POLITECNICO
MILANO 1863

Earth Observation Advanced - Supervised Classification for Change Detection in Built-Up Areas

Sveva Zanetti

Code Repository: Link

September 3, 2025

1 Introduction

"During the last two decades Dubai built up area expanded like never before [...] The total built up area increased from only 54 square Kms in 1975 to 977 square Kms in 2015, as Dubai increased (1700 % in only 38 years) that high percentage make Dubai one of the fastest growing cities in the world" [1]
This rapid urban growth makes Dubai the ideal study area for investigating land cover changes using Sentinel 2 images. Sentinel 2 mission was chosen because it provides level-2A orthorectified atmospherically corrected surface reflectance images with a 10 m spatial resolution for RGB bands, that allows to detect finer details than Landsat 8 mission, that has a spatial resolution of 30 m.
The main downside is that the dataset is available only from 2017-03-28, so in order to see significant changes in the expansion of the built-up area it was crucial to find a region that has been rapidly expanding in the last 8 years.
Figure 1 shows in green the area of interest considered for this project.

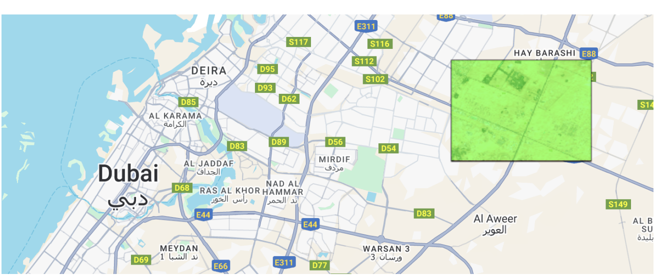

Figure 1: Area of Interest in green

The study is divided in 3 phases, the first one deals with the preparation of a hand-annotated dataset for the classification into 4 land cover classes: vegetation, water, soil and urban. The second phase deals with the selection of the best supervised model, choosing from Artificial Neural Network, Random Forest and Support Vector Machine. The third phase uses the final model to classify images at different epochs and detect the changes in built-up area.

2 Tools and libraries

The main tool used for the preparation of the hand-annotated dataset and the selection of the images too be classified is Google Earth Engine, a cloud-based platform for planetary scale geospatial analysis. The training and selection of the best model, prediction of land cover classes for each image and comparison of the results has been done in a Jupiter Notebook, using specific libraries for ML and for the visualization of geospatial results, such as:

- ee: a client library for Python that translate complex geospatial analyses to Earth Engine requests. [2]
- tensorflow: a leading open-source library designed for developing and deploying state-of-the-art machine learning applications. [3]
- scikit-learn: an open source machine learning library that provides various tools for model fitting, data preprocessing, model selection, model evaluation, and many other utilities. [4]

- shapely: a Python package for manipulation and analysis of planar geometric objects. [5]
- geemap: a Python package for interactive geospatial analysis and visualization with Google Earth Engine. [6]

3 Data

This project is entirely based on Sentinel 2 level-2A orthorectified atmospherically corrected surface reflectance data, taken from the Earth Engine Data Catalog.
Starting from 2018, one image every two years have been created, taking the median over the original collection filtered for the month of December and considering only images with a cloudy percentage lower than 10 %.
Sentinel 2 has 13 spectral bands, but this project only considers B2, B3, B4, B8, B11, B12, that respectively correspond to Blue, Green, Red, NIR, SWIR1, SWIR2. Additionally, 3 indexes have been computed and added as bands to the images: NDVI, NDWI and NDBI, that respectively help recognize vegetation, water and built-up areas.
The area of interest is a rectangle north-east of Dubai (Figure 1), it is in the outskirts of the city, but it is rapidly expanding.

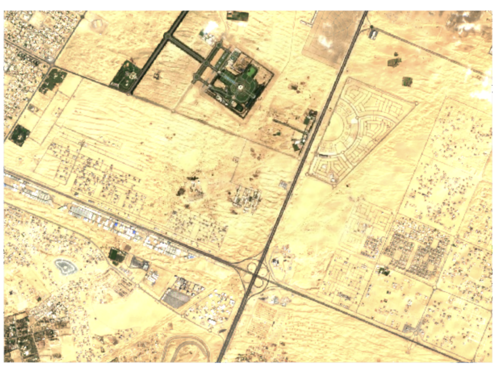

Figure 2: Median image of December 2018 of the AOI

As shown in Figure 2, the area is dry and desert, for this reason the project classifies just in 4 classes:

- water: private pools or artificial small lakes
- vegetation: grass and trees
- soil: soil and sand
- urban: buildings, streets, factories

4 Methods

This project uses supervised learning to classify pixels in 4 different land cover classes. The images to be classified are all relative to the month of December, so that the condition of vegetation and soil are as similar as possible across different images.
The aim is to classify images of 2018, 2020, 2022, 2024 using a single model trained on the image of 2018, and then check if the accuracy is still acceptable when the model is used on the last image, relative to 2024.

4.1 Dataset creation

To train and test the model, two hand annotated datasets where needed:
The first one is built starting from points of the 2018 image, and is divided in training, validation and test. The validation dataset will be used to choose the best model, and the test dataset will be used to assess the accuracy of the final model.
The second dataset is a test dataset built using points of the 2024 image and is needed to check if the model maintains good performance on images at a different epoch than the one used for training.
Both datasets are obtained sampling random points from polygons drawn on the image, in this way is possible to get a good number of points from different locations in the image.

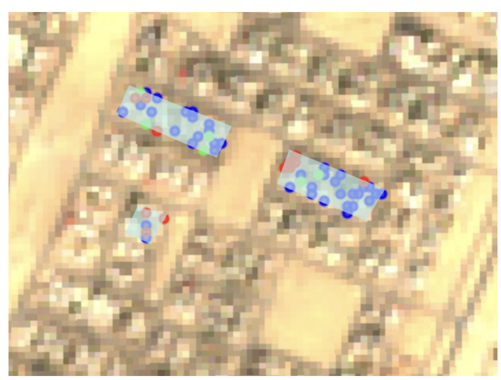

Figure 3: Points randomly extracted from urban polygons

Figure 3 shows in light blue some urban polygons, from which training (blue), validation (green) and testing (red) points have been sampled. The number of points in the final dataset of 2018 is shown in Table 1. Two classes have less data, because water and vegetation are present just is small areas of the image.

Table 1: Samples in 2018 dataset

Vegetation	Water	Soil	Urban
679	49	983	1025

Once all the points have been sampled, 70 % has be randomly extracted and used for training, 15 % has been used for validation and 15 % has been used for testing.

4.2 Model selection

The project compares the performance of 3 different supervised approaches: ANN, RF and SVM. Two different ANN have been implemented to see the difference between a simple and a more complex model. The first one has just one hidden layer with 64 neurons, followed by a dropout layer, while the second one has 3 hidden layers with 64, 32 and 16 neurons respectively, all followed by a dropout layer, and implements early stopping and adaptive learning rate callbacks.
All the 4 models have been trained using the same training set and evaluated using the validation set, and the most promising model has later been trained on the unified dataset of training and validation using cross-validation to tune the hyperparameter, and once the hyperparameters have been chosen the model has been trained one last time using all the data of training and validation, checking the final performance on the test set to see if the model overfitted or not the data. The final model has been tested also on the test dataset from 2024, to check if the performance decreased or not on images at a different epoch than the one used for training.

4.3 Prediction

The final model has been used to predict the class of all pixels in the images of 2018, 2020, 2022, 2024 and then the results have been used to investigate the expansion of the built-up area.

5 Results

5.1 Model selection

Table 2 shows the accuracy of the 4 different models on the validation set, and Figure 4 shows the corresponding confusion matrices.

Table 2: Accuracy of the 4 models

ANN1	ANN2	SVM	RF
0.9098	0.9774	0.9774	**0.9975**

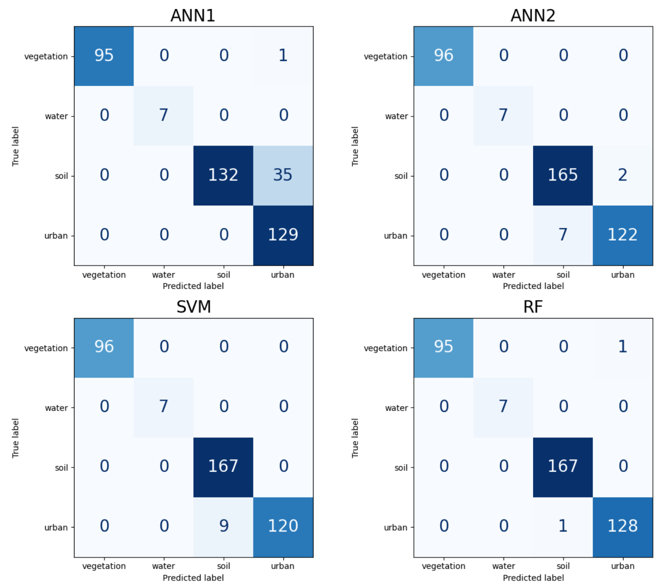

Figure 4: Confusion matrix of the 4 models

These results show very similar performance, but Random Forest, an ensemble method, outperformed the other models, and has also the advantage of been quickly trainable compared to the ANN models.
Cross validation has been performed on the RF model using a unified dataset of training and validation, and the combination of $n - estimators = 100$ and $max - depth = 10$ as hyperparameters obtained the best performance.

5.2 Performance evaluation

Table 3: Accuracy on the test sets

Test 2018	Test 2024
1.000	0.9798

The final RF model, trained using both training and validation dataset, has obtained good performance on both test sets, as shown in Table 3 .
(*See Discussion section for a possible explanation of why the accuracy is so high*)

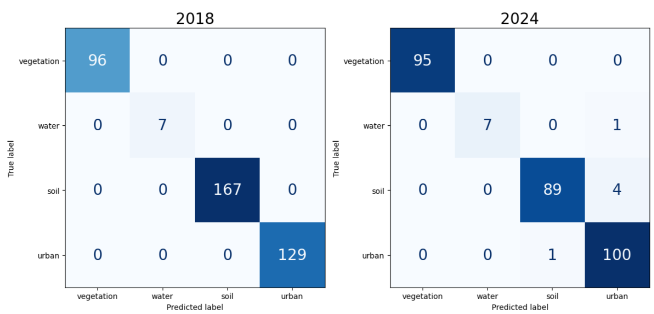

Figure 5: Confusion matrix on the 2 test sets

The confusion matrix of the test dataset of 2024 in Figure 5 shows that a few soil pixels have been misclassified as urban, but the performance did not drastically decrease on the test dataset of 2024, so the model has been used to classify images at different epochs.

5.3 Prediction

Figure 6 shows the prediction of the classes in the 4 images of 2018, 2020, 2022, 2024. As shown in Table 4 built-up area significantly expanded over the years and vegetation almost doubled its extent from 2018 to 2024.

Table 4: Total pixels and areas km^2

	Veg	Water	Soil	Urban
Pixel 2018	15707	336	663782	190643
Pixel 2024	29957	180	493858	346473
Area 2018 km^2	1.57	0.034	66.38	19.06
Area 2024 km^2	3.00	0.018	49.39	34.65
Percentage of change	+91%	-46%	-25%	+82%

Figure 6: Predicted label at different epochs

6 Outcomes

Figure 7 highlights in red new built-up areas from 2018 to 2024, while in blue are shown the areas that were already built in 2018, and it's clear that the extent of the built-up areas almost doubled, as stated in Table 4 .

Figure 7: Difference in built-up area 2018-2024

Most of the urbanized areas have increased building density, but there are two specific areas, at the bottom-left and center-right of the image, where the change is particularly significant, with two neighbourhoods built in areas that in 2018 contained just bare soil. Figure 8 compares a zoomed RGB image of 2018 and of 2024 of the new neighbourhood highlighted in a black box in Figure 7 .

Figure 8: Zoomed RGB 2018 vs RGB 2024

By looking at the confusion matrix on the test set of 2024 in Figure 5 , some pixel belonging to the soil class have been incorrectly classified as urban. These pixels are located in the thin red stripes shown in the comparison between labelled and RGB image of Figure 9 and correspond to the shadowed areas of the sand dunes.

Figure 9: Predictions 2024 vs RGB 2024

Even if there are these misclassified pixels, the accuracy on the test set of 2024 was very high and this indicates that the results obtained are a good estimate of the expansion of the urban area.

7 Discussion

All the models used in this project have shown very good performance, but the accuracy has probably been positively biased by the construction of the dataset: the polygons from which the random points used to build the dataset were drawn corresponded to areas that clearly belonged to one of the 4 classes. Ambiguous pixels have been left out of the training, validation and testing dataset because it was difficult to assign a label just by looking at the picture. This implies that point used for validation or testing might be easier to classify also for the model, and the real accuracy over all the pixels in the image is probably lower than the one obtained.

References

[1] F. Elessawy. The boom: Population and urban growth of dubai city. *Horizons in Human and Social Sciences*, 2:26–41, 2017.

[2] Earth engine. `https://developers.google.com/earth-engine/guides/client_server`, 2024. Accessed: 2 Sept 2025.

[3] Tensorflow. `https://www.nvidia.com/en-us/glossary/tensorflow/`. Accessed: 2 Sept 2025.

[4] Scikit-learn. `https://scikit-learn.org/stable/getting_started.html`. Accessed: 2 Sept 2025.

[5] Shapely. `https://shapely.readthedocs.io/en/stable/`. Accessed: 2 Sept 2025.

[6] Geemap. `https://geemap.org`. Accessed: 2 Sept 2025.