



Predicting of Flooding in the Mekong Delta Using Satellite Images

Hiep Xuan Huynh¹, Tran Tu Thi Loi², Toan Phung Huynh^{1(✉)},
Son Van Tran³, Thu Ngoc Thi Nguyen³, and Simona Niculescu⁴

¹ Cantho University, Cantho 94000, Vietnam
{hxhiep, hptuan}@ctu.edu.vn

² Fsoft Cantho, Can Tho, Vietnam
tutran5985@gmail.com

³ Kiengiang Medical College, Rạch Giá 91000, Kiengiang, Vietnam
tvsonkl3@gmail.com, ntngocthu@gmail.com

⁴ Université de Bretagne Occidentale - UBO, Brest, France
simona.niculescu@univ-brest.fr

Abstract. Flooding is a natural risk, large floods have occurred almost every year. These are major issues that researchers are interested and to identify flooded areas or assess the risk of flooding, the researchers using image LiDAR or image RADAR to flood mapping, flood risk management, observation and change detection in floodable area. However, flood modeling or flood assessment don't solve the problem of flood risks. Therefore, in this paper we propose a new approach of processing methodology based on time series analysis that enables predicting of the floodable areas in the Mekong Delta using new satellite images such as Landsat 7 ETM+, Landsat 8 OLI and sentinel-2 MSI.

Keywords: Assess the risk of flooding · Satellite image · Modeling of image classification · Time series analysis · Random forest · Decision trees · Determining and predicting the flooding area

1 Introduction

Flood inundation is the natural disaster, damage to human life and agriculture [13]. In the near years, the flood inundation changes due to the effect of climate change. Therefore, we need the best solutions for monitoring and management to forecast flooding capacity and also to limit obstacles to people and productions.

Nowaday, most of the researchers have mapped the floods [25] by using normalized difference water index are used to separate water and soil [8]. Supervised learning methods and classification algorithms such as decision trees [10], support vector machine [5], random forest [11] applied to the remote sensing images have successfully implemented image classification showing that the uniform pixels are grouped in layers.

Remote sensing technology such as radar remote sensing sensor (SAR) [20] seems to be one of the fastest and most effective ways to observe and provide information on flooding levels but it is expensive and not available for public use. Meanwhile Landsat

images (Landsat 7 ETM+, Landsat 8 OLI) and optical images (Sentinel-2 MSI) are provided free of charge by NASA, is one of the scientific achievements that is widely used for all fields because it has high time resolution and rich spectral information. Landsat's images known in flood mapping [15], followed and monitored major flood by repeating of period cycle. It is capable of quickly assessing the change of objects [15] in monitoring the situation of floods events [21], and previously flooded areas could be identified based on spectral attribute changes between before floods and after floods [24]. Besides, Sentinel-2 optical is a new approach for improving the area to be analyzed such as the mapping of restored areas relevant for the classification of wetland vegetation [22].

The aim of this paper uses new satellite images such as Landsat 7 ETM+, Landsat 8 OLI and Sentinel-2 MSI [3] to analyze the objects of the images based on normalized difference water index (NDWI) [19], and establish region of interest (ROI) [18]. Combination with supervised classification algorithm, random forest [11] classifies and generalises for all images. The classification results are accurate with overall accuracy from 90% to 100% in before floods, during floods and after floods periods in the Mekong Delta with new satellite images.

This paper is organized as follows. The first section is introduction. The second presents about regions of interest for flooding. The third and fourth sections introduce the discovering floods area and determine flooding disposition over time. The fifth section presents results experiment. And the end is conclusion and future work.

2 Regions of Interest for Flooding

2.1 Image Characteristics

Digital image is a two-dimensional matrix that is displayed in the form of small squares of equal size, each square has a value (numeric type) representing a color. There are two types of images that we observe are grayscale images (each pixel counts from 0 to 255, which means that there are 8 bits or 1 byte to represent each of these pixels) and color image, a combination of three colors such as green, blue, red. Each color is between 0 and 255, meaning that each pixel needs 24 bits or 3 bytes to be represented as illustrated in Fig. 1.

The spectral characteristics of objects such as soil, clear water, turbid water and vegetation are described as follows:

- Clear river water absorbs a lot and is less reflective, so its color is very dark on the image.
- Turbid water is more reflective than the clear water because its reflexivity depends on the ability of the objects in the water to reflect (such as silt or algae)
- Soil is very reflective and the ability to reflect depends on the length of the wavelength
- Vegetation are reflected at 0.54 μm wavelength and infrared. The spectral reflectance of vegetation in the infrared is many times larger than the visible light.

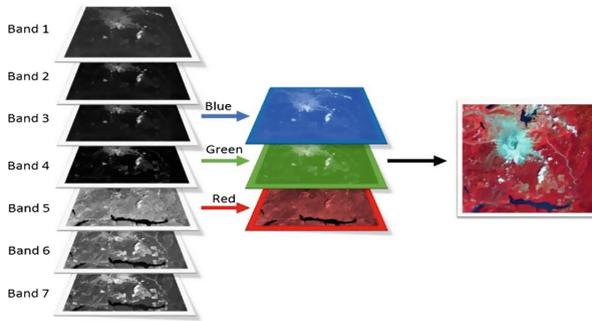


Fig. 1. Color image with combination of band red, green, blue (Color figure online)

Pixel images are square and display a certain area on an image as illustrated in Fig. 2. It is important to distinguish between pixel size and spatial resolution, they can not be interchangeable. If a sensor has a spatial resolution of 20 m and an image from that sensor is displayed at full resolution, each pixel represents an area of 20 m × 20 m on the ground. In this case the pixel size and resolution are the same. However, it is possible to display images of pixel size other than resolution. Landsat image have 30 m × 30 m resolution.

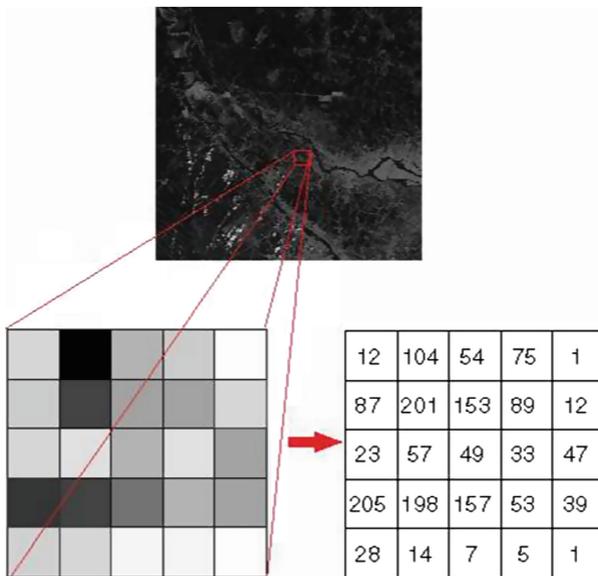


Fig. 2. Number of regions (Source: CCRS/CCT)

2.2 Creating Regions of Interest

Region of interest (ROI) [18] is a subset of an image or a set of data is determined for a particular purpose. ROI is used to determine the boundary of an object to be considered. In Fig. 3, the center pixel is used as seed (figure a) for the development area of spectral (figure b) with spectral distance parameter = 0, 1; Similar pixels are selected to create the training area (figure c và figure d). For example, sea water (blue) different soil (grey), river water (turbid water) different sea water (clearer water). Floods often involve rivers, lakes, flooded rice fields or urban areas.

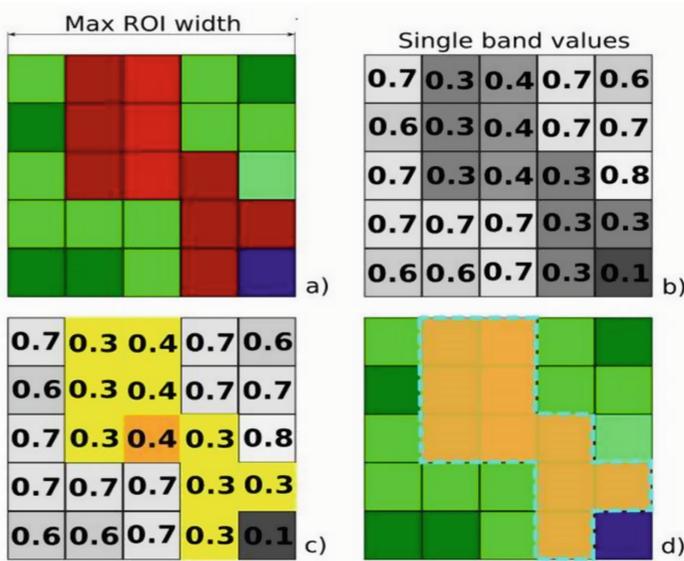


Fig. 3. Region of interest (Source: semi automatic classification manual) (Color figure online)

3 Identity Flood Areas

3.1 Features Extraction in Combined Satellite Images

The mathematical approach to identifying which image bands relate to merge spectral bands into one final image from which the needed information can be obtained and to determine of flood inundation area [16]. With image in Fig. 1 show different multi-band combinations such as water, urban, vegetation, unclassification.

With Blue-Green-Red, this is natural colors show, vegetation is brown and yellow, shorelines and clouds appear white and very difficult to distinguish. Water is white gray, not as easily detected to sparsely vegetated areas and urban.

With Blue-NIR-SWIR1, the vegetation is light green, not clear. The city is bright purple. Clouds are white with a red border, the shadow is black. Red water is easy to mistake with clouds.

With Green-NIR-SWIR2, the vegetation and grasslands is green, the soil is dark brown and urban areas appear in varying shades of purple. Clouds are white with yellow borders, the shadow is black. The water is red brown, shallow water or flooded on rice fields will appear as green yellow.

The multi-band combination [16] will show the object such as water, urban, vegetation, cloud, shadow. Therefore, the merge images is necessary for each request as illustrated in Fig. 4.

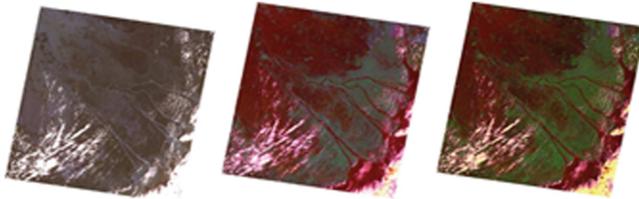


Fig. 4. Various features of the area with different band combinations

3.2 Object Extraction with Normalized Difference Water Index

The Normalized difference water index (NDWI) [1] that is shown in Fig. 5 is calculated from Near-Infrared (NIR) and Short Wave Infrared (SWIR) channels. The NDWI allows the separation of two soil and water objects [4], [23] in satellite images. In addition, the NDWI [19] of McFeeters asserted that values of NDWI smaller or equal than to zero are assumed to non-water surfaces, while values greater than zero are assumed to be water surfaces.

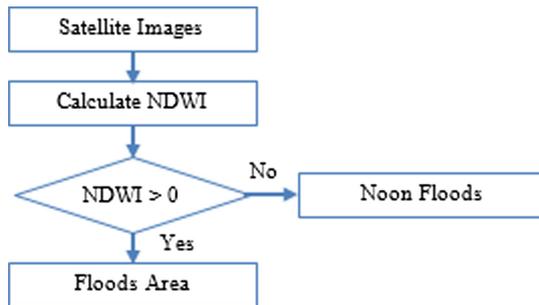


Fig. 5. Flowchart calculate NDWI

The NDWI is calculated using Eq. (1):

$$\text{NDWI} = (\text{Green} - \text{NIR}) / (\text{Green} + \text{NIR}) \quad (1)$$

Below image (Fig. 6) is the satellite image calculated NDWI with the threshold [1, 1] [2].

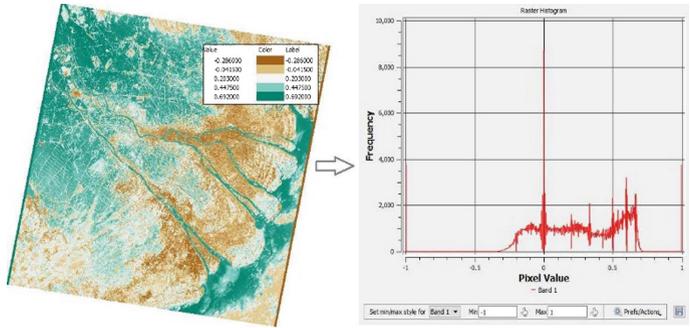


Fig. 6. NDWI for Landsat (ETM+, OLI) with threshold [-1,1]

4 Determine Flooding Disposition over Time

In this study, we used a random forest algorithm in the Orfeo toolbox combining QGIS as well as three different sensors. This algorithm can be used to assign the pixels in the image to the various map classes as illustrated in Fig. 7. This algorithm creates decision trees for each pixel. Each of these decision trees votes on what the pixel should be classified.

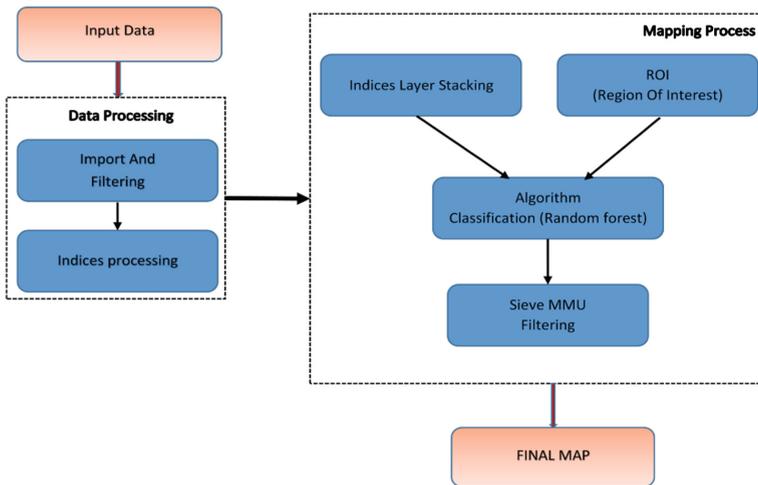


Fig. 7. Supervised classification for floods area

The method of determining the period before floods, during floods and after floods, is observation of remote-sensing images from 2014–2016 to determine the distribution of floods over time. The next, images classification based on NDWI [26] (extract the flooding) and supervised classification (random forest) will separate water, urban,

Sentinel 2 MSI [3] is multi-spectral image will be collected in 2015 and 2017 for floods analysis (there are 91 images, but use 7 images in 2015-09-19; 2015-12-28; 2016-02-26; 2016-09-26; 2016-11-12; 2017-03-02; 2017-07-10).

Table 2. Class information

Color	Value	Legend
Blue	1	Water
Green	2	Vegetation
Pink	3	Urban
Black	4	Unclassification

In datasets, we have 4 class are water, vegetation, urban (built-up) and unclassification (cloud, shadow, other,..). A class have 20 polygons, the polygons is divided into two datasets, is learning dataset (training about 30% of the polygons) and validation (testing about 70% of the polygons). The Table 2 is the class information.

5.2 Tool Used

We use combination of Orfeo Toolbox (OTB version 2.14) [7, 14] and Quantum GIS (QGIS version 2.14) [17] with the Grass Plugin. Inside, OTB is a cross-platform software that can process high-resolution optical, multi-spectral, radar images and have some supervised classification algorithms as SVM, random forest is used to classify images based on homogeneous groups of pixels in the remote sensing images. These applications is open source software displays the precision, recall, F-score and kappa indices and can output results as an *.CSV file.

5.3 Study Area

Mekong delta has 12 provinces and Can Tho City. This is the largest rice growing area of Vietnam, the region also has great potential for tourism. However, the region is also affected by floods that occur every year especially Tan Chau and Hong Ngu districts in An Giang and Dong Thap provinces. The Mekong river floods [9] was formed in May and divided into 3 period [15]. July to August is flood water flowing into the channels of Tan Chau, Hong Ngu district. Flood peak happened in September and October (in 2000, 2011 years). In November and December, the water level will be lowered.

5.4 Scenario 1: Determining the Flooding Area

5.4.1 Floods Area in 2000–2001 with Landsat 7 ETM+

The statistics water levels at Tan Chau, Hong Ngu in 2000–2001 [6] rather higher than normal water levels with the maximum water level of 5.06 m and 4.86 m as illustrated in Fig. 9.

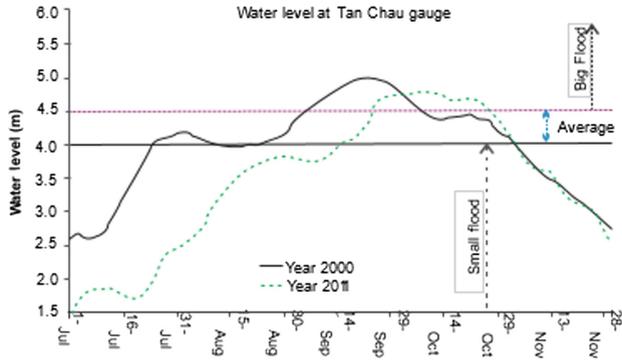


Fig. 9. Flood hydrograph at Tan Chau in 2000 and 2001 [6]

In Fig. 10, water color is clear and widely distributed in flooded areas such as Tan Chau and Hong Ngu district and summarized in Table 3.

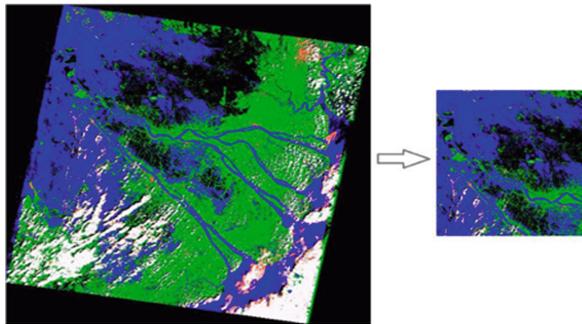


Fig. 10. Result classification of floods area (Tan Chau and Hong Ngu district)

Table 3. Result of area and PixelSum at Tan Chau, Hong Ngu district in 2000

Over time	PixelSum	Area [m ²]
Before floods	380981	342747979
During floods	10708276	9637448400
After floods	1628777	1465386081

With below image (Fig. 11), we will see the change of water surface (blue) through the three images, the middle image has the most water because this is the period of floods in October-2000 in Vietnam (Tan Chau, Hong Ngu district).

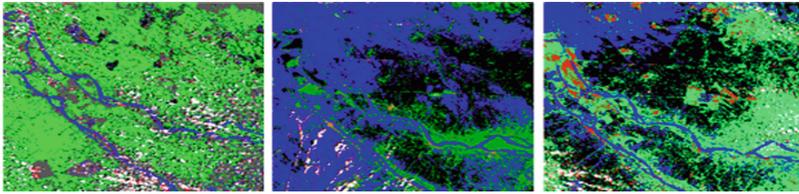


Fig. 11. Change of floods zone in Feb-2000, Oct-2000, Dec-2000 at Tan Chau and Hong Ngu district

5.4.2 Floods Area in 2014–2016 with Landsat 8 OLI

We predict the time of the floods in 2014–2016, and 2017 similar to 2000 above. The right image in Fig. 12 is combined of bands 4, 5, 6. This images are features related to clear water, turbidity water, and vegetation (tree, rice field và vegetable). The classification image in the left of a period before the floods (Feb-2014), results blue pixels less than green and pink. And it summarized in Table 4.

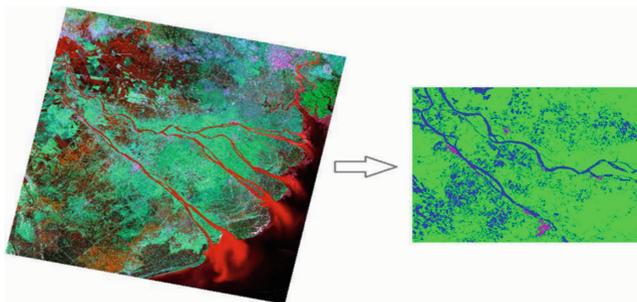


Fig. 12. Result classification of floods area Tan Chau and Hong Ngu (image in the left)

Table 4. Result average of pixel classification and area water

Over time	PixelSum	Area [m ²]
Before floods	1807200	1626480000
During floods	7547690	6792921000
After floods	6429232	5786308800

5.4.3 Floods Area in 2015–2016 with Sentinel-2 MSI

With Sentinel-2 MSI similar to Landsat-8 OLI, floods season analysis during 2000 was theoretical to analyze the Sentinel-2 image in three stages.

In 2015, the period before the floods, no images have a condition for smaller clouds 10%, during floods (image in Sep-2015) and after floods (image in Dec-2015). In 2016, the period after the floods, no images have a condition for smaller clouds 10%, image in Feb-2016 (before floods) and image in Sept-2016 (during floods) is illustrated in Table 5.

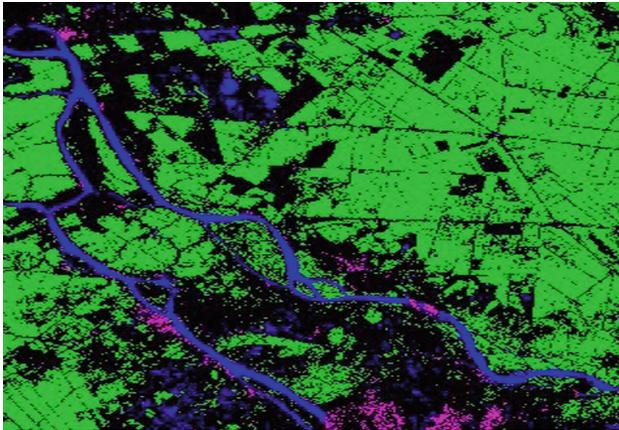
Table 5. Result of area floods

Over Time	Area [m ²]	
	2015	2016
Before floods	–	676784079.39
During floods	1085437975.34	1951083316.47
After floods	1023188829.29	1004844619.61

5.5 Scenario 2: Predicting the Flooding Area

5.5.1 Floods Area in 2017 with Landsat 8 OLI

In this section, we use the Landsat 8 flood classification model (2014–2016) to predict the Landsat 8 image at Mar-2017 as Fig. 13. The results of the classification and prediction accounted for 98.02% of the overall and the lowest water area was 246291280.216 m² with 273861 pixels.

**Fig. 13.** Image classification in March-2017

With the information classified below shows high accuracy 97.59% for water (class 1), 98.59% for urban and 100% for vegetation (class 2 and class 3) and then 96.15% for unclassification (class 4) as Table 6.

Table 6. Result of classification

Class	Area	User accuracy	User accuracy uncertainty
1	246291280.216	97.59	3.3
2	56859209.2352	98.59	2.74
3	2338038201.06	100.0	0.0
4	2349089162.13	96.15	4.27

5.5.2 Floods Area in 2017 with Sentinel-2 MSI

In this section, we use a model to classify floodplains through the phases of Sentinel-2 (2015–2016) to predict the sentinel-2 image by March-2017. The classification and prediction results in 97.67% and overall area is 885939671.67 m² with a pixel number of 8861339.

Classification image (Fig. 14) of Sentinel-2 MSI display clearly colors than Landsat-8 OLI, water objects in fixed areas (the river, lake) are classified correctly (blue). Beside, other objects are also categorized into training class (green: vegetation, pink: urban, black: unclassification).

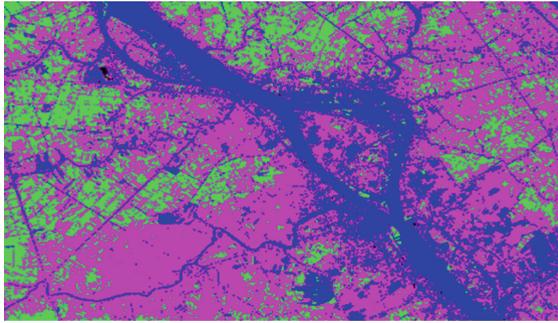


Fig. 14. Image classification in March-2017 (Sentinel-2 MSI) at Can Tho city

With the information classified below shows high accuracy 98.71% for water (class 1), 100% for vegetation and 96.44% for urban (class 2 and class 3) and then 99.58% for unclassification (class 4) as Table 7.

Table 7. The result of classification

Class	Area	User accuracy	User accuracy uncertainty
1	885939671.671	98.71	1.45
2	559115422.744	100	0.0
3	1839076111.51	96.44	2.28
4	26326128.4107	99.58	0.82

6 Conclusion

In this study, we evaluated the performances of the three sensor (Landsat 7 ETM+, Landsat 8 OLI và sentinel-2 MSI) highly effective with random forest classification has been correctly classified in the before floods (Jan to Aug), during floods (Sep to Oct) and after floods (Nov to Dec) in two main areas are Tan Chau and Hong Ngu district.

We chose the greatest flooding times of 2000 to observe and evaluate the flood classification in combination with the flood season in recent years (2014 to 2016). As a

result, the area of flooding in Tan Chau and Hong Ngu district, before the floods were lowest, during floods were highest and after floods decreased compared to during the floods.

In addition, we predict that flooding in 2017 in the Can Tho area will result in accurate classification with a probability of 97.67% (Sentinel-2), 98.02% (Landsat 8) and area of floods (March-2017) is 246291280.216 m² (Landsat 8), 885939671.67 m² (Sentinel-2). We have evaluated Landsat 7 is not as high as Landsat 8 OLI and sentinel-2 MSI when classifying using random forest algorithms. However, the Landsat 7 ETM + image data has more.

References

1. Gao, B.: NdwI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **58**(3), 257–266 (1996)
2. Brivio, P.A., Colombo, R., Maggi, M., Tomasoni, R.: Integration of remote sensing data and GIS for accurate mapping of flooded areas. *Int. J. Remote Sens.* **12**, 429–441 (2002)
3. Dechoz, C., et al.: Sentinel 2 global reference image (2015)
4. Cassé, C., Viet, P.B., Nhung, P.T.N., Phung, H.P., Nguyen, L.D.: Remote sensing application for coastline detection in CA MAU, Mekong delta. In: *International Symposium on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences* (2012)
5. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995)
6. Duong, V., Van Trinh, C., Nestmann, F., Oberle, P.: Land use based flood hazards analysis for the mekong delta. In: *Proceedings of the 19th IAHR-APD Congress 2014*, Hanoi, Vietnam (2014)
7. Christophe, E., Inglada, J., Giros, A.: Orfeo toolbox: a complete solution for mapping from high resolution satellite images. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII, Part B4, pp. 1263–1268, Beijing (2008)
8. Bonn, F., Dixon, R.: Monitoring flood extent and forecasting excess run off risk with radarsat-1 data. *J. Int. Soc. Prev. Mitig. Nat. Hazards* **35**(3), 377–393 (2005)
9. Nguyen, H.Q., et al.: Water quality dynamics of urban water bodies during flooding in Can Tho city, Vietnam. *Water* **9**(4), 260 (2017)
10. Quinlan, J.R.: Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986)
11. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
12. NASA: What are the band designations for the Landsat satellites? URL <https://landsat.usgs.gov/what-are-band-designations-landsat-satellites>
13. Natural Disaster Risk Assessment and Area Business Continuity Plan Formulation for Industrial Agglomerated Areas in the ASEAN Region. Japan International Cooperation Agency OYO International Corporation Mitsubishi Research Institute, Inc. CTI Engineering International Co., Ltd. March (2015)
14. OTB Development Team. Orfeo toolbox 5.4. July 2016
15. Dao, P.D., Liou, Y.-A.: Object-based flood mapping and affected rice field estimation with landsat 8 oli and modis data. *Remote Sens.* **7**(5), 5077–5097 (2015)
16. Potcoava, M.C., Stancalie, G., Raducanu, D.: The using of satellite image data from optic and microwaves data for development of a methodology for identification and extraction of flooded area. *Int. Arch. Photogrammetry Remote Sens.* **6**, 1185–1190 (2000)

17. QGIS Development Team: QGIS Geographic Information System. Open Source Geospatial Foundation (2009). <http://qgis.osgeo.org>
18. Brinkmann, R.: The Art and Science of Digital Compositing. The Morgan Kaufmann Series in Computer Graphics, 2 edn., pp. 149–188 (2008)
19. McFEETERS, S.K.: The use of the normalized difference water index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **17**(7), 1425–1432 (1996)
20. Niculescu, S., Lardeux, C., Hanganu, J., David, L., Mercier, G.: Change detection in floodable areas of the Danube delta using radar images. *Nat. Hazards J. Int. Soc. Prev. Mitig. Nat. Hazards* **78**(3), 1899–1916 (2015)
21. Niculescu, S., Lardeux, C., Hanganu, J.: Synergy between SENTINEL-1 radar time series and Sentinel-2 optical for the mapping of restored areas in Danube delta. In: Proceedings of the International Cartographic Association, vol. 1 (2017)
22. Niculescu, S., Lardeux, C., Guttler, F., Rudant, J.-P.: Multisensor systems and flood risk management application to the danube delta using radar and hyperspectral imagery. *Teledetection* **9**(3–4), 271–288 (2010)
23. McFeeters, S.K.: Using the normalized difference water index (NDWI) within a geographic information system to detect swimming pools for mosquito abatement: a practical approach. *Remote Sens.* **5**(7), 3544–3561 (2013)
24. Sakamoto, T., Van Nguyen, N., Kotera, A., Ohno, H., Ishitsuka, N., Yokozawa, M.: Detecting temporal changes in the extent of annual flooding within the cambodia and the vietnamese mekong delta from modis time-series imagery. *Remote Sens. Environ.* **109**(3), 295–313 (2007)
25. Bertin, X.: A modeling-based analysis of the flooding associated with xynthia, central bay of biscay. *Coast. Eng.* **94**, 80–89 (2014)
26. Zhou, Y., et al.: Open surface water mapping algorithms: a comparison of water-related spectral indices and sensors. *Water* **9**, 256–272 (2017)