Classifying Urban Areas Over Time: A Case Study of Surat, India

Ana Caklovic, Marta Kozłowska, and Jirush John Suja Prema Rajini

INTRODUCTION:

- The rapid expansion of urban areas has become an increasingly important issue in today's world.
- The UN has predicted that by 2050, 68% of the world's population will live in urban areas, and combined with population growth over the years, it could result in an increase of 2.5 billion people in urban areas.
- Classification of urban areas has important ramifications for many fields of study, from urban planning to landscape development.
- Land-use Land-Cover classification (LULC) is an important part of classifying urban areas, but to increase its efficiency and accuracy, the use of machine learning models has become a popular tool.

STUDY AREA:

This study will focus on the area of Surat, and the neighboring areas in the north. The city is in Western India (21°10' Latitude, 72°49' Longitude) and its size is 474,2 km². This area was chosen has one of the highest rate of development in the world. The population of Surat went from 5,670,924 in 2015 to 7,784,276 in 2022. In addition, it's near the river so the water and sand change, affecting the landscape and spread of urban areas.

OBJECTIVE:

- The main goal of this study is to analyze changes in urban areas of the fast-developing city of Surat, India, based on classification of urban areas with the Random Forest model and then post-classification change detection across three time steps (2015, 2108, 2022).
- Amongst the machine learning, the Random Forest model has proven to have a high accuracy when it comes to land cover classification. Advantages of Random Forest include the ability to examine feature contribution and the fact that it is fast and powerful, and has been used before for the remote sensing of urban areas.

METHODS

Data – Landsat 8, 7 bands

4 Landcover classes – water, built-up, vegetation, sand

AOI – vectorization in QGIS

Calculating NDVI, NDBI, NDWI

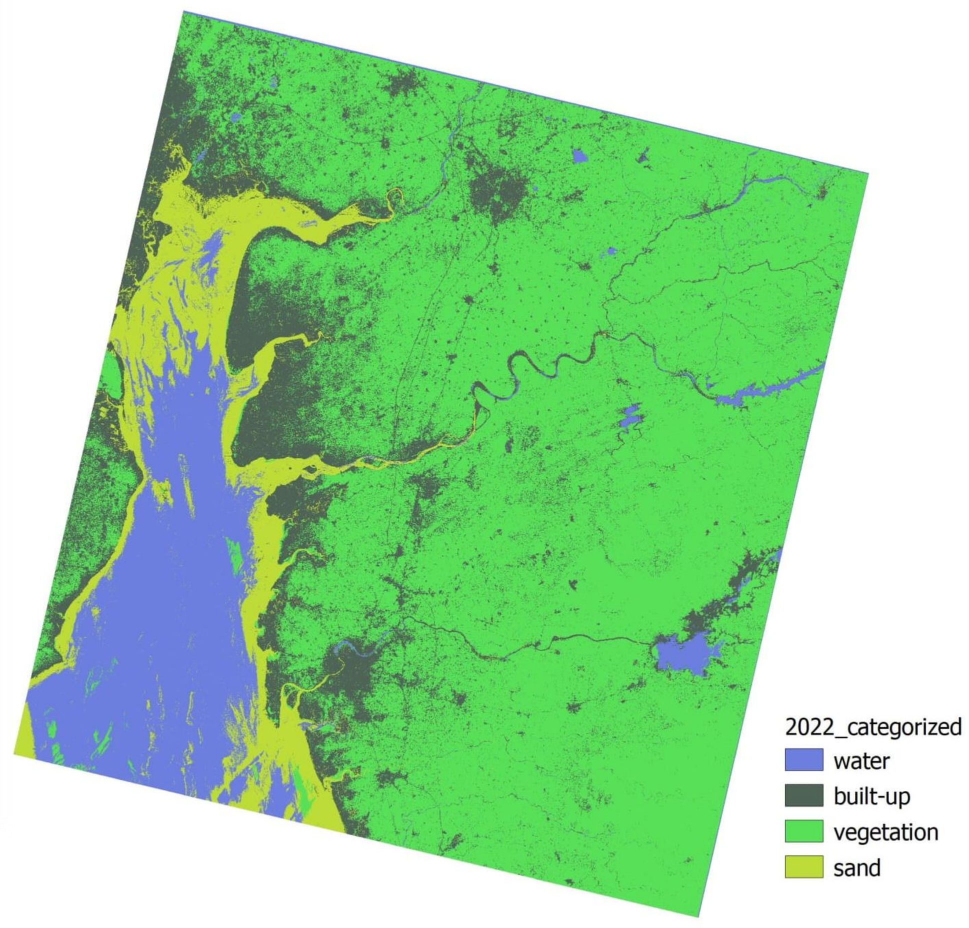
Vectorization of AOI and uniformly sampling the pixels

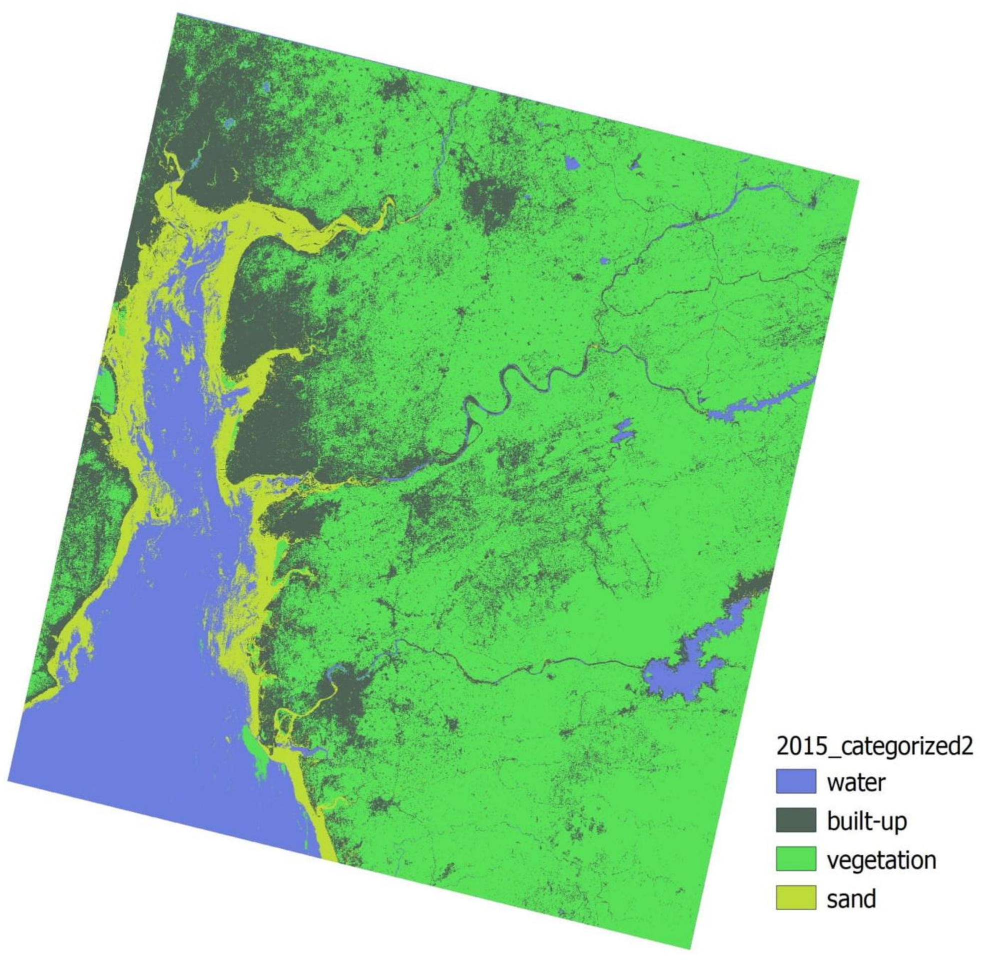
Training 100 random forest on AOIs on 7 bands + indices

Doing prediction for landcover classes on the whole image

RESULTS:

- Random Forest Classifier, accuracy (for test data) for each time step: **96.6286% for 2015, and 95.5097% for 2018., and 95.3931% for 2022.**
- The results show that most of the area contains water and vegetation for all the time steps, while built-up and sand are the smallest areas.
- From 2015 to 2018: large decrease in water, a decrease in built-up, a large increase in vegetation, and a small increase in sand.
- From 2018 to 2022, there is a large increase in water, an increase in built-up, a decrease in vegetation, and a slight decrease in sand.
- However, interestingly, the vegetation once again decreases a lot from 2018 to 2022. It is possible that for 2018 the sudden and large increase in vegetation (which is much higher than for 2015 and 2022) could be due to misclassification errors in the model
- When analyzing the classified image visually and compering them to Google satellite images for this time step, there was a lot more misclassification of vegetation areas and water for built-up, then for other time steps.
- Feature Contribution - for 2015: the band B5 (NIR) contributes the highest for the classes - water and vegetation. In contrast, the band B3 (Green) contributes more for the built-up class, and the NDWI index has is the highest contributing feature for the sand category.
- Feature Contribution - for 2022: the bands B5 remains the highest contributor for the water and vegetation classes, but for the built-up class, the band B2 (Blue) contributes more. For the sand class, the band B7 (SWIR) has the highest contributing factor.





Class Prediction Percentages

Classes	2015	2018	2022
1 - Water	42.11%	11.10%	40.86%
2 – Built-Up	13.06%	4.08%	11.74%
3 - Vegetation	40.27%	75.51%	41.77%
4 - Sand	4.55%	9.31%	5.63%

DISCUSSION:

- Arbitrary choice of the number of classes to be included in the Areas of Interest (AOI). As there was no publicly available data for the chosen city, polygons representing the various AOIs were manually drawn on QGIS and digitized for the reference data-set.
- the accuracy of the testing data slightly decreases over the years - it was 96.62% in 2015 followed by 95.5% in 2018, and at 2022, the accuracy decreases even further to 95.39%.
- Our assumption was that the NDBI, NDVI, and NDVWI indices would have the highest contributions compared to the other spectral bands, but most of the classes had one of the seven spectral bands as their highest contributing feature.
- The most misclassified area was vegetation and the best classified class was water across all the time steps. Vegetation changed the most and water changed the least across the time step.
- When it came to urban areas, the amount of built-up areas decreased a little bit instead of increasing from 2015 to 2022 (probably due to misclassification error).

CONCLUSION:

- The results of our study show that classification of urban areas with the Random Forest model can produce fairly accurate results (with the highest accuracy of a time step being 96.63%).
- Vegetation proved to be the most misclassified area across time steps, with vegetation unexpectedly increasing from 2015 to 2018, but then decreasing from 2018 to 2022.
- NDBI, NDVI, and NDWI did not contribute as much to the model as expected, when comparing feature contributions.
- Urban areas surprisingly decreased a little bit instead of increasing over time, which did not fit with the fact that Surat is one of the fastest growing cities in the world.
- Future research should focus on comparing different machine learning models (examples include SVM and Neural Networks), the tuning of parameters, and examining new indices instead of simply using one model to perform the classification and change detection. In addition, future studies could examine other cities and countries.

Feature Contributions - 2015

	('water', 1)	('buildup', 2)	('vegetation', 3)	('sand', 4)
B1	0.214431	0.730655	-0.848171	-0.096916
B2	0.291873	1.092124	-1.258748	-0.125248
B3	-0.058058	1.984720	-1.978688	0.052026
B4	0.267695	0.309762	-0.334375	-0.243083
B5	0.835359	-0.638694	0.202000	-0.398666
B6	-0.543324	0.270613	-0.273282	0.545993
B7	-0.128648	0.292276	-0.594082	0.430454
NDBI	0.159029	0.314102	-0.290086	-0.183046
NDVI	-0.329254	0.462366	-0.463503	0.330390
NDWI	-0.541952	0.057056	-0.250282	0.735179

Feature Contributions – 2022

	('water', 1)	('buildup', 2)	('vegetation', 3)	('sand', 4)
B1	0.383418	0.195368	-0.035925	-0.542861
B2	0.212163	0.622442	-0.345541	-0.489063
B3	-0.169150	0.119250	-0.130782	0.180682
B4	0.539857	0.012898	0.267373	-0.820128
B5	0.874382	0.082019	0.583356	-1.539758
B6	0.048450	-0.471534	0.097999	0.325085
B7	0.243654	-0.316958	-0.433475	0.506779
NDBI	0.556606	-0.090772	0.484442	-0.950276
NDVI	0.215529	0.241528	-0.373395	-0.083662
NDWI	0.304311	0.318435	-0.848832	0.226086

BIBLIOGRAPHY:

1. Saeid Zare Naghadehi, Milad Asadi, Mohammad Maleki, Seyed Mohammad Tavakkoli Sabour, John van Genderen, and Samira-Sadat Saleh. Prediction of urban area expansion with implementation of mlc, sam and svms' classifiers incorporating artificial neural network using landsat data. ISPRS International Journal of Geo-Information, 10:513, 07 2021.
2. Wanliu Mao, Debin Lu, Li Hou, Xue Liu, and Wenze Yue. Comparison of machine-learning methods for urban land-use mapping in hangzhou city, china. Remote Sensing, 2020.
3. R Asy'Ari A Ranti and T H Ameiliani. Detection of urban forest change in jabodetabek megacity using sentinel 2 and landsat 8 imagery through google earth engine cloud computing platform. IOP Conf. Series: Earth and Environmental Science, 2022.
4. Pablo Pozzobon de Bem, Osmar Abilio de Carvalho Junior, Renato Fontes Guimaraes, and Roberto Arnaldo Trancoso Gomes. Change detection of deforestation in the brazilian amazon using landsat data and convolutional neural networks. Remote Sensing, 12(6), 2020.

Utrecht University