Spatiotemporal land use land cover (LULC) change analysis of urban narrow river using Google Earth Engine and Machine learning algorithms in Monterrey, Mexico

Kevin David Rodríguez González[1], Luis Eduardo Arista Cázares[1], Fabiola Doracely Yépez Rincón[1]

[1]Geomatics Department, Instituto de Ingeniería Civil, UANL, Av. Universidad s/n, Anáhuac, C.P. 66455, San Nicolás de los Garza, Nuevo León, México.

Keywords: Spatiotemporal, Machine Learning, LULC, Change Analysis, Urban narrow river.

Abstract

This study evaluates four Machine Learning Algorithms—Random Forest (RF), K-Means Clustering, Support Vector Machine (SVM), and Classification and Regression Trees (CART)—for precise land use and land cover (LULC) classification in the Monterrey Metropolitan Area. During the period 2016-2019, and with alternating wet and dry season classifications, the research addresses challenges in identifying narrow rivers, using geospatial tools and it does notably the Pesqueria River, which is specially the most narrow and shallow river in the area. Five classes—Water, Vegetation, Urban, and Soil—were classified, achieving precision rates above 85%. Remarkably, SVM exhibited an excellent accuracy, particularly for narrow rivers, showcasing its utility in complex urban landscapes. The study utilizes high resolution satellite imagery with a spatial resolution of 4.7m, contributing to the reliability of the results. Emphasizing temporal dynamics, the research links LULC changes to urbanization, infrastructure, and seasonal variations, offering vital insights for sustainable urban development.
.

1. Introduction

LULC is a global phenomenon with profound environmental implications. The research of Bajocco et al. (2012) demonstrate that, together with climate change and soil deterioration, LULC alters the hydrological cycle, progressively degrading the ecosystem and reducing the quality of land resources, biodiversity and agriculture. Riparian vegetation also plays a vital role in maintaining and improving the environmental quality of these aquatic environments. This strip of vegetation along the banks, known as the riparian zone, offers a series of benefits that contribute to the sustainability and resilience of urban rivers (Canales G., 2023).

In Mexico rivers face many problems that are preceded by the consequences of anthropogenic activities, some of these are the modification of the cross section of the channel, water quality, decrease in riparian vegetation, occupation of floodplains, reduction of biological diversity, loss of aquatic ecosystems or climate change. In addition to this, in Nuevo León, Mexico, demographic growth, industries and illegal urban settlements are constantly changing the soil conditions and in turn the flow and continuity of its river and its riparian vegetation (Zhao et al., 2021). The management and conservation of urban rivers represents a challenge for the country that can be addressed with new technologies that allow presenting a precise environmental response and in relevant times, such as the use of geospatial Earth Observation tools and models, and high-precision technologies based on artificial intelligence (AI) applied to the environment (Khan A & Sudheer M., 2022).

The optimal way to obtain LULC data is through remote sensing techniques, such as multi-temporal satellite images that are processed using Machine Learning techniques.

In this study, the objective is to classify the use and coverage of the land in the section of the Pesqueria River that is part of the Metropolitan Area, because, in recent years, the urban area surrounding the Pesqueria River has noticed a notable increase in human settlements.

2. Study case

The Pesqueria River Basin is located within the northeastern region of Mexico (Figure 1), covering parts of the state of Nuevo León, although its basin covers a part of the municipality of Saltillo, Coahuila. Geographically, the Pesqueria River is located between the latitudes of 25° 22' 12.30" N and longitudes 100° 50' 24.00" W. The Pesqueria River rises in the Sierra Madre Oriental, at an approximate altitude of 1,800 m.a.s.l. It descends through canyons and ravines until it flows into the Santa Catarina River, a tributary of the San Juan River, in the metropolitan area of Monterrey at about 500 meters above sea level.
For most of the year, this river behaves like an intermittent current, especially after entering the metropolitan area of Monterrey, which in turn is the area with the greatest population growth in recent years. Within this urbanized area there are industries, Wastewater Treatment Plants (WWTP) and many illegal settlements in the areas surrounding the river, which can considerably alter the nature of this aquatic ecosystem.

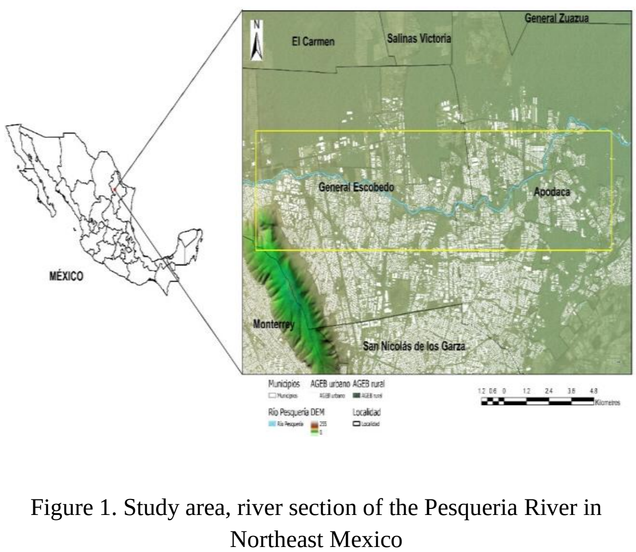

Figure 1. Study area, river section of the Pesqueria River in Northeast Mexico

The landscape within the Pesqueria River in the mostly urban area exhibits a diverse pattern of land use due to the changes it has presented in recent years. The classification of land use in the urban area of the Pesqueria River is essential to better understand the dynamics that affect this important river channel. In this context, the use of remote sensors becomes a very valuable tool to address this type of large-scale studies. scale coupled with subsets of AI.

3. Data and method

3.1 Data

This project used Norway's International Climate and Forest Initiative (NICFI) satellite imagery for its LULC. Unlike very high-resolution commercial satellite images (Table 1) which can have high costs, NICFI images are available through a free license which can be linked directly to Google Earth Engine (GEE) for continuous use and advanced.

Data Layer	Source	N° Band	Name	Description	Spatial resolution (m)
NICFI	GEE, Planet	1	B	Blue	4.7
		2	G	Green	4.7
		3	R	Red	4.7
		4	N	NIR	4.7

Table 1. NICFI band information used for LULC classification

3.2 Data description

In this research project, a time series of satellite data acquired between the years 2016 and 2019 was used, in order to carry out a multi-temporal analysis of the spatio-temporal dynamics of land cover and land use in the study area.

These temporalities included images captured during different phenological periods (Table 2), covering both dry and rainy periods, this allows characterizing the seasonal variations presented by the different types of soil and the Pesqueria River throughout the urban area. In general, five seasons were taken into account, of which three are dry and two are rainy.

Satellite Images	Season	Years
NICFI	Dry	2016-2017
	Wet	2017-2017

Table 2. NICFI seasonal specification

The delimitation of the seasons was based on the analysis of historical data on temperature and total monthly precipitation, for the period between 2008 and 2017, carried out by Meléndez (2020). These records were obtained from the meteorological stations of the Environmental Monitoring System (SIMA as in Spanish) located within the study area. Based on this analysis, two well-defined seasonal periods were established: the dry season, which covers the months of November to April, and the rainy season, which covers the months of May to October.

3.3 Methodology

The methodology used in this study follows a workflow for land use/cover classification using remote sensing data (Figure 2). The main source of input data is NICFI satellite images. The first step involves the collection of these images within the stipulated seasons, followed by geometric and atmospheric corrections using the GEE platform. The corrected images undergo the LULC classification process, which involves two main approaches (Table 3): supervised classification. There are three supervised classification algorithms: CART, SVM and RF. These algorithms are trained with a set of representative samples for the 5 classes Water, Vegetation, Agricultural Zones, Urban Growth and Soil, allowing them to learn the spectral patterns and discriminate between different types of coverage.

LULC Classes	*Study periods*	
	2016-2017	2018-2019
Water	44	45
Vegetation	50	50
Agricultural zone	52	62
Built-up Urban area	50	54
Fallow land	50	40
Total	246	251

Table 3. Number of reference data for each LULC

Likewise, an unsupervised classification was developed using the K-Means clustering algorithm. This method groups image pixels based on their spectral similarity without prior knowledge of land cover classes. The resulting groups are interpreted and labeled as different types of coverage.

To give greater veracity and reliability to the results of this classification, a validation process was carried out. Spectral indices (NDVI, NDWI, NDBI) were employed as key indicators of land cover in the study area, complemented by Earth observation data for validation purposes. The NDVI response image facilitated more accurate vegetation data acquisition, while NDWI and NDBI were utilized for water and urban class identification, respectively. Twenty percent of the acquired data was allocated for validation. This methodological process was significantly expedited through the use of GEE platform, its Application Programming Interface (API), and JavaScript programming language.

In order to evaluate the performance and determine the optimal algorithm for the prediction of different land uses and the identification (continuity) of intermittent and low-flow rivers in the urbanized area of the Pesqueria River, confusion matrices were generated, Kappa indices and overall accuracies for each of the supervised and unsupervised classification models. This rigorous analysis was carried out in Python, applied to the different satellite images corresponding to the dry and rainy seasons. In this way, it was possible to obtain a precise quantification of the degree of precision of each model, allowing the most appropriate Machine Learning algorithm to be selected to map with high accuracy the various land covers present in the study area.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-2024-371-2024

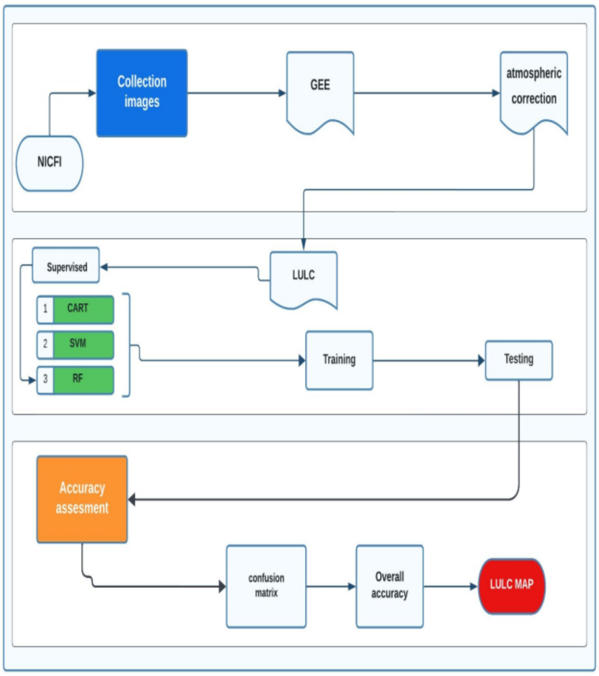

Figure 2. Workflow of the LULC.

4. Results

4.1 LULC Classification

Land use is a socioeconomic and environmental criterion that refers to the dominant activity that a biophysical indicator describes the materials that cover the territory to be studied. In this research, 5 different seasons and 4 Machine Learning models were taken into account for the classification. of land cover of the urban area of the Pesqueria River, Taking into account the classes of Water, Vegetation, agricultural areas, urban growth and soil.

Within these algorithms, The Supervised models, for their part, managed to show the different class contours and the environment that exists in the surrounding areas of the channel (Figure 3).

Supervised classification models are capable of identifying the contours of each of the classes studied effectively. However, when analyzing intermittent rivers with low flow, such as the case of the Pesqueria River, certain slight errors are observed in the classification of water and vegetation classes. Additionally, satellite images that have pronounced shadows, especially due to mountainous topography, tend to cause some confusion between water class and others due to the relationship between pixels.

K-means clustering (Figure 4), being an unsupervised model, managed to provide an indication of the different classes that can be found in the study area, being key to identifying the clusters in the different dry seasons. or rain in the surrounding areas of the Pesqueria River.

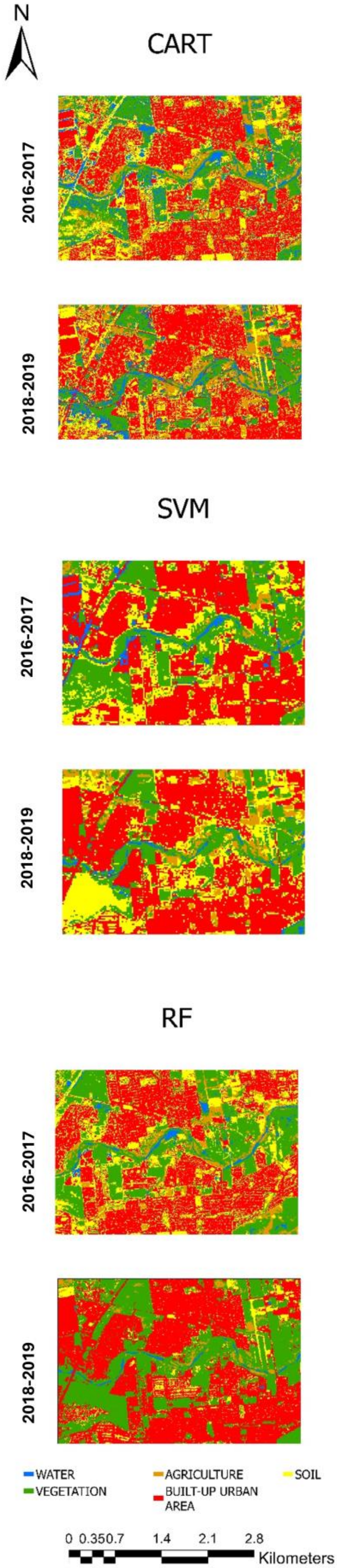

Figure 3. Maps of LULC.

https://doi.org/10.5194/isprs-annals-X-3-2024-371-2024 | © Author(s) 2024. CC BY 4.0 License.

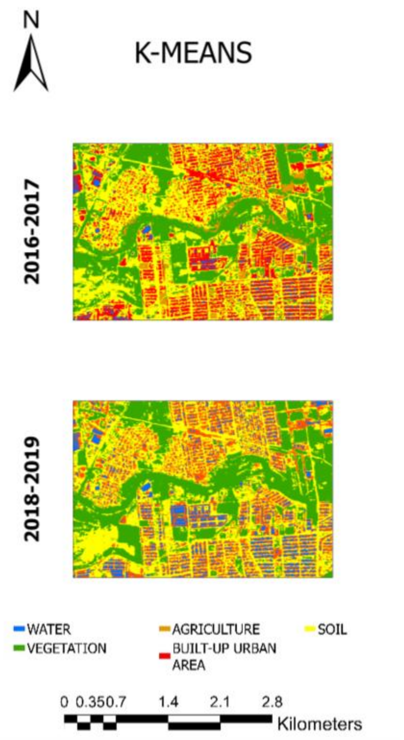

Figure 4. K-Means Clustering.

Supervised classification models are capable of identifying the contours of each of the classes studied effectively. However, when analyzing intermittent rivers with low flow, such as the case of the Pesqueria River, certain slight errors are observed in the classification of water and vegetation classes. Additionally, satellite images that have pronounced shadows, especially due to mountainous topography, tend to cause some confusion between water class and others due to the relationship between pixels.

During the seasons evaluated in this research, Kappa coefficient and general precision methods were used to identify the precision of the three models considered. For the first season, it is observed that RF demonstrated the best performance with an accuracy of 91% and a Kappa coefficient of 0.89, followed closely by SVM with an accuracy of 89% and a Kappa coefficient of 0.9.

These results underline the effectiveness of Machine Learning algorithms in classifying land use and narrow rivers such as the one evidenced in this project.

In the case of the second season, RF continues to present the best precision with 93% and a Kappa index of 0.91 (Table 4), thus guaranteeing the importance of this type of algorithms in the identification of the different classes and specifically in water.

Methods	Kappa coefficient	Overall accuracy
SVM	0.9	0.89
CART	0.88	0.84
RF	0.89	0.91

Table 4. Accuracy assessment and Kappa coefficient 2016-2017

SVM presented accuracies of 0.9, 0.89 and CART of 0.86 and 0.82 as shown in Table 5.

RF and SVM are robust and reliable approaches for land use classification, with RF showing slightly better performance in both seasons. CART, although acceptable, may require additional adjustments to improve its performance and stability in different conditions.

Methods	Kappa coefficient	Overall accuracy
SVM	0.91	0.89
CART	0.86	0.82
RF	0.93	0.91

Table 5. Accuracy assessment and Kappa coefficient 2018-2019

The confusion matrix was one of the tools used to evaluate the accuracy of the Machine Learning models in this study. Tables 6 and 7 present the matrices that exhibited greater accuracy in the training and validation data used. For the 2016-2017 season, high precision is observed in the classification of the classes "Water", "Vegetation", "Agricultural zone" and "Built-up urban area", with minimal confusion between them. However, there is some difficulty in differentiating between "Agricultural zone" and "Soil", since a sample of "Soil" was incorrectly classified as "Agricultural zone" by the contours of the pixels of the processed image.

One of the classification results are presented in Tables 6 and 7, which show the confusion matrices corresponding to the RF algorithm. This method demonstrated the highest overall accuracy, and the most elevated Kappa index values compared to the other classifiers evaluated in this study.

Water	Vegetation	Agricultural zone	Built up urban area	Soil
12	3	0	0	0
1	13	1	0	0
0	0	14	0	1
0	0	0	14	1
0	0	0	0	15

Table 6. Confusion matrix for 2016-2017

On the other hand, the confusion matrix for 2018-2019 still presented great precision of the observed data and only some differences between the contour of "Soil" with those of agriculture and urban growth.

Water	Vegetation	Agricultural zone	Built up urban area	Soil
15	1	0	0	0
0	14	1	0	0
0	0	14	0	1
0	0	0	15	1
0	0	1	1	9

Table 7. Confusion matrix for 2018-2019

Both confusion matrices reflected good performance of the Random Forest algorithm in land use classification, with generally high accuracies for most classes. However, it is important to highlight the specific difficulties that the model faced in each season. These confusions could be attributed to factors such as the spectral similarity between the classes and even to the specific environmental conditions of each season used.

To address these confounds, strategies such as adding additional spectral or texture features and integrating auxiliary data could be explored.

5. Conclusions

The results obtained in this study demonstrate the effectiveness of machine learning algorithms, specifically SVM, CART and RF, in the classification of land use in the urban area of Monterrey, Nuevo León, Mexico. The general accuracies achieved, which ranged between 0.86 and 0.93, as well as the Kappa indices, in the range of 0.82 to 0.90, for the two seasons analyzed (2016 and 2019), endorse the robustness of these methods to map and monitor changes in land use.

The analysis revealed that the RF model obtained better accuracy compared to SVM and CART. However, it was observed that SVM excelled in differentiating dark contours and plant shadows in relation to water, which represents a challenge due to the interdependence of these elements, and which commonly generates biases in Machine Learning models.

These findings are particularly relevant in the context of water resources management and understanding the behavior of the Pesqueria River, a small tributary with limited flow. By having accurate and up-to-date land use maps, it is possible to more accurately evaluate the impacts of human activities and land cover changes on the hydrological regime of this watershed.

It should be noted that these results are specific to the study area and the particular conditions of the urban area of Monterrey. However, the methodology used could be replicated in other regions or hydrographic basins, adapting it to local characteristics and using appropriate remote sensing data. In order to improve the veracity of the models created, it is expected that future projects will perform revalidations through different satellite images of greater or lesser spatial resolution.

6. Bibliography

Bajocco, S., Etchevers, T., Maggi, S., Cretaz, J., 2012. Land use/land cover changes and hydrological response of a mountain basin in Argentina. *Water Resources Management*, 26(7), 1807-1822

Canales-G, E., Peña-J, K. E., Cruz-R, B., Téllez-L, J., 2023. Loss of riparian vegetation cover due to dewatering of the Pitillal River in Jalisco. *Revista Chapingo Serie Ciencias Forestales y del Ambiente*, 29(2).

Gorelick., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.03.

Khan A, Sudheer M., 2022. Machine learning-based monitoring and modeling for spatio-temporal urban growth of Islamabad. *The Egyptian Journal of Remote Sensing and Space Sciences*. 25 541-550.

Melendez, E., 2020. "Diurnal Butterflies (Lepidoptera: Papilionoidea) as Indicators of Environmental Quality in the Metropolitan Area of Monterrey, Mexico." Doctoral Thesis. *Universidad Autónoma de Nuevo León* - Faculty of Forestry Sciences.

Talukdar S, Singha P, Mahato S, Pal S, Liou Y, Rahman A., 2020. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote sensing* 2020, 12, 113.

Zhao, Y.; An, R.; Xiong, N.; Ou, D.; Jiang, C., 2021. Spatio-Temporal Land-Use/Land-Cover Change Dynamics in Coastal Plains in Hangzhou Bay Area, China from 2009 to 2020 Using Google Earth Engine. *Land*, 10, 1149.