

Classification of Surface Water Using Machine Learning Methods from Landsat Data in Nepal [†]

Tri Dev Acharya ¹, Anoj Subedi ², He Huang ³ and Dong Ha Lee ^{1,*}

¹ Department of Civil Engineering, Kangwon National University, Chuncheon 24341, Korea; tridevacharya@kangwon.ac.kr

² Institute of Forestry, Pokhara Campus, Tribhuvan University, Pokhara 33700, Nepal; anojsubedi99@gmail.com

³ School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture, Beijing 100044, China; huanghe@bucea.edu.cn

* Correspondence: geodesy@kangwon.ac.kr; Tel.: +82-33-250-6232

[†] Presented at the 5th International Electronic Conference on Sensors and Applications, 15–30 November 2018; Available online: <https://ecsa-5.sciforum.net>.

Published: 15 November 2018

Abstract: With over 6000 rivers and 5358 lakes, surface water is one of the most important resources in Nepal. However, the quantity and quality of Nepal's rivers and lakes are decreasing due to human activities and climate change. Therefore, the monitoring and estimation of surface water is an essential task. In Nepal, surface water has different characteristics such as varying temperature, turbidity, depth, and vegetation cover, for which remote sensing technology plays a vital role. Single or multiple water index methods have been applied in the classification of surface water with satisfactory results. In recent years, machine learning methods with training datasets, have been outperforming different traditional methods. In this study, we tried to use satellite images from Landsat 8 to classify surface water in Nepal. Input of Landsat bands and ground truth from high resolution images available from Google Earth is used, and their performance is evaluated based on overall accuracy. The study will be helpful to select optimum machine learning methods for surface water classification and therefore, monitoring and management of surface water in Nepal.

Keywords: classification; machine learning; surface water; Landsat; Nepal

1. Introduction

Nepal is a geographically diverse country with flats in the south and increasing hills towards the north, to the mighty Himalayas. In Nepal, around 70% to 90% of the total annual rainfall occurs during the monsoon period resulting in high runoff and sediment discharge and causing surface water area change [1]. Thus, it is rich in water resources with about 600 rivers [2] and 5358 lakes [3]. Due to such seasonal variation and large surface water area, it is difficult to track changes in surface water. Furthermore, the quantity and quality of Nepal's rivers and lakes are decreasing due to human activities and climate change. Therefore, the monitoring and estimation of surface water is an essential task.

In Nepal, surface water has different characteristics such as varying temperature, turbidity, depth, and vegetation cover. In such cases, remote sensing satellite images are very well used to classify water bodies. For the identification of surface water using Landsat image, we used various techniques in our previous studies, such as, water index methods, single or combined [4,5], decision tree-based classification [6], and segmentation of scene [7] in diverse areas of Nepal.

In recent years, machine learning methods with training datasets, have been outperforming different traditional methods. In this study, we tried to use satellite image from Landsat 8 to classify surface water in Nepal.

2. Materials and Methods

2.1. Case Study

In order to achieve a more accurate comparison of the classification work, a Landsat scene from a previous study [7] was used. The scene contains various types of surface water, which can be compared with each other for classification results. Figure 1 shows the Landsat 8 scene in natural color composite. A total of 800 ground truths with 614 non-water and 186 water points within the scene were extracted from high resolution images available from Google Earth. These were used for training purposes as well as later validation in the classification process.

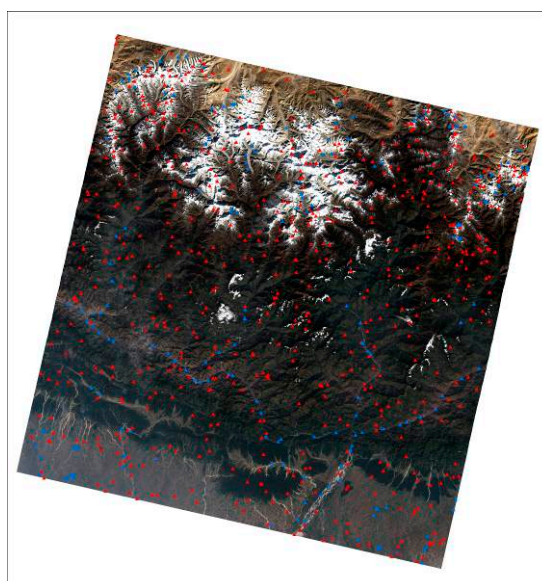


Figure 1. Landsat 8 scene in natural color composite with training points (blue: Water and red: Non-water) acquired on 8 December 2017.

2.3. Method

After preprocessing the at-satellite reflectance, the images were first used to extract all Operational Land Imager (OLI) bands values for the training dataset. After forming the training dataset, four machine learning methods were used to train different models in R and later applied to the full scene to classify the image into binary water and non-water map.

A Random Forest (RF) is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control overfitting.

Recursive Partitioning (RPART) is a type of binary tree used for classification or regression tasks. It performs a search over all possible splits by maximizing an information measure of node impurity, selecting the covariate showing the best split.

Support Vector Machine (SVM) is a data classification method that separates data using a hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which separates only one type of data. SVM technique is generally useful for data which is non-regularity, which means data whose distribution is unknown.

Neural Network (NNET) in R, is a feed-forward neural network with a single hidden layer flowing left to right. Feed-forward neural networks were the first type of artificial neural network invented, and are simpler than their counterparts, recurrent neural networks. They are called feed-forward networks because information only travels forward in the network (no loops), first through

the input nodes, then through the hidden nodes (if present), and finally through the output nodes. These are primarily used for supervised learning in cases where the data to be learned is neither sequential nor time-dependent.

After the classification, the given dataset was reclassified to evaluate for overall accuracy.

3. Results and Discussion

After the models were developed, the surface water using the selected four machine learning methods was derived according to the full scene as shown in Figure 2.

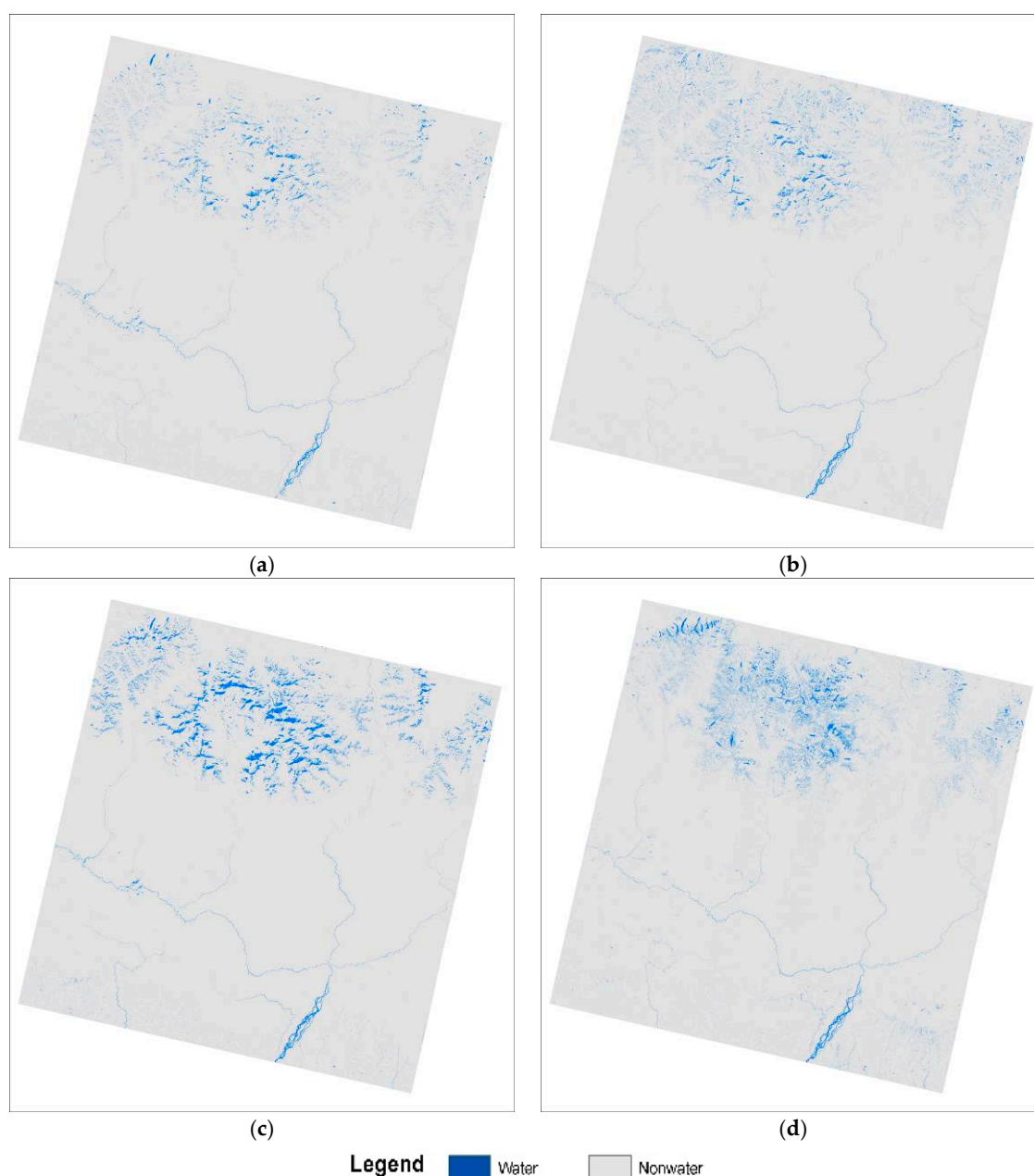


Figure 2. Surface water derived from Landsat 8 OLI image using various machine learning methods: (a) Random Forest; (b) Recursive Partitioning Tree; (c) Support Vector Machine; (d) Neural Network.

In Figure 2, we can see that the resulting water maps show a visible similarity in the lower lands for rivers, whereas variation exists in the upper Himalaya regions. These variations seem to be mostly in cold icy water areas, hill shades, and forest areas.

Table 1 shows the results of the overall accuracy from all the four methods, in which RF performed at 1, the highest level, and SVM performed lowest with 0.926. Both RPART and NN

showed an overall accuracy of 0.95. Comparing the result of these methods with the previous study [7] with the same training points, there seems to be a vast improvement against single or combined index methods. However, segmentation accuracy was still higher i.e., 0.96 against machine learning methods, except RF.

Table 1. Formulae of spectral Indices and their threshold applied in the study area.

S. No.	Machine Learning Methods	Overall Accuracy
1	Random Forest	1.00
2	Recursive Partitioning Tree	0.950
3	Support Vector Machine	0.926
4	Neural Network	0.956

4. Conclusions

In this study, we applied four machine learning methods: RF, RPART, SVM, and NN to derive the surface water map using a Landsat 8 OLI images in Nepal. Using previous training data and Landsat scenes, surface water was modeled and applied. Our results show that the snow and cold-water areas with hill shadows in the Himalayas caused a difference in water detection among the four used methods. In addition, RF has shown a maximum overall accuracy of 1 for the scene within the given dataset. It seems that machine learning methods could be very useful for the accurate automated binary classification of surface water in complex geographies, such as Nepal.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Khanal, N.R.; Chalise, S.R.; Pokhrel, A.P. Ecohydrology of river basins of Nepal, Ecohydrology of High Mountain Areas. In Proceedings of the International Conference on Ecohydrology of High Mountain Areas, ICIMOD, Kathmandu, Nepal, 24–28 March 1996; Chalise, S.R., Ed.; ICIMOD: Kathmandu, Nepal, 1996; pp. 49–61.
2. WECS. *Water Resources of Nepal in the Context of Climate Change*; Water and Energy Commission Secretariat, Government of Nepal: Kathmandu, Nepal, 2011.
3. Bhujju, U.R.; Khadka, M.; Neupane, P.K.; Adhikari, R. Lakes of Nepal: 5358-A Map Based Inventory; Ministry of Tourism and Civil Aviation, Government of Nepal; Kathmandu, Nepal, 2009.
4. Acharya, T.D.; Yang, I.T.; Subedi, A.; Lee, D.H. Change Detection of Lakes in Pokhara, Nepal Using Landsat Data. *Proceedings* **2017**, *1*, 17.
5. Acharya, T.D.; Subedi, A.; Yang, I.T.; Lee, D.H. Combining Water Indices for Water and Background Threshold in Landsat Image. *Proceedings* **2018**, *2*, 143.
6. Acharya, T.D.; Lee, D.H.; Yang, I.T.; Lee, J.K. Identification of Water Bodies in a Landsat 8 OLI Image Using a J48 Decision Tree. *Sensors* **2016**, *16*, 1075.
7. Acharya, T.D.; Subedi, A.; Lee, D.H. Evaluation of Water Indices for Surface Water Extraction in a Landsat 8 Scene of Nepal. *Sensors* **2018**, *8*, 2580, doi:10.3390/s18082580.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).