Master's Thesis

INVESTIGATION OF DEFORESTATION

USING MULTI-SENSOR SATELLITE

TIME SERIES DATA IN NORTH KOREA

Yeeun Jang

Department of Urban and Environmental Engineering
(Environmental Science and Engineering)

Graduate School of UNIST

2017

INVESTIGATION OF DEFORESTATION

USING MULTI-SENSOR SATELLITE

TIME SERIES DATA IN NORTH KOREA

Yeeun Jang

Department of Urban and Environmental Engineering
(Environmental Science and Engineering)

Graduate School of UNIST

INVESTIGATION OF DEFORESTATION

USING MULTI-SENSOR SATELLITE

TIME SERIES DATA IN NORTH KOREA

A thesis

submitted to the Graduate School of UNIST

in partial fulfillment of the

requirements for the degree of

Master of Science

Yeeun Jang

12.27.2016

Approved by
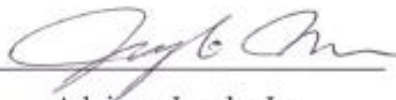
—————————————————

Advisor

Jungho Im

INVESTIGATION OF DEFORESTATION

USING MULTI-SENSOR SATELLITE

TIME SERIES DATA IN NORTH KOREA

Yeeun Jang

This certifies that the thesis of Yeeun Jang is approved.

12.27.2016

Advisor: Jungho Im

Kyoung-Min Kim

Kyung Hwa Cho

Abstract

North Korea is very vulnerable to natural disasters such as floods and landslides due to institutional, technological, and other various reasons. Recently, the damage has been more severe and vulnerability is also increased because of continued deforestation. However, due to political constraints, such disasters and forest degradation have not been properly monitored. Therefore, using remote sensing based satellite imagery for forest related research of North Korea is regarded as currently the only and most effective method. Especially, machine learning has been widely used in various classification studies as a useful technique for classification and analysis using satellite images.

The aim of this study was to improve the accuracy of forest cover classification in the North Korea, which cannot be accessed by using random forest model. Indeed, another goal of this study was to analyze the change pattern of denuded forest land in various ways.

The study area is Musan-gun, which is known to have abundant forests in North Korea, with mountainous areas accounting for more than 90%. However, the area has experienced serious environmental problems due to the recent rapid deforestation. For example, experts say that the damage caused by floods in September 2016 has become more serious because denuded forest land has increased sharply in there and such pattern appeared even in the high altitude areas. And this led the mountain could not function properly in the flood event.

This study was carried out by selecting two study periods, the base year and the test year. To understand the pattern of change in the denuded forest land, the time difference between the two periods was set at about 10 years.

For the base year, Landsat 5 imageries were applied, and Landsat 8 and RapidEye imageries were applied in the test year. Then the random forest machine learning was carried out using randomly extracted sample points from the study area and various input variables derived from the used satellite imageries. Finally, the land cover classification map for each period was generated through this random forest model. In addition, the distribution of forest changing area to cropland, grassland, and bare-soil were estimated to the denuded forest land. According to the study results, this method showed high accuracy in forest classification, also the method has been effective in analyzing the change detection of denuded forest land in North Korea for about 10 years.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

North Korea has a high geographical distribution of slopes and high mountainous areas of 1,500 m or more, and had a significantly higher forestage ratio than that of South Korea at the time of liberation in 1945. However, as it is already known, serious forest degradation is under way. The causes are diverse, but largely due to indiscriminate tenant farming to overcome the food shortage, deforestation of trees due to fuel shortages, and formal and informal timber exports.
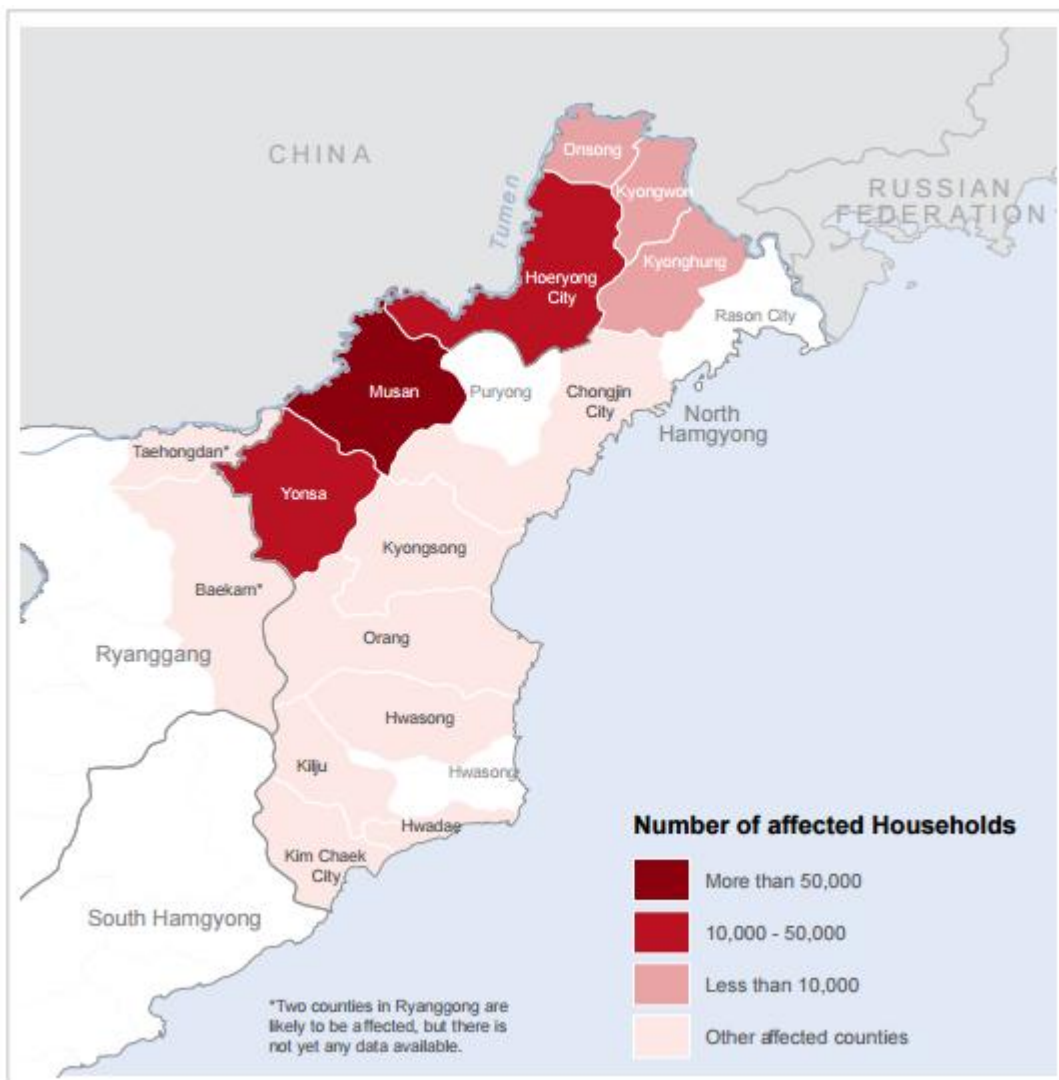
Figure 1 2016 flood damaged area in Hamgyong buk-do, North Korea (Source: 20160916 DPRK North Hamgyong floods Joint Assessment Report, UN OCHA)

As a result of deforestation, forest ecosystem has been destroyed and soil loss has become serious, making it more vulnerable to natural disasters such as floods and landslides. Like South Korea, North Korea is located in the monsoon climate zone where rainfall is concentrated in summer, thus the damage in case of natural disasters caused by deforested forests is increasing. For example, severe floods occurred in North Korea due to unusually heavy rains in September 2016. In particular, there were intense damage to the Musan-gun area, Yonsa-gun area, and Hoeryeoung City that were near to the Tumen River in North Hamgyong Province. According to the Joint Assessment North Hamgyong Floods 2016 report of United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA), the victims were more than 50,000 families in Musan-gun area, and that of Yonsa-gun area and Hoeryong City were between 10,000 and 50,000, respectively (Figure 1). And this figure indirectly indicates how severe the forest degradation in this area is.

1.2   Purpose of study

It is very necessary to establish forest information to effectively manage the forests of the unified Korean peninsula by classifying the denuded forest land in each region of North Korea. However, since North Korea is not accessible as an isolated region politically and economically, satellite images are the only research data. Actually, the use of satellite image alone has many limitations in that it is not possible to obtain field data for verification. For example, due to the nature of the forest which the boundaries of the species are vague and mixed forms appear in many areas, there is a high risk of misclassification when using only satellite images. Especially, even though unstocked forest land area which is abandoned as scrubland is important that should be selected or grasped first in support projects for future forest restoration, it is not easy to distinguish those areas by only using spectroscopic characteristics of satellite images because it has spectroscopically similar properties in the alpine region, which is a good grassland commonly found in high mountains. Nonetheless, using satellite image is the most effective and available method for accurately classifying and analyzing denuded forest land in North Korea. Therefore, remote sensing has been applied most commonly in the studies regarding forest of North Korea.

This study aims to develop an improved technique to classify land cover of North Korea, which is forest dominant area by using high and middle resolution satellite imageries. Also, this study examines the status of denuded forest land in North Korea and extracts classification maps of those areas based on machine learning approach.

1.3   Definition of denuded forest land

According to the Korea Forest Service, denuded forest land refers to a mountainous area where normal forest production activities are not being carried out properly, or a forest area that is used for other purposes, such as cultivated land, thereby making it unable to perform its own intrinsic functions and rendering it impossible to be recovered to stocked land. It thus indicates an area that requires restoration and afforestation. The representative types of denuded forest land can be categorized as (a) deforested cropland in the form of sloping fields where crops are cultivated by forest clearing, (b) unstocked forest land that remains scrubland due to the extraction of firewood, the grazing of livestock, and the outbreak of forest fires, or a barren-mountain area where a cultivated forest has been left abandoned, and (c) bare land which remains in the state of a bald mountain or bare ground with few or no trees and cover vegetation, in accordance with the standard of United Nations Food and Agriculture Organization (FAO) (Figure 2).

| (a) Deforested cropland | (b) Unstocked forest land | (c) bare land |

Figure 2 Main types of denuded forest land in North Korea (Lee et al., 2014)

2. Literature Review

Various studies have been researched for classification of the deforested area at home and abroad. North Korea is one of the most serious forest degradation countries in recent decades. However, despite the severe forest degradation, direct field survey is limited due to inaccessible political reasons. Thus, most studies related to the status of denuded forest land in North Korea have been conducted based on remote sensing technique.

For example, a study using high resolution satellite images was examined by Korean National Institute of Forest Science recently. In that study, the ISO classification method was used to extract a deforestation map as a final output. ISO is a widely used traditional remote sensing classification

method, but actually that method has a high risk of misclassification. Also, Hong et al. (2008) analyzed land cover classification in North Korea using Landsat Thematic Mapper (TM) imagery, and Part and Yu (2009) performed land cover classification using multi sensor satellite images and analyzed deforestation status and trend in North Korea. Yoo et al. (2011) conducted image classification and deforestation analysis in Hamheung, North Korea, using two satellite images of Landsat TM taken in 1988 and SPOT taken in 2007, and finally, selected an optimized reforestation site based on the classification result. Kim et al. (2010) performed image classification using SPOT satellite images and developed a technique to maximize the separation between types of deforestation by using texture bands for optimal forest cover classification. Lee et al. (2015) analyzed deforestation status of North Korea through NDVI and modified NDVI using r, g, b bands. Also, Yu and Kim (2015) utilized various spatial variables based on satellite remote sensing to examine the reason of deforestation and the spatial and temporal forest changing pattern in North Korea, and confirmed potential reforestation area. Lee (2004) reviewed forest change status and forest degraded area in North Korea by summarizing previously published materials.

Meanwhile, various studies on the classification of denuded forest land have been carried out abroad. Eliakim et al. (2016) conducted a study of techniques to reduce seasonal variations of Landsat data and improve the accuracy of forest degradation detection using spatial context, and Johannes et al. (2015) proposed a pixel-based Multi-sensor Time-series Correlation and Fusion Approach (MulTiFuse) technique using NDVI data of Landsat and ALOS PALSAR L-band SARs. Especially, Johannes confirmed that the accuracy of forest degradation detection was improved and deforestation map was prepared as a final product by using the fused NDVI-PALSAR image. Cesar et al. (2015) conducted a study to analyze the change of land cover in detail by using AViFS sensor, which is divided into color, tone, texture and context. Eckert et al. (2015) conducted a study to detect vegetation change area and land degradation area in Mongolia using MODIS time series NDVI data, and the trend of NDVI was found to affect vegetation growth of forest land cover. Also, Arai et al. (2011) conducted a study to analyze deforestation in the Amazon River basin by using multi-sensor time series satellite data. Terra MODIS and Landsat 7 imageries were used to apply the linear spectral mixing model. The results showed the accuracy was improved when using both MODIS and Landsat 7 than using single sensor data and it also shows every Landsat 7 ETM + bands were highly related to the area that estimated as forest degraded.

Recently, studies regarding Reducing Emissions from Deforestation and forest Degradataion (REDD) have been increasing as a new trend of forest-related research. Thus, many relevant studies have been reported around the world to estimate the forest degradation area and estimate the potential of REDD project. REDD is an international environmental policy which aims to reduction of carbon emissions

in developing and undeveloped countries. It is a project to prevent forest loss and to secure carbon credits at the same time by applying opportunity cost to the country that preserves its good quality forest area without losing or destroying them. Therefore, in the near future of a unified Korean peninsula, REDD will be an essential issue and now it is significantly important to monitor forest and deforestation status of North Korea. Choi et al. (2012) conducted a study to analyze a status of forest degradation in the area near Mt. Geumgang in North Korea and to estimate carbon emissions in there by using Landsat TM. Kim et al. (2014) quantified forest loss of North Korea by estimating carbon emissions based on satellite remote sensing data.

By the way, machine learning approach has been used steadily in land cover classification area based on remote sensing (Gislason et al., 2006; Rodriguez-Galiano et al., 2012; Pal Mahesh, 2005). Several researches revealed that random forest showed good performance in forest cover classification (Markus et al., 2012; Baccini et al., 2012). Precedents studies indicate several relevant variables as like NDVI were effective to distinguish land classification. Also, those studies show time series multi-sensor data were widely used in forest cover classification and its analysis.

Although many studies have been done on forest degradation through machine learning technique so far, but there are not many related studies on North Korea. Recently, Jin et al. (2016) examined deforestation status of North Korea by using random forest technique. In this study, phenology index was applied as input variables of random forest. However, this study was a comprehensive analysis of whole North Korea by using Moderate Resolution Imaging Spectroradiometer (MODIS) low resolution satellite imageries, and it is necessary for analysis more specific areas for improving accuracy of classification. Therefore, this study selected Musan-gun as a test area where has been seriously degraded area in North Korea, and analyzed denuded forest land in there. Also, this study tried to increase accuracy of land cover classification, including denuded forest land mapping through random forest machine learning approach by using middle resolution Landsat series and high resolution RapidEye satellite imageries.

3. Study Area and Study Data

3.1 Study Area

Despite the widespread and the extensive occurrence of deforestation throughout North Korean territory, Musan-gun was selected as the research object region in this study, because it is one of the areas where deforestation is most severe in North Korea. Musan-gun is geographically located on the north side of the border with China through the Tumen River, and as part of the Baekmu Plateau in the north-west side of Hamgyeong Mountain range along with the Baekdu Plateau, it is an inland area surrounded by steep mountains over 1,500m above sea level (Figure 3). More than 90% of this mountainous area remains uninhabited, and it is also one of the coldest areas in North Hamgyeong Province. The annual average temperature of Musan-gun is 5.2 ˚C, the average temperature in January is -13.4 ˚C, the average temperature in July is 21.3 ˚C, and the frost period is more than 200 days per year. Though Musan-gun has been one of the most forested areas in North Korea, it is also one of the most severely damaged areas due to reckless exploitation of forests. As mentioned earlier, due to continued food shortages, even forests in high altitude have been deforested and converted as crop land and this caused such vulnerable conditions to natural disasters.
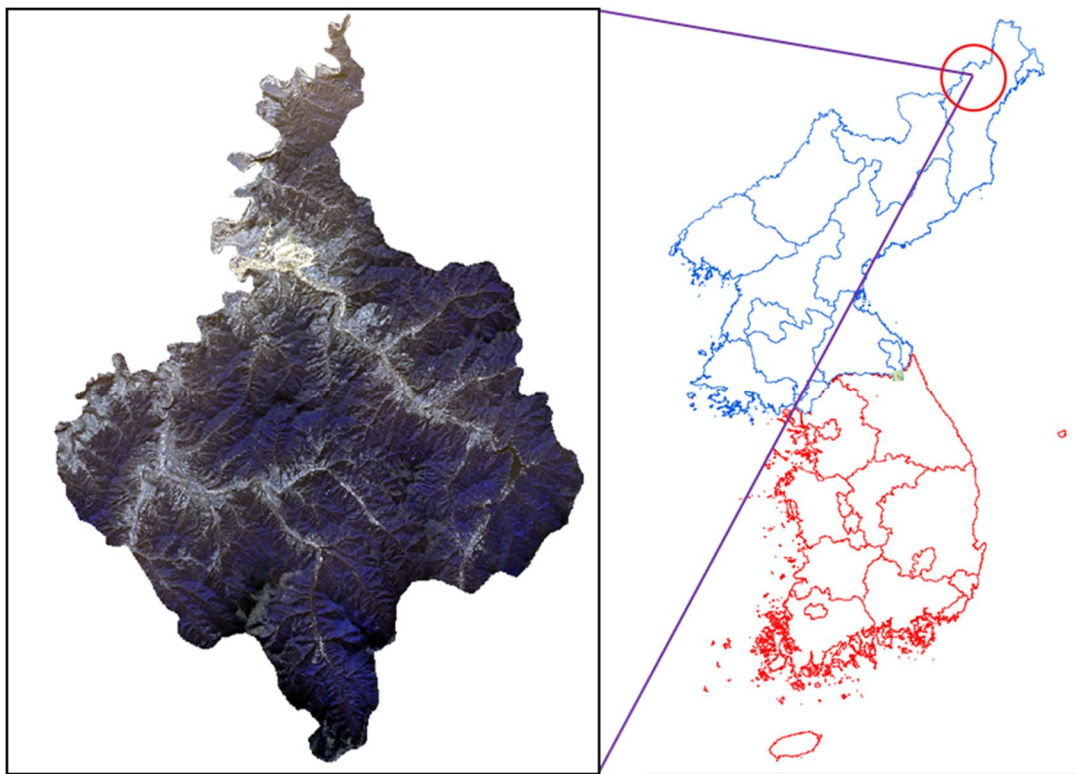
Figure 3 Location of study area, Musan-gun in North Korea

3.2 Study Data

3.2.1   Satellite images

This study was divided into two periods. After reviewing all available satellite images taken on a clear,

cloudless day between the two periods, several images were selected per each period. First period was base year, two clear days Landsat 5 TM satellite images acquired in the study area on October 5, 2005 and September 22, 2006. The second period was test year, and Landsat 8 OLI and RapidEye satellite images were used in this period. Four Landsat 8 OLI images acquired on September 28, 2014, October 14, 2014, October 17, 2015, and May 28, 2016 were selected. Also four RapidEye images acquired on October 29, 2014, September 21, 2015, May 5, 2016, and May 28, 2016 were applied. The reason of separating base year and test year is to compare how forest cover has changed and degraded between two periods. For this, the base year was selected considering 10 years before than a test year to confirm changes easily. Table 1 shows the list of satellite images used in this study.

Table 1 List of satellite images used in this study

Period	Satellite	DOY	Date
Base year	Landsat 5 TM	278	October 5, 2005
		265	September 22, 2006
Test year	Landsat 8 OLI	271	September 28, 2014
		287	October 14, 2014
		290	October17, 2015
		149	May 28, 2016
Test year	RapidEye	302	October 29, 2014
		264	September 21, 2015
		126	May 5, 2016
		149	May 28, 2016

As like Table 1 shows, Landsat 5 TM imageries were used for base year data. Landsat 5 was launched by NASA on March 1, 1984 at the Vandenberg Air Force Base site in California. It has two sensors, Thematic Mapper (TM) and Multispectral Scanner (MSS). After launched, Landsat 5 had been operated during 29 years and TM sensor operational imaging ended in November 2011. It has 8 spectral bands such as Red, Green, Blue, NIR, and captures data through these bands at 30 m resolution.

For test year data, Landsat 8 OLI imageries were used at first. Landsat 8 is the representative Earth observation satellite which has launched on February 11, 2013. It had developed as the Landsat Data Continuity Mission by NASA and has 30 m spatial resolution. The Landsat 8 has 2 more extra bands including costal – aerosol and cirrus bands compared than previous Landsat 7 ETM+ (Table 2). This

satellite has been used various fields such as change detection, classification and so on. However, it has to be considered cloud and shadow of clouds at first cause those barriers can effects significantly on the quality of satellite image when using the Landsat data series. In general, as like South Korea, North Korea also has a moist and humid climate in summer and that's why no Landsat 8 imageries had obtained during the summer season unfortunately.

Table 2 Landsat 5 TM, Landsat 7 ETM+ and Landsat 8 OLI spectral bands comparison

Landsat 5 TM and Landsat 7 ETM+ Spectral Bands (µm)			Landsat 8 OLI and TIRS Spectral Bands (µm)		
Band #	Band name	Wavelength	Band #	Band name	Wavelength
Band 1	Blue (30 m)	0.441 – 0.514	Band 1	Coastal/ Aerosol(30 m)	0.435 – 0.451
Band 2	Green (30 m)	0.519 – 0.601	Band 2	Blue (30 m)	0.452 – 0.512
Band 3	Red (30 m)	0.631 – 0.692	Band 3	Green (30 m)	0.533 – 0.590
Band 4	NIR (30 m)	0.772 – 0.898	Band 4	Red (30 m)	0.636 – 0.673
Band 5	SWIR-1 (30 m)	0.547 – 1.749	Band 5	NIR (30 m)	0.851 – 0.879
Band 6	TIR (60 m)	10.31 – 12.36	Band 6	SWIR-1 (30 m)	1.566 1.651
Band 7	SWIR-2 (30 m)	2.064 – 2.345	Band 10	TIR-1 (100 m)	10.60 – 11.19
Band 8	Pan (15 m)	0.515 – 0.896	Band 11	TIR-2 (100 m)	11.50 – 12.51
			Band 7	SWIR-2 (30 m)	2.107 – 2.294
			Band 8	Pan (15 m)	0.503 – 0.676
			Band 9	Cirrus (30 m)	1.363 – 1.384

Another satellite data applied for test year is RapidEye. This satellite was launched on August 29, 2008 by the German Aerospace Center for the purpose of using agricultural area. The spatial resolution of RapidEye is 5 m and it consists of five identical satellites. Thus, it has an advantage to obtain a wide range of time series images for the same area in a short re-visit period. Also, RapidEye has five bands such as Red, Green, Blue, Near InfraRed, and Red Edge (Table 3). Since it is possible to acquire data from Red Edge wavelength, RapidEye is effective in various fields as like agricultural crop monitoring and forest monitoring.

Table 3 RapidEye technical specifications

Characteristic	Description
Number of Satellites	5
Spacecraft Lifetime	7 years
Orbit Altitude	630 km

Global Revisit Time	1 day
Inclination	97.8 degrees
Ground Sampling Distance (nadir)	6.5 m
Pixel Size	5 m
Swath Width	77 km
On Board Data Storage	Up to 1,500 km of image data per orbit
Image Capture Capacity	4 million km^2 / day
Spectral Bands	440 – 510 nm (Blue)
	520 – 590 nm (Green)
	630 – 685 nm (Red)
	690 – 730 nm (Red Edge)
	760 – 850 nm (Near IR)

3.2.2   Reference data

In this study, multi-time Google Earth imageries were used as references. This is because of the isolated situation of North Korea. Unfortunately, it is difficult to obtain measured data from relevant governmental organizations or conduct field surveys directly. Indeed, even if data are obtained, the reliability of the data is very low. Therefore, Google Earth can be considered as the only and the most reliable reference data in North Korea.

4. Methodology

4.1 Image preprocessing

The Landsat satellite images used in this study were downloaded from United States Geological Survey (USGS) EarthExplorer website. USGS has been providing free satellite imageries around the world on a regular basis through the Landsat series since 1972. The downloaded Landsat images were Level 1T type which applied systematic radiometric correction and geometric correction by incorporating ground control point (GCP) and Digital Elevation Model (DEM) data (USGS home page). Generally Landsat satellite image is provided with quantified radiance type. Thus, it is required to convert radiance data to Top of Atmosphere (ToA) reflectance to use the features of each band. In this study, ENVI 5.2 software was applied to convert Landsat 5 and 8 radiance values to ToA

reflectance. Also Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) atmospheric correction program was conducted by ENVI 5.2 software. FLAASH is a commonly used program for atmospheric correction of multi spectral and hyper-spectral satellite data. It extracts atmospheric water vapor images, cloud maps, and visible images by means of MODTRAN simulation of spectra which calculated based on various atmospheric, vapor, and sun altitude conditions (Jensen, 1986).
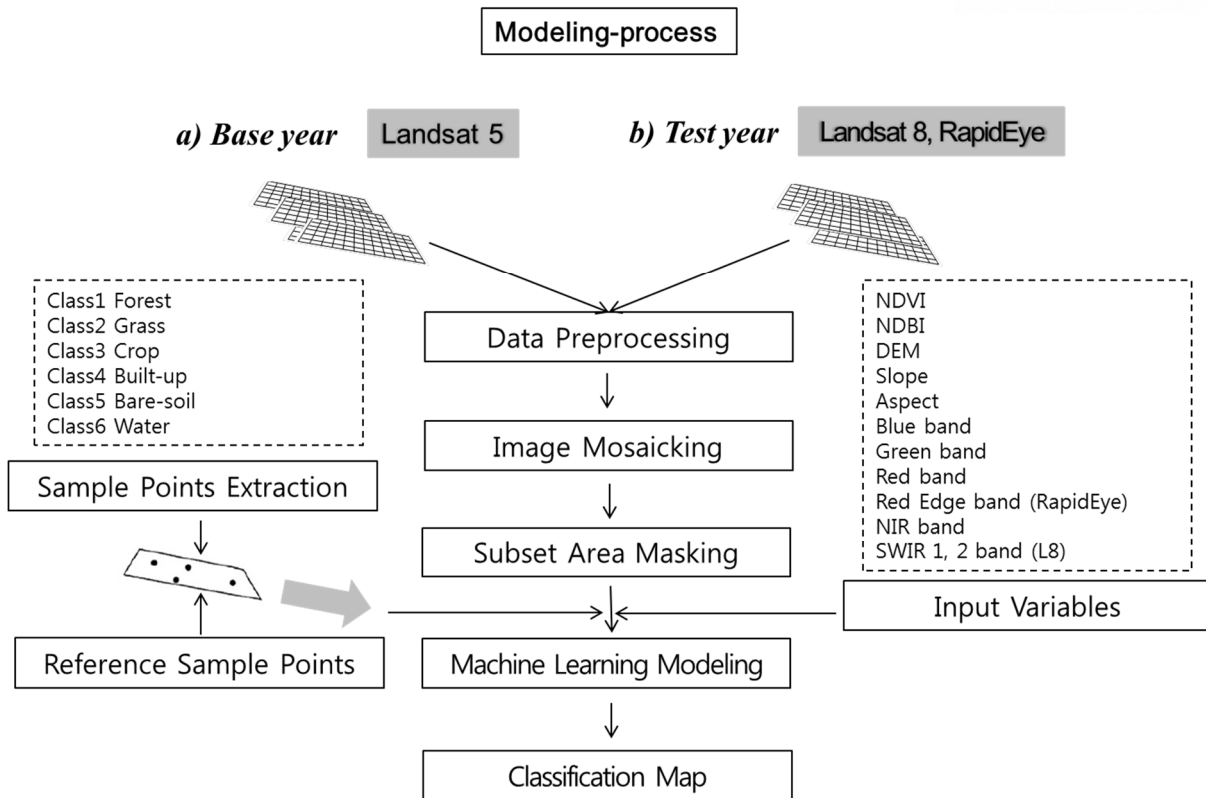
By the way, the RapidEye satellite imageries used in this study were orthorectified Level 3A products so the radiometric correction applied, and also rectified by DEM based on appropriate GCPs. RapidEye imageries were also conducted FLAASH atmospheric correction by ENVI 5.2 software like Landsat.

4.2 Modeling Process

The overall modeling process applied in this study can be categorized into several parts like Figure 4.

The first step is data preprocessing. For this, FLAASH atmospheric correction was applied on Landsat 5, Landsat 8 and RapideEye satellite imageries by using ENVI 5.2 software. In case of 5 m high resolution RapidEye image, the study area consists of several image tiles so the divided tiles had been conducted FLAASH atmospheric correction processing separately. As an output of FLAASH process the reflectance data were produced.

The second step is image mosaicking. In this process, the preprocessed reflectance images were combined into a single image per each study site through Merge Tool in ArcGIS Version 10.4.1 software. Also the projection of each image tiles was defined and confirmed as WGS_1984_UTM _Zone_52N before and after mosaicking process.

Modeling-process

a) Base year Landsat 5   *b) Test year* Landsat 8, RapidEye

| Class1 Forest |
| Class2 Grass |
| Class3 Crop |
| Class4 Built-up |
| Class5 Bare-soil |
| Class6 Water |

Data Preprocessing

NDVI
NDBI
DEM
Slope
Aspect
Blue band
Green band
Red band
Red Edge band (RapidEye)
NIR band
SWIR 1, 2 band (L8)

Sample Points Extraction

Image Mosaicking

Subset Area Masking

Reference Sample Points

Input Variables

Machine Learning Modeling

Classification Map

Figure 4 Overview of the methodological workflow

The third step is making subset areas by masking tool in ENVI 5.2 software. The shape files based on administration area which was projected as WGS_1984_UTM_Zone_52N were used to fit the subset areas for Musan-gun.

Fourth step is machine learning modeling based on supervised classification. In this study, Random forest machine learning method was applied. There are two more sub processes in this course. One is an extraction of sample points per each class and the other is an extraction of input variables for applying next machine learning.

In this study, 6 representative classes were selected, and those 6 classes consisted of forest, grass, crop, built-up, bare-soil, and water respectively. By using create random points function in the data management tool of ArcGIS, 1,200 random samples were generated on the Musan-gun study site. Then those points were classified through visual inspection. Additionally, multi times Google Earth imageries were used as reference data during classifying and confirming sample points. Due to the isolated situation in North Korea, verification can only avail through Google Earth images and this is regarded as a limitation of this study unfortunately.

11

The last step is producing classification maps as final outputs. The produced base year classification map was compared to the map of the test year and google earth imageries.

4.3 Input Variables

Input variables were selected considering factors that could affect the classification of deforestation area. Normalized Difference Vegetation Index (NDVI), which is representative index of vitality of vegetation, was used. Also, the Normalized Difference Built-up Index (NDBI), which is representative urbanization index, was used because the study area includes some built-up area. NDVI and NDBI were calculated for Landsat 5, Landsat 8 and RapidEye imageries respectively according to the equation described in the Table 4. By the way, topographic variables such as DEM, slope, and aspect were also applied in this study. Since the study site belongs to the mountainous area, the influence of topographic variables on the deforestation is assumed to be large. DEM data is provided by Shuttle Rader Topography Mission (SRTM) in USGS. Slope and aspect data were derived by means of Spatial Analyst tool in ArcGIS Version 10.4.1 with DEM data respectively. For aspect, the beers transformation was applied for define correct degrees (Beers et al., 1966). Furthermore, bands of each satellite image were applied as input variables. Each band has its own unique characteristics, and it includes the parts that affect the forest, so it is necessary to check each band as input variables. Thus, blue, green, red, near infrared (NIR), short-wave infrared (SWIR) 1, and swir2 band for Landsat 5 and Landsat 8, and blue, green, red, red edge, and NIR band for RapidEye were used as input variables. Specifically, in case of blue band, it has a high water permeability and distinctness of soil from plants or deciduous tree from coniferous trees. Also, green band shows healthy vegetation which means the most vigorous vegetation. Red band is one of the most important bands in distinguishing vegetation and it is available to discriminate vegetated slope. NIR Band is very sensitive to plant biomass, so it is good to figure out soil from crop, and land from water. By the way, SWIR is useful to discriminate soil and plant moisture, and this feature can be used to analysis of denuded forest land (Jensen, 1986).

Table 4 NDVI and NDBI calculation equation per satellite image

	NDVI	NDBI
Landsat 5	$\dfrac{NIR(b4) - RED(b3)}{NIR(b4) + RED(b3)}$	$\dfrac{SWIR1(b5) - NIR(b4)}{SWIR1(b5) + NIR(b4)}$
Landsat 8	$\dfrac{NIR(b5) - RED(b4)}{NIR(b5) + RED(b4)}$	$\dfrac{SWIR1(b6) - NIR(b5)}{SWIR1(b6) + NIR(b5)}$

| RapidEye | $\dfrac{NIR(b5) - RED(b3)}{NIR(b5) + RED(b3)}$ | - |

4.4 Machine Learning Approach

In this study, three testing was done and compared each other according to the period. The summary of these machine learning approaches is shown in Table 5. The first scheme is for the base year, random forest was conducted by means of Landsat 5 derived input variables, including spectral bands, DEM, slope, and aspect. The second scheme is for the test year. In this scheme Landsat 8 derived variables such as NDVI and NDBI per selected date were applied as well as 30 m resolution DEM, slope, and aspect. The last scheme is for the test year and it was conducted by using Landsat 8 and RapidEye derived input variables with topographic variables such as DEM, slope, and aspect. The machine learning results of these three schemes were compared respectively.

As mentioned above, random forest technology was applied in this study. Random Forest is a potent machine learning technique based on ensemble algorithms. It uses many decision trees simultaneously during classification process and normally provides increased accuracy results than previous methods such as maximum likelihood or single tree decision tree method (Jensen, 1986). R for Windows 3.0.3 software program was used for installing random forest package tool and it is available to download from internet as freely. Through this tool, accuracy of classification can be calculated as a confusion matrix form and also variable importance can be extracted easily.

Table 5 Machine Learning scheme of this study

No.	Satellite Image		Input Variables			
	Base year	Test year	Bands	NDVI	NDBI	DEM, Slope, Aspect
1	Landsat 5		○	○	○	○
2		Landsat 8	○	○	○	○
3		Landsat 8/ RapidEye	○	○	○ / -	○

5. Study Results

5.1 Phenology pattern Analysis of NDVI and NDBI input variables

Machine learning techniques are heavily influenced by data and its input variables. Thus, it is necessary to examine the characteristics of the input variable. Among others, NDVI can be easily confirmed its seasonal variation. Since it represents the vegetation's vitality it has a great correlation with forest and deforestation status, too. NDBI represents urbanization status and it also appears to be an important factor in identifying deforestation. That's why the phenology of these two variables was reviewed in this study. Figure 5 to 7 displays the results those reviews.

Figure 5 shows the seasonal changes of NDVI and NDBI for the base year, respectively. These two variables generally tend to be opposite in the seasons. Thus, NDVI was higher in September, when the vegetation was more active, and NDBI was lower, in contrast to October. NDVI was higher in order of forest, grass, crop, bare-soil, built-up, and water.

Figure 6 displays NDVI and NDBI variation graphs of test year in case of using only Landat8 imageries. As like base year, NDVI was higher in order of forest, grass, crop, bare-soil, built-up, and water. And NDBI tended to be opposite to NDVI.
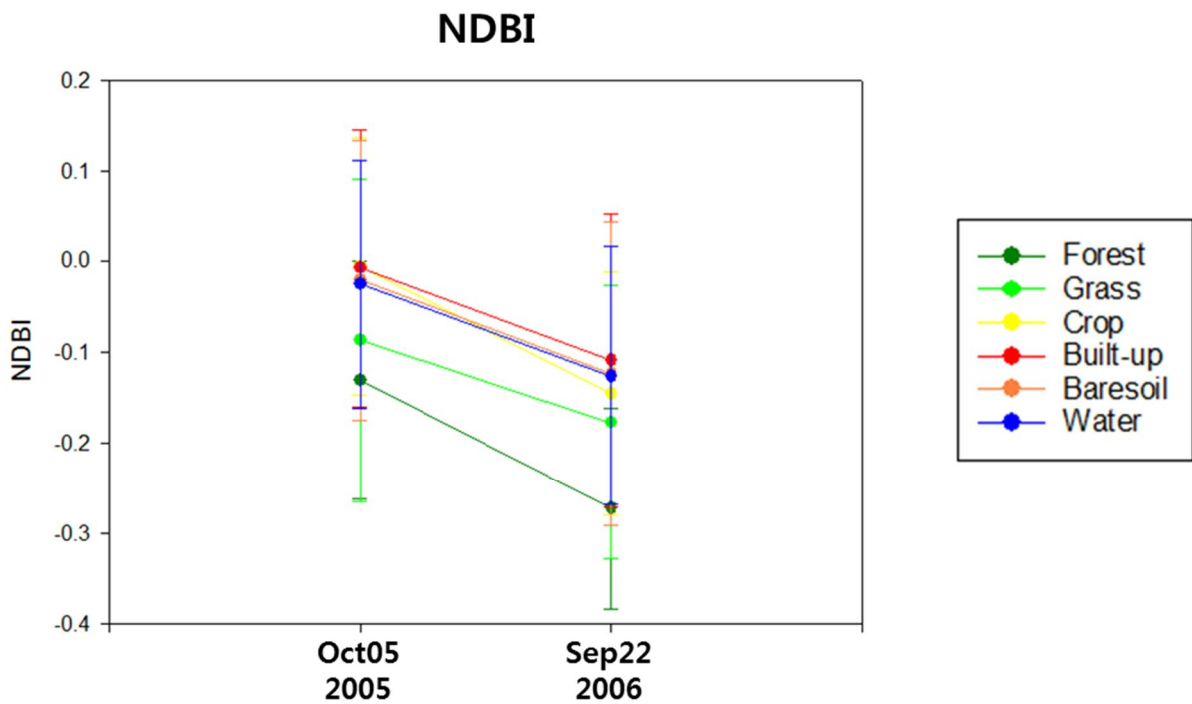
By the way, Figure 7 means the seasonal changes of NDVI and NDBI for the test year in case of using Landsat 8 and RapidEye satellites imageries together. And it also showed a similar tendency.

When comparing all three cases, it is confirmed that these variables show clear characteristics according to the land cover, and also show a certain change pattern according to season and time variation. As the variables that can best represent the sensitivity of vegetation and urbanization index, they are considered and utilized as meaningful variables in this analysis because the denuded forest land is also closely related to vegetation and forest loss.
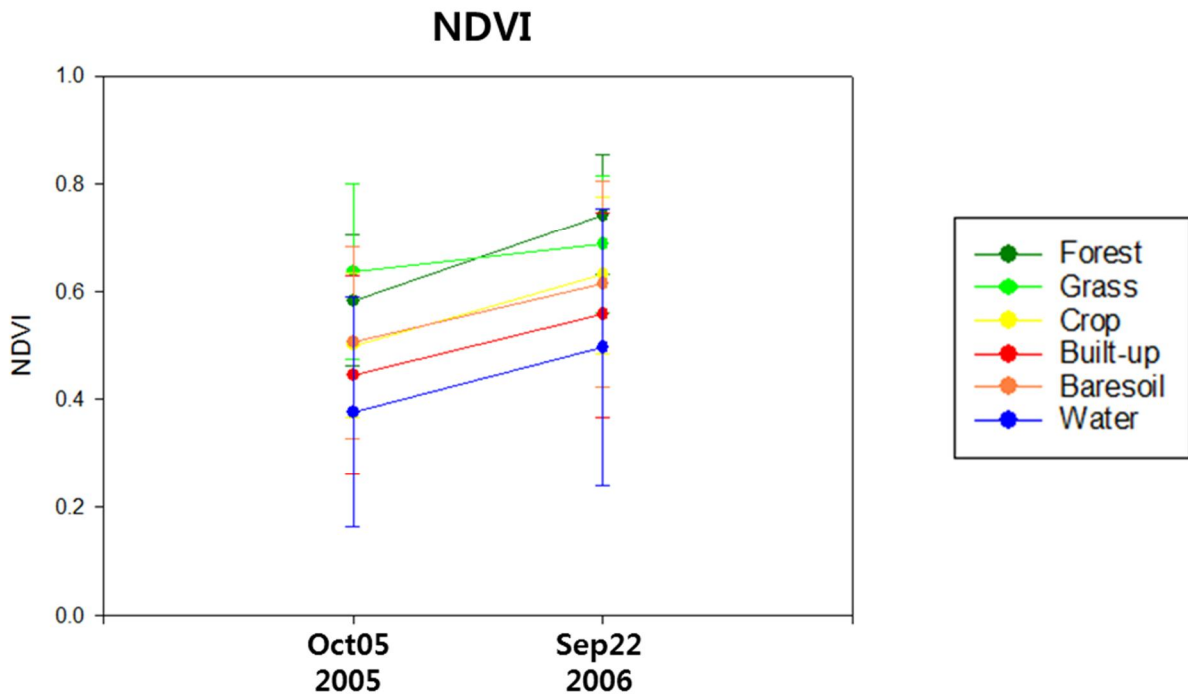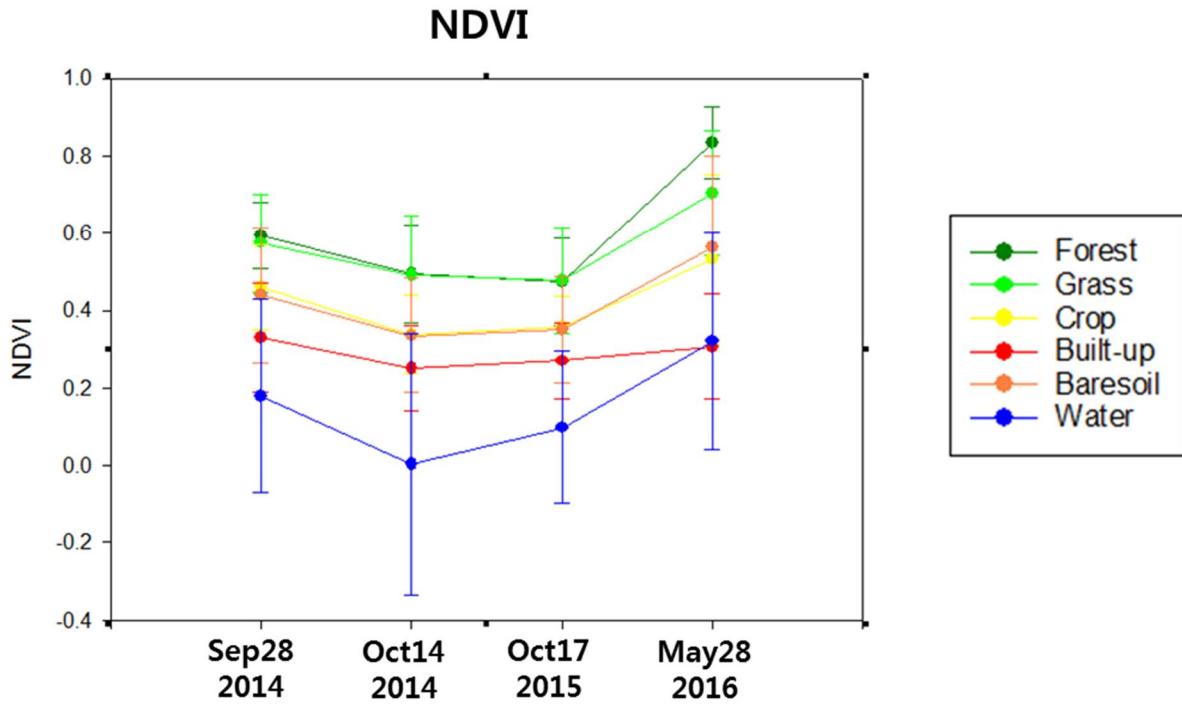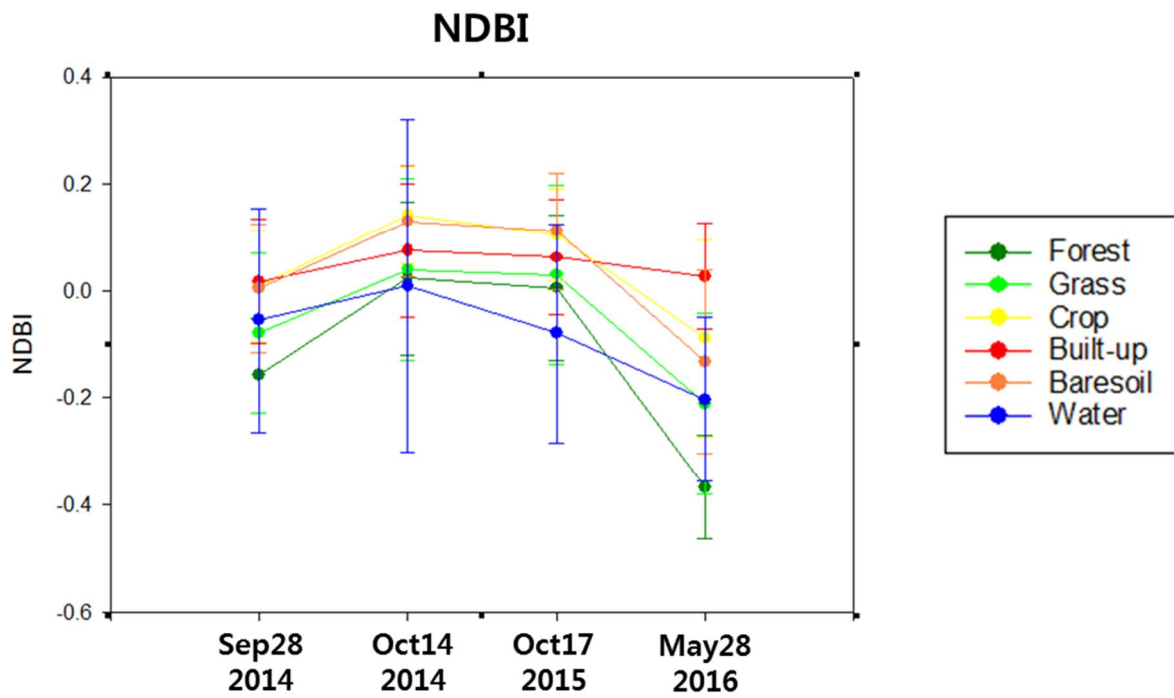
(a) NDVI variation

(b) NDBI variation

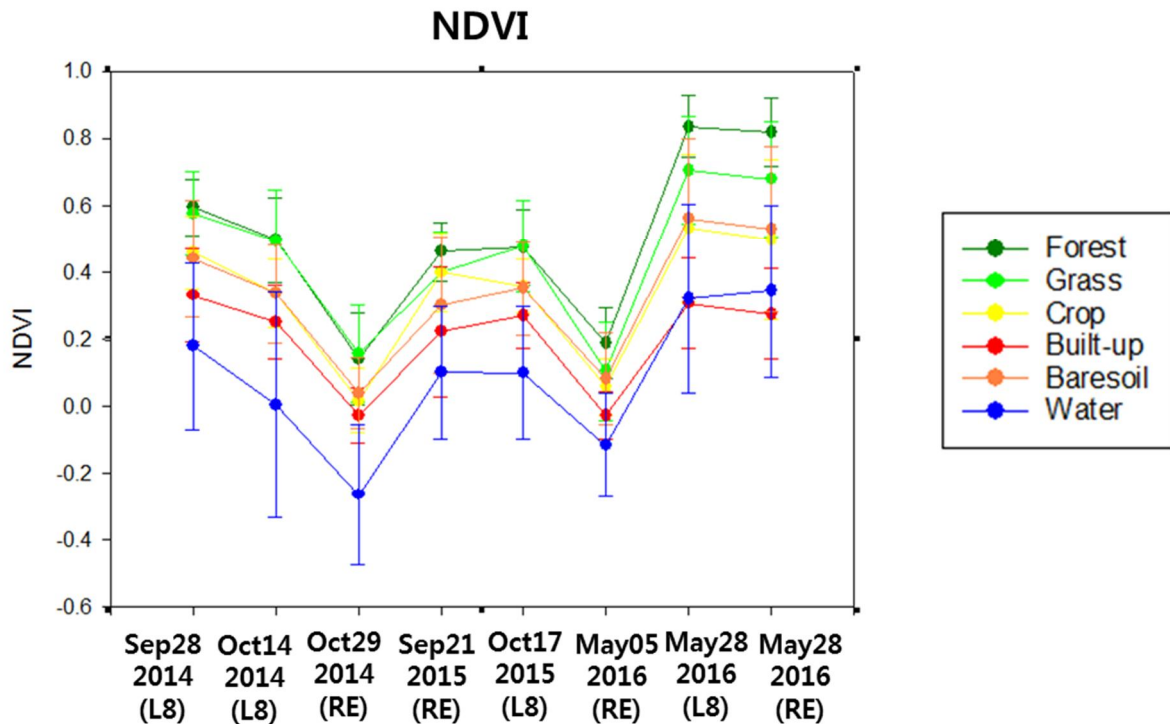**Figure 5 Variation of input variables for base year**

NDVI

(a) NDVI variation

NDBI

(b) NDBI variation

Figure 6 Variation of input variables for test year using Landsat 8

Figure 7 NDVI variance for test year using both Landsat 8 and RapidEye

## 5.2 Accuracy analysis and importance of variables

To calculate overall classification accuracy, random forest was applied in this study. And for this, the dataset was obtained from sample points where all variable information was combined. And the dataset was divided into 7:3 for training and verification data. After performing random forest model with the prepared dataset, accuracy matrix was calculated and then the overall accuracy was confirmed finally. The overall accuracy was the mean value of ten times random forest applying.

The overall accuracy of base year was 0.81, which was calculated with the Landsat 5 TM derived variables. In addition, as a result of random forest, importance of variables was also calculated as shown in Figure 8 and Figure 9 as well as overall accuracy. In the base year, total 19 input variables were used, including topographic variables such as DEM, slope, aspect, and the variables derived from 2 days Landsat 5 imageries. And the top 10 important variables were ranked in order of SWIR2 band (September 22, 2006), DEM, NDVI (September 22, 2006), B band (September 22, 2006), NDBI (October 05, 2005), NDBI (September 22, 2006), B band (October 05, 2005), SWIR1 (October 05, 2005), R band (September 22, 2006), and NIR band (September 22, 2006) as like Figure 8. In the base year, the variables derived from the satellite images taken on September 22, 2006 were highly

17

significant.

One the other hand, the overall accuracy of test year was 0.74 when using only Landsat 8 OLI derived variables. During the test year, total 35 input variables were used, including topographic variables such as DEM, slope, aspect, and the variables derived from 4 days Landsat 8 imageries. In that case, the importance of variables was high in order of SWIR1 (May 28, 201), DEM, NDVI (May 28, 2016), R band (May 28, 2016), NIR band (May 28, 2016), NDVI (October 14, 2014), NDBI (May 28, 2016), SWIR1 (May 28, 2016), B band (May 28, 2016), and NDVI (October 17, 2015) from Figure 9, (a). In this case, the variables derived from the satellite images taken on May 28, 2016 were highly significant. In addition, NDVI variables showed high importance.

Also the overall accuracy was 0.73 when using both Landsat 8 OLI and RapidEye derived variables as input. Total input variables were 59 in this case, including topographic variables such as DEM, slope, aspect, and the variables derived by both 4 days Landsat 8 imageries and 4 days RapidEye imageries. The importance of variables in this case was high in order of R band (May 28, 2016), NDVI (October 14, 2014), SWIR2 (May 28, 2016), NDVI (May 28, 2016), NIR band (May 28, 2016), NDVI-RapidEye (October 29, 2014), DEM, NDVI-RapidEye (May 28, 2016), NDVI (October 17, 2015), and NDBI (October 14, 2014) as like shown in Figure 9, (b). This case also the variables derived from the satellite images taken on May 28, 2016 were highly significant, and NDVI variables were also important.
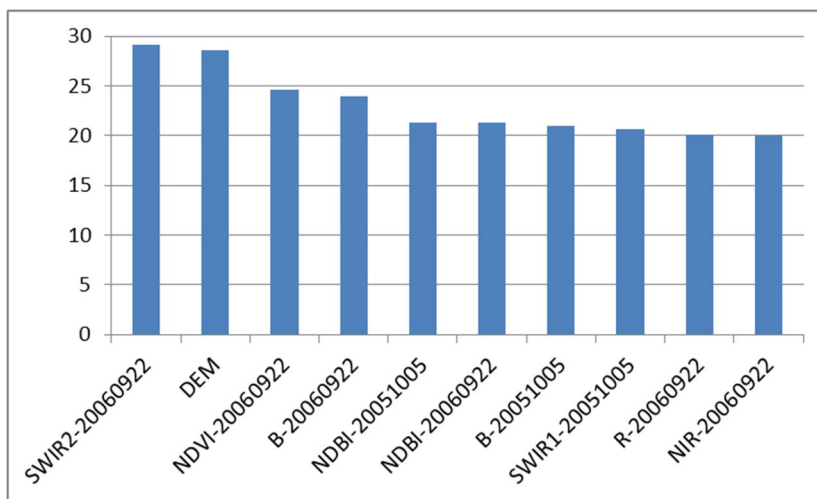
Figure 8 Importance of variables for base year
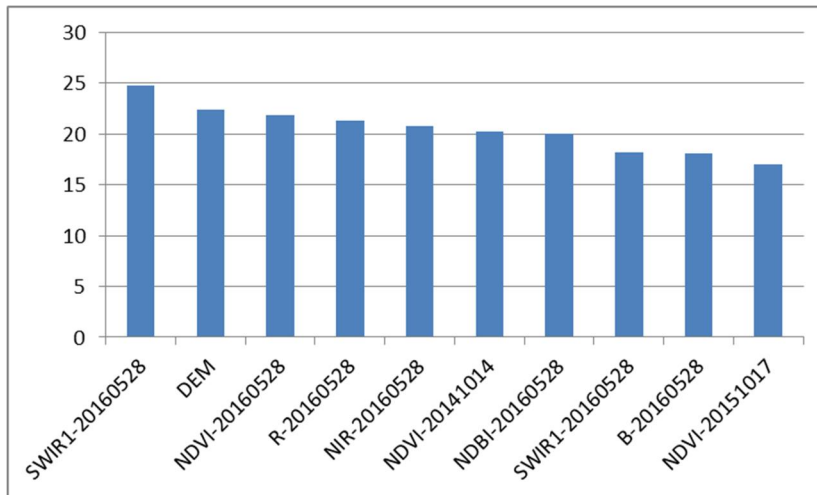
(a) Importance of variables for test year in case of using Landsat 8
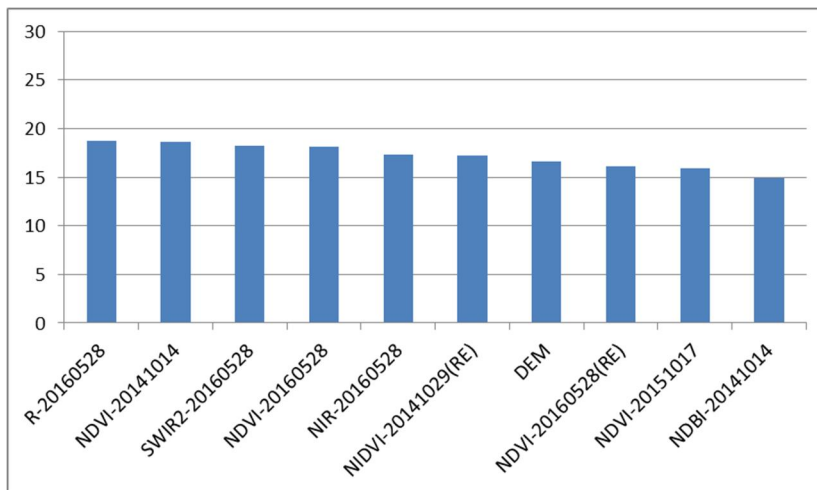
**(b) Importance of variables for the test year in case of using Landsat 8 and RapidEye**

Figure 9 Importance of variables for test year

5.3 Final classification map

The random forest model was applied to the entire study area to extract final classification map. Through the Matlab program, whole pixel point values of the entire study area were extracted as csv format. And the random forest prediction model was practiced with the extracted csv files. The prediction model was based on the random forest model which was generated by training and verification sample data above.

The final classification map includes 6 main land cover classes such as forest, grass, crop, built-up, bare-soil, and water. However, due to the regional feature of the study area, forest class showed the most dominant classes compared to other types of land cover. Next dominant class is crop land and

the other remaining classes are very small amount. Especially, grass land is rarely distributed in this area. Figure 10 shows the comparison of class distribution between the base year and test year. Overall, there is no significant change during 10 years between two periods, but crop, grass, and bare-soil classes increased slightly. The similar tendency can be found in the classification map, too.
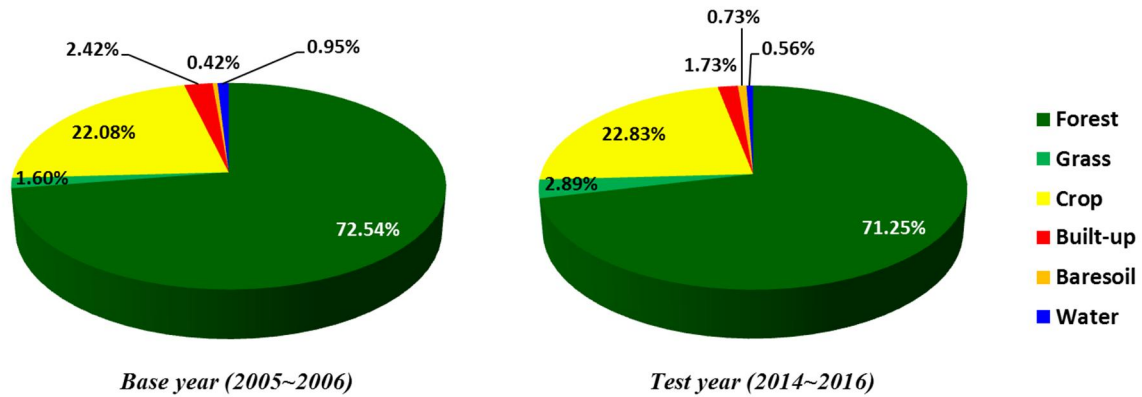
**Figure 10 Comparison of class distribution between base year and test year**

By the way, Figure 11 (a) shows the land cover classification map which was made by the Ministry of Environment in 2010. This land cover map was produced at 30 meter spatial resolution based on Landsat level. Even though there are some uncertainty due to the different image year between land cover map and final classification map of the base and test year, the land cover map can be used as a rough reference in this study. Comparing the land cover map with the final classification map, the classification map tends to overestimate the agricultural land in comparison with the land cover map. However, the extracted classification map is able to estimate some parts of grass, crop, and bare-soil more exactly than the land cover map where seems to be the results of forest degradation. And this is shown in the classification maps on Figure 11 to Figure 12.

In this study, three main denuded forest types of North Korea were categorized into each class as like deforested cropland to crop, unstocked forest land to grass, and bare land to bare-soil respectively. Compared to a base year, the classes which classified as denuded forest land increased slightly during the test year. For example, in the southern high mountain zone of Musan-gun, when observed through the actual Google Earth, the number of unstocked forest land has increased significantly. And those parts could be identified through the classification map actually.

(a) 2010 Land cover map   **(b) Base year classification map**
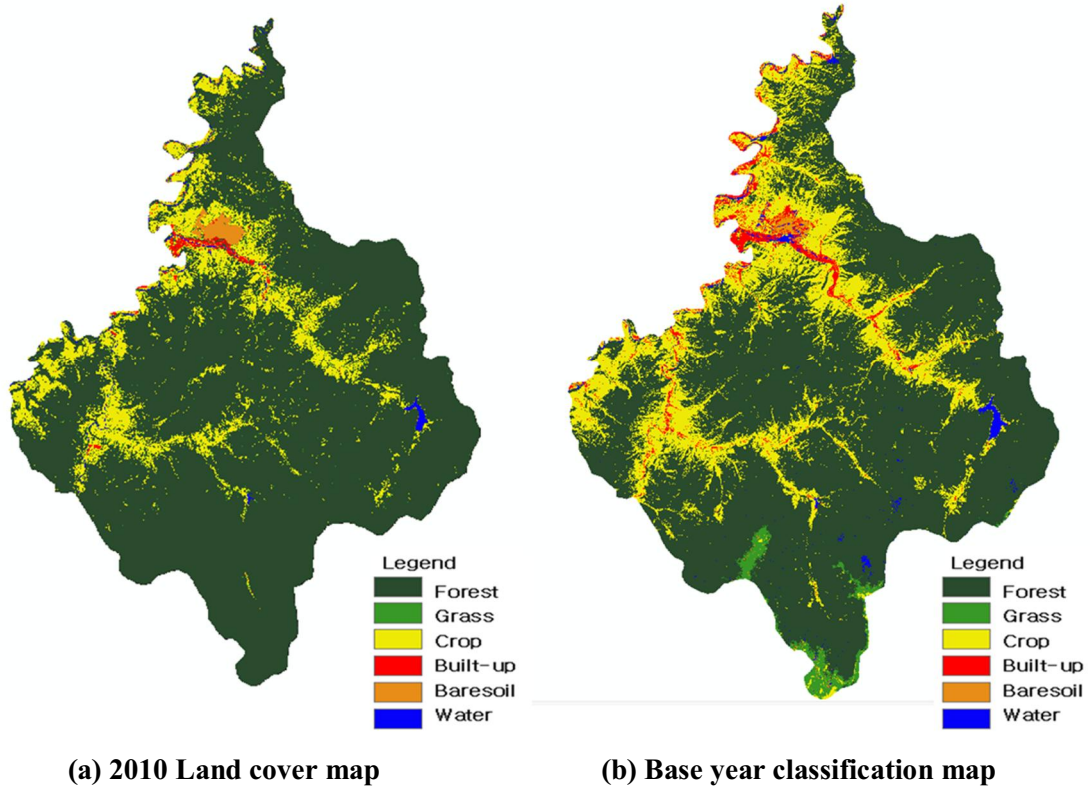
(Source: Ministry of Environment, South Korea)

Figure 11 Final classification map of base year (2005~2006)

(a) In case of using Landsat 8   **(b) In case of using Landsat 8 and RapidEye**
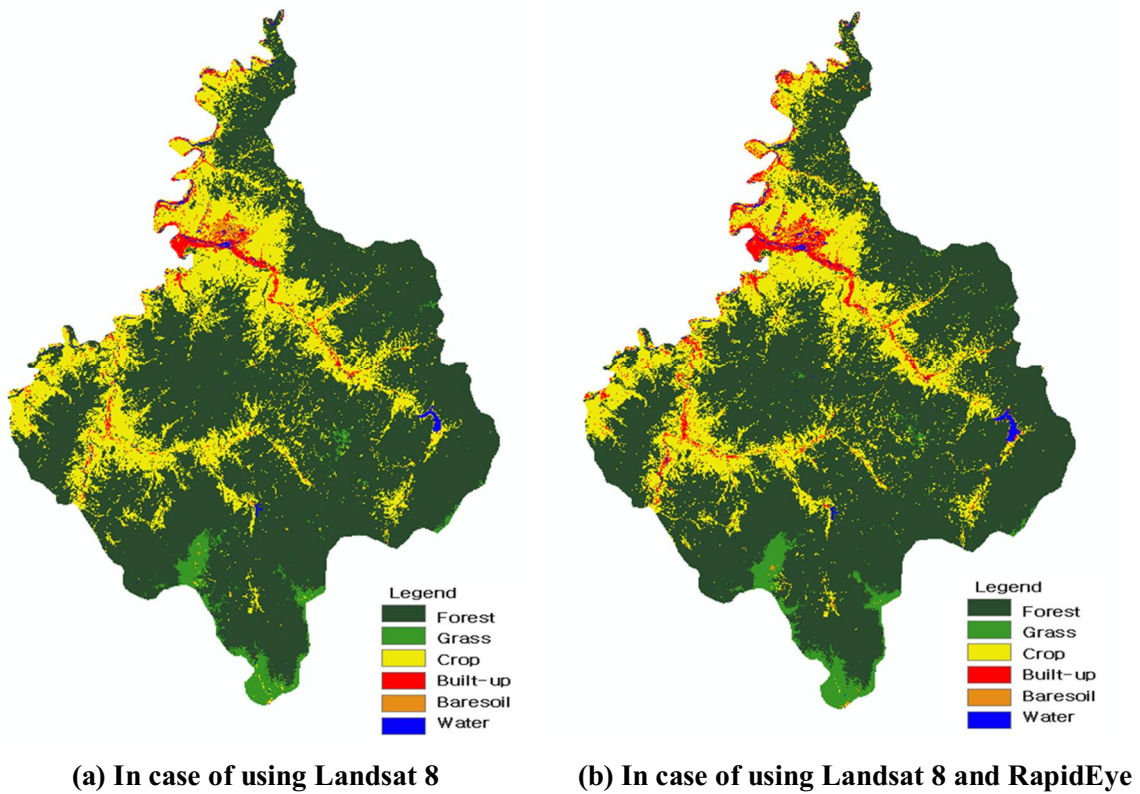
Figure 12 Final classification map of test year (2014~2016)

5.4 Forest changing analysis

In order to analyze the denuded forest land, this study first compares the classification map extracted through the machine learning random forest model and analyzed the overall forest area. According to Table 6, forest area during ten years period was 957.76 ㎢ and nonforest area during ten years period was 328.56 ㎢. Also the area changed from forest to nonforest area was 84.02 ㎢, and the area changed from nonforest to forest was 65.44 ㎢ respectively. As a result, the total forest area estimated by the change area has decreased by 18.58 ㎢ during ten years period, although there may be various reasons. The estimated changing area from forest to forest is like Figure 13 (a). And the changed and unchanged areas are distributed as like Figure 13 (b). In general, most of land cover changes are appearing along crop land or built-up area, and it shows that change has occurred considerably even in the southern high mountain area of Musan-gun.

As a next step, estimate denuded forest lands were calculated as like Table 7. Of course, it is difficult to say that all the changes have turned into denuded forest land, but in case of the change types shown in Table 7, almost all due to forest degradation and it can be estimated as the most significant cause of forest change in this area. According to Table 7, estimated change area from forest to crop is 67.62 ㎢ during ten years period, and this is the biggest change among others. This changed area is assumed as deforested cropland. Also, the estimated change area from forest to grass is 15.16 ㎢ during ten years period, and this changed area can be assumed as unstocked forest land. The last changing type in Table 7 is forest to bare-soil. And the estimated change area from forest to bare-soil is 0.73 ㎢ during ten years period. This can be assumed as bare land. Through forest change analysis, it is revealed that deforested cropland is the most prominent in Musan-gun among other types of deforestation.
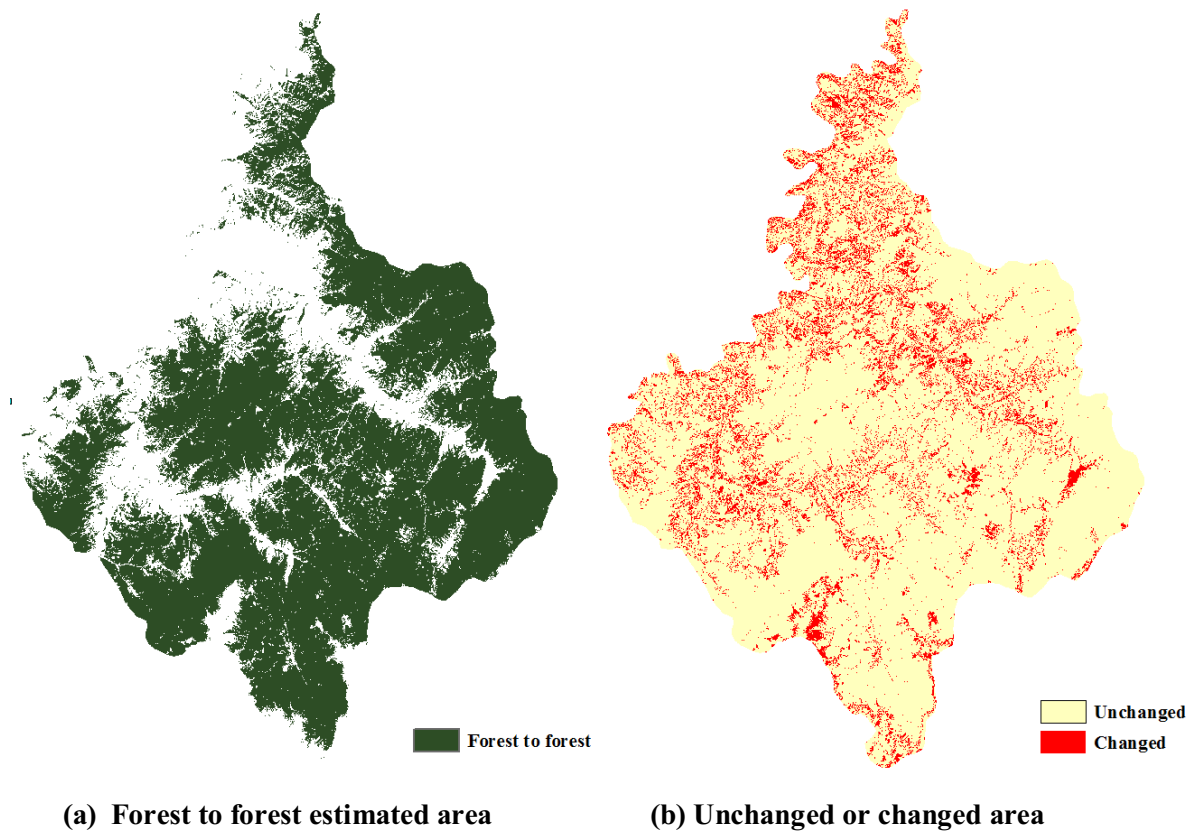
Table 6 Forest area changing analysis between base year and test year

		T2 (Test year)			
		Forest	**Area (㎢)**	**Nonforest**	**Area (㎢)**
T1	**Forest**	1064179	957.76	93356	84.02
(Base year)	**Nonforest**	72714	65.44	365069	328.56

Table 7 Estimated denuded forest land between base year and test year

Type	Pixel (Qty)	Area (㎢)
Forest to crop	75131	67.62
Forest to grass	16846	15.16
Forest to bare-soil	812	0.73

Forest to forest

Unchanged
Changed

(a) Forest to forest estimated area   (b) Unchanged or changed area

Figure 13 Estimation of changing area

Figure 14 shows the spatial distribution of estimated denuded forest land. As mentioned above, Figure 14 (a) forest to crop can represent deforested cropland which is the most major denuded forest land, (b) forest to grass shows the spatial distribution of unstocked forest land, and (c) forest to bare-soil shows the distribution of estimated bare land.

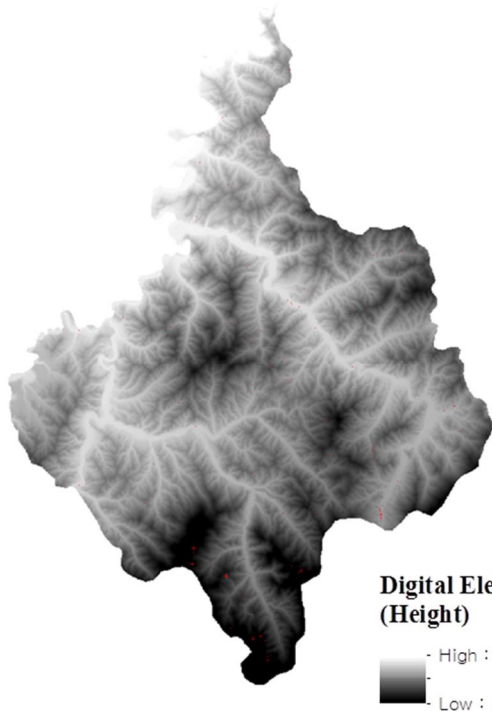In addition, changing areas per land cover class was compared in Figure 15 between the base year and test year. In the test year, grass and bare-soil classes were increased compared than base year. This indirectly indicates that forest degradation in the region has increased.
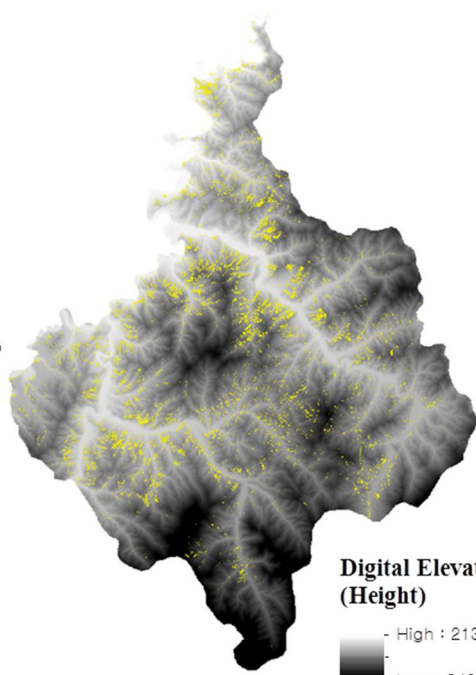
(a) Forest to crop
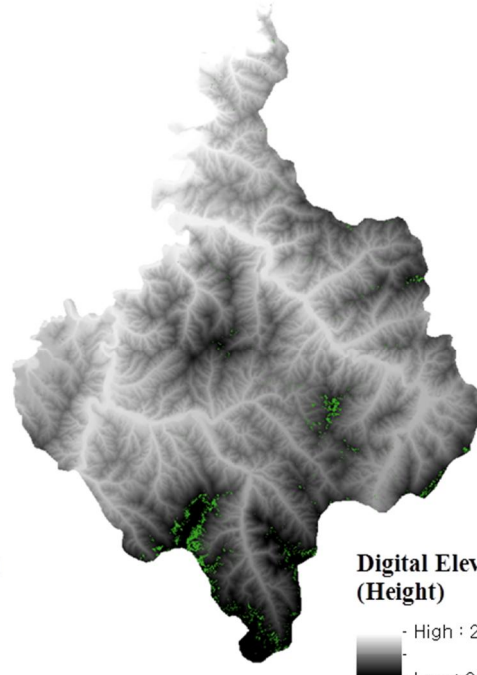
(b) Forest to grass

(c) Forest to bare-soil

Figure 14 Estimated denuded forest land between base year to test year
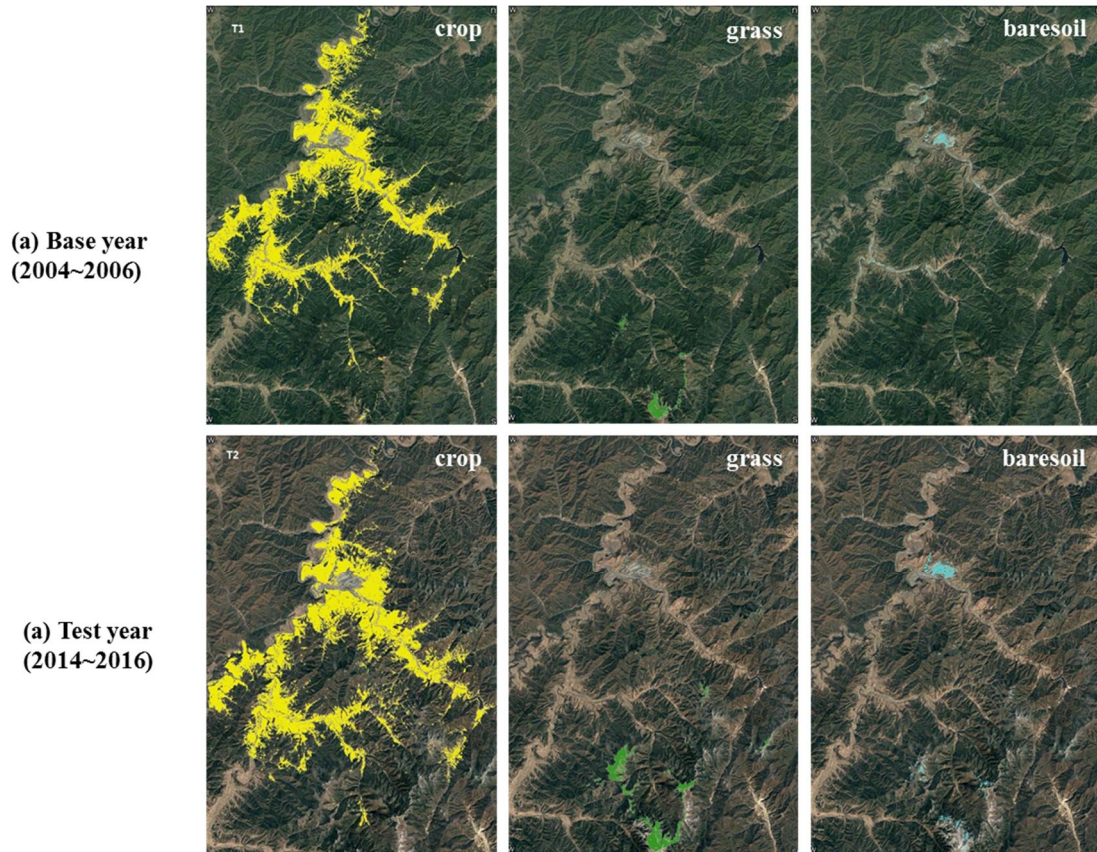
Figure 15 Comparison of changing area per land cover class

6. Discussion

As mentioned above, in this study, forest cover classification of Musan, the study area, was conducted through machine learning, and this method was used to analyze the status of deforestation in North Korea. The study examined the status of denuded forest land in North Korea between the base year and test year based on 10 years changes. Among the areas estimated as denuded forest land, particularly the changing areas of forest to crop land were high in Musan-gun. This implicates that forest degradation has been occurring by the indiscriminate clearing of mountain areas to crop land due to serious food shortages in North Korea.

By the way, in this study 30 m spatial resolution Landsat and 5 m high resolution RapidEye satellite imageries were used. In case of RapidEye, both growing and non-growing season imageries were available. But in case of Landsat, it was difficult to obtain clear images due to the influence of weather and clouds. Therefore, only a certain season, mostly fall season, satellite imageries were obtained. Such lack of available image and seasonal data is one of limitations in this study. It is expected that the accuracy and study results will be improved if the research is carried out by supplementing these limitations and adding available satellite imageries in the future.

7. Summary and Conclusion

North Korea is one of the most serious deforestation countries in recent decades. Especially, Musan-gun was one of the most vulnerable areas in North Korea by suffering rapid forest degradation. As a result of such poor forest condition, Musan-gun was extremely damaged by floods and landslides recently. However, due to the isolated political situation, few studies on the relevant data measuring or monitoring of denuded forest land in North Korea have been conducted so far. Nevertheless, it is very necessary to analyze and study North Korean forests because it could be as a valuable resource for the future Korean peninsula. Based on this view, it can be said that this study has an important meaning. Also, to monitor denuded forest land, using remote sensing satellite images is the only and most effective way in the case of North Korea because field survey is not available there. Thus, Google Earth multi time imageries were used as an indirect reference for the verification.

Therefore, this study aims to develop an improved technique to classify land cover and denuded forest land in North Korea. And this study monitored and analyzed the status of denuded forest land in

North Korea through remote sensing based machine learning approach. In this study, random forest machine learning technique was applied by using extracted random sample points and various input variables.

The denuded forest land, which is mainly found in North Korea, can be classified into three types according to the standard of United Nations, Food and Agriculture Organization (FAO). First denuded forest type is deforested cropland where crops are cultivated by forest clearing on the sloping fields. The second type is unstocked forest land that remains scrubland due to the extraction of firewood, grazing of livestock and so on. Last one is bare land which remains in the bald mountain or bare ground status with no trees.

This study examined by separating two periods, base year during 2005 to 2006 and test year during 2014 to 2016. This is because to analyze forest change status in the study site during 10 years. For the base year, 2 days Landsat 5 satellite imageries were applied while 4 days Landsat 8 and 4 days RapidEye were applied to the test year. The accuracy and the importance of variables were analyzed in accordance with the input variables and satellite data during operating random forest model. The overall accuracy was 0.81, 0.74, and 0.73 for the base year, test year in case of using Landsat 8 only, and test year when using Landsat 8 and RapidEye together, respectively. As a final result, the classification maps were derived which was categorized as 6 main land cover classes such as forest, grass, crop, built-up, bare-soil, and water.

Also forest changed areas were analyzed to estimate denuded forest land in this study. From the results, the area preserved as forest during ten years period was 957.76 ㎢ and the area changed from forest to nonforest was 84.02 ㎢. Specifically, the area changed from forest to crop, forest to grass, and forest to bare-soil were estimated as like 67.62 ㎢, 15.16 ㎢, 0.73 ㎢, respectively. And those changes considered as estimates denuded forest land

The purpose of this study is to improve the accuracy of land classification by using machine learning model and the high resolution satellite imageries. In addition, this study aims to find out the improved and effective techniques to examine denuded forest land in North Korea by using forest change detection method. From the study results, the presented method revealed high accuracy in forest

classification, also the results showed good performance in change detection of denuded forest land in North Korea during approximately 10 year period.

REFERENCES

1. Hong, S.Y., Rim, S. K., Lee, S.H., Lee, J.C, & Kim, Y. H. (2009). Spatial Analysis of Agro-Environment of North Korea Using Remote Sensing. *Korean Journal of Environmental Agriculture*, 27(2), 120-132.

2. Park, J. H. & Yu, J. S. (2009). Investigation of Deforestation in North Korea using Remote Sensing (in Korean). *Journal of environmental studies*, 48(8), 3-24.

3. Yoo, S. J., Lee, W. K., Lee, S. H., Kim, E. S., & Lee, J. Y. (2011). Approach for Suitable Site Selection and Analysis for Reforestation CDM using Satellite Image and Spatial Data in North Korea. *Journal of Korean Society for Geospatial Information System*, 19(3), 3-11.

4. Kim, E. S., Lee, S. H., & Cho, H.. K. (2010). Segment-based Land Cover Classification using Texture Information in Degraded Forest Land of North Korea. *Korean Journal of Remote Sensing*, 26(5), 477-487.

5. Lee, S. M., Park, S. H., & Jung, H. S. (2015). Monitoring in North Korea using Modified NDVI Method. *The 36th Asian Conference on Remote Sensing*.

6. Yu, J. S., & Kim, K. M. (2015). Spatio-Temporal Changes and Drivers of Deforestation and Forest Degradation in North Korea. *Journal of the Korea Society of Environmental Restoration Technology*, 18(6), 73-83.

7. Lee, S. H. (2004). Situation of Degraded Forest Land in DPRK and Strategies for Forestry Cooperation between South and North Korea. *J Agriculture & Life Sciences*, 38(3), 101-113.

8. Hamunyela, E., Verbesselt, J., & Herold, M. (2016). Using spatial context to improve early detection of deforestation from Landsat time series. *Remote Sensing of Environment* 172, 126-138.

9. Reiche, J., Verbesselt, J., Hoekman, D., & Herold, M. (2015). Fusing Landsat and SAR time series to detect deforestation in the tropics. *Remote Sensing of Environment* 156, 276-293.

10. Diniz, Cesar Guerreiro, et al. (2015). DETER-B: the new amazon near real-time deforestation detection system. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.7, 3619-3628.

11. Eckert, S., Husler, F., Liniger, H., & Hodel, E. (2015). Trend analysis of MODIS NDVI time series for detecting land degradation and regeneration in Mongolia. *Journal of Arid Environments* 113, 16-28.

12. Arai, E., Shimabukuro, Y. E., Pereira, G., & Vijaykumar, N. L. (2011). A multi-resolution multi-temporal technique for detecting and mapping deforestation in the Brazilian Amazon rainforest. *Remote Sensing* 3.9, 1943-1956.

13. Choi, J. H. & Um, J. S. (2012). Application of Satellite Image to Evaluate UN-REDD

Registration Potential of North Korea : a Case Study of Mt. Geumgang. *Journal of the Korean Society for Geospatial Information System*, 20(4), 77-87.

14. Kim, O.S. & Youn, Y.C (2014). An Identification of Project Sites for Lowering Carbon Emissions and Saving Forests in DPR Korea. *Journal of the Korean Geographical Society* 49.2, 264-274.

15. Oskar, G. P., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters* 27(4), 294-300.

16. Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 67, 93-104.

17. Mahesh, P. (2005). Random forest classifier for remote sensing classification. International *Journal of Remote Sensing,* 26(1), 217-222.

18. Markus, I., Atzberger, C., & Koukal, T. (2012). Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sensing* 4(9), 2661-2693.

19. Baccini, A. G. S. J., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., ... & Samanta, S. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, *2*(3), 182-185.

20. Jin, Y., Sung, S., Lee, D. K., Biging, G. S., & Jeong, S. (2016). Mapping Deforestation in North Korea Using Phenology-Based Multi-Index and Random Forest. *Remote Sensing*, *8*(12), 997.

21. Beers, T. W., Dress, P. E., & Wensel, L. C. (1966). Notes and observations: aspect transformation in site productivity research. *Journal of Forestry*, *64*(10), 691-692.

22. Jensen, J. R. (1986). *Introductory digital image processing: a remote sensing perspective*. (The 4th Edition, pp. 556-56, 219-220, 443-444). Univ. of South Carolina, Columbus.

23. Joint Assessment North Hamgyong Floods 2016 report. UN OCHA.

24. Lee, Y. G. (2014). A Study on Inter-Korean Economic Cooperation Measures Using the UN Response Mechanism for Climate Change - Final Report (in Korean). Climate Change Center.