ACCURACY COMPARISON OF LAND COVER MAPPING USING THE OBJECT-ORIENTED IMAGE CLASSIFICATION WITH MACHINE LEARNING ALGORITHMS

Shota Mochizuki *1 and Takuhiko Murakami 2

1Doctoral Student, Graduate school of science and technology, University of Niigata,
8050, Ikarashi-2no-cho, Nishi-ku, Niigata-shi, Niigata 950-2181, Japan;
Tel: +81-25-262-6627; Fax: +81-25-262-6627
E-mail: mochi.sho@gmail.com

2Associate Professor, Graduate school of science and technology, University of Niigata,
8050, Ikarashi-2no-cho, Nishi-ku, Niigata-shi, Niigata 950-2181, Japan;
Tel: +81-25-262-6627; Fax: +81-25-262-6627
Email: muratac@agr.niigata-u.ac.jp

KEY WORDS: Classification and regression tree, Bagging, Boosting, Random Forest, Ensemble classifier

ABSTRACT: Land cover mapping provides basic information for advanced science such as ecological management, biodiversity conservation, forest planning and so on. In remote sensing research, the process of creating an accurate land cover map is an important subject. Recently, there has been growing research interest in the object-oriented image classification techniques. The object-oriented image classification consists of multi-dimensional features including object features and thus requires multi-dimensional image classification approaches. For example, a linear model such as the maximum likelihood method of pixel-based classification cannot characterize the patterns or relations of multi-dimensional data. In multi-dimensional image classification, data mining and ensemble learning have been shown to increase accuracy and flexibility. This study examined the use of the object-oriented image classification by the multiple machine learning algorithms for land cover mapping. We applied four classifiers: Classification and regression tree (CART), Decision tree with Boosting, Decision tree with Bagging, and Random Forest. The study area was Sado Island in Niigata Prefecture, Japan. Pan-sharpened SPOT/HRG imagery (June 2007) was used and classified into the following eight classes: broad-leaved deciduous forest, Japanese cedar, Japanese red pine, bamboo forest, paddy field, urban area, road, and bare land. We prepared four data sets with the object based features including textural information. The number of features is increased from data set I through IV. As the result, CART was unsuitable for multi-dimensional classification. Random Forest and Decision tree with Boosting showed high classification accuracies. Furthermore, in the data set with the limited features, Decision tree with Boosting was the accurate classifier. Finally, we propose two machine learning algorithms to every datasets. Random Forest is effective in the case of the multi-dimensional image classification such as data set II, III, and IV. Decision tree with Boosting is effective in the case of the image classification with the limited features such as data set I.

INTRODUCTION

Remotely sensed data has become essential tool for global environmental observation which includes global climate change, land-cover/land-use change, sustainable forest management, natural resources management, and disaster management (Szuster et al., 2011). In particular, image classification of multispectral data to detect land cover and land use (LCLU) is developing in importance. LCLU maps represent valuable information for evaluating the natural and artificial environments through quantifying vegetation structure from various landscape levels at a specific time point or over a continuous period (Xie et al., 2008). Using remotely sensed data for LCLU mapping, we have the advantage of rapid data acquisition at a lower cost than ground survey methods (Pal and Mathur, 2004).

When preparing LULC maps, the various classification techniques, such as the maximum likelihood method, the multilevel slice method, the ISODATA method, and k-nearest neighbor method have been applied. Also about satellite data, we can employ several data from high spatial resolution imagery (e.g., GeoEye, Worldview-2) to middle spatial resolution imagery (e.g., ALOS/AVNIR-2, SPOT/HRG, Landsat series) and low spatial resolution imagery (e.g., Terra/MODIS, Aqua MODIS). In the image classification, however, the optimal classification technique and the spatial resolution demanded change with the design of project or analysis. Therefore, to improve the LCLU maps efficiently, it is necessary to systematize information based on many case studies.

In recent years, in the image classification of remotely sensed data, the researches using the object-oriented image classification has increased. The object-oriented image classification have multi-dimensional features, such as multiple object and textural information (Baatz and Schäpe, 2000). Some research has suggested that the image patterning and relational aspects of multi-dimensional data are not properly characterized by traditional methods, such as the maximum likelihood method, because of complexity caused by factors such as the variability of features in each land cover type, the cause of the multidimensionality as the large number of image features, and possible correlation among features to be classified (Melgani and Bruzzone, 2004).

In the classification of multi-dimensional data, data mining and ensemble learning are considered effective approaches (Gislason et al., 2006). Many ensemble methods have been proposed (Hansen and Salamon, 1990; Benediktsson and Swain, 1992), with the most widely used being boosting (Schapire, 1999) and bagging (Breiman, 1994). An ensemble is itself a supervised learning algorithm, because it can be trained and then used to make predictions. The trained ensemble, therefore, represents a single hypothesis. This hypothesis, however, is not necessarily contained within the hypothesis space of the models from which it is built. Thus, ensembles can be shown to have more flexibility in the functions they can represent. In the main subject of this study, the object-oriented image classification which used Bagging (BAGG), Boosting (BOOST), and Random Forest (RF) was tried. Moreover, Classification and Regression Tree (CART) was used as a candidate for comparison of classification accuracy. We performed the image classification from the ensemble learning methods, and aimed at clarifying the optimal classification technique from accuracy comparison. Simultaneously, in each classification technique, the effect of the image features was also evaluated by multiple data analysis.

METHODS

Study area and data correction

The study area was Sado Island, located in Niigata Prefecture, Japan (37°50’–38°20’N, 138°10’–138°33’E; Fig. 1). The island covers 855.26 km^2 and has a population density of 74 people per square kilometer. Elevation ranges from 1 to 1,165 m above sea level. Sado Island is divided into two regions: the Osado region and the Kosado region (Fig. 1). The Osado region has higher mountains and small settlements with paddy fields. The Kosado region has low mountains and gently sloping valleys with paddy fields. Forest covers 76% of the island and is composed mostly of secondary forest dominated by oak (Quercus L.) and conifer plantations of Japanese cedar (Cryptomeria japonica) and Japanese red pine (Pinus L.). The mean annual temperature is 14.3°C, and the mean annual precipitation is 1301.3 mm. We established a test site (size: 9 km×9 km) in the Kosado region and performed a field survey (Fig. 1) in which we collected vegetation data and located the data using a Global Positioning System device (GPS: GPSMAP60Cx, GARMIN, Inc.). Location data were used for training samples and validation of the LCLU mapping. Image classification was conducted for this test site.

SPOT/HRG satellite imagery, acquired on 3 June 2006, was used for LCLU mapping. The SPOT data were pan-sharpened at 2.5m spatial resolutions and geo-referenced to the Universal Transverse Mercator system using ERDAS IMAGINE 9.3 (ERDAS, Inc.) with a root mean-square error within one pixel. Shade caused by topographic relief (i.e., topographic effect) can create serious obstacles for the analysis of remote sensing data. A dual partitioning regression method was applied to correct topographic effects (Sakamoto et al., 2009). We used the digital numbers (DNs) for the observed wavelength bands (band 1: green, band 2: red, band 3: near infrared (NIR), band 4: shortwave infrared (SWIR)).

Image classification and accuracy assessment

The object-oriented image classification was applied for LCLU mapping. We used the commercial software eCognition ver. 4 (Definiens Imaging, Inc.) to conduct the object-oriented method. Using this method, the object-oriented classification can be performed in lieu of traditional methods, such as pixel-based classification. Segmentation represents the first step of any object-oriented classification. In this segmentation technique, individual pixels are considered the initial regions. A region-growing procedure for segmentation was used for image classification. In eCognition, the segmentation is a bottom-up, region-merging technique, where the smallest object contains one pixel. In this process, adjacent pixels in image objects are totaled by considering spectral and shape features. This process stops when the smallest growth exceeds the threshold defined by the scale parameter (Benz et al., 2004). Segmentation analysis, at fine and coarse scales, is important in the object-oriented image classification to extract the boundaries of dominant objects occurring at corresponding scales (Hall et al., 2004).

Pan-sharpened SPOT/HRG data were then segmented into homogeneous objects. In the segmentation process, the object-amount, object-form, and object-size features were used. The parameters were scale (10), shape (0.3), and compactness (0.4). For the image classification features, we selected various features, and built four set of data to evaluate the effect of image features in LCLU mapping (Table 1). The number of features is increased from data set I through IV. In classification, we defined eight classes: broad-leaved deciduous forest, Japanese cedar, Japanese red pine, bamboo forest, paddy field, urban area, road, and bare land. We applied four classifiers: Classification and regression tree (CART), Decision tree with Boosting, Decision tree with Bagging, and Random Forest (details of these techniques are described in late sections). Classification training data were extracted from the area in which land cover could be checked using an aerial photograph from 2006, combined with field verification as reference data. The accuracy of LCLU maps was assessed using an aerial photograph and field survey data. Class accuracy (user's accuracy: UA), overall accuracy (OA), and kappa coefficient (KAPPA) were used for accuracy assessment.

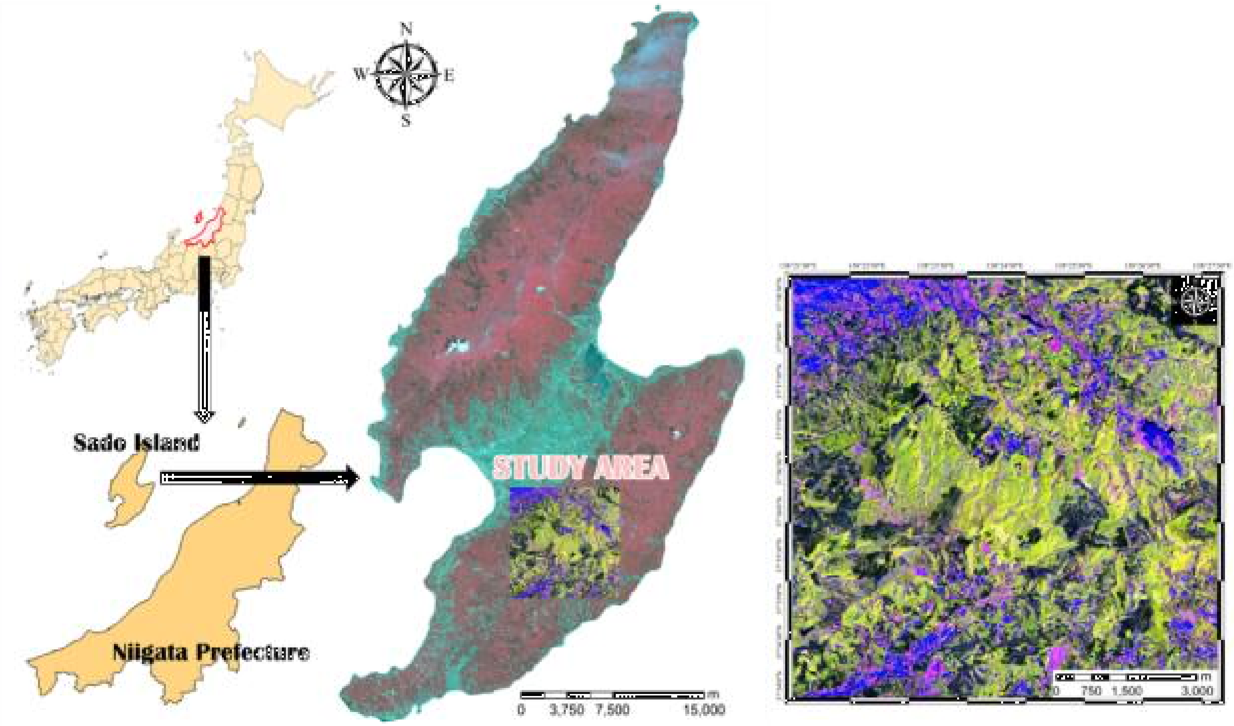

Fig. 1 Study area and SPOT/HRG (acquired in 2007/06/03) imagery within test site

Classification algorithms

Classification and regression trees (CART): Breiman et al. (1984) developed CART, a classification tree technique based on binary recursive splitting. CART divides data into homogeneous groups. In the classification tree analysis, the predicted outcome is the class to which the data belongs. The tree is developed initially having one node. The training data are divided into two groups and these nodes are each split into two child nodes. In this study, Gini impurity (Breiman et al., 1984) is a measure of how often a randomly chosen element from a set would be incorrectly labeled if it were randomly labeled according to the distribution of labels in the subset. We used the rpart package in R (Therneau and Atkinson, 1997) for CART image classification.

Decision tree with bagging (BAGG): Bagging is a method proposed by Breiman (1994) to improve the performance of prediction models. Given a classification model, bagging draws some independent samples with replacement from the available training set (bootstrap samples), fits a model to each bootstrap sample, and finally aggregates the all models by majority voting. Bagging tends to be a very effective procedure when applied to unstable learning algorithms (i.e., ‘‘small changes in the data can cause large changes in the predicted values’’, Breiman, 1994), such as classification and regression trees and neural networks. In the model construction, the out-of-bag (OOB) estimates were developed using the one-third of the data. When a bootstrap resample is drawn, about 37% of the data is excluded from the sample, but other data are replicated to bring the sample to full size. The portion of the data drawn into the sample in a replication is known as the ‘‘in-bag’’ data, whereas the portion not drawn is the ‘‘out-of-bag’’ data. We detected 500 bootstrap samples and the models were built using all image features. The R package caret which computes bagged tree models was used in our study.

Decision tree with boosting (BOOST): The idea of boosting appeared in the Machine Learning literature in the 1980’s. Boosting was proposed in order to combine the outputs of many ‘‘weak’’ classifiers thereby producing a powerful ‘‘committee’’, in an attempt to improve the generalization performance of weak algorithms. The various models are fitted to differently reweighted samples. At each step, those observations that were misclassified by the previous classifier have their weights increased, whereas the weights of those correctly classified are decreased. One of the most popular boosting algorithms was AdaBoost.M1 (Freund and Schapire, 1997). Since only two-class classification of AdaBoost.M1 corresponds, in this study, we employed the R package caret which computes multi-class boosting.

Random Forests (RF): The Random Forest method, as proposed by Breiman (2001), is an ensemble classifier that consists of numerous decision trees; the output class is the mode of the classes for the individual trees. In training, the RF algorithm, like CART, creates multiple trees, each trained on a bootstrapped sample of the training data. It then searches only across a randomly selected subset of input variables to determine a split. In this study, we developed RF models with 500 classification trees. The number of features used at each tree split was optimized based on OOB estimates of error (Liaw and Wiener, 2002). The OOB estimates were developed using the one-third of the data that was randomly excluded from the contraction of each of the 500 classification trees. Although the algorithm of RF resembles BAGG, the difference is selecting at random rather than using all image features. The RF model was developed using training data and the accuracy was evaluated using test data. We used the Random Forest package in R (Liaw and Wiener, 2002) for the RF image classification.

Table 1. Data set for image classification

Data set I	Statistic within objects	
	Mean	
Data set II	Statistic within objects	
	Mean; Standard deviation; Ratio	
Data set III	Statistic within objects	Shape patterns of objects
	Mean; Standard deviation; Ratio	Area; Length; Width; Length/Width; Compactness; Shape index; Density
Data set IV	Statistic within objects	Shape pattern of objects
	Mean; Standard deviation; Ratio	Area; Length; Width; Length/Width; Compactness; Shape index; Density
	Texture within objects	
	Homogeneity; Contrast; Dissimilarity; Entropy; Angle Second Moment; Mean; Standard deviation; Correlation	

RESULTS

The results of the classification accuracy in each classification technique and data set were shown in Figure 2 and Table 2. The combination which was the best as for classification accuracy was the classification of the data set II which used RF algorithm. Subsequently, in the classification which applied BOOST and RF algorithm in the data sets III and IV, it was high-precision. In the classification using CART method, the Kappa statistics value of every data set was less than 0.6. The CART method shows that prediction power is weak compared with ensemble learning methods, such as BOOST and RF algorithm. BAGG algorithm showed the middle accuracy in the classification techniques of this study.

The difference in the classification accuracy between the used amounts of the features (data set) was remarkable between the data set I and other data sets. When there are few image features used for the classification, this results are shown that a highly precise classification is not expectable. Moreover, even if there was much information for the image classification, the accuracy fell. In the data set I, the accuracy of BOOST algorithm was better than other classification techniques, and the accuracy of RF algorithm was high in the data set II. Moreover, in the data sets III and IV, BOOST and RF were the same accuracy in Kappa value.

DISCUSSION

This research is a case study of the object-oriented image classification to a SPOT/HRG pan-sharpened imagery. As the results of evaluating the classification accuracy obtained from CART, BAGG, BOOST, and RF algorithm, high accuracies were acquired in BOOST and RF. The lowest classification technique was CART method in our results. This result shows that the ensemble learning is effective in the image classification for LCLU map. RF algorithm was highly precise than BAGG which the algorithm resembles. The difference between RF and BAGG is in the decision mechanism of the features to be used in classification. Using RF algorithm, in order to sample the image features at random, it led to mitigation of the correlation which exists between variables. This is clear also from the difference in the accuracy between data sets. In the data set I with few image features to be used in the classification, although the accuracy of RF and BAGG was comparable, when the image features increased, the difference arose in the classification accuracy (Table 2).

From the classification accuracy according to data set, the high-precision image classification has been performed by the data sets II and III (Table 2). The data set II is a basic statistics value of the objects, and consists of the averaged DNs within an object, and its standard deviation. The data set III consists of the basic statistics values and object patterns (e.g., shape of object, edge density, compactness and so on) of an object. We suggest the necessity of using the basic statistics value of an object at worst, when performing the object-oriented image classification for LCLU mapping. The data set I of only the averaged DNs within an object had the lowest classification accuracy among four data sets. Moreover, the classification accuracy fell also the data set IV which applied the textural features within an object. This means that the textural features cannot be used effectively by the SPOT data with 2.5 m of spatial resolution. To use the textural features in the object-oriented image classification, it is necessary to employ high resolution data rather than SPOT data.

In every data set, BOOST and RF algorithm took high classification accuracies (Table 2). It will be necessary to use both classification techniques by the image classification using ensemble learning. Especially when the image features to be used are in multi-dimensions, RF algorithm serves as an effective technique for LCLU mapping. On the other hand, when there are few image features to be used in the classification (e.g., the data set I), the accuracy of BOOST excels RF algorithm. This can be explained from the difference in both algorithms. BOOST algorithms consist of iteratively learning weak classifiers with respect to a distribution and adding them to a final strong classifier. Therefore, predictive accuracy is secured when there are few image features to be used. In RF algorithm, reweighting is not performed to weak classifiers. This mechanism is considered to have led to the difference in the classification accuracy in the data set I. Since correlation between variables becomes a problem when using a certain amount of image features, RF algorithm becomes effective. Finally, according to the numerousness of the image features which can be used in the classification process, it is necessary to decide which ensemble learning method is adopted.

CONCLUSIONS

In the object-oriented image classification with ensemble learning method for accurate LCLU mapping, we proposed the optimal classification technique according to a situation. When the image features to be used in classification is limited, BOOST algorithm is recommended. When the image features in the classification have multi-dimension, RF algorithm could be recommended for LCLU mapping. There is no merit which uses CART and BAGG algorithm for image classification from our results. Furthermore, textual features are not effective under the image classification of SPOT/HRG pan-sharpened data. In order to use textual features effectively, it is necessary to examine high resolution satellite data (e.g., IKONOS, GeoEye-1, Worldview-2 and so on). In the spatial resolution which tested in our study, the statistical features such as averaged DNs and standard deviation are effective for the image classification. The image features of object patterns which include shape index, edge density, and compactness are also effectual variables in the object-oriented image classification. From our results, we suggest that the basic statistic values within objects are indispensable, when using the object-oriented image classification. However, this result is only one of the case studies, and the further verification will be required for accurate LCLU mapping.

REFERENCES:

Baatz, M., Schäpe, M. 2000. Multiresolution segmentation- an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T. Griesebner, G. (eds) Angewandte Geographische Informations- Verarbeitung XII. Wichmann Verlag, Karlsruhe, pp. 12–23

Benediktsson, J.A., Swain, P.H. 1992. Consensus theoretic classification methods. IEEE Trans. Systems Man Cybernet 22, pp. 688-704.

Benz, U., Hofmann, P., Willhauck, G., Lingenfelder, L. Heynen, M. 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS Journal of Photogrammetry and Remote Sensing 58, pp. 239-258.

Breiman, L. 1994. Bagging predictors. Technical Report No. 421, Department of Statistics, University of California, Berkeley.

Breiman, L. 2001. Random Forests. Machine Learning 40, pp. 5–32.

Breiman, L., 1994. Bagging predictors. Technical Report No. 421, Department of Statistics, University of California, Berkeley.

Breiman, L., Freedman, J., Olshen, R. 1984 Classification and Regression Trees. Chapman & Hall, New York, 368 pp.

Freund, Y., Schapire, R.E. 1997. A Short Introduction to Boosting. Journal of Japanese Society for Artificial Intelligence, 14, pp. 771-780.

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R. 2006. Random Forests for land cover classification. Pattern Recognition Letter 27, pp. 294-300.

Hall, O., Hay, G.J., Bouchard, A., Marceau, D.J. 2004. Detecting dominant landscape objects through multi scales: an integration of object-specific methods and watershed segmentation. Landscape Ecology 19, pp. 59-76.

Hansen, L.K., Salamon, P. 1990. Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence 12, pp. 993-1001.

Liaw, A., Wiener, M. 2002. Classification and regression by Random Forests. R News. 2/3: 18-22.

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.

Pal, M., Mather, P.M. 2004. Assessment of the effectiveness of support vector machines for hyperspectral data. Future Generation Computer Systems, 20, pp. 1215-1225.

Sakamoto, K., Nakayama, D., Matsuyama, H. 2009. New topographic correction methods of satellite image in the season of low solar elevation. Journal of The Remote Sensing Society of Japan 29, pp. 471-484.

Schapire, R.E. 1999. A brief introduction to boosting. In: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence. pp. 1401–1406.

Szuster, B.W., Chen, Q., Borger, M. 2011. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. Applied Geography 31, pp. 525-532.

Therneau, T. M., Atkinson, E. J. 1997. An introduction to recursive partitioning using the rpart routine. Technical Report 61, Section of Biostatistics, Mayo Clinic, Rochester, 52 pp.

Xie, Y., Sha, Z.Y., Yu, M. 2007. Remote sensing imagery in vegetation mapping: a review. Journal of Plant Ecology. 1, pp. 9-23.

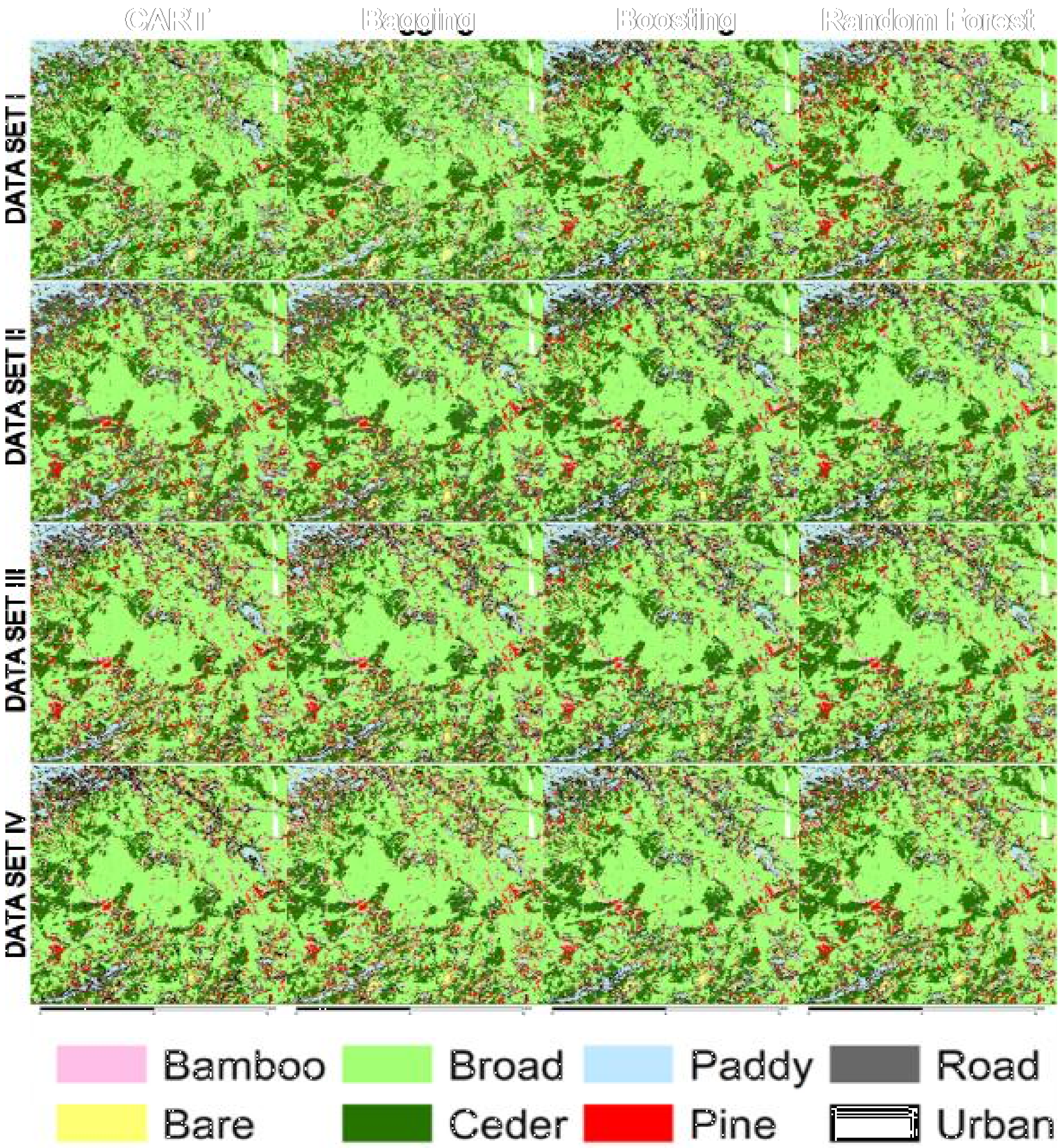

Figure 2. LCLU maps of the classification results with each ensemble learning algorithm and data set.

Table 2. Accuracy comparison by the overall accuracy, kappa, and the class accuracy.

			OA		Class accuracy (User's Accuracy: UA) (%)							
Rank	DATA	Algorithm	(%)	K	Bamboo	Bare	Broad	Cedar	Paddy	Pine	Road	Urban
1	II	Random Forest	72.4	0.684	76.1(2)	64.0(4)	**85.8(1)**	63.7(3)	92.6(4)	70.5(3)	59.2	**83.0(1)**
2	III	Random Forest	71.2	0.671	72.1(4)	60.0	82.7(4)	63.1(4)	90.1	67.6(4)	66.6	72.2
3	III	Boosting	71.0	0.670	67.4	58.2	85.2(2)	63.8(2)	92.2	66.7	66.9(3)	75.0(3)
4	IV	Random Forest	70.8	0.661	61.2	55.6	84.9(3)	**63.9(1)**	93.1(3)	**79.6(1)**	**72.3(1)**	75.4(2)
5	IV	Boosting	69.8	0.655	70.7	52.7	81.8	62.1	89.2	66.2	69.1(2)	68.4
6	III	Bagging	68.5	0.640	70.5	61.0	81.8	60.1	93.7(2)	57.6	59.6	69.8
7	I	Boosting	68.3	0.634	66.6	66.7(3)	73.3	56.2	74.6	76.1(2)	66.7(4)	66.7
8	II	Boosting	67.3	0.630	75.3(3)	64.0	81.1	61.9	**93.8(1)**	60.5	47.6	67.9
9	IV	Bagging	66.3	0.615	63.5	53.3	81.6	61.7	92.1	62.8	59.0	61.2
10	I	Random Forest	65.3	0.604	63.5	73.1(2)	77.8	56.2	71.5	56.3	57.9	69.8
11	II	Bagging	65.2	0.601	**87.7(1)**	61.0	81.8	55.2	90.5	57.7	47.0	73.4(4)
12	I	Bagging	65.2	0.600	62.9	**79.7(1)**	79.2	55.7	71.7	59.0	56.1	67.7
13	IV	CART	61.7	0.563	57.9	47.3	81.8	55.7	76.9	55.7	58.9	52.2
14	III	CART	61.7	0.563	57.9	47.3	81.8	55.7	76.9	57.7	58.9	52.2
15	II	CART	60.1	0.544	55.1	54.2	81.8	55.7	85.7	57.7	34.4	61.6
16	I	CART	57.6	0.515	60.0	38.0	80.9	48.1	67.7	60.0	50.7	64.0