International Journal of Innovations in Engineering and Technology (IJIET)
http://dx.doi.org/10.21172/ijiet.133.10

Evaluation of the Different Parameters of the KNN Classifier for Identifying the Various Land Covers Using the Landsat Satellite Image Acquired in Osong and Sejong City Areas

Louis Sungwoo Cho[1], Yoon-Je Jeong[2]
[1]*Asia Pacific International School, Seoul, 01874, Republic of Korea*
[2]*GIS Research Center, U&GIT Co. Ltd., Seoul, 05717, Republic of Korea*

Abstract- The generation of the land cover map is necessary for identifying the various land covers in huge areas. The satellite image is useful for generating the land cover map because it can provide the surface information of huge areas without human access. In this research, the different parameters (10 and 1000) of the K-nearest neighbor (KNN) classifier, a widely used machine learning technique, were evaluated to generate the land cover maps for identifying the various land covers (soil, urban, vegetation and water) using the Landsat satellite image acquired in the Osong and Sejong city areas, where significant land cover changes have been experienced due to the construction of the Osong Korean Train eXpress (KTX) station and the Sejong government complex. The experiment results showed that the land cover map generated with the parameter 10 of the KNN classifier had 96% overall accuracy while the land cover map generated with the parameter 1000 had 56% overall accuracy for identifying the various land covers in the Osong and Sejong city areas using the given Landsat-8 satellite image.
Keywords – Land cover map, Landsat satellite image, Machine learning, K-nearest neighbor (KNN) classifier

I. INTRODUCTION

Land cover is defined as “how land is utilized” [1]. The land cover map documents how much of the region is covered by vegetation, urban areas, soil, water, and other land types. Land cover mapping is useful for monitoring the important variables on earth surface [2]. Satellite images are generally useful for generating the land cover maps because it is capable of detecting land cover changes in the study area in a certain time period. Fonji and Taff [3] generated a land cover map using Landsat satellite images acquired in northeastern Latvia. Sheffield et al. [4] generated land cover maps using the MODIS satellite images acquired in Australia. Malarvizhi et al. [5] utilized Google Earth imagery to generate a land cover map.

The land cover map of the Osong and Sejong city areas is recently necessary because these areas have the significant land cover changes due to the construction of the Osong Korean Train eXpress (KTX) station and the Sejong government complex (see Figure 1).

(a) (b)

Figure 1 (a) Osong KTX station [6]; and (b) Sejong government complex [7]

The machine learning techniques have advantages for generating land cover maps using satellite images because the various land covers can be automatically classified from the satellite images with high accuracy [8]. The K-nearest neighbor (KNN) classifier is a widely used machine learning technique for detecting the various land covers from the satellite images by assigning the appropriate class to each test point based on the number of training samples lying in proximity to the test point [9]. Research on determining the appropriate parameter of the KNN classifier for detecting the various land covers has not been carried out.

This research aimed to evaluate the different parameters of the KNN classifier for mapping the land covers using the Landsat satellite image acquired in the Osong and Sejong city areas, through the following procedure: In the first step, the Landsat satellite image acquired in the Osong and Sejong cities was given. The next step was to evaluate the different parameters of the KNN classifier to generate the different land cover maps using the given Landsat satellite image. Finally, the accuracy of each land cover map was assessed for determining the appropriate parameter of the KNN classifier for identifying the land covers using the given Landsat satellite image.

II. DATASETS AND STUDY AREA

This research selected the Osong and Sejong cities as the study area because this area recently suffered significant land cover changes since Osong Korean Train eXpress (KTX) Station and the Sejong government complex were constructed in 2010 and 2012, respectively. Figure 2 shows the Osong and Sejong city areas selected as the study area in this research.

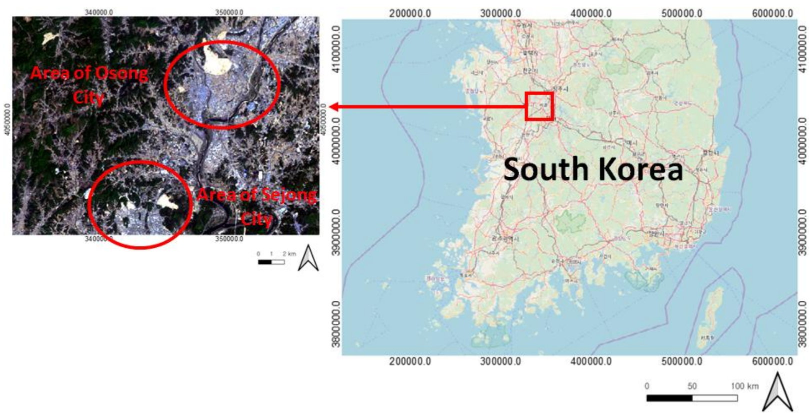


Figure 2 Osong and Sejong city areas, selected as the study area in this research

As can be seen in Figure 2, the selected study area consists of the multiple land covers such as urban areas, soil areas, vegetation areas, and water areas.

The Landsat satellite images are widely used for monitoring the land use changes because the Landsat satellites revisit the same areas every 16 days [10]. The Landsat satellite image used in this research was acquired on April 27, 2017, with the minimal cloud coverage. The given Landsat satellite image consists of the six spectral bands: the blue band (wavelength: 0.45-0.51μm), green band (wavelength: 0.53-0.59μm), red band (wavelength: 0.64-0.67μm), near-infrared band (wavelength: 0.85-0.88μm), shortwave infrared band 1 (wavelength: 1.57-1.65μm), and shortwave infrared band 2 (wavelength: 2.11-2.29μm), and the spatial resolution of the given Landsat satellite image is 30 meters [11].

III. METHODOLOGY

This section illustrates the procedure of the proposed methodology for evaluating the different thresholds of the KNN classifier to detect the land covers from the given Landsat satellite image acquired in the Osong and Sejong city areas. Figure 3 shows a flowchart presenting the procedure.

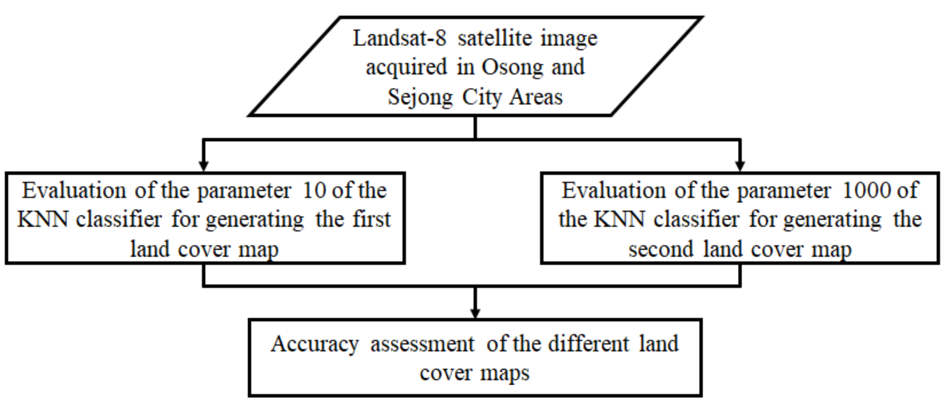
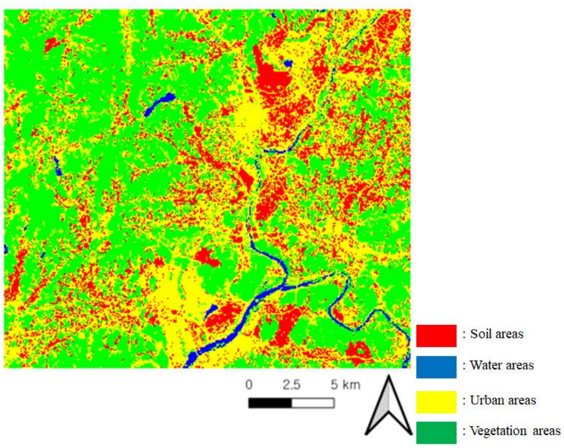


Figure 3 Flowchart presenting the procedure for evaluating the different parameters of the KNN classifier to detect the land covers from the given Landsat satellite image acquired in the Osong and Sejong city areas

The KNN classifier is a simple machine learning algorithm that assigns the appropriate class to each test point considering the number of training samples lying in proximity to the test point, and the class assigned to the test point is determined by the number of training samples, called "K" [9]. In the first step of the proposed methodology, the training sample group is set to include 1,000 pixels for each land cover class (urban, bare soil, vegetation, and water). Then, we separately employed the different parameters, such as 10 and 1000, selected as the K values for generating the land cover maps using the given Landsat-8 satellite image. Figure 4 shows the land cover maps generated from the given Landsat-8 satellite image by separately employing the different parameters 10 (called as the first land cover map) and 1000 (called as the second land cover map) of the KNN classifier.



(a) (b)

Figure 4 Land cover maps generated from the given Landsat-8 satellite image by separately employing the different parameters of the KNN classifier: (a) First land cover map generated by employing the parameter 10 of the KNN classifier; and (b) Second land cover map generated by employing the parameter 1000 of the KNN classifier

IV. RESULTS AND DISCUSSION

In this section, the accuracy of the generated land cover map was assessed using the 100 ground truth points manually digitized and located on the given Landsat satellite image. Figure 5 shows examples of such ground truth points.

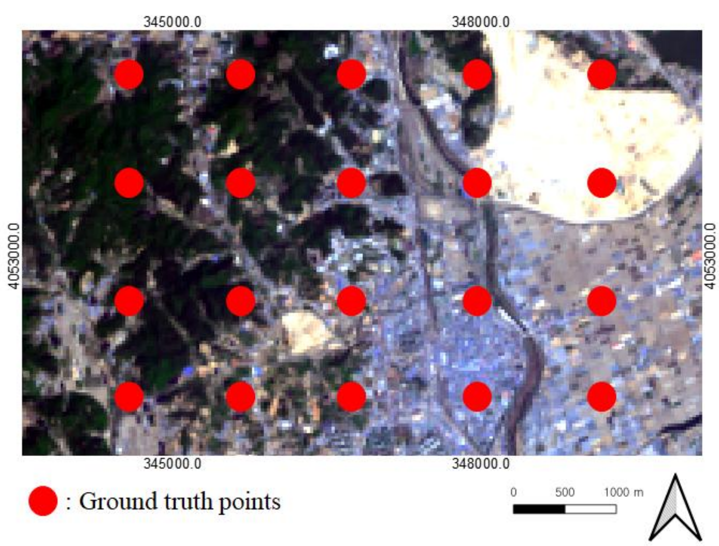


Figure 5 Examples of the ground truth points manually digitized and located on the given Landsat satellite image

Table 1 shows the classification error matrices of the identified land covers in the first and second land cover maps shown in Figure 4.

Table -1 Classification error matrix of the identified land covers (a) in the first land cover map; and (b) in the second land cover map

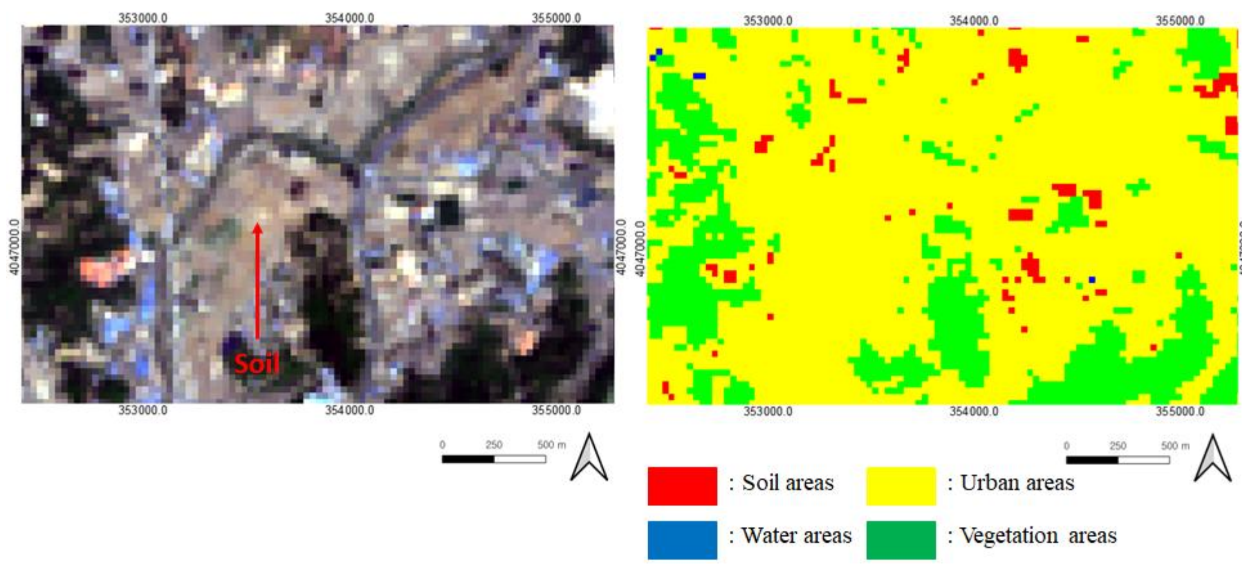
(a)

Class	Soil (Reference)	Urban (Reference)	Vegetation (Reference)	Water (Reference)	User's accuracy
Soil (Classified)	34	0	0	0	100%
Urban (Classified)	4	9	0	0	69%
Vegetation (Classified)	0	0	37	0	100%
Water (Classified)	0	0	0	16	100%
Producer's accuracy	89%	100%	100%	100%	96%

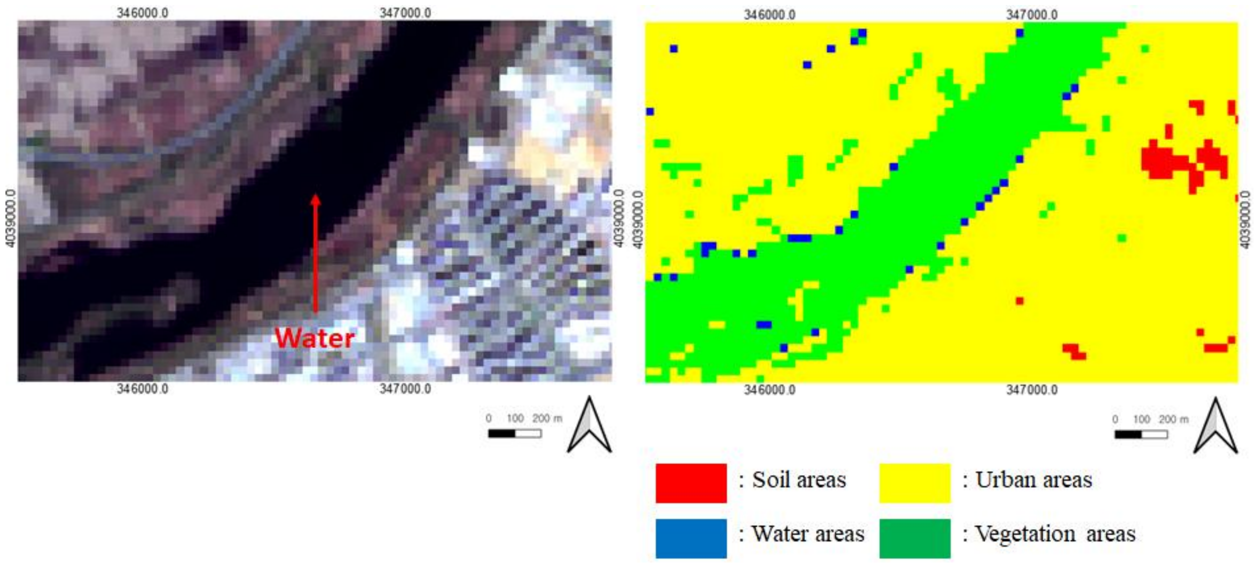
(b)

Class	Soil (Reference)	Urban (Reference)	Vegetation (Reference)	Water (Reference)	User's accuracy
Soil (Classified)	9	0	0	0	100%
Urban (Classified)	29	9	0	0	24%
Vegetation (Classified)	0	0	37	15	71%
Water (Classified)	0	0	0	1	100%
Producer's accuracy	24%	100%	100%	6%	56%

As can be seen in Table 1, the first land cover map had 96% overall accuracy while the second land cover map had 56% overall accuracy for identifying the various land covers. Figure 6 shows the misclassification errors occurred in the second land cover map where the soil features were misclassified into the urban features (a); and the water features were misclassified into the vegetation features (b).



(a)



(b)

Figure 6 Examples of the misclassification errors occurred in the second land cover map: (a) where the soil features were misclassified into the urban features; and (b) where the water features were misclassified into the vegetation features

As can be seen in Table 1 and Figure 6, the parameter 10 of the KNN classifier provides better accuracy than the parameter 1000 of the KNN classifier for detecting the various land covers from the given Landsat-8 satellite image.

V. CONCLUSIONS AND FUTURE WORKS

The generation of an accurate land cover map is necessary for identifying the various land covers and monitoring the sustainable land cover changes. In this research, an experiment was performed to evaluate the different parameters 10 and 1000 of the KNN classifier for identifying the various land covers (soil, urban, vegetation, and water) using the Landsat-8 satellite image acquired in the Osong and Sejong city areas, which recently experienced significant land cover changes due to the constructions of Osong KTX Station and the Sejong government complex.

The experiment results showed that the land cover map generated with the parameter 10 of the KNN classifier had 96% overall accuracy while the land cover map generated with the parameter 1000 of the KNN classifier had 56% overall accuracy in terms of identifying the various land covers in the Osong and Sejong city areas. In future research, the “multi-temporal” Landsat satellite images will be utilized for monitoring the land cover changes in the study area after the constructions of Osong KTX Station and the Sejong government complex.

VI. REFERENCES

[1] UC Davis, Land Use and Land Cover - Teacher’s Guide, 2019, http://cstars.metro.ucdavis.edu/files/5413/8542/6968/Land_Use_and_Land_Cover_Teachers_Guide.pdf
[2] South African Environmental Observation Network (SAEON), “Creating land cover maps to monitor the most important variable of global change”, 2019, http://www.saeon.ac.za/enewsletter/archives/2008/february-2008/document-actions-creating-land-cover-maps-to-monitor-the-most-important-variable-of-global-change
[3] S.F. Fonji, and G.N. Taff, “Using satellite data to monitor land-use land-cover change in North-eastern Latvia,” Springerplus, vol. 3, no. 61, 2014.
[4] K. Sheffield, E. Morse - Mcnabb, R. Clark, S. Robson, and H. Lewis, “Mapping dominant annual land cover from 2009 to 2013 across Victoria, Australia using satellite imagery”. Scientific Data. Volume 2, Article Number 150069, 2015.
[5] K. Malarvizhi, S.V. Kumar, and P. Porchelvan, “Use of High Resolution Google Earth Satellite Imagery in Land use Map Preparation for Urban Related Applications,” Procedia Technology, vol. 24, pp. 1835-1842, 2016.
[6] Chosun-Ilbo, A KTX train at Osong station, 2019, http://news.chosun.com/site/data/html_dir/2018/11/20/2018112002988.html
[7] Mcnews, Government Complex at Sejong City, South Korea, 2015, http://www.mcnews.co.kr/44946
[8] Y.-J. Choung, and M.-H. Jo, “Comparison between a Machine-Learning-Based Method and a Water-Index-Based Method for Shoreline Mapping Using a High-Resolution Satellite Image Acquired in Hwado Island, South Korea,”. Journal of Sensors, vol. 2017, pp. 1-13, 2017.
[9] R. Gandhi, “K Nearest Neighbours - Introduction to Machine Learning Algorithms”, Medium, 2019, https://towardsdatascience.com/k-nearest-neighbours-introduction-to-machine-learning-algorithms-18e7ce3d802a
[10] Earth Observation System, Landsat-8, 2019, https://eos.com/landsat-8/
[11] United States Geological Survey (USGS), Landsat 8 band designations, 2019, https://www.usgs.gov/media/images/landsat-8-band-designations