Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 11, No. 2, June 2023, pp. 431~441
ISSN: 2089-3272, DOI: 10.52549/ijeei.v11i2.4270

A Novel Approach for Improving Post Classification Accuracy of Satellite Images by Using Majority Analysis

Swasti Patel[1], Dr. Priya Swaminarayan[2]
[1]Department of Computer Science and Engineering, Parul University, India
[2]Department of Computer Application, Parul University, India

Article Info

Article history:

Received Oct 29, 2022
Revised Jan 14, 2023
Accepted Feb 21, 2023

Keywords:

Change Detection
Image Classification
SVM
Majority Analysis
Forest Fire

ABSTRACT

In past one year, due to climatic changes and some anthropogenic activities, the forests of Uttarakhand are burning. To identify the damage caused by the forest fires, an area of Nainital district has been taken for the study. Multi temporal Landsat 7 images were taken from April - 2020 and April – 2021. This paper shows a novel approach to increase the accuracy of the classified image. The Support Vector Machine classification is first done and then to improve the accuracy of the classified image, a post-classification technique called Majority Analysis is applied. This method helps to classify the unclassified pixel and it also smoothens out the boundary of the classified pixels, leading to higher accuracy rate. The classification accuracy has improved significantly for April 2020 and April 2021 images from 89.35% to 98.71% and from 88.52% to 99.76% respectively. The change detection study showed a drastic increase in the barren land due to the forest fires and on the contrary, the forest, scarce forest and the shrub land areas have decreased.

Corresponding Author:

Swasti Patel,
Department of Computer Science and Engineering,
Parul University,
Parul University, P.O.Limda, Ta.Waghodia – 391760, Dist. Vadodara, Gujarat (India).
Email: swastiwala16@gmail.com

1. INTRODUCTION

Uttarakhand has seen a number of fire incidents in the past. It is a Himalayan state, due to which the ecology is quite delicate. It has largest number of forest cover in the northern states. It is also a habitat wide range of flora and fauna that are preserved by sanctuaries, national park and bio-sphere reserve. The forest fire in state is classified in three types: 1) Surface fires, 2) Ground fires and 3) Crown fires. Surface fire spread very quickly and mostly consumes small vegetation and surface litter. Ground fire consumes the organic matter. It mostly reaches out to musk, duff, or peat available underneath the surface litter of the forest floor. The Crown fire burns the top of the trees and shrubs without having any close link with surface fire [1].

Change detection (CD) is essential for determining the changes a specific land cover underwent in the due course of time. For performing CD, image analysis and classification of satellite images is necessary [2]. CD is contemplated as one of the biggest challenges in the field of remote sensing application areas [3, 4]. CD has crucial importance in various applications such as deforestation, forest burning, study of land cover dynamics, vegetation change and surveillance of shifting cultivation [5]. These issues involve examining of larger topographical areas [6]. CD is used to ascertain change in a geographical area by incorporating multiple images taken at different time instances [7]. The considerable steps to comprehend the CD process are 1) image pre-processing, 2) finding difference image (DI), and 3) difference image analysis. Here, the image classification is done using supervised machine learning technique – Support Vector Machine (SVM). The pixels of same spectral signature are classified as one class [8-12]. For post – classification, a hybrid technique is developed by combination of SVM and Major Analysis (MA) is proposed. In this technique, the unclassified pixels are compared with a group of pixels (neighbourhood) and based on the majority of the surrounding pixels, the unclassified pixel is labelled accordingly. The objective of this technique is to improve the accuracy of the classified image. Lastly, a CD is performed on an area that was severely affected during the forest fires in Nainital, Uttarakhand, India.

2. STUDY AREA

For the research purpose, an area of Nainital District, of Uttarakhand state, India; has been taken. It covers a geographical area of 195 km^2. The study area is bounded by 29° 13' 22'' to 29° 18' 30" North latitude and 79° 13' 22" to 79° 29' 30" East longitudes as shown in Figure. 1. The area has vast variety of terrain such as agricultural fields, dense forest, scarce forest, infrastructure, etc. The area has prominent towns such as Haldwani, Kaladhungi, Chausala, Basani and Fatehpur. According to the forest survey of India; Nainital, Pauri Garhwal and Almora are the most fire prone zones of the Uttarakhand state. For comparing the effect of fire from 2020 to 2021, multi-temporal imagery has been taken from Landsat 7. The images captured are dated 7th April, 2020 and 26th April, 2021.

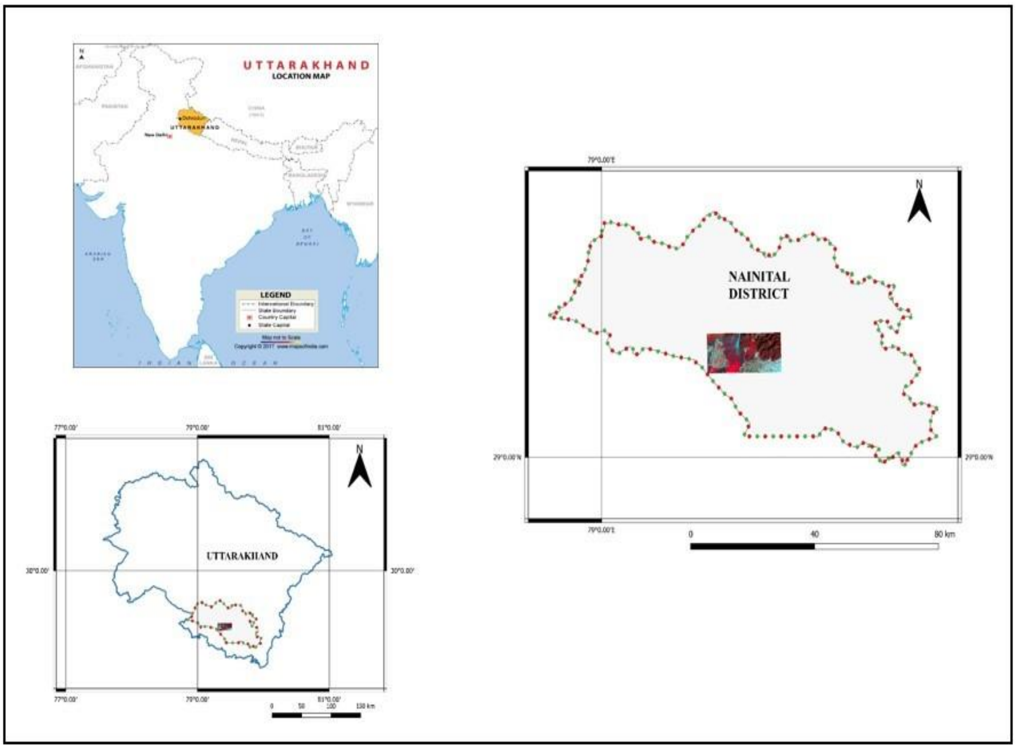

Figure 1. Location of Nainital District – Study Area

3. PROPOSED METHODOLOGY

As shown in the Figure 2, the first and foremost step is to perform forest fire study to understand the nature of it, factors that cause it, impacts of the fire, and so on. Next it is important to decide the image dates from which we want the images to be. For this study two images have been chosen with span of one year in-between them. The next step is to perform image pre-processing, in this, atmospheric corrections are made. Since the study area is in Nainital District, extraction needs to be done for that area from the satellite image, for this a mask is created using QGIS 3.18, and layered over the satellite image. This layering will give the exact coordinates of the study area, which is then cropped. Once the image is cropped, it needs to be classified into various classes based on the spectral signature reflected by the pixels. Here, seven different signatures have been identified and based on that supervised classification is performed having seven classed namely: Sediments, Agriculture, Forest, Scarce Forest, Shrub land, Barren land and Infrastructure. Afterwards, samples from each class are to be collected which will act as training data for SVM. After the sampling is completed, the classification takes place and the accuracy of the classified image is checked. The accuracy assessment is followed by the post classification technique: Majority Analysis (MA). This process will iterate until satisfactory accuracy is achieved. Finally, with images having higher accuracy are used for identifying the changes the area underwent due to the forest fires.

Image Pre-processing

Panchromatic sharpening (PAN sharpening) is a famous technique that is widely used to improve the quality satellite image. It combines low resolution colour bands with corresponding high resolution grayscale bands. To improve the image resolution, PAN sharpening of the images is performed for stronger visibility of the pixels as shown in Figure.3. Panchromatic sharpening utilizes spatial information available in the high-resolution grayscale band and the colour information present in the multispectral bands to create a high-resolution colour image. This will prominently lead to increase of resolution of the colour information in the data set. The range of the panchromatic band is 15m, whereas other low-resolution bands have range of 30m. But the panchromatic band will capture the light twice as much as compared with the other bands in the Landsat 8 satellite. This results into much sharper images.

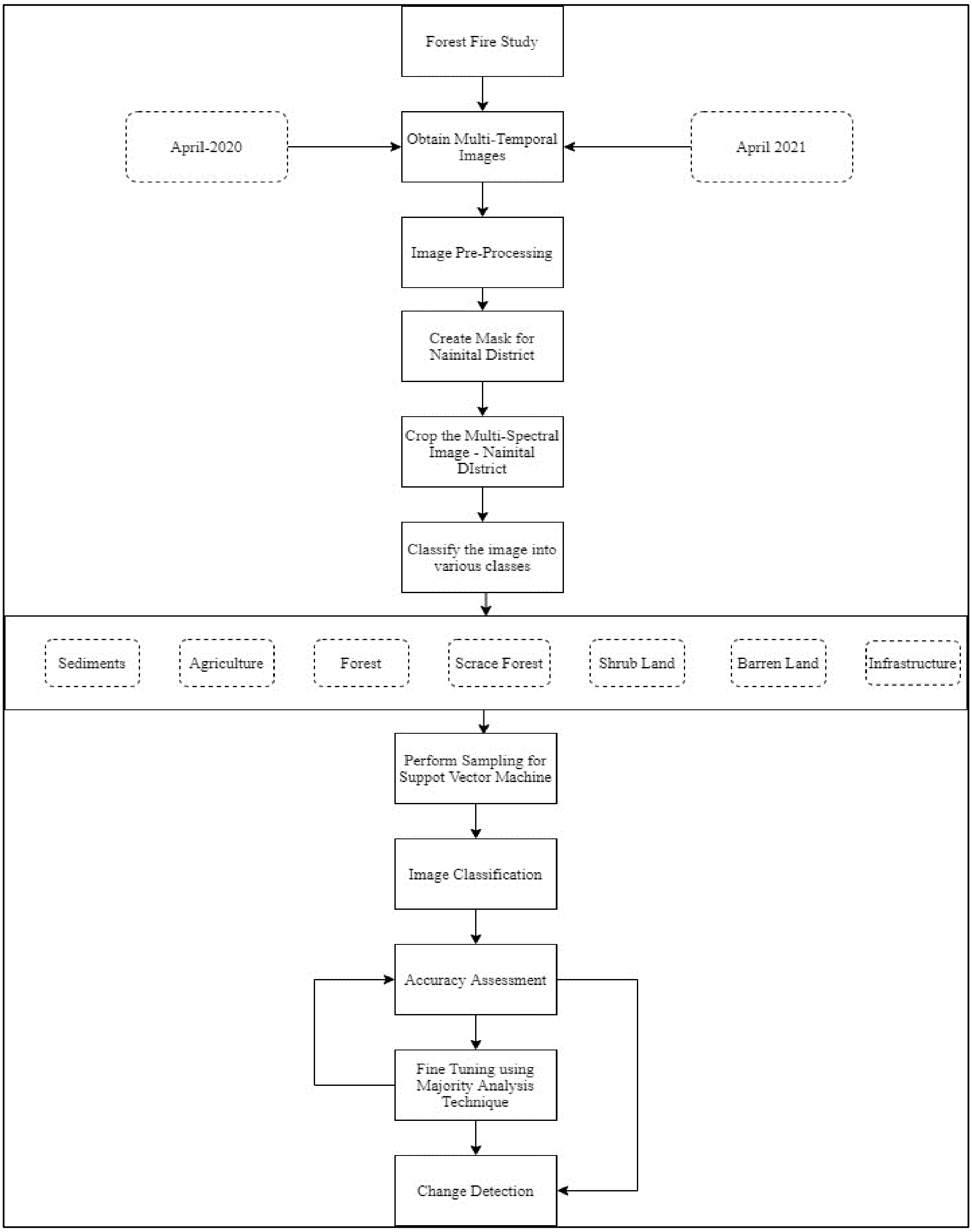

Figure 2. Flow of the proposed system

Figure 3. PAN sharpening of the study area

Algorithm

Step 1	**Randomly determine a size of the group of pixels.**
Step 2	Form an analysis window around pixel p. Let this be called as a neighbourhood of the misclassified pixel.
Step 3	Count the number of pixels belonging to different classes.
Step 4	If there is a majority of pixels, then label the misclassified pixel with the majority. a. Initialize count = 1 and majindex=0. b. Iterate each pixel and maintain a count of majority pixel. c. Maintain a majority index, majindex. d. If next pixel is same as majority, then increment majindex by 1. e. Else decrement maj_index by 1. f. If count=0 then do; majindex = currpixel and count =1
Step 5	Traverse through neighbourhood and find the count of the majority element. If count is greater than half the size of neighbourhood then print pixel as majority of the labelled class.
Step 6	Repeat steps 4 and 5 till all pixels are labelled.

Tools and Technology

For supervised classification technique, Support Vector Machine (SVM) is used. The kernel used is Radial Basis Function (RBF) and the gamma value given is 0.125. For MA, the neighbourhood size is 5X7 and the pixel weight is 5. The pixel weight determines the number of times the unclassified pixel will iterate before it is labelled. For training set of SVM, the image is classified into seven different classes based on the spectral signatures: Sediments, Agriculture, Forest, Scarce Forest, Shrub Land, Barren Land and Infrastructure. The classified image is different from what normal eyes would perceive, table 2 shows the colour interpretation of the classified image. Table 1 shows the number of samples collected from each image.

Table 1. Training sample collection

Class Name	**Before Image (07-04-2020)**	**After Image (26-04-2021)**
	Number of Training Samples	**Number of Training Samples**
Sediments	557	558
Agriculture	1455	1167
Forest	935	301
Scarce Forest	495	660
Shrub Land	740	260
Barren Land	645	905
Infrastructure	188	384
Total	**5015**	**4235**

Table 2. Colour interpretation of the classified image

Sr. No.	Color	Area Classified
1	Brown	Sediments
2	Sea Green	Agriculture
3	Dark Green	Forest
4	Light Green	Scarce Forest
5	Yellow	Shrub Land
6	Maroon	Barren Land
7	White	Infrastructure

4. RESULTS AND DISCUSSION

Initially the multi-spectral image was classified by machine learning algorithm, to improve the accuracy further, majority analysis was applied on the classified image. Figure 4 and 5 show the supervised classification and post-classification results. Some of the prominent differences are highlighted in Figure 6. For the year 2020, as shown in Figure 6 - 1(a), 1(b) and 3(a), 3(b), it is clearly seen that after applying majority analysis technique, the sediments were more clearly classified. Similarly comparing 2(a) and 2(b), the scarce forest area is correctly classified instead of forest area.

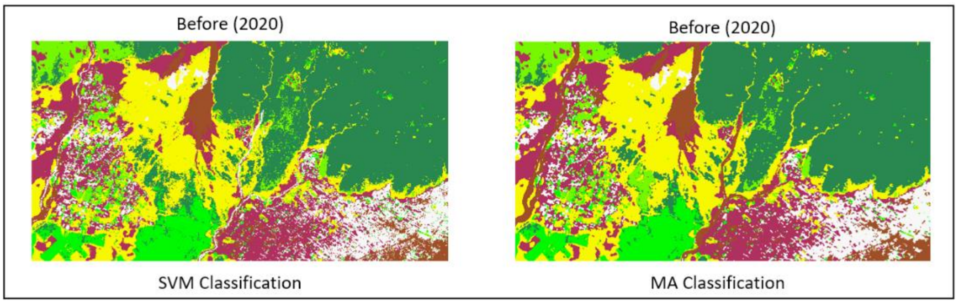

Figure 4. Classification comparison of SVM and MA (2020)

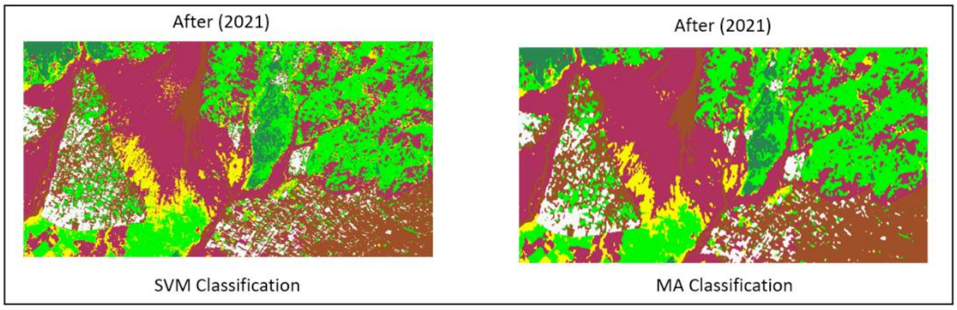

Figure 5. Classification comparison of SVM and MA (2021)

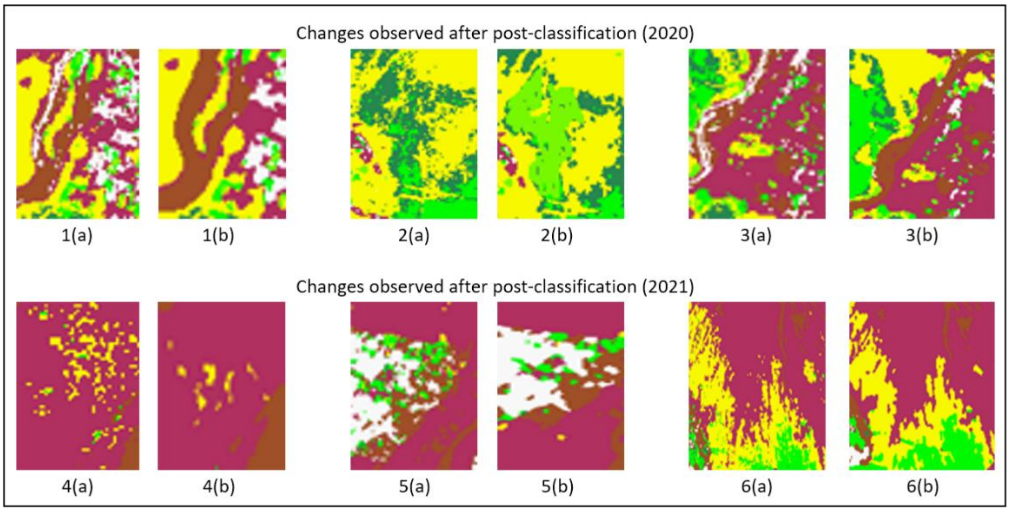

Figure 6. Changes observed after post-classification is performed on SVM

Similar changes are observed for the year 2020, as shown in Figure 6 – 4(a) and 4(b) depict that some pixels which were previously classified as shrub land were labelled as barren land in the post classification phase. For 5(a), 5(b), 6(a) and 6(b), it is clearly seen that the boundary has smoothened and minor changes in the labelling of the classification is also observed. Confusion matrix has been generated for measuring the performance of the existing machine learning algorithm and the proposed method. This matrix helps in knowing the performance of the classification model on a set of test data for that the true values are known.

Table 3. Confusion matrix of SVM – 2020

	SVM Ground Truth - 2020								
	Class	Sediments	Agriculture	Forest	Scarce Forest	Shrub Land	Barren Land	Infrastructure	TOTAL
SVM Predicted- 2020	Sediments	449	8	11	12	0	11	28	519
	Agriculture	12	1374	0	11	7	9	13	1426
	Forest	7	7	891	27	15	8	16	971
	Scarce Forest	10	22	11	397	13	23	12	488
	Shrub Land	54	14	7	12	694	13	5	799
	Barren Land	12	11	12	26	11	569	7	648
	Infrastructure	13	19	3	10	0	12	107	164
	TOTAL	557	1455	935	495	740	645	188	5015
	Overall Accuracy	**89.35%**							
	Kappa Coefficient	**0.870**							

Table 4. Confusion matrix of MA – 2020

	MA Ground Truth - 2020								
	Class	Sediments	Agriculture	Forest	Scarce Forest	Shrub Land	Barren Land	Infrastructure	TOTAL
MA Predicted - 2020	Sediments	697	0	0	0	0	0	13	710
	Agriculture	0	1419	0	0	0	0	0	1419
	Forest	0	0	864	9	0	0	0	873
	Scarce Forest	0	0	4	563	0	7	0	574
	Shrub Land	41	0	0	0	705	1	1	748
	Barren Land	3	0	0	5	0	570	3	581
	Infrastructure	5	0	0	0	0	0	105	110
	TOTAL	746	1419	868	577	705	578	122	5015
	Overall Accuracy	**98.17%**							
	Kappa Coefficient	**0.978**							

As shown in table 3, 4, 5 and 6, each row of the matrix is a number of predicted pixels in the class and each column of the matrix corresponds to the actual ground truth pixels. The sum of correct predictions for a class are assigned into the predicted column and expected row for that class value whereas the sum of incorrect predictions for a class are assigned to the expected row for that class value and the predicted column for that specific class value. By looking the values in red colour, we can say that the ground truth and predicted classes are achieved better in the majority analysis technique. Based on the detailed accuracy report overall accuracy and the kappa coefficient have been calculated for both the methods and there is definitely an increase when the post classification technique is applied on the SVM classified image. The overall accuracy for images in year 2020 for the proposed technique is 98% and kappa coefficient is 0.97 as compared to only 89% accuracy and kappa coefficient 0.87 is achieved in alone SVM classification. Similarly for the images in 2021, the overall accuracy for the proposed technique is 99% and kappa coefficient is 0. as compared to only 88% accuracy and kappa coefficient 0.86 is achieved in alone SVM classification.

Table 5. Confusion matrix of SVM – 2021

	SVM Ground Truth - 2021								
	Class	Sediments	Agriculture	Forest	Scarce Forest	Shrub Land	Barren Land	Infrastructure	TOTAL
SVM Predicted- 2021	Sediments	501	11	25	0	18	4	12	571
	Agriculture	13	1098	7	8	13	28	13	1180
	Forest	4	11	224	15	4	9	3	270
	Scarce Forest	15	8	28	589	8	19	11	678
	Shrub Land	7	9	6	7	202	13	22	266
	Barren Land	12	13	2	14	6	816	4	867
	Infrastructure	6	17	9	27	9	16	319	403
	TOTAL	558	1167	301	660	260	905	384	4235
	Overall Accuracy	**88.52%**							
	Kappa Coefficient	**0.860**							

Table 6. Confusion matrix of MA - 2021

	MA Ground Truth - 2021								
	Class	Sediments	Agriculture	Forest	Scarce Forest	Shrub Land	Barren Land	Infrastructure	TOTAL
MA Predicted - 2021	Sediments	506	0	0	0	0	0	0	506
	Agriculture	0	1263	0	0	0	0	0	1263
	Forest	0	0	235	0	0	0	0	235
	Scarce Forest	0	0	4	621	0	0	0	625
	Shrub Land	0	2	0	0	165	1	0	168
	Barren Land	1	0	0	0	2	1103	0	1106
	Infrastructure	0	0	0	0	0	0	332	332
	TOTAL	507	1265	239	621	167	1104	332	4235
	Overall Accuracy	**99.76%**							
	Kappa Coefficient	**0.997**							

Based on the confusion matrix, Figure 7 and 8 represent the individual classification results of SVM and MA. By studying the comparison graph it can be concluded that for the image of 2020, there has been rise in the accuracy of the classified image for classes: sediments, agriculture, forest, scarce forest, shrub land, barren land and infrastructure by 13%, 6%, 5%, 17%, 7%, 10% and 30% respectively. Similarly, for image of 2021, there has been rise in the accuracy of the classified image for classes: sediments, agriculture, forest, scarce forest, shrub land, barren land and infrastructure by 10%, 5%, 26%, 11%, 19%, 9% and 17% respectively. Another comparison in terms of the capacity of the classification technique is also represented in the Figure 9 and 10. These charts show the performance of the existing method and the proposed technique. It can be observed that, the total percentage of the unclassified pixels was 11%, which reduced to 2% for image of 2020 and 0% for image of 2021 when MA method was applied on the classified image.

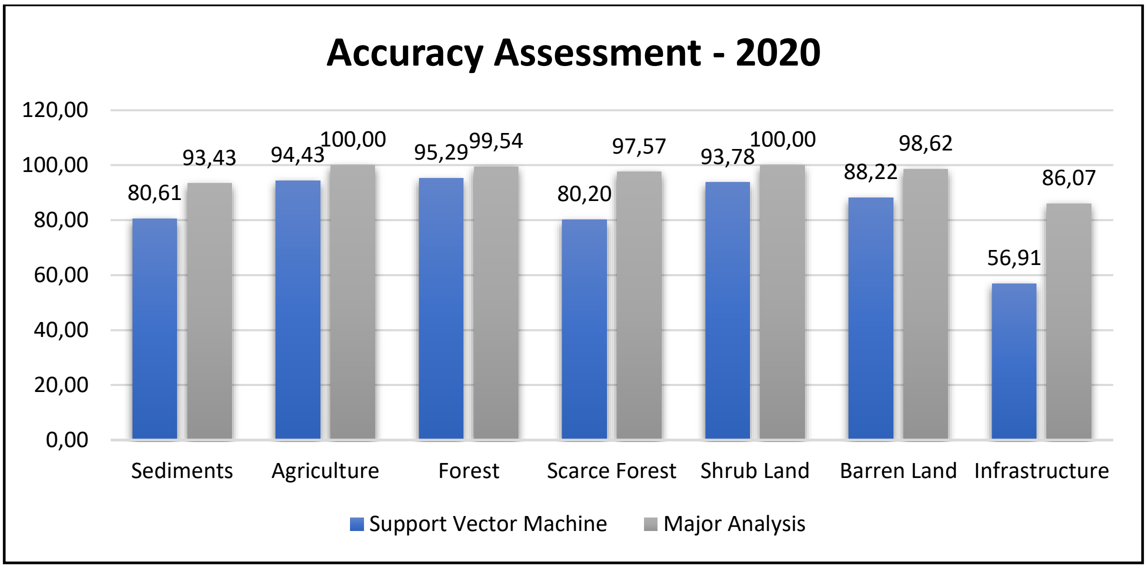

Figure 7. Accuracy assessment between SVM and MA – 2020

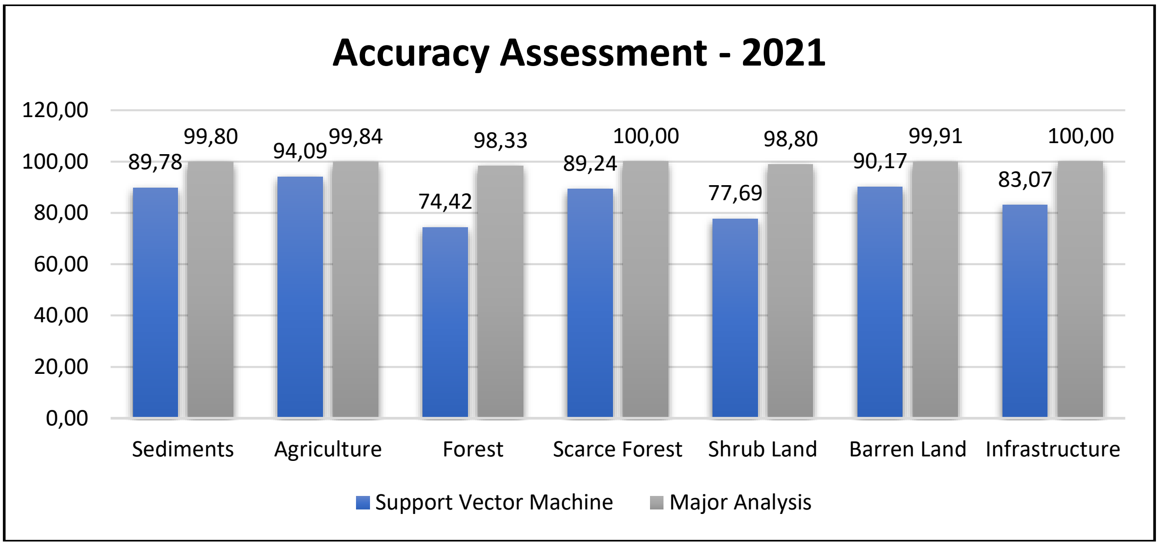

Figure 8. Accuracy assessment between SVM and MA – 2021

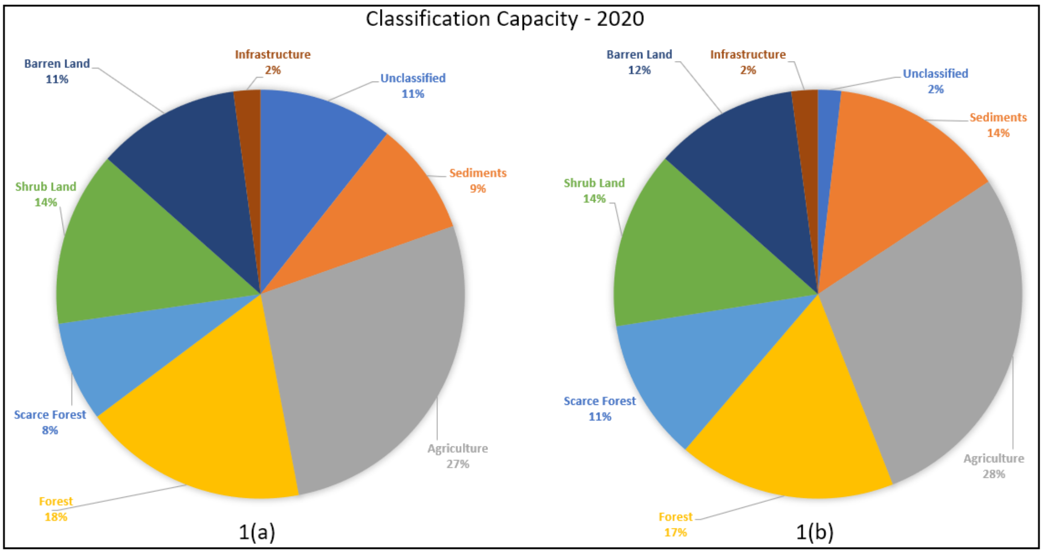

Figure 9. Classification capacity of SVM and MA – 2020. 1(a) SVM and 1(b) MA

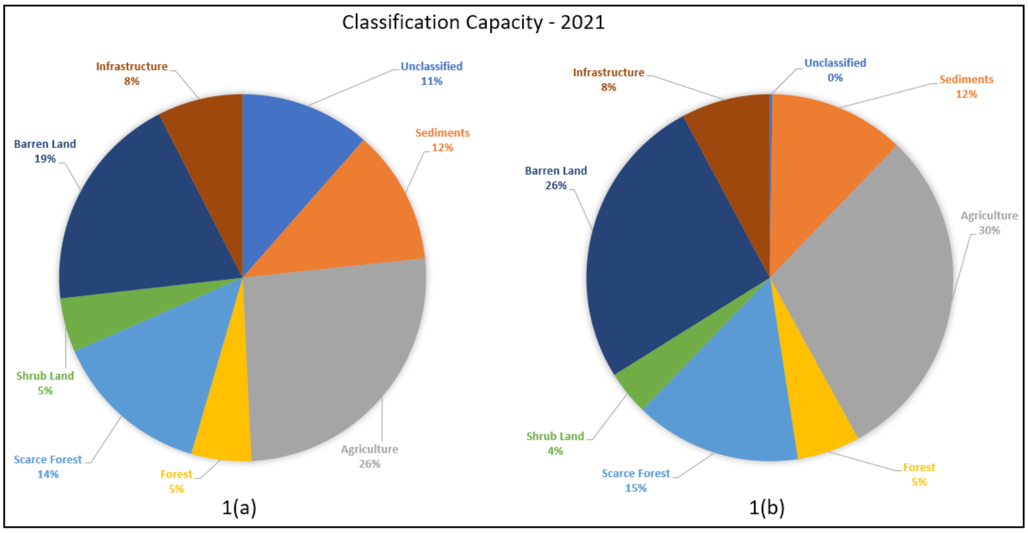

Figure 10. Classification capacity of SVM and MA – 2021. 1(a) SVM and 1(b) MA

Change Detection

To identify the damage caused by the forest fires, change detection has been performed on the multi-temporal images. Table 7 shows the readings of the area that underwent changes due to the fires caused in the forest and nearby regions. The change statistics tables will list the initial state classes in the columns and the final state classes in the rows. The total class row shows the total number of pixels in each initial state class, and total class column shows the total number of pixels in each final state class. The total row column is a class-by-class addition of all final state pixels that come under the selected initial state classes. The class changes row displays the total number of initial state pixels that changed from one class to another class. The image difference row is the difference of total number of correctly classed pixels in the two images, calculated by subtracting the initial state class totals from the final state class totals.

Table 7. Change detection

		INITIAL STATE 2020								
		Sediments	Agriculture	Forest	Scarce forest	Shrub land	Barren land	Infrastructure	Total row	Total class
FINAL STATE 2021	Sediments	86.53	10.62	0.03	4.79	5.05	32.23	81.39	100	100
	Agriculture	0.59	63.79	0.79	19.89	2.97	3.54	2.14	100	100
	Forest	0.04	0.64	55.56	17.68	1.75	0.57	0.08	100	100
	Scarce forest	0.01	0.72	5.74	37.71	0.13	0.65	0.03	100	100
	Shrub Land	0.01	13.59	8.45	5.41	8.71	1.13	0.14	100	100
	Barren Land	11.81	6.02	29.28	10.25	75.20	41.17	6.37	100	100
	Infrastructure	0.95	4.61	0.14	4.25	6.19	20.71	9.86	100	100
	Total class	100	100	100	100	100	100	100		
	Class changes	13.41	36.21	44.44	62.29	91.29	58.83	90.14		
	Image difference	315.30	3.58	-40.29	-19.65	-70.21	91.17	-33.40		

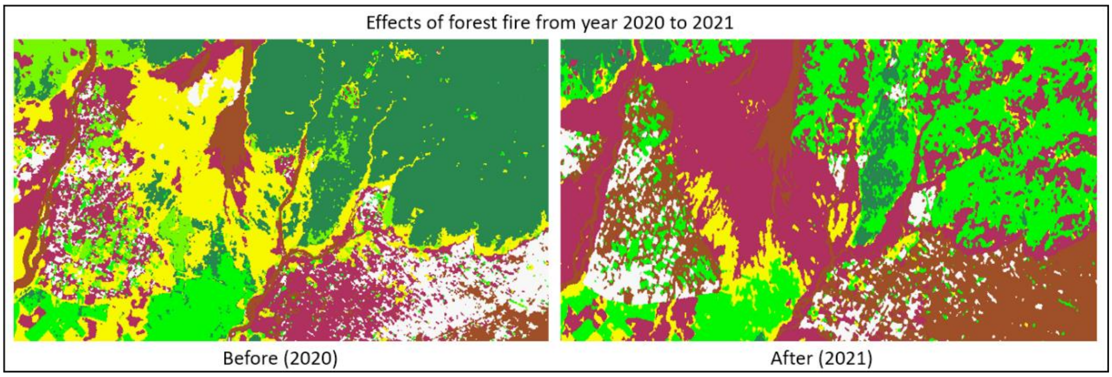

Figure 11. Changes observed during 2020 and 2021

The positive value of image difference represents increase in the number of pixels and negative value means there is decrease in the pixel count. A comparison of the effects of before fire and after fire has been shown in Figure 11. It is evident that, due to forest fires there has been a decline in shrub lands, forest and scarce forest. On the contrary, the barren land has increased as an after effect. The sediments i.e. ashes and debris from the fire has settled into the lower lying areas. The health of the vegetation has also been severely affected as a consequence. Figure 12 shows the percentage of increase or decrease due to increased number of forest fire in State of Uttarakhand in last year, changes observed are:

- Sediment settlement has increase up to 320%.
- Very minor change is observed in agricultural lands with 4% change.
- As severe fires were reported, the barren land coverage has increased up to 90%, on the contrary forest, scarce forest and shrub lands have decreased up to -40%, -20% and -70% respectively.
- Since some of the town areas were also affected by the forest fires, the human settlement as also reduced to -33%.

5. CONCLUSION

From this study, we can say that our proposed post-classification technique yields better accuracy than standalone SVM. The Support Vector Machine (SVM) classification is first done and then to improve the accuracy of the classified image, a post-classification technique called Majority Analysis (MA) is applied. When the multi spectral image is classified using SVM only the accuracy achieved is 89.35% and 88.52% for before and after images respectively. But when during post-classification by applying the MA technique, the unclassified pixels were classified into respective other classes and the boundary of the classes also smoothened out. This has led to higher accuracy rate for before and after images 98.71% and 99.76% respectively. The change detection study showed a drastic increase in the barren land due to the FF and on the contrary, the forest, scarce forest and the shrub land area has decreased. For future scope, this method can be clubbed with other existing machine learning techniques and other natural calamities such as earthquake, volcanic eruption, etc. can be used for identifying the changes.

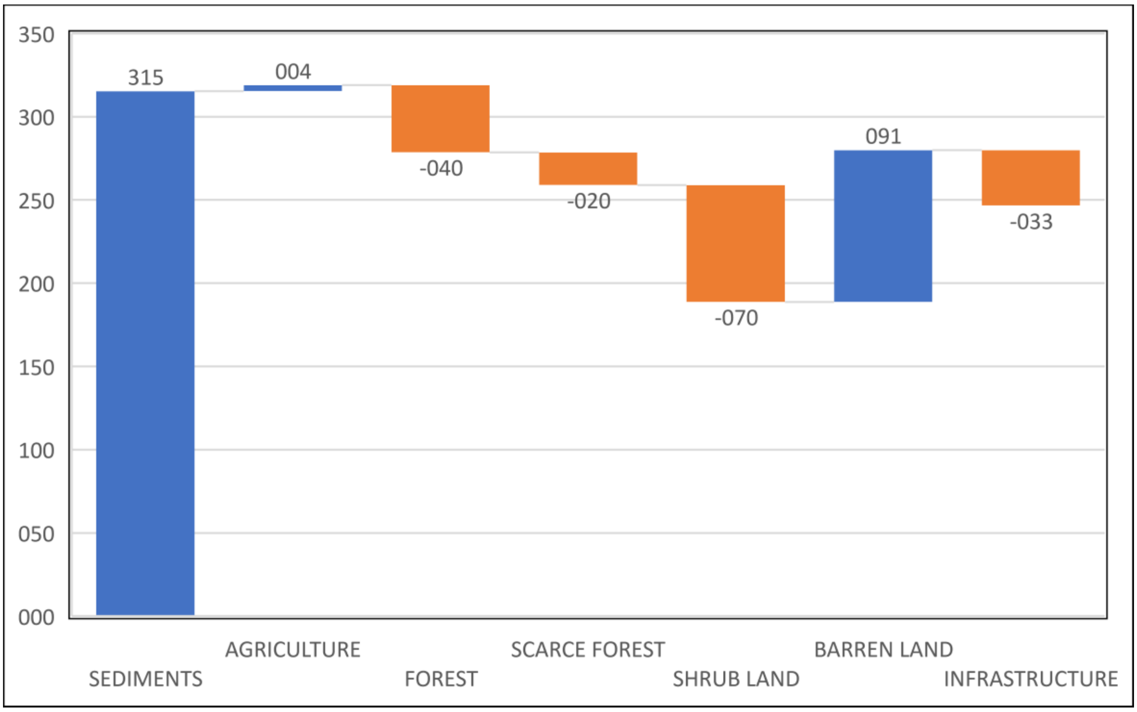

Figure 12. Change detection statistics

REFERENCES

[1] Muchkund Pant and Vinay Purohit, "Forest Fire-A Case Study on the Four National Park of Uttarakhand", International Journal of Engineering Research & Technology (IJERT), vol. 24(12), pp. 347-354, 2019.

[2] Muchkund Pant and Vinay Purohit, "Forest Fire-A Case Study on the Four National Park of Uttarakhand", International Journal of Engineering Research & Technology (IJERT), vol. 24(12), pp. 347-354, 2019.

[3] M. Zanetti and L. Bruzzone, "A generalized statistical model for binary change detection in multispectral images," 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016, pp. 3378-3381, doi: 10.1109/IGARSS.2016.7729873.

[4] Andrew P. Tewkesbury, Alexis J. Comber, Nicholas J. Tate, Alistair Lamb and Peter F. Fisher, "A critical synthesis of remotely sensed optical image change detection techniques", Remote Sensing of Environment, vol. 160, pp. 1-14, 2015, doi: 10.1016/j.rse.2015.01.006.

[5] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo and A. Ghosh, "A Context-Sensitive Technique for Unsupervised Change Detection Based on Hopfield-Type Neural Networks," in IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 3, pp. 778-789, March 2007, doi: 10.1109/TGRS.2006.888861.

[6] Chinmayee Pati, Ashok K. Panda, Ajaya Kumar Tripathy, Sateesh K. Pradhan and Srikanta Patnaik, "A novel hybrid machine learning approach for change detection in remote sensing images", Engineering Science and Technology, an International Journal, vol. 23(5), pp. 973-981, 2020, doi: 10.1016/j.jestch.2020.01.002.

[7] Kamrul Islam, Mohammed Jashimuddin, Biswajit Nath and Tapan Kumar Nath, "Land use classification and change detection by using multi-temporal remotely sensed imagery: The case of Chunati wildlife sanctuary, Bangladesh", The Egyptian Journal of Remote Sensing and Space Science, vol. 21(1), pp. 37-47, 2018, doi: 10.1016/j.ejrs.2016.12.005.

[8] Shivesh Kishore Karan and Sukha Ranjan Samadder, "Accuracy of land use change detection using support vector machine and maximum likelihood techniques for open-cast coal mining areas", Environmental Monitoring and Assessment, vol. 188(8), pp. 486, 2016, doi: 10.1007/s10661-016-5494-x.

[9] Anju Asokan and J. Anitha, "Change detection techniques for remote sensing applications: a survey", Earth Science Informatics, vol. 12(2), pp. 143–160, 2019, doi: 10.1007/s12145-019-00380-5.

[10] Keshtkar, H., Voigt, W. and Alizadeh, E, "Land-cover classification and analysis of change using machine-learning classifiers and multi-temporal remote sensing imagery", Arabian Journal of Geosciences, vol. 10, pp. 154, 2017, doi:10.1007/s12517-017-2899-y.

[11] Somnath Bar, Bikash Ranjan Parida and Arvind Chandra Pandey, "Landsat-8 and Sentinel-2 based Forest fire burn area mapping using machine learning algorithms on GEE cloud platform over Uttarakhand, Western Himalaya", Remote Sensing Applications: Society and Environment, vol. 18, 2020, 100324, doi: 10.1016/j.rsase.2020.100324.

[12] M. Pal and P. M. Mather, "Support vector machines for classification in remote sensing", International Journal of Remote Sensing, vol. 26(5), pp. 1007-1011, 2006, doi: 10.1080/01431160512331314083.

BIOGRAPHIES OF AUTHORS

Ms. Swasti Patel is a Ph.D. scholor at Parul University. She received a gold medal for her outstanding performance in academics during her post graduation. She has collaborated with BISAG-N (Bhaskaracharya National Institute for Space Applications and Geo-informatics) for her Ph.D. research work. Her area of research includes remote sensing, machine learning, image processing, land cover mapping and GIS.

Dr. Priya Swaminarayan is working as a Dean - Faculty of IT & CS; Director - MCA Department, Parul Institute of Engineering & Technology-MCA,; Principal - BCA Department, Parul Institute of Computer Application at Parul University. SHE is having 21 years of teaching experience, 11 years of research & administrative experience. SHE has completed PhD in the field of Semantic Web in 2012. Under her guidance, 5 research scholars have completed their PhD & currently 6 research scholars are pursuing their PhD under her supervision. She has published 50+ Research Papers in reputed and peer-reviewed National/International Journals, presented 35+ research papers in State/National/International conferences, delivered 25+ expert talks in various refresher course/orientation course/workshop conducted by Sardar Patel University, Saurashtra University, Anand Agricultural University, Marwadi group of Institutes etc., attended 40+ workshops/conferences/refresher course/orientation programme. She has published 4 books with Prof. Digvijay, Prof. Mayur Raj, Prof. Unnati Patel.