Annals of Emerging Technologies in Computing (AETiC)
Vol. 6, No. 4, 2022

Research Article

An Empirical Approach to Monitor the Flood-Prone Regions of North India Using Sentinel-1 Images

Mohammed Siddique[1,2,*], Tasneem Ahmed[1] and Mohd Shahid Husain[3]

[1]Department of Computer Application, Integral University, India
siddiquep@student.iul.ac.in; tasneemfca@iul.ac.in
[2]Majan University College, Oman
muhammed.siddique@majancollege.edu.om
[3]CAS-Ibri University of Technology and Applied Sciences, Muscat, Oman
mshahid.ibr@cas.edu.om
*Correspondence: siddiquep@student.iul.ac.in

Received: 30th march 2022; Accepted: 14th September 2022; Published: 1st October 2022

Abstract: Floods in India is among the perilous natural disasters with a high impact on its economic sectors. One of the critical factors to handle such hazardous events is monitoring the affected areas and changes in flood patterns. Flood management is a very complex issue, largely owing to the growing population and investments in flood-affected regions. Satellite images especially Synthetic Aperture Radar (SAR) images are very useful and effective because SAR images are acquired day and night in all types of weather conditions. This research analyzes a combination of machine learning algorithms implemented on Sentinel-1A (SAR) data using supervised classification techniques to monitor the flooded areas in the North Indian region. Random Forest (RF) and the K-nearest neighbour (KNN) classification is applied to classify the different land covers such as water bodies, land, vegetation, and bare soil land covers. The outcomes of the presented work depict that the SAR data provides efficient information that helps in monitoring the flooded extents and the analysis shows that Sentinel-1 images are quite effective to detect changes in flood patterns in urban, vegetation, and regular water areas of the selected regions. The distribution of flooded areas was 16.6% and 16.8% in the respective region which is consistent with the resultant images of the proposed approach using RF and KNN classifiers. The obtained results indicate that both classifiers used in the work generate higher classification accuracy. These classifiers define the potential of multi-polarimetric SAR data in the classification of flood-affected areas. For a thorough evaluation and comparison, the RF and KNN are utilized as benchmarked classifiers. The classification accuracies based on the investigated results from the three SAR images can be improved by incorporating spatial and polarimetric features. In the future, the deep-learning classification techniques using ensemble strategies are expected to achieve an increased accuracy level with an overall classification strategy of urban and vegetation mapping.

Keywords: ***Flood mapping; Image Classification; K-Nearest Neighbour; Random Forest; SAR; Sentinel-1***

1. Introduction

The rescue operations and assessing the damages caused by the floods can be carried on SAR data (Sentinel-1A) images downloaded before, during, and after the floods have occurred. The north Indian regions are vulnerable to floods and the flood management practices are required to be regularly reviewed for effectiveness [1]. The cities and villages of India especially in the state of Uttar Pradesh and Bihar in particular are at risk every year. It is particularly very important to identify specific patterns by means of monitoring and analyzing the situation using credible data sources. SAR image processing is a traditional approach used by flood management agencies for a long time to gather imminent pieces of evidence to

Mohammed Siddique and Tasneem Ahmed and Mohd Shahid Husain, "An Empirical Approach to Monitor the Flood-Prone Regions of North India Using Sentinel-1 Images", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 1-14, Vol. 6, No. 4, 1st October 2022, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2022.04.001, Available: http://aetic.theiaer.org/archive/v6/v6n4/p1.html.

design efficient disaster management control mechanisms. There is an increase in the frequency of floods occurring every year and one of the factors affecting the crisis is the effects of climate change. In August 2020, most of the villages and districts in the cities of Gorakhpur, Ayodhya, and Basti in the northeast Indian region were severely hit by the floods. The year 2020 and 2021 saw severe flooding with reportedly many river banks being broken. Flood-affected areas' waterlogged images are shown in Figure 1. The SAR data from the Sentinel-1 series satellite offers crucial information services to tackle calamities. Additionally, many other natural calamities such as earthquakes, tsunamis, avalanches, and wildfires can also be monitored.

Figure 1. Photos at the time of Flooding in Uttar Pradesh, India, August 2021[1]

The SAR data processing helps in identifying flood-prone regions and monitoring the flooded areas. This could further provide an opportunity to create robust flood monitoring and early alert systems. With effective SAR data processing, satellite image-based surveillance and monitoring of the distinct water, urban, vegetation, and bare soil areas can be explored.

Artificial Intelligence (AI) can essentially reduce errors in functionalities and provides logical reasoning to make applied models more accurate. Automated computer systems have functionalities that can execute intelligently programmed tasks. Sophisticated techniques like Data mining (DM), Machine Learning (ML), and Deep Learning (DL) are the subsets of AI which make use of a wide range of data sets consisting of audio, video, or images to build such an automated system [2]. The image analysis to extract visual features such as texture, color function as input, training and validated information in implementing supervised classifications [3]. It is very well known that classification plays an important role in monitoring flood-affected areas. There are various classification methods like regression, Naïve Bayes, ANN, RF, Support Vector Machine (SVM), and **KNN** are available [4-6]. Juliana Maria *et al*. have conducted land use and land cover (LULC) mapping by using RF and SVM classifiers on Sentinel-1A data in a Brazilian Amazon region for evaluating backscattering, interferometry, and polarimetry approaches. Their results achieve global accuracy of 82.7% in RF and 74.5% in SVM classification [7]. As stated by Heydari and Mountrakis [8], the performance of RF, KNN, and SVM classifiers using Sentinel-2 with several training data was evaluated. They performed evaluations to review the effects of training data on the accuracy of the results achieved. The KNN was among the best classification approaches for land use and land cover mapping. Khatami *et al*. [9] observed that the supervised classifiers RF, KNN, and SVM outperform other traditional classification approaches. Hence, RF and KNN classifiers are among the common ML algorithms used when implementing image classification techniques.

During the land-cover classification, the Sentinel and the L-band ALOS-2 PALSAR-2, datasets are pre-processed and machine learning algorithms have been implemented for image segmentation for a false-color composite RGB image generation. The Sentinel-2 and the ALOS-2 PALSAR-2 data can be efficiently used by dividing the RGB images into 500*500 pixel portions. Feature vectors of the datasets used are generated using Gaussian, Median, Scharr, and Gabor filters to train the RF, KNN, and SVM machine learning. Based on outcomes from different patches used for testing, the RF classifier outperformed when compared with the KNN and SVM classifiers. The false-color composite based optical image dataset works marginally well compared to ALOS-2 PALSAR-2 when classified [10]. The Ikonos and QuickBird format of

[1]Davies, R., 2022, "India – Hundreds of Villages Flooded in Uttar Pradesh, 9 People Dead – FloodList", Floodlist.com. Available: https://floodlist.com/asia/india-floods-uttar-pradesh-august-2021 [Accessed 11 August 2022].

optical images have four multi-spectral bands. The Ikonos covers urban regions and the QuickBird covers both rural as well as urban regions. However, due to the windy weather, the latter had noisy patterns over the Black sea. The Initial results from the evaluation using the RF algorithm gave better classification accuracy when compared with the SVM, Maximum Likelihood Classification (MLC), and Gentle AdaBoost (GAB). The RF classifier for the Ikonos image provided a 10% higher classification than the SVM which was better than the GAB algorithm. Although there was a noisy pattern, the RF provided 11% higher overall accuracy for the QuickBird image [11]. In contrast, the systems which are compact polarimetry (CP) based SAR images are recommended for monitoring larger areas as it supports a wide swath when compared with the full polarimetry (FP) systems.

With the Interferometric Wide-Swath modes, the SAR data with geometric resolution can be utilized to detect changes in the targeted regions by designing an automated high-frequency algorithm. This coherence-based algorithm is supported by the sentinel-1's six days repeat cycle to detect changes in urban regions with a high-frequency algorithm. The urban region's flood mapping can be efficiently conducted by using the SAR data. This was confirmed after qualitatively evaluating the results derived by photo-interpretation from Digital Globe and Federal Emergency Management Agency's (FEMA) inundation model [12]. In the work reported by Abazaj *et al*. an ultrasonic sensor measures the distance of the water level to detect the changes. The Arduino micro-controller processes the signals from the sensors along with a GSM module sending the data from the micro-controller to the server. A web-based flood monitoring system was developed with an SMS alert option that accesses the server data for analyzing the flood situation [13]. The Lo Ra (Long Range) and NB- IoT (Narrowband) technologies that are sensor information communication methods can be used to detect water levels. These sensor devices detect river flooding and the risk map is categorized by using the flood damage insurance maps [14].

The water level change-detection on interferometric coherence maps based on multitemporal sentinel-1 C-band VV-polarized SAR backscatter images is dependent on the temporal baseline. The Sentinel-1 C-VV data's perpendicular baseline was clearly negative between the sentinel-1 coherence and the water level. As tested by Pakhrur *et al.* [15] in marsh wetlands of China's Momoge National Nature Reserve, the outcomes of changes detected from InSAR processing showed that there is no mapping with the insite measurements. Additionally, the data is not suitable for monitoring changes in water levels [16]. The spatial inundation levels and changes periodically can be detected on the datasets transformed from GTS station datasets to a raster image and the thematic format. These are prepared from AMSU-B (Advanced Microwave Sounding Unit) and the SSM/I (Special Sensor Microwave Imager) microwave sensors. The archive is retrieved to get the spatial inundation layer matching with the estimated water levels.

This paper addresses the issue of analyzing and monitoring flooded areas by creating regions of interest (ROI) in the SAR images using a data-driven approach with the help of training samples. Two different types of classification techniques are applied demonstrating the new methodology for land cover type discrimination. Furthermore, the RGB composite of the SAR images are generated for clear flood mapping visibly generated on Google Maps to analyze the flood situation first-hand. Section 1 provides the literature review of various classification techniques, the importance of SAR data utilization, additional modules that can benefit in flood mapping, and classifying images for clear identification of water and non-water bodies. Section 2 discusses the source of SAR data and the study areas in detail. Section 3 contains the methodology through case study 1 which is an implementation of an RF classification, case study 2 with the implementation of KNN classification, K-Means classification, and in-depth analysis of the processed images. This is followed by Section 4, which provides a detailed evaluation of the results and discussion. Sections 5 and 6 exhibit the conclusions and future scope of the conducted study.

2. Study Areas and Satellite (SAR) Data Used

The study areas taken to implement various classification techniques for identifying flood-prone areas include the vulnerable regions of the city of Basti (26.8140° N, 82.7630° E), the city of Ayodhya (26.7922° N, 82.1998° E), and the city of Gorakhpur (26.7606° N, 83.3732° E).

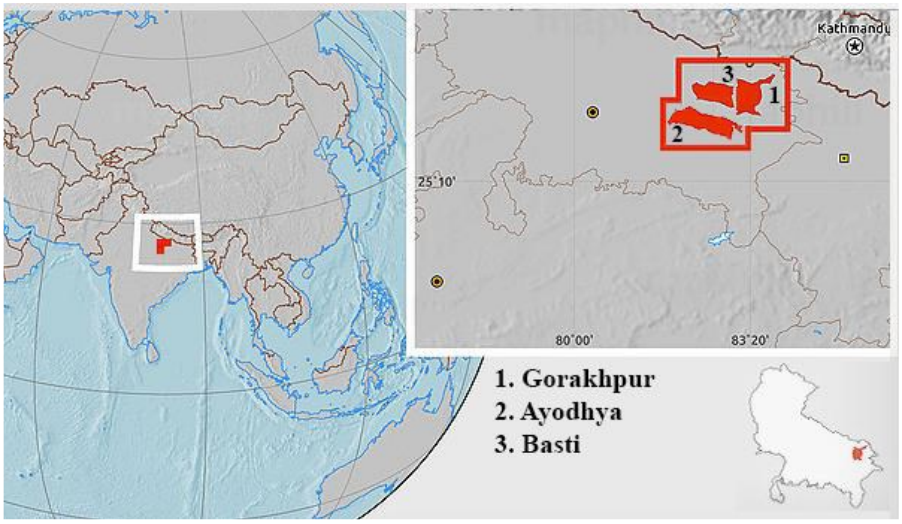


Figure 2. Extents of SAR data study area of North-East India[2]

The acquired images on the set are characterized as before the floods, during the floods, and after the floods respectively. As referred to in the disaster mapping applications, these images are categorized as an archive1 image retrieved before the floods, the crisis image retrieved during the floods, and the third image is the archive2 image retrieved shortly after the crisis. In this study, the classification of various land covers is conducted through Sentinel-1 data mainly by using the SNAP application platform, ENVI classic, and QGIS Desktop with Grass application. The details of SAR data acquisition are listed in Table 1 below.

Table 1. Sentinel-1 (SAR data) 2020 Acquisition details.

#	Acquisition-ID	Acquisition-Date	Remarks	City
1	S1A_IW_GRDH_1SDV_20200517T123002_20200517T123027_032605_03C6C2_6B3E	17-May-2020	Archive1 Image (Before floods)	Basti
2	S1A_IW_GRDH_1SDV_20200805T002754_20200805T002819_033764_03E9E5_716C	05-Aug-2020	Crisis Image (During floods)	
3	S1A_IW_GRDH_1SDV_20201219T123008_20201219T123033_035755_042F37_2718	19-Dec-2020	Archive2 Image (After floods)	
1	S1A_IW_GRDH_1SDV_20200428T123819_20200428T123844_032328_03BDC6_1856	28-April-2020	Archive1 Image (Before floods)	Ayodhya
2	S1A_IW_GRDH_1SDV_20200802T123825_20200802T123850_033728_03E8C0_481F	02-Aug-2020	Crisis Image (During floods)	
3	S1A_IW_GRDH_1SDV_20200910T002821_20200910T002846_034289_03FC44_EB16	10-Sep-2020	Archive2 Image (After floods)	
1	S1A_IW_GRDH_1SDV_20200517T123002_20200517T123027_032605_03C6C2_6B3E	17-May-2020	Archive1 image (Before floods)	Gorakhpur
2	S1A_IW_GRDH_1SDV_20200821T123008_20200821T123033_034005_03F250_FA1D	21-Aug-2020	Crisis Image (During floods)	
3	S1A_IW_GRDH_1SDV_20201219T123008_20201219T123033_035755_042F37_2718	19-Dec-2020	Archive2 Image (After floods)	

The SAR data is acquired from Copernicus Open Access Hub's Sentinel-1A constituting the geographical coordinates taken from space. Sentinel-1 images can be used for flood mapping and analysis of the required regions precisely [17]. The Sentinel's SAR sensor-based images are not affected due to the atmospheric conditions and provide data that is easier to distinguish between water and land regions. They have their own source of illumination and these Ground Range Detected (GRD) products focus on distinct and properly defined resolution.

[2]Team, M., 2022. "India: Maps", Maphill.com. Available: http://www.maphill.com/india [Accessed 11 August 2022].

3. Theoretical Background

3.1. Processing of Sentinel-1 images

The processing carried on SAR data includes image pre-processing and post-processing categories. The binarisation, band maths expressions, and color manipulation on the subset of the images create water images. Further, the radiometric calibration, speckle filtering, and terrain correction refine the images for in-depth image investigation. The SNAP tool supports batch processing that automates all the required image classification techniques and supports post-processing for image analysis.

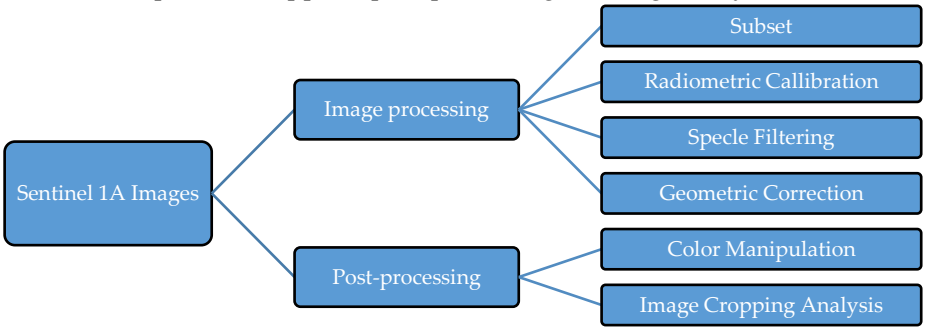
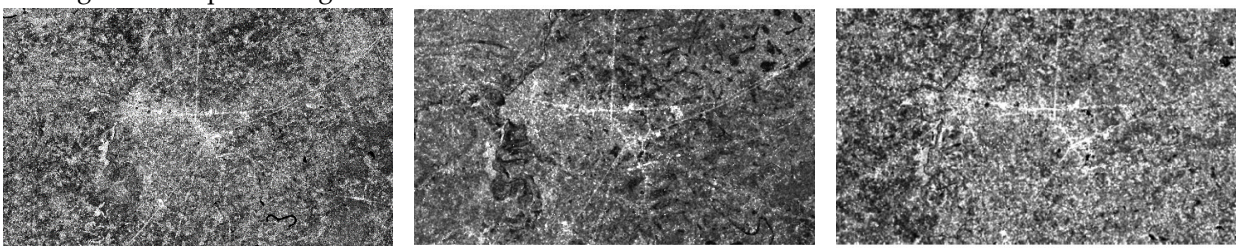


Figure 3. SAR Image pre and post processing [18]

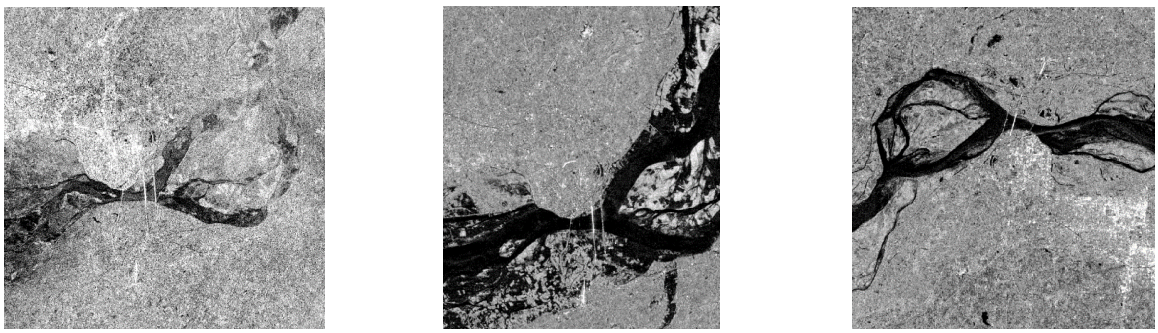
As presented in Figure 3, the processed images are necessary for the clear identification of flood-prone areas using RF, KNN classification techniques, or an RGB composite image. The accuracy of the data depends on the SAR image segmentation of its various regions to further the mapping of water areas and classify them with detected changes. The images are acquired with VH and VV bands, IW mode having 10m spatial resolution compatible to be processed in the vector / (ROI) supported image processing tools with land cover classes defined such as water bodies, land, vegetation, and the bare soil land covers.

The sub-datasets of archive and crisis images are generated by loading an input SAR data of the cities of Basti and Ayodhya. The subset pixel ranges to 11115 X and 13338 Y coordinates at the start and between 17784 X and 16759 Y coordinates at the end respectively. The radiometric calibration with a spackle filter is processed on the images. This process is implemented on all subset images as shown in Figure 4(a) - 4(i) and is converted to decibel. This supports in reducing the speckle reduce in relevance to the dimensions of the image and the processing time.



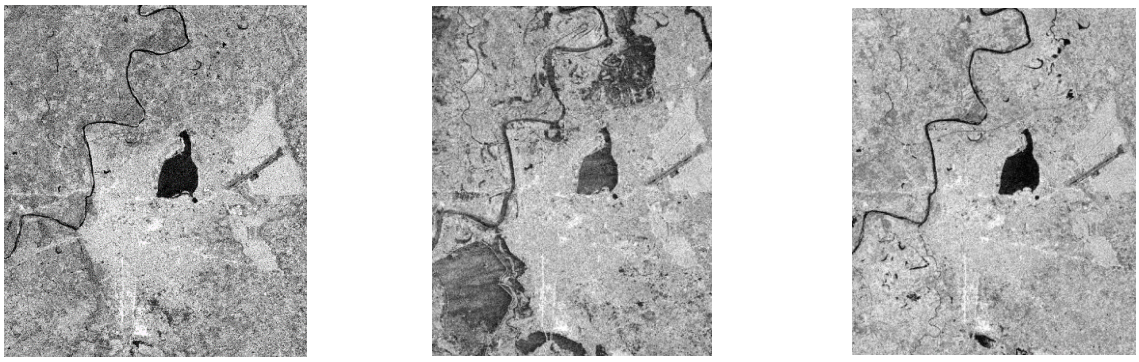
(a) Subset of Archive1 image (b) Subset of Crisis image (c) Subset of Archive2 image

Archive1, Crisis, and Archive2 image (Basti)



(d) Subset of Archive1 image (e) Subset of Crisis image (f) Subset of Archive2 image

Archive1, Crisis, and Archive2 image (Ayodhya)

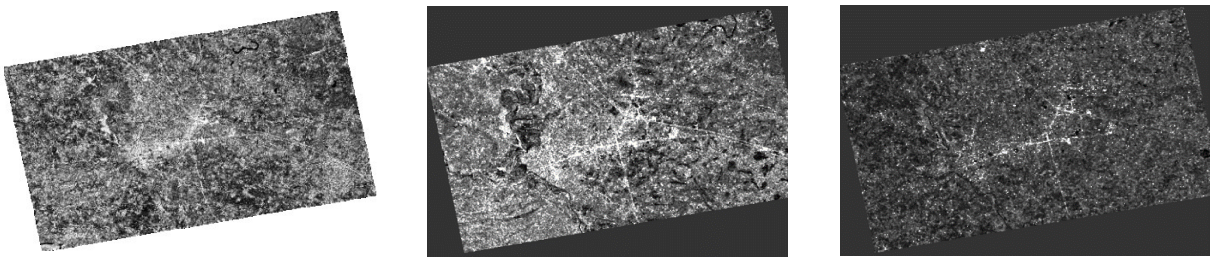


(g) Subset of Archive1 image (h) Subset of Crisis image (i) Subset of Archive2 image

Archive1, Crisis, and Archive2 image (Gorakhpur)

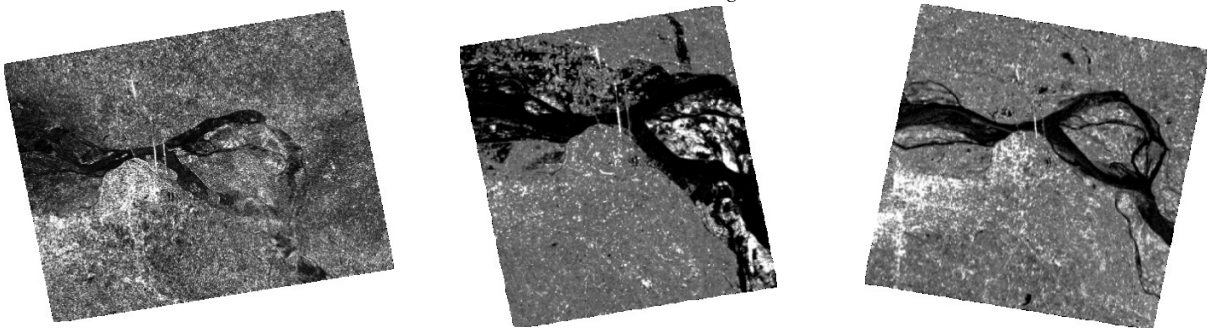
Figure 4. Calibrated Subsets of Archive1, Crisis, and Archive2 images with Speckle Filter

As displayed in Figure 5, the terrain correction is applied with Range Doppler Terrain Correction feature to correct the distortion caused due to the terrain by projecting these images onto a map system as shown in Figure 5(a) - 5(i). The different land covers are categorized into the water, urban, vegetation, and bare soil. The multi-looking feature is enabled on the data subsets to analyze deeper parts of urban areas, although that would affect the image resolution. The terrain corrected images contain the mentioned processes with a conversion from virtual to a file system band.



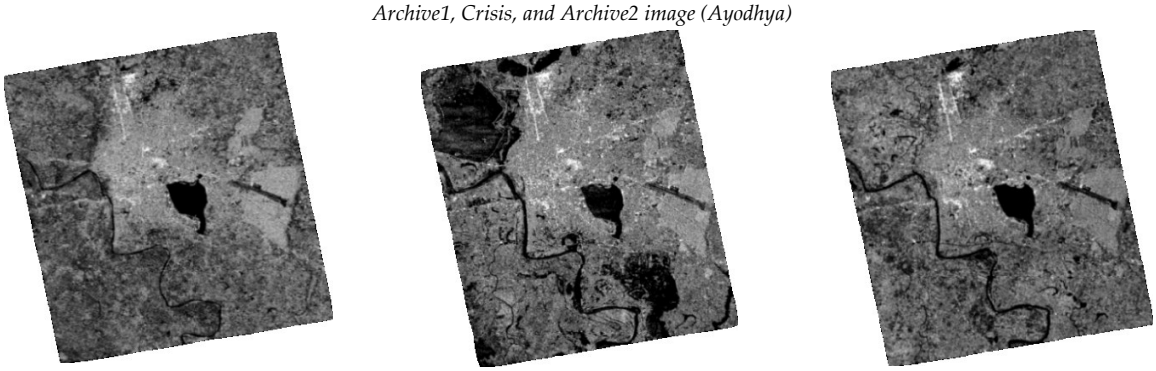
(a) Terrain Corrected Archive1 image (b) Terrain Corrected Crisis image (c) Terrain Corrected Archive2 image

Archive1, Crisis, and Archive2 image (Basti)



(d) Terrain Corrected Archive1 image (e) Terrain Corrected Crisis image (f) Terrain Corrected Archive2 image

Archive1, Crisis, and Archive2 image (Ayodhya)



(g) Terrain Corrected Archive1 image (h) Terrain Corrected Crisis image (i) Terrain Corrected Archive2 image

Archive1, Crisis, and Archive2 image (Gorakhpur)

Figure 5. Terrain correction of Archive1, Crisis, and Archive2 images

The RF classifier generates several classification trees, these trees get trained over bootstrapped data samples used for training. This algorithm determines the split at all nodes by finding across the subset of variables randomly. The input vector is submitted to all the trees in the forest to classify and determined through a majority vote [19]. The RF algorithm is among the most efficient classification techniques, due to its inherent interdisciplinary environment and it is a common choice by researchers from various disciplines. Akar and Güngör [11] investigated the RF algorithm's performance on multi-spectral SAR data with diverse spatial resolutions and characteristics. The prediction function (f(x)) for the RF Classification is generated using:

$$f(x) = C_{full} + \sum_{k=1}^{K} contribution(x, k)) \quad (1)$$

Where C_{full} is the average of complete dataset and K is the number of features. RF classifier generates the analysis for predictive capacity of features in terms of the important score.

Another classification technique is the kNN algorithm, which is a distance-based ML algorithm and instance-based learning method in which the elements are classified according to the nearest training samples [20]. In addition to the adjustment parameter, the data has a vital role in spatial forecasting. The KNN classification technique has been largely used for mapping flooded areas. In this classifier, the class of a data-point is determined by using the principle of majority voting, and based on the majority class, predictions are made. The color composition-based images are divided into patterns that reflect different land covers. In order to classify images, the KNN algorithm work on patterns and symbols. The feature space is built as long as the algorithm image crosses the input image. The scatter plot is created with samples for each class spread over the complete image. The position of points appears far away in the input image. However, these points are close to each other in the feature space. The closest neighbours are classified as members of the same path. The max distance parameter is used as a distance between the neighbours and input pixels. A majority vote-based case is allotted to common classes amongst the K-nearest neighbours. This is measured through a distance function. The distance function for Manhattan distance valid for continuous variables is calculated using:

$$D_H = \sum_{i=1}^{K} |x_i - y_i| \quad (2)$$

When the first step to select the k value is completed, the prediction for the KNN classification is calculated based an average of k number of closest points and is generated. The changes in water level were detected based on water images from different periods of time (before floods/after floods and during floods). The confusion matrix, overall accuracy, and kappa statistics are commonly used performance metrics in image classification techniques.

4. Methodology

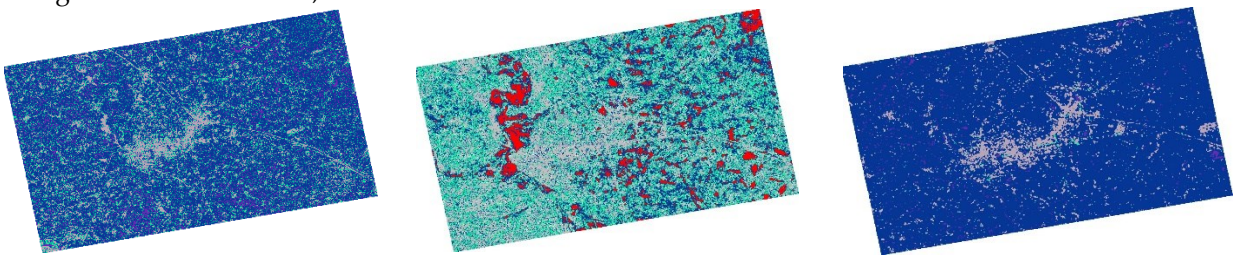
The satellite images from the Sentinel-1A series are one of the best methods to identify flood-prone areas and differentiate the changes in inundation by taking images over different periods of time. The RF and KNN are the ML algorithms that are used to classify the sub-datasets of geographical regions of Ayodhya, Basti, and Gorakhpur. The overall accuracy and the Kappa statistics of data predicted are generated as a metric for in-depth analysis and compared to attain better inter-class discrimination.

4.1. Random-Forest-based Classification of Sentinel-1 Images

The RF classification is implemented to classify the Basti and Ayodhya images into the vegetation, bare soil, water, and urban vectors as segments. The water, urban, vegetation, and bare soil categories for different land covers are created and the appropriate segments are identified manually in the crisis image. This RF classification is implemented on the vector-based image segments, the cross-validation outcome provides four types of classes, and classified images are shown in Figure 6(a) - 6(f).

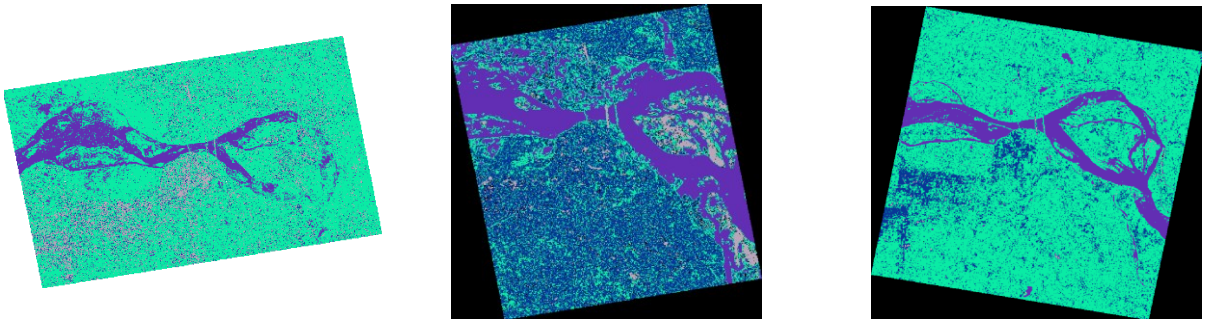
The comparative analysis of each of these classes is carried out in terms of accuracy, precision, correlation, and error rates, and the outcome of the distribution of water class is retrieved [21]. The confidence level was found to be greater than or equal to 0.5. Based on the testing sub-datasets of 9024, the correction predictions for the city of Basti before floods are 97.48%, during the crisis are 61.85%, and after

floods are 94.40%. Similarly, the correction predictions for the city of Ayodhya before floods are 97.48%, during the crisis are 61.85%, and after floods are 99.65%.



(a) Labelled classes of Archive1 image (b) Labelled classes of Crisis image (c) Labelled classes of Archive2 image

Archive1, Crisis, and Archive2 image (Basti)



(d) Labelled classes of Archive1 image (e) Labelled classes of Crisis image (f) Labelled classes of Archive2 image

Archive1, Crisis, and Archive2 image (Ayodhya)

Figure 6. Segmentation of various land covers - Supervised Classification based on Labelled classes

The classifier depicts that the flooded area with water bodies stretches beyond the areas of permanent water bodies as shown in Figure 7(a) and 7(b) and the distribution of land cover classes are given in Table 2. The correct prediction percentages are averaged from thrice partially inconsistent feature scores achieved from the RF classification technique. It is evident that the water areas of Basti which normally constitute 8% approximately increased to over 16% during floods that spread over urban populated areas and vegetation fields. Similarly, the water areas of Ayodhya that normally are around 54% approximately increased to 94.1% which confirms that the large areas of land covers were inundated with flood waters.

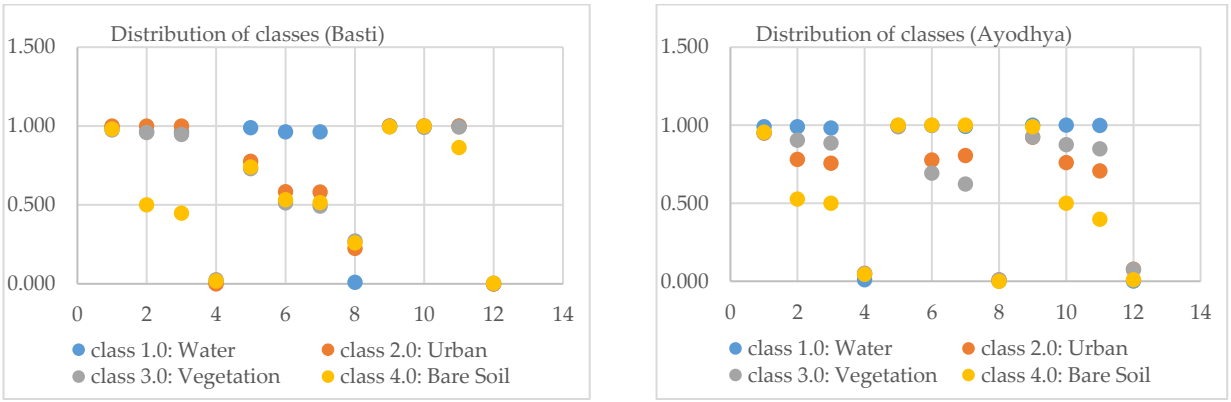


Figure 7. Distribution of classes – Cross-validation

Table 2. Distribution of Classes – RF classification

SAR data (Basti)	WATER AREA	URBAN AREA	VEGETATION	BARE SOIL
Archive1 image (Before Floods)	10.2%	52.4%	35.4%	1.9%
Crisis	26.8%	27.7%	17.7%	27.7%
Archive2 image (After Floods)	6.4%	47.0%	44.9%	1.6%
SAR data (Ayodhya)	**WATER AREA**	**URBAN AREA**	**VEGETATION**	**BARE SOIL**
Archive1 image (Before Floods)	57%	11%	27%	5%
Crisis (During Floods)	94.1%	2.1%	1.4%	2.5%
Archive2 image (After Floods)	49.2%	14.8%	35.1%	1.0%

The accuracy assessment conducted provided the overall statistics that include the producer and user accuracy of each class with the Kappa coefficient (T) confirming that the dataset matches the schema. Due

to floods, there is a huge variation in the water level and the non-water levels as seen in the classified images. The overall accuracy of the classification identified for archive1 and archive2 images of Basti was 96.47% and 97.57% with the kappa coefficient of 0.95% and 0.95% respectively. Similarly, the overall accuracy of the classification found for archive1 and archive2 images of Ayodhya is 67.54% and 86.39% with the kappa coefficient of 0.54% and 0.80% respectively. The overall accuracy of classification for the crisis images of Basti and Ayodhya was found to be 75.22% and 88.38% respectively.

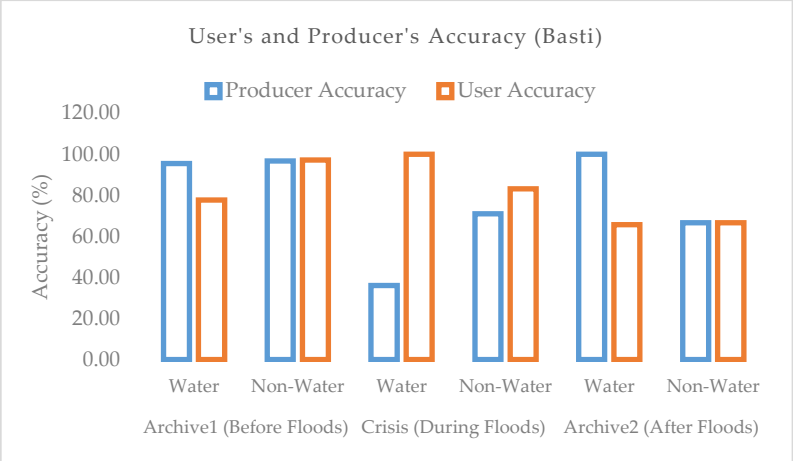


Figure 8. User's and producer's accuracy (Basti)

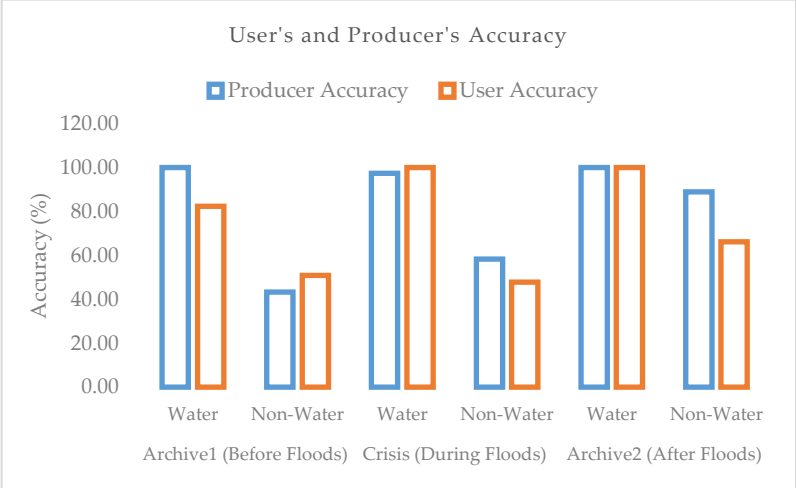


Figure 9. User's and producer's accuracy (Ayodhya)

4.2. KNN based Classification of Sentinel-1 Images

ML based KNN classification technique on the segmented images of the city of Basti is implemented. Few pixels shown below are not similar to the pattern provided in the samples. Some of the pixels are visibly classified in the result as per the KNN max parameter considered part of certain clusters. A total of 3500 samples with five neighbours are initialized to classify the image. KNN, the supervised classification is implemented on the vector-based image of the city of Basti with all water, urban, vegetation, and bare soil segments [22]. The distribution of classes is given in Figure 10, where it is evident that the flooded areas constitute over 27% that spread over urban populated areas and vegetation fields. Similarly, the flooded areas in the city of Ayodhya constituted approximately spread over to 90% regions of urban and vegetation areas.

The correction predictions using this ML algorithm for the city of Basti before floods are 97.12. By using the testing dataset of total samples of 9273, the correct predictions during floods are found to be 69.29% and the correct predictions after the floods are 97.24. By using a testing dataset of total samples of 2353, the correction predictions for the city of Ayodhya before floods are found to be 90.80%, and during floods 98.89%, and 92.73% after the floods, which re-confirms that a major area of land cover was waterlogged during the floods. The classification tested various k values to select the perfect parameter for the KNN classifier. Using different subsets of SAR data, this was based on the lowest estimate of the RMSE which is

the Root Mean Square Error. This is one of the criteria for choosing the ideal k value. The RMSE value of archive images of the city of Basti has been consistent with 0.3 whereas, the crisis image shows a deviation of 1.0. Similarly, the value based on archive images for the city of Ayodhya also ranges between 0.3 to 0.4 whereas, it deviated to 0.1 during the crisis.

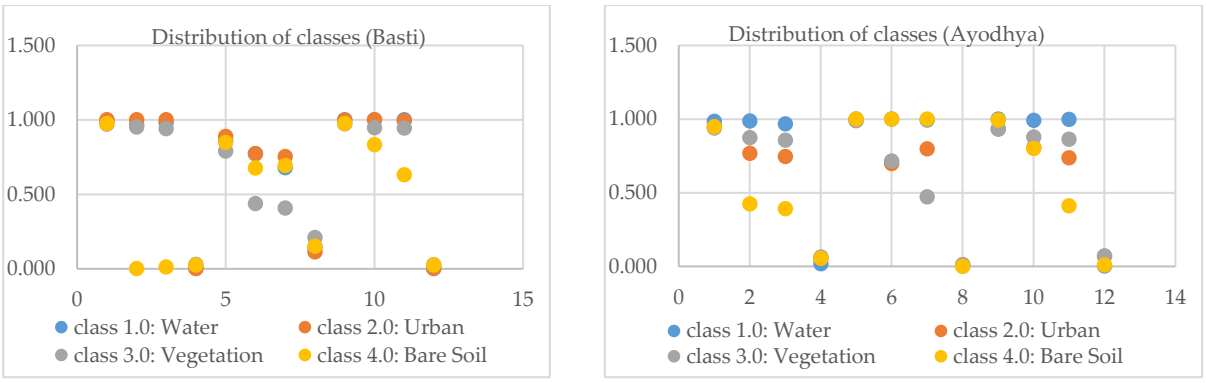


Figure 10. Distribution of classes – Cross-validation

Table 3. Distribution of classes – KNN classification

SAR Image (Basti)	Water Area	Urban Area	Vegetation	Bare Soil
Archive1 image (Before Floods)	10%	52%	35%	2%
Crisis (During Floods)	27.0%	27.0%	19.1%	27.0%
Archive2 image (After Floods)	14.3%	39.3%	42.3%	4.1%
SAR Image (Ayodhya)	**Water Area**	**Urban Area**	**Vegetation**	**Bare Soil**
Archive1 image (Before Floods)	55.9%	11.1%	28.9%	4.1%
Crisis (During Floods)	93.1%	2.1%	2.6%	2.2%
Archive2 image (After Floods)	47.1%	14.8%	37.2%	1.0%

The flooded area polygons are generated on the SAR data images that have been used as a reference in this paper. The analysis is based on pixels with a confusion matrix as shown in Table 4. The true class data is referred to as positive or negative (flooded or unflooded). The true-positive (TP) is considered if the instance and the classified polygon are positive, which means the area is flooded. The false-positive (FP) is considered if the instance and classified polygon are positive, which means the area is unflooded.

Table 4. Confusion matrix - Flooded/Unflooded classified areas

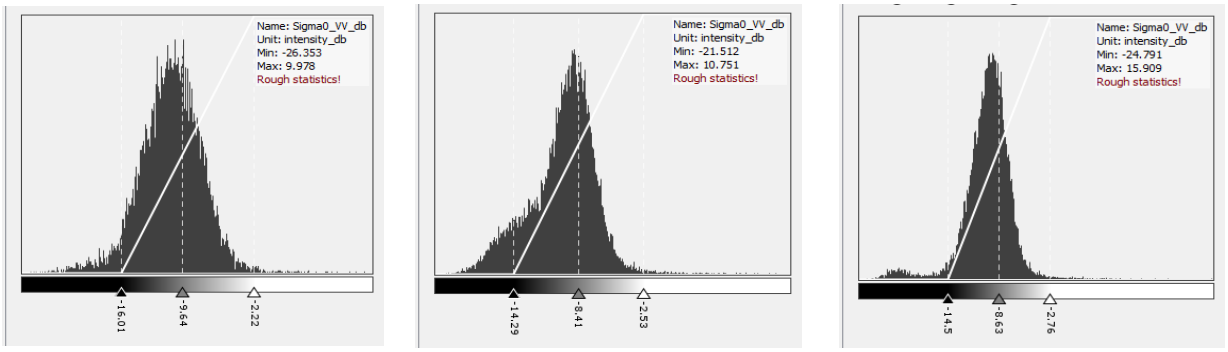
REFERENCE MAP				
		Flooded	**Unflooded**	**Total**
SAR Data	**Flooded**	TP	FN	TP + FN
	Unflooded	FP	TN	FP + TN
	Total	TP+FP	FN+TN	TP+FN+FP+TN

Additionally, if the instance and classified polygon are negative which means unflooded areas, then it will be true-negative (TN) whereas, if only classified as positive, it will be considered as false-negative (FN).

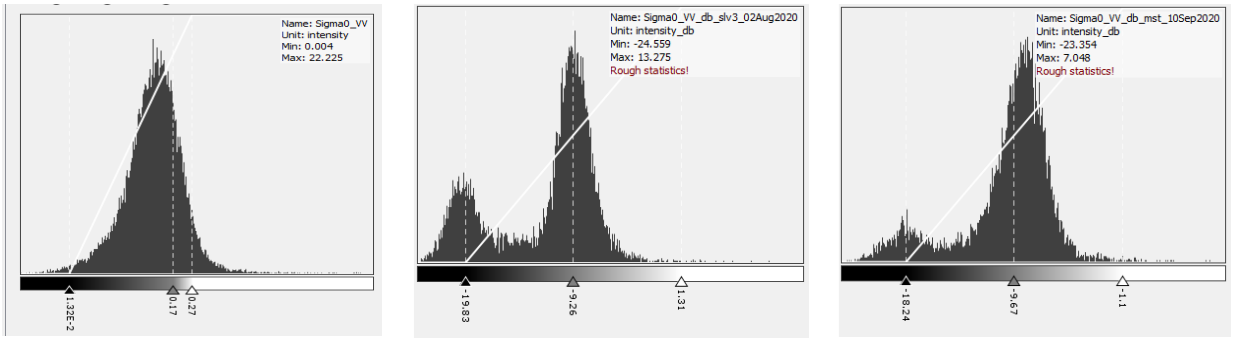
4.3. Flood Mapping – RGB Composite of Gorakhpur and Ayodhya

The management of data related to floods, communication and improving the accuracy of flood mapping could be examined by "Mapping the Zone" by evaluating the affecting factors such as advantages and cost of reviewing additional flood maps [23]. Flood maps or flood mapping exercises are vital for a range of activities carried out by public, private, and third-sector actors including establishing and enforce zoning, land use, and building standards, when planning and building infrastructure and transportation networks, for flood warning, evacuation, and emergency management and planning.

It is observed in the color manipulation process that many of the pixels had low backscatter values and hence, there is a switch from digital numbers to the physical quantity which is sigma naught backscatter in this case. Hence, the pixels are converted from a linear scale to decibel which is non-linear logarithm scale to have a clear visualization.



Archive1, Crisis, and Archive2 image (Gorakhpur)



Archive1, Crisis, and Archive2 image (Ayodhya)

Figure. 11. Flood Mapping - Histograms for images

The histogram generated with this process is easier to manipulate for achieving clear image that distinguished between water and other areas in the city. The crisis and archive images can further be refined with the contrast stretch on respective histograms. To overlay images over the map together, they are stacked into output extends using the product geolocation to ensure they are well registered. To do the interferometry, there must be much precise co-registration. These images are simply combined to create an RGB composite to differentiate between the flooded areas and the permanent water areas. The archive image is selected as a red band to ensure a red channel highlighting the flooded region as a high-radar response as these will be land cover.

The backscatter return will be higher in archive images since the water areas are lesser during the normal situation. In the crisis image, the backscatter return will be low in the flooded areas and hence it is assigned with green and blue bands. The appearance of flooded areas will be in red i.e. there will be a high response in the red channel and will have a low response in the blue and green channels.

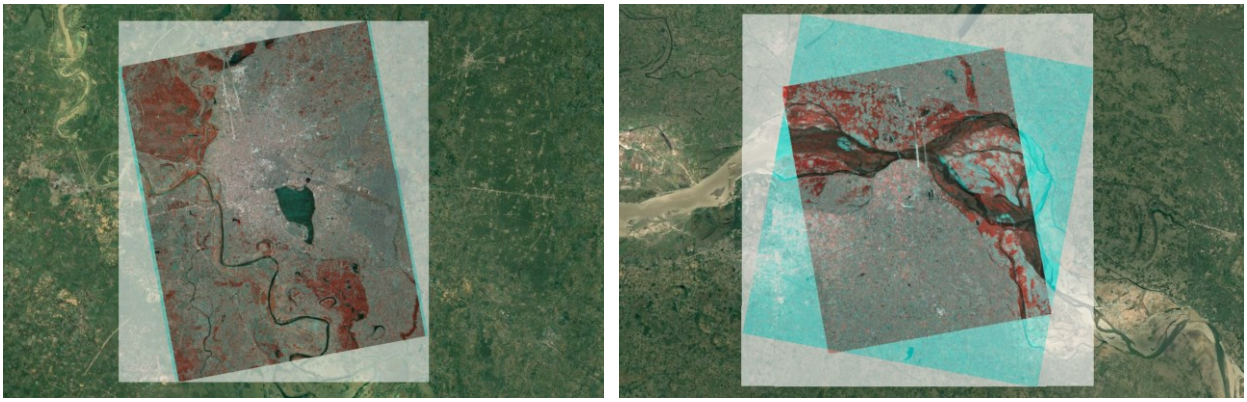


Figure. 12. Google Earth Pro-Flood areas (Red highlighted) of Gorakhpur and Ayodhya

The image is exported in a GeoTiff format and also in Google Earth KMZ format to overlay it onto the imagery of Google Earth. In Google Earth Pro, the flooded areas are highlighted in red which covers a large portion of the cities.

5. Results and Discussion

The sentinel-1 image is utilized to implement ML algorithms such as RF and KNN classifications. The outcomes produced are compared for the sub-datasets of different geographical regions of northeast India. The number of training sample sizes varied from 550 to 9250 pixels/classes with results (OA %) ranging from 22.1% to 27.6% that are trained using MLC. The percentage of flooded areas generated by both the classifiers are found to be closely matched. The statistics differentiating the distribution of flooded areas with the regular water bodies are categorized and displayed in the below table.

Table 5. Comparison of flooded areas based on RF and KNN classifiers

Distribution of Flooded area (%)			
Classifiers	**Parameters**	**Basti**	**Ayodhya**
RF	Water area	10.2%	56.9%
	Flooded area	16.6%	37.2%
KNN	Water area	10.2%	56.9%
	Flooded area	16.8%	41.6%

The urban and the vegetation land covers are classified as more than the water area in the city of Basti whereas, the water area covers more or equal to the vegetation, urban and bare soil combined in the city of Ayodhya. The root-mean-square error (RMSE) values range between 0.3 and 0.4 for the city of Ayodhya and Basti respectively which depicts that the implementation technique can predict the data quite accurately.

Table 6. Overall accuracy and Kappa statistics for the study areas

Study Area	**Overall accuracy (%)**	**Kappa coefficient**	**RMSE**
Basti	89.8	0.8	0.4
Ayodhya	80.8	0.7	0.3

The overall accuracy using supervised classification is found to be 89.8% for the city of Basti and 80.8% for the city of Ayodhya which confirms the agreement with the different data sets. According to the results, KNN has a slightly better performance when compared with the RF classification. However, to cross-verify the land cover classification, the SAR data images are recommended to have a training sample size of at least 0.25% of the regions. Moreover, the performance of both classifiers on different training samples is slightly different as depicted in Table 6. The accuracy level will be higher with the larger training sample. The Kappa statistics produced by the classifier gave better inter-class discrimination as a metric than the overall accuracy.

6. Conclusion

The results and the evaluation observed that the approach to identifying the flood-prone areas and the changes in flooded areas is quite eminent for further analysis. From the RF and KNN classifiers, the land cover classification of water bodies, vegetation, urban, and bare soil areas have been successfully detected. The accuracy of the data is confirmed from the similarly mapped data retrieved from both these techniques. Further, this is supported through flood mapping of different regions by developing RGB composite images and unsupervised classification techniques. The outcomes from additional ML and DL algorithms for the processed image datasets of chosen regions from different crisis periods would enhance the data collection techniques for deeper time-series analysis of flood-prone areas. This supports the purpose of this article which is to focus on a comparative analysis of results on SAR data images of particularly north-east Indian regions based on various classification techniques to further support the development of a flood monitoring and early warning system. The time-series analysis based on periodically acquired Sentinel-1 images can be easily carried out using the Flood Monitoring and Early Warning System (FMEWS) which can help continuous monitoring of flood situations and alert the concerned authorities for necessary action.

7. Future Studies

The results illustrated suggest that the classification approach and the data analysis outperform the traditional processes in conditions where the ML algorithms are significantly utilized on SAR data during different periods of time. The unsupervised ML algorithms and the DL algorithms-based classification

techniques could be implemented in future research to detect the changes in flood patterns more efficiently and utilize the results for developing the proposed FMEWS model.

Acknowledgement

Authors are thankful to the authorities of Integral University, Lucknow for providing their kind support. The manuscript number issued by the University is IU/R&D/2022-MCN0001656.

References

[1] Mohit Prakash Mohanty, Sahil Mudgil and Subhankar Karmakar, "Flood management in India: A focussed review on the current status and future challenges", *International Journal of Disaster Risk Reduction (IJDRR)*, ISSN: 2212-4209, pp. 648-658, Vol. 49, 2020, Published by Elsevier, DOI: 10.1016/j.ijdrr.2020.101660.

[2] Abdulla All Noman, Umma Habiba Akter, Tahmid Hasan Pranto and AKM Bahalul Haque, "Machine Learning and Artificial Intelligence in Circular Economy: A Bibliometric Analysis and Systematic Literature Review", *Annals of Emerging Technologies in Computing (AETiC)*, Print ISSN: 2516-0281, Online ISSN: 2516-029X, pp. 13-40, Vol. 6, No. 2, 1st April 2022, Published by International Association for Educators and Researchers (IAER), DOI: 10.33166/AETiC.2022.02.002, Available: http://aetic.theiaer.org/archive/v6/v6n2/p2.html.

[3] Mohammed Waleed Ashour, Fatimah Khalid, Alfian Abdul Halin, Samy H. Darwish and M. M. Abdulrazzaq, "A Review on Steel Surface Image Features Extraction and Representation Methods", in *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Emerging Technologies in Computing (iCETiC)*, Vol. 332, Online ISBN: 978-3-030-60036-5, Print ISBN: 978-3-030-60035-8, Series Print ISSN: 1867-8211, Series Online ISSN: 1867-822X, DOI: 10.1007/978-3-030-60036-5_17, pp. 239–250, 2020, Published by Springer, Available: https://link.springer.com/chapter/10.1007/978-3-030-60036-5_17.

[4] Pravesh Kumar Singh and Mohd Shahid Husain, "Methodological study of opinion mining and sentiment analysis techniques", *International Journal on Soft Computing (IJSC)*, Print ISSN: 2229-7103, Online ISSN: 2229-6735, pp. 11-21, Vol. 5, No. 1, 2014, Published by Academy and Industry Research Collaboration Center (AIRCC), DOI: 10.5121/ijsc.2014.5102.

[5] Nawazish Naveed, Hayan T. Madhloom and Mohd Shahid Husain, "Breast cancer diagnosis using wrapper-based feature selection and artificial neural network", *Applied Computer Science (ACS)*, Print ISSN: 1895-3735, Online ISSN: 2353-6977, pp. 19-30, Vol. 17, No. 3, 2021, Published by Polskie Towarzystwo promocji Wiedzy, DOI: 10.23743/acs-2021-18.

[6] Pravesh Kumar Singh and Mohd Shahid Husain, "Analytical study of feature extraction techniques in opinion mining", *Computer Science & Information Technology (CS & IT)*, ISSN: 2231-5403, pp. 85-94, Vol. 3, 2013, Published by AIRCC Publishing Corporation, DOI: 10.5121/csit.2013.3410.

[7] Juliana Maria Diniz, Fabio Gama Furlan and Marcos Adami, "Evaluation of polarimetry and interferometry of sentinel-1A SAR data for land use and land cover of the Brazilian Amazon Region", *Geocarto International*, Print ISSN: 1010-6049, Online ISSN: 1752-0762, pp. 1482-1500, Vol. 37, No. 5, 2022, Published by Informa UK Limited, DOI: 10.1080/10106049.2020.1773544.

[8] Shahriar S. Heydari and Giorgos Mountrakis, "Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites", *Remote Sensing of Environment (RSE)*, ISSN: 0034-4257, pp. 648–658, Vol. 204, 2018, Published by Elsevier, DOI: 10.1016/j.rse.2017.09.035.

[9] Reza Khatami, Giorgos Mountrakis and Stephen V. Stehman, "A meta-analysis of remote sensing research on supervised pixel-based land cover image classification processes: General guidelines for practitioners and future research", *Remote Sensing of Environment (RSE)*, ISSN: 0034-4257, pp. 89-100, Vol. 177, 2016, Published by Elsevier, DOI: 10.1016/j.rse.2016.02.028.

[10] Rajat Garg, Anil Kumar, Manish Prateek, Kamal Pandey and Shashi Kumar, "Land cover classification of spaceborne multifrequency SAR and optical multispectral data using machine learning", *Advances in Space Research (ASR)*, Print ISSN: 0273-1177, Online ISSN: 1879-1948, pp. 1726-1742, Vol. 69, No. 4, 2022, Published by Elsevier, DOI: 10.1016/j.asr.2021.06.028.

[11] Özlem Akar and Oguz Güngör, "Classification of multispectral images using Random Forest algorithm", *Journal of Geodesy and Geoinformation*, Print ISSN: 2147-1339, Online ISSN: 2096-1650, pp.105-112, Vol. 1, 2012, Published by Surveying and Mapping Press, DOI: 10.9733/jgg.241212.1.

[12] C. M. Bhatt and G. S. Rao, "Ganga floods of 2010 in Uttar Pradesh, north India: a perspective analysis using satellite remote sensing data", *Geomatics, Natural Hazards and Risk*, Print ISSN: 1947-5705, Online ISSN: 1947-5713, pp. 747-763, Vol. 7, No. 2, 2016, Published by Informa UK Limited, DOI: 10.1080/19475705.2014.949877.

[13] J. G. Natividad and J. M. Mendez, "Flood Monitoring and Early Warning System Using Ultrasonic Sensor", in *Proceedings of the International Conference on Information Technology and Digital Applications (ICITDA 2017)*, 8–9

November 2017, Yogyakarta, Indonesia, ISBN: 978-1-5108-5984-5, DOI: 10.1088/1757-899x/325/1/012020, pp. 012020, Published by IOP Publishing Ltd., Available: https://iopscience.iop.org/article/10.1088/1757-899X/325/1/012020.

[14] S. Yeon, J. Kang and I. Lee, "A study on real-time flood monitoring system based on sensors using flood damage insurance map", *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)*, Print ISSN: 1682-1750, Online ISSN: 2194-9034, pp. 569-571, Vol. XLII-3/W4, 2018, Published by Copernicus GmbH, DOI: 10.5194/isprs-archives-XLII-3-W4-569-2018.

[15] Razi Pakhrur, Sri Sumantyo Josaphat, Yulkifli, Widodo Joko, Daniele Perissin *et al.*, "Land deformation modelling of Taiwan earthquake using interferometry technique", in *Proceedings of The 2nd International Conference on Research and Learning of Physics*, 9-10 August 2019, East Sumatra, Indonesia, Online ISBN: 1742-6588, E-ISBN: 1742-6596, DOI: 10.1088/1742-6596/1481/1/012009, pp. 1-5, 2020, Published by IOP Publishing, Available: https://iopscience.iop.org/article/10.1088/1742-6596/1481/1/012009.

[16] Y. Chen, S. Qiao, G Zhang, Xu YJ, L. Chen *et al.*, "Investigating the potential use of Sentinel-1 data for monitoring wetland water level changes in China's Momoge National Nature Reserve", *PeerJ*, ISSN: 2167-8359, pp. e8616, Vol. 8, 2020, Published by PeerJ, DOI: 10.7717/peerj.8616, Available: https://peerj.com/articles/8616/.

[17] Abazaj Freskida and Hasko Gëzim, "Use of Sentinel-1 Data in Flood Mapping in the Buna River Area", *International Journal of Environment and Climate Change (IJECC)*, ISSN: 2581-8627, pp. 147-156, Vol. 10, 2020, Published by Sciencedomain International, DOI: 10.9734/IJECC/2020/v10i1030257.

[18] Tasneem Ahmed, "Monitoring and Mapping of Flash Flood of Patna City Using Sentinel-1 Images: A Case of India's Most Flood Prone State", *Academia Letters*, ISSN 2771-9359, Article 1349, 2021, Published by Academia Publishing, DOI: 10.20935/AL1349.

[19] P. O. Gislason, J. A. Benediktsson and J. R. Sveinsson, "Random Forest classification of multisource remote sensing and geographic data", in *Proceedings of 2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2004)*, 20-24 September 2004, Anchorage, AK, USA, Print ISBN:0-7803-8742-2, DOI: 10.1109/IGARSS.2004.1368591, pp. 1049-1052, Published by IEEE, Available: https://ieeexplore.ieee.org/document/1368591.

[20] Phan Thanh Noi and Martin Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery", *Sensors*, ISSN: 1424-8220, pp. 18, Vol. 18, 2017, Published by Multidisciplinary Digital Publishing Institute (MDPI), DOI: 10.3390/s18010018.

[21] Viktoriya Tsyganskaya, Sandro Martinis and Philip Marzahn, "Flood Monitoring in Vegetated Areas Using Multitemporal Sentinel-1 Data: Impact of Time Series Features", *Water*, ISSN: 2073-4441, pp. 1938, Vol. 11, 2019, Published by IWA Publishing, DOI: 10.3390/w11091938.

[22] Francisco Carreño Conde and Maria De Mata Muñoz, "Flood Monitoring Based on the Study of Sentinel-1 SAR Images: The Ebro River Case Study", *Water*, ISSN: 2073-4441, pp. 2454, Vol. 11, 2019, Published by IWA Publishing, DOI: 10.3390/w11122454.

[23] National Research Council of the National Academies, "*Mapping the Zone: Improving Flood Map Accuracy*", Illustrated ed. Washington, DC, USA: National Academies Press, 2009, ISBN: 978-0-309-13057-8, Available: https://nap.nationalacademies.org/catalog/12573.