

Decision-tree-based Mapping of Erosion-prone Areas in Hilly Regions of Kangwon Province, North Korea

Tri Dev Acharya,^{1,2} Kyung Nam Kim,³ Kwang Youn Lee,³
Anoj Subedi,⁴ He Huang,² and Dong Ha Lee^{1*}

¹Department of Civil Engineering, Kangwon National University,
1 Kangdaehak-gil, Chuncheon 24341, Republic of Korea

²School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture,
15 Beijing 100044, China

³Research Institute for Gangwon 5, Jungang-ro, Chuncheon 24265, Republic of Korea

⁴Institute of Forestry, Pokhara Campus, Tribhuvan University, Hariyokharka, Pokhara 33700, Nepal

(Received November 14, 2019; accepted November 28, 2019)

Keywords: J48, pruned decision tree, forest, open land, erosion, Landsat, Kangwon, North Korea

After the division of the Korean peninsula, North Korea overexploited their natural resources especially the forest. It lost about 23% of the total forest from 1990 to 2011, which continues today. However, the country is inaccessible to monitor such changes. Hence, in this study, we aim to use Landsat 8 imagery with the aid of Google Earth to map erosion-prone areas in a subset area of Kangwon Province, North Korea. Pruned Decision Tree (DT) modeling was used in selecting the optimum ratio/index and threshold based on ground truth points extracted for Landsat scenes from May, October, and both months combined. Pruned DT resulted in applying the normalized green, near-infrared (NIR), green ratio vegetation index (GRVI), red-green ratio index (RGRI), infrared percentage vegetation index (IPVI), and slope with the optimum threshold for the segmentation of the study area with reasonable accuracy. The result shows that combining the ground truths from different seasons resulted in rules giving higher overall accuracy (OA) and kappa coefficient than the individual rule results. However, interchanging ground truths of different months is not effective. On average, out of the total land, high and medium erosion-prone areas are 15 and 20%, respectively. The remaining 65% is covered by forest. The result can be useful for estimating loss and restoring resources such as forest and land in the future.

1. Introduction

Following the division of the Korean peninsula soon after World War II, North Korea remained isolated to the rest of the world. Natural resources, such as forests, have been exploited extensively in many parts of North Korea owing to a limited supply of energy. According to the United Nations Environment Programme, it was reported that in 2003, forest covered about 73% of the total area of North Korea, 70% of which is on slopes greater than

*Corresponding author: e-mail: geodesy@kangwon.ac.kr
<https://doi.org/10.18494/SAM.2019.2707>

20 degrees.⁽¹⁾ Between 1990 and 2000, it lost 127 thousand hectares of forest with an annual average change rate of -1.7% . Moreover, between 2000 and 2010, it again lost 127 thousand hectares of forest with an annual average change rate of -2% .⁽²⁾ In total, between 1990 and 2011, North Korea lost 23.12% of its forest cover, which is the highest among countries in East Asia.^(2,3) The accelerated deforestation was due to the economic crisis in the 1990s; people turned to the forest to provide firewood and food. Also, the government control over croplands has made people convert cultivable lands in remote hilly forest areas secretly. These activities have put a lot of pressure on hilly areas, thus leading to soil depletion and erosion. As the nation cannot be freely accessed, field work to investigate the position, type, and activity of erosion to develop an inventory and continuous monitoring system is impossible. However, satellite-based remote sensing technology can make this possible. In this study, we assessed the impact of deforestation on hillslope erosion using remote sensing technologies.

Remote sensing is a powerful and cost-effective technology that allows us to collect data and assess the spatial and temporal dynamics of Earth's surface processes and hazards.^(4–8) Remotely sensed data from legacy satellites such as Landsat have been continuously obtained for the past forty years. With such a large collected redundant database, it is possible to monitor deforestation and thus erosion-prone areas in remote and inaccessible areas. For mid-resolution satellite imagery, two approaches have been widely used to map the spatial distribution and characteristics of vegetation cover. The first is by thresholding remote-sensing-based vegetation indexes, and the second is the classification of multiband satellite imagery either by using clustering techniques (i.e., ISODATA unsupervised classification) or using training data (i.e., supervised classification). Although classification techniques have been widely used, they are unable to account for the image variation in land cover, cloud cover in different seasons, and weather conditions. Moreover, the collection of data for each time and condition is costly and time-consuming.

To address these drawbacks, new sets of input variables were derived for robust mapping. Owing to their simplicity, remote-sensing-image-based indices, such as NDVI, have widely been used as an effective approach to map certain land cover types.^(7,9–12) Despite being simple and fast, a major issue of this approach is the requirement of a threshold that could be used to segment target and nontarget classes. Depending on the ratio of target to nontarget classes in the study area, neither standard nor automatic thresholds such as Otsu's are always successful. Thus, choosing the exact optimal threshold is also a challenging task depending on the large amount of ground truth and expert knowledge. To assign an optimum threshold for a given set of attributes, the Decision Tree (DT) is better suited. DT establishes a rule for the determination of a target and exhibits a high accuracy across many environments, allocating more homogenous datasets based on binary splits.⁽¹³⁾ Binary splitting nodes are based on conditions of explanatory variables that can be easily understood and implemented in Geographic Information System (GIS).⁽¹⁴⁾

Despite few studies on land cover and forest monitoring in the past,^(15–17) not much studies have efficacy in the assessment of erosion-prone areas using remote sensing and DT modeling approaches for highly rugged steep mountainous landscapes. Moreover, studies that consider the remote sensing index in North Korea are also rare. Hence, the aim of this study is to map

erosion-prone areas over an area of Kangwon Province, North Korea, using Landsat Imagery and DT. Satellite-based digital elevation model (DEM) and imagery were selected, indices were derived along with slope, ground truth points were carefully chosen for optimum threshold and band using DT, and accuracy assessment was carried out. The overall flow of the study is shown in Fig. 1.

2. Study Area and Data

2.1 Study area

A rectangular area from Kangwon Province, North Korea, has been selected for this study (Fig. 2). The area is geographically bounded between $38^{\circ}18'41.45''\text{N}$ to $38^{\circ}57'59.38''\text{N}$ and $126^{\circ}47'49.73''\text{E}$ to $128^{\circ}0'14.97''\text{E}$ and is 7520 sq. km. It has a rugged topography and a dendritic drainage pattern. The elevation ranges from 0 to 1510 m. Most of the area is forested with deforested patches, agricultural lands, and forest roads with few urban area. It also has dammed reservoirs and coastal areas covering $\sim 2.45\%$ of its total area. It was selected for two reasons; first, it is densely forested with extensively eroded landscapes that can be seen in Google Earth imagery, and it has very similar climate and topography to Gangwon Province of South Korea, so that a field observation of an area in South Korea adjacent to the study area can be carried out to better understand the geologic and geomorphic environments.

2.2 Data

Three types of data, namely, Shuttle Radar Topography Mission (SRTM) Global DEM, Landsat Operational Land Imager (OLI) imagery, and high-resolution images available from

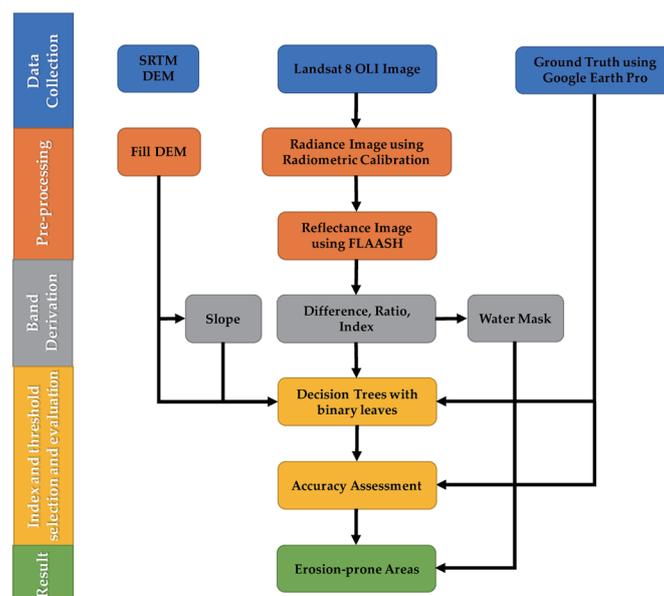


Fig. 1. (Color online) Overall flow of the study to map erosion-prone areas.

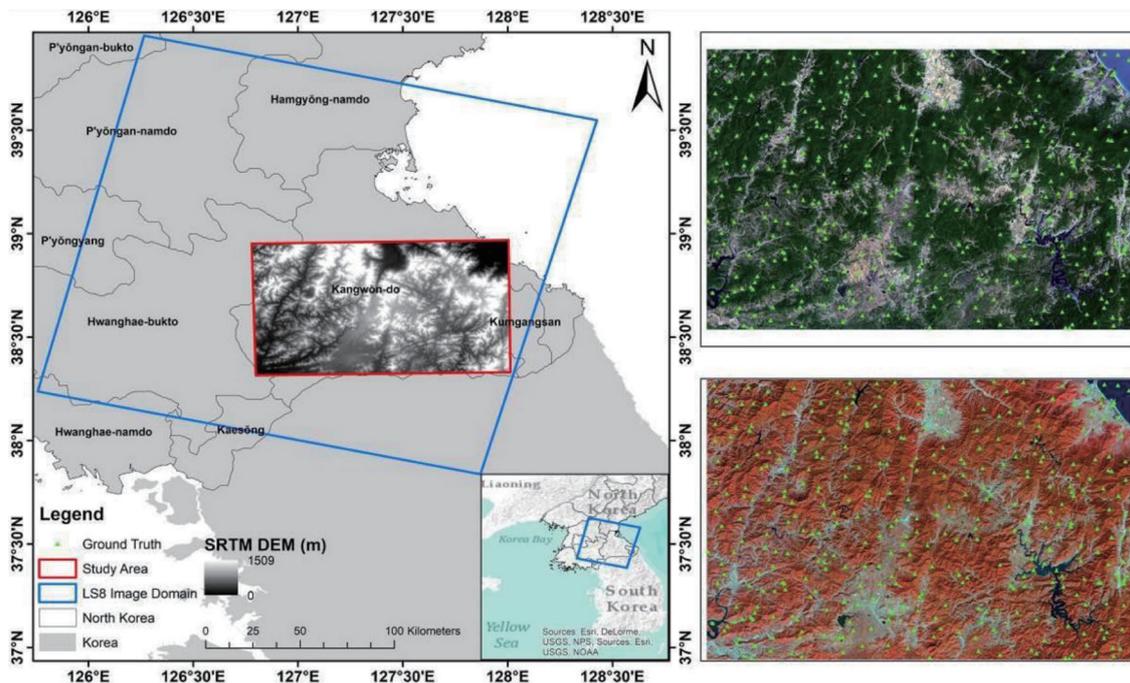


Fig. 2. (Color online) Location of the test area in North Korea with DEM and ground truth over Landsat images (right): natural color composite (R-G-B) acquired on 19 May 2016 (top) and false-color composite (NIR-G-B) acquired on 10 October 2016 (bottom).

Google Earth (GE) ProTM (Google Inc., Menlo Park, CA, USA), were used. The selection was carried out carefully considering cloud cover, phenology, and dryness of the ground that covers most parts of Kangwon Province, North Korea. Landsat data from early May and October were selected because the study area in these images was observed with the lowest vegetation cover, thereby providing a relatively unobstructed view of the surface. Both Landsat and DEM were resampled to a 30-m-resolution subset scene of 3541 rows and 2360 columns after preprocessing the data.

For the extraction of elevation for the scene, the SRTM Global DEM of 30 m⁽¹⁸⁾ was obtained from the OpenTopography portal.⁽¹⁹⁾ Its version 3.0 product is void-filled and the most complete high-resolution digital topographic database of Earth. Their evaluation showed good representation with other data in hilly⁽²⁰⁾ and coastal⁽²¹⁾ regions of Korea. The data was used to provide the ground elevation and slope of the study area.

Two Landsat scenes from path 116 and row 33 were obtained from the United States Geological Survey (USGS) Global Visualization Viewer (GLOVIS) portal.⁽²³⁾ Table 1 shows the metadata of the Landsat scene used. The Level 1 terrain-corrected multiband image contained coastal blue, blue, green, red, near-infrared (NIR), shortwave infrared 1 (SWIR1), and shortwave infrared 2 (SWIR2) bands at a resolution of 30 m. The digitally numbered images were first converted to top-of-atmosphere radiance and then reflectance using the Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) atmospheric correction tool in Environment for Visualizing Images (ENVI) version 5.3 (Exelis Visual Information Solutions, Boulder, CO, USA). The required coefficients and values were obtained from the Landsat response function, metadata file, and SRTM DEM.

Table 1
Metadata of Landsat 8 OLI bands used.⁽²²⁾

WRS Path/Row	Bands	Date acquired	Scene cloud cover (%)	Land cloud cover (%)
116/33	Coastal Aerosol (CA): 0.435–0.451	2016/05/19	7.86	0.03
	Blue: 0.452–0.512			
	Green: 0.533–0.590			
	Red: 0.636–0.673	2016/10/10	1.65	1.34
	NIR: 0.851–0.879			
	SWIR1: 1.566–1.651			
SWIR2: 2.107–2.294				

Google Earth has been widely used in various studies,^(5,24–28) even in North Korea,⁽¹⁷⁾ as a reference and validation data. Because access to the area is limited, this study used Google Earth imagery for training and validation purposes in this study. A total of 350 points were randomly sampled in the study area for forest cover, hill, and open land surface (Fig. 2).

3. Methods

3.1 Remote sensing indices

Various vegetation indices based on remote sensing products or technologies were used to characterize vegetation cover (Table 2). All these indices were taken from the Harris Geospatial Docs Centre website⁽⁴¹⁾ where more details on them and their applications can be found.

3.2 DT

DT is a hierarchical model that computes the probability of occurrence in various situations based on known information (Fig. 3). It can also be a rule-based structure giving an answer for what happens if something in the model is modified. The main advantage of the method is that it is a graphical representation of the overall model that could be easily constructed and is an intuitive application of probability analysis. However, it does not allow multiple outputs and is limited to reducing the noise in data.⁽⁴²⁾ There are various available algorithms such as Classification and Regression Tree,⁽⁴³⁾ Chi-square Automatic Interaction Detector Decision Tree,⁽⁴⁴⁾ ID3, J48,⁽⁴⁵⁾ and C4.5.⁽⁴⁶⁾

The classification of a new item in the algorithm first requires a DT based on the attribute values of the available training data. On the basis of the available set of items in the training data, it identifies the attribute that classifies the various instances most clearly. The feature that tells us most about the data instances, which could lead to the best classification, is said to have the highest information gain. On the basis of the possible values of the feature, branches are terminated, and a target value is assigned. In other cases, the algorithm searches for other attributes that provide the highest information gain. The process continues until a clear decision regarding the combination of attributes that gives us a rule for the determination of a target value is achieved. With the help of this DT, all the respective attributes and their values undergo checking, thereby assigning or predicting the target values of all new instances.

Table 2
Multiband indices used in this study.

Multiband Index	Formula	Reference
Difference Vegetation (DV)	$\text{NIR} - \text{red}$	29
Green Difference Vegetation (GDV)	$\text{NIR} - \text{green}$	30
Simple Ratio (SR) also known as Ratio Vegetation Index (RVI)	NIR / red	31
Green Ratio Vegetation Index (GRVI)	$\text{NIR} / \text{green}$	30
Red Green Ratio Index (RGRI)	$\text{Red} / \text{green}$	32
Infrared Percentage Vegetation Index (IPVI)	$\text{NIR} / (\text{NIR} + \text{red})$	33
Normalized NIR (NNIR)	$\text{NIR} / (\text{NIR} + \text{red} + \text{green})$	30
Normalized Red (NR)	$\text{Red} / (\text{NIR} + \text{red} + \text{green})$	30
Normalized Green (NG)	$\text{Green} / (\text{NIR} + \text{red} + \text{green})$	30
Normalized Difference Vegetation Index (NDVI)	$(\text{NIR} - \text{red}) / (\text{NIR} + \text{red})$	34
Green Normalized Difference Vegetation Index (GNDVI)	$(\text{NIR} - \text{green}) / (\text{NIR} + \text{green})$	35, 36
Soil Adjusted Vegetation Index (SAVI)	$[(\text{NIR} - \text{red}) / (\text{NIR} + \text{red} + L)] \times (1 + L)$, where $L = 0.5$	37
Green Soil Adjusted Vegetation Index (GSAVI)	$[(\text{NIR} - \text{green}) / (\text{NIR} + \text{green} + L)] \times (1 + L)$, where $L = 0.5$	30
Enhanced Vegetation Index (EVI)	$2.5 * (\text{NIR} - \text{Red}) / (\text{NIR} + 6 * \text{Red} - 7.5 * \text{Blue} + 1)$	38
Normalized Difference Water Index (NDWI) also known as Normalized Difference Moisture Index (NDMI)	$(\text{NIR} - \text{SWIR1}) / (\text{NIR} + \text{SWIR1})$	39, 40

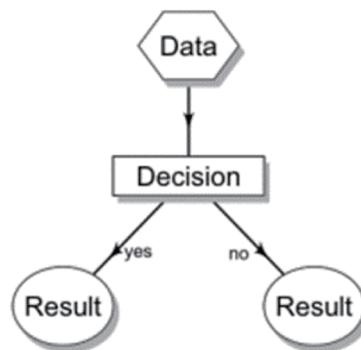


Fig. 3. Schematic of a basic DT.

If DT becomes complicated, the tree structure and rules might be difficult to interpret and may result in a low accuracy. A simple technique of pruning is not to continue splitting if the nodes receive a very small or minimum number of instances per leaf. Building a full tree and pruning back are more reasonable than trying to forward pruning as a statistical test can also be applied to each stage. The test considers a confidence factor of 0.25 by default, where a small value incurs more pruning. It simplifies the results and gives manageable trees in addition to overfitting the training data without affecting the performance considerably.⁽⁴⁷⁾

In this study, pruned DT was used to generalize data from the root, i.e., the best attribute to split the first decision. We used a minimum object of 10% of ground truth as the pruning

condition, i.e., 35 for May and October and 70 for both months combined. It will help to pick a layer of band or ratio with the optimum threshold to segregate the target. J48 DT, a java implementation of C4.5 DT in Waikato Environment for Knowledge Analysis (WEKA) Version 3.8.1 (University of Waikato, Hamilton, New Zealand),⁽⁴⁵⁾ was used for this study.

3.3 Accuracy assessment

To prevent overfitting and biased assessment, a k -fold cross-validation method has widely been used. For given k folds, the method divides whole data into k parts, one piece, and trains the rest. However, there are many ways to divide the data into k -folds. The stratified n -fold cross-validation is preferred to ensure that each fold has the right portion of each class value, which reduces the variance in the estimate. In this study, we used the stratified 10-fold cross-validation implanted in WEKA.

An accuracy assessment was carried out using a confusion matrix.⁽⁴⁸⁾ Table 3 shows a typical layout of the matrix, with the columns representing the classified data and the rows representing the ground truth values, although both can be interchanged. In the table, the diagonal elements are the pixels of agreements, whereas the off-diagonal elements are disagreements. From the confusion matrix, overall accuracy (OA) was calculated by dividing the sum of the entries along the diagonal by the total number of validation samples. OA represents the current predictions and ranges from 0 to 1, where 1 is perfect. However, it does not consider the agreements between datasets, which are due to chance alone. Thus, the kappa coefficient (kappa) of the agreement was also derived from the confusion matrix. Kappa is a measure of agreement based on the difference between the actual agreement in the error matrix and the chance agreement.⁽⁴⁸⁾ Usually, kappa can range from -1 to $+1$, where 0 represents the degree of agreement that can be expected from random chance and 1 represents a perfect agreement between the raters. In very rare cases, kappa can show a negative value, which signifies that there is no effective agreement between the two raters.⁽⁴⁹⁾ OA and kappa are calculated using Eqs. (1) and (2), respectively. Aside from OA, producer's accuracy (PA) and user's accuracy (UA) were also calculated on the basis of the confusion matrix.

$$\text{Overall accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (1)$$

$$\text{Kappa coefficient} = \frac{\text{Observed accuracy} - \text{Chance agreement}}{1 - \text{Chance agreement}} \quad (2)$$

Table 3
Typical layout of a confusion matrix.

		Classified	
		Target	Nontarget
Ground Truth	Target	True Positive	False Negative
	Non target	False Positive	True Negative

3.4 Erosion-prone area

In this study, three conditions were considered to identify erosion-prone areas. First, erosion occurs in open land; second, erosion is accelerated in hilly areas; third, erosion is controlled by forest/vegetation cover. Hence, three separate maps were prepared and overlaid as per their relationship to erosion. To better understand the erosion hazard, open space was given a value of 1, hill as 1, and nonforest areas as 1, such that their sum will give the hazard, as shown in Table 4. Pixels fulfilling all the conditions were high erosion-prone areas and with two conditions fulfilled as medium and only one or none as low.

4. Results and Discussion

After preprocessing the data, different band ratios and indices were derived from Landsat 8 images as per Table 2. A similar slope in degree was also derived from DEM. For a better analysis of the erosion-prone areas, water bodies were masked from the study area using NDWI. For the ground truth points, the corresponding bands, indices, and slopes were extracted for both 20160519 and 20161010 scenes. In addition to the two extracted data, an additional dataset combining both May and October extracts was also prepared to understand the generalized ability of derivative bands in all-year-round scenes. All these comma-separated values were converted to the attribute-relation file format (ARFF) for input in WEKA.

With the above-mentioned ARFF points, first, the ground truth data were evaluated for their baseline accuracy by stratified 10-fold cross-validation in WEKA. The baseline accuracy was tested using the ZeroR rule, which predicts the basic model for a class, i.e., if the data were all mode class, what could be the OA? This gives the idea of how the mode class is in the input data. From Table 5, it can be understood that out of the total ground truth data, 78.86% (276) are hills, 60.86% (213) are nonopen land and, 58% (253) are forest class.

For each class, open land, hill, and forest, first, J48 pruned trees with a minimum of 10% ground truth data for two objects were built for three conditions i.e., May, October, and the two

Table 4

(Color online) Overlaying class for determination of erosion-prone hazard (low: green; medium: yellow; red: high).

Open land	0	0	0	0	1	1	1	1
Hill	0	1	0	1	0	1	0	1
Nonforest	0	0	1	1	0	0	1	1
Erosion-prone hazard	0	1	1	2	1	2	2	3

Table 5

(Color online) OA, number of leaves, and size of tree based on ground truth using pruned DT. White color is the lowest and deep color is the highest in value.

Ground Truth / Class	OA (%)			No. of leaves			Size of tree		
	Open land	Hill	Forest	Open land	Hill	Forest	Open land	Hill	Forest
May	86.00	92.57	87.14	12	8	5	23	15	9
October	89.71	86.57	93.14	15	15	9	29	29	17
Combined	89.86	91.71	89.43	22	20	24	43	39	47

months combined, as shown in Table 6. The table shows good OA of around 90% or more with relatively large numbers of leaves and trees. However, these trees are difficult to interpret. For example, Fig. 4 shows the J48 pruned DT for a forest in the 20161010 scene using the ground truth extracted from the scene itself. Hence, in this work, we focused on the use of binary left pruned DT for easy interpretation and application in mapping erosion-prone areas in the inaccessible area of North Korea. This helps in the regular short-interval monitoring of land covers such as surface water, open land, and forest in areas with a high rate of change.

As a result, from a pruned tree for 10% ground truth data, Table 6 shows the application of one band for each class segmentation. On the basis of Table 6, three different maps were derived for both the 20160519 and 20161010 scenes. Figures 5 and 6 respectively show the results of open land, hill, and forest cover for 20160519 and 20161010 scenes derived from the optimum band and threshold in Table 6. Variations in optimum ratio/index and threshold for the different classes are seen in the table. For the 20160519 scene, RGRI above 0.7822 was found to be the optimum band and threshold to segment open land from the rest. It was interesting that NG was able to segment both the hill and forest classes with different thresholds. As the May

Table 6
Pruned DT for open land, hill, and forest covers using ground truth data from May, October, and the two months combined.

	May	October	Combined
Open land	RGRI > 0.7822	GRVI <= 5.4451 GNDVI <= 0.6897	NNIR <= 0.711
Hill	NG <= 0.2105	Slope > 6.9263	Slope > 6.9263
Forest	NG <= 0.1692	IPVI > 0.8227	NNIR > 0.711

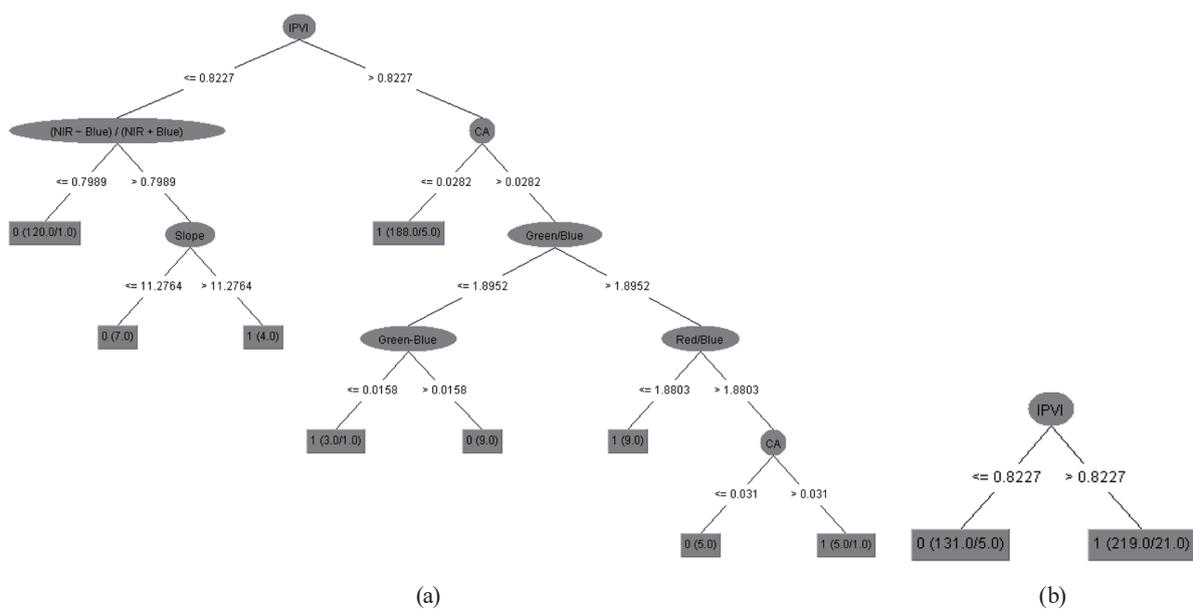


Fig. 4. J48 pruned DT for 20161010 scene produced by ground truth data of October: (a) DT with minimum 2 objects and (b) DT with minimum 35 objects, i.e., 10% of ground truth data.

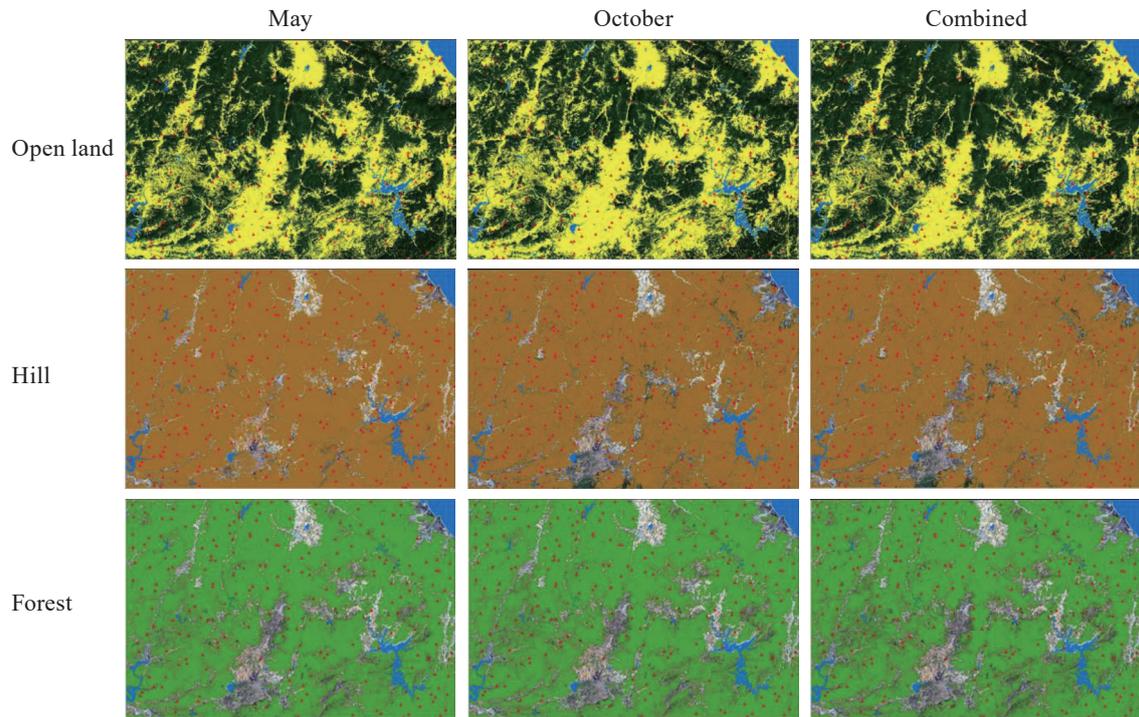


Fig. 5. (Color online) Open land, hill, and forest for 20160519 scene that was derived from the optimum band and threshold in Table 6.

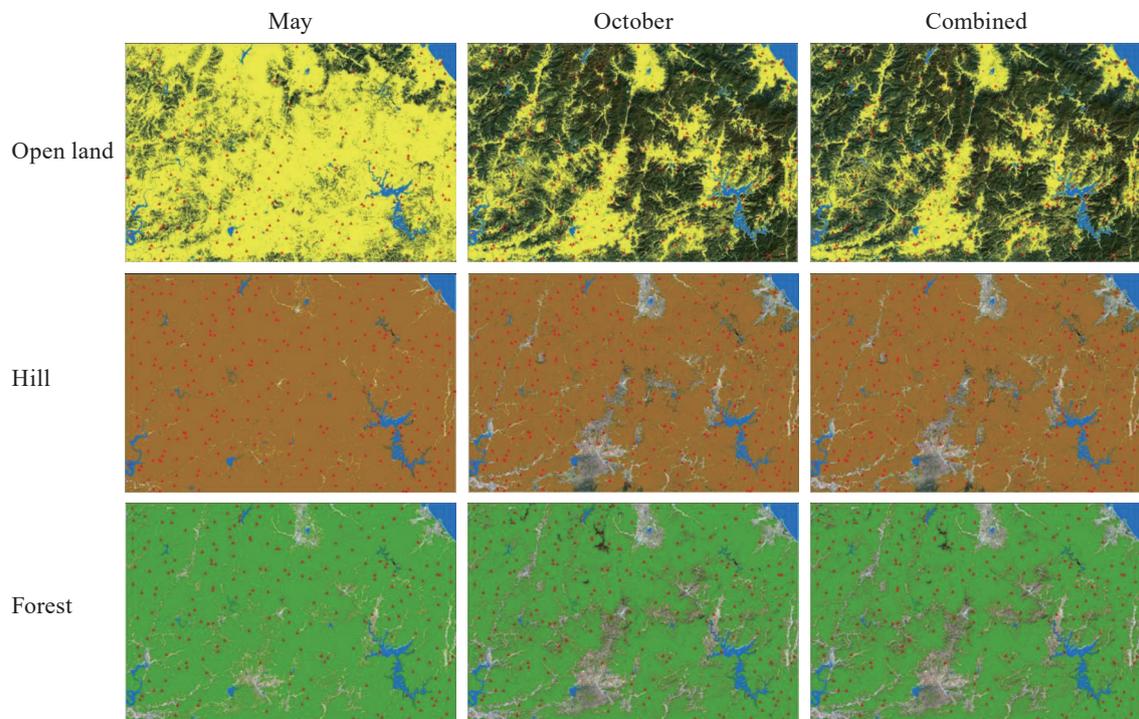


Fig. 6. (Color online) Open land, hill, and forest for the 20161010 scene that was derived from the optimum band and threshold in Table 6.

season lacked fully grown agricultural crops, the resulting land cover was mostly open land. Only the newly grown leaves of the forest absorbed blue and red lights but reflected NIR and green lights. Thus, there is a sharp increase in reflectance between the red and NIR regions, which is known as the red edge and is used in plant stress detection.⁽⁵⁰⁾ This process verifies the adoption of RGRI and NG for the 20160519 scene, while in the case of the 20161010 scene, the forest loses its chlorophyll; thus, deciduous plants reflect more NIR than in an earlier stage. This verifies the use of IPVI in segmenting forest. The slope above 6.9263 degrees is selected as optimum to segment hill and flatlands. As the autumn leaves are colorful, GRVI, i.e., red/green ratio, is optimum for open land segregation. Moreover, the GNDVI above 0.6897 was able to produce a similar result, which used NIR and green bands for the index.

To further understand the effect of these optimum indexes, we combined both ground truths for the selection of the optimum ratio/index for segregation. Using a minimum of 70 objects as pruning conditions, we found that NNIR alone is sufficient for segregating open land from forest cover, while the slope is optimum for hilly areas. An NNIR value below and above the threshold segregates open land and forest; in other words, they are complementary in terms of cover. The study areas can be considered complementary in terms of forest or open land as the area is rural with little build-up and some surface water.

For accuracy assessment, on the basis of the confusion matrix, UA, PA, OA, and kappa were derived for each threshold in Table 6 for both scenes. Table 7 shows the results of accuracy assessment based on ground truth points for open land, hill, and forest cover for 20160519 and 20161010 scenes using the stratified 10-fold cross-validation. For better understanding, we also plotted in Fig. 7 the densities of each target and nontarget based on the ground truth for the optimum band in Table 6.

In Table 7, most of the measures are around 90% except for hill. As the terrain is not known well and SRTM also has only a 30 m resolution, a misjudgment might occur while labeling the ground truth. In Fig. 7, it is clearly visible that hill and nonhill ground truths have relatively large overlapping values in terms of NG and slope. UA for the nonforest cover and PA for

Table 7

Accuracy assessment based on ground truth points for open land, hill and forest cover for 20160519 and 20161010 scenes using stratified 10-fold cross-validation.

Ground Truth	Measures	Open land	Nonopen land	Hill	Nonhill	Forest	Nonforest
May	UA (%)	87.79	86.86	75.68	91.67	80.27	95.07
	PA (%)	91.22	82.07	70.89	93.36	92.19	86.94
	OA (%)		87.43		88.29		88.86
	Kappa		0.74		0.66		0.77
October	UA (%)	87.79	84.67	66.22	93.84	85.71	94.09
	PA (%)	89.90	81.69	74.24	91.20	91.30	90.09
	OA (%)		86.57		88.00		90.57
	Kappa		0.72		0.63		0.80
Combined	UA (%)	89.67	87.23	76.92	93.12	86.39	95.81
	PA (%)	91.61	84.45	74.32	93.97	93.73	90.68
	OA (%)		88.71		89.78		91.86
	Kappa		0.93		0.80		0.82

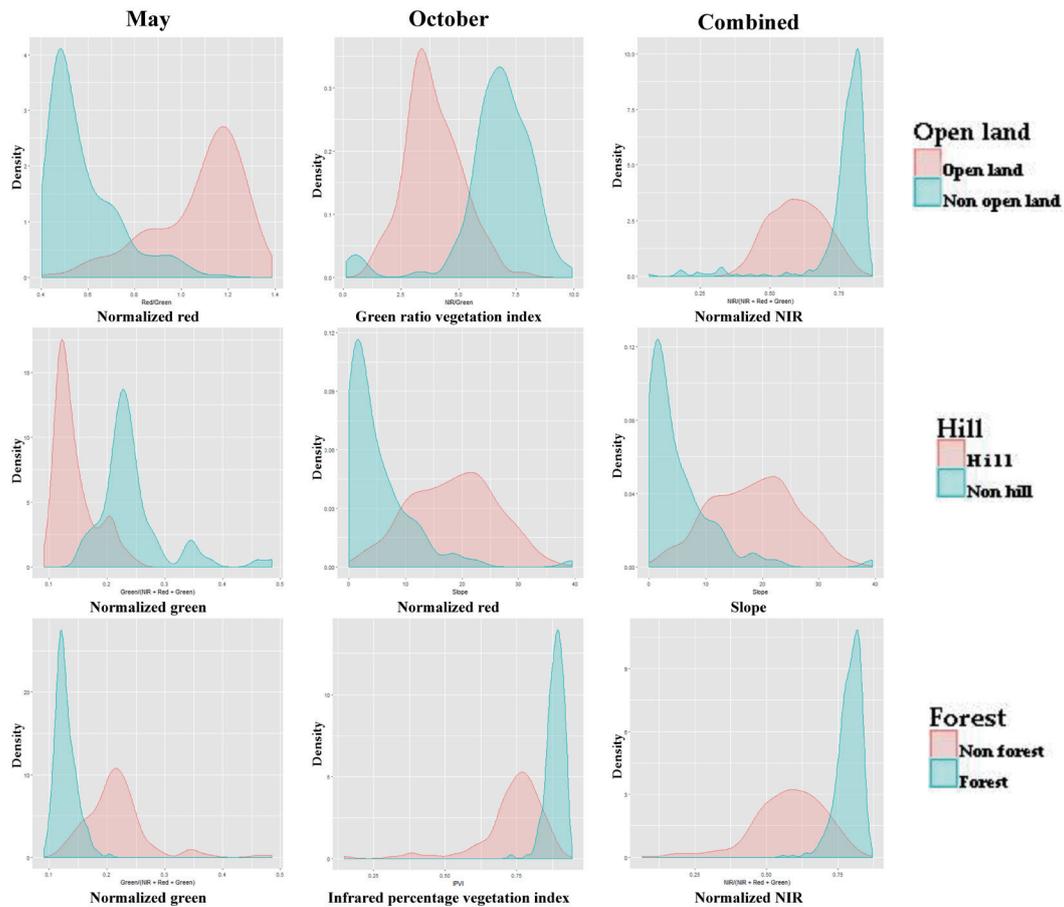


Fig. 7. (Color online) Density plots based on ground truth for optimum band ratio in Table 6.

nonhill and forest covers are highest in the study area. The combined ground truth shows better UA and PA for both scenes. From the combined ground truth, forest cover produced the highest OA of 91.86% whereas open land was more consistent with 0.93 kappa compared with the others. Using the combined ground truth improved OA and kappa in all the classes. Hence, using a multiple-season ground truth could be a better option than selecting only one for land cover mapping. However, the improvement was clear, that is, without actual knowledge of the study area and field works, it is difficult to avoid the error, as shown in Fig. 7. Most of the target and nontarget classes were overlapping. This suggests that a careful selection of the ground truth should be carried out to ensure results with higher accuracy even for inaccessible areas.

Finally, erosion-prone area maps were derived from binary maps, i.e., Figs. 5 and 6 of open land, hill, and forest derived from Table 6. The erosion-prone areas were indexed as low, medium, and high hazards, as per the rule in Table 4. The resulting maps of the erosion-prone areas are shown in Fig. 8 with the corresponding areal coverage being given in Table 8. Except for the 20161010 scene with the May ground truth, all others show consistent erosion-

prone areas. Because we were unable to access the actual site, it is impossible to judge the accuracy of the results obtained; however, a decision can be made on the basis of the average areal distribution. As shown in Table 4, we used the average of all the results to approximate the erosion-prone area in the study area. Out of the total land area of 7338.61 km² (excluding surface water) in the study area, around 20 and 15% are moderately and highly erosion-prone areas, respectively. The remaining 65%, which does not include erosion-prone areas, is mostly forested.

5. Conclusions

In this study, we applied remote sensing technology to investigate erosion-prone areas over an inaccessible area. A subset area from Kangwon Province, North Korea, was used as a test study site and Landsat Imagery from May and October as data for mapping. Satellite-based DEM and imagery were selected, indices were derived along with slope, ground truth points were carefully chosen for optimum threshold and band using pruned DT, and accuracy assessment was carried out. The results of this study can be summarized as follows:

- (a) Using remotely sensed data from Landsat and high-resolution imagery in Google Earth, it is possible to map different land covers and thus derive erosion-prone maps with reasonable accuracy.
- (b) Pruned J48 DT helped in the segmentation of the imagery based on single-band data of ratio/index without degrading accuracy. The optimum thresholds were much easier to understand than a complex tree with a minimum of 2 objects for binary classification.
- (c) Although the optimum band and threshold derived from the October ground truth were able to give acceptable results, they seemed ineffective for the interchange rule as demonstrated for the 20161010 scene using May rules. However, combining the ground truths from different seasons resulted in rules giving better OA and kappa than the individual rule results.
- (d) From the UA, PA, and density plots, we can understand that, despite a careful selection of ground truths from a high-resolution image, good knowledge of the study area and field investigations are necessary to select ground truths.
- (e) The average of erosion-prone maps, except for the 20161010 scene with the May ground truth, shows that 15 and 20% are high and medium erosion-prone areas out of the total land, respectively. The remaining 65% is due to forest cover.

On the basis of these results, we found that remotely sensed satellite imagery could be used to obtain land cover and erosion-prone area maps with reasonable accuracy. Although limited to one sensor image subset in a hilly area, satellite images could be very useful for monitoring changes in forest and open land covers over inaccessible areas.

Such studies can be very useful for calculating the total cost for planning the recovery of North Korean mountains, especially the bare land identified from satellite images. Future works will compare the green coverage and land cover pattern between Kangwon Province of North Korea and Gangwon Province of South Korea.

Acknowledgments

We would like to thank USGS for the free data and Google for the Google Earth Pro license. We would also like to thank Dr. Netra Regmi of the Oklahoma Geological Survey, University of Oklahoma as well as all the anonymous reviewers for their constructive comments, and MYU K.K., Japan for the English correction service. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2018R1A2B6009363).

References

- 1 UNEP: DPR KOREA: State of the Environment (2003) 12. <http://hdl.handle.net/20.500.11822/9690>
- 2 FAO: FAO A-I2000E (2011) 112. <http://www.fao.org/3/a-i2000e.pdf>
- 3 FAO: FAO A-X6953E (1995) 36. <http://www.fao.org/3/a-x6953e.pdf>
- 4 T. D. Acharya, I. T. Yang, and D. H. Lee: *Sens. Mater.* **30** (2018) 1753. <https://doi.org/10.7848/ksgpc.2018.36.1.17>
- 5 J. K. Lee, T. D. Acharya, and D. H. Lee: *Sens. Mater.* **30** (2018). <https://doi.org/10.18494/sam.2018.1934>
- 6 T. D. Acharya, J. Parajuli, K. Shahi, D. Poudel, and I. T. Yang: *Int. J. IT, Engineering and Applied Sciences Research (IJIEASR)* **4** (2014) 1.
- 7 T. D. Acharya, D. H. Lee, I. T. Yang, and J. K. Lee: *Sensors* **16** (2016) 1075. <https://doi.org/10.3390/s16071075>
- 8 A. M. Melesse, Q. Weng, P. S. Thenkabail, and G. B. Senay: *Sensors* **7** (2007) 3209. <https://doi.org/10.3390/s7123209>
- 9 G. L. Feyisa, H. Meilby, R. Fensholt, and S. R. Proud: *Remote Sens. Environ.* **140** (2014) 23. <https://doi.org/10.1016/j.rse.2013.08.029>
- 10 S. J. Goetz, G. J. Fiske, and A. G. Bunn: *Remote Sens. Environ.* **101** (2006) 352. <https://doi.org/10.1016/j.rse.2006.01.011>
- 11 Y. Zha, J. Gao, and S. Ni: *Int. J. Remote Sens.* **24** (2003) 583. <https://doi.org/10.1080/01431160304987>
- 12 J. M. Chen: *Can. J. Remote Sens.* **22** (1996) 229. <https://doi.org/10.1080/07038992.1996.10855178N>.
- 13 Mueller, A. Lewis, D. Roberts, S. Ring, R. Melrose, J. Sixsmith, L. Lymburner, A. McIntyre, P. Tan, S. Curnow, and A. Ip: *Remote Sens. Environ.* **174** (2016) 341.
- 14 C. Baker, R. Lawrence, C. Montagne, and D. Patten: *Wetlands* **26** (2006) 465. [https://doi.org/10.1672/0277-5212\(2006\)26\[465:mwarau\]2.0.co;2](https://doi.org/10.1672/0277-5212(2006)26[465:mwarau]2.0.co;2)
- 15 Y. Jin, S. Sung, D. K. Lee, G. S. Biging, and S. Jeong: *Remote Sens.* **8** (2016) 997. <https://doi.org/10.3390/rs8120997>
- 16 S. Kang and W. Choi: *Reg. Environ. Change* **14** (2014) 347. <https://doi.org/10.1007/s10113-013-0497-4>
- 17 S. Cha and C. Park: *Korean J. Remote Sens.* **23** (2007) 483.
- 18 T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, and D. Alsdorf: *Rev. Geophys.* **45** (2007) 1. <https://doi.org/10.1029/2005rg000183>
- 19 S. Krishnan, C. Crosby, V. Nandigam, M. Phan, C. Cowart, C. Baru, and R. Arrowsmith: *Proc. 2nd Int. Conf. Computing for Geospatial Research & Applications (COM.Geo, 2011)* 7. <https://doi.org/10.1145/1999320.1999327>
- 20 T. D. Acharya, I. T. Yang, and D. H. Lee: *J. Korean Soc. Surv. Geodesy Photogramm. Cartogr.* **36** (2018) 17. <https://doi.org/10.7848/ksgpc.2018.36.1.17>
- 21 I. T. Yang, D. H. Lee, T. D. Acharya, and D. S. Han: *J. Korean Asso. Geog. Inf. Stud.* **21** (2018) 34. <https://doi.org/10.11108/kagis.2018.21.2.034>
- 22 U. S. Geological Survey: <http://landsat.usgs.gov/landsat8.php> (accessed August 2019)
- 23 J. B. Campbell: *J. Geogr.* **106** (2008) 239. <https://doi.org/10.1080/00221340701849799>
- 24 T. D. Acharya, I. T. Yang, and D. H. Lee: *J. Korean Soc. Surv. Geodesy Photogramm. Cartogr.* **35** (2017) 357.
- 25 Y. Zhou, X. Xiao, Y. Qin, J. Dong, G. Zhang, W. Kou, C. Jin, J. Wang, and X. Li: *Int. J. Appl. Earth Obs. Geoinf.* **46** (2016) 1. <https://doi.org/10.1016/j.jag.2015.11.001>
- 26 D. Costanzo, C. Cappadonia, C. Conoscenti, and E. Rotigliano: *Nat. Hazards* **61** (2012) 103. <https://doi.org/10.1007/s11069-011-9870-0>

- 27 J. Knorn, A. Rabe, V. C. Radeloff, T. Kuemmerle, J. Kozak, and P. Hostert: *Remote Sens. Environ.* **113** (2009) 957. <https://doi.org/10.1016/j.rse.2009.01.010>
- 28 H. P. Sato and E. L. Harp: *Landslides* **6** (2009) 153. <https://doi.org/10.1007/s10346-009-0147-6>
- 29 C. J. Tucker: *Remote Sens. Environ.* **8** (1979) 127. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- 30 R. P. Sripada, R. W. Heiniger, J. G. White, and A.D. Meijer: *Agron. J.* **98** (2006) 968. <https://doi.org/10.2134/agronj2005.0200>
- 31 G. S. Birth and G. R. McVey: *Agron. J.* **60** (1968) 640. <https://doi.org/10.2134/agronj1968.00021962006000060016x>
- 32 J. A. Gamon and J. S. Surfus: *New Phytol.* **143** (1999) 105. <https://doi.org/10.1046/j.1469-8137.1999.00424.x>
- 33 R. E. Crippen: *Remote Sens. Environ.* **34** (1990) 71. [https://doi.org/10.1016/0034-4257\(90\)90085-z](https://doi.org/10.1016/0034-4257(90)90085-z)
- 34 J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering: *SP-351* (1973) 309.
- 35 C. Buschmann and E. Nagel: *Int. J. Remote Sens.* **14** (1993) 711. <https://doi.org/10.1080/01431169308904370>
- 36 A. A. Gitelson and M. N. Merzlyak: *Adv. Space Res.* **22** (1998) 689. [https://doi.org/10.1016/s0273-1177\(97\)01133-2](https://doi.org/10.1016/s0273-1177(97)01133-2)
- 37 A. R. Huete: *Remote Sens. Environ.* **25** (1988) 295. [https://doi.org/10.1016/0034-4257\(88\)90106-x](https://doi.org/10.1016/0034-4257(88)90106-x)
- 38 A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira: *Remote Sens. Environ.* **83** (2002) 195. [https://doi.org/10.1016/s0034-4257\(02\)00096-2](https://doi.org/10.1016/s0034-4257(02)00096-2)
- 39 B. Gao: *Remote Sens. Environ.* **58** (1996) 257. [https://doi.org/10.1016/s0034-4257\(96\)00067-3](https://doi.org/10.1016/s0034-4257(96)00067-3)
- 40 T. J. Jackson, D. Chen, M. Cosh, F. Li, M. Anderson, C. Walthall, P. Doriaswamy, and E. R. Hunt: *Remote Sens. Environ.* **92** (2004) 475. <https://doi.org/10.1016/j.rse.2003.10.021>
- 41 Harris Geospatial Solutions, Inc.: 2018 <http://www.harrisgeospatial.com/docs/alphabeticallistspectralindices.html> (accessed August 2019).
- 42 Y. Zhao and Y. Zhang: *Adv. Space Res.* **41** (2008) 1955. <https://doi.org/10.1016/j.asr.2007.07.020>
- 43 L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen: *Classification and Regression Trees* (CRC Press, Boca Raton, 1984) Chap. 2. <https://doi.org/10.2307/2530946>
- 44 D. Kobayashi, O. Takahashi, H. Arioka, S. Koga, and T. Fukui: *Am. J. Geriatric Psychiatry* **21** (2013) 957. <https://doi.org/10.1016/j.jagp.2012.08.009>
- 45 M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten: *SIGKDD Explor. Newsl.* **11** (2009) 10. <https://doi.org/10.1145/1656274>
- 46 J. R. Quinlan: *C4. 5: Programs for Machine Learning* (Morgan Kaufmann, San Mateo, 2014) p. 35.
- 47 I. H. Witten and E. Frank: *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann, 2005)
- 48 R. G. Congalton and K. Green: *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices* (CRC Press, Boca Raton, 2008) p. 105. <https://doi.org/10.1201/9781420055139>
- 49 M. L. McHugh: *Biochemia Medica* **22** (2012) 276. <https://doi.org/10.11613/bm.2012.031>
- 50 A. R. Huete: *Remote Sensing for Environmental Monitoring* (Elsevier, Burlington, 2004) p. 183.

About the Authors



Tri Dev Acharya received his B.E. degree in Geomatics from Kathmandu University, Nepal, in 2011, and his combined M.S. and Ph.D. degree from Kangwon National University, Korea, in 2018. He is now a postdoctoral researcher at Kangwon National University, Korea. His research interests are in geospatial data preparation, modelling, and simulation of land cover, surface water, and natural hazards using various machine learning algorithms. (tridevacharya@kangwon.ac.kr)



Kyoung Nam Kim received his B.E., M.S., and Ph.D. degrees from the Kangwon National University, Korea, in 1993, 1995, and 2001, respectively. Since 2003, he has been a senior research fellow at the Research Institute for Gangwon. His research interests are in disaster management using aerial photos, surveying, and geospatial information. (robert00@rig.re.kr)



Kwang Youn Lee received his B.E., M.S., and Ph.D. degrees from Kangwon National University, Korea, in 2014, 2017, and 2019, respectively. Since 2014, he has been a researcher at the Research Institute for Gangwon. His research interests are in geospatial and topographic information. (kwangyounlee@rig.re.kr)



Anoj Subedi received his B.S. degree in forestry from the Institute of Forestry, Pokhara, which is affiliated with Tribhuvan University, Nepal. He has been involved in various national and international projects and has well contributed to various wildlife and forest surveys. His research interests are related to forest, wildlife, water bodies, land use/cover, and climate change. (anojsubedi99@gmail.com)



He Huang received his B.S. degree from Wuhan University in 2000 and his M.S. and Ph.D. degrees in Geodetic Surveying from Sungkyunkwan University. Currently, he is an associate professor in the School of Geomatics and Urban Spatial Information at Beijing University of Civil Engineering and Architecture. His research interests include GNSS technology and applications, indoor navigation, and positioning. (huanghe@bucea.edu.cn)



Dong Ha Lee received his B.E., M.S., and Ph.D. degrees from Sungkyunkwan University, Korea, in 2000, 2003, and 2008, respectively. Since 2015, he has been an assistant professor at Kangwon National University, Korea. His research interests are in geodesy, surveying, geospatial information, and natural hazard analysis. (geodesy@kangwon.ac.kr)