Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Cultivating Resilience: Sentinel-2 Remote Sensing for Precision Flood Detection and Susceptibility Mapping in Agricultural Landscapes

Nilkamal More[1], Suchitra Patil[1], Bhakti Palkar[1], Seema Talmale[1],V.Venkatramanan[1],Biswajit Sarkar[2]
K.J.Somaiya School of Engineering, Somaiya Vidyavihar University, Vidyavihar,Mumbai,India[1]
Yonsei University, South Korea[2]

ARTICLE INFO

Received: 27 Dec 2024

Revised: 15 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

In agricultural areas, flooding is a frequent and destructive natural calamity that seriously harms crops, infrastructure, and property. Early detection and prediction of floods is crucial for disaster management and response efforts in agricultural areas. In this study, we present a novel method for Sentinel-2 remote sensing data-based flood detection and susceptibility mapping in agricultural farms, based on the application of different unsupervised clustering methods and supervised machinelearning algorithms. The approach involves pre-processing the Sentinel-2 images to extract features, such as spectral reflectance values, and transforming the images into feature vectors. These feature vectors are then used as inputs to several differentalgorithms to identify regions that represent flooded areas. This work presents two approaches namely, unsupervised clustering approach, where 5 different algorithms are used and supervised machine learning approach, where 3 different algorithms areused in identifying the flooded regions under different conditions. Accuracy and water percentage increase of both approaches are compared. The study's findings demonstrate that the suggested strategy is efficient for detecting and mapping flood susceptibility in agricultural areas using Sentinel-2 remote sensing data, with an overall accuracy of up to 92% for the Support Vector machine(SVM) algorithm in the supervised algorithms category and DBScan algorithm with an accuracy of 84% in the unsupervised alorithms category.

Keywords: *Flood detection, Flood Susceptibility Mapping, Sentinel-2,Remote Sensing, Unsupervised Clustering, Supervised Classification*

INTRODUCTION

One of the most frequent and destructive natural disasters that can affect different areas including agricultural areas is flooding. Agricultural production is critical for food security and the economy, and floods can cause significant losses in crop yields and infrastructure, leading to major social and economic disruptions. Accurate and timely detection and mapping of floods and their impacts are essential for mitigating their impacts on agriculture and reducing the risk to agricultural production. Remote sensing technology provides an opportunity to monitor and map floods and their impacts on agricultural areas at a regional scale. High-resolution photos of the surface of the earth, including agricultural regions, are provided by the Sentinel-2 satellite, a European Space Agency satellite. These images can be used to detect and map floods and their impacts on agricultural areas, providing valuable information for farmers, decision-makers, and researchers. This study aims to identify and map the vulnerability of agricultural crops to flooding using Sentinel-2 remote sensing data. Five different unsupervised clustering algorithms, DBSCAN, OPTICS, CLARANS, Gaussian Mixture Model(GMM) and K-Means, and three different supervised classification algorithms namely Support Vector Machine Algorithm (SVM), Random Forest and Decision tree (CART Method) ere applied to the pre-processed Sentinel-2 data to identify the best method for detecting and mapping the susceptibility of agricultural farms to flooding.The algorithms were selected based on their suitability for

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

identifying and mapping clusters in remote sensing data. The performance of the algorithms was evaluated using overall accuracy based on different parameters involved for unsupervised and supervised approaches, which are commonly used metrics for evaluating the accuracy of land cover classification results. The results of both the approaches were used to create flood susceptibility maps of the study area, which showed the extent and distribution of flooded and non-flooded areas in the study region. The findings of this study have significant ramifications for managing flood risk and making decisions in agricultural areas. This study offers useful information for farmers and decision-makers in flood-prone areas by showcasing the potential of Sentinel-2 remote sensing data and clustering algorithms for flood detection and susceptibility mapping in agricultural farms. This data can be utilised to enhance flood risk management plans and decision-making procedures, ultimately lessening how much flooding affects communities and agricultural output. In conclusion, using Sentinel-2 remote sensing data and clustering algorithms is a viable strategy for identifying and mapping agricultural farms' susceptibility to flooding. The findings of this study contribute to the development of more effective techniques for reducing the effects of floods on agriculture by offering useful information for enhancing flood risk management and decision-making procedures in agricultural areas.A number of practical issues highlight the necessity for smaller, more efficient neural network structures.

OBJECTIVES

- The study tries to address the problem of flood detection and susceptibility mapping in agricultural areas, a common and destructive issue for crops, infrastructure, and property.
- Sentinel-2 satellite imagery is utilized for detecting floods, leveraging its spectral reflectance values to capture relevant features from the landscape.
- The images undergo pre-processing to extract features, and these are converted into feature vectors for further analysis. This step is critical for preparing data for subsequent machine learning methods.
- The study concludes that the proposed method is highly effective for detecting and mapping flood susceptibility in agricultural regions, showcasing the potential of Sentinel-2 data combined with advanced machine learning techniques.

LITERATURE SURVEY

One of the most destructive natural disasters, floods everyyear result in significant economic losses, population displacement, and fatalities. For disaster management and mitigation measures, flood detection is essential. Because it can cover a large area and is revisited frequently, remote sensing is a useful tool for flood detection. Using satellite data for flood monitoring can be an efficient and cost-effective way to monitor large areas, particularly in remote or difficult-to-access regions [1]. Flood detection algorithms are computer programs that are designed to analyze satellite or aerial imagery to identify and map areas that have been flooded. CDAT and NDFI are two such algorithms that have been developed specifically for use in Bangladesh [2]. The authors were able to determine whether new floods are unusual events or instead are part of annual seasonal variation by using the observation made by orbital remote sensing and flood detection algorithms to preserve the extent of flooding and define local flood histories [3]. With the help of remote sensing data and GIS tools, an efficient approach was established in the Naogoan District of Bangladesh to precisely delineate the flooded areas. This approach describes the relationship between the flood in those four observation years as well as the percentages of loss connected to the spread, height, and land usage of the flood. The damaged area was located using pixel value in a matrix union after receiving two categorised maps from before and during the flood [4]. Additionally, studies demonstrate the use of a BN to detect flooded areas using the data fusion of SAR intensity imaging, InSAR coherence, and auxiliary data [5]. The Bagging-Cubic-KNN model should be used more frequently for the long-term management of flood-prone areas, it is suggested. We are able to recognise flooding disturbances on agricultural productivity by taking into account the influence of natural changes during crop growth in processing flood effect signals.The temporal pathways of EVI and peak EVI during the crop growth periods are compared between actual and expected normal values in order to identify and remove any interference caused by the crop's intra-annual natural fluctuations [6].With reference to Asia's monsoon regions, the use of GIS and satellite imagery has proven to be a very successful and reasonably priced method of managing floods [7].A study done in Bangladesh assesses the trends in land use/cover change for Dhaka over the previous 45 years, providing useful information for planners to create sustainable land use.All of the data were resampled to a 30 m pixel size using the closest neighbour approach, and a first-order polynomial fit was performed.In addition to being essential for understanding the historical and present state of the land, timely and accurate land cover data is also utilised to help design integrated resource management policies, accomplish sustainable urban development, and create solid environmental planning [8]. Change detection is carried out using a separate method based on pixel-based change detection. Before doing a flood study, it is crucial to understand the backscatter properties of various ground objects. Rapid flood mapping and flood degree estimation are helpful. It transforms the categories of land cover backscatter from change detection of thresholds [9]. This technique was used in Northern Iran to identify floods, where models employing bagging and KNN classifiers were developed to spatially forecast flooding in the Haraz watershed [10]. Hybrid algorithms were used to advance hazard modelling further. Machine learning ensemble models fall under this category; compared to the approaches outlined above, they are more adaptable and more suited for advanced flood modelling. It has been demonstrated that machine learning ensemble models can offer more accurate flood danger predictions. Historically, flood detection and associated remote sensing applications have used Sentinel-1(SAR) and Sentinel-2 optical images. Studies were done to determine how clouds and vegetation affected surface obstacles. These studies created flood maps for two case studies using Sentinel-2 and Landsat data, and then compared the accuracy of the maps with and without cloud and vegetation cover modifications [11]. Despite the Sentinel-2 mission's distinctive features and its imagerys high potential for mapping and monitoring floods, its usage is difficult and confined due to the heavy cloud cover during flood occurrences. However, if there is little to no cloud cover during flood events, which is more likely given the satellite's high revisit coverage, Sentinel-2 imagery's multispectral optical images can be utilised efficiently for mapping and monitoring floods. Sentinel-2 multispectral optical data has a much higher degree of accuracy than Sentinel-1 SAR when there is no cloud cover, little cloud cover and significant hilly terrain, high raised buildings, flat highly reflective surfaces that appear dark in the SAR image, or any combination of these. Various techniques and tools are now available for making atmospheric modifications to decrease and hide the clouds [11][12]. Recently, a novel approach for flood detection using a combination of Sentinel-1 and Sentinel-2 satellite imagery. The authors address the challenge of detecting floods in cloudy scenes by integrating radar and optical data. They utilize the backscattering coefficient from the Sentinel-1 radar data and the normalized difference water index from Sentinel-2 optical data to generate flood maps. These two datasets are combined by the authors to produce a more thorough flood map that can precisely identify floods in foggy areas. According to the study's findings, this fusion strategy can efficiently detect floods in areas with heavy cloud cover and can be a useful tool for managing floods and responding to disasters. The fusion of these two datasets improves the accuracy and reliability of flood detection in cloudy regions. The findings suggest that the fusion approach can effectively detect floods in regions with high cloud cover and could be a valuable tool for flood management and disaster response [13].

METHODS

The work includes two approaches for flood detection: Unsupervised and Supervised Learning. Several algorithms are used in each of the methods. These approaches are explained further in the flowchart in Figure 1.

Below, each step of the workflow is discussed in brief.Sentinel 2 images of Pre and Post Flood Period: Sentinel-2 remote sensing data of the particular region is collected using Google Earth Engine. The data was acquired during a specific time period, typically covering the period of interest, and during a specific season, such as the rainy season, when flooding is most likely to

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

occur. The images of the same region were taken twice (once before the flooding and once after to make comparisons between the two). Create image chips of decided resolution: The images collected are broken down into equal size image chips (256x256 resolution), which is an optimal size for images to be fed to the Machine Learning Models. We collected 500 pre-flood and 500 post-flood image chips of size 256X256.

Image Pre-processing: The pre-processing of the Sentinel-2 data is an important step in ensuring the quality and accuracy of the results. This step involves removing noise and enhancing the quality of the images by applying atmospheric correction. This step also involves removal of cloud cover, removal of NaN values and applying standardisation on the dataset.Sentinel-2 uses the Normalised Difference Water Index (NDWI) to detect and map surface water in satellite images, making it a crucial index. The Short wave Inrared(SWIR-Band 12) and nearinfrared (NIR-Band 8) spectral regions of the Sentinel-2 satellite are used to determine the NDWI along withe the green band(Band-4).

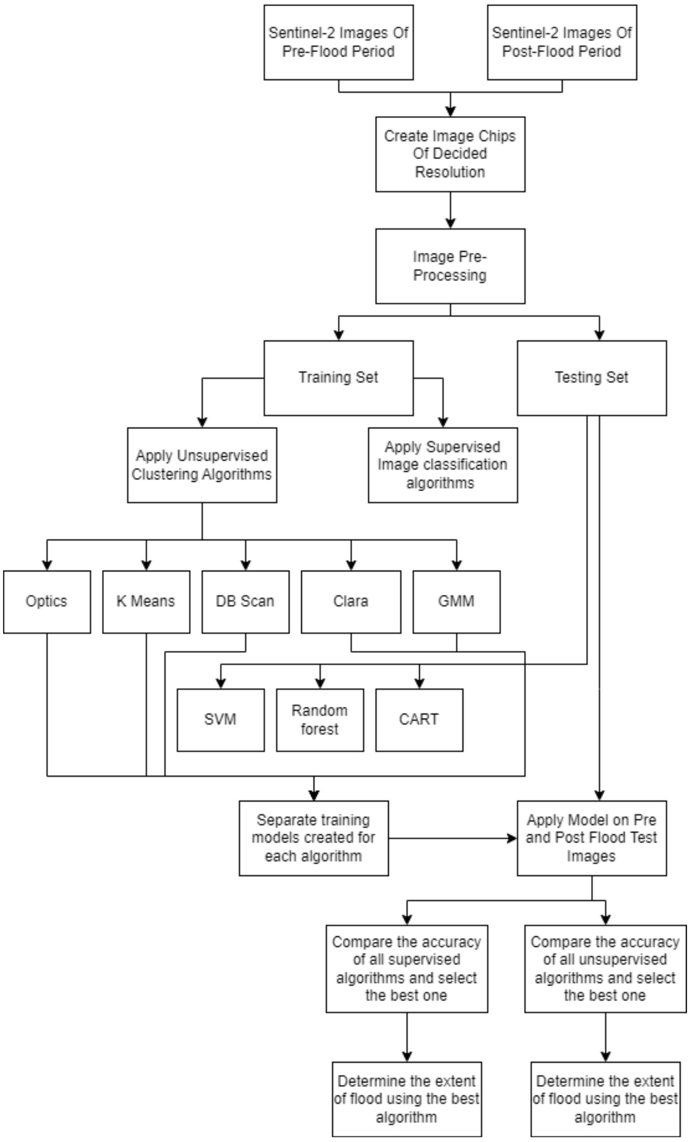


Fig. 1: Workflow of Flood Detection using both supervised and unsupervised learning models

Cloud Masking: Sentinel-2 performs cloud masking using a combination of spectral and spatial

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

techniques. Spectral approaches separate clouds from the underlying surface by comparing the reflectance values of clouds and other surface features. On the other hand, spatial approaches make use of the texture and pattern of cloud cover to recognise and mask clouds. The blue, green, red, and near-infrared spectral bands are those that are most affected by cloud cover. The reflectance levels in these bands are typically subjected to a threshold as a common method of cloud masking in Sentinel-2. Pixels are identified as clouds and then hidden if their reflectance values are higher than the threshold. Splittig into Training and Test Set: The dataset considered is split into two sets namely Training and Test in the 4:1 ratio. The training set will be used to train the Machine Learning Models while the test set will be utilised to verify the results.

Applying Unsupervised Learning Algorithms: Five different unsupervised clustering algorithms were applied to the pre-processed training set data to identify the best method for detecting and mapping the susceptibility of agricultural farms to flooding. The algorithms used were KMeans, GMM, Clara, DB-Scan and Optics. Each Algorithm is explained further here: K-Means Algorithm: K-means is an unsupervised clustering algorithm that attempts to divide a dataset into K user-defined clusters. It is an iterative algorithm that reduces the sum of the squared distances between the assigned centroid and each data point. This is how the k-means algorithm operates:

1) Initialise the algorithm with a set of data points and the desired number of clusters.

2) Select K data points at random as the initial cluster centroids.

3) According to the Euclidean distance, assign each data point to the nearest centroid.

4) Recalculate the cluster centroids as the mean of the designated data points.

5) Repeat steps 3 and 4 until convergence is achieved, which typically occurs when the assignment of data points to clusters no longer varies substantially.

6) The final centroids and the data points designated to each centroid define the resulting clusters. We applied the elbow method to determine the optimal number of clusters, which in our case was two.

Fig. 2: Ground Truth Image

- **CLARA Algorithm**: CLARA (Clustering Large Applications) is a clustering algorithm that is designed to handle large datasets by sampling subsets of the data and clustering them separately. It is a variant of the k-medoids algorithm that works by selecting a subset of data points, computing the medoids for the subset, and using them as the initial centroids for the k-medoids algorithm. This is how the CLARA algorithm operates:

1) Initialise the algorithm with a collection of data points, the desired number of clusters, and a sample size.

2) Select at random a subset of the data points (of size equal to the sample size) and compute the medoids for the subset using the PAM (Partitioning Around Medoids) algorithm.

3) Use the resulting medoids as the initial centroids for the k-medoids algorithm on the full

dataset.

4) Assign each data point to its closest centroid and recalculate the medoids for each cluster.

5) Repeat steps 3 and 4 a predetermined number of times or until convergence is attained.

6) Choose the best clustering solution based on a clustering validity measure such as the silhouette coefficient.

We used 2 clusters for this algorithm as well.

- **GMM Algorithm**: GMM (Gaussian Mixture Model) is a statistical clustering algorithm that assumes that the data points are generated from a mixture of Gaussian distributions. It is a generative probabilistic model that aims to identify the underlying probability distributions that generate the data. This is how the GMM algorithm operates:

1) Initialise the algorithm with a collection of data points and the desired number of clusters.

2) Means, covariances, and mixing coefficients for the Gaussian distributions that will comprise the mixture model are initialised at random.

3) Using Bayes' theorem, determine the likelihood of each data point belonging to each Gaussian distribution.

4) Update each Gaussian distribution's mean, covariance,and mixing coefficient based on the probabilities of the data points.

5) Repeat steps 3-4 until convergence is achieved, typically when the likelihood of the data no

longer changes significantly.

6) Assign each data point to the Gaussian distribution with the highest likelihood. The resulting

clusters are defined by the means and covariances of the Gaussian distributions that form the mixture model.

We used 2 clusters for this algorithm as well.

- **DBSCAN Algorithm:** The well-known clustering algorithm DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is used to locate data point clusters within a given dataset. It is a density-based clustering algorithm that identifies high-density regions that are separated by lowdensity regions and labels these regions as clusters. The algorithm is capable of discovering clusters of any configuration and is able to deal with noise and outliers. Two input parameters are required by the DBSCAN algorithm: a distance metric and the minimum number of points required to construct a dense region. The algorithm is implemented as follows:

1) Select at random a data point from the dataset that has never been visited previously.

2) Retrieve all adjacent data points that are within eps (the distance metric) of the selected point.

3) If the number of neighbouring data points is less than the minimum required to form a dense region, designate the selected data point as noise and proceed to the next point.

4) Mark the chosen data point as a core point and add all its neighbours to a cluster if the number

- **OPTICS Algorithm**: The clustering algorithm OPTICS (Ordering Points To Identify the Clustering Structure) is quite similar to DBSCAN. OPTICS is a density-based clustering technique, that can recognise clusters of any shape and handle noise/outliers. However, while DBSCAN builds clusters based on density-connected points, OPTICS produces a hierarchical clustering based on a set of reachability distances. In order to function, the OPTICS algorithm generates a reachability plot for the dataset. The reachability plot shows the distance at which a data point can be reached from its nearest neighbor, and it is used to identify regions of high density. The algorithm proceeds as follows:

1) Initialize the algorithm with a set of data points and a distance metric.

2) Select a random data point that has not been visited before.

3) Compute the reachability distance of the selected point to its nearest neighbor. If the nearest neighbor is a core point (i.e., it has enough neighboring points), add the selected point to the core points cluster.

4) Otherwise, add the selected point to a priority queue ordered by reachability distance.

5) Retrieve the next point from the priority queue and compute its reachability distance to its nearest neighbor. If the point is a core point, add its neighbors to its cluster and update their reachability distances.

6) If the point is not a core point, add it to the priority queue and continue.

7) Repeat steps 5-6 until all points have been visited. Parameters similar to DBSCAN algorithm were used here.

Applying Supervised Learning Algorithms: Three different unsupervised learning algorithms were applied to the pre-processed training set data to identify the best method for detecting and mapping the susceptibility of agricultural farms to flooding. The algorithms used were SVM, CART and Random Forest Algorithm.

Each Algorithm is explained further here:

- **SVM Algorithm**: The supervised machine learning algorithm SVM (Support Vector Machine) is frequently used for classification and regression analysis. To do this, a hyperplane or group of hyperplanes that maximally divide data points from various classes are constructed. The SVM algorithm works as follows:

1) Initialize the algorithm with a set of labelled training data points, where each data point is represented by a feature vector and a class label.

2) Determine the optimal hyperplane or set of hyperplanes that separates the data points of different classes. The optimal hyperplane is the one that maximizes the margin between the two classes, where the margin is the distance between the hyperplane and the closest data points.

3) Classify new data points by assigning them to the class that corresponds to the side of the hyperplane they fall on. By utilising kernel functions to change the feature space into a higher-dimensional space where the data points are separable, the SVM technique can be expanded to accommodate non-linearly separable data. Polynomial, Gaussian, and sigmoid functions are typical kernel functions.

- **CART Algorithm:** Classification and regression analysis can be performed using CART (Classification and Regression Trees), a decision tree method. It achieves its goals by iteratively subdividing the data set into smaller and smaller chunks according to the values of features and attributes of interest. The CART algorithm works as follows:

1) Initialize the algorithm with a set of labelled training data points.

2) Select the best feature or attribute to split the data based on a splitting criterion, such as Gini index or entropy for classification, or mean squared error for regression.

3) Split the data into two or more subsets based on the selected feature.

4) Till a stopping requirement is satisfied, such as getting to a minimal amount of data points or a maximum depth of the tree, repeat steps 2-3 iteratively for each subgroup.

5) Using the values of the specified features, classify fresh data points by recursively traversing the tree. The resulting tree can be visualised graphically as a hierarchy of if-then expressions, with each leaf node denoting a class label or regression value and each interior node denoting a choice based on a feature.

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- **Random Forest Algorithm**: Multiple decision trees are used in Random Forest, an ensemble learning technique, to increase forecast stability and accuracy. On a collection of randomly selected training data subsets, a set of decision trees are built, and the predictions from each tree are then combined. The Random Forest algorithm works as follows:

1) Initialize the algorithm with a set of labelled training data points.

2) Randomly select a subset of the training data points with replacement to produce a bootstrap sample.

3) Randomly select a subset of the features, without replacement, to create a subset of features to consider for splitting at each node of the tree.

4) Construct a decision tree on the bootstrap sample using the subset of features, using the chosen feature that maximises a splitting criterion such as Gini index or information gain, the data is split recursively.

5) Repeat steps 2-4 to create a forest of decision trees.

6) Classify new data points by aggregating the predictions of the individual trees, such as by taking the majority vote for classification or the average for regression.

The resulting Random Forest is a set of decision trees that have been trained on different subsets of the data and features, and therefore capture different aspects of the underlying patterns in the data. This can improve the accuracy and stability of the predictions, and also provide a measure of feature importance based on the frequency and depth of the splits.

Step1 • Apply Model on Test Images: The trained model is then applied to the test dataset.We use the test dataset to find the accuracy of our trained models.

Step 2• Compare Accuracy: The accuracy is compared between the 5 unsupervised models and 3 supervised models. The accuracy measure for unsupervised models was silhouette score and for supervised model were precision and recall.

The models with the highest scores (1 for unsupervised and 1 for supervised) were declared as the best and used in the next step.

Step 3 • Determine Extent of Flood using best algorithms: For both the best unsupervised and supervised models, the extent of the flood is determined by comparing the water percentage in pre-flood and post-flood images of the same region to determine if and by how much the water coverage has increased.

RESULTS

This section reviews the findings obtained after applying each algorithm and compares their performances based on the accuracy obtained. The accuracy for unsupervised clustering algorithms is calculated based on a parameter called Silhouette Score.The silhouette value represents an object's cohesion or similarity to its own cluster in comparison to other clusters or separation. A high number on the silhouette indicates that the object is well-matched to its own cluster but inadequately matched to clusters in the vicinity. The silhouette's range is between 1 and +1.The accuracy for supervised classification algorithms were calculated on the basis of Precision and Recall scores obtained.

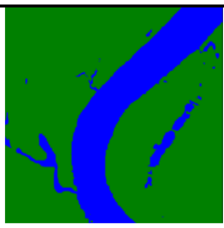
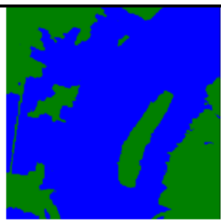
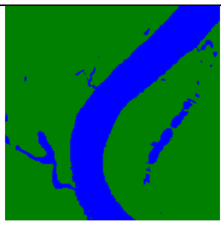
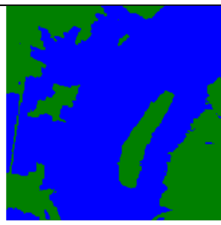
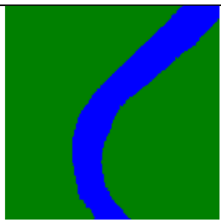
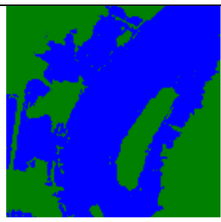
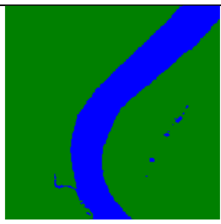
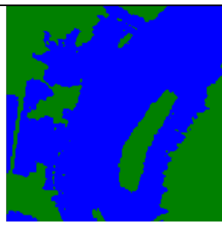
The model's precision indicates how accurately it can predict a specific category. Recall indicates how frequently the model correctly identified a particular category.

A. Results obtained with Unsupervised Clustering Algorithms

Name of Algoritm	Before Flood Image	After Flood Image

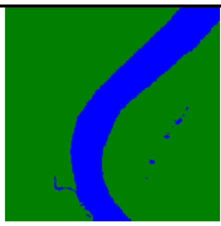
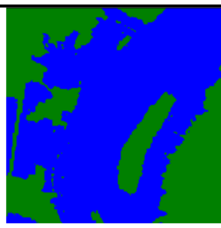
Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

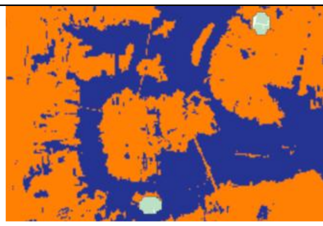
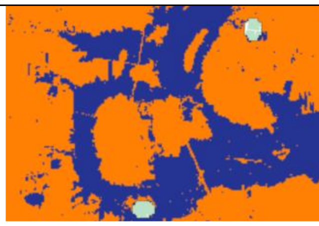
KMeans Accuracy: 80.317% Silhouette Score: 0.60634	 Fig. 3: Before Flood Before Flood Water Percentage: 23.47%	 Fig. 4: After Flood After Flood Water Percentage: 66.72%, Water Percentage Increase: 43.25%
CLARA Accuracy: 80.317% Silhouette Score: 0.60634	 Fig. 5: Before Flood Before Flood Water Percentage: 23.47%	 Fig. 6: After Flood After Flood Water Percentage: 66.72% Water Percentage Increase: 43.25%
GMM Silhouette Score: 0.48531 Accuracy: 74.266%	 Fig. 7: Before Flood Before Flood Water Percentage: 17.79%	 Fig. 8: After Flood After Flood Water Percentage: 56.65% Water Percentage Increase: 38.86%
DBSCAN Silhouette Score: 0.68119 Accuracy: 84.059%	 Fig. 9: Before Flood Before Flood Water Percentage: 19.40%	 Fig. 10: After Flood After Flood Water Percentage: 63.04% Water Percentage Increase: 43.64%

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

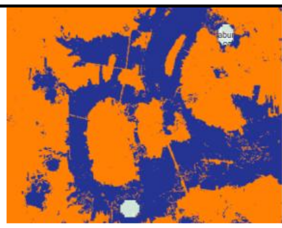
Research Article

OPTICS Silhouette Score: 0.67946 Accuracy: 83.973%	 Fig. 11: Before Flood Before Flood Water Percentage: 19.38%	 Fig. 12: After Flood After Flood Water Percentage: 62.98% Water Percentage Increase: 43.60%

B. Reults obtained with Supervised Classification Algorithms

Name of Algoritm	Before Flood Image	After Flood Image
Support Vector Machine Accuracy: 91.66%	 Fig. : Before Flood Before Flood Water Percentage: 20.91%	 Fig. : After Flood After Flood Water Percentage: 66.24%, Water Percentage Increase: 45.33%
Random Forest Classifier Accuracy=87.57%	 Fig. : Before Flood Before Flood Water Percentage: 20.91%	 Fig. : After Flood After Flood Water Percentage: 66.24%, Water Percentage Increase: 45.33%

Journal of Information Systems Engineering and Management
2025, 10(35s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Decision Tree(CART method) Accuracy: 84.23%	 Fig. : Before Flood Before Flood Water Percentage: 20.91%	 After Flood Water Percentage: 63.032%· Water Percentage Increase: 43.71% .

The results obtained after running all the algorithms are summarized in the table below:

Approach	Category	Before Flood Water Percentage	After Flood Water Percentage	Water Percentage Increase	Accuracy
K-Means	Unsupervised Clustering	23.47%	66.72%	43.25%	80.317%
CLARA	Unsupervised Clustering	23.47%	66.72%	43.25%	80.317%
GMM	Unsupervised Clustering	17.79%	56.65%	38.86%	74.266%
DBSCAN	Unsupervised Clustering	19.40%	63.04%	43.64%	84.059%
OPTICS	Unsupervised Clustering	19.38%	62.98%	43.60%	83.973%
SVM	Supervised Classification	20.91%	66.24%	45.33%	91.66%
Random Forest	Supervised Classification	20.45%	64.92%	44.47%	87.57%
Decision tree-CART	Supervised Classification	19.63%	63.032%	43.71%	84.23%

TABLE 1: Summary Of Results Obtained

CONCLUSION

In conclusion, our study examined the efficacy of supervised classification and unsupervised clustering algorithms that can be used in flood detection systems. Our results demonstrate that, in terms of accuracy, supervised algorithms outperform unsupervised methods.The Support Vector Machine (SVM) classifier gives the best accuracy in results with a 91.66% accuracy and performs better than other two supervised algorithms (Random Forest and Decision Tree) as well as all the unsupervised algorithms. Using Supervised algorithms has the drawback of necessitating labelled data for training, which can be time- and money-consuming to acquire. Unsupervised algorithms, on the other hand, are more adaptable and use less labelled data, but they could result in more false alarms and miss certain floods. Among all the unsupervised algorithms, DBScan and OPTICS were found to be most accurate with 84.059% and 83.973% accuracy respectively. The decision between supervised and unsupervised algorithms ultimately comes down to the particular application at hand as well as the accessibility of labelled data. To improve accuracy and flexibility in flood detection using remote sensing data, future research may investigate the use of hybrid techniques that incorporate both supervised and unsupervised algorithms. These algorithms can be incorporated into flood detection websites and mobile applications which can be used to provide real-time information to the owners of agricultural farms about floods and take necessary precautions in case of this natural hazard.

REFERENCES

[1] C. G. Boryan, Z. Yang, A. Sandborn, P. Willis and B. Haack". Operational Agricultural Flood Monitoring with Sentinel-1 Synthetic Aperture Radar," IGARSS 2018 - 2018 IEEE

International Geoscience and Remote Sensing Symposium, 2018, pp. 5831-5834, doi: 10.1109/IGARSS.2018.8519458.

[2] Mrinal Singha, Jinwei Dong, Sangeeta Sarmah, Nanshan You, Yan Zhou, Geli Zhang, Russell Doughty, Xiangming Xiao,Identifying floods and flood-affected paddy rice fields in Bangladesh based on Sentinel-1 imagery and Google Earth Engine,ISPRS Journal of Photogrammetry and Remote Sensing, Volume 166,2020,Pages 278-293,ISSN 0924-2716.

[3] G. R. Brakenridge, E. Andersona, S. V. Nghiemb, S. Caquard, T. B. Shabaneh, "Flood Warnings, Flood Disaster Assessments, and Flood Hazard Reduction", Proceedings of the 30th International Symposium on Remote Sensing of Environment,2020.

[4] Abdullah-Al Faisal, Abdulla-Al Kafy,Sumita Roy, "Integration of Remote Sensing and GIS Techniques for Flood Monitoring and Damage Assessment: A Case Study of Naogaon District, Bangladesh", Journal of Remote Sensing GIS,2018,ISSN: 2469-4134,DOI: 10.4172/2469-

4134.1000236

[5] A. D'Addabbo, A. Refice, G. Pasquariello, F. P. Lovergine, D. Capolongo and S. Manfreda, "A Bayesian Network for Flood Detection Combining SAR Imagery and Ancillary Data," in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 6, pp. 3612-3625, June 2018, doi: 10.1109/TGRS.2016.2520487.

[6] Huili Chen, Qiuhua Liang, Zhongyao Liang, Yong Liu, Shuguang Xie,Remote-sensing disturbance detection index to identify spatiotemporal varying flood impact on crop production,Agricultural and Forest Meteorology,Volumes 269-270, 2019,Pages 180-191,ISSN 0168-1923

[7] Sanyal, J., Lu, X.X. Application of Remote Sensing in Flood Management with Special Reference to Monsoon Asia: A Review. Natural Hazards 33, 283-301 (2018). https://doi.org/10.1023/B:NHAZ.0000037035.65105.95

[8] Dewan, A.M., Yamaguchi, Y. Using remote sensing and GIS to detect and monitor land use and land cover change in Dhaka Metropolitan of Bangladesh during 1960-2017. Environ Monit Assess 150, 237 (2018). https://doi.org/10.1007/s10661-008-0226-5

[9] Huang, M.; Jin, S. Rapid Flood Mapping and Evaluation with a Supervised Classifier and Change Detection in Shouguang Using Sentinel- 1 SAR and Sentinel-2 Optical Data. Remote Sens. 2020, 12, 2073. https://doi.org/10.3390/rs12132073

[10] Shahabi, H.; Shirzadi, A.; Ghaderi, K.; Omidvar, E.; Al-Ansari, N.; Clague, J.J.; Geertsema, M.; Khosravi, K.; Amini, A.; Bahrami, S.; Rahmati, O.; Habibi, K.; Mohammadi, A.; Nguyen, H.; Melesse, A.M.; Ahmad, B.B.; Ahmad, A. Flood Detection and Susceptibility Mapping Using Sentinel-1 Remote Sensing Data and a Machine Learning Approach: Hybrid Intelligence of Bagging Ensemble Based on K-Nearest Neighbor Classifier. Remote Sens. 2020, 12, 266. https://doi.org/10.3390/rs12020266

[11] Apoorva Shastry, Elizabeth Carter, Brian Coltin, Rachel Sleeter, Scott McMichael, Jack Eggleston,Mapping floods from remote sensing data and quantifying the effects of surface obstruction by clouds and vegetation,Remote Sensing of Environment, Volume 291,2023, 113556, ISSN 0034-4257,https://doi.org/10.1016/j.rse.2023.113556.

[12] Kuntla, Sai Kiran. (2021). An era of Sentinels in flood management: Potential of Sentinel-1, -2, and -3 satellites for effective flood management. Open Geosciences. 13. 1616-1642. 10.1515/geo-2020-0325.

[13] Chen, Qiuyang," Detecting Floods from Cloudy Scenes: A Fusion Approach Using Sentinel-1 and Sentinel-2 Imagery" in NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning, 2022. https://www.climatechange.ai/papers/neurips2022/104