accuracy and both ensemble classifiers outperformed Artificial Neural Network (ANN). Tigges et al. [4] utilized multispectral RapidEye data for vegetation mapping in an urban environment. This study utilized multi-temporal data for vegetation classification and the results demonstrated that the Red-Edge information is beneficial for class separability. Schuster et al. [15] examined the potential of the Red-Edge channel using RapidEye data for land use classification for a study area located in western Berlin (Germany). In this study, SVM and Maximum Likelihood Classifiers (MLC) were employed. This study demonstrated that the additional information provided by Red-Edge band contributed significantly to improving the accuracy of classification, particularly for vegetation classes. Another study performed by Inglada et al. [16] implemented supervised classification techniques to produce accurate crop maps on a global scale. This study used high-resolution optical imagery (multisensory data SPOT4, Landsat-8, RapidEye). Twelve different sites were selected all over the world. In this study, the selected test sites had different characteristics, so the outcome of the study may apply to different types of landscape. The results show that the RF classifier achieved maximum accuracy for most of the sites. Sonobe et al. [17] carried out a study for mapping of crops using Landsat-8 OLI data by employing multi-Grained Cascade Forest (deep forest), CART, and RF classifiers. In this study, 57 spectral indices have been computed on Landsat-8 data and evaluated their potential in crop mapping at Hokkaido (Japan). The results show that maximum accuracy is achieved after the inclusion of spectral indices by RF followed by deep forest. In the literature it is clear that various vegetation indices are used for a variety of application using remotely sensed data [21–26].

Kim and Yeom [27] carried out a study to evaluate the sensitivity of vegetation indices on crops (three paddy rice classes) using RapidEye data of two different seasons. The results of the study have shown that NDVI and EdgNDVI were most appropriate vegetation indices (VI) to differentiate between the selected crops. Another study performed by Ustuner et al. [28] utilized RapidEye data for crop classifications using the SVM classifier. This study used three VI namely NDVI, GNDVI and NDRE. The results indicated that the highest Overall Accuracy (OA) of 87.46% was obtained by including all three VI. Furthermore, the authors claim that NDRE VI contributed most in terms of the classification accuracy as compared to other selected indices. Otunga et al. [29] evaluated the potential of Red-Edge data from two satellites i.e., RapidEye and Sentinel-2, to distinguish grass (festuca C3). This study used NDVI and NDRE VI and Maximum Likelihood Classifier (MLC) to classify grass species. The results have shown that the integration of Red-Edge enhances the accuracy for RapidEye (+4.76%) and Sentinel-2 (+5.95%) satellite data. Furthermore, the integration of NDVI and NDRE is beneficial to classify grass species more accurately. Peng et al. [30] combined two satellite data (RapidEye and GF-2) to map mangrove species. The authors used three ensemble classifiers i.e., AdaBoost, Random Forest (RF) and Rotation Forest (RoF). The results demonstrated that RF and RoF performed better in comparison to the AdaBoost classifier.

The literature indicates that ensemble techniques provide excellent classification results, however, the utilization of boosting methods is limited for applications such as vegetation mapping and monitoring. Therefore, this study aims to examine the potential of boosting methods (AdaBoost), specifically for mapping vegetation classes and comparing the statistical results with another popular ensemble method i.e., Random Forest. In addition, this study will also analyse the potential of several spectral indices for vegetation classes within the selected study area. Furthermore, in order to obtain optimal results in terms of classification accuracies, optimization has been carried out for the tuning parameters associated with both of the classifiers.

2. Study Area

The selected study area is located in the Dehradun district of Uttarakhand state in India (Fig. 1). This study region covers about 55.224 km^2 with upper left 30°18′2.46″ N and 78°2′7.21″ E to lower left 30°14′17.68″ N and 78°7′2.81″ E. The data used consists of a RapidEye satellite multispectral image acquired on 7 March 2013. RapidEye represents a constellation of five satellites and provides images in five optical bands (Tab. 1) within the wavelength range of 400–850 nm. The RapidEye satellite provides multispectral data, as well as good spatial resolution (5 m). Another advantage of RapidEye is that it includes a Red-Edge band, which provides additional information that is useful for vegetation mapping [14]. With such properties, RapidEye represents a good option for vegetation mapping. The key features of RapidEye imagery are shown in Table 1.

Table 1. Charactertics of the RapidEye satellite

Characteristics	Details
Number of satellites	5
Orbit altitude	630 km (Sun-synchronous orbit)
Sensor type	Multi-spectral push broom imager
Swath width	77 km
Spectral channels or bands	Blue (440–510 nm) Green (520–590 nm) Red (630–685 nm) Red-Edge (690–730 nm) NIR (760–850 nm)
Ground sample distance	6.5 m
Pixel size	5.0 m
Revisit time	Daily (off-nadir; always less than 20°) / 5.5 days (nadir)
Dynamic range	12 bits

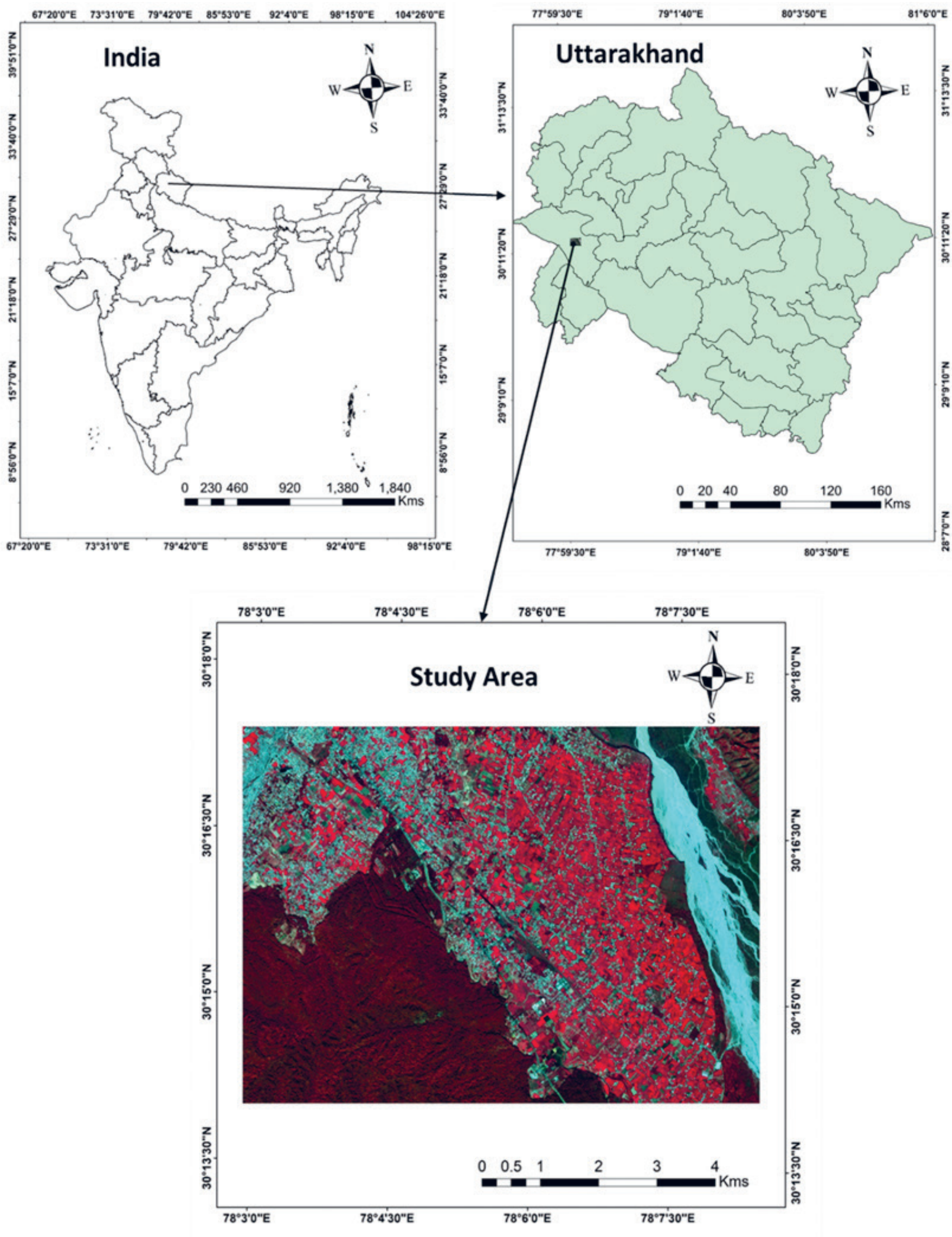

Fig. 1. Selected study area located in India

3. Methodology

For this study, the methodology adopted is depicted in Figure 2, which represents the complete workflow of the study.

In order to obtain classification maps, first the stacking operation was performed to prepare the data, which consists of five spectral bands. Then, subset computation was done to get the selected study area. Then computation of vegetation indices (VI) was done.

Vegetation indices are commonly computed using the reflectance of two or more spectral bands. Vegetation indices are widely used to obtain quantitative information about the biophysical variable of vegetation using remotely sensed data [17]. The importance of VI has been recognized by many studies for the classification of crops [15, 17, 18]. In this study, six vegetation indices have been computed using spectral bands of RapidEye. Vegetation indices and their formulas are listed in Table 2. After the computation of selected indices, a final dataset was prepared, consisting of both the selected VIs (the six listed in Table 2) and spectral bands.

For classification purposes, the Random Forest and AdaBoost methods were implemented. More precisely, the research design was performed using following steps:

Step 1. Input data preparation includes several operations. Selection of study area followed by a stacking operation. The stacking operation was performed in ArcGIS software by combining all five spectral bands (Red, Green Blue, Red-Edge, NIR). Thereafter, a subset operation was performed in ArcGIS to obtain the desired dataset (Spectral bands).

Step 2. Computation of the vegetation indices (NDVI, NDRE, GNDVI, SR-RE, MTVI2, RTVIcore using the spectral bands from the RapidEye satellite. Computation of the VI was performed in the R Programming Environment. Preparation of second dataset (spectral + vegetation indices).

Step 3. Preparation of complete reference sample datasets by means of the stratified random sampling approach.

Step 4. Implementation and parameter optimization of both the Machine Learning models (Random Forest and AdaBoost) in the R programming environment. Partition of reference data, 70% data was used for training and 30% for testing and training both classifiers.

Step 5. Prediction by both Machine Learning classifiers on unseen datasets and assessment of accuracy using various accuracy measures. Computation of the importance score for vegetation classes.

Step 6. Resultant accuracy measures and classified Land Use Land Cover (LULC) maps obtained by both the classifiers.

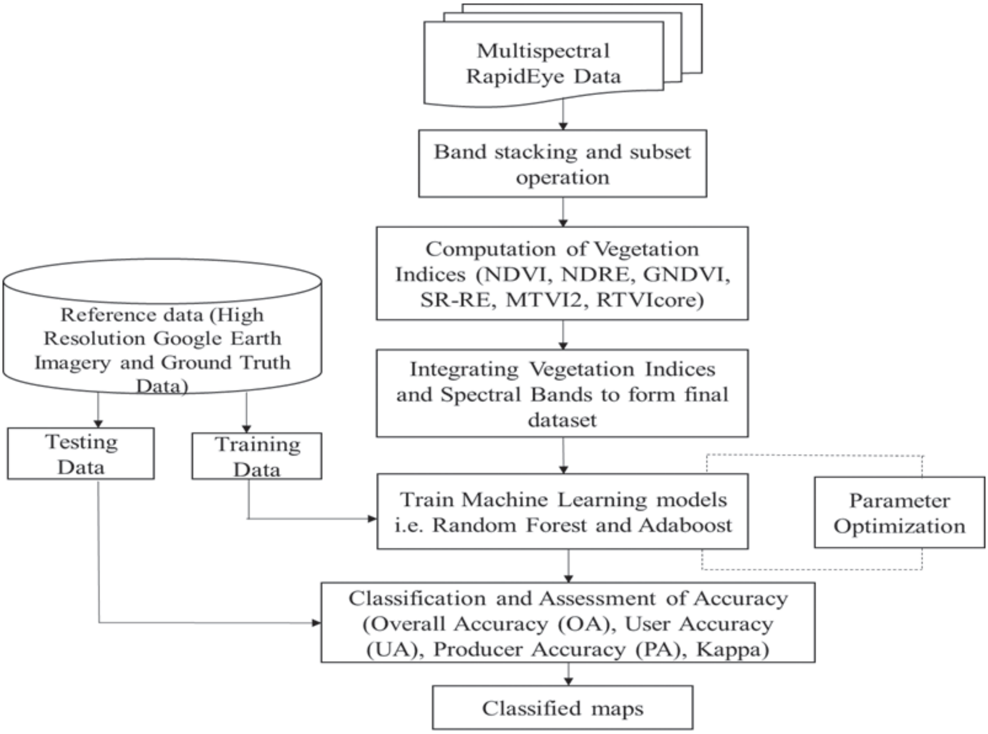

Fig. 2. Classification process flow methodology

Table 2. Used vegetation indices along with their formulas

No.	Vegetation indices	Acronym	Formulas
1	Normalized Difference Vegetation Index	NDVI	$(NIR - R)/(NIR + R)$
2	Normalized Difference Vegetation Index – Red-Edge	NDRE	$(NIR - RE)/(NIR + RE)$
3	Green Normalized Difference Vegetation Index	GNDVI	$(NIR - G)/(NIR + G)$
4	Red-Edge Simple Ratio	SR-RE	NIR/RE
5	Modified Triangular Vegetation Index	MTVI2	$\frac{1.5[1.2(NIR - G) - 2.5(R - G)]}{\sqrt{(2RNIR + 1)^2 - \left(6NIR - 5\sqrt{(R)}\right) - 0.5}}$
6	Red-Edge Triangular Vegetation Index (core only)	RTVIcore	$100(NIR - RE) - 10(NIR - G)$

3.1. Random Forest Classifier

The Random Forest classifier has been applied successfully to various kinds of classification problem using a variety of remotely sensed data. A number of studies have indicated that Random Forest produces excellent classification results [7, 8, 11]. Random Forest is a kind of ensemble approach and was proposed by Breiman [10]. This method works on the concept of bagging or bootstrap aggregation. This technique builds multiple decision trees using bootstrap samples with replacement policy. The predictive performance of Random Forest depends on the individual tree used to build the forest. In this method, each tree contributes its vote for the classification, whereas the final prediction is made by a voting scheme. Decision tress in the Random Forest are decorrelated because this method selects features randomly at each node, which increases the prediction efficacy of this method. The Random Forest algorithm also work well when the input data is of a higher dimension, at the same time this method has a low tendency to overfitting [10]. RF has two tuning parameters, namely Mtry and ntree. The Mtry parameter represents the number of features used to split the node, whereas the ntree parameter indicates the total number of trees.

3.2. AdaBoost Classifier

AdaBoost is an ensemble Machine Learning method which comes under the category of boosting approach and was introduced by Freund and Schapire [13]. The AdaBoost classifier works on the principle of constructing strong classifiers using weak classifiers. Here, a strong classifier is formed by combining basic or weak classifiers. The algorithm work on the concept of an adaptive resampling method to select training samples. Initially, equal weights are assigned to all samples. At each iteration of the algorithm, it assigns weight to the samples in such a manner that the coming integration emphasizes on the samples which were not accurately classified by the previous attempt and the final outcome is the weighted sum of the prediction [1]. The AdaBoost Machine Learning technique has been used for the classification of a variety of applications, however, the evaluation for this algorithm for vegetation mapping is limited. The AdaBoost classifier has two parameters, namely Mfinal and maxdepth. Mfinal indicates the total number of trees used to build the final model whereas the maxdepth parameter represents maximum depth of any node.

In this study, reference data have been collected by field visits as well as using Google Earth imagery for reference sample generation. In this work, a stratified random sampling approach has been employed. Both the selected classifiers i.e., Random Forest and AdaBoost, have been trained using labelled samples. The assessment of accuracy is performed by calculating the most commonly used accuracy measures i.e., Overall Accuracy (OA) and kappa coefficient. Furthermore, to evaluate the classifier's performance on each individual class, two other popular accuracy measures i.e., Producer Accuracy (PA) and User Accuracy (UA) measures have been used. For the implementation purposes of both the classifiers, the R programming language has been used.

4. Results and Discussion

In this research work, vegetation mapping was performed using multispectral satellite data of Dehradun, India. Here, with an objective of achieving optimal classification results, the collective benefits of spectral features and vegetation indices were utilized to train the selected Machine Learning techniques (Random Forest and AdaBoost). In this study, two datasets were prepared, with the first dataset consisting of only spectral bands. The second dataset was formed by combining the original spectral bands and computer vegetation indices. Thus, the first dataset consists of five spectral features and the second dataset includes a total of eleven features. In order to perform the supervised classification, the stratified random sampling technique was adopted for training and testing samples, these samples were selected in such a manner that they are mutually exclusive. In this study, parameter optimization is done in order to utilize the classifier's maximum potential for accurate classification results. The optimal value of tuning parameters has been determined and used to build Machine Learning models for the final predictions (Tab. 3).

Table 3. Optimal values of parameters for Random Forest and AdaBoost models

Dataset	Random Forest	AdaBoost
Spectral bands	Mtry = 2, ntree = 200	maxdepth = 3, mfinal = 150
Spectral bands + vegetation indices	Mtry = 3, ntree = 350	maxdepth = 3, mfinal = 100

The results of this investigation revealed that the Random Forest method obtained a higher OA as compared to the AdaBoost ensemble classifier. This observation is similar for both datasets (spectral data and spectral bands + vegetation indices). More specifically, Random Forest achieved +3.09% higher OA and an increase of +0.03 in kappa value as compared to AdaBoost classifier using only spectral data. For the second dataset (spectral + vegetation indices), the Random Forest classifier has shown an increase of +2.64% and +0.04 in OA and kappa value respectively in comparison to the AdaBoost classifier.

First, we discuss the outcomes of both the classifiers using spectral data (Tab. 4). The results have shown that using only spectral data, an OA of 87.36% with a kappa value of 0.85 and 84.27% with a kappa value of 0.82 has been achieved by Random Forest and AdaBoost classifiers respectively. The class-specific accuracy measures in terms of UA and PA are listed in Table 4. This study mainly focused on vegetation mapping using multispectral RapidEye data. A total number of six LULC classes, namely Forest, Sand area, Crop land, Fallow land, Shrub land, and Built-up have been considered for classification. As far as the class-specific performance is concerned, three classes come under the umbrella of vegetation i.e., Forest, Crop land, and Shrub land. Therefore, the class-specific accuracy is only discussed regarding these vegetation classes.

Table 4. Accuracy measures obtained by Random Forest and AdaBoost classifiers using only spectral information

LULC classes	Random Forest		AdaBoost	
	UA	PA	UA	PA
Forest	90.76	88.01	85.30	82.17
Sand area	86.08	85.14	83.93	84.54
Crop land	87.88	89.45	79.14	83.16
Fallow land	93.77	90.23	90.36	87.29
Shrub land	88.78	86.95	90.18	87.42
Built-up	80.32	84.45	78.12	81.34
OA	87.36		84.27	
Kappa	0.85		0.82	

The results demonstrated that the Random Forest classifier achieved UA of 90.76% for Forest, 87.88% for Crop land and 88.78% for Shrub land. The PA of 88.01%, 89.45% and 86.95% were obtained for Forest, Crop and Shrub land respectively. The AdaBoost classifier obtained UA of 85.30% for Forest, 79.14% for Crop land and 90.18% for Shrub land. The PA obtained by AdaBoost is 82.17%, 83.16% and 87.42% for Forest, Crop and Shrub land respectively.

Table 5 demonstrate the results obtained by Random Forest and AdaBoost classifiers using second dataset i.e., the integration of vegetation indices and spectral bands together. It was found that an OA of 91.23% with a kappa value of 0.89 and 88.59% with a kappa value of 0.85 was achieved by the Random Forest and AdaBoost classifiers respectively (Tab. 5). The results of this investigation have shown that both classifiers performed well, however, Random Forest outperformed AdaBoost ensemble method for both datasets. Random Forest achieved a 2.64% higher classification accuracy and 0.04 higher value of kappa coefficient in comparison to AdaBoost. The main objective of this study was to map vegetation using multispectral RapidEye data. As discussed above, the class-specific performance is evaluated for three classes in terms of vegetation i.e., Forest, Crop land and Shrub land. Therefore, the class-specific accuracy is discussed regarding these vegetation classes only. The Random Forest classifier achieved UA of 96.53% for Forest, 91.52% for Crop land and 92.59% for Shrub land. Whereas, the PA of 93.05%, 96.77% and 90.61% was obtained for Forest, Crop land and Shrub land respectively. The AdaBoost classifier obtained a UA of 91.80% for Forest, 87.12% for Crop land and 94.28% for Shrub land. The PA obtained by AdaBoost is 86.07%, 90.01% and 92.37% for Forest, Crops and Shrub land respectively.

Table 5. Accuracy measures obtained by Random Forest and AdaBoost classifiers by integrating selected vegetation indices

LULC classes	Random Forest		AdaBoost	
	UA	PA	UA	PA
Forest	96.53	93.05	91.80	86.07
Sand area	89.16	85.03	86.79	89.06
Crop land	91.52	96.77	87.12	90.01
Fallow land	93.65	96.92	92.03	90.45
Shrub land	92.59	90.61	94.28	92.37
Built-up	83.27	86.91	80.07	83.62
OA	91.23		88.59	
Kappa	0.89		0.85	

The findings of this study clearly demonstrated the positive impact of including vegetation indices in the classification process. The results have shown an increase of 3.87% in OA after inclusion of selected vegetation indices by Random Forest classifier. The AdaBoost classifier showed a rise of 4.32% after integrating vegetation indices with spectral bands. For the vegetation classes it has been observed that accuracies of all the considered LULC classes have been increased after the inclusion of vegetation indices. Furthermore, it has been found that the class specific accuracies of vegetation classes (Forest, Crop land and Shrub land) have been increased significantly.

In order to perform the in-depth evaluation, we have analyzed the impact of integration of vegetation indices on class-specific accuracies for both the classifiers. It was found that Random Forest obtained UA of 96.53% and PA of 93.05% for Forest by integrating vegetation indices. More specifically, for the Forest class UA and PA increased by +5.77% and +5.04% respectively after inclusion of vegetation indices. For Crop land, UA and PA of 91.52% and 96.77% has been obtained, which indicates an increase of 3.64% and 6.54% in the value of UA and PA respectively. For Shrub land, the value of UA and PA values obtained were 92.59% and 90.61% respectively, which is 3.81% (UA) and 3.66% (PA) higher as compared to spectral data only.

The AdaBoost classifier mapped forest with the UA of 91.80% and PA of 86.07% by combining spectral bands and vegetation indices. The results indicate that, in comparison with spectral data, UA increased by 6.50% whereas PA increased by 3.90% for the Forest class. Crop land is classified with UA of 87.12% and PA of 90.01%, which indicates an increase of 7.98% in the value of UA and 6.85% in the value of PA after the inclusion of vegetation indices. For Shrub land, the UA and PA values are 94.28% and 92.37% respectively. It can be seen from the outcome of both datasets that an increase of 4.10% in the value of UA and 4.95% in the value of PA were achieved for shrub land class after including vegetation indices in the input dataset. It can be observed from the classification measures that the accuracies (UA and PA) of all the considered LULC classes have been increased by including selected vegetation indices in the classification procedure (Tabs. 4, 5). A similar trend can be observed for both the selected classifiers (Random Forest and AdaBoost). Furthermore, it was found that the maximum positive impact of vegetation indices is on vegetation classes i.e., Forest, Crop and Shrub land using Random Forest, with the highest accuracy enhancement of 6.54% in the value of PA reported for the Crop class. Whereas, AdaBoost produced the highest accuracy enhancement of 7.98% in UA for the Crop class.

If we compare the outcomes of this study with previous ones, Chan and Paelinckx [1] demonstrated that the AdaBoost classifier and Random Forest obtained an approximately similar classification accuracy. In contrast, Inglada et al. [16] found that the Random Forest classifier achieved maximum accuracy for most of the selected study sites. Sonobe et al. [17] also observed that RF outperformed in terms of crop mapping using Landsat-OLI data. In contrast to the study [1], our results indicated that Random Forest performed better for both datasets, which is similar to the findings observed by Inglada et al. [16] and Sonobe et al. [17]. In context to the integration of vegetation indices, previous studies also observed that the integration of vegetation indices is beneficial for classification, which is similar to the observation of our study. Elsewhere in the literature, the study performed by Kim and Yeom [27] for classification of three paddy rice classes, Ustuner et al. [28] for crop classification, Otunga et al. [29] to distinguish grass (festuca C3) found that integration of vegetation indices improves classification accuracy. Therefore, based on the obtained statistical results it can be concluded that inclusion of vegetation indices improves the classification results of all vegetation classes significantly. LULC maps obtained by AdaBoost and Random Forest integrating spectral bands and vegetation indices are shown in Figures 3 and 4 respectively.

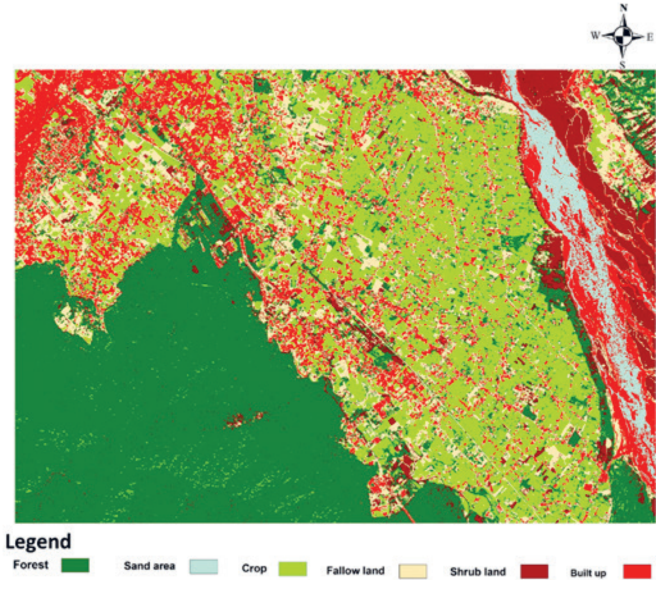

Fig. 3. Classification results obtained by the AdaBoost classifier

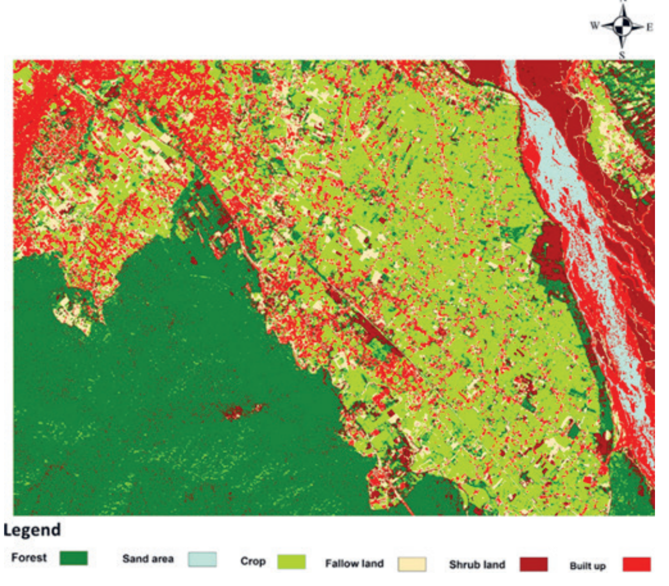

Fig. 4. Classification results obtained by the Random Forest classifier

As discussed in the methodology section, six vegetation indices computed using spectral bands (Tab. 2) and incorporated in the second dataset. Therefore, a total number of eleven features were used to perform the classification. In order to extend the evaluation procedure, feature importance have been computed for each considered feature for all vegetation classes i.e., Forest, Crop and Shrub land. The results of feature importance are based on Mean Decrease Gini Score [26]. The importance graphs for Forest, Crop land and Shrub land is shown in Figures 5a, b, and c respectively. It was found that for all the vegetation classes, different features obtained different importance measure. The results indicate that NDRE obtained highest importance for the Forest class, and the NIR band had the second highest importance. For Crop land, the maximum importance is reported by GNDVI followed by NDRE. In the case of the shrub class, NDRE and NDVI are the most important features. A study performed by Ustuner et al. [28] also revealed that NDVI, GNDVI and NDRE are important vegetation indices for the classification of crops. Apart from the discussed features in this study, Red-Edge, NIR, Green and RTVIcore also obtained higher importance. Furthermore, it was found that the Blue, Red and SE-RE Vegetation Index spectral bands obtained a minimum importance score.

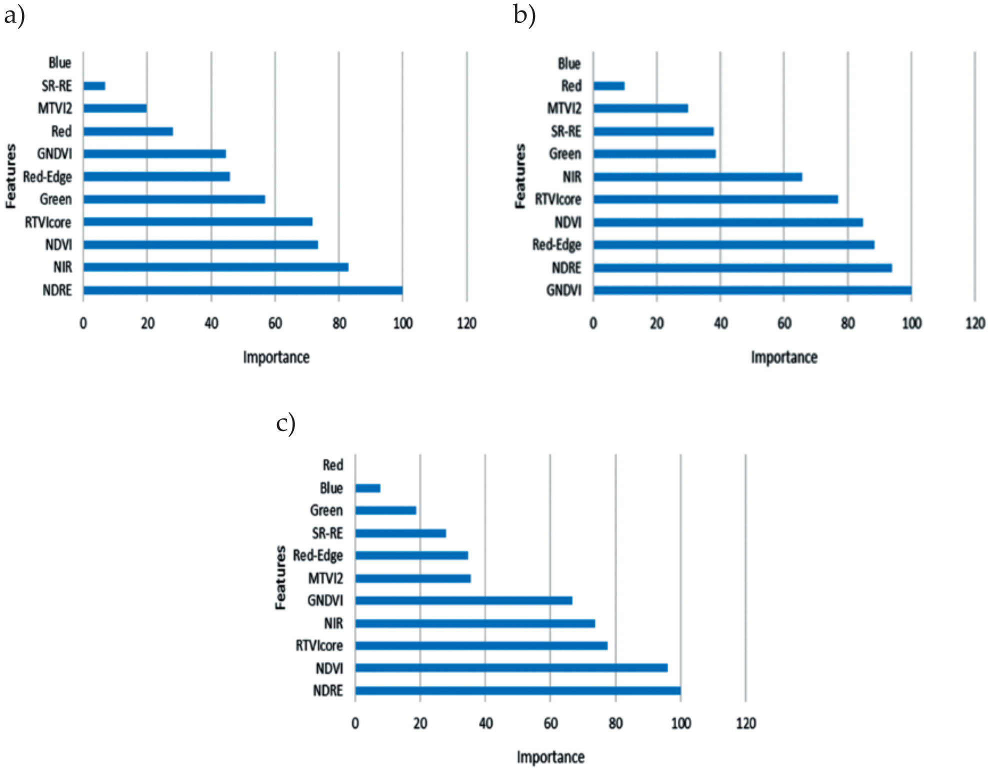

Fig. 5. Feature importance graph for Forest (a), Crop land (b) and Shrub land (c)

5. Conclusion

In this study, vegetation mapping was performed using multispectral RapidEye imagery by combining several vegetation indices and spectral bands. In this investigation, the Random Forest and AdaBoost methods were considered for classification purposes.

The major findings of this study are listed below:

– Results demonstrated that Random Forest performed better in terms of OA and kappa value for both the datasets. The Random Forest classifier has shown a rise of +3.09% and +2.64% in OA compared to AdaBoost for spectral data and integrated spectral bands plus vegetation indices.
– The inclusion of vegetation indices has a significant positive impact on classification. The highest OA of 91.23% and 88.59% has been obtained by Random Forest and AdaBoost classifier respectively using spectral bands plus vegetation indices.
– The results have shown a rise of +3.87% and +4.32% in OA after the inclusion of selected vegetation indices in Random Forest and AdaBoost respectively.
– For class-specific accuracy, Random Forest reported a maximum increase of +6.54% in the value of PA for the Crop class while AdaBoost produced the highest accuracy enhancement of +7.98% in UA for Crop land.
– Although Random Forest achieved greater OA as compared to AdaBoost, interestingly, for the classes of Shrub land and sand area AdaBoost provided better class-specific accuracy (UA and PA) compared to Random Forest.
– Furthermore, this study also evaluated the importance of each individual feature used in the classification. The results of the importance measure clearly indicated that vegetation indices such as NDRE, GNDVI, RTVIcore, are the most important features for the target classes.
– Feature importance results also revealed that spectral bands NIR and Red-Edge bands contributed more for the classification in comparison to other spectral bands.
– A future scope for research may include the in-depth evaluation of other ensemble classification techniques and the evaluation of their performance for a specific class.

References

[1] Chan J.C.-W., Paelinckx D.: *Evaluation of Random Forest and AdaBoost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery*. Remote Sensing of Environment, vol. 112(6), 2008, pp. 999–3011. https://doi.org/10.1016/j.rse.2008.02.011.

[2] Liu J., Feng Q., Gong J., Zhou J., Liang J., Li Y.: *Winter wheat mapping using a random forest classifier combined with multi-temporal and multi-sensor data*. International Journal of Digital Earth, vol. 11(8), 2018, pp. 783-802. https://doi.org/10.1080/17538947.2017.1356388.

[3] Adam E., Mutanga O., Odindi J., Abdel-Rahman E.M.: *Land-use / cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers*. International Journal of Remote Sensing, vol. 35, 2014, pp. 3440–3458. https://doi.org/10.1080/01431161.2014.903435.

[4] Tigges J., Lakes T., Hostert P.: *Urban vegetation classification: Benefits of multitemporal RapidEye satellite data*. Remote Sensing of Environment, vol. 136, 2013, pp. 66–75. https://doi.org/10.1016/j.rse.2013.05.001.

[5] Micheletti N., Foresti L., Robert S., Leuenberger M., Pedrazzini A., Jaboyedof M., Kanevski M.: *Machine Learning feature selection methods for landslide susceptibility mapping*. Mathematical Geosciences, vol. 46, 2014, pp. 33–57. https://doi.org/10.1007/s11004-013-9511-0.

[6] Saini R., Verma S.K., Gautam A.: *Implementation of Machine Learning classifiers for built-up extraction using textural features on Sentinel-2 data*. [in:] *2021 IEEE 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Vol. 1*, IEEE, Piscataway 2021, pp. 1394–1399. https://doi.org/10.1109/ICACCS51430.2021.9441713.

[7] Saini R., Ghosh S.K.: *Exploring capabilities of Sentinel-2 for vegetation mapping using Random Forest*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-3, 2018, pp. 1499–1502. https://doi.org/10.5194/isprs-archives-XLII-3-1499-2018.

[8] Saini R., Ghosh S.K.: *Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date Sentinel-2A imagery*. Geocarto International, vol. 36(19), 2019, pp. 2141–2159. https://doi.org/10.1080/10106049.2019.1700556.

[9] Lu D., Weng Q.: *A survey of image classification methods and techniques for improving classification performance*. International Journal of Remote Sensing, vol. 28(5), 2007, pp. 823–870. https://doi.org/10.1080/01431160600746456.

[10] Breiman L.: *Random Forests*. Machine Learning, vol. 45(1), 2001, pp. 5–32. https://doi.org/10.1023/A:1010933404324.

[11] Belgiu M., Drăguţ L.: *Random forest in remote sensing: A review*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 114, 2016, pp. 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

[12] Zhang C., Xie Z.: *Data fusion and classifier ensemble techniques for vegetation mapping in the coastal Everglades*. Geocarto International, vol. 29(3), 2013, pp. 228–243. https://doi.org/10.1080/10106049.2012.756940.

[13] Freund Y., Schapire R.E.: *A decision-theoretic generalization of on-line learning and an application to boosting*. Journal of Computer and System Sciences, vol. 55(1), 1997, pp. 119–139. https://doi.org/10.1006/jcss.1997.1504.

[14] Colkesen I., Kavzoglu T.: *Ensemble-based canonical correlation forest (CCF) for land use and land cover classification using Sentinel-2 and Landsat OLI imagery*. Remote Sensing Letters, vol. 8(11), 2017, pp. 1082–1091. https://doi.org/10.1080/2150704X.2017.1354262.

[15] Schuster C., Förster M., Kleinschmit B.: *Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data*. International Journal of Remote Sensing, vol. 33(17), 2012, pp. 5583–5599. https://doi.org/10.1080/01431161.2012.666812.

[16] Inglada J., Arias M., Tardy B., Hagolle O., Valero S., Morin D., Dedieu G. et al.: *Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery*. Remote Sensing, vol. 7(9), 2015, pp. 12356–12379. https://doi.org/10.3390/rs70912356.

[17] Sonobe R., Yamaya Y., Tani H., Wang X., Kobayashi N., Mochizuki K.I.: *Evaluating metrics derived from Landsat 8 OLI imagery to map crop cover*. Geocarto International, vol. 34(8), 2019, pp. 839–855. https://doi.org/10.1080/10106049.2018.1425739.

[18] Frampton W.J., Dash J., Watmough G., Milton E.J.: *Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 82, 2013, pp. 83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007.

[19] Rotjanakusol T., Laosuwan T.: *An Investigation of Drought around Chi Watershed during Ten-Year Period Using Terra/Modis Data*. Geographia Technica, vol. 14(2), 2019, pp. 74–83. https://doi.org/10.21163/GT_2019.142.07.

[20] Uttaruk Y., Laosuwan T.: *Drought detection by application of remote sensing technology and vegetation phenology*. Journal of Ecological Engineering, vol. 18(6), 2017, pp. 115–121. https://doi.org/10.12911/22998993/76326.

[21] Rouse J.W., Haas R.H., Schell J.H.: *Monitoring the Vernal Advancement and Retrogradation (Greenwave Effect) of Natural Vegetation*. Texas A & M University, College Station, 1974.

[22] Gitelson A., Merzlyak M.N.: *Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L. and Acer platanoides L. leaves. Spectral features and relation to chlorophyll estimation*. Journal of Plant Physiology, vol. 143(3), 1994, pp. 286–292. https://doi.org/10.1016/S0176-1617(11)81633-0.

[23] Gitelson A.A., Kaufman Y.J., Merzlyak M.N.: *Use of a green channel inremote sensing of global vegetation from EOS-MODIS*. Remote Sensing of Environment, vol. 58(3), 1996, pp. 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7.

[24] Haboudane D., Miller J.R., Pattey E., Zarco-Tejada P.J. Strachan I.B.: *Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture*. Remote Sensing of Environment, vol. 90(3), 2004, pp. 337–352. https://doi.org/10.1016/j.rse.2003.12.013.

[25] Chen P.F., Tremblay N., Wang J.H.., Vigneaulta P.: *New index for crop canopy-fresh biomass estimation*. Spectroscopy and Spectral Analysis, vol. 30(2), 2010, pp. 512–517 [in Chinese]. https://doi.org/10.3964/j.issn.1000-0593(2010)02-0512-06.

[26] Forkuor G., Dimobe K., Serme I., Tondoh J.E.: *Landsat-8 vs. Sentinel-2: examining the added value of sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso*. GIScience & Remote Sensing, vol. 55(3), 2018, pp. 331–354. https://doi.org/10.1080/15481603.2017.1370169.

[27] Kim H.-O., Yeom J.-M.: *Sensitivity of vegetation indices to spatial degradation of RapidEye imagery for paddy rice detection: a case study of South Korea*. GIScience & Remote Sensing, vol. 52(1), 2015, pp. 1–17. https://doi.org/10.1080/15481603.2014.1001666.

[28] Ustuner M., Sanli F.B., Abdikan S., Esetlili M.T., Kurucu Y.: *Crop type classification using vegetation indices of RapidEye imagery*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XL-7, 2014, pp. 195–198. https://doi.org/10.5194/isprsarchives-XL-7-195-2014.

[29] Otunga C., Odindi J., Mutanga O., Adjorlolo C.: *Evaluating the potential of the red edge channel for C3 (Festuca spp.) grass discrimination using Sentinel-2 and Rapid Eye satellite image data*. Geocarto International, vol. 34(10), 2019, pp. 1123–1143. https://doi.org/10.1080/10106049.2018.1474274.

[30] Peng L., Liu K., Cao J., Zhu Y., Li F., Liu L.: *Combining GF-2 and RapidEye satellite data for mapping mangrove species using ensemble machine-learning methods*. International Journal of Remote Sensing, vol. 41(3), 2020, pp. 813–838. https://doi.org/10.1080/01431161.2019.1648907.