VALIDATION OF AUTOMATIC FLOOD DETECTION ALGORITHM IN GOOGLE EARTH ENGINE CLOUD PLATFORM USING SYNTHETIC APERTURE RADAR DATA AND RANDOM FOREST METHOD

Sa'ad Ibrahim[1]

[1]Department of Geography, School of Art and Social Sciences, Adamu Augie College of Education, Argungu PMB 1012, Kebbi State, Nigeria

*Correspondence: saadarg1@yahoo.com

DOI : https://www.doi.org/10.56726/IRJMETS35943

ABSTRACT

A great number of communities in Africa are threatened by flood disasters. While mapping the spatial extent of flooding is necessary for emergency response as well as for adaptation decisions, accurately mapping the extent of these floods across regions requires a significant amount of training data, usually obtained via field surveys. Field surveys can be time-consuming, costly and impractical in inaccessible terrain. This necessitates the application of automatic algorithms for flood detection. Therefore, it is important to assess the effectiveness of the current automated techniques to guarantee their precision. This study employed the RF method to delineate flood extent as a basis for the validation of automatic flood detection algorithms using Sentinel-1 data within the Google Earth Engine (GEE) platform. RF's overall accuracy was 99% while Otsu's automatic flood detection algorithms were 76%. The validation results highlight how combining machine learning techniques with SAR data might improve flood monitoring and aid in disaster management efforts.

1. INTRODUCTION

Globally, there is a collective effort to tackle environmental challenges to promote sustainable living. In Nigeria, for example, conservation policy aims to maintain a high-quality environment to promote good health and well-being. It also aims to conserve and responsibly utilise natural resources for the benefit of current and future generations. The policy supports the development of detailed land suitability/capability inventories, comprehensive land classifications, assessments of current land use practices, water damage prevention through flood control, identification of causes and extent of damage, and the creation of a regulatory framework for sustainable land use (NESREA, 2017). However, Mashi et al. (2019) emphasized that Nigeria's emergency management legislation is weak in providing necessary action plans, highlighting that effective legislation and action plans should cover prevention, mitigation, preparedness, response, resilience, and recovery comprehensively. Flooding is a serious disaster that requires urgent attention to safeguard food security.

One of the key approaches to addressing this issue is tracking the magnitude or extent of the event. This is a crucial step in quantifying its impact on land use. Remote sensing techniques are also quite viable for flood detection and easily accessible and transferable to climate change research partly due to the advancement in Earth observation coupled with the proliferation of satellite missions (Jones et al., 2023). For example, machine learning techniques such as decision trees, K-means clustering, K-nearest neighbour (kNN), Support vector machines (SVN), neural networks and random forests (RF) have become widely used for land use land cover (LULC) and flood and non-flood classifications. RF has shown promise in many previous studies (Balzter et al., 2015; Ibrahim, 2023; Zare & Schumann, 2021).

Billah et al. (Billah et al., 2023) perform land use mapping to assess rapid flood damage using Sentinel-1 and Sentinel-2 data by comparing random forest (RF) maximum likelihood classification methods. Another recent study by Ganjirad and Delavar (2023) provides a comparative analysis of RF and support vector machines (SVN) to estimate flood risk in Louisiana (United States) by integrating spectral indices (e.g., Modified Normalized Difference Water Index (MNDWI) and topographical features. results show that RF has a low mean absolute (MAE) error (0.04) compared to SVN (0.09). Loukika et al. (Loukika et al., 2021) classified LULC in a river basin by comparing SVM, RF, and CART based on Landsat and Sentinel-2 data. The results indicate that RF outperformed both classification methods, achieving accuracy rates of 94.85% and 95.8% compared to SVM's 90.88% and 93.8% and CART's 82.88% and 86.4% for Landsat and Sentinel-2, respectively. Ren et al. (Ren et al., 2024) evaluated flood susceptibility using RF, XGBoost, SVM, and ANN. The findings indicate that RF achieved the highest accuracy, 0.87. XGBoost followed closely at 0.84, outperforming both ANN (0.83) and SVM (0.82). The RF approach is highly robust because of its numerous advantages over other machine learning (Maxwell et al., 2018). Flooding is becoming more frequent and unpredictable due to uncertainty in climate extremes (Cred, 2020; Tellman et al., 2021), upsetting established crop cycles and putting yields at risk (Raza et al., 2019) and further worsening, food security, and subsistence and livelihoods of local communities (FAO, 2018).

Remote sensing approaches for mapping the spatial extent of flood events are highly viable, but the need for extensive training data over large areas through field campaigns is quite costly. At the same time, the advantages provided by the Google Earth Engine cloud computing platform present a viable option for mapping land use and land cover (LULC). This study aims to validate one of the most widely used automatic flood detection algorithms using Sentinel-1 Synthetic Aperture Radar (SAR) data within the Google Earth Engine (GEE) platform.

2. Material and method

2.1 Study area

The study area is located in Argungu, Kebbi State, in north-western Nigeria. This region is known for the cultivation of rice, as well as other cereals and vegetable crops, with a large portion of the population engaged in rice farming during both the rainy and dry seasons. However, annual flood events significantly impact rice production, making it challenging for farmers to sustain their livelihoods. Mapping the spatial extent of these floods is a crucial step in addressing this issue.
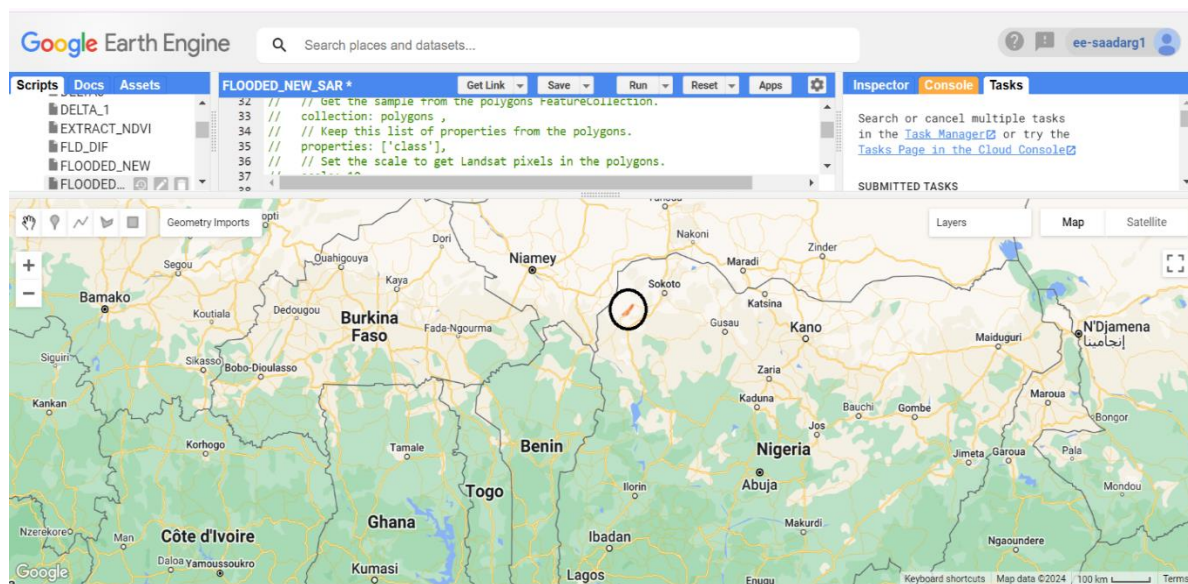
Figure 1: the study area based on a high-resolution true-color imagery

2.2 Sentinel-1 data

Sentinel-1 data, which is provided by the European Space Agency's Copernicus Programme, is a dual-polarization C-band Synthetic Aperture Radar (SAR) data. In this study, publicly available Sentinel-1 data on Google Earth Engine (GEE) will be utilized. The data is available in both single polarization (VV) and dual polarization (VV and VH). Figure 2 indicates the Sentinel-1 backscatter in VV polarization.
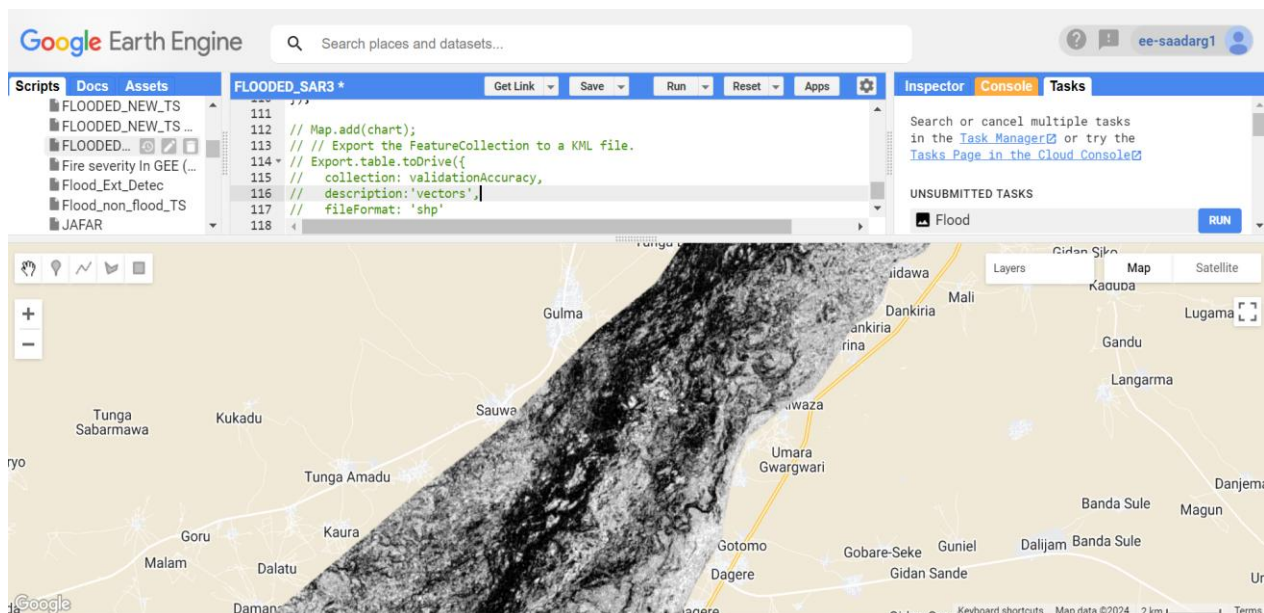
Figure 2: Sentinel-1 backscatter (VV) polarization

2.3 Method

2.3.1 RF classification

For the Random Forest (RF) classification, training areas were carefully selected based on high-resolution imagery, and 100 decision trees were chosen for the classification process. The classification was performed within the Google Earth Engine (GEE) platform, leveraging its cloud computing capabilities to efficiently analyze the large spatial datasets

2.3.2 Automatic water detection method using Otsu algorithm.

The Otsu algorithm is commonly used for automatic image thresholding, which is useful for detecting water in remote sensing images. This method finds the threshold value that minimizes intra-class variance (or, equivalently, maximizes inter-class variance) between two classes. The method usually converts the image to grayscale, calculates the histogram of the grayscale image and estimates probabilities. Based on these scenarios, it computes the mean pixel values for the two classes (Cao et al., 2019; Sang et al., 2024; Xu et al., 2011).

2.3.3 Validation

To validate the Sentinel-I data, RF classification of flooded and non-flooded areas was used. 500 pixels were extracted from the Sentinel-1 RF classified image and their coincidence pixels were extracted from the Sentinel-1 Otsu automatic flood algorithm for validation. This method was tested for its reliability by computing producers, consumers, and overall accuracy.

2. RESULTS

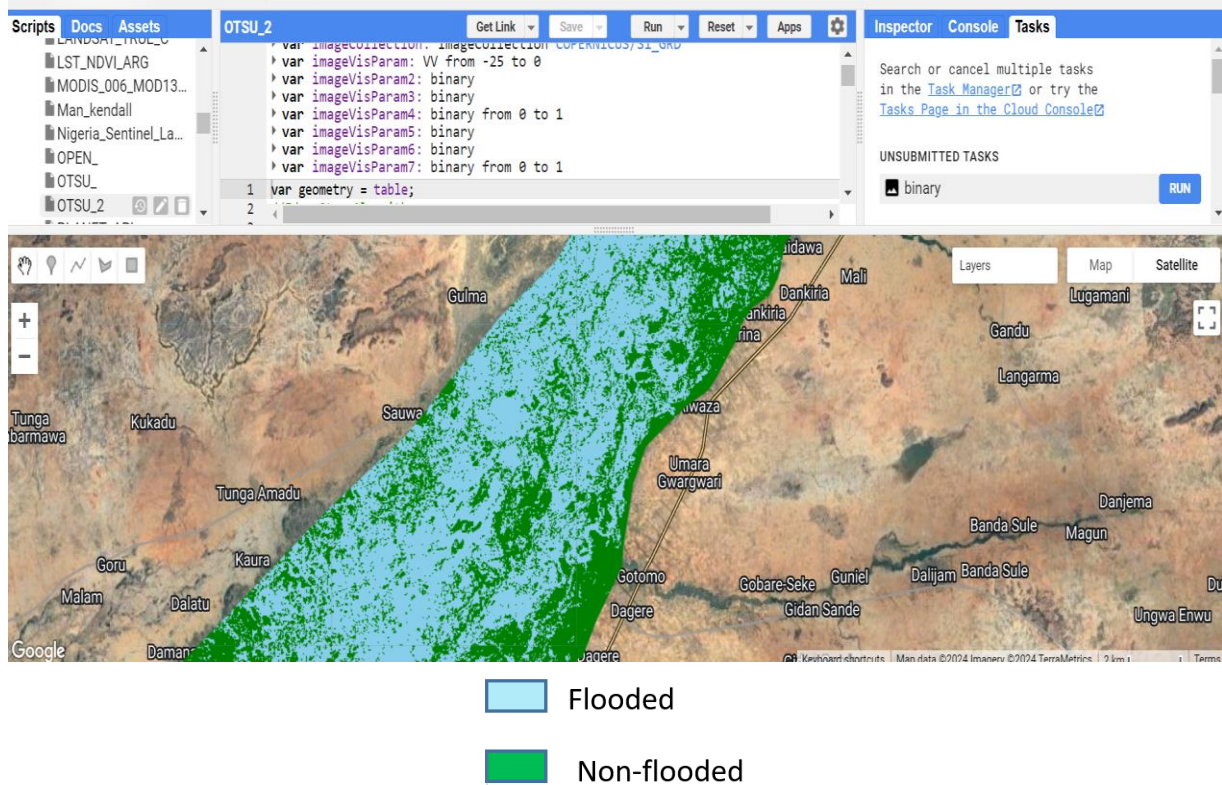### 3.1 Sentinel-1 flood classification based on RF

Figure 3 indicates results derived using the RF classification method for flood detection. The findings show spatial patterns that allow flooded and non-flooded areas to be identified. Areas affected by floods are shown in cyan, denoting locations where water has accumulated or remained as a result of flooding incidents. Lower backscatter values in the SAR data, which normally indicate surfaces covered in water, may be indicative of these places. However, the places that are not inundated are indicated in green, indicating that during the observation time, these areas were either dry or unaffected by the flood. The variable importance of the RF shows that VV polarization is more important to the model (Figure 3).
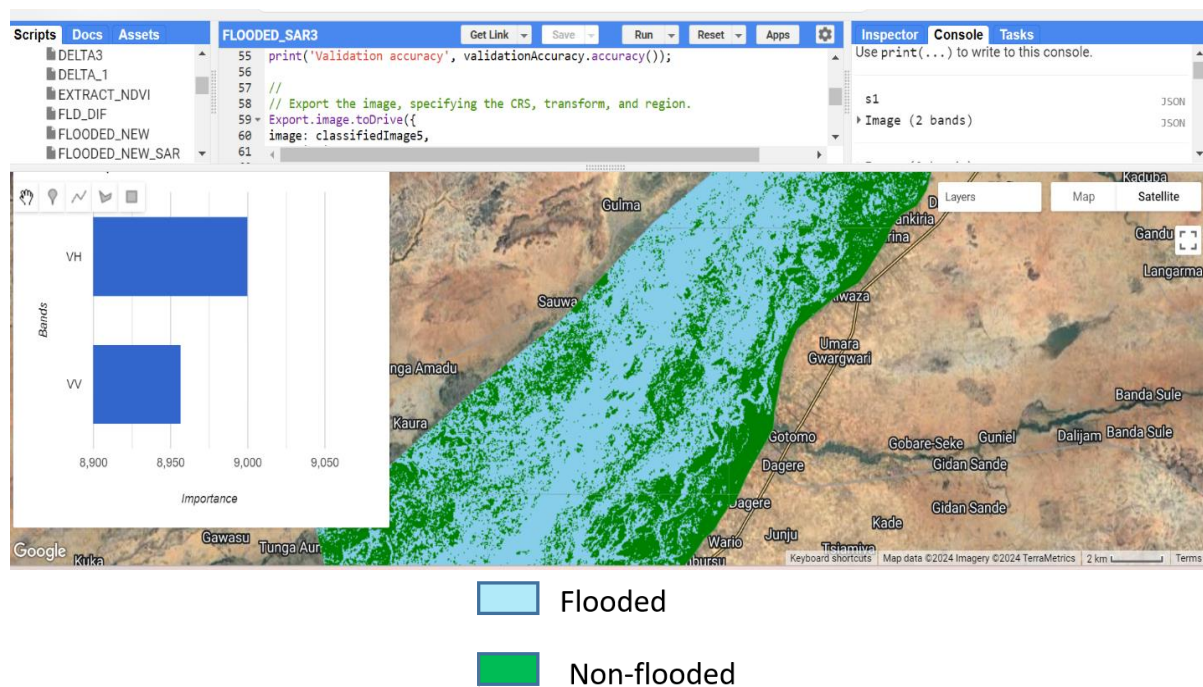
Figure 3: Sentinel-1 RF classification showing flooded and non-flooded sites and the variable importance based on feature selection

3.2 Sentinel-1 flood classification based on Otsu

Similarly, Figure 4 which was derived using Otsu's algorithm shows spatial patterns that allow flooded and non-flooded areas to be identified. Areas affected by floods are shown in cyan, denoting locations where water has accumulated or remained as a result of flooding incidents. Lower backscatter values in the SAR data, which normally indicate surfaces covered in water, may be indicative of these places. However, the places that are not inundated are indicated in green, indicating that during the observation time, these areas were either dry or unaffected by the flood.

Flooded

Non-flooded

Figure 4: Sentinel-1 automatic water detection-based Otsu algorithm showing flooded and non-flooded sites

3.3 Accuracy assessment

Table 1 presents the accuracy assessments and validation results for the Sentinel-1-based RF classification, alongside the validation of Otsu's water detection algorithm. The RF classification showed excellent performance, with both the user's and producer's accuracy for the flooded and non-flooded classes reaching 99%, and an overall accuracy of 99%. In comparison, the validation of Otsu's water detection algorithm, based on the RF results, was slightly lower. The user's and producer's accuracy for Otsu's method were 74% and 78%, respectively, with an overall accuracy of 76%.

Table 1: Accuracy assessment/validation of flooded/non-flooded classifications

Accuracy metric	RF Sentinel-1	Otsu's Sentinel-1
Users	0.99	0.74
Producers	0.99	0.78
Overall	0.99	0.76

3. DISCUSSION

The results of this study demonstrate the effectiveness of using Sentinel-1 data and the RF method for flood detection within the GEE platform. The RF-based classification of flooded and non-flooded areas demonstrated high accuracy, with both user and producer accuracy reaching 99% and overall accuracy of 99% (Table 1). These results underline the robustness of the RF method in distinguishing flooded and non-flooded areas and make it a highly reliable tool for flood detection (Table 1 and Figure 3). In a recent study, by Shilengwe et al. (2023) who assessed flood damage using radar and optical sensors in GEE indicated that their flood extent delineation based on RF achieved an overall accuracy of 95% The Otsu-based floodplain classification algorithm provided lower accuracy in comparison. The user and producer accuracy for the Otsu method was 74% and 78% respectively, with an overall accuracy of 76% (Table 1). Although the Otsu method can still be useful, its relatively lower performance compared to the RF approach suggests that it is suitable for large-scale or complex flood detection tasks (Figure 4 and Table 1). Our results corroborate the findings of Vanama et al. (2020) who mapped large flood areas using the GEE4FLOOD framework in India. Their results indicate promising accuracy with 82% overall accuracy and 78.5% accuracy for flood class alone compared to the ground truth data. However, it is recommended that further studies should compare different automatic detection algorithms (e.g., K-means and the unsupervised Gaussian Mixture Model) to assess their applicability. It is evident from the spatial patterns of flooded and non-flooded areas are distinct from one another. The flood extent is reflected in these spatial patterns, emphasizing the significance of using algorithms such as RF for precise flood mapping.

4. CONCLUSION

This study employed the RF method to delineate flood extent as a basis for the validation of automatic flood detection algorithms using Sentinel-1 data within the GEE platform. RF's overall accuracy was 99% while Otsu's automatic flood detection algorithms were 76%. The validation results highlight how combining machine learning techniques with SAR data might improve flood monitoring and aid in disaster management efforts.

5. REFERENCES

[1] Balzter, H., Cole, B., Thiel, C., & Schmullius, C. (2015). Mapping CORINE Land Cover from Sentinel-1A SAR and SRTM Digital Elevation Model Data using Random Forests. Remote Sensing, 7(11), 14876-14898. https://www.mdpi.com/2072-4292/7/11/14876

[2] Billah, M., Islam, A. K. M. S., Mamoon, W. B., & Rahman, M. R. (2023, 2023/04/01/). Random forest classifications for land use mapping to assess rapid flood damage using Sentinel-1 and Sentinel-2 data. Remote Sensing Applications: Society and Environment, 30, 100947. https://doi.org/https://doi.org/10.1016/j.rsase.2023.100947

[3] Cao, H., Zhang, H., Wang, C., & Zhang, B. (2019). Operational Flood Detection Using Sentinel-1 SAR Data over Large Areas. Water, 11(4), 786. https://www.mdpi.com/2073-4441/11/4/786

[4] Cred, U. (2020). Human Cost of Disasters. An Overview of the last 20 years: 2000–2019. CRED, UNDRR, Geneva.

[5] FAO. (2018). The impact of disasters and crises on agriculture and food security. Report.

[6] Ganjirad, M., & Delavar, M. R. (2023). FLOOD RISK MAPPING USING RANDOM FOREST AND SUPPORT VECTOR MACHINE. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., X-4/W1-2022, 201-208. https://doi.org/10.5194/isprs-annals-X-4-W1-2022-201-2023

[7] Ibrahim, S. a. (2023). Improving Land Use/Cover Classification Accuracy from Random Forest Feature Importance Selection Based on Synergistic Use of Sentinel Data and Digital Elevation Model in Agriculturally Dominated Landscape. Agriculture, 13(1), 98. https://www.mdpi.com/2077-0472/13/1/98

[8] Jones, A., Kuehnert, J., Fraccaro, P., Meuriot, O., Ishikawa, T., Edwards, B., Stoyanov, N., Remy, S. L., Weldemariam, K., & Assefa, S. (2023, 2023/06/08). AI for climate impacts: applications in flood risk. npj Climate and Atmospheric Science, 6(1), 63. https://doi.org/10.1038/s41612-023-00388-1

[9] Loukika, K. N., Keesara, V. R., & Sridhar, V. (2021). Analysis of Land Use and Land Cover Using Machine Learning Algorithms on Google Earth Engine for Munneru River Basin, India. Sustainability, 13(24), 13758. https://www.mdpi.com/2071-1050/13/24/13758

[10] Mashi, S. A., Oghenejabor, O. D., & Inkani, A. I. (2019, 2019/02/01/). Disaster risks and management policies and practices in Nigeria: A critical appraisal of the National Emergency Management Agency Act. International Journal of Disaster Risk Reduction, 33, 253-265. https://doi.org/https://doi.org/10.1016/j.ijdrr.2018.10.011

[11] Maxwell, A. E., Warner, T. A., & Fang, F. (2018, 2018/05/03). Implementation of machine-learning classification in remote sensing: an applied review. International Journal of Remote Sensing, 39(9), 2784-2817. https://doi.org/10.1080/01431161.2018.1433343

[12] NESREA. (2017). National Policy on Environment. Retrieved 26/02/2024 from Raza, A., Razzaq, A., Mehmood, S. S., Zou, X., Zhang, X., Lv, Y., & Xu, J. (2019). Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review. Plants, 8(2), 34. https://www.mdpi.com/2223-7747/8/2/34

[13] Ren, H., Pang, B., Bai, P., Zhao, G., Liu, S., Liu, Y., & Li, M. (2024). Flood Susceptibility Assessment with Random Sampling Strategy in Ensemble Learning (RF and XGBoost). Remote Sensing, 16(2), 320. https://www.mdpi.com/2072-4292/16/2/320

[14] Sang, G., Wang, X., & Zhang, J. (2024). Improved Otsu Theory of Image Multi-threshold Segmentation by Incorporating Ant Colony Algorithm. Informatica, 48(9).

[15] Shilengwe, C., Nyimbili, P. H., Msendo, R., Banda, F., Mukupa, W., & Erden, T. (2023). Synthetic Aperture Radar and Optical Sensor Techniques Using Google Earth Engine For Flood Monitoring and Damage Assessment–A Case Study of Mumbwa District, Zambia. Zambia ICT Journal, 7(1), 7-15.

[16] Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A., & Slayback, D. A. (2021, 2021/08/01). Satellite imaging reveals increased proportion of population exposed to floods. Nature, 596(7870), 80-86. https://doi.org/10.1038/s41586-021-03695-w

[17] Vanama, V. S. K., Mandal, D., & Rao, Y. (2020). GEE4FLOOD: rapid mapping of flood areas using temporal Sentinel-1 SAR images with Google Earth Engine cloud platform. Journal of Applied Remote Sensing, 14(3), 034505. https://doi.org/10.1117/1.JRS.14.034505

[18] Xu, X., Xu, S., Jin, L., & Song, E. (2011). Characteristic analysis of Otsu threshold and its applications. Pattern recognition letters, 32(7), 956-961.

[19] Zare, M., & Schumann, G. J. P. (2021). Chapter 15 - Emerging Techniques in Machine Learning for Processing Satellite Images of Floods. In G. J. P. Schumann (Ed.), Earth Observation for Flood Applications (pp. 321-336). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-819412-6.00015-8