FLOOD MAPPING IN MOUNTAINOUS AREAS USING SENTINEL-1 & 2 DATA AND GLCM FEATURES

Beste Tavus[1,2], Sultan Kocaman [2*]

[1] Hacettepe University, Graduate School of Science and Engineering, Ankara, Türkiye - beste.tavus@hacettepe.edu.tr
[2] Hacettepe University, Department of Geomatics Engineering, 06800 Beytepe Ankara, Türkiye - sultankocaman@hacettepe.edu.tr

KEY WORDS: Flood Mapping, Inundated (Flooded) Vegetation, Random Forest, GLCM, Sentinel-1, Sentinel-2

ABSTRACT:

Flooding events have been frequently observed throughout the world and may have devastating effects on the environment. Mapping of flood extent is important for taking the necessary mitigation measures for future. The freely available Sentinel-1 radar and Sentinel-2 optical images allow analysis of flood extent with adequate spatial resolution. However, temporal resolution may be insufficient, and currently only events with the suitable satellite orbital passes can be analyzed with Sentinel sensors. In addition, clouds are often in scene during and shortly after a flood event, which hinders the use of optical imagery. Here, we investigated the complementary use of Sentinel-1 and Sentinel-2 data with a land cover classification approach based on random forest over a part of northern Türkiye, which frequently confronts floods and landslides. We expanded the feature set with principal components (PC) of gray-level co-occurrence matrix (GLCM) variables obtained from Sentinel-1 polarization and Sentinel-2 spectral bands, and also the normalized difference vegetation index (NDVI) and modified normalized difference water index (MNDWI) images produced from the optical data. The training and test data were manually extracted from pre- and post-event optical data. The findings demonstrated that using GLCM PCs significantly increased the overall accuracy (OA = 99% with GLCM and OA = 93% without GLCM) of the classification. Furthermore, the flooded vegetation differs in textural features when compared with the other inundated surfaces, and also permanent water. Therefore, by allowing the separation of flooded vegetation and the other flooded areas, the GLCM data considerably increased the map quality.

1. INTRODUCTION

The occurrence of floods is on the rise due to urbanization and population growth, leading to significant negative consequences for society, economy, and ecosystems worldwide (EMDAT, 2022). In order to mitigate these impacts, to effectively plan disaster management, and to administrate the emergency response and insurance processes, it is crucial to generate reliable spatial and temporal information regarding the extent of flooding. Several studies have demonstrated the effectiveness of widely utilized Earth Observation (EO) datasets and diverse mapping approaches in accurately identifying smooth open water bodies and flooded areas (Tavus et al., 2022).

In most flooding hazard events in rural areas, inundated vegetation accounts for more than three-quarters of the total flooded area. In the literature, approaches utilizing backscattering intensity from different polarizations have been used to determine the extent of inundated vegetation (Cazals et al., 2016). In backscatter-based approaches, the utilization of polarimetric synthetic aperture radar (PolSAR) and interferometric SAR (InSAR) coherence were also preferred for reducing of the confusion between inundated vegetation and urban areas, as well as the distinction between shadow areas and open water (Gallant et al., 2014).

In recent years, studies aiming to determine the flooded vegetation have focused on utilizing the complementary potential of SAR and optical data together in the literature. The studies using only Sentinel-2 data (Bhatnagar et al., 2018) and both Sentinel-1&2 data complementarily (Chatziantoniou et al., 2017) have demonstrated the advantages of both sensors for flood mapping. In addition, many studies investigated the potential of multi-temporal Sentinel-1 (Huang et al., 2017; Mleczko and Mróz, 2018; Tsyganskaya et al., 2018) and Sentinel-2 datasets for this purpose.

A recent study by Tavus et al. (2022) tested various data availability scenarios with Sentinel-1 & 2 data taken at pre- and post-event conditions. The main rationale behind was that Sentinel-2 data is usually unavailable due to the cloud cover, and they evaluated the mapping performances with machine learning classification with and without Sentinel-2 images in addition to Sentinel-1 data. The proposed methodology involved an in-depth analysis on the contribution of textural features for classification accuracy. It was stated that the use of pre-event Sentinel-2 data improved the Sentine1 results in scenarios where post-event optical data cannot be used. In addition, textural features obtained from gray-level co-occurrence matrix (GLCM) from both Sentinel-1 and Sentinel-2 data were utilized in the processing. In this context, a multi-temporal feature space was created by generating pre- and post-event GLCM textures and various spectral indices and modelled with the random forest (RF) classifier. As a result of the validation w.r.t. 3 m spatial resolution PlanetScope imagery, it was observed that the GLCM features were highly effective in determining the flooded areas, inundated vegetation and urban structures.

In this study, the method proposed by Tavus et al. (2022) was applied to an area with dense forests and rugged topography in the northern part of Türkiye, and the performance of the method

* Corresponding author

was assessed. When compared with Tavus et al. (2022), the main difference here is the topography and land cover, as the former study was carried out in a relatively flat region and mapped the flood caused by a dam break in Sardoba, Uzbekistan. In this study, an in-depth analysis of the contribution of textural features to the classification accuracy for the detection of floods occurred in Ordu and Samsun Provinces of Türkiye on 8 August 2018 is presented and discussed accordingly.

2. MATERIALS AND METHODS

Under the following sub-headings, the study area, the data and the methodology are described in detail.

2.1 Study Area

The study area investigated here is located in Fatsa and Unye districts of Ordu Province, and Terme and Carsamba districts of Samsun Province, Türkiye. Both Ordu and Samsun Provinces are located in the northern part of Türkiye, in the Black Sea Region. The study area covers approximately 3037 km^2. Figure 1 illustrates the sub-areas (involving town centers) used for the analysis together with land use land cover (LU/LC) map and e EU-DEM v1.1 (2023). The study area comprises mainly tree and crop cover according to the ESA WorldCover product (ESA-WorldCover, 2020).

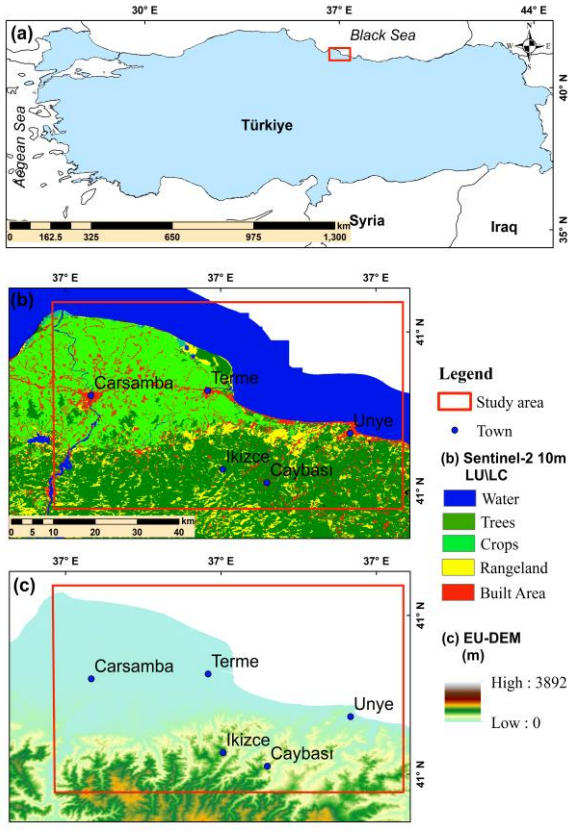

Figure 1. The study site location (a), the LULC map obtained from the ESA WorldCover (b) and (c) EU-DEM v1.1 (25 m).

Ordu Province has approximately 100 km shoreline, mostly with sand. The region exhibits predominantly wet climate, which is typical of the Black Sea Region. Precipitation is observed throughout the year. In addition, Ordu has 36 rivers and streams with different degrees (Demir et al., 2016). According to information and statistics gathered between 1950 and 2011 in the province, landslides are the most frequently observed natural hazard (80%), followed by floods (9%) and rockfall (8%) (Demir et al., 2016).

On August 8, 2018 in the late afternoon, flooding occurred in the area and damaged houses, agricultural areas, and infrastructure. There have been numerous landslides and a total of 8 bridges that have been destroyed. In Fatsa and Unye Towns, about 80 mm of precipitation fell per square meter and caused flooding of streams and rivers. The flooding caused the Cevizdere Bridge in Unye Town to collapse, and the Black Sea Coastal Highway was shut down to traffic (TRT News, 2018). Further details about the event can be found in Tavus et al. (2019, 2020) and Kocaman et al. (2020).

2.2 Study Datasets

In this study, we utilized Sentinel-1 and Sentinel-2 datasets provided by the ESA Copernicus Programme (Copernicus, 2020). Table 1 summarizes the data characteristics such as ground conditions, the time of the data acquisition, usage purpose. The selected Sentinel-1 and Sentinel-2 data accurately represented the pre-and post-flood conditions in terms of time and resolution.

Table 1. Basic specifications of the datasets used in the study.

	Acquisiton Date	Condition	Usage
Sentinel-1	18/05/2018 (DS1)	Pre-event	Feature extraction & Classification
	18/05/2018 (DS2)		
	10/08/2018 (DS3)	Post-event	
	10/08/2018 (DS4)		
Sentinel-2	16/05/2018 (DS5)	Pre-event	
	09/08/2018 (DS6)	Post-event	Train data delineation
	09/08/2018 (DS7)		

DS1: S1B_IW_GRDH_1SDV_20180518T034024_20180518T034049_010970_014159_08A7
DS2: S1B_IW_GRDH_1SDV_20180518T034049_20180518T034114_010970_014159_B477
DS3: S1B_IW_GRDH_1SDV_20180810T034029_20180810T034054_012195_016770_7612
DS4: S1B_IW_GRDH_1SDV_20180810T034054_20180810T034119_012195_016770_841C
DS5: S2A_MSIL2A_20180516T081611_N0207_R121_T37TCF_20180516T102815
DS6: S2B_MSIL2A_20180809T081559_N0208_R121_T37TCF_20180809T112729
DS7: S2B_MSIL2A_20180809T081559_N0208_R121_T37TCF_20180809T132329

The post-flood Sentinel-1 (S1) and Sentinel-2 (S2) data have been acquired on August 09 and May 10, 2018, respectively (DS3, DS4, DS6, DS7). The first cloud-free data before the flood date was selected for the pre-event Sentinel-2 (DS5). For pre-event Sentinel-1, DS1 and DS2 were selected, which correspond to the date of the DS5 data.

2.3 Methodology

The methodology of this study consists of three basic stages as (i) data pre-processing and feature extraction, (ii) feature selection, and (iii) modelling, mapping and accuracy assessment, as depicted in Figure 2. In the first stage of data pre-processing and feature extraction, several methods such as noise filter and removal of systematic errors caused by terrain were applied to the Sentinel-1 data. We used blue (B2), green

(B3), red (B4), NIR (B8) and SWIR (B11) of pre- and post-event Sentinel-2 data. The lower resolution band of Sentinel-2 (B11) was upsampled as well. Afterwards, cloud and cloud shadow pixels mask out the data by using masks supply from Sentinel-2.

Shadow and layover effects in Sentinel-1 radar data were also removed at this stage. As the last pre-processing step, mosaics were created from the DS1 - DS2, DS3 - DS4, and DS6 -DS7 datasets covering the study area.

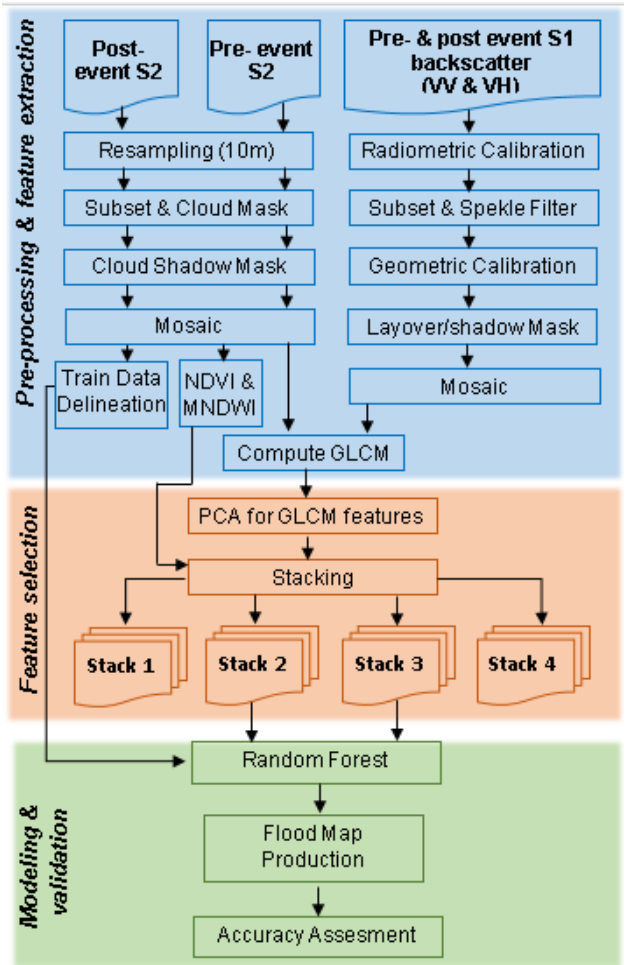

Figure 2. Overall methodology of the study.

For the feature extraction, a total of 10 GLCM texture features introduced by Haralick et al. (1973) were applied to each of the pre- and post-event Sentinel-1 and pre-event Sentinel-2 band data. The variables include dissimilarity, homogeneity, contrast, angular second moment (ASM), energy, maximum probability, entropy, mean, variation and standard deviation. Thus, 50 GLCM texture variables were produced from pre-event Sentinel-2 data and 40 GLCM texture features were produced from pre- and post-event Sentinel-1 data.

In addition, the normalized difference vegetation index (NDVI) and modified normalized difference water index (MNDWI) were produced from the pre-event Sentinel-2 data. The pre-event NDVI and MNDWI images, and their histograms are presented in Figure 3. Both index values were in the range of (-1:1). As can be seen from the NDVI given in Figure 3, the study area covers dense agricultural and forest areas. Likewise, many streams with different characteristics between forests are seen in MNDWI. In the flood disaster that occurred, the towns in the forest areas were flooded because of the overflow of these streams.

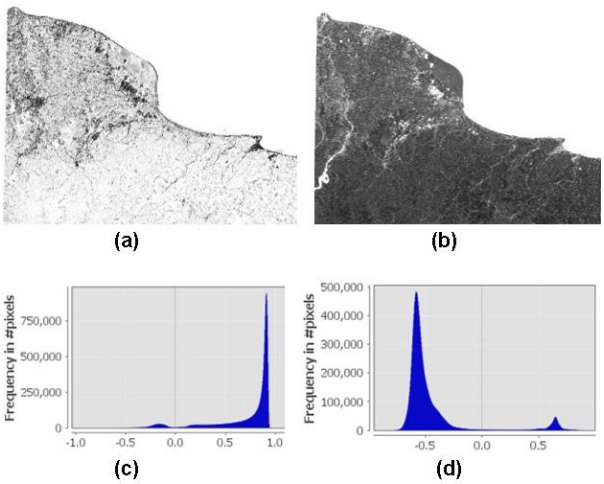

Figure 3. (a) Pre-event NDVI and (b) pre-event MNDWI indices; (c) pre-event NDVI and (d) pre-event MNDWI histograms.

In the feature selection stage, the Principle Component Analysis (PCA) was applied to the GLCM texture features in order to reduce the dimensionality as there were a total of 90 of them. The PCA was separately applied to GLCM texture features produced from Sentinel-1 and Sentinel-2 and 3 principle components (PCs) were produced from each. As a result of the PCA, a total of 6 GLCM principle components (GLCM PCs) obtained from the analysis were used as additional information to the original pre- and post-event S1 and pre-event S2 bands.

In order to assess the contribution of the GLCM feature components (GLCM PCs) to the prediction of classes, original S1 and S2 data, NDVI, MNDWI and produced GLCM PCs were stacked with different combinations (Table 2). For this purpose, while the bands in Stack 1 were pre-event S2, NDVI and MNDWI, pre- and post-event S1, Stack 2 was obtained by adding S1 GLCM PCs and S2 GLCM PCs to Stack 1. In addition, in order to separately analyse the effect of GLCM PCs produced from Sentinel-1 and Sentinel-2 on the prediction results, Stacks 3 and 4, which contain S1 GLCM PCs and S2 GLCM PCs components, respectively, were created (Table 2).

Table 2. Data Stack and their components.

	Stack1	Stack2	Stack3	Stack4
Pre- S2	✓	✓	✓	✓
Pre- S2 NDVI	✓	✓	✓	✓
Pre- S2 MNDWI	✓	✓	✓	✓
Pre- S1	✓	✓	✓	✓
Post- S1	✓	✓	✓	✓
S1 GLCM PCs		✓	✓	
S2 GLCM PCs		✓		✓

In the third stage, the RF method proposed by Breiman (2001), which is based on decision trees, was used for learning from data formed in the previous stage. For this purpose, a total of 69,544 training samples manually delineated from pre- and

post-event Sentinel-2 were used with a tree size of 300 and 3-fold cross-validation. The RF parameters were previously tested by Tavus et al. (2021) and found suitable. Previous studies carried out by Tavus et al. (2018, 2020, 2022) have shown that instead of applying a binary classification approach for flooded areas, applying a holistic LU/LC classification increases the accuracy and reliability of flood extent maps. Thus, the six classes namely flooded vegetation (FV), flooded area (FL), permanent water (PW), urban area (Ur), crop (CR), and forest (FR), were identified from the pre- and post-event Sentinel-2 images.

In Tavus et al. (2022), the effect of post-event S2 data on the prediction results was analyzed and it was suggested to use pre-event S2 in cases even when post-event S2 cannot be used. Because the post-event S2 data have a high amount of cloud here, it was used only for the delineation of FL and FV training samples and was not included in other processes to obtain ground-truth accurately. The results of the classification were assessed in terms of overall accuracy for all classes.

3. RESULTS AND DISCUSSIONS

Figure 4 (a and b) shows the classification results from the Stack 1 and Stack 2 features and their detailed views of sub-areas. The maps include the classes of flooded vegetation (FV), flooded area (FL), permanent water (PW), urban area (Ur), crop (CR), and forest (FR). The overall accuracies (OA) achieved from the Stacks 1, 2, 3 and 4 were 93% (without GLCM), 99% (with S1&S2 GLCM), 97% (with S1 GLCM) and 96% (with S2 GLCM), respectively. The OA values show that the integration of the Sentinel-1 and Sentinel-2 feature sets provided the best prediction performance for the classification of all classes. No significant change was observed in the PW class within all classes based on the visual inspection.

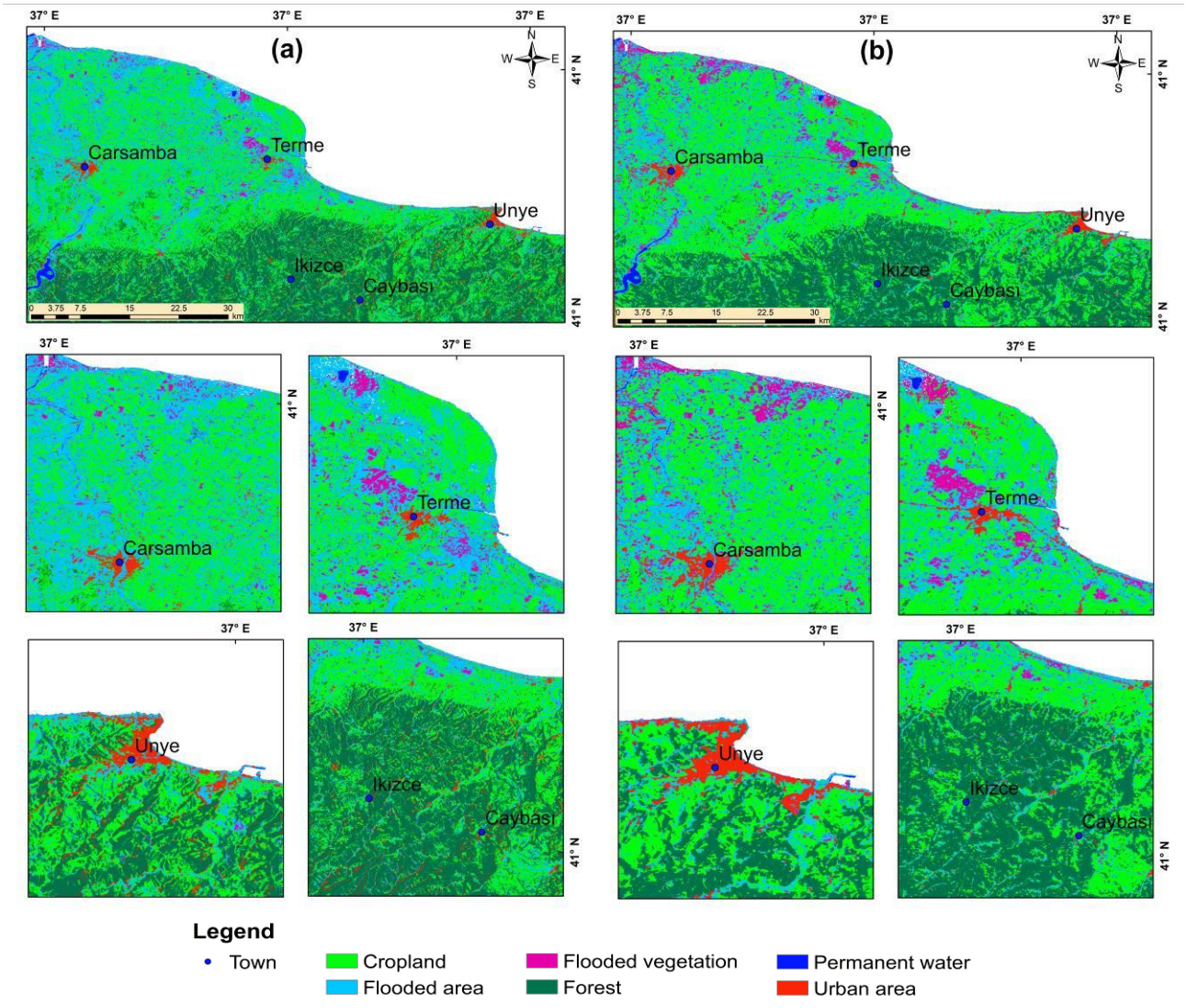

Figure 4. The RF classification results obtained from (a) Stack 1 (without GLCM) and (b) Stack 2 (with S1&S2 GLCM).

As a result of the visual evaluation of the flood map given in Figure 4 (a), it was seen that all classes especially the flood and agricultural areas exhibit a noisy pattern. It was also observed that the FV class was labelled as FL according to the Carşamba and Terme sub-areas given in Figure 4 (a) and (b). On the other hand, the Unye and Terme sub-areas were inspected in the detailed views, and it was observed that the floods in the urban areas were better determined with the Stack 1.

The use of texture information has led to the fact that flood pixels occurring in urban areas were largely labelled as urban pixels. Also, while Stack 2 (with S1&S2 GLCM) gives better results in detecting the urban pixels (Unve and Terme) in a rather open area, Stack 1 (without GLCM) is more effective in detecting the urban pixels (Ikizce and Caybaşı) in the dense forest area.

Figure 5 (c and d) shows the classification results from the Stack 3 and Stack 4 features and their detailed views of sub-areas. Stack 3 contains texture components produced from Sentinel-1 (S1 GLCM PCs), while Stack 4 includes texture components produced from Sentinel-2 (S2 GLCM PCs). Stack 3 is more sufficient in labelling the flood pixels that occur caused by the overflow of existing streams in forest areas (see Figure 5c). When the FV class in the Carşamba and Terme sub-areas are compared, it was seen that the texture components produced in Sentinel-1 (S1 GLCM PCs) contribute to the detection of FV pixels. When Stack 3 and Stack 4 are compared in terms of floods in urban areas, it can be said that Stack 3 is more effective based on visual inspection. On the other hand, using only Sentinel-1 data (Stack 3) resulted in labelling most crop pixels and forests.

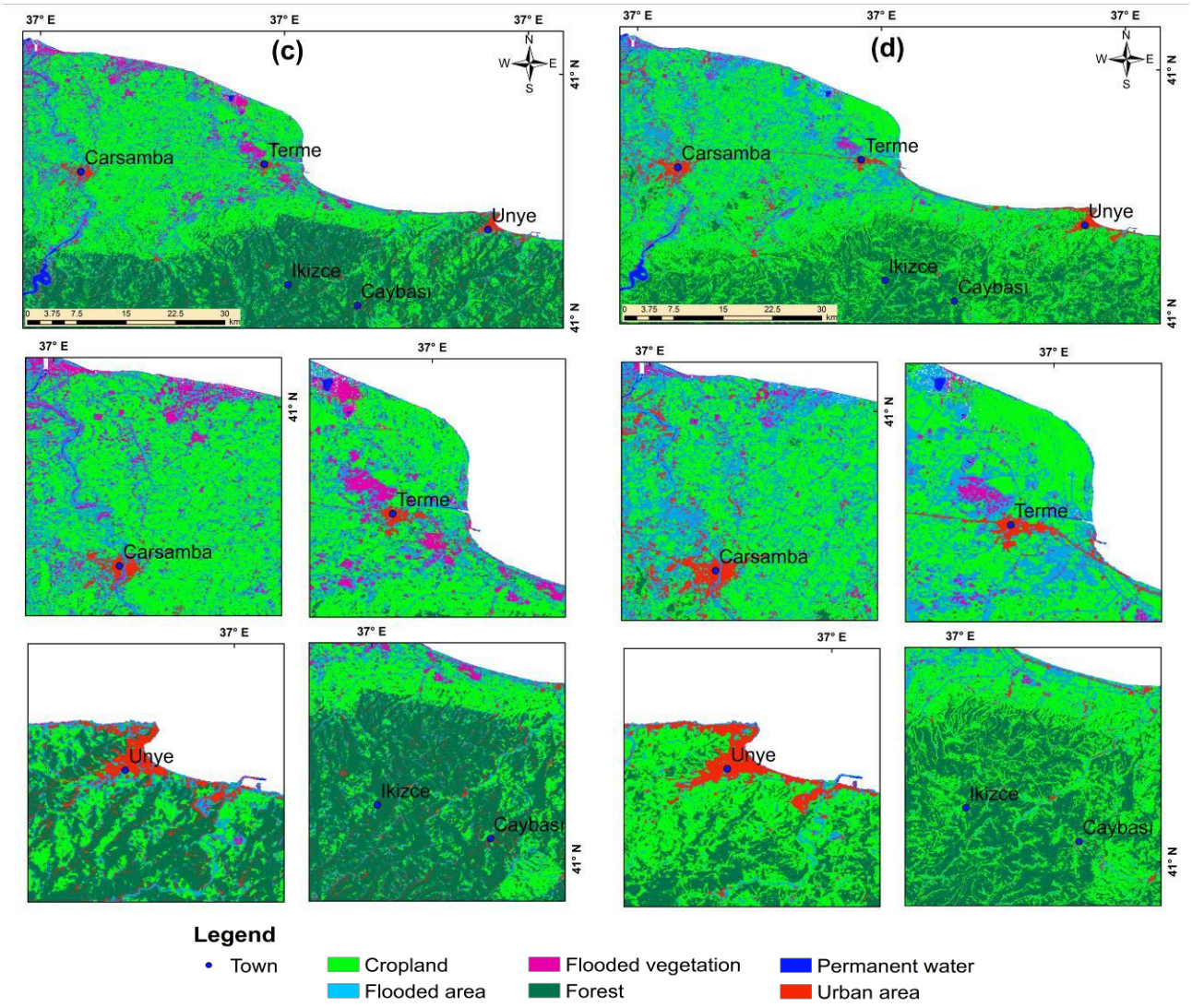

Figure 5. The RF classification results obtained from (c) Stack 3 and (d) Stack 4.

4. CONCLUSIONS AND FUTURE WORK

In the present study, the contribution of GLCM textural features for flood extent mapping including flooded vegetation were evaluated with the RF classifier applied to the learning set obtained from various Sentinel-1 polarization and Sentinel-2 spectral bands. The study area is located in Ordu and Samsun Provinces of Türkiye, which are frequently affected by flash flooding, and comprises mainly cropland and dense forest area. In the study area, partially cloud-free Sentinel-2 images were available representing the post-event status and the topography is rugged. Four sets of learning variables, one containing GLCM textural information in the form of principal components and two separately containing GLCM textural information produced from Sentinel-1 and Sentinel-2, and the last one without GLCM textures were produced. A LULC classification for a total of six classes was followed here. The results were

assessed using test data manually delineated from pre-and post-Sentinel-2 datasets.

The results showed that the use of GLCM PCs greatly contributed to increase the overall classification accuracy (OA= 99% with GLCM and OA=93% without GLCM). It is evident that the inundated vegetation shows a different scattering mechanism in radar data compared to other classes. This particular finding highlights the difference in texture properties of the inundated vegetation compared to the floods, water, or the other agricultural areas in the region. Thus, GLCM data significantly contributes to increasing the classification accuracy by preventing the mixture between the FV and FL classes specifically.

As future work, it is planned to produce accuracy metrics by comparing the classification results with external reference data such as satellite imagery with higher resolution. In addition, as a result of visual inspections, it was clear that the textural features produced from Sentinel-1 were effective in the analysis of floods in urban and dense forest areas, which require further in-depth analysis.

ACKNOWLEDGEMENTS

This study is part of the Ph.D. thesis research of Beste Tavus.

REFERENCES

Bhatnagar, S., Ghosh, B., Regan, S., Naughton, O., Johnston, P., Gill, L. 2018. Monitoring environmental supporting conditions of a raised bog using remote sensing techniques. Proc. IAHS 380, 9–15. https://doi.org/10.5194/piahs-380-9-2018.

Breiman, L., 2001. Random forests. Machine Learning, 45(1), 5-32.

Cazals, C., Rapinel, S., Frison, P.L., Bonis, A., Mercier, G., Mallet, C., Corgne, S., Rudant, J.-P. 2016. Mapping and Characterization of Hydrological Dynamics in a Coastal Marsh Using High Temporal Resolution Sentinel-1A Images. Remote Sensing, 8, 570. https://doi.org/10.3390/rs8070570

Chatziantoniou, A., Petropoulos, G.P., Psomiadis, E. 2017. Co-Orbital Sentinel 1 and 2 for LULC mapping with emphasis on wetlands in a Mediterranean setting based on machine learning. Remote Sensing. 9, 1259. https://doi.org/10.3390/rs9121259.

Demir, A., Ilgen, H.G., Isik, A.,2016. Ordu İlinde, 04-06/07/2016 Tarihleri Arasinda Meydana Gelen Sel-Taşkin-Su Baskini ve Heyelan Olaylarinin Genel Değerlendirmesi, Presentation at 4. Ulusal Taşkın Sempozyumu, 21-24 November, Rize, Turkey.

EU-DEM v1.1, 2023. https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1 (10 June 2023).

Gallant, A., Kaya, S., White, L., Brisco, B., Roth, M., Sadinski, W., Rover, J. 2014. Detecting Emergence, Growth, and Senescence of Wetland Vegetation with Polarimetric Synthetic Aperture Radar (SAR) Data. Water, 6, 694–722. https://doi.org/10.3390/w6030694

Haralick, R.M., Shanmugam, K., Dinstein, I. 1973. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics. 6, 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Huang, W., Devries, B., Huang, C., Jones, J., Lang, M., Creed, I. 2017. Automated extraction of inland surface water extent from sentinel-1 data. International Geoscience and Remote Sensing Symposium (IGARSS) 2259–2262. https://doi.org/10.1109/ IGARSS.2017.8127439.

Kocaman, S., Tavus, B., Nefeslioglu, H. A., Karakas, G., & Gokceoglu, C., 2020. Evaluation of floods and landslides triggered by a meteorological catastrophe (Ordu, Turkey, August 2018) using optical and radar data. Geofluids, 2020, 1-18.

Mleczko, M., Mróz, M. 2018. Wetland mapping using SAR data from the Sentinel-1A and TanDEM-X missions: a comparative study in the Biebrza Floodplain (Poland). Remote Sensing. 10, 78. https://doi.org/10.3390/rs10010078

Tavus, B., Kocaman, S., Gokceoglu, C., Nefeslioglu, H.A. 2018. Considerations on the Use of Sentinel-1 Data in Flood Mapping in Urban Areas: Ankara (Turkey) 2018 Floods. ISPRS Comm. V Symposium, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 42(5), 575-581. https://doi.org/10.5194/isprs-archives-XLII-5-575-2018.

Tavus, B., Kocaman, S., Nefeslioğlu, H., Gökçeoğlu, C. 2019. Flood Mapping Using Sentinel-1 SAR Data: A Case Study of Ordu 8 August 2018 Flood. International Journal of Environment and Geoinformatics, 6(3), 333–337, 2019. https://doi.org/10.30897/ijegeo.666212

Tavus, B., Kocaman, S., Nefeslioglu, H.A., Gokceoglu, C. 2020. A Fusion Approach for Flood Mapping Using Sentinel-1 and Sentinel-2 Datasets. ISPRS Virtual Congress 2020. https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-641-2020

Tavus, B., Kocaman, S., Gokceoglu, C. 2022. Flood damage assessment with Sentinel-1 and Sentinel-2 data after Sardoba dam break with GLCM features and Random Forest method. Science of the Total Environment, 151585. doi.org/10.1016/j.scitotenv.2021.151585

The International Disaster-Emergency Events Database (EMDAT). Disasters Year in Review 2022; Available online: https://www.emdat.be/publications (accessed on 12 April 2022).
Tsyganskaya, V., Martinis, S., Marzahn, P., Ludwig, R. 2018. Detection of temporary flooded vegetation using Sentinel-1 time series data. Remote Sensing. 10, 1286. https:// doi.org/10.3390/rs10081286.

T. R. T. News, “Ordu’da Sel Felaketi. Online news portal of Turkish radio television agency on 8 august 2018,” 2018, May 2023, https://www.trthaber.com/haber/turkiye/ordudasel-felaketi-379211.html.

Tsyganskaya, V., Martinis, S., Marzahn, P., Ludwig, R. 2018. Detection of temporary flooded vegetation using Sentinel-1 time series data. Remote Sensing. 10, 1286. https:// doi.org/10.3390/rs10081286