PROJECT REPORT

ON

A Time Series Analysis of Spatial Data for Examining the Land Transformation Bangalore

For the Partial Completion of

Course AI-703- Geographical Information Systems 2020

Under the guidance of Prof. Uttam Kumar

ज्ञानमुत्तमम्

Group Members
Nimisha Karnatak (DT2019006)
Kaushal Mittal (IMT2017024)
Ravi Jain (MT2019087)
Mohith Sarma (IMT2017512)

Link to all the resources, report and sheet
https://drive.google.com/drive/folders/1F75h62XfzOFxPSdrShkIqdWVcUQqD2I7?usp=sharing

INDEX

INTRODUCTION:

Urbanisation in India is growing at unprecedented rates. Percentage population residing in urban areas was 28.53% in the 2001 census, and is now 34% in 2017 according to The World Bank. This has led to significant changes in and around urban areas with respect to land cover and usage patterns. This report presents the changes in land use and land cover pattern in the city of Bangalore for the year 2003, 2010, 2015 and 2020 using Multitemporal Landsat Dataset with respect to multiple algorithms for classification.

(Keywords: GIS, Urbanisation, Temporal Patterns, Bangalore, Time Series Analysis)

LITERATURE RESEARCH

Fazal et al., in 2000, used Remote Sensing along with field checks and surveys to analyse the extent of urbanisation in the Saharanpur City at the expense of agricultural land. The study uses the data of 10 years, from 1988 and 1998.This study of land-transformation because of rapid urbanisation uses the data set obtained from IRS-1C geo-coded panchromatic satellite analogue imagery along with field checks and surveys.
Another study done by Pandey, & Seto,2015, investigates the loss of agriculture land using the econometric time series analysis, because of urbanisation, a phenomenon commonly known as “urban conversion of agricultural lands” from June 2000 to May 2011 across India (Pandey & Seto,2015). The dataset used in the study is the MODIS MOD13Q1 time series dataset (Pandey & Seto,2015).
A similar study was done in Ranchi by Kumar et al, in 2011 to understand the urban expansion in Ranchi. The dataset used includes Landsat MSS for 1975,Landsat TM for 1986 & 1996 , IRS P6-LISS IV for 2005 and USGS topographical sheet of 1927.

Studies done by Rawat & Kumar (2015), Rahman et. al, 2012 and Dewan et.al, 2008 uses Maximum Likelihood classifier to explore the land use and land cover patterns of Hawalbagh block, district Almora,North-West District of Delhi and Dhaka Metropolitan of Bangladesh.

In our project, we have used Supervised Maximum Likelihood classification, but along with this algorithm, we have used 4 other algorithms, in order to find the most efficient one.

The following table enlist the summary of the few papers we studied for the project.

Paper Name	Tools and Dataset	Methodology	Common Grounds

1. Land-cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing	1.multitemporal Landsat Thematic Mapper (TM) data 2.1986, 1991, 1998, and 2002. 3. Image processing: ERDAS Imagine, version 8.5	-Guided clustering - Post classification refinements	1.Change detection model in ERDAS Imagine
2.A Multi-Temporal Landsat Data Analysis for Land-use/Land-cover Change in **Haridwar** Region using Remote Sensing Techniques	multispectral Landsat-8 OLI, Landsat-7 ETM+, and Landsat-5 TM	1. **USGS Earth Explorer (1996, 2003, 2010 & 2017)** 2. Classification: **orchards, vegetation, agricultural land, rangeland, urban land, water bodies, and watershed.** **3. pixel-based Supervised classification method**	1.Pixel based supervised classification
3.Using remote sensing and GIS to detect and monitor land use and land cover change in Dhaka Metropolitan of Bangladesh during 1960–2005	-topographic maps -multi-temporal remotely sensed data (1960 -2005)	**1.Maximum likelihood supervised classification techniqu**e **2.post-classification change detection method** 3. Classification:Bare soil, built-up.cultivated land, vegetation, waterbodies, wetland/lowlands	1. multi-temporal remotely sensed data 2. Maximum likelihood supervised classification technique
4. Assessment of Land use/land cover Change in the North-West District of Delhi Using Remote Sensing and GIS Techniques	-Aster image of 2003, spatial resolution of 15 m - Survey of India (SOI),1972 toposheet,scale of 1:50,000.	1.Supervised digital classification (maximum likelihood classifier) 2.Change detection model in ERDAS Imagine 3. Classification: Urban, Rural, Industrial,Forest, River, Pond/Lake, Agriculture,Wasteland	**1.Supervised digital classification using maximum likelihood classifie**r 2.Change detection model in ERDAS Imagine
5.Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district **Almora, Uttarakhand, India**	**Landsat Thematic Mapper (TM) of 1990 and 2010**	**1. Global Land Cover Facility Site (GLCF) and earth explorer site** **2.Supervised classification methodology** : **maximum likelihood technique** in ERDAS 9.3 Software. 3.Classification 1) vegetation, 2)agriculture,	1. Landsat satellite imageries 2. **Maximum Likelihood supervised classification technique**

		3)barren 4)built-up and 5)water body	

OBJECTIVE OF PROBLEM

To analyse the patterns of land transformation in Bangalore in the year 2003, 2010, 2015 and 2020 and compare different classification algorithms like maximum likelihood classification, random forest, decision trees and support vector machine. The land usage pattern includes :

1. Water
2. Open Area
3. Urban Area
4. Vegetation
5. Agricultural lands

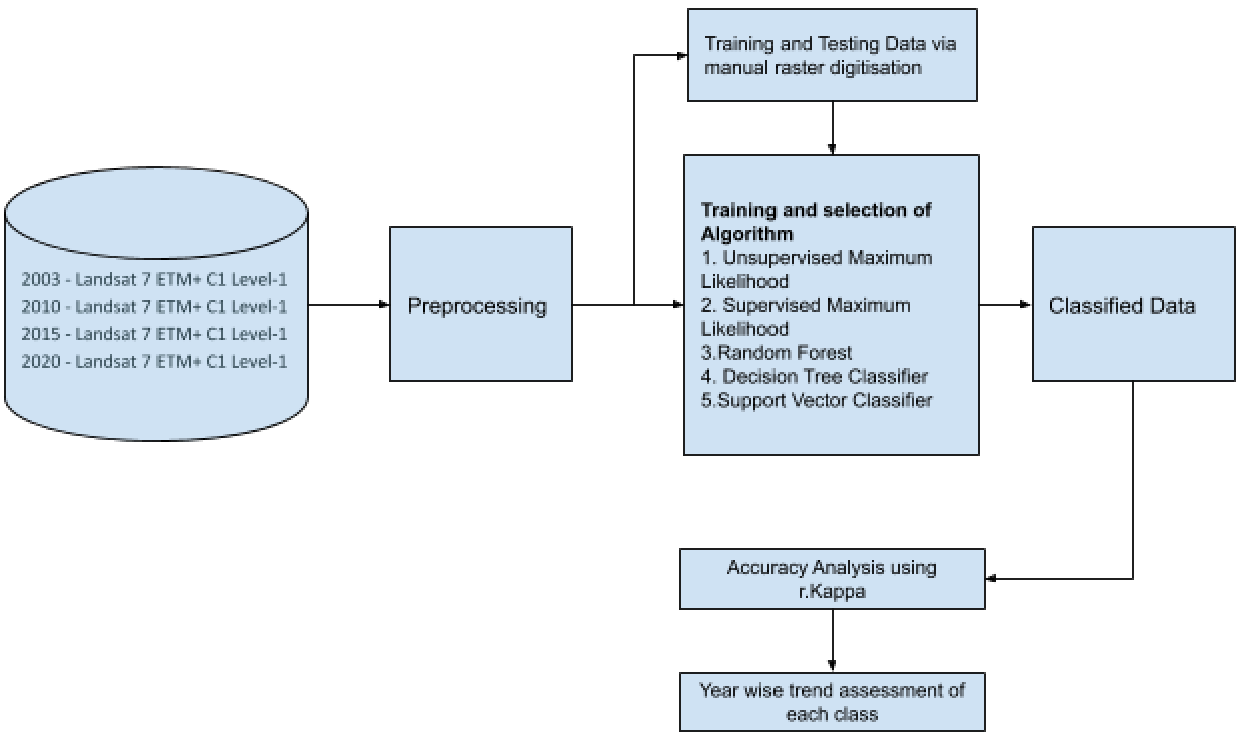
PROBLEM DEFINITION

To analyse the land transformation of Bangalore from 2003, 2010, 2015 and 2020 using the Multitemporal Landsat data and the most efficient algorithm to get accurate results for the classification of land use/cover. We used 5 algorithms to find the most accurate one, in order to understand the trends of land transformation in Bangalore.

1. Unsupervised Maximum Likelihood
2.Supervised Maximum Likelihood
3.Random Forest
4.Decision Tree Classifier
5. Support Vector Classifier

METHODOLOGY

In this section, the methodology we followed is represented in the form of the below flowchart and discussed as bullet points.



Step 1 - Obtaining Dataset

We downloaded the dataset for Bengaluru, Karnataka, India (Latitude - 12.9716, Longitude - 77.5946) from https://earthexplorer.usgs.gov. Multitemporal Landsat data is considered for following years as follows -

- 2003 - Landsat 7 ETM+ C1 Level-1
- 2010 - Landsat 7 ETM+ C1 Level-1
- 2015 - Landsat 7 ETM+ C1 Level-1
- 2020 - Landsat 7 ETM+ C1 Level-1

The raw dataset has a resolution of 30m pixel, and consists of 7 bands with specific wavelengths (in nm)450-520nm (B1-Blue), 520-600nm (B2-Green), 630nm-690nm (B3-Red), 770nm-900nm (B4-NIR), 155nm-175nm (B5-SWIR), 1040nm-1250nm (B6-TIR) and 209nm-235nm (B7-MIR).

Step 2 - Preprocessing Dataset

For the preprocessing step we use the administrative shape of bangalore urban from https://diva-gis.org/gdata to crop the obtained dataset as per our region of interest using the shape file. Scan Line Corrector in landsat-7 (SLC) failed in 2003 but the data coming from the sensors is still usable. The data from satellites data will include a Gap mask for all the bands. We use ***R.fill.gaps*** to fill the gaps with the interpolated values. This tool by default will use spatially weighted mean to fill the 'no data' cells in the data.

Step 3 - Preparing for Classification

We consider different sets of bands to digitize the image and create training and test data. Combinations we use are

1. B3 - Red, B2 - Green, B1 - Blue to create natural colour RGB,
2. B4 - NearIR, B3 - Red, B2 - Green used to create false colour composite to identify Vegetation, Agriculture
3. B4 - NearIR, B5 - Short Wave IR, B3 - Red used to create false colour composite to identify Water Bodies

After the train and test data has been created, we create signature files for each of the classification algorithms.

Step 4 - Classification and post classification

Next we run the corresponding classification algorithm with each of the signature files. Post classification we reclassify the image into 5 expected classes. Next step is to assign corresponding colours to each class. We then display the classified map with legend and scale to understand/ interpret the classified image.

Step 5 - Statistics, Accuracy and Inferences

We then calculate the statistics for each of the class with respect to area in hectares and percentage cover of the map. Then we use Sheets to tabulate the data and construct graphs for visual representation of the interpretations. Also using the test maps we calculate the overall accuracy, kappa coefficient and omission and commission scores.

Step 6 - Year-wise trend analysis of each classification

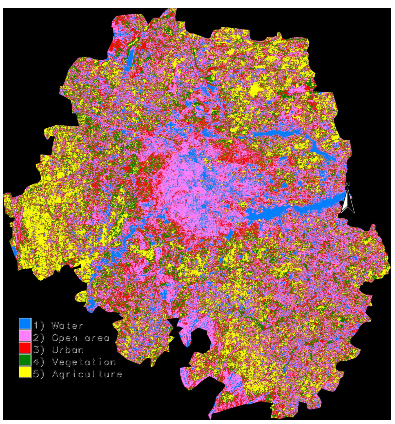
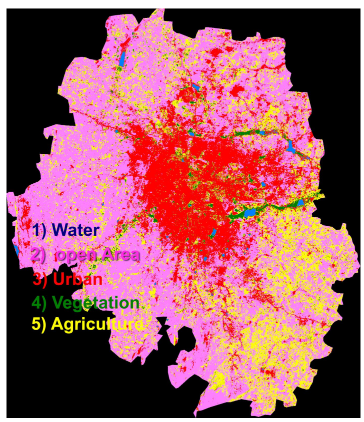
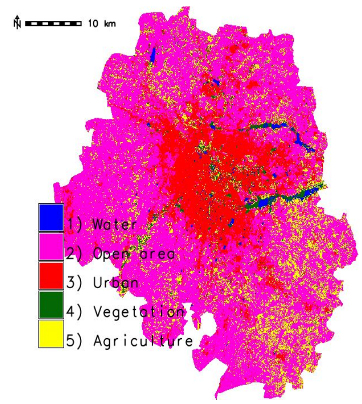
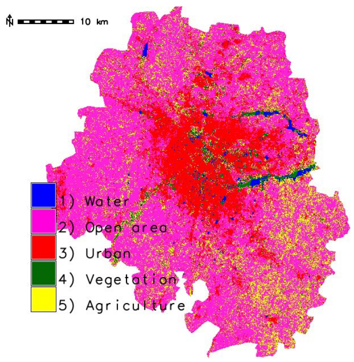
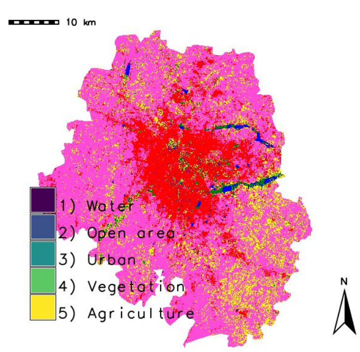
Using the most accurate algorithms : Random Forest and SVM, we then analysed the year-wise trend of each classification. Thus completing the Time Series Analysis of Spatial Data to understand the Land Transformation of Bangalore.

Results:

This section presents the result of the 5 classification algorithms in the form of images for the year 2003, 2010, 2015 and 2020 for Bangalore city (Latitude - 12.9716, Longitude - 77.5946) . The 5 classification algorithms used are enlisted below:

1. Unsupervised Maximum Likelihood
2. Supervised Maximum Likelihood
3. Random Forest
4. Decision Tree Classifier
5. Support Vector Classifier

Year 2003

<table>
<tr>
<td>Image 1 Unsupervised classification
</td>
<td>Image 2 supervised maximum likelihood
</td>
<td>Image 3 random forest
</td>
</tr>
<tr>
<td>Image 4 decision tree classifier
</td>
<td>Image 5 support vector classification
</td>
<td></td>
</tr>
</table>

Year 2010

Image-1 Unsupervised classification

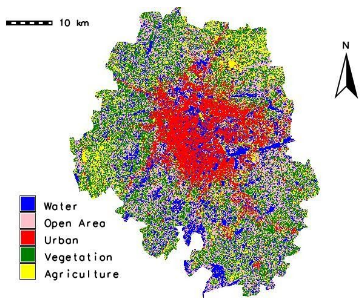


Image-2 supervised maximum likelihood

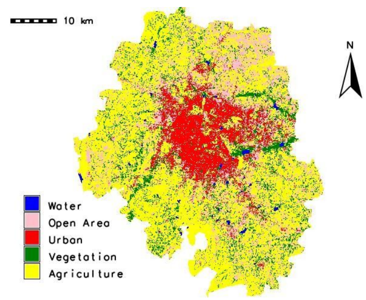


Image-3 random forest

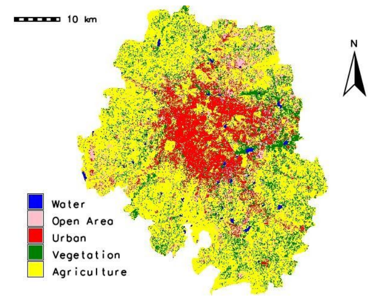


Image-4 decision tree classifier

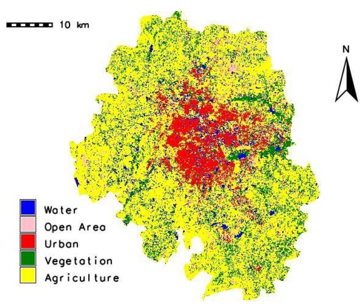
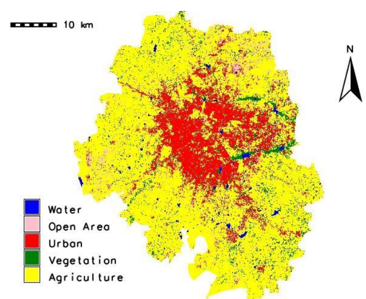


Image-5 support vector classification



Year 2015

Image-1 Unsupervised classification

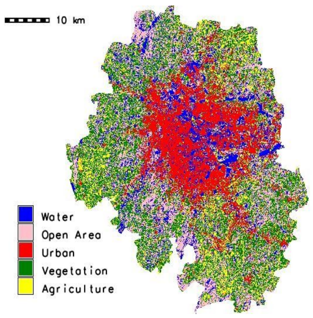


Image-2 supervised maximum likelihood

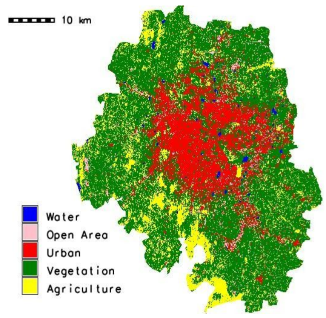


Image-3 random forest

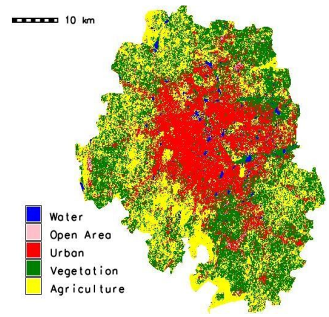


Image-4 decision tree classifier

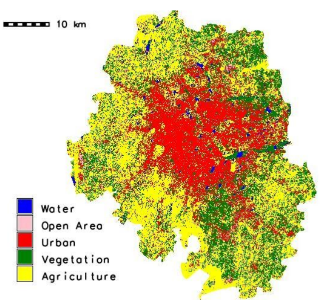
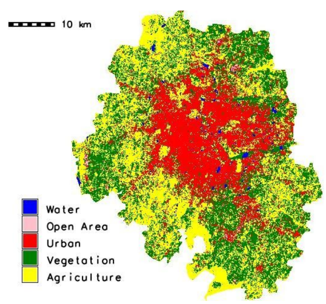


Image-5 support vector classification



Year 2020

Image 1
Unsupervised classification

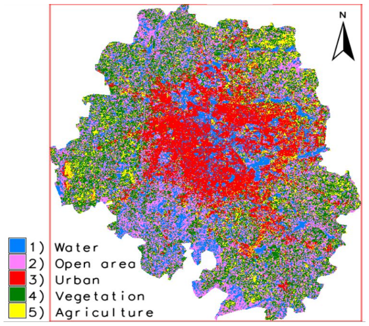


Image 2 supervised maximum likelihood

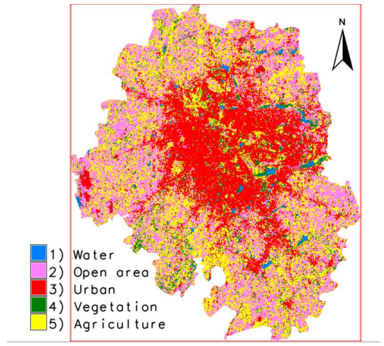


Image 3
Random forest

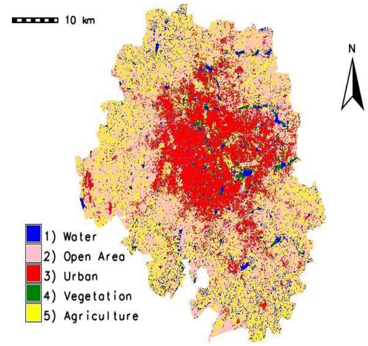


Image 4
Decision tree classifier

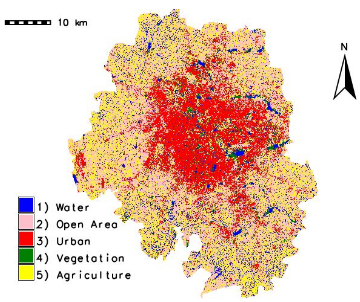
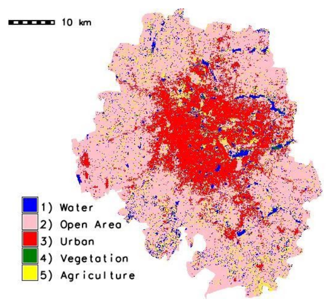


Image 5 support vector classification



Accuracy Assessment

1. Overall Accuracy:

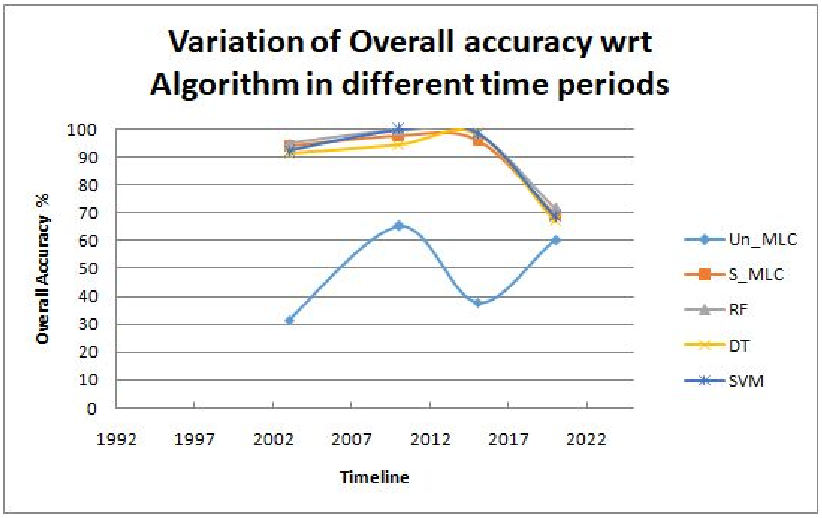
We used r.kappa analysis in GRASS GIS to assess the accuracy of the classified images. The table below represents the overall accuracy of the classified images:

Year	Un_MLC	S_MLC	RF	DT	SVM
2003	31.599	93.689	94.87	91.16	92.3
2010	65.2	97.52	99.68	94.32	99.76
2015	37.88	95.79	98.341	98.341	98.56
2020	60.27	69.08	71.66	66.79	68.46
Avg	45.3195	84.68	89.10	86.372	89.77

Here Un_MLC = Unsupervised Maximum Likelihood
S_MLC = Supervised Maximum Likelihood
RF= Random Forest
DT= Decision Tree
SVM = Support Vector Machine

As evident from the above table, Random Forest and SVM gives the maximum accuracy across years, followed by Decision Tree and Supervised Maximum Likelihood. The unsupervised maximum likelihood gives the least accuracy, as no training data is available that can be assumed as ground truth because the algorithm randomly picks 5 clusters for classification.

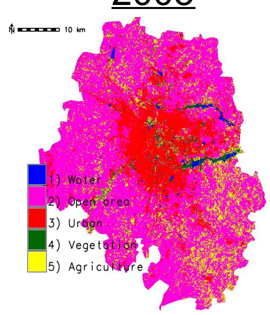
Another important trend to note is that the overall accuracy of Supervised Maximum Likelihood,Random Forest, Decision Tree and Support Vector Machine is above 90 percent for the year 2003, 2010 and 2015. However, for 2020, it drops drastically to around 70 percent. The graph represents the variation of overall accuracy with respect to different algorithms (Unsupervised Maximum Likelihood,Supervised Maximum Likelihood,Random Forest, Decision Tree and Support Vector Machine) across the years 2003, 2010, 2015 and 2020.



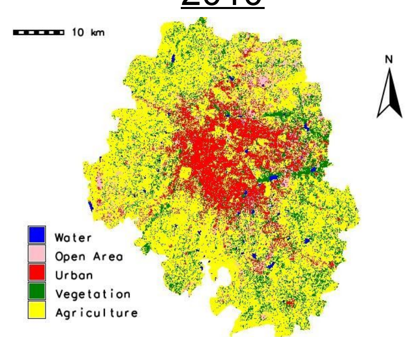
Hence, from the accuracy analysis we conclude that the two most accurate algorithms are:
1. Random Forest
2. Support Machine Vector
The image presents the classification of Bangalore city for the year 2003, 2010, 2015 and 2020, done using Random Forest.

Random Forest Classification

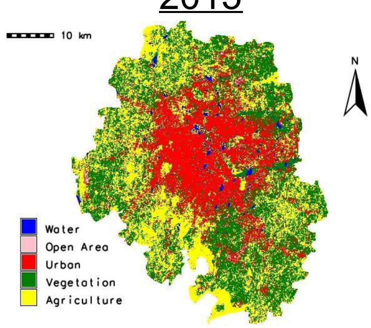
2003



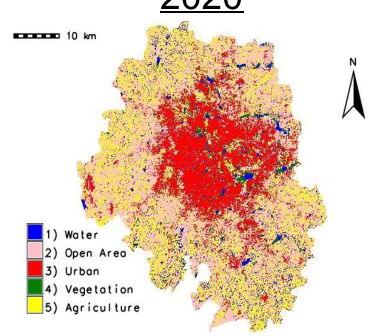
2010



2015



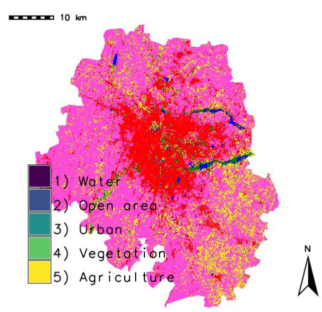
2020



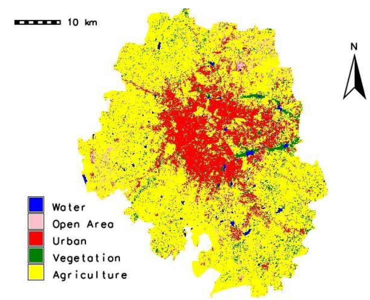
The image presents the classification of Bangalore city for the year 2003, 2010, 2015 and 2020, done using Support Machine Vector Algorithm.

Support Vector Machine Classification

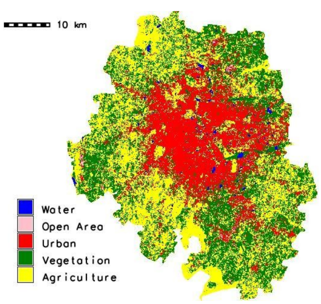
2003



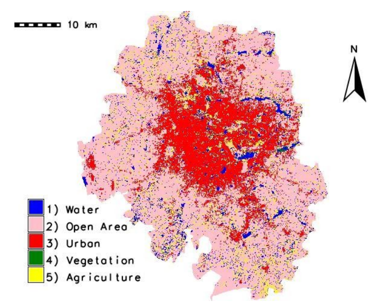
2010



2015



2020



The image presents the classification of Bangalore city for the year 2003, 2010, 2015 and 2020, done using Support Machine Vector Algorithm.

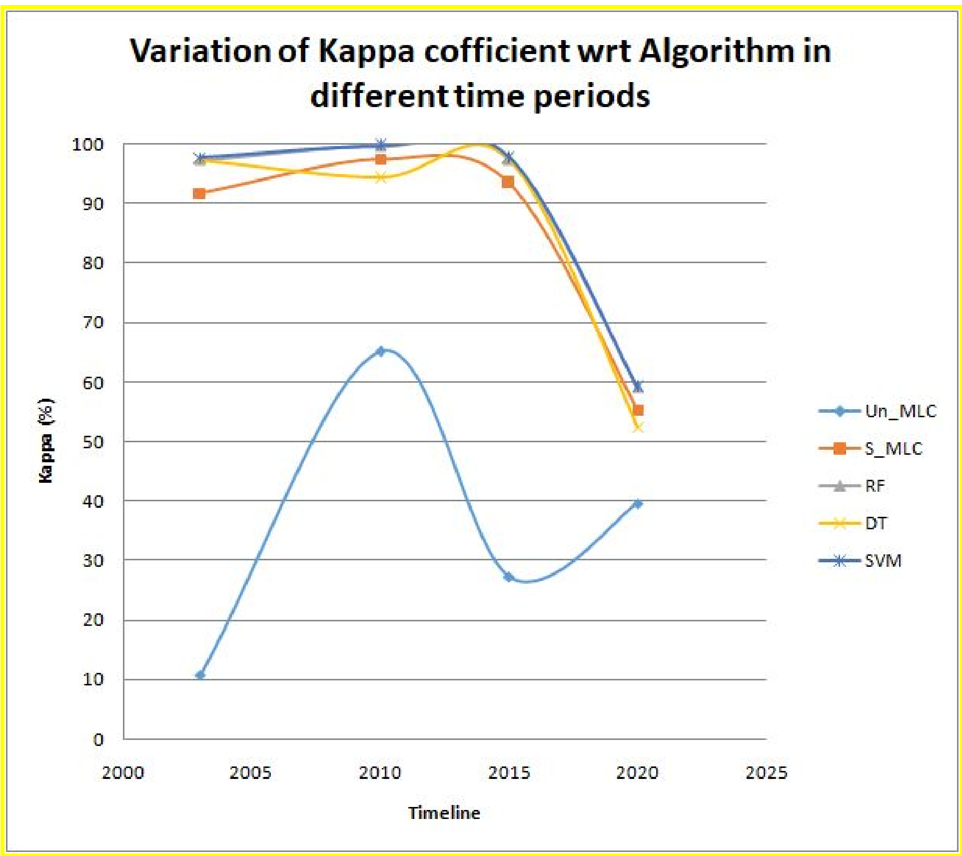
2. Kappa coefficient

The range of the Kappa coefficient varies between 0 and 1.
If Kappa coefficient = 0, no agreement between classified images and reference image
If Kappa coefficient = 1, classified images and the reference image are identical
The higher the Kappa coefficient, the more is the accuracy.
The table below gives the list of all the Kappa coefficients (in percentage) across the year 2003, 2010, 2015, 2020, for all 5 classifications.

Year	Un_MLC	S_MLC	RF	DT	SVM
2003	10.834	91.747	97.38	97.38	97.776
2010	65.2	97.52	99.68	94.32	99.76
2015	27.28	93.48	97.38	97.38	97.73
2020	39.65	55.31	59.11	52.4661	59.13

The algorithms are arranged in descending order as per their Kappa coefficient:

Support Vector Machine> Random Forest> Decision Tree> Supervised Maximum Likelihood > Unsupervised Maximum Likelihood



3. Error of Commission and Error of Omission :

Error of omission refers to the percentage of the pixels which are not classified into the respective class (and classified into the wrong class).

Another parameter associated with Error of Omission is Producer's Accuracy.
Producer's accuracy = 100%-Omission Error

Whereas, error of commission refers to the percentage of pixels which are misclassified into the corresponding class (but belong to a different class).

Another parameter associated with Error of Commission is User's Accuracy.
User's accuracy = 100%-Commission Error

In this section we present the error of omission, error of commission, producer's accuracy and user's accuracy for the year 2003 for the 2 most accurate algorithms, namely: Random Forest and Support Vector Machine algorithm. All the other accuracies have been tabulated in the sheet linked.

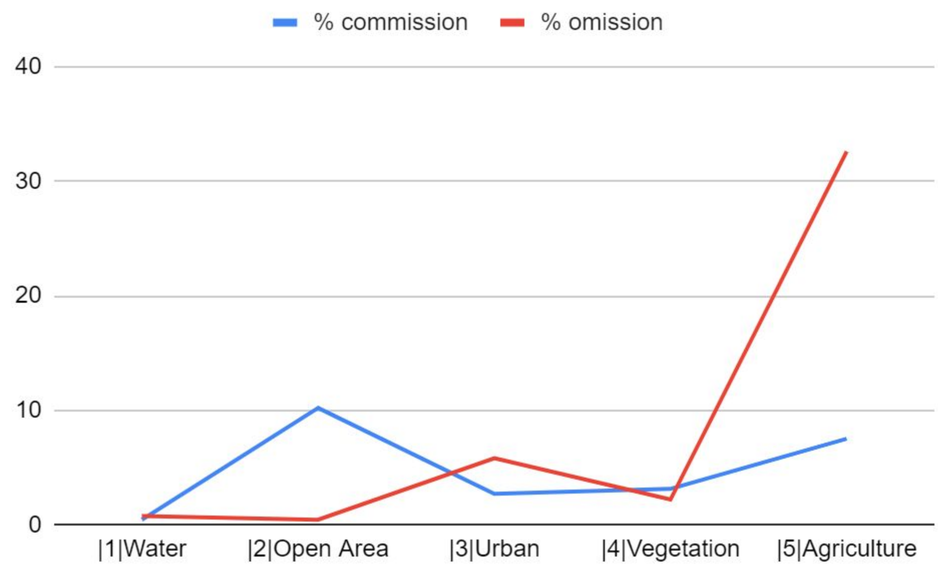
2003 - Random Forest

	% commission	% omission	Producer's accuracy	User's accuracy
|1|Water	0.44843	0.782123	99.217877	99.55157
|2|Open Area	10.21062	0.456853	99.543147	89.78938
|3|Urban	2.724177	5.824176	94.175824	97.275823
|4|Vegetation	3.147954	2.224576	97.775424	96.852046
|5|Agriculture	7.524752	32.61183	67.38817	92.475248

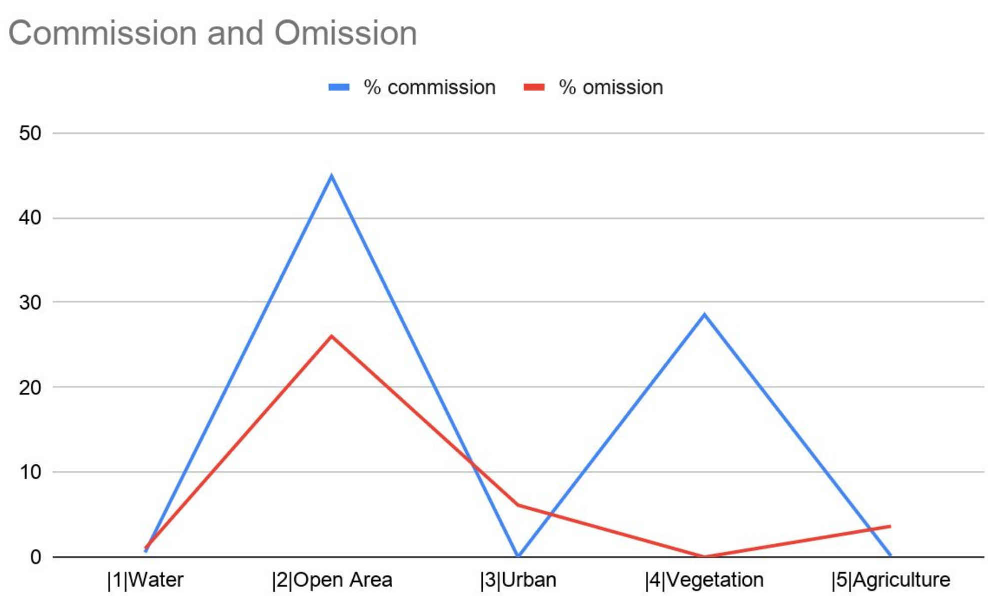
2003 - Support Vector Machine

	% commission	% omission	Producer's accuracy	User's accuracy
|1|Water	0.44843	0.782123	99.217877	99.55157
|2|Open Area	10.21062	0.456853	99.543147	89.78938
|3|Urban	2.724177	5.824176	94.175824	97.275823
|4|Vegetation	3.147954	2.224576	97.775424	96.852046
|5|Agriculture	7.524752	32.61183	67.38817	92.475248

2003 - Random Forest



2003 - Support Vector Machine



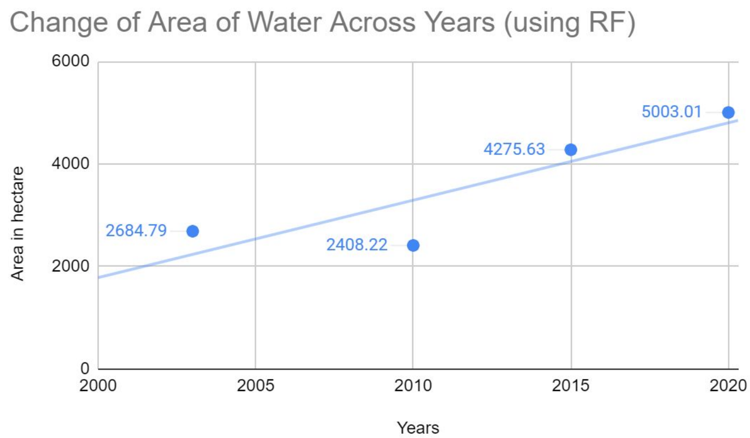
4. Year Wise classification of each class:

From the accuracy tests, we found that the Random Forest and Support Vector Machine algorithm have the highest overall accuracy across all the years.
Using these two algorithms, we found the year wise change in the area of 5 classes, namely :Water, Open Land, Urban, Vegetation, Agriculture.

4.1 Year-wise classification of each class using Random Forest Algorithm:

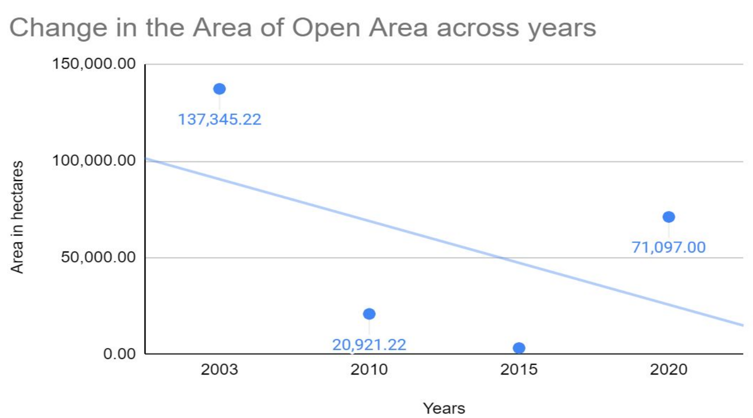
1. Change in area of water (hectare) over the year 2003, 2010, 2015, 2020

2003	2010	2015	2020
2684.79	2408.22	4275.63	5,003.01



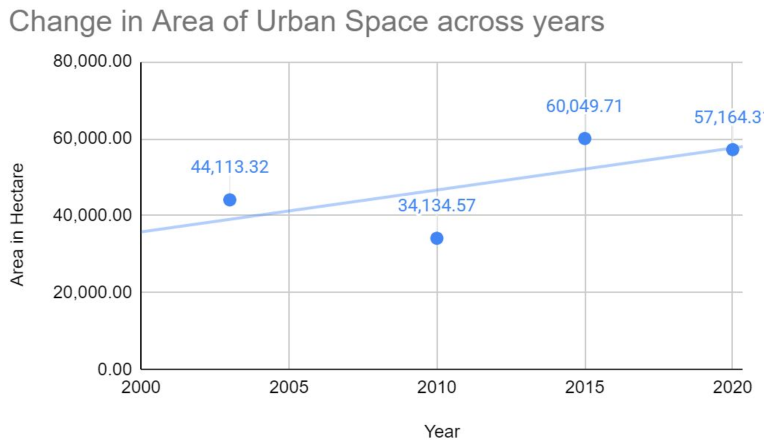
2. Change in area of open area (hectare) over the year 2003, 2010, 2015, 2020

2003	2010	2015	2020
137,345.22	20,921.22	3265.2	71,097.00



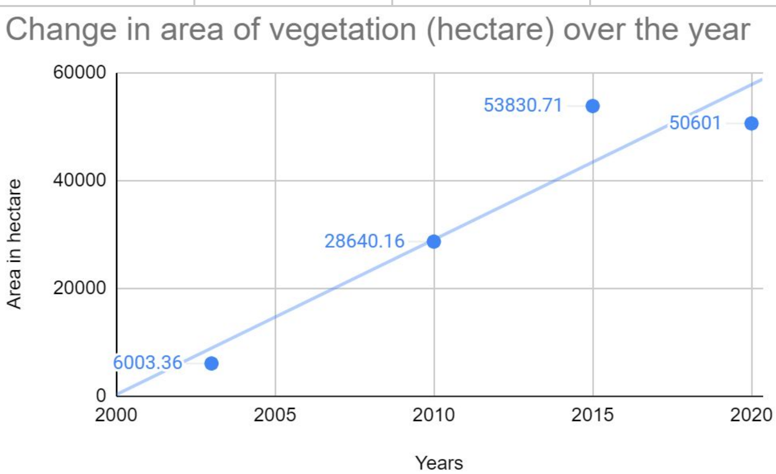
3. Change in area of urban space (hectare) over the year 2003, 2010, 2015, 2020

2003	2010	2015	2020
44,113.32	34,134.57	60,049.71	57,164.31



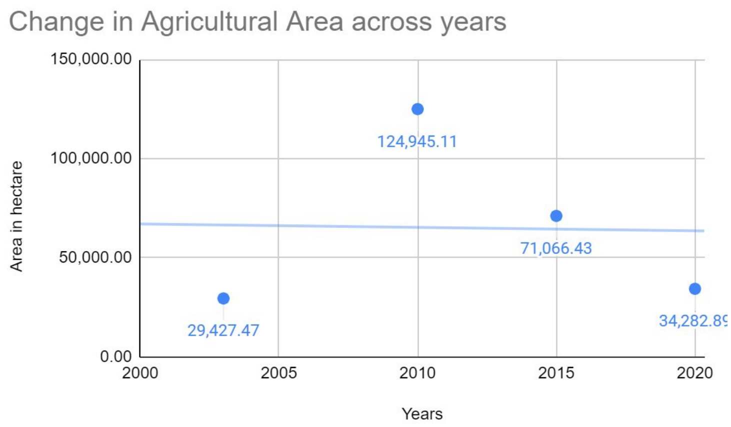
4. Change in area of vegetation (hectare) over the year 2003, 2010, 2015, 2020

2003	2010	2015	2020
6003.36	28,640.16	53,830.71	50601



5. Change in area of Agriculture (hectare) over the year 2003, 2010, 2015, 2020

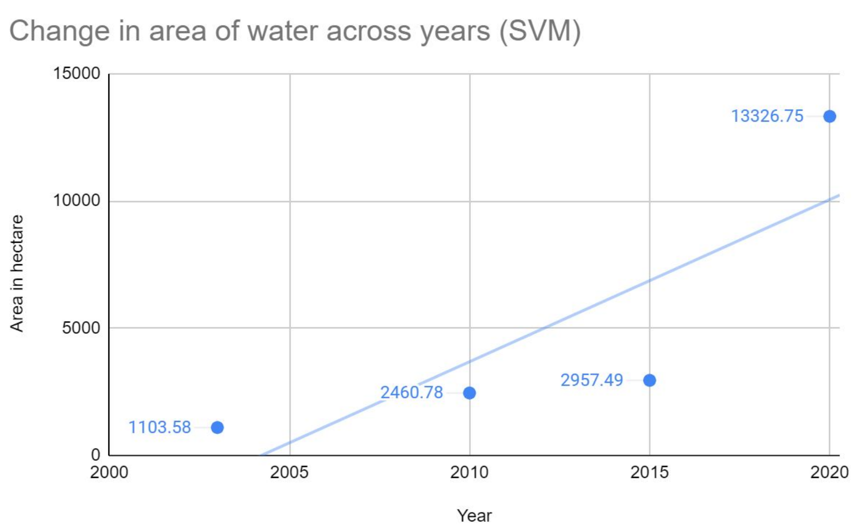
2003	2010	2015	2020
29,427.47	124,945.11	71,066.43	34,282.89



4.2 Year-wise classification of each class using SVM Algorithm:

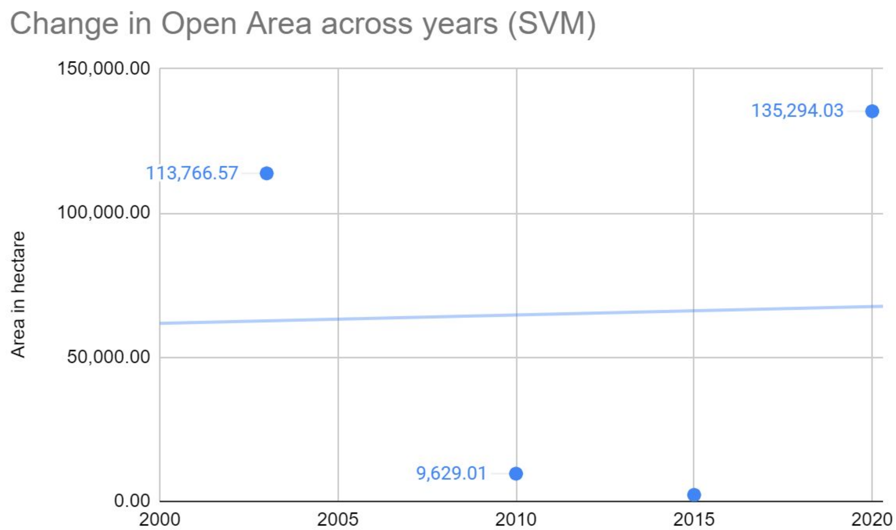
1. Change in area of water (hectare) over the year 2003, 2010, 2015, 2020

2003	2010	2015	2020
1103.58	2460.78	2957.49	13,326.75



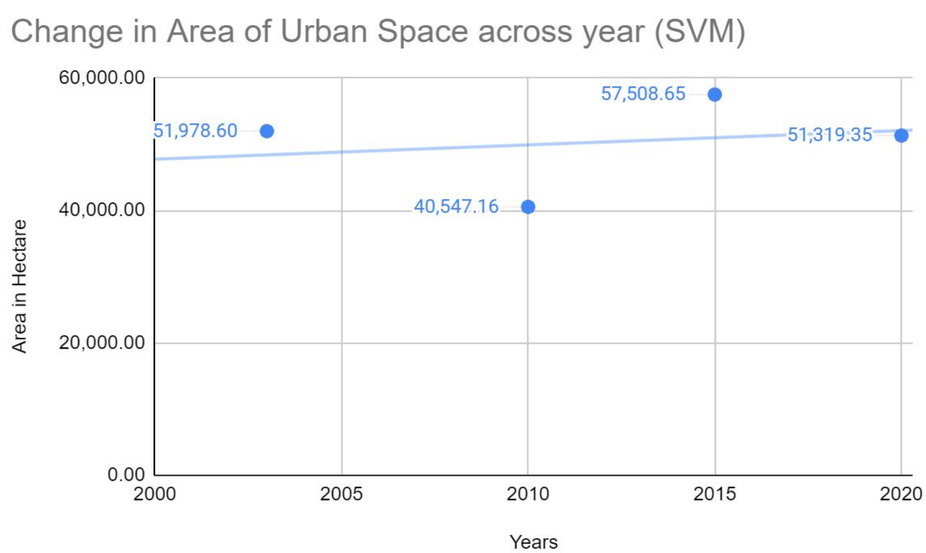
2. Change in open area (hectare) over the year 2003, 2010, 2015, 2020

2003	2010	2015	2020
113,766.57	9629.01	2276.46	135,294.03



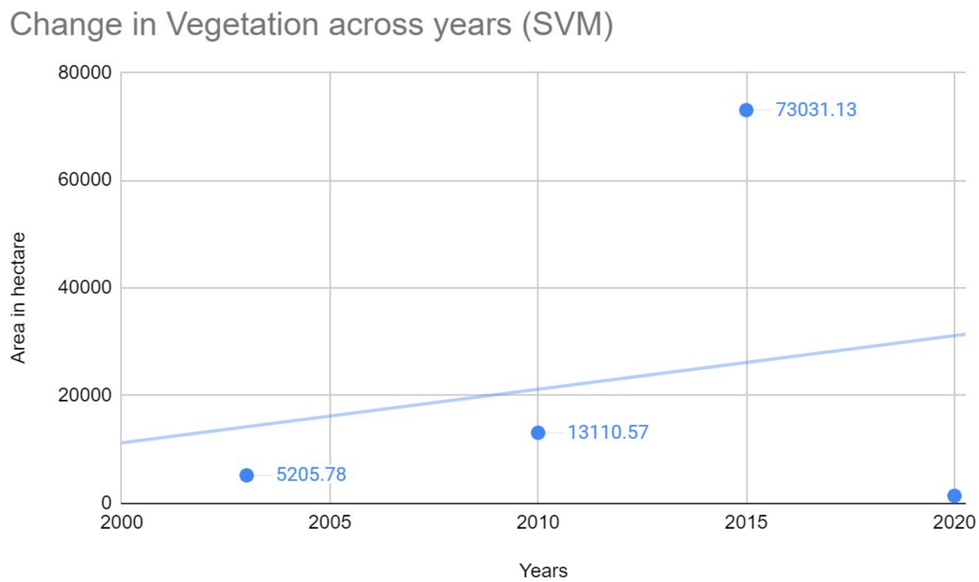
3. <u>Change in area of urban space (hectare) over the year 2003, 2010, 2015, 2020</u>

2003	2010	2015	2020
51,978.60	40,547.16	57,508.65	51,319.35



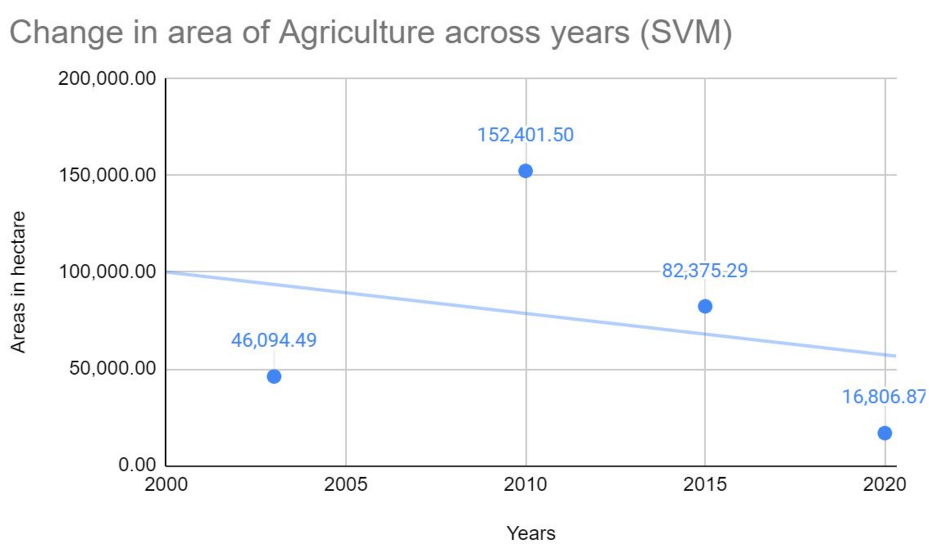
4. <u>Change in area of vegetation (hectare) over the year 2003, 2010, 2015, 2020</u>

2003	2010	2015	2020
5205.78	13,110.57	73,031.13	1402.02



5. Change in area of agricultural area (hectare) over the year 2003, 2010, 2015, 2020

2003	2010	2015	2020
46,094.49	152,401.50	82,375.29	16,806.87



Discussion:

From the accuracy assessment, we found that Random Forest and SVM gives the maximum accuracy for all classes across the years.
This section discusses the patterns that will help us analyze the change in the area of class, from 2003 t0 2020.

1. Trends from Random Forest

a) There is an increase in water area (hectare) from
2003 - 2684.79
2010 - 2408.22
2015 - 4275.63
2020 - 5,003.01

b) The open area (hectare) is decreasing from 2003 to 2020
2003 - 137,345.2
2010 - 20,921.22
2015 - 3265.2
2020 - 71,097.00

As seen from graph 2, in section 4.1, there is a decreasing trend in the open area,i.e. the open area is decreasing from 2003 to 2015, however there is an anomaly with the data of 2020.

c) There is an increasing trend in the area of urban space too, as represented in graph 3 of section 4.1.

d) As per the random forest classification algorithm, there is an increase in the area of vegetation from 2003 to 2020.

e) The agricultural land witnessed a decreasing trend from 2003 to 2020, as represented in graph 5 of section 4.1

2003 - 29427.47
2010 - 124,945.1
2015 - 71,066.4
2020 - 34,282.8

2. Trends from Support Vector Machine :

a) The area of water increases across the year, but not very drastically.
2003 - 1103.58
2010 - 2460.78
2015 - 2957.49
2020 - 13, 326.75
Here, we ignore the anomaly : the 2020 data, before understanding the trends.
b) There is a decreasing trend in open areas, ignoring the anomaly data of 2020.

c) Three is an increasing trend in the area of urban space, ignoring the anomaly data of 2020.
d) The area under vegetation is found to be increasing, ignoring the anomaly data of 2020.
e) There is a decreasing trend in the agricultural land.

In conclusion, we saw that the trend we found using the two most accurate algorithms : Random Forest and SVM remains the same. Across years (2003-2020), we noticed these trends :

1. Increase in area of water
2. Decrease in open space
3. Increase in urban space
4. Increase in vegetation area
5. Decrease in agricultural area

Limitations :

1. Since we used manual raster digitization to obtain test data and training data, there is a good possibility that the accuracy of the training data sets and testing data sets might be influenced because of the human-induced errors.
2. Limitation in the number of clusters taken per classification while creating test and training datasets: We used 5 to 8 clusters (samples) for each class while creating training and test datasets. In hindsight, we now understand that the accuracy would have been better if we had taken 15 to 20 clusters per class.
3. We performed the classification of the year 2020 twice for all algorithms and made its test data and training data again. Even after that, we were not able to reach a good accuracy level. The below table shows the overall accuracy calculated using all algorithms:

Year: 2020	Unsupervised MLC	Supervised MLC	Random Forest	Decision Tree	Support Vector Machine
Earlier accuracy	46.599	51.76	63.548	61.67	68.46
New accuracy	60.27	69.087838	71.669884	66.795367	68.463612

From the above table we see that even if the ‘new accuracy’ for the year 2020 is higher than the previous accuracy, it is not as high as 90 percent, as seen in the case of other years with algorithms like random forest and support vector machine.

REFERENCES :

1. Fazal, S. (2000). Urban expansion and loss of agricultural land-a GIS based study of Saharanpur City, India. Environment and Urbanization, 12(2), 133-149
2. Pandey, B., & Seto, K. C. (2015). Urbanization and agricultural land loss in India: Comparing satellite estimates with census data. Journal of environmental management,

148, 53-66.

3. Kumar, A., Pandey, A. C., Hoda, N., & Jeyaseelan, A. T. (2011). Evaluating the long-term urban expansion of Ranchi urban agglomeration, India using geospatial technology. Journal of the Indian Society of Remote Sensing, 39(2), 213-224.
4. Yuan, F., Sawaya, K., Loeffelholz, B., & Bauer, M. (2005). Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. Remote Sensing Of Environment, 98(2-3), 317-328. doi: 10.1016/j.rse.2005.08.006
5. Kumar, S., Shwetank, & Jain, K. (2020). A Multi-Temporal Landsat Data Analysis for Land-use/Land-cover Change in Haridwar Region using Remote Sensing Techniques. Procedia Computer Science, 171, 1184-1193. doi: 10.1016/j.procs.2020.04.127
6. Dewan, A., & Yamaguchi, Y. (2008). Using remote sensing and GIS to detect and monitor land use and land cover change in Dhaka Metropolitan of Bangladesh during 1960–2005. Environmental Monitoring And Assessment, 150(1-4), 237-249. doi: 10.1007/s10661-008-0226-5
7. Rahman, A., Kumar, S., Fazal, S., & Siddiqui, M. (2012). Assessment of Land use/land cover Change in the North-West District of Delhi Using Remote Sensing and GIS Techniques. Journal Of The Indian Society Of Remote Sensing, 40(4), 689-697. doi: 10.1007/s12524-011-0165-4
8. Rawat, J., & Kumar, M. (2015). Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. The Egyptian Journal Of Remote Sensing And Space Science, 18(1), 77-84. doi: 10.1016/j.ejrs.2015.02.002