ACRS 2022
MONGOLIA

CHANGE DETECTION USING OPTICAL DATA IMAGE FOR TORRENTIAL RAIN DISASTER IN DILI, TIMOR-LESTE

Pedro Junior Fernandes[1], Masahiko Nagai[2]

[1]Graduate School of Sciences and Technology for Innovation, Yamaguchi University,
2-16-1, Tokiwadai, Ube, Yamaguchi 755-8611, Japan
Email: pedrojunior160295@gmail.com
[2]Center for Research and Application of Satellite Remote Sensing, Yamaguchi University,
2-16-1, Tokiwadai, Ube, Yamaguchi 755-8611, Japan
Email: nagaim@yamaguchi-u.ac.jp

KEY WORDS: natural hazards, remote sensing, supervised classifier, random forest,

ABSTRACT: In April 2021, Timor-Leste was hit by Seroja cyclone which caused massive rainfall and resulted in massive flooding and landslides. As a new country, Timor-Leste is more vulnerable to such disaster. Natural hazards have become disasters that cause many casualties and cause huge losses in various regions and places, destroying people's homes, infrastructure, agricultural crops, etc. The objective of this study is to detect and analyze the changes that occur during natural events in Dili, Timor-Leste, based on satellite remote sensing. Remote sensing has an important role to play in terms of monitoring, observing, and analyzing earth conditions. In this paper, remote sensing optical imagery is used to perform supervised classification for pre-disaster, post-disaster and recovery phases to look at changes during these periods using Planet Image. In addition, this change detection approach uses stacked images from two different periods to extract changes in the images. In the case of image classification, since the area of interest is already known how the ground conditions are, supervised classification is a reliable approach and it is essential to the development of the training data and digitize polygons within the area with several land cover classes. The training area is then used to train the algorithm. The random forest algorithm is applied to this multi-imagery classification because it is an ensemble model that combines results from different models and logically the results will be better than a single model. Then, with the training data developed, it is then used to feed the algorithm with a focus on obtaining classification results. Based on the classification results, the changes are automatically classified along with their respective classes. Moreover, post-processing steps still need to be considered to obtain more accurate change detection results. This study will provide valuable information that can be used to develop a sustainable disaster management policy in Timor-Leste.

1. INTRODUCTION

Timor-Leste is one of the youngest countries in Southeast Asia, declaring its independence on May 20, 2022. With 1.3 million people and an area of 14,874 km2, it occupies half of Timor Island and the only Asian countries located entirely in the Southern Hemisphere. The country is also small and mountainous, making it prone to natural disasters. On April 4, 2021 Timor-Leste was inundated by flash floods, washing away houses along the Comoro river and causing thousands of people to flee their homes in search of safety. The heavy rains had already caused problems in several municipalities across the country and destroyed houses and other affected infrastructure, including roads and bridges. There are several natural hazards that often occur in the area such as heavy rains, tropical cyclone, droughts, flash floods, landslides. These natural hazards will continue to pose an ongoing threat to communities, and will also damage basic infrastructure, agriculture and livestock due to its geographical location.

Flooding is one of the most frequent and destructive natural hazards, which often cause property and life loss (Huang et al. 2020). In recent decades, remote sensing has played an important role in dealing with natural disaster management. Detecting change in satellite imagery generally refers to the conversion of the landscape from one dominant feature type to another. The type of information you can get from satellite imagery is when and where has change taken place, how much and what king of change has occurred, and what are the cycles and trends in the change.

The random forest classifier consists of a combination of tree classifiers where each classifier is generated using a random vector sampled independently from the input vector, and each tree casts a unit vote for the most popular class to classify an input vector (Breiman, 1999). The random forest algorithm is an example of an ensemble model which means that it combines the results from multiple models and the logic for that is that the results for the combination would be better than from a single model. It incorporates supervised learning component which means it gets better as it learns.

The main objective of this research is to analyze land cover changes during disasters using satellite imagery in order to obtain information related to the disaster, which will then be used by decision makers as a consideration in making policies in dealing with the situation.

2. STUDY AREA AND DATA USED

2.1 Focus area

This study focuses on Timor-Leste, specifically Dili and specifically Dili and some eastern parts of Liquiça municipality. Dili is the capital of the country, the most populous city, with over 230,000 people currently living in the city. Dili has a small area of plains and is also surrounded by hills and mountains. The city also has several large rivers that flow through the city, making it a disaster-prone area. On April 4, 2021, Dili was the city most affected by the devastating floods as a result of the large number of people who moved in search of a safe place to stay. There was a total of 44 deaths, including 20 in Dili, and over 80% of the affected households were in the Dili municipality.

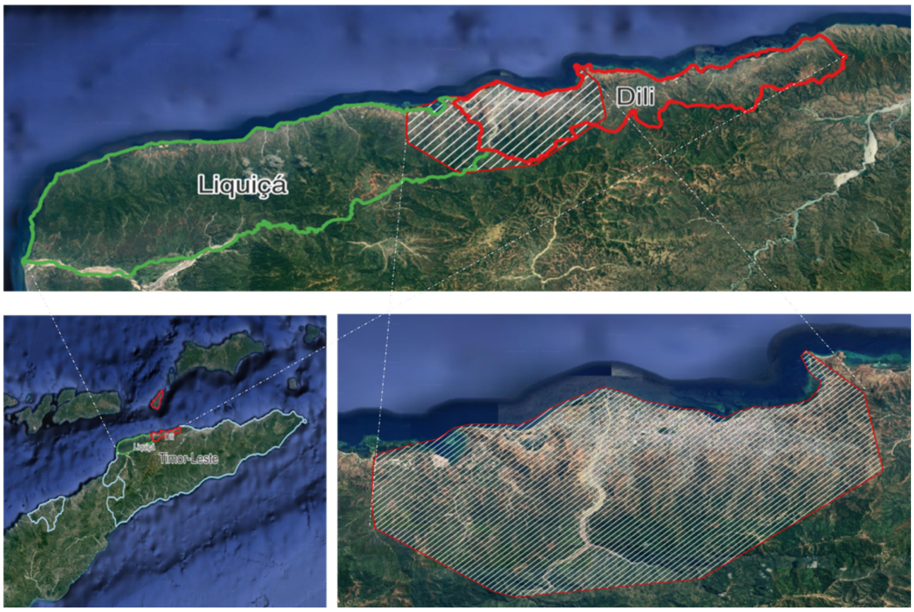


Figure 1: Study area (Dili City and Liquiça Tibar, Timor-Leste)

2.2 Data used

This research has used Planet Scope for the before and after images. Planetary imagery consists of many micro-satellites operating in open space. It has high spatial resolution, high frequency as well as high coverage rate. It provides 4 bands in surface reflectance.

Table 1. Information of two Planet images.

Product	Date	Spatial Resolution	Bands
4233514_5135126	2021-Mar06	3m	Red, Green, Blue, Infrared
4355850_5135126	2021-Apr09	3m	Red, Green, Blue, Infrared

3. METHODOLOGY

The main idea of this methodology is to use the two-date supervised image classification and change detection approach to delineate land cover change. This approach uses the analyst's knowledge to generate representative areas of known vegetation or land cover types. These are referred to as training data and are then used to train the classification algorithm. The algorithm then uses this known information to identify and label areas that are similar to the training data.
The detailed preprocessing step is pictured below and the flowchart of the entire processing chain in fig...

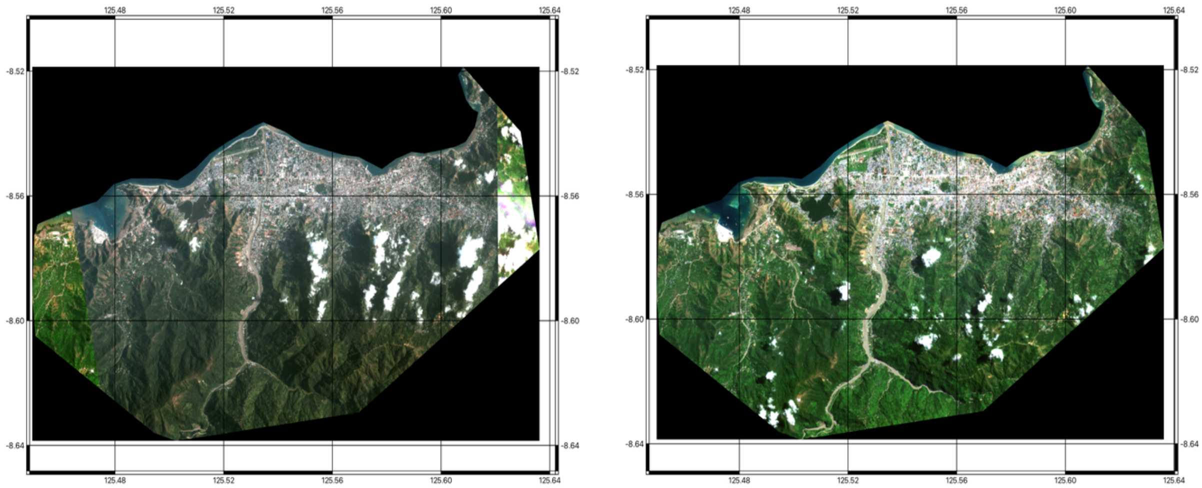
(1) Importing images and creating two-date image stacks;
(2) Image enhancement, stretching, and exploiting different band combinations;
(3) Development of training data sites;
(4) Use R Studio to run Random Forest Classification;

3.1 Band Stacking of two dates

This step includes taking an image of two dates that is Planet Image with 4 bands (Red, Green Blue and NIR) and then stacking them into a combined image of both dates with all bands. Thus, we will have a new 8-band image where bands 1 to 4 are the before image and bands 5 to 8 are the after image.

3.2 Image enhancement and band combination

Performing various image enhancements interprets the image and essentially highlights the various features present. We stretch the image by adjusting the range of minimum and maximum pixel values used and using the standard deviation of the pixels. To highlight the usefulness of different band combinations, we use several band combination approaches. The first is true color band combinations is to capture how our eyes perceive colors, whereas the second is false color band combinations or color composites we use to highlight vegetation by exploiting the different reflectance properties of the vegetation. So, in this infrared color band combination we will see the reflectance of healthy vegetation increase dramatically in the near infrared.



(1) (2)

Figure 2: (1) and (2) are the original color images for before and after, respectively.

3.2.1 Multi-date image enhancement

Instead of 4 bands in the one image, we are having the 8 bands image. The advantage of the multi-date image stack image is that we can highlight changes between the two dates. We use this to develop training data sites that's intern to be used to train the algorithm.

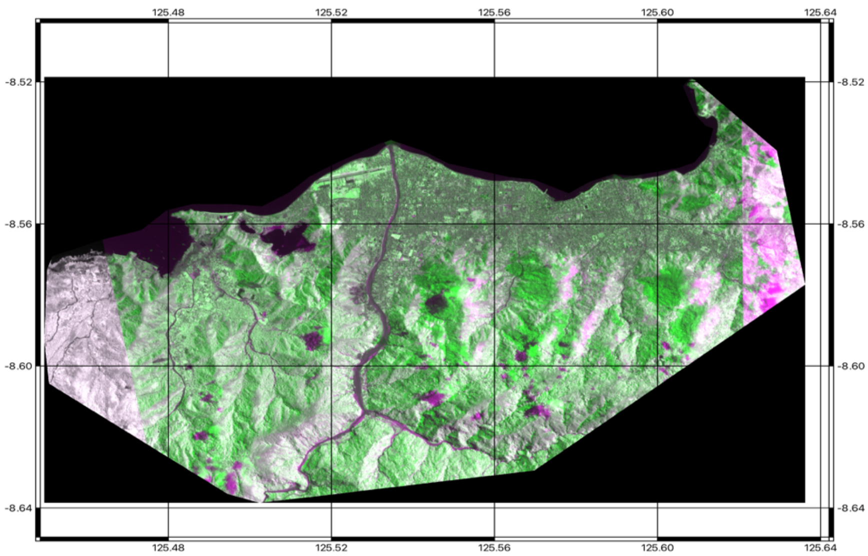


Figure 3: Two-stacked images band combination

3.3 Change land cover classes

It's important to consider when dealing with the collection of training data sites to ensure the consistency in the naming and labeling used. In a single date classification, we would normally use a single digit for the land cover class code, for example, number 1 is forest, number 2 is non-forest, number 3 is water, and so on. In the two-date classification, we are interested in capturing the transition changes that have occurred between the dates, therefore, the codes for the land cover classes are code 11 for forest-to-forest or stable forest where one is for the first date and the other is for the second date, code 12 will be used to capture the change from forest on the first date to non-forest on the second date, and code 21 will be used to capture non-forest on the first date and forest on the second date, and so on.

Table 2. Land cover classes for two-dates classification

Land cover classes	Code
Vegetation - Vegetation	11
Vegetation - Non-Vegetation	12
Non-vegetation - Vegetation	21
Non-Vegetation – Non-Vegetation	22
Water - Water	33
Non-Water - Water	43
Urban area	55
Bare ground	66
Shadow	77

3.4 Training sites

The supervised classification approach, we define the training data based on their knowledge of what is on the ground and then they generate the training sites which can be done by digitizing a polygon using QGIS software. This is all based on exploiting the reflectance properties of different materials whether that is water, bare ground, or vegetation. Each material has a spectral signature in which we can exploit when we are classifying imagery. This information is then able to build training data set of mean spectral signatures for the various classes of interest and then the classification algorithm uses this information to match unknown pixels with the mean spectral signatures of the classes of interest and identify the best match of class for that pixel.

3.4.1 Training data preparation

The training data includes ensuring the training sites are distributed across the image, ensuring that both changing and unchanging areas are captured, and the size of the training data is largely driven by the spectral characteristics of the image. In general, large training locations make sense for simple images, conversely smaller training sites are required for spectrally complex images.

3.4.2 Training sites for two-date classification

As we use a decision tree classification algorithm which is a random forest algorithm, the location of the training data does not have to be homogeneous meaning that when collecting training data, it is possible to mix forest types and non-forest land cover types. However, it is important to capture the spectral range across these classes.

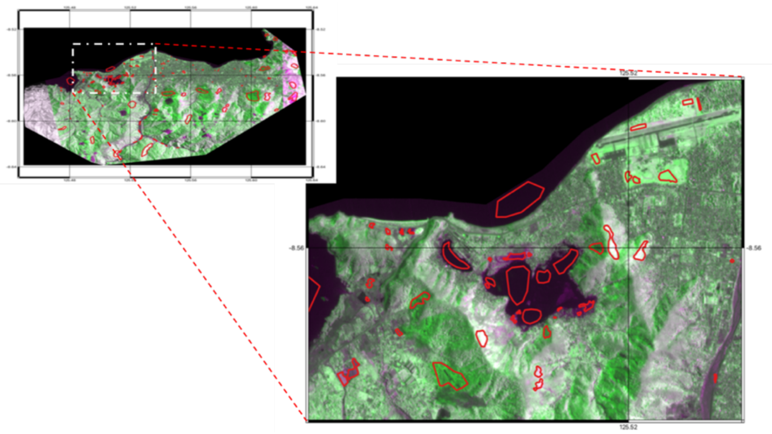


Figure 4: Training data for two-dates images

3.5 Random Forest algorithm

It basically uses two steps that take a random set of training sites usually 2/3 and build some decision trees or classification trees, and the remaining 1/3 will be used to estimate the error as well as highlight the importance of each predictor variable. The trees developed have branches and these are called nodes as well as leaves called class labels. The random component for each decision regards this as a coin toss and there is a continuous learning component. The pixels are assigned to classes based on majority rule, so each decision tree votes on which class the pixel belongs to. In this study using R to run the algorithm.

4. RESULTS AND DISCUSSION

4.1 Results

This result is still a preliminary result for the random forest classification, it also shows that the satellite image data has been classified based on the predefined land cover classes. In the preliminary results of this change map, we can see quite a lot of green areas referring to stable forests, some yellow which means before there was vegetation and the vegetation was lost after the disaster, as well as changes in other classes.

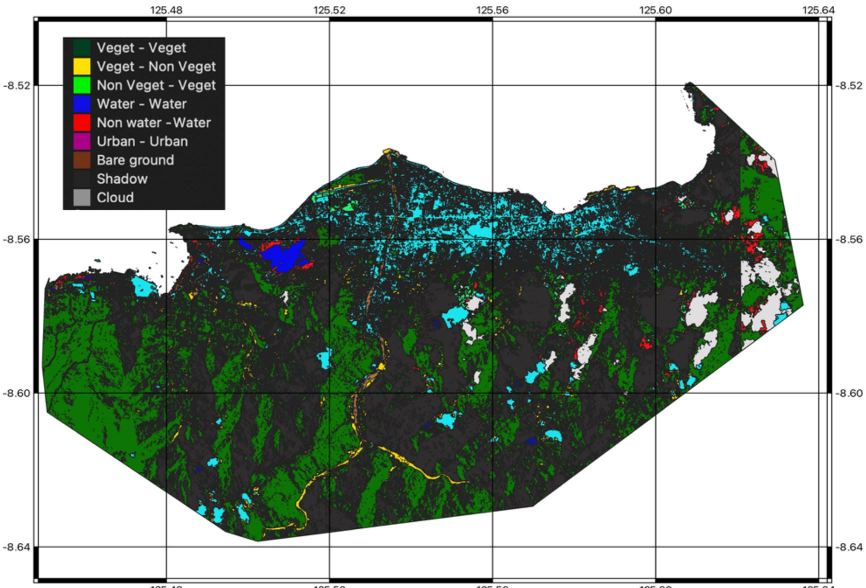
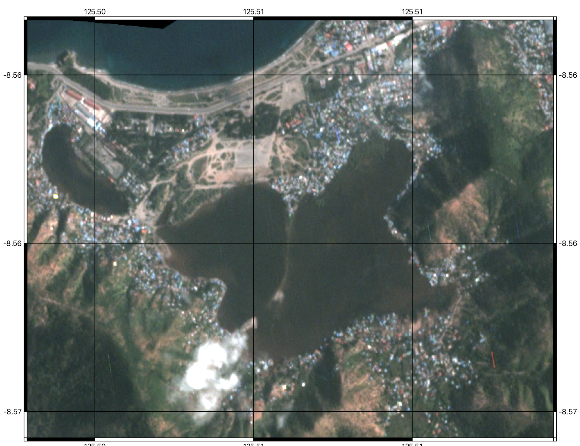
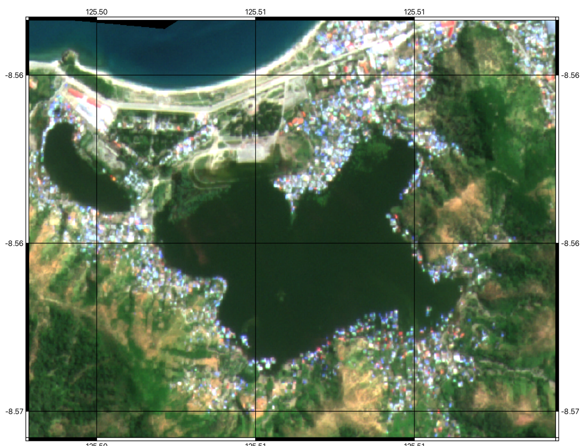


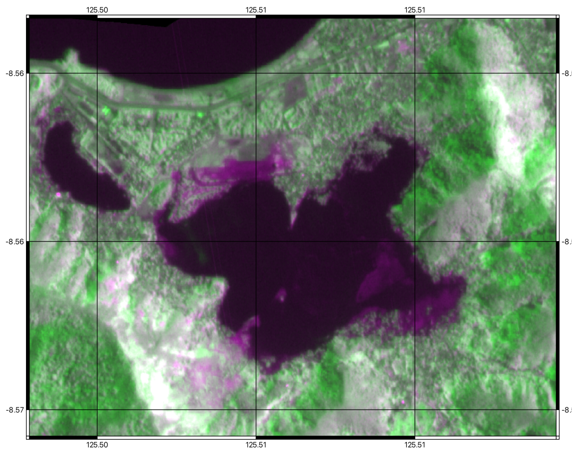
Figure 5: Preliminary results of Random Forest Classification

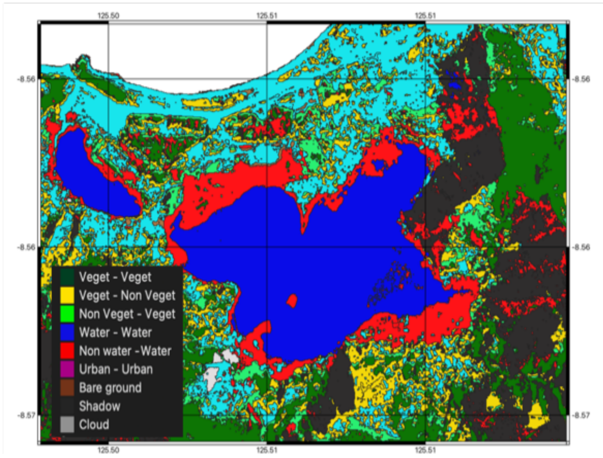


(1)



(2)


(3)



(4)

Figure 6: (1) Fist date image, (2) Second date image, (3) Band combination, (4) Classification results

In figure 6, it represents the difference between four images that we can interpret visually. Image number 3 highlights the changes between the two different images in purple. In the bottom right corner of the image, the classification results show a significant expansion of water in Lake Tasi-Tolu, Dili.

4.2 Discussion

As an interim result of this study, we can see on the classification result map, there are small areas of classification results that appear within a large area. While the image has more pixel variability which causes some misclassification or errors. Under this circumstance, we need to review the results, check the accuracy of the map, need to change or update the training data site as needed, re-run the random forest algorithm with new training data.

5. CONCLUSION

This is a preliminary result of the research that still needs to be continued to conduct a more in-depth analysis using different approaches and calculations to obtain better accuracy and also ensure that the results obtained are feasible to use. On the other hand, the training data should be distributed throughout the image as it will be used to feed the classification algorithm. Post-processing steps need to be performed after performing the analysis. Finally, an accuracy assessment needs to be done to get better classification results.

6. REFERENCES

BREIMAN, L., 1999, Random forests—random features. Technical Report 567, Statistics Department, University of California, Berkeley, ftp://ftp.stat.berkeley.edu/pub/users/breiman.

Huang, M., Jin, S. Rapid Flood Mapping and Evaluation with a supervised Classifier and Change detection in Shouguang Using Sentinel-1 SAR and Sentinel-2 Optical data. MDPI, 2020, pp 1-5.

Timor-Leste: Floods, UN Resident Coordinator's Office (RCO), Situation Report No.9 May 21, 2021 https://timorleste.un.org/en/127366-timor-leste-floods-un-situation-report-no-9-21-may-2021.

TIMOR-LESTE Disaster Management Reference Handbook, October 2019, https://www.preventionweb.net/publication/timor-leste-disaster-management-reference-handbook-2019.