ISSN: 0974-5823 Vol. 7 (Special Issue, Jan.-Mar. 2022)

International Journal of Mechanical Engineering

A Relative Study of Estimation of Pre-flood area of Flood-prone regions using Maximum Likelihood Classifier (MLC) and Minimum Distance to Means Classifier (MDM) in Cuddalore District, Tamil Nadu, India

M Saikrishna[1], Vidhya Lakshmi Sivakumar[2*]

[1]Research Scholar,Department of Civil Engineering, Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.

[2]Project Guide, Department of Civil Engineering, Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India.
Pincode: 602105.

* Corresponding author

ABSTRACT

Aim: The aim of this study is to quantify the various Land use/Land cover (LULC) regions, in particular, pre-flood regions using Maximum Likelihood Classification (MLC) and Minimum Distance to Means (MDM) classifiers on Cuddalore district, Tamil Nadu. **Materials and Methods:** Sentinel 2A data acquired from the satellite repository, USGS Earth Explorer for Cuddalore district, Tamil Nadu. After pre-processing, the image is classified using both the classifiers. Eleven numbers of classes have been chosen per group and MLC and MDM classification were performed. **Results**: On performing an independent-samples-t-test on two groups, it is revealed that there is a statistical significant difference between MLC and MDM classifiers**.** The mean and standard deviation is higher in MLC (1096.0±1564.49) than that of MDM (1096.0±1201.55). Based on the statistical analysis, it is observed there is an insignificant difference between the two groups, MLC and MDM, $p=0.256$ ($p<0.05$) **Conclusion:** The results revealed that the novel supervised flood classification using Maximum Likelihood Classifier has performed better than Minimum Distance to Mean Classifier.

Keywords: Maximum Likelihood Classifier, Minimum Distance to Means Classifier, Sentinel 2A, Flood Mapping, Classification Accuracy, Novel Supervised Flood Classification Method.

INTRODUCTION

Remote Sensing and Geographic Information System (GIS) is a computer-based tool for mapping and analyzing data and phenomena on earth. It is the idea and science of making measurements of the earth using sensors on satellites. Remotely sensed images are classified to identify, map and monitor Land Use/Land Cover features on the Earth’s surface. One such classification technique, Maximum Likelihood Classification (MLC) assumes that the statistics for each class in each band are normally distributed and calculate the probability of a pixel belonging to each LULC. Minimum Distance to Means (MDM) is used to classify based on the minimum distance between the mean of each class and the current pixel under consideration. Floods happen because of heavy storms and the common natural calamities and significant floods occur frequently and deteriorate the place that is being flooded. The term may also be referred to as the flow of a wave depending on the flowing water. In terms of hydrology, floods are a very important area of study related to agriculture, rivers, groundwater table, and civil engineering. Overflowing of water may lead to death and damage to the people (Sanyal and Lu 2004; Lin et al. 2016). It is required to identify the flood prone zones to prevent any further destruction caused (Notti et al. 2018). Flood mapping helps in the assessment of hazards (Kumar 2016; Bhatt et al. 2014), examination of potential damage caused (Kumar et al. 2018) and serves as a means of information for policy makers for relief and mitigation ("Governance and Disaster: Analysis of Land Use Policy with Reference to Uttarakhand Flood 2013, India" 2019).

According to (Díez-Herrero and Garrote 2020), a large number of articles have been published on Flood Risk Analysis is available in the Web of Science database. Total number of articles published on flood mapping over the past five years (2016-2021) in two databases, such as the IEEE Transactions on and Google Scholar are 150 and 190, respectively. A number of articles have highlighted the importance of supervised classification algorithms such as MLC and MDM. A majority of them have observed that MLC performs better than MDM. In supervised classification, the image examiner supervises the pixel arrangement process by determining the computer algorithm and numerical descriptor representing various land cover types in a scene (Patil, Desai, and Umrikar 2012). It describes a synergetic use of satellite radar image and ancillary information to detect flooded areas at their peak (Brivio et al. 2002). The central focus in the field revolves around delineation of flood zones and preparation of flood hazard maps (Sanyal and Lu 2006). The study points out some drawbacks of flood mapping based on the use of SAR and multispectral satellite data (Klemas 2015).

It is, therefore, observed from previous studies that availability of high spatial resolution data is uneconomical and hence, free-of-charge medium resolution data with appropriate classification techniques may reveal potential flooded regions (Notti et al. 2018). The aim of this study is to test and compare the accuracy and the kappa coefficient of two supervised classifiers, MLC and MDM in terms of mapping flooded zones in Cuddalore district, Tamil Nadu, India.

MATERIALS AND METHODS

A novel supervised flood classification method was used to conduct this study in the Geographic Information System (GIS) laboratory, Department of Civil Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Tamil Nadu, Chennai. The two groups considered for comparison in this study are supervised classifiers, group A-Maximum Likelihood Classification and group B-Minimum Distance to Mean. The total number of samples per group is 11 which is (N=11). The eleven classes are fallow land, deep water, shallow water, soil 1, soil 2, agricultural land, barren land, cloud, cloud shadow, coastal water and forest. The sample size was taken as 20 after being calculated for a pre-test power of 80% with an alpha value of 0.05 in clinical.com with a mean of 0.0668±0.0125 for MLC and a mean of 0.0003 (Kane, Phar, and BCPS n.d.).

The data obtained is a geometrically corrected image of the study region. One sample and eleven sample sizes of Region Of Interest (ROI) on subset data with Sentinel 2A data was the sample preparation for both the group A and group B. Layer Stacking of all the bands namely, band 3, band 4 and band 8 corresponding to central wavelengths of 559 nm, 664.0, and 832.0 nm was carried out. This is used to create a false colour composite using RGB combination. A Region of Interest consisting of the Cuddalore district is used to derive the subset resulting in the study area image. Several ROIs depicting the various LULC identified in the study region are given as training sample pixels for performing the Maximum Likelihood Classification and Minimum Distance to Means classification.

Cuddalore district covers an area of 3,564 km² and it is bounded on the north by Villupuram district, on the east by the Bay of Bengal and on the south by Nagapattinam district. The district coordinates are bounded between 11°11" and 12° 5"N latitude; and between 78° 38" and 80° 00" E longitude. Being a coastal zone, Cuddalore is usually covered by plain terrain without any high relief zone except some sedimentary high ground in Virudachalam, Cuddalore and Panruti taluks. Floods occurred in December 2020 with heavy waves and started with moderate rainfall as heavy rains (Ravikumar, Bhaskaran, and Others 2018).

In supervised classification, all the land cover and land use classes are given *a-priori* training samples or Regions of Interest (ROIs), and the image is then categorised based on the input ROIs. These ROIs are critical to the accuracy of the classification process since they define which class each pixel belongs to, in the input image. There are several classifiers under Supervised classification; the two classification algorithms used are as follows:

Maximum Likelihood Classification Algorithm :

The MLC algorithm assumes that the statistics for every class in each band are normally distributed and calculates the probability that a given pixel belongs to a specific class. Each pixel is assigned to the class that has the highest probability (that is, the maximum likelihood) (Abburu and Golla 2015) .

Minimum Distance to Means Classification Algorithm

The MDM uses the mean vectors for each class and calculates the Euclidean distance from each unknown pixel to the mean vector for each class. The pixels are classified to the nearest class (Gomathi, Greetha Priya, and Krishnaveni 2018)

The eleven classes considered for the purpose of classification are fallow land, deep water, shallow water, soil 1, soil 2, agricultural land, barren land, cloud, cloud shadow, coastal water and forest. Sample pixels or training sites for each class were collected from the FCC using the ROI tool of Envi 5.3 software. The selected ROIs are taken as input into the classifier to produce the LULC map of the study area

Statistical analysis

A statistical analysis between the two groups is performed using SPSS version 25. An independent-samples-t-test was carried out with independent variables as kappa coefficient and overall accuracy. Dependent variable is Region of Interest and Independent variables are classified images of Maximum likelihood and Minimum distance mean.

RESULTS

Figure 1 shows the conceptual flowchart of the process of image classification and how it is being subdivided into main types in classification like supervised and unsupervised classification. A layer stacked satellite false colour composite image of the Sentinel 2A data for the study region is shown in Fig. 2.

Figures 3 and 4 are the images of Maximum and Minimum Classifiers and give the results of the class of region of interest which is divided into 11 classes and its percentage of area and all these are done in the GIS and Remote Sensing lab. Figure 5 portrays that on the basis of both overall accuracy and kappa coefficient MLC has performed better than MDM.

Table 1 shows the results after performing Maximum distance classifiers and percentage of land covered as it shows Agricultural land covered by 4048.85 km^2, coastal water covered by 3687.74 km^2 and followed by the rest of classes. Table 2 is similar to Table 1 which is of Minimum distance to mean covered land area from Coastal land (3818.62 km^2) to smallest Soil2 (12.64 km^2). The overall accuracy obtained for the classification carried out in the study for both the classifiers is shown in Table 3. Table 3 displays the result of Kappa coefficient and Overall accuracy of both MLC and MDM .It has shown that the MLC has got more accuracy than the MDM with the difference of 0.069 and Kappa coefficient with difference of 0.0124.

Tables 4 and 5 exhibit the mean and standard error of two algorithms, MLC and MDM of 11 samples . Mean of both the algorithms are the same (Mean =1096.0036) and standard deviation is higher in MLC than MDM (std.deviation =1564.49013&1201.55170) and independent t test shows with significant value is same for both (Sig =1).

DISCUSSIONS

From the analysis of the overall accuracy and that of the kappa coefficients derived from the MLC and MDM methods, it is observed that the MLC performs better than the MDM method for the mapping of the flooded regions in the Cuddalore district.

Similar results related to this study were observed in several studies. (Vidhya Lakshmi and Dilli Babu 2014) observed that MLC led to the best result in terms of both overall accuracy and kappa coefficient. (Ahmad and Quegan 2012) achieved a classification accuracy of 80% and a kappa coefficient of 0.97 for MLC in comparison to Mahalanobis and MLC classifiers. (Akgün, Hüsnü Eronat, and Türk 2004) discussed that MLC has a higher correlation value of 0.79 than that of MDM fisher (linear discriminant) method and the parallelepiped methods. In another study by (Yousefi et al. 2015), MLC gave an overall classification accuracy of 94% compared to the MDM method yielding 65% for mapping dry climate regions in Iran. Perhaps, the authors also argued that the Kappa coefficients were the best metrics for use in LULC mapping. (Sisodia, Tiwari, and Kumar 2014) comment that MLC is the most robust method available for the classification of satellite images. MLC led to a classification accuracy range of 90 to 96% in UAV images for the identification of cotton rot diseases (Wang et al. 2020).

The factors affecting image classification process are pre-processing of data set and resolution of an image according to (Pal 2002), these affect the image classification to get the accuracy. Limitations of these methods are that medium to coarse resolution data provided limited results. In addition, selection of training pixels for classification also plays a vital role in overall accuracy. Hybrid classification algorithms may be resorted to, for better accuracy (Kantakumar and Neelamsetti 2015). This study can be extended to include high resolution images along with a new range of machine learning and/or deep learning based classification to provide better information of the pre-flooded regions. , (The Egyptian Journal of Remote Sensing and Space science).

CONCLUSIONS

An innovative supervised classification has been applied and tested extensively for flood prone regions. It has been concluded that supervised classification of Maximum likelihood Classifiers performed better than Minimum Distance to Mean classifier. Maximum Likelihood classifier has yielded more accuracy of image classification compared to Minimum distance to mean within the limits of the study.

DECLARATIONS

Conflict of interests

No conflict of interest in manuscript.

Authors Contributions

Author MSK was involved in data analysis, data, manuscript writing. Author SVL was involved in conceptualization, data validation, and critical review of the manuscript.

Acknowledgement:

The authors would like to thank Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly Known as Saveetha University) for providing all the necessities that helped us to finish the project.

Funding: We would like to thank the following organisations for providing financial assistance to complete the study.

1. Venkata Narasimhulu Farms, Andhra Pradesh
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

REFERENCES

1. Abburu, Sunitha, and Suresh Babu Golla. 2015. "Satellite Image Classification Methods and Techniques: A Review." *International Journal of Computer Applications*. https://doi.org/10.5120/21088-3779.
2. Ahmad, Asmala, and Shaun Quegan. 2012. "Analysis of Maximum Likelihood Classification Technique on Landsat 5 TM Satellite Data of Tropical Land Covers." *2012 IEEE International Conference on Control System, Computing and Engineering*. https://doi.org/10.1109/iccsce.2012.6487156.
3. Akgün, Aykut, A. Hüsnü Eronat, and Necdet Türk. 2004. "COMPARING DIFFERENT SATELLITE IMAGE CLASSIFICATION METHODS: AN APPLICATION IN AYVALIK DISTRICT, WESTERN TURKEY," January. http://dx.doi.org/.
4. Bhatt, G. D., Komal Sinha, P. K. Deka, and Ajay Kumar. 2014. "Flood Hazard and Risk Assessment in Chamoli District, Uttarakhand Using Satellite Remote Sensing and GIS Techniques." *International Journal of Innovative Research in Science, Engineering and Technology*. https://doi.org/10.15680/ijirset.2014.0308039.
5. Brivio, P. A., R. Colombo, M. Maggi, and R. Tomasoni. 2002. "Integration of Remote Sensing Data and GIS for Accurate Mapping of Flooded Areas." *International Journal of Remote Sensing*. https://doi.org/10.1080/01431160010014729.
6. Díez-Herrero, Andrés, and Julio Garrote. 2020. *Flood Risk Assessments: Applications and Uncertainties*.
7. Gomathi, M., M. Greetha Priya, and D. Krishnaveni. 2018. "Supervised Classification for Flood Extent Identification Using Sentinel-1 Radar Data." In *The 39th Asian Conference on Remote Sensing*, 32773284.
8. "Governance and Disaster: Analysis of Land Use Policy with Reference to Uttarakhand Flood 2013, India." 2019. *International Journal of Disaster Risk Reduction* 36 (May): 101090.
9. Kane, Sean P., Phar, and BCPS. n.d. "Sample Size Calculator." Accessed June 26, 2021. https://clincalc.com/stats/samplesize.aspx.
10. Kantakumar, Lakshmi N., and Priti Neelamsetti. 2015. "Multi-Temporal Land Use Classification Using Hybrid Approach." *Egyptian Journal of Remote Sensing and Space Sciences* 18 (2): 289–95.
11. Klemas, Victor. 2015. "Remote Sensing of Floods and Flood-Prone Areas: An Overview." *Journal of Coastal Research*. https://doi.org/10.2112/jcoastres-d-14-00160.1.
12. Kumar, Rajesh. 2016. "Flood Hazard Assessment of 2014 Floods in Sonawari Sub-District of Bandipore District (Jammu&Kashmir): An Application of Geoinformatics." *Remote Sensing Applications: Society and Environment*. https://doi.org/10.1016/j.rsase.2016.10.002.
13. Kumar, Rajesh, Rupendra Singh, Hemant Gautam, and Manish Kumar Pandey. 2018. "Flood Hazard Assessment of August 20, 2016 Floods in Satna District, Madhya Pradesh, India." *Remote Sensing Applications: Society and Environment*. https://doi.org/10.1016/j.rsase.2018.06.001.
14. Lin, L., L. Di, E. G. Yu, L. Kang, R. Shrestha, M. S. Rahman, J. Tang, et al. 2016. "A Review of Remote Sensing in Flood Assessment." In *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–4.
15. Notti, Davide, Daniele Giordan, Fabiana Caló, Antonio Pepe, Francesco Zucca, and Jorge Galve. 2018. "Potential and Limitations of Open Satellite Data for Flood Mapping." *Remote Sensing*. https://doi.org/10.3390/rs10111673.
16. Pal, Mahesh. 2002. "Factors Influencing the Accuracy of Remote Sensing Classifications: A Comparative Study," January. http://dx.doi.org/.
17. Patil, Manisha B., Chitra G. Desai, and Bhavana N. Umrikar. 2012. "Image Classification Tool for Land Use/land Cover Analysis: A Comparative Study of Maximum Likelihood and Minimum Distance Method." *International Journal of Geology, Earth and Environmental Sciences* 2 (3): 189–

96.

18. Ravikumar, P., G. Bhaskaran, and Others. 2018. "Assessment of Changes in the Land Use and Land Cover Pattern Using Matrix Union Method: A Case Study of Cuddalore District, Tamil Nadu, India." *International Journal for Research in Applied Science and Engineering Technology* 6 (4): 4828–32.
19. Sanyal, Joy, and X. X. Lu. 2004. "Application of Remote Sensing in Flood Management with Special Reference to Monsoon Asia: A Review." *Natural Hazards*. https://doi.org/10.1023/b:nhaz.0000037035.65105.95.
20. ———. 2006. "GIS-Based Flood Hazard Mapping at Different Administrative Scales: A Case Study in Gangetic West Bengal, India." *Singapore Journal of Tropical Geography*. https://doi.org/10.1111/j.1467-9493.2006.00254.x.
21. Sisodia, Pushpendra Singh, Vivekanand Tiwari, and Anil Kumar. 2014. "Analysis of Supervised Maximum Likelihood Classification for Remote Sensing Image." *International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014)*. https://doi.org/10.1109/icraie.2014.6909319.
22. Vidhya Lakshmi, S. V., and S. Dilli Babu. 2014. "A Comparison of Image Classifiers for Detecting Change in Chennai Using Remotely Sensed Data" 9 (22): 5672–78.
23. Wang, Tianyi, J. Alex Thomasson, Chenghai Yang, Thomas Isakeit, and Robert L. Nichols. 2020. "Automatic Classification of Cotton Root Rot Disease Based on UAV Remote Sensing." *Remote Sensing*. https://doi.org/10.3390/rs12081310.
24. Yousefi, Saleh, Somayeh Mirzaee, Mehdi Tazeh, Hamidreza Pourghasemi, and Haji Karimi. 2015. "Comparison of Different Algorithms for Land Use Mapping in Dry Climate Using Satellite Images: A Case Study of the Central Regions of Iran." *Desertification Control* 20 (1): 1–10.

FIGURES AND TABLES

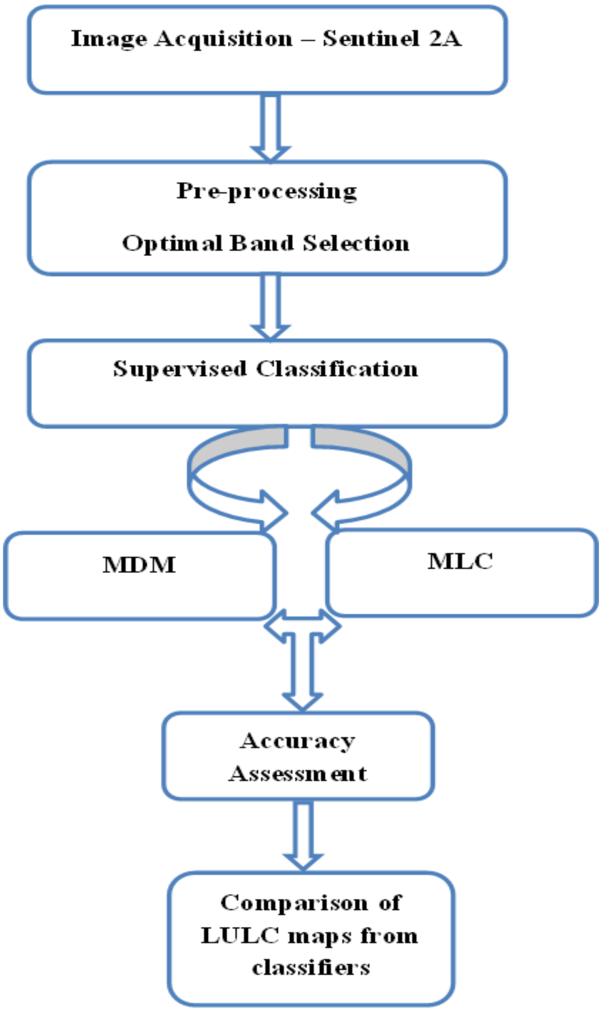

Fig. 1. A flow chart showing the methodology adopted in this study.

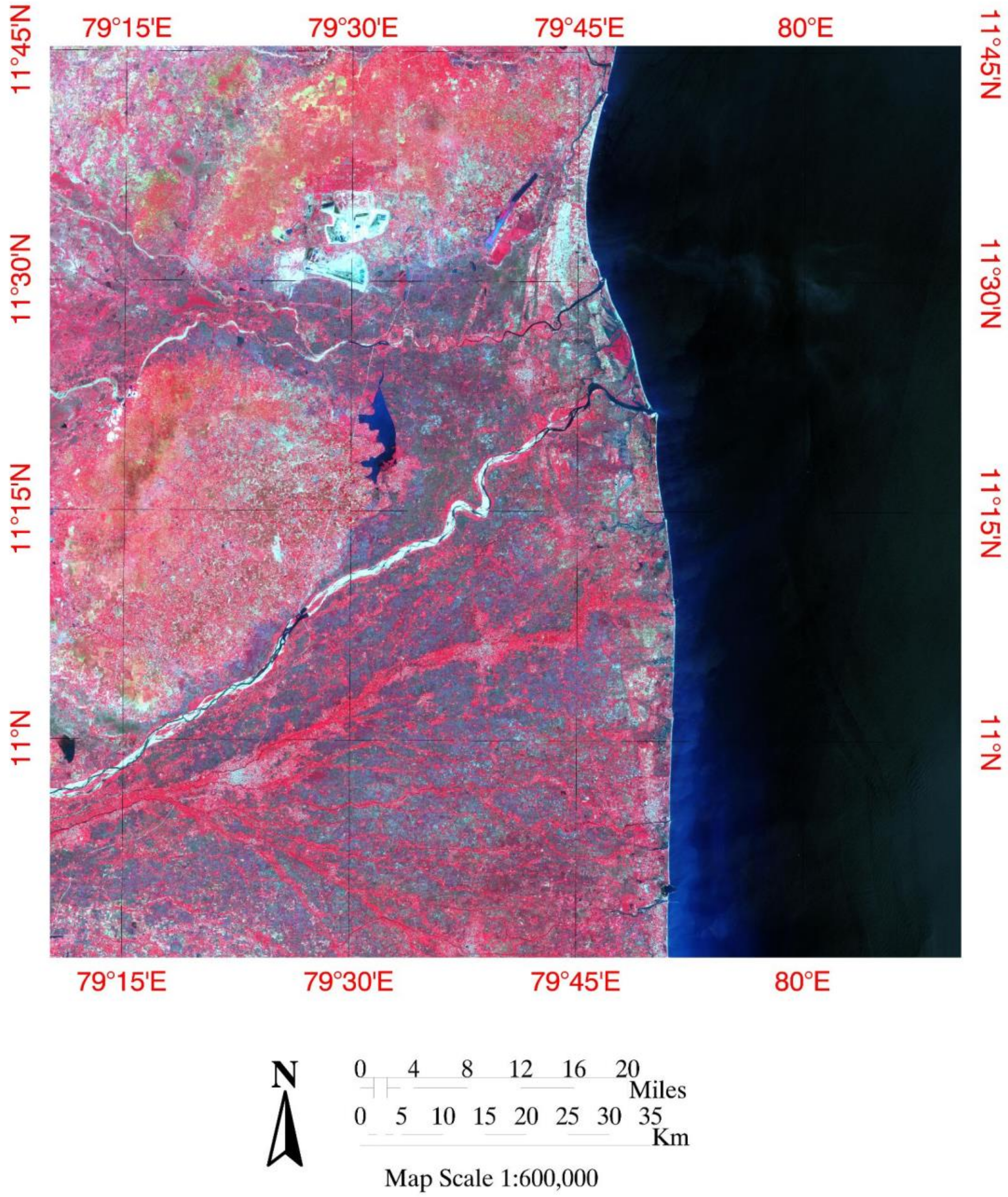

Fig. 2. A false colour composite image of the study region generated after layer stacking from Sentinel 2A captured on (Source: U.S. Geological Survey).

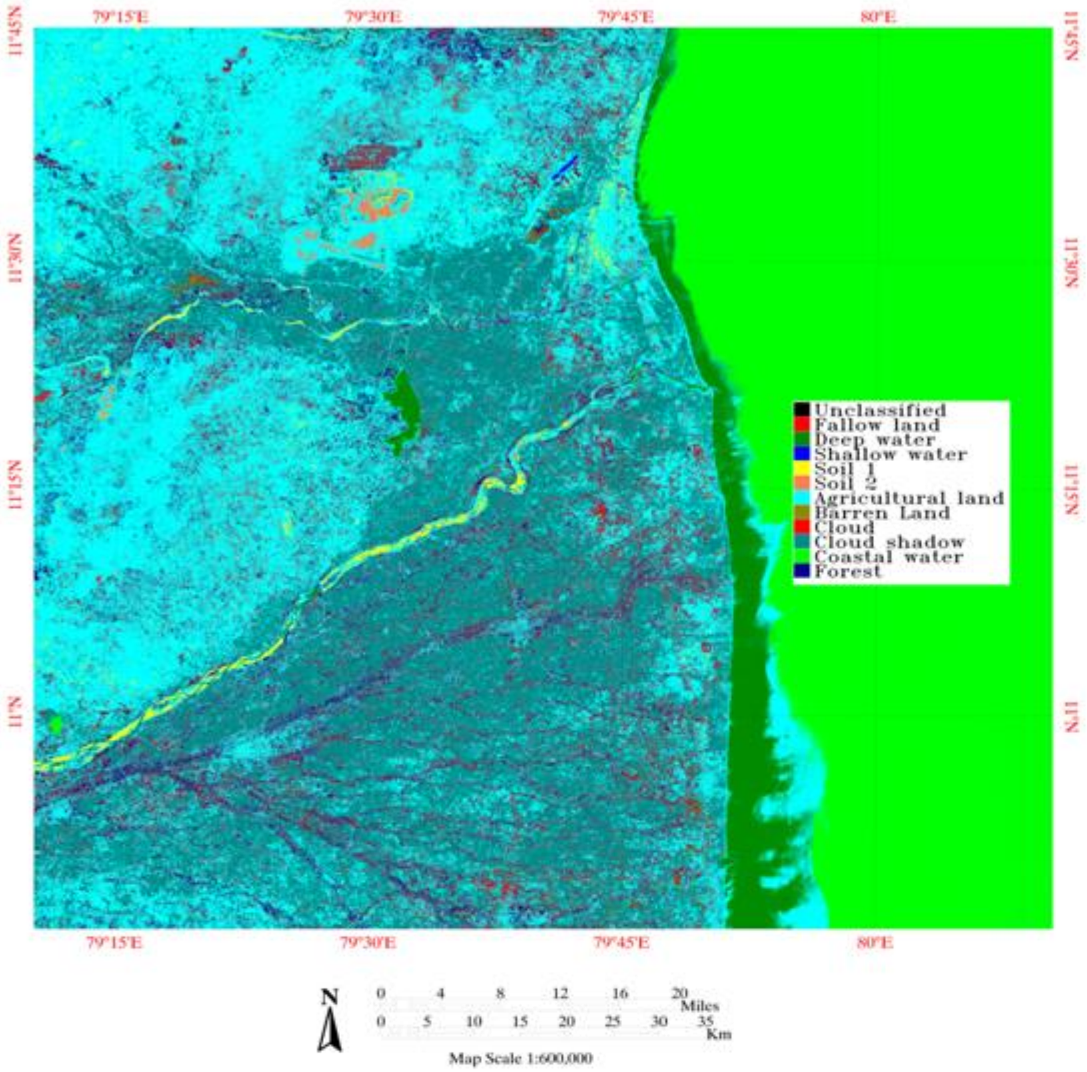

Fig. 3. Maximum Distance Likelihood Classifier of Sentinel 2A of pre-flood using Envi 5.3

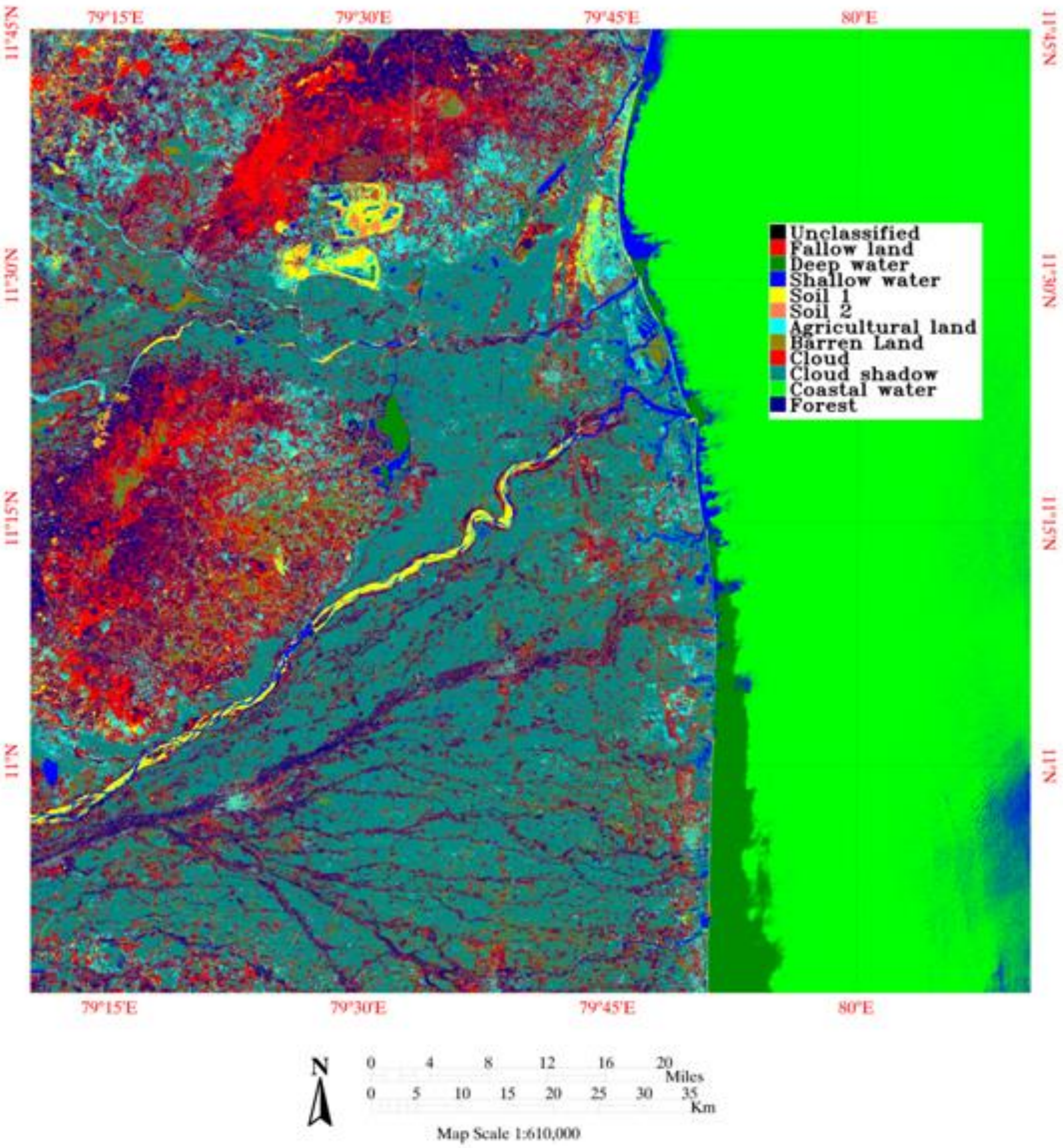

Fig. 4. Minimum Distance to Mean Classifier of Sentinel 2A of pre-flood using Envi 5.3

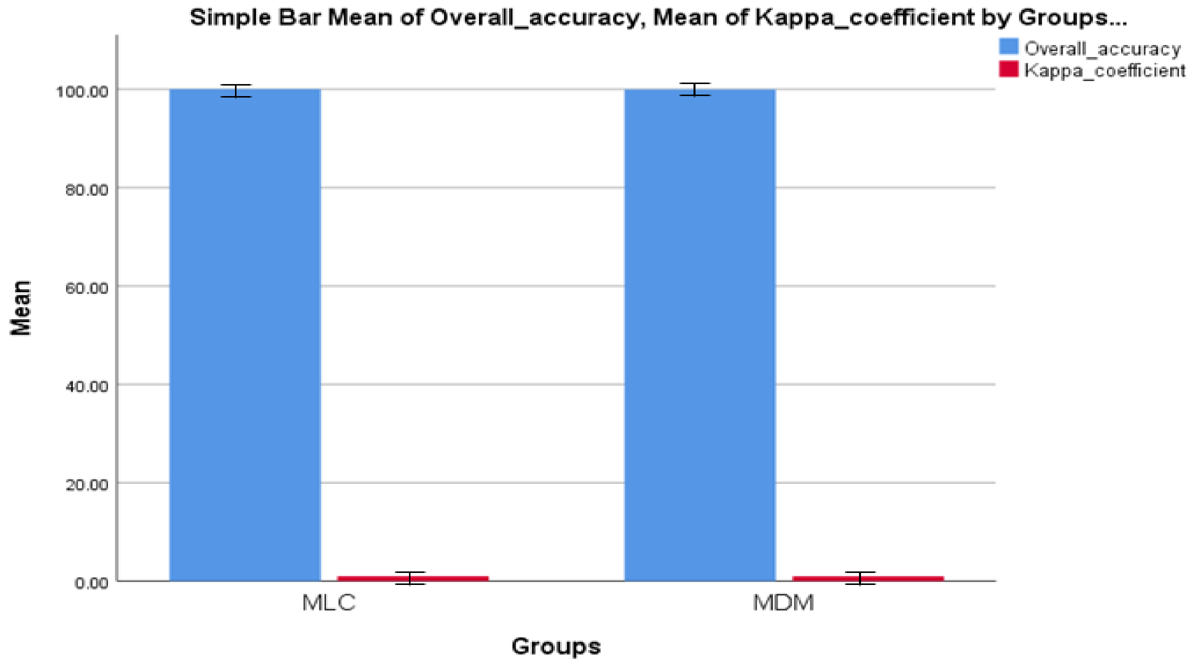

Fig. 5: There are two metrics obtained as a result,accuracy and kappa coefficient. Based on the accuracy assessment , accuracy MLC is better than that of MDM. The same stands correct in the case of the kappa coefficient (+/- 1SD)

Table 1: Result of the classified area of the various LULC considered in this study.

Class Summary	**Pixel count**	**Percentage of area in the satellite image**	**Area in km^2**
Unclassified	0	0	0
Fallow land	35383	0.02	3.53
Deep water	3714388	3.08	371.43
Shallow water	30597	0.02	3.05
Soil1	738848	0.61	73.88
Soil2	328101	0.27	32.81
Agricultural land	40488535	33.58	4048.85
Barren land	552378	0.45	55.23
Cloud	7271645	6.03	727.16
Cloud shadow	26130010	21.67	2613.00
Coastal land	36877406	30.58	3687.74
Forest	4393109	3.64	439.31

Table 2: Class statistics obtained as output after performing MDM classification on the satellite image. The Table represents the number of pixels, percentage of area, and area in km2 under each class.

Class Summary	Pixel count	Percentage of area in the satellite image	Area in km^2
Unclassified	0	0	0
Fallow land	4548293	3.77	454.82
Deep water	2403869	1.99	240.38
Shallow water	3014457	2.50	301.44
Soil1	2158368	1.79	215.83
Soil2	126474	0.10	12.64
Agricultural land	8061759	6.68	806.17
Barren land	6920222	5.74	692.02
Cloud	14711167	12.20	1471.11
Cloud shadow	27963448	23.19	2796.34
Coastal land	38186218	31.67	3818.62
Forest	12466125	10.34	1246.61

Table 3: Overall accuracy and Kappa coefficient from ground truth table. The area extent calculated from MLC outperformed the MLC algorithms and the pre-flood area of 374 km^2 was noted.

	Maximum Likelihood classifier	Minimum Distance to Mean
Overall Accuracy (%)	99.995	99.926
Kappa coefficient (no units)	0.9992	0.9868

Table 4: Group Statistics showing the mean, standard deviation and standard error mean values for the two groups in the study with 11 samples for each group such as MLC and MDM.The confidence interval is kept at 95% and alpha value at 0.05. It shows MLC performed better than that of MDM in terms of accuracy and kappa coefficient.

Groups	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**
MLC	11	1096.0036	1564.49013	471.71152
MDM	11	1096.0036	1201.55170	`362.28147

Table 5: Independent-samples-t-test statistical results derived using SPSS statistical software and showing an insignificant statistical difference (p = 0.256; p > 0.05) between the groups, which means that the area of the different LULC in the images as observed by both the classifiers are identical.

	Levene's Test for Equality of Variances		**t-test for Equality of Means**						
	F	**Sig**	**t**	**df**	**Sig (2-tailed)**	**Mean Difference**	**Std Error Difference**	**95%Confidence interval of the Difference**	
Areal_extent								**Lower**	**Upper**
Equal Variances assumed	1.367	0.256	.000	0.20	1.000	.00000	594.7770	-1240.68	1240.683
Equal Variances Not assumed			.000	18.752	1.000	.00000	594.7770	-1246.00	1245.998