Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings

Volume 15 *Seoul, South Korea* Article 5

2015

CROWD-2-CLOUD – Remote Sensing Land Cover Verification With Crowd-Sourcing Data

Moataz Ahmed
School of Geography, University of Nottingham

Dai Huynh
School of Geography, University of Nottingham

Darshana Wickramasinghe
School of Geography, University of Nottingham

Tuong-Thuy Vu
School of Geography, University of Nottingham, Malaysia campus (UNMC)

Recommended Citation

Ahmed, Moataz; Huynh, Dai; Wickramasinghe, Darshana; and Vu, Tuong-Thuy (2015) "CROWD-2-CLOUD – Remote Sensing Land Cover Verification With Crowd-Sourcing Data," *Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings*: Vol. 15 , Article 5.
DOI: https://doi.org/10.7275/R5WQ020P
Available at: https://scholarworks.umass.edu/foss4g/vol15/iss1/5

PO-11

CROWD-2-CLOUD – REMOTE SENSING LAND COVER VERIFICATION WITH CROWD-SOURCING DATA

Moataz Ahmed[1],Dai Huynh[2],Darshana Wickramasinghe[3],Tuong Thuy Vu[4]

[1234]School of Geography, University of Nottingham
43500 Semenyih, Selangor Darul Ehsan, Malaysia
Email[1]: khgx4mad@nottingham.edu.my,[2]isxcdh@nottingham.ac.uk,
[3]khgx4dcs@nottingham.edu.my,[4]tuongthuy.Vu@nottingham.edu.my

ABSTRACT

Nowadays, advanced remote sensing technologies provide huge amount of Earth Observation (EO) data timely. Growing quickly in terms of size and structure, EO data require a new way of handling and processing as it is considered big data. Cloud-computing platform proved to be a reliable and scalable platform that suits various user demands in remote sensing data processing. To verify the ambiguity of information derived solely from remote sensing, ground data is vital. The only way to keep pace with big remote sensing data is to exploit the crowdsourced data, which has been recently proposed elsewhere. In this study, we developed a prototype of an integrated location based service on top of cloud computing platform to detect land cover features and engage the crowd of volunteers during training and verification process. Relying on open-source tools, the proposed system provides location-based data collection and satellite image classification. The prototype was tested over the rapid on-going landscape surrounding the University of Nottingham, Malaysia campus. More advanced functions will be developed and a full system will be deployed and tested in further study.

1. INTRODUCTION

Remote Sensing technologies have a wide scope of applications. Urban planning, agriculture, environmental management, or climate change, remote sensing products are the indispensable sources of information in our daily life activities (Robila 2006). In the past, lesser remote sensing data were available and limited in spatial, spectral and temporal resolutions. This limitation makes it difficult to analyze data and derive the proper information about the Earth surface (Ramapriya*n 2013*).

Until recently, the picture has changed. As the number of operating satellites increase periodically and each of which has a specific mission to achieve with different types of attached sensors. Furthermore, advanced sensor technologies generate higher spectral and spatial resolution images than before, which can contribute effectively in better understanding of the Earth as an integrated system. Different remote sensing datasets become more complicated in terms of structure and size. Therefore, bigger volume of data received from satellites, more storage and computing power is required to analyze these datasets(Christophe 2011).

High-performance computing system is obviously a solution and cloud-computing platform is the today system for big data. Cloud computing provides unlimited storage capacity and elastic computing services in form of Software As A Service, (SAAS) Platform

As A Service (PAAS) and Infrastructure As A Service (IAAS) models. Also, Cloud Computing can reduce the costs, as it's possible to pay as use and on demand (Xiaoqiang 2010), instead of purchasing advanced hardware and software licenses to process these big data. Moreover, the scalability options could make it easier to scale up or down the number of processors required to achieve EO data processing and analyzing.

Earth observation agencies are moving toward cloud computing platforms to host and deliver their remote sensing data like ESA Helix Nebula (Helix Nebula Online), NASA NEX (NASA NEX Project Online) and MODIS Azure (MODISAzure Online). The established system so far is merely a storage platform whereas few data processing services have been deployed to utilize the efficiency of cloud-computing platform. Various applications and user demands require different ways to derive the information from satellite EO data and it seems complicated to provide a general processing service. Let’s work on the best resolution available 30-m Landsat data set; the land-cover product seems a generally required product. Land-cover classification processing service can be developed, but a critical issues remains, validation. The challenge here is how to validate such big data? We cannot rely on a small team of experts conventionally, but a bigger group of people.

Recent advances in positioning technologies, location-based services, broadband Internet connections, and Web 2.0 allow every human to easily become an intelligent sensor and collect geospatial data accurately (Goodchild, 2007b). Goodchild also states that surprisingly among those six billion sensors exist thousands of volunteers who are willing to contribute, share, and update their collected geographic data. This new innovation of generating geographic information is termed *volunteered geographic information – VGI* (Goodchild, 2007a), or *crowdsourcing*. Thereafter, we have observed the rise of crowdsourcing and its applications in various areas including land cover (Comber *et al.*, 2013, Foody and Boyd, 2012, Fritz *et al.*, 2012).

This paper reports our first step of application development taking advantage of current advancements in remote sensing data evolution and the rise of using crowdsourced data and its applications. Here, we utilized cloud computing elastic architecture to build supervised classification service with crowd-based accuracy assessment model. The second section describes the development steps and the system architecture starting from data acquisition till the classification results, while the third section explains the initial results and and fourth section presents the further development and future work to be done.

2. Methodology

This paper proposes a cloud-based framework utilizing crowdsourcing data to validate ambiguity land cover information derived from remote sensing supervised Support Vector Machine (SVM) classification algorithm. For this purpose, we present an open-source mobile application, *Crowd-2-Cloud*, which is used to collect crowd-sourced land cover data. The crowd-sourced database later has two important roles in assisting the validation of remote sensing land cover data: (i) as a training sample set and (ii) as a validation data set of our proposed classification method.

2.1 Data Acquisition

Crowd-sourced volunteers can use the mobile app to capture landscape photos embedded with Geo-location data. They also can decide which land cover features dominate in the photo. Finally, submitted data from the crowd will be stored on a cloud based Postgresql server using Nottingham University private cloud. Figure (1) shows a screenshot of the mobile app main page.

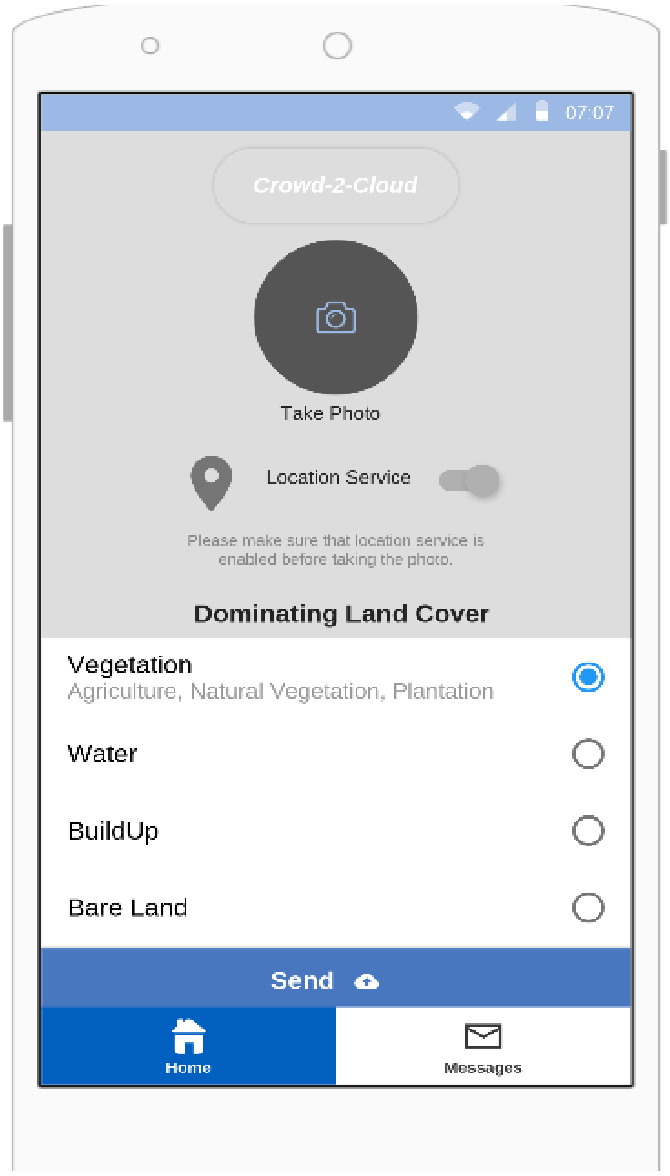

Figure 1. Screenshot of the mobile application

2.2 Processing

The overall system architecture consisting of all components is shown in Figure (2). The main components of the system are: 1) A mobile app to collect crowdsourced data from users (as shown in figure (1)). 2) A Postgresql Database Server on top of the cloud to receive the sent data from users. 3) A large computing node to run the supervised classification algorithm and publish results to be assessed using 10-fold cross-validation method.

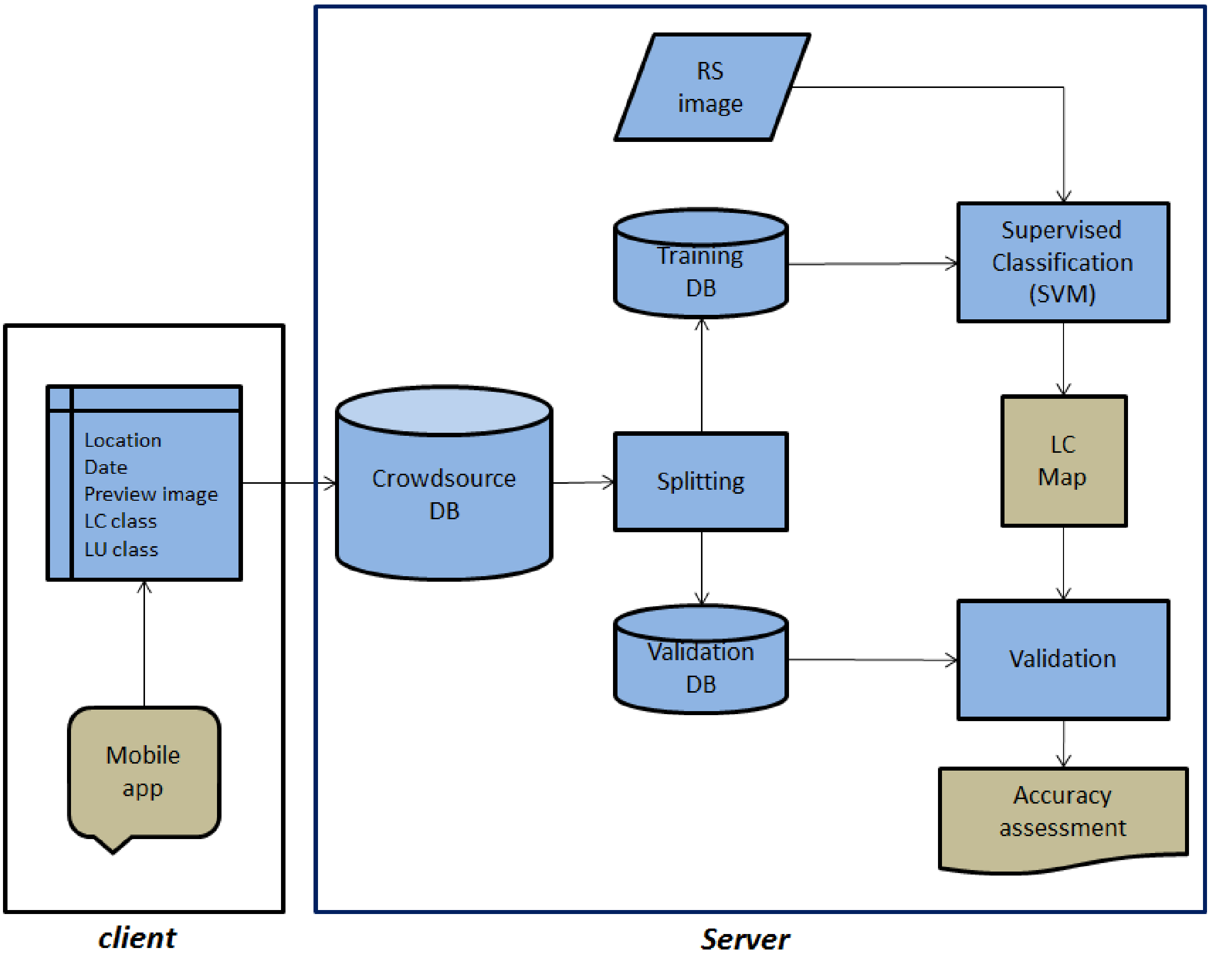

Figure 2: System Architecture and Data Flow Between Client and Server

2.2.1 Mobile Application

A simple Android-based mobile application is used for crowdsourced data collection from the users. The application's main page asks the user to take a photo and enables the location service to retrieve the user's current location. The app also asks the user to pick one choice as a dominating land cover class from list contains "Vegetation, Water, Built-up, Bare land" then the user can press "send" to upload the data to the server side which is a cloud-based Postgresql server.

2.2.2 Database Server

We set up a fully Open Source enabled database server configured with Postgresql 9.3 and Ubuntu 14.10 Server on top of Nottingham University private cloud infrastructure as shown in figure (3) from the dashboard of OpenNebula (OpenNebula Online). The aim from using Postgresql server is to store the crowdsourced data coming from the mobile app and then split it into two different databases. Using 10-fold cross-validation method (Ljumovi 2015), crowd-sourced data could be divided into training sample database and verification database.

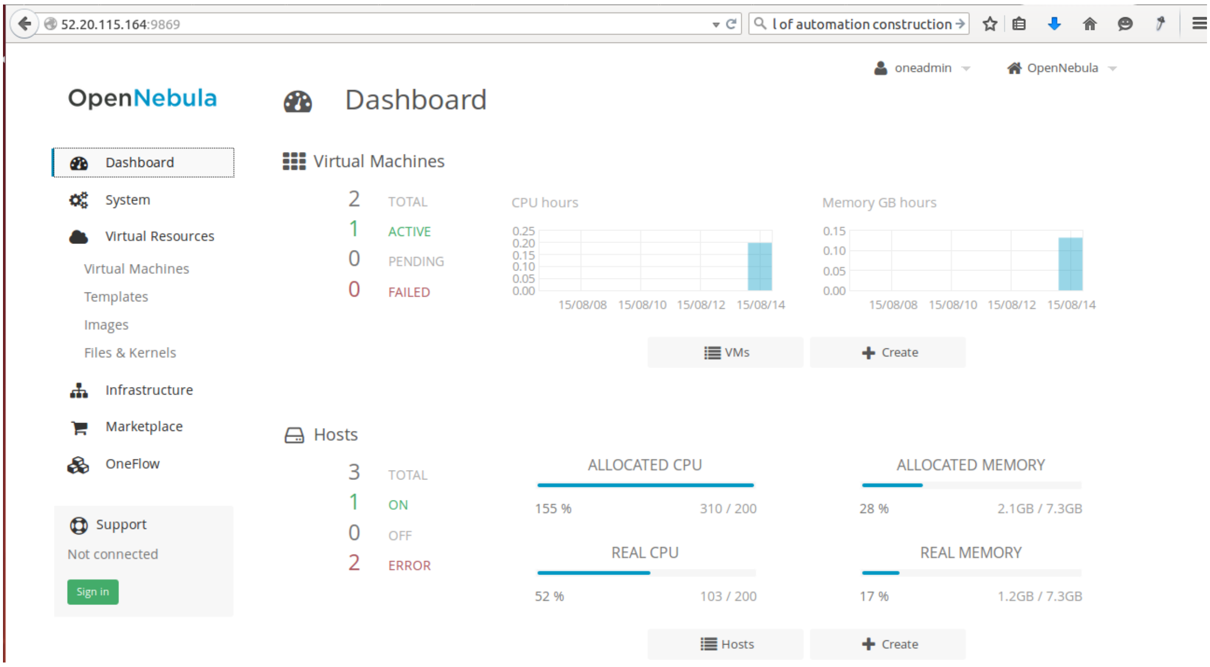

Figure 3: OpenNebula Dashboard for Nottingham University Private Cloud

2.2.3 Supervised Classification Algorithm

The proposed method use crowdsourced (CS) data for both model training and verification purposes. Hence initial CS data set split into two set randomly by using k-th cross validation method.

First set is considered as training samples. Pixel-based image classification methods are still applicable and well working on moderate resolution image like Landsat. The study adopted a proven SVM image classification algorithm for the image classification but it is extendable to any type of supervised learning methods. As a dynamic system, the collected training data are regularly updated and refined. The set of spectral signature library including temporal factor for a specific study area like Malaysia will be made available on our system. Though it is uneasy to have a big and properly distributed training data from begining but we assume that it would be sufficient in the long-run of the system operation thanks to the crowd contribution, In our initial development here for Malaysia study site, we proposed to classify the image into four dominant land cover classes: built-up, water, vegetation and bare-land.

The second set is validation data, used to validate the resulted classification maps acquired by SVM supervised learning (Liu 2014). The system extracts the classified land cover class of each validation points and calculates the confusion matrix by comparing classified and CS validation class. The accuracy assessment will be delivered in a form of a text document, containing overall accuracy, user and producer accuracy represented as percentage. The temporal factor needs to be taken into account in both training and validation via checking the time stamp of the crowdsourced photos and the satellite image acquisition date. Though it is not critical in tropical Malaysia but elsewhere.

3- Results

In this initial pilot version, we collected crowdsourced data by using the mobile application. The mobile application distributed over internally and gets the support of student and staff to collect the crowd sourcing data around Kuala Lumpur and surrounding city areas.

The resulted crowd source data contains around 60 locations with their dominating land cover description. Lesser number of samples selected as training points and used to train the SVM classification model and gets the classifier file. Figure 4 showing the training sample distribution over the testing area on top of Landsat8 image which going to be classified. Once the user selected the required satellite image, system automatically will run the SVM classifier and outputs the classified colored land cover classification image.

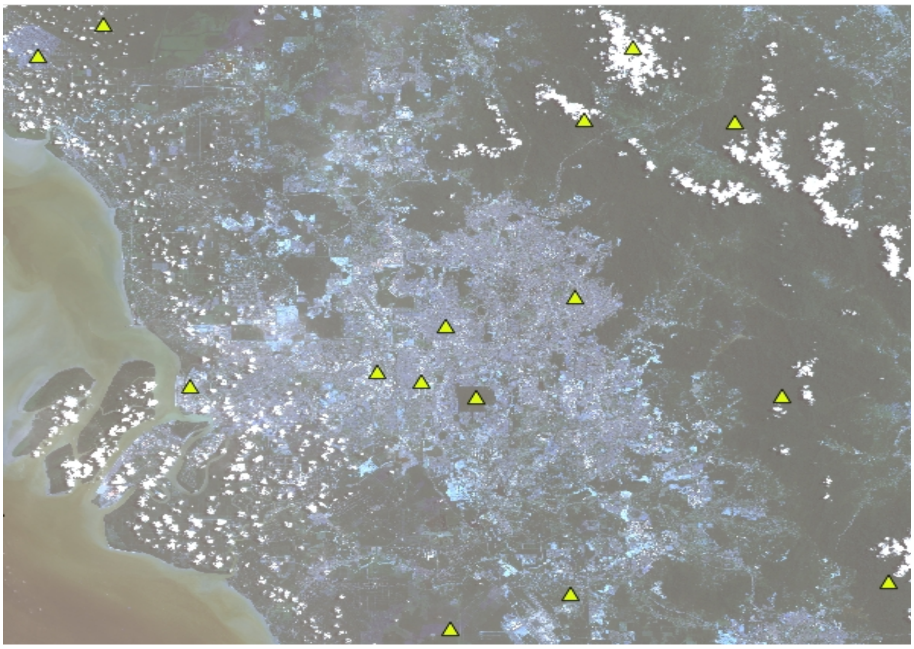

Figure 4. Training sample distributed over satellite image

Landsat 8 images used for classification with bands 2, 3, 4 and 5 only . Figure 5 shows the resulted SVM classification map with four dominant classes. There is few cloud classes in the satellite image and they are masked out using quality assessment band. Hence, cloud areas which are mixed with other land cover types have been minimized.

The accuracy assessment carried out using the rest of the CS data which contains 54 sample points for classification validation purpose. The system automatically extracts the corresponding land cover type in classified map and calculates the confusion matrix and Table 1 shows the resulted confusion matrix.

According to this result, the classification got 81.48% as an overall accuracy. Water classes accurately identified while Vegetation class also got the highest accuracy level however in some places it is mixed with bare-land and built-up areas. The built-up and bare-land areas also mixed with each other classes and got the lesser accuracy.

4- Conclusion

The development of land cover database met new challenge when Earth Observation (EO) data accelerate both in size and structure thanks to advanced remote sensing technologies. Although a deeper understanding of the Earth surface is made possible with a growing number of satellites in operation and higher spectral and spatial resolution images, such large volume of data require sufficient storage and computing power to handle. Cloud computing arises as a solution due to its advantages and reliability. This paper addressed the usage of cloud computing in detecting land cover features and data verification through this system architecture.

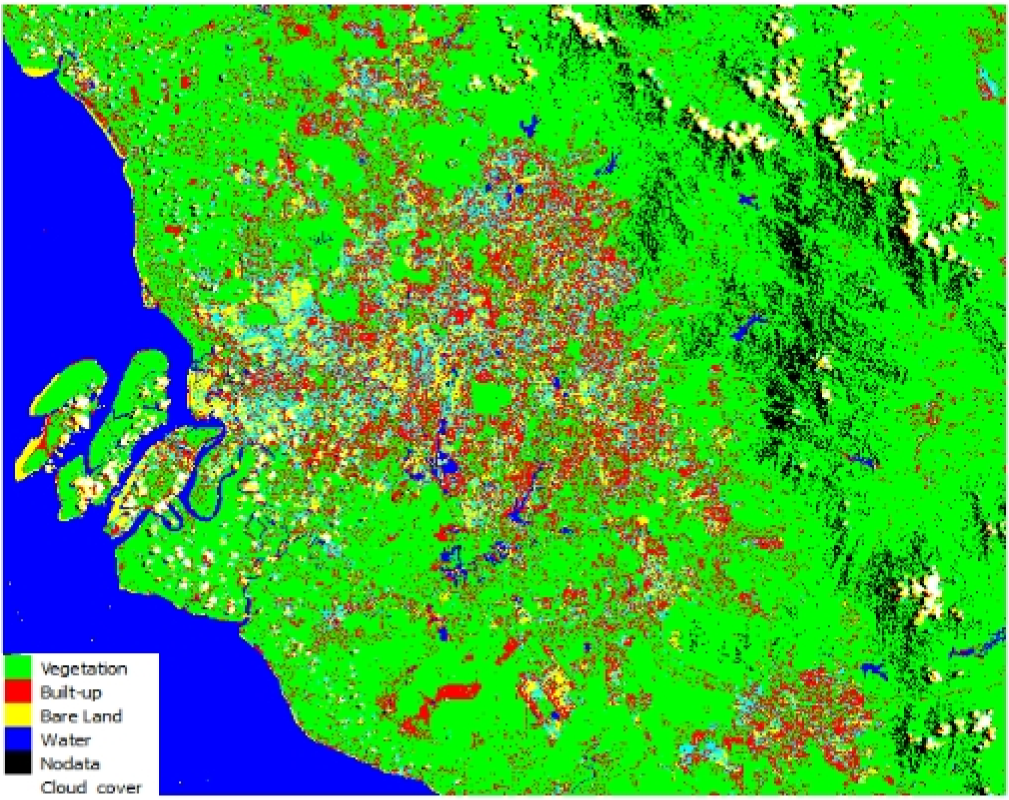

Figure 5. classified satellite image with four land cover classes

A mobile application is designed to collect data, and send it to the cloud based server for analysis which is supported by collected crowdsourcing data, machine-learning algorithms to identify land cover features from Landsat8 image. At this initial stage, implementation and trial test applied to the areas around Kuala Lumpur, Malaysia and future expansion will be reported in later research.

5- Future Work

This paper presented the initial steps of development which have been taken to provide land cover detection using Landsat 8 data sets and verification using crowd source data on top of cloud computing resources. Future work will include modifications to let the system work in 2 ways communication to let the users get notifications on their mobile application once a nearby spot detected to be changed from type of land cover to another. furthermore, we will provide a comparison study between using cloud computing and shared hosting solutions for hosting the service.

Table 1: Confusion Matrix

LC Class		Classified Image				Total	Producer's Accuracy
		Vegetation	Water	Built-up	Bare-land		
Crowd source data	Vegetation	13	0	1	1	15	86.66667
	Water	0	7	0	0	7	100
	Built-up	2	0	16	3	21	76.19048
	Barren	1	0	2	8	11	72.72727
Total		16	7	19	12	54	
User's Accuracy		81.25	100	84.21053	66.66667		**81.48148**

6- References

Comber, A., SEE, L., Fritz, S., Velde, M. V. D., Perger, C. & Foody, G. 2013 Using control data to determine the reliability of volunteered geographic information about land cover. *International Journal of Applied Earth Observation and Geoinformation,* 23**,** 37-48.

Foody, G. M. & Boyd, D. S. (2012) Exploring the potential role of volunteer citizen sensors in land cover map accuracy assessment. *Proceeding of the 10th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*.

Fritz, S., Mccallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., Velde, M. V. D., Kraxner, F. & Obersteiner, M. (2012) Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software,* 31**,** 110-123.

Goodchild, M. F. 2007a Citizens as Sensors: The World of Volunteered Geography. *GeoJournal,* 69**,** 211-221.

Goodchild, M. F. 2007b Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *International Journal of Spatial Data Infrastructures Research,* 2**,** 24-32.

Jackson, M. J., Rahemtulla, H. A. & Morley, J. 2010 The Synergistic Use of Authenticated and Crowd-sourced Data for Emergency Response. *The 2nd international workshop on validation of geoinformation products for crisis management (Valgeo)* 91-99.

S. A. Robila, "Use of Remote Sensing Applications and its Implications to the Society," in *2006 IEEE International Symposium on Technology and Society*, 2006, pp. 1–6.

H. Ramapriya*n* "Managing Big Data « Earth Imaging Journal: Remote Sensing, Satellite Images, Satellite Imagery." [Online]. Available: http://eijournal.com/print/articles/managing-big-data. [Accessed: 13-July-2015].

Christophe E, Michel J, Inglada J. Remote sensing processing: From multicore to GPU. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2011;4(3):643-652. doi:10.1109/JSTARS.2010.2102340.

Y. Xiaoqiang and D. Yuejin, "Exploration of cloud computing technologies for geographic information services," *2010 18th Int. Conf. Geoinformatics*, pp. 1–5, Jun. 2010.

"ESA SSEP Use Case | Helix Nebula." [Online]. Available: http://www.helix-nebula.eu/usecases/esa-ssep-use-case. [Accessed: 17-July-2015].

"NASA NEX." [Online]. Available: http://aws.amazon.com/nasa/nex/. [Accessed: 23-July-2015].

"MODISAzure - Microsoft Research." [Online]. Available: http://research.microsoft.com/en-us/projects/modisazure/. [Accessed: 12-Aug-2015].

OpenNebula | Flexible Enterprise Cloud Made Simple. Available at: http://opennebula.org/. Accessed August 15, 2015.

Ljumovi M, Gmbh RB. Estimating Expected Error Rates of Random Forest Classifiers : A Comparison of Cross-Validation and Bootstrap. 2015:212-215.

Liu Q, Guo Y, Liu G, Zhao J. Classification of Landsat 8 OLI image using support vector machine with Tasseled Cap Transformation. In: *2014 10th International Conference on Natural Computation (ICNC)*. IEEE; 2014:665-669. doi:10.1109/ICNC.2014.6975915.