Proceedings

Classification of Surface Water using Machine Learning Methods from Landsat Data in Nepal †

Tri Dev Acharya [1], Anoj Subedi [2], Huang He [3] and Dong Ha Lee [1,*]

1 Department of Civil Engineering, Kangwon National University, Chuncheon 24341, Korea; tridevacharya@kangwon.ac.kr (T.D.A.)
2 Institute of Forestry, Pokhara Campus, Tribhuwan University, Pokhara 33700, Nepal; anojsubedi99@gmail.com (A.S.)
3 School of Geomatics and Urban Spatial Information, Beijing University of Civil Engineering and Architecture, Beijing, 100044 China; huanghe@bucea.edu.cn (H.H.)
* 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://sciforum.net/conference/ecsa-5.

Received: date; Accepted: date; Published: date

Abstract: With over 6,000 rivers and 5358 lakes, surface water is one of the important resources in Nepal. However, their quantity and quality are decreasing due to human activities and climate change. Hence, 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 vital role. Single or multiple water index methods has been applied in classification of surface water with satisfactory results. In recent years machine learning methods with training dataset, have been outperforming different traditional methods. In this study, we tried to use satellite image from Landsat 8 to classify surface water in Nepal. Input of Landsat bands and ground truth from high resolution images available form Google Earth will be used. And their performance will be evaluated based on overall accuracy. The study will be will helpful to select optimum machine learning method for surface water classification and therefore, monitoring and management of the surface water in Nepal.

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

1. Introduction

Nepal is a geographically diverse country with flats in south and increasing hills towards north to mighty Himalayas. In Nepal, around 70-90% of the total annual rainfall occurs during 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, changes in them is a difficult task to track. Also, their quantity and quality are decreasing due to human activities and climate change. Hence, 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 case, remote sensing satellite images are very well used. For the identification of surface water using Landsat image, we used various techniques in our previous studies, such as, using water index methods single or combined [4,5], decision tree based classification [6] and segmentation of scene [7] in diverse area of Nepal.

In recent years machine learning methods with training dataset, 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

For better comparison of the classification work, Landsat scene from 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 colour composite. A total of 800 ground truths with 614 non-water and 186 water points within the scene were extracted as from high resolution images available form Google Earth. These were used as training as well as later validation in the classification process.

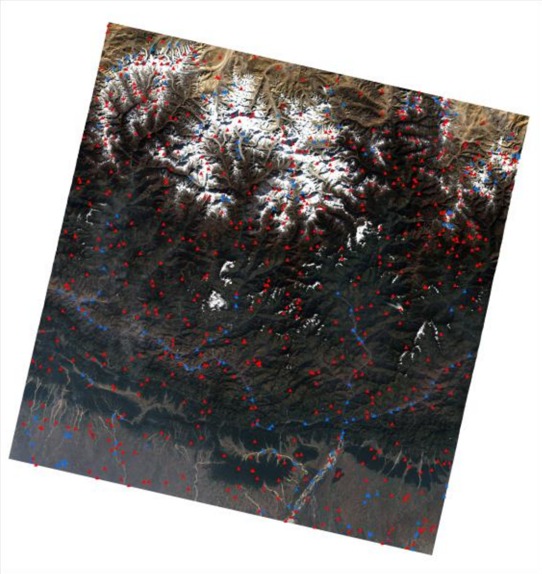

Figure 1. Landsat 8 scene in natural colour composite with training points (blue: water and red: non-water) acquired on acquired on 8 December 2017.

2.3. Method

After preprocessing to at-satellite reflectance, the image were first used to extract all OLI bands values for training dataset. After forming training dataset, four machine learning methods were used to train different models in R and later applied in 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 over-fitting.

A Recursive Partitioning (RPART) is a type of binary tree 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 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 has non-regularity which means, data whose distribution is unknown.

Neural Network (NNET) in R, is a feed-forward neural networks with a single hidden layer flowing left to right. Feedforward neural networks were the first type of artificial neural network invented and are simpler than their counterpart, recurrent neural networks. They are called feedforward 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, given dataset was reclassified to evaluate for overall accuracy.

3. Results and Discussion

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

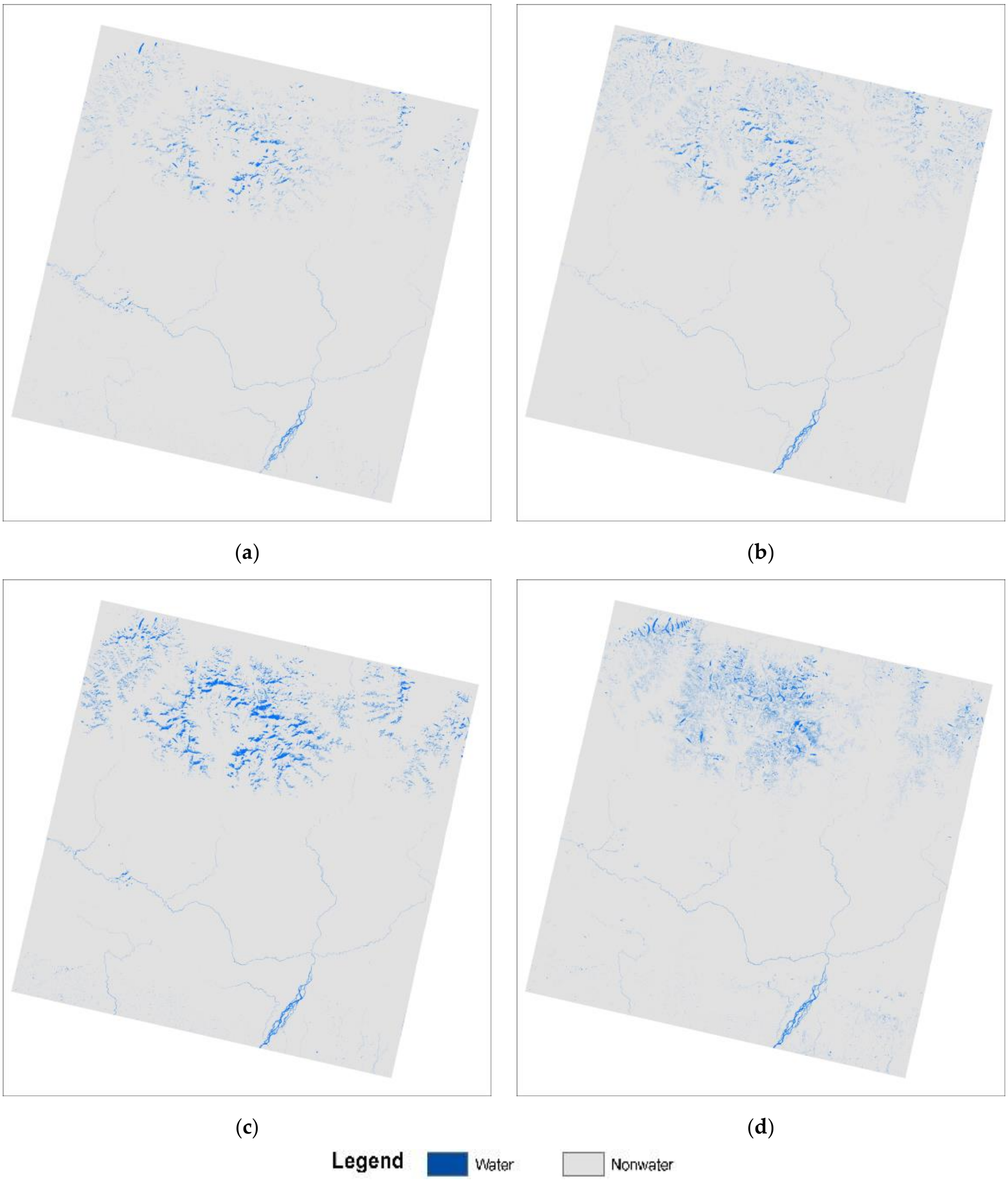

Figure 2. Surface water derived from Landsat 8 OLI image using various machine learning methods: (**a**) Random Forest; (**b**) Recursive Partitioning Tree; (**c**) Support Vector Machine; (**d**) Neural Network.

In Figure 2, we can see that the resulting water maps shows similarity in the lower lands for rivers whereas variation in the upper Himalaya regions. These variations seems to be mostly in cold icy water, hills shades and forest areas.

Table 1 shows the result of the overall accuracy from all the four methods, in which random forest performs with 1 as highest and SVM performs lowest with 0.926. Both RPART and NN showed overall accuracy of 0.95. Comparing the result of these methods with previous work [7] with same training points, there seems to be vast improvement against index methods single or combined. However, the segmentation accuracy is still higher i.e. 0.96 against machine learning methods except random forest.

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

S. No.	Machne 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, application of four machine learning methods: RF, RPART, SVM and NN were used to derive the surface water map using a Landsat 8 OLI image in Nepal. Using previously used training data and Landsat scene, surface water were modeled and applied. Result shows that the snow cold water in Himalayas with hill shadows caused difference in water detection among methods. In addition, random forest has shown maximum overall accuracy 1 for the scene with given dataset. It seems that machine learning methods could be very useful in automated binary classification of surface water accurately in complex geography of Nepal.

Conflicts of Interest: 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: Proceedings of the International Conference on Ecohydrology of High Mountain Areas, , ICIMOD, Kathmandu, Nepal, 24-28 March; S.R. Chalise, Ed.; ICIMOD: Kathmandu, Nepal, 1996; , pp. 49-61.
2. WECS Water Resources of Nepal in the Context of Climate Change. 2011.
3. Bhuju, U.R.; Khadka, M.; Neupane, P.K.; Adhikari, R. Lakes of Nepal: 5358-A Map Based Inventory. 2009, SP01.
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. Available online: http://www.mdpi.com/1424-8220/18/8/2580.