IDENTIFICATION OF OPTIMUM METHOD FOR SENTINEL-1 SAR IMAGE FLOOD MAPPING DURING TROPICAL STORM KROVANH IN AGUSAN DEL SUR, PHILIPPINES

Krisha Ann N. Oller[1], Kimberly A. Pastor[1], Monalaine M. Bermoy[1, 2], and Kendel P. Bolanio[1, 2]
[1]Department of Geodetic Engineering, Caraga State University, Ampayon, Butuan City 8600, Philippines,
Email: krishaann.oller@carsu.edu.ph, kimberly.pastor@carsu.edu.ph
[2]Caraga Center for Geo-informatics, Caraga State University, Ampayon, Butuan City 8600, Philippines,
Email: mmbermoy@carsu.edu.ph, kendel.bolanio@gmail.com

KEY WORDS: Flood mapping, Sentinel-1, binarization, RF, SVM

ABSTRACT: Flooding is one of the most severe and destructive natural hazards, posing significant threats to human lives, properties, and the economy. An accurate flood mapping is a vital component to understand the risks to plan for a mitigation measure. This study utilized Sentinel-1 SAR images to map the flood extent that occurred during December 2020, the onslaught of Tropical Storm Vicky (International Name: Krovanh) due to the sensor's capability to penetrate clouds which enables flood mapping during typhoon events. Before actual flood mapping of the area, backscattering parameters such as VV and VH Polarizations and NDDPI, VV+VH, and VV-VH indices were used to generate maps using machine learning algorithms, random forest (RF), and support vector machine (SVM). The traditional way of flood mapping, which is binarization, was also used to generate maps. The flood maps generated from these methods were assessed, identifying which has the highest flood mapping accuracy. To determine the practical way in flood mapping, the results from the combination of Sentinel-1 polarizations have to undergo an accuracy assessment using the actual ground truth validation points. The results of flood mapping showed that SVM best classified the flood extent. The flood mapping process was then applied to the study area using the identified combination of the polarization and SVM method.

1. INTRODUCTION

Flooding has been affecting humanity for so long in different aspects of the society with effects such as losses in the economy, injuries brought about to the population affected, and in worst cases, numerous casualties (Bosello et al., 2018). It is dangerous when it occurs, and the people are not totally equipped and fully informed on the risks. This case happened in the Caraga Region during the onslaught of Tropical Storm "Krovanh" (local name: Vicky) last December 2020 that unfortunately took four lives in the region only (Davies, 2020).

To aid in preparing such problems, mapping the extent and strength of flooding that occurred or currently occurring is employed. One way of mapping the extent of the flood is through using Sentinel-1 Satellite data, which produces images that contain SAR data. Synthetic Aperture Radar (SAR) data can penetrate through thick clouds and provides all-weather and day-to-night imagery. Thus, it can reach different earth objects which have other backscatter properties. Sentinel 1 SAR data characterized by 10-meter resolution images and six (6) days revisit time with dual-polarization HH and VV backscattering or a combination of both is sensitive to floodwaters. Therefore, SAR is suitable for mapping the flood extent.

In the past two decades, several studies have been conducted to map the extent of proposed solutions and vary only to the sources and algorithm used. Machine learning approaches provide new possibilities for flood detection as more data becomes available, computing power increases, and machine learning algorithms improve. The supervised machine learning classifiers train and evaluate the proposed algorithm using Sentinel-1 image and flood delineation maps. The critical strength of machine learning classifiers is the ability to provide reliable, fast computation and handle many data inputs to generate flood maps (Torino et al., 2020). This study will be significant for mapping the flood extent of Agusan del Sur.

The main objective of this study is to map the extent of flooding in Agusan del Sur during the Tropical Storm "Krovanh" using the flood mapping method with the highest accuracy. Specifically, this study aims: (1) to generate Sentinel-1 SAR image flood map using the traditional method binarization, (2) to generate Sentinel-1 SAR image flood map using machine learning classifiers, namely, Random Forest and Support Vector Machine, and (3) to validate and assess the resulting maps from the flood mapping methods used in this study – Binarization and machine learning classifier; using ground truth points.

2. METHODOLOGY

2.1 Overview of the Study

Shown in Figure 1 is the methodological flowchart containing three (3) significant activities. These are the following: (1) generation of flood maps using SAR images from Sentinel-through binarization, (2) generation of flood maps using machine learning classifiers, and (3) accuracy assessment of the generated flood maps using the ground truth data. The specific methods and the specific activities will be discussed systematically in the sub-sections.

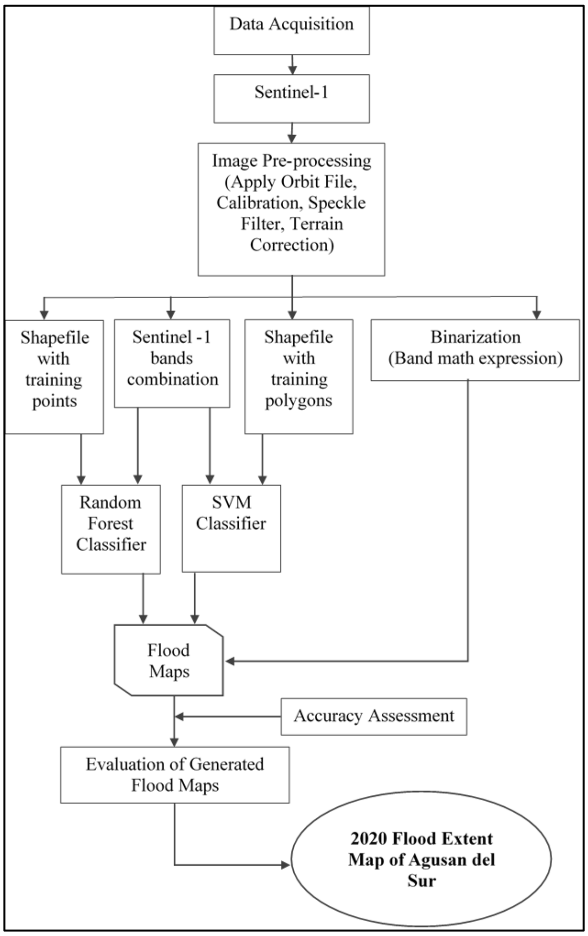

Figure 1. Methodological Flowchart

2.2 Study Area

The province's area stretches up to 8,568 square kilometers, making it the 7th largest province in the country. Agusan del Sur also houses parts of the Agusan marsh which is near the municipality of Bunawan. It is a vast complex of freshwater wetlands and watercourses with numerous other water bodies with extensive flooding from November to March.

2.3 Data Collection

Images were collected from the Sentinel-1 satellite. The images were downloaded through the online platform of Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus/#/home).

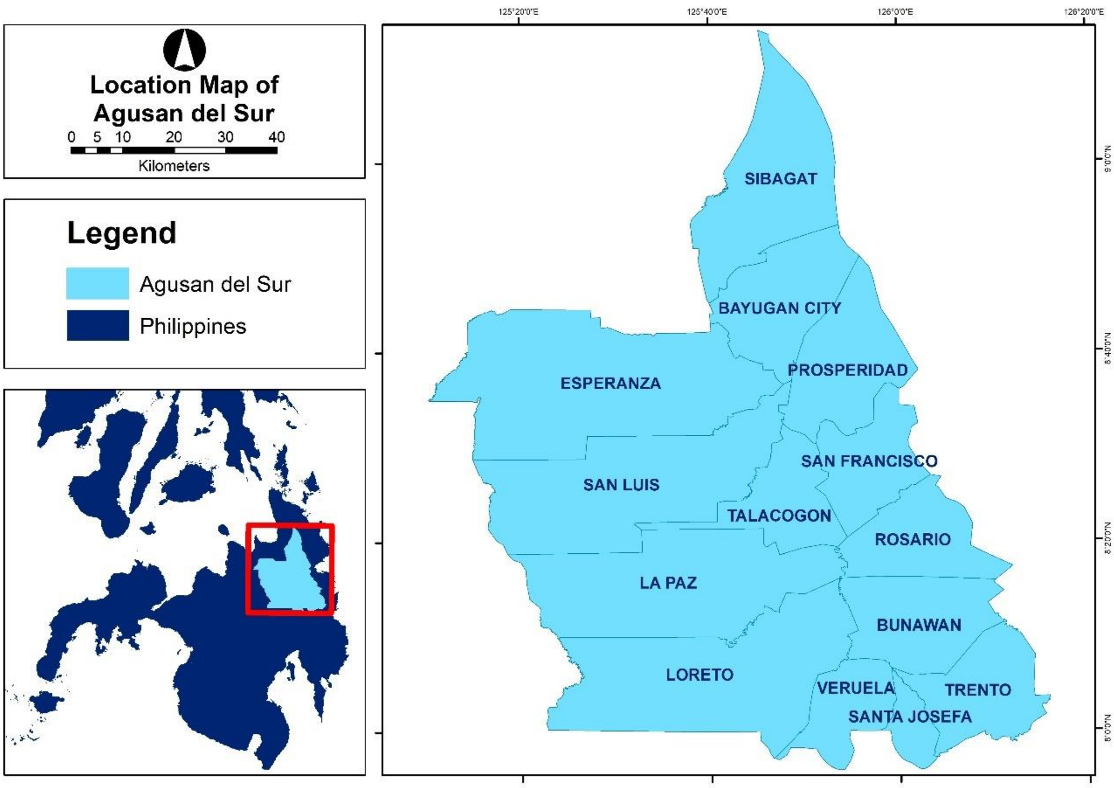

Figure 2. Study Area

Table 1. List of Sentinel 1 images

Images	Date of Acquisition	Mode	Polarization
S1A_IW_GRDH_1SDV_20170121T212336_ 20170121T212401_014935_0185FC_F29C	January 21, 2017	IW	VV and VH
S1A_IW_GRDH_1SDV_20201219T212403_ 20201219T212428_035760_042F6F_D2A4	December 19, 2020	IW	VV and VH

The researchers downloaded the flooded images acquired on January 21, 2017 (TECF 2017) and December 19, 2020 (Tropical Storm Krovanh), considering the dates where flooding is most visible for the duration of the event. The first product on the list is the flooded image for the TECF 2017 phenomena used to evaluate the three different methods. The TECF 2017 flooding event was considered in assessing the flood mapping methods because the available ground truth points for the accuracy assessment are only from the TECF 2017 event. These ground points were requested from the Caraga Center for Geo-Informatics (CCGeo). The other image is for the mapping of the 2020 flooding event for Tropical Storm Krovanh.

2.4 Sentinel-1 Image Pre-processing

The Sentinel application platform (SNAP), a typical architecture for all satellite toolboxes of Sentinel, the Sentinel-1 images were pre-processed. Pre-processing is essential for several applications using earth observations. This study would utilize the SAR data, which contains dual polarimetric mode, specifically HH-HV and VV-VH, but will focus on the VV-VH polarization from Sentinel-1. This step is needed to make the data ready for higher processing levels (Mandal, 2019).

2.4.1 Apply Orbit File

Apply Orbit File was used to improve the geolocation accuracy since it automatically updates the orbit state vectors for the SAR data after over the internet, compiling days to weeks of precise orbits of the source satellite. This was done since the orbit state vector included in the metadata information of SAR products is generally not accurate (Filipponi, 2019).

2.4.2 Radiometric Calibration

Uncalibrated SAR imagery can be utilized for qualitative use. Still, to provide imagery that can be used for quantitative purposes, the pixel values need to be directly related to the radar backscatter of the reflecting surface. The Sigma nought calibration option calibrates the backscatter returned to the antenna from a unit area on the scene and related to a ground range. Scientists also use this to interpret surface scattering and reflection and surface properties, and this is also what is better in land and water separation in the data.

2.4.3 Speckle Filter

As an output of the sensor, the image is composed of both more robust and weaker signals that interfere with one another and create what seems like random patterns of brighter and darker pixels, making the image appear grainy. This is called speckle, and speckle filtering is applied using Lee filter 15x15 which efficiently suppresses extremely high and low values while preserving texture information (Rana and Suryanarayana, 2019).

2.4.4 Terrain Correction

Due to the angle of incidence towards the ground from satellite during image acquisition and the changes in topography on the area of study, there are geometric distortions present in SAR images. Terrain correction and orthorectification are done using the tools in SNAP. This specific study will be the Range Doppler Terrain Correction, where the DEM used is the auto-downloaded SRTM 3-Sec (Conde and Muñoz, 2019).

2.5 Flood Mapping Using SAR Images

Synthetic Aperture Radar is a radar sensing system capable of producing high-resolution radar images in any weather condition and even at night time. This is why it is used for mapping a specific area regardless of heavy rains, flooding, and even fogs.

2.5.1 Binarization

Using SAR images directly in SNAP is the traditional way of showing flood extent, which is done by separating water and non-water pixels using threshold values which in this case for the VV image is -15.33 and for the VH image is -21.95. To classify the image and directly separate water and non-water pixels, the band math expression "if sigmaVV > -15.33 then 1 else 0" is used to separate the values into two for the VV image. The same expression using the other threshold value was used for the VH image, where sigmaVV signifies the values for the pixels in the image.

2.6 Flood Mapping Using Machine Learning Classifiers

Machine learning, over the years, has proved to provide new possibilities for flood detection and is even considered to be one of the better instruments in generating flood maps and predicting floods. It has been recognized due to its potential of improved accuracy with much-reduced computation and model development time than other traditional methods.

2.6.1 Using Support Vector Machine Classifier

For training samples, 150 polygons each for the flooded and non-flooded areas were created using the pre-processes Sentinel-1 image. The parameter used in the processing was the radial basis function for the kernel type. Then the inputs used are the flooded images in VV and VH polarization, generated Sentinel 1 indices, namely NDDPI, VV+VH, and VV-VH, and the shapefile containing the training samples. The process was repeated three times using different combinations: VV, VH, NDDPI; VV, VH, VV+VH; and lastly, VV, VH, VV-VH.

2.6.2 Using Random Forest Classifier

Using the same condition used in the SVM classification, the flooded images and the indices computed from the images are layer-stacked. Using the image generated for each combination, the spectral profile of one flooded and non-flooded each are extracted and a decision tree is built using the following conditions; for NDDPI, if pixel value

> -0.083178, all true values = non-flooded while false = flooded, for VV+VH, if pixel value > -47.343304, all true values = non-flooded while false = flooded, and lastly for VV-VH, if pixel value > 2.909464, all true values = non-flooded while false = flooded. The process was repeated for each of the conditions to generate flood map extent.

2.7 Validation

For the validation using the ground truth data and the resulting flood maps from all the methods and combinations or polarization, their accuracy was assessed by rasterizing the validation points. The raster datasets were then used as inputs in the semi-automatic classification plugin found in the QGIS software.

3. RESULTS AND DISCUSSION

3.1 Sentinel 1- Preprocessing

Figure 3a shows the Sentinel-1 VH backscattering with values ranging from -31.9876 to 13.8501, where the highest values appear white and likely represents built-ups while the lowest values appear white, indicating water bodies. In Figure 3b, backscattering values from the VV polarization are shown. The values range from -28.8894 to 21.807 and indicate higher values than that of the VH polarization. Both represent built-ups with the highest positive value and water bodies with negative values.

a.

b.

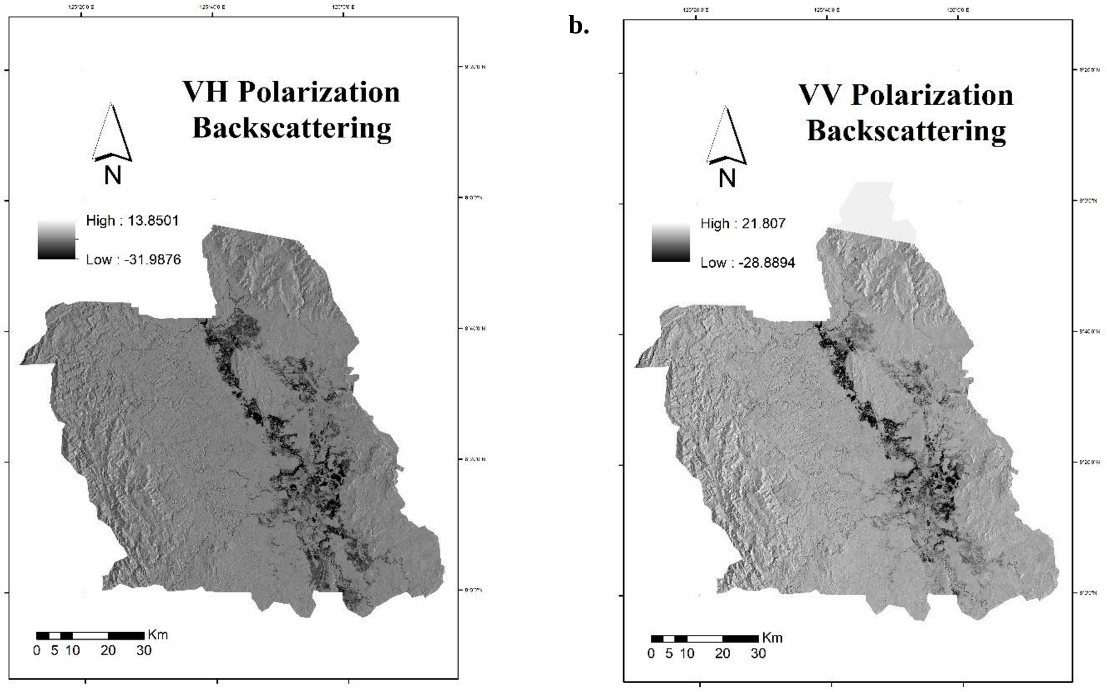

Figure 3. 2017 Flooded Image in (a) VH Polarization, and (b) VV Polarization.

3.2 Validation of Flood Maps

The accuracy of the generated maps would be validated using the ground truth points from the TECF phenomena in 2017 since it is the latest flood event with actual flood points. Through the accuracy assessment process, the optimal method for flood mapping would be identified and used to map the flooding event last December 2020.

3.2.1 Binarization Flood Maps and Error Matrices

The values assigned to the area are flooded and non-flooded layers (actual and model-derived). The shapefile for the validation points contained 75 points for flooded and non-flooded pixels. The result from the binarization method in VH polarization, as shown in Table 2, shows 87.33% overall accuracy where 65 out of 75 pixels corresponded correctly with the flooded areas in the reference image and 66 out of 75 pixels also compared accurately for the non-flooded areas. The flood map generated using Binarization in VV polarization generated 66 out of 75 flooded pixels and 68 out of 75 non-flooded pixels precisely from the validation points bringing the overall accuracy to 89.33%.

Table 2. Error Matrix of Flood Map using Binarization in VH and VV Polarizations.

BINARIZATION (VH)					**BINARIZATION (VV)**				
	Actual Flooded	Actual Not Flooded	Total	User Accuracy		Actual Flooded	Actual Not Flooded	Total	User Accuracy
Model Flooded	65	9	74	87.84	Model Flooded	66	7	73	90.41
Model Not Flooded	10	66	76	86.84	Model Not Flooded	9	68	77	88.31
Total	75	75	150		Total	75	75	150	
Producer's Accuracy	86.67	88.00			Producer's Accuracy	88.00	90.67		
Overall Accuracy	**87.33%**				**Overall Accuracy**	**89.33%**			

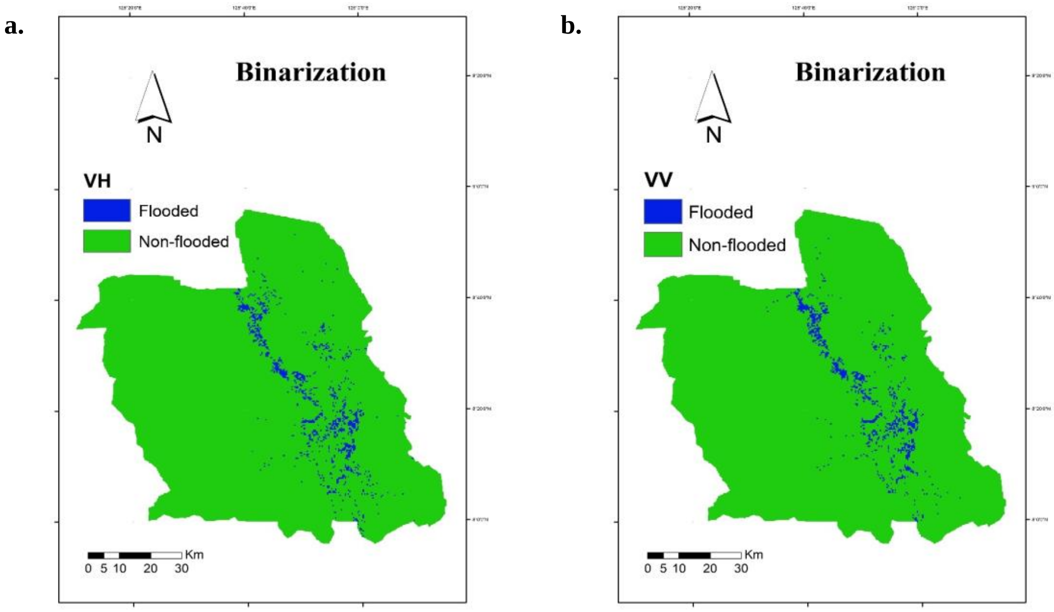

Figure 4. Flood Map using Binarization in (a) VH Polarization, and (b) VV Polarization.

3.2.2 SVM Classification Flood Maps and Error Matrices

For the flood map extent generated using NDDPI index combination, 66 out of 75 pixels were classified for flooded while 70 out of 75 pixels were correctly for non-flooded resulting in high overall accuracy of 90.67%. Using the same ROI shapefile but a different image classified through the VV-VH index, the resulting matrix showed that only 65 out of 75 flooded pixels corresponded with the reference image and 67 out of 75 non-flooded pixels checked as well. With those values, the overall accuracy is 88.00%. Of the three combinations utilized to generate flood map extent using the Support Vector Machine classifier, the most promising result was from the VV+VH index. 68 out of 75 pixels were assigned to the flooded area and 72 out of 75 in the generated map using SVM classifier with VV+VH index. These values yielded an overall accuracy of 93.33%.

Table 3. Error Matrix of Flood Map using SVM Classification in NDDPI and VV-VH Combination

nSVM (NDDPI)					**SVM (VV-VH)**				
	Actual Flooded	Actual Not Flooded	Total	User Accuracy		Actual Flooded	Actual Not Flooded	Total	User Accuracy
Model Flooded	66	5	71	92.96	Model Flooded	65	8	73	89.04
Model Not Flooded	9	70	79	88.61	Model Not Flooded	10	67	77	87.01
Total	75	75	150		Total	75	75	150	
Producer's Accuracy	88.00	93.33			Producer's Accuracy	86.67	89.33		
Overall Accuracy	**90.67%**				**Overall Accuracy**	**88.00%**			

Table 4. Error Matrix of Flood Map using SVM Classification in VV+VH Combination.

SVM (VV+VH)				
	Actual Flooded	Actual Not Flooded	Total	User Accuracy
Model Flooded	68	3	71	95.77
Model Not Flooded	7	72	79	91.14
Total	75	75	150	
Producer's Accuracy	90.67	96.00		
Overall Accuracy	**93.33%**			

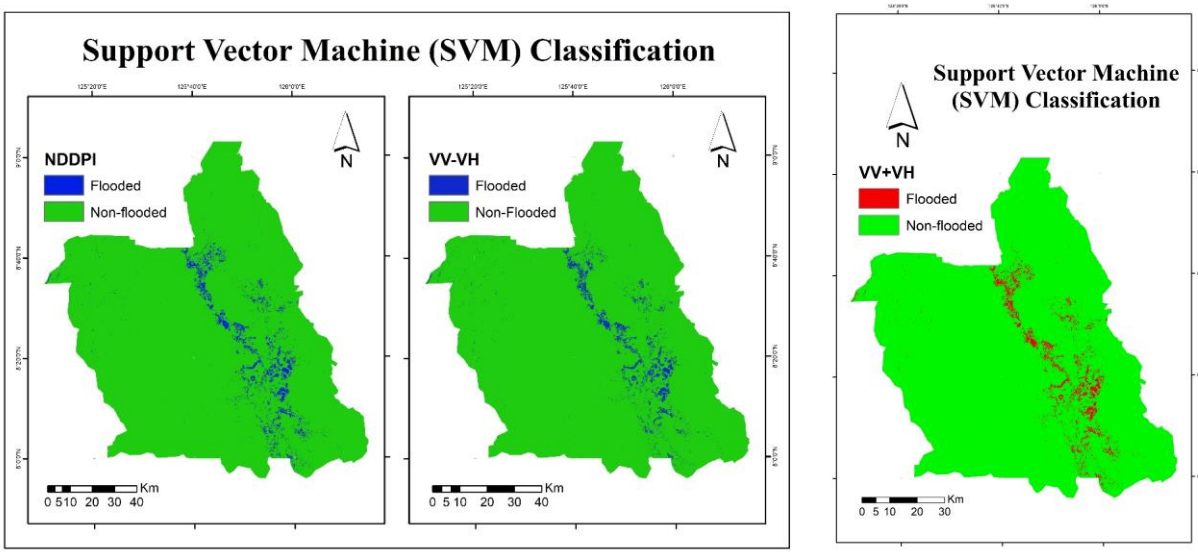

Figure 5. Flood Map using SVM Classification in NDDPI, VV-VH Combination, and VV+VH Combination.

3.2.3 RF Classification Maps and Error Matrices

The same parameters are used, the flooded and non-flooded layers (actual and model-derived). For VV-VH, there were very few pixels classified as flooded. As shown in the error matrix, 64 out of 75 pixels of the flooded area classified from the reference image were correctly classified using this combination of indices, 66 out of 75 pixels were correctly classified for the non-flooded. Even with high user and producer accuracy, the overall accuracy for this combination is only 86.67%. In NDDPI combination, 63 out of 75 flooded pixels corresponded with the training points, and 68 out of 75 pixels were checked correctly with non-flooded areas in reference image with an overall accuracy of 87.33%. The results from error matrices of the flood map extent generated through the Random Forest classifier show low accuracy values compared to the pixel values obtained from the raw images. With that, the highest overall accuracy is only 88.67% which is from the VV+VH index. The error matrix for VV+VH indices using Random Forest classifier shows that 65 out of 75 pixels of the flooded areas in the generated training ROIs were correctly classified in this index combination. Meanwhile, the non-flooded classified 68 out of 75 non-flooded sites obtaining an overall accuracy of 88.67%, the highest among the combinations applied to the random forest classifier.

Table 5. Error Matrix of Flood Map using RF Classification in VV-VH Combination and NDDPI.

RF (VV-VH)					**RF (NDDPI)**				
	Actual Flooded	Actual Not Flooded	Total	User Accuracy		Actual Flooded	Actual Not Flooded	Total	User Accuracy
Model Flooded	64	9	73	87.67	Model Flooded	63	7	70	90.00
Model Not Flooded	11	66	77	85.71	Model Not Flooded	12	68	80	85.00
Total	75	75	150		Total	75	75	150	
Producer's Accuracy	85.33	88.00			Producer's Accuracy	84.00	90.67		
Overall Accuracy	**86.67%**				**Overall Accuracy**	**87.33%**			

Table 6. Error Matrix of Flood Map using RF Classification in VV+VH Combination.

RF (VV+VH)				
	Actual Flooded	Actual Not Flooded	Total	User Accuracy
Model Flooded	65	7	72	90.28
Model Not Flooded	10	68	78	87.18
Total	75	75	150	
Producer's Accuracy	86.67	90.67		
Overall Accuracy	**88.67%**			

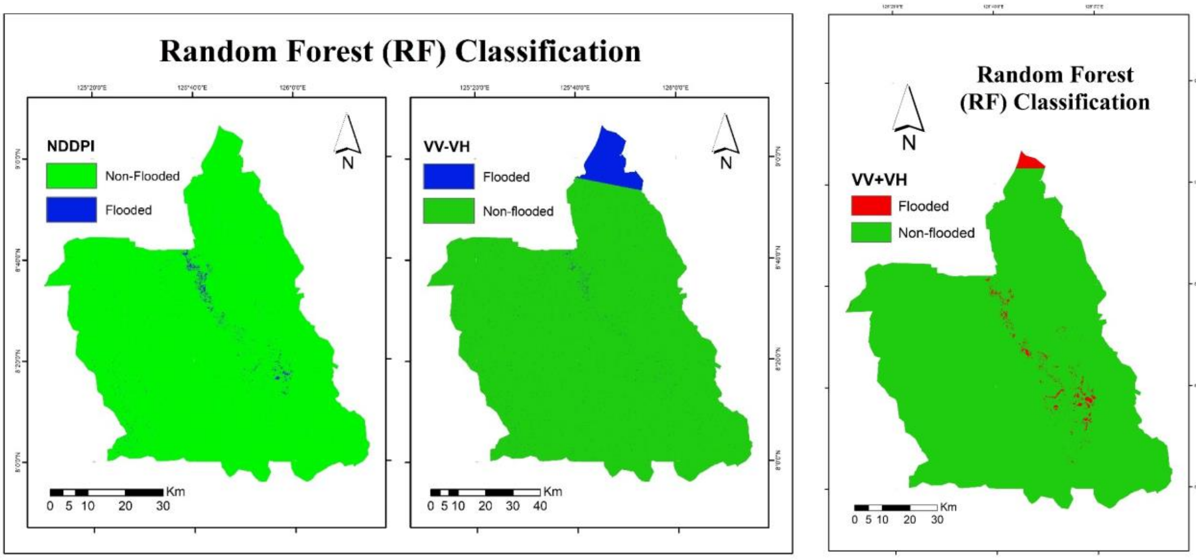

Figure 6. Flood Map using SVM Classification in NDDPI, VV-VH Combination, and VV+VH Combination.

3.3 Flood Extent Map of Agusan del Sur in 2020

Figure 7a shows the terrain corrected and pre-processed 2020 flooded image in VH polarization. The highest value shown is 2.63704, and the lowest value representing the water bodies is -30.4672. Figure 7b depicts the study area in VV polarization with values that range from -29.1203 to 11.4974. These images were used as inputs to generate the final map for the 2020 flooding event using SVM and the VV+VH parameter. The flood map is generated through SVM classification with the VV+VH parameter (see Figure 8). The flooded areas are shown in blue, where most of the pixels occupy San Francisco and Bunawan, Agusan del Sur. Table 7 presents the locations in Agusan del Sur affected by Tropical Storm Krovanh last December 2020. The most affected area was San Francisco, with a total flooded area of 26.7337 km^2.

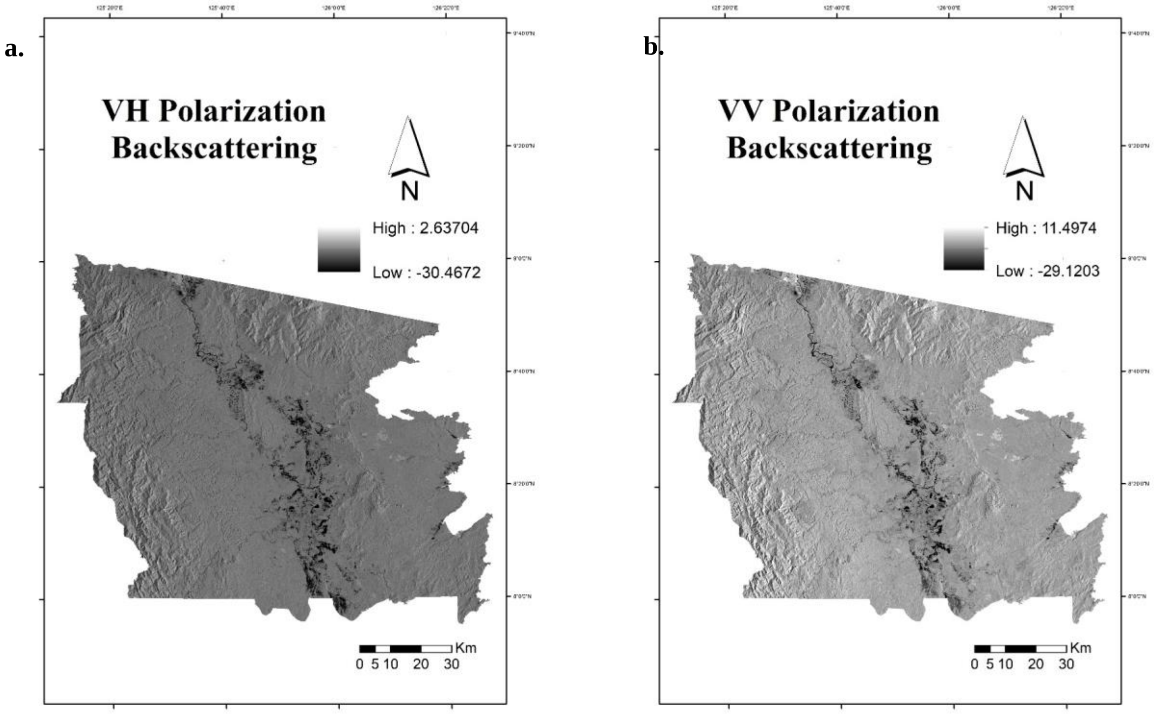

Figure 7. Pre-processed Image of the 2020 Flood Event in (a) VH polarization and (b) VV polarization.

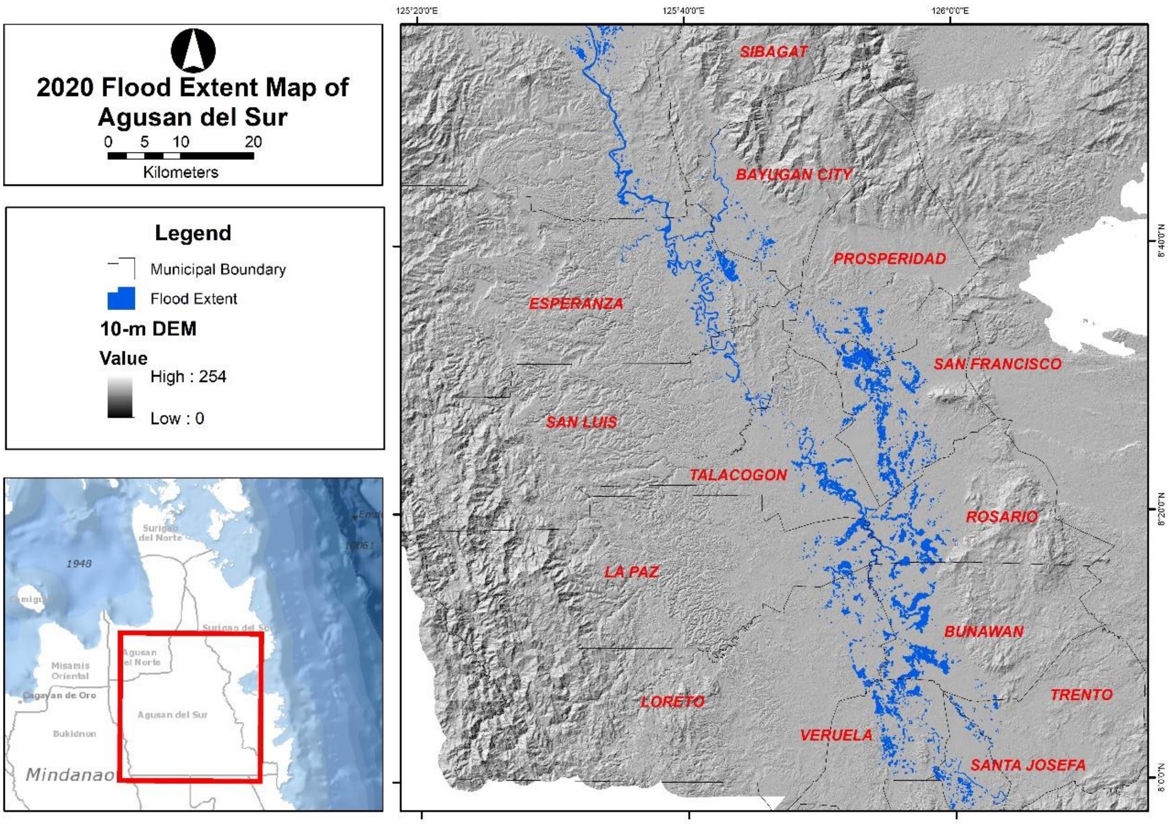

Figure 8. Flood Map of Agusan del Sur.

Table 7. Flooded Areas in Agusan del Sur.

Municipality	**Flooded Areas (km^2)**
Bunawan	21.70298
City of Bayugan	2.745126
Esperanza	14.26877
La Paz	6.912236
Loreto	5.462286
Prosperidad	6.772985
Rosario	13.12684
San Francisco	26.7337
San Luis	2.997277
Santa Josefa	4.724845
Sibagat	0.018713
Talacogon	10.49076
Trento	1.038619
Veruela	12.67637

4. CONCLUSION AND RECOMMENDATION

4.1 Conclusion

This study aimed to assess the efficiency of different machine learning classifiers in mapping flood extent and compare the results from the traditional flood extent mapping method. Binarization, which can be performed directly on the SNAP software, was used as the conventional method. At the same time, the machine learning classifiers included Random Forest and Support Vector Machine classifiers where three combinations of Sentinel-1 backscattering values were used, namely VV+VH, VV-VH, and NDDPI (VV-VH/VV+VH) as inputs to produce three maps for each classifier. The accuracy assessment results using the validation points revealed using SVM classifier with the VV+VH combination to have the highest accuracy, which is 93.33%.

5. REFERENCES

Bosello, F., Iglesias, A., Termansen, M., Jeuken, A., Winsemius, H., De Cian, E., Chiabai, A., and Garrote, L., 2018. "Economy-Wide Impacts of Climate Mitigation and Adaptation Strategies Across European Regions".

Conde, F. C., and Muñoz, M. d. M., 2019. "Flood Monitoring Based on the Study of Sentinel-1 SAR Images: The Ebro River Case Study," Flood Risk Assessments: Applications and Uncertainties, no. 2454, pp. 11-12.

Davies, R., 2020. "Philippines – 8 Killed in Floods and Landslides, Thousands Evacuate," floodlist, 20 December 2020. Retrieved March 27, 2021 from http://floodlist.com/asia/philippines-floods-caraga-cagayan-december-2020.

Filipponi, F. 2019. "Sentinel-1 GRD Preprocessing Workflow †," in Istituto Superiore per la Protezione e la Ricerca Ambientale, Rome, 2019.

Mandal, D., 2019. "Sentinel-1 SLC preprocessing workflow for polarimetric applications: A generic," Preprints, 30 November 2019.

Rana, V. K. and Suryanarayana, T., 2019. "Evaluation of SAR speckle filter technique for inundation mapping," Remote Sensing Applications: Society and Environment, vol. 16, no. 100271.

Torino, P. D., Palomba, G., Farasin, A., and Rossi, C., 2020. "Sentinel-1 Flood Delineation with Supervised Machine Learning," August 2020.