



Machine learning algorithms for land cover mapping in a Protected Natural Area

Algoritmos de aprendizaje automático para el mapeo de coberturas en un Área Natural Protegida

Natalia Martínez de León¹, Ignacio González Gutiérrez^{1*}, Xóchitl Celeste
Ramírez Campanur², Mario Rocandio Rodríguez¹, Arturo Medina Puente¹

Fecha de recepción/Reception date: 6 de junio de 2025.

Fecha de aceptación/Acceptance date: 12 de enero de 2026.

¹Facultad de Ingeniería y Ciencias, Universidad Autónoma de Tamaulipas. México.

²Facultad de Arquitectura, Universidad Michoacana de San Nicolás de Hidalgo. México.

*Autor para correspondencia; correo-e: Ignacio.gonzalez@uat.edu.mx

*Corresponding author; e-mail: Ignacio.gonzalez@uat.edu.mx

Abstract

Protected natural areas contribute to biodiversity conservation and to climate change mitigation, and provide ecosystem services. Accuracy assessment information on land cover and land use distribution is essential for managing these areas, and Sentinel-2 mission data are well-suited for monitoring them. Therefore, the objective of the study was to compare the performance of four machine learning algorithms —Support Vector Machine (SVM), Random Forests (RF), Gradient-Boosted Decision Trees (GBDT), and Classification and Regression Trees (CART)—, integrating spectral indices and topographic variables. The Sentinel-2 collection and a stratified sample set were used for validation ($n=641$). Accuracy was assessed using area-weighted confusion matrices. A two-proportion Z-test was used to compare the algorithms globally, and a McNemar *chi*-square test was used to compare predictions for each class. The results showed that SVM and GBT had the highest overall accuracy, of 88 % and 86 %, respectively. Comparison of the Z-test algorithms showed that half of the algorithm pairings were statistically different. McNemar's *chi*-square test showed that 46 % of the comparisons by class between paired algorithms were statistically significant ($p \leq 0.05$). In conclusion, machine learning algorithms enable the generation of accurate land cover and land use (LCLU) maps. Its implementation in decision-making is recommended due to its ability to recognize complex patterns.

Keywords: Thematic reliability, Google Earth Engine, Olofsson, McNemar test, Z test, Sentinel-2A.

Resumen

Las áreas naturales protegidas contribuyen a la conservación de la biodiversidad, la mitigación del cambio climático y brindan servicios ecosistémicos. La información precisa sobre la distribución de las coberturas y usos del suelo es fundamental para la gestión de estas zonas, y los datos de la misión *Sentinel-2* son adecuados para su monitoreo. Por ello, el objetivo del estudio fue comparar el rendimiento de cuatro algoritmos de aprendizaje automático: Máquina de Vectores de Soporte (SVM), Bosques Aleatorios (RF), Árboles de Gradiente Aumentado (GBT) y Árboles de Clasificación y Regresión (CART), integrando índices espectrales y variables topográficas. Se utilizó la colección *Sentinel-2* y un conjunto de muestras estratificadas para su validación ($n=641$). La fiabilidad temática se evaluó mediante matrices de confusión ajustadas al área. Se utilizó una prueba Z de dos proporciones para comparar los algoritmos a nivel global y una prueba χ^2 de *McNemar* para comparar las predicciones clase por clase. Los resultados mostraron que SVM y GBT presentaron la mayor fiabilidad global de 88 % y 86 %, respectivamente. La comparación de algoritmos de la prueba Z evidenció que la mitad de los emparejamientos de algoritmos eran estadísticamente diferentes. La prueba χ^2 de *McNemar* mostró que 46 % de las comparaciones por clase entre algoritmos pareados fueron estadísticamente significativas ($p \leq 0.05$). En conclusión, los algoritmos de aprendizaje automático permiten generar mapas precisos de cobertura y uso del suelo (CUS). Se recomienda su implementación en la toma de decisiones por su capacidad para reconocer patrones complejos.

Palabras clave: Fiabilidad temática, *Google Earth Engine*, *Olofsson*, prueba *McNemar*, prueba Z , *Sentinel-2A*.

Introduction

The use of satellite platforms such as Sentinel, Landsat, MODIS, Gaofen, and Worldview has grown considerably, and these have established themselves as essential tools for a wide range of applications, including the analysis of changes in land cover and land use (LCLU) (Zhao et al., 2022). In this context, machine learning algorithms have proven effective at generating accurate LCLU maps from satellite data (Zhao et al., 2024), meeting the need for reliable, up-to-date information. For example, the National Institute of Geography and Statistics (Inegi) provides LCLU maps, but only infrequently; the latest version dates back to 2017 and has an intermediate level of detail (scale 1:250 000). In contrast, Sentinel images allow for more frequent and detailed updates, with scales of up to 1:50 000 or less, a pixel size of 10 m.

In terms of accuracy assessment, traditional algorithms such as minimum distance achieve overall accuracy rates of around 83 % (Montenegro & Díaz, 2021), while the maximum likelihood method and Landsat images have recorded values of 75 % to 90 % (Camacho-Sanabria *et al.*, 2015; Escandón-Calderón *et al.*, 2018).

Among machine learning methods, Support Vector Machines (SVM) remain one of the most robust techniques due to their effectiveness in handling large-dimensional spaces (Sheykhmousa *et al.*, 2020). Random Forests (RF) remains relevant because they are easy to implement, as they require minimal hyperparameter tuning, and their data are easy to interpret (Ghatkar *et al.*, 2019). Despite advances in deep learning models, SVM and RF continue to be widely used in the scientific community, with overall accuracy rates exceeding 90 % (Zhao *et al.*, 2024).

In Mexico, studies that apply these algorithms are scarce. For example, Rodríguez-González *et al.* (2024) analyzed LCLU changes in the Northern part of the *Monterrey* metropolitan area using Planet images (4.7 pixels) for the 2016-2019 period. They concluded that RF performed best, with 91 % overall accuracy, followed by SVM, with 89 %, and Classification and Regression Trees (CART), with 83 %. Likewise, Rodríguez-Rosales *et al.* (2024) evaluated LCLU changes in *Huehuetla, Puebla*, using Landsat images from 2002 and 2021. The method they used was RF, and their accuracy rates were 92.5 % and 92.3 %, respectively. Keskes *et al.* (2025) mention that the Gradient-Boosted Trees (GBT) algorithm is the most visually accurate for environments with high variability.

Biodiversity loss and climate change represent the greatest environmental challenges today; therefore, having accurate LCLU classifications is essential for optimal ecosystem management. One of the platforms that has greatly facilitated the application of these algorithms is Google Earth Engine (GEE), because it enables the visualization and analysis of spatial data through an Application Program Interface (API) that provides access to extensive collections of satellite images and enables complex analyses to be performed in the cloud (Zhao *et al.*, 2021). The integration of machine learning algorithms such as SVM or decision tree algorithms into GEE is not always available in commercial or open-source software; therefore, they open up new

opportunities for image classification, as well as for the automated real-time monitoring of large areas, with high levels of accuracy at low cost (Yang et al., 2022).

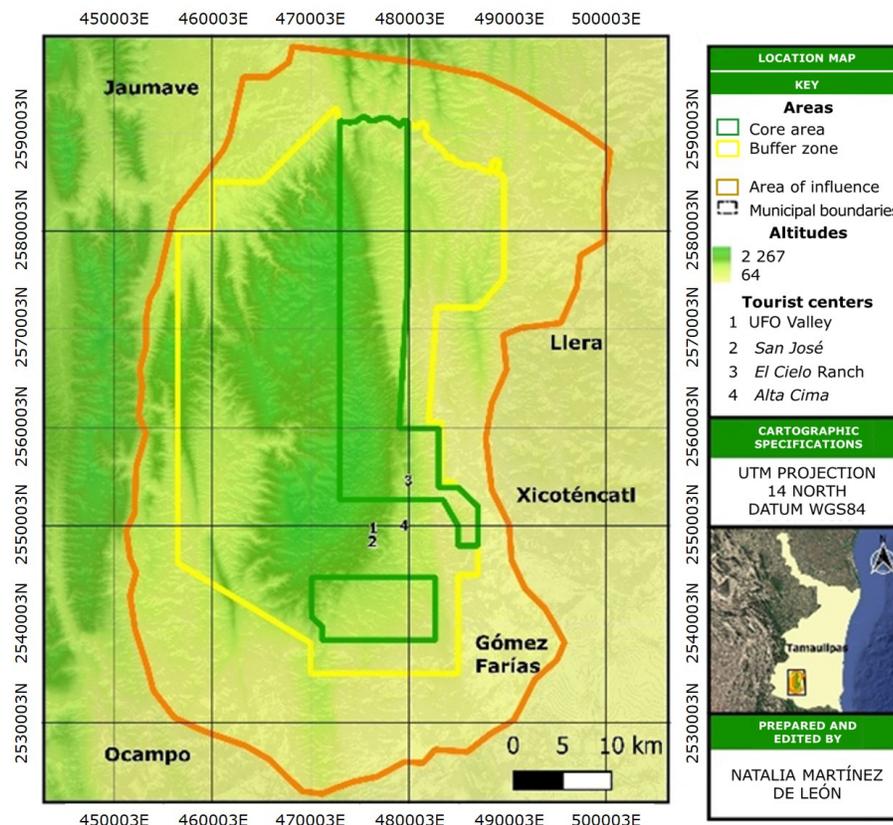
The *El Cielo* Biosphere Reserve (ECBR) in the state of *Tamaulipas*, Mexico, is an example of a heterogeneous area because it is home to a wide variety of ecosystems, including Tropical deciduous forest, pine and oak forests, Montane cloud forest, and submontane shrubland. Based on classification using machine learning algorithms, it is possible to analyze this ecological complexity, which is useful for studying species adaptation, the interaction between biomes and anthropogenic factors, and the effects of the altitudinal gradient on biodiversity. Thus, the objective of this study was to classify land use and vegetation types in the ECBR, as well as to compare various machine learning algorithms in order to evaluate the accuracy of classifications that can be adapted to heterogeneous landscapes. The implementation of machine learning algorithms in Google Earth Engine makes it possible to generate highly accurate LCLU classifications in large and ecologically complex areas.

Materials and Methods

Study area

The study area is the ECBR, which was declared a biosphere reserve in 1985 and is considered a priority region for the conservation of different types of vegetation. It is located in the Southwest of the state of *Tamaulipas*, Mexico (Figure 1), with an area of 2 695 km² (Rangel-Lucio, 2024). The predominant climates in the different types of vegetation are: semi-warm-subhumid with rainfall in summer, for tropical

deciduous forest; semi-warm-humid, with abundant rainfall in summer, in the montane cloud forest; temperate subhumid with rainfall in summer for pine-oak forests; and semi-dry-semi-warm with rainfall in summer for submontane shrubland (Vargas-Contreras & Hernández-Huerta, 2001). On the other hand, the study area is located between the transitional zone of the Nearctic and Neotropical biogeographic regions. The ECBR is divided into two central core areas covering a total of 376 km², a buffer zone of 1 446 km², and an area of influence comprising the entire protected area.



Jaumave = *Jaumave* municipality; *Ocampo* = *Ocampo* municipality; *Gómez Farías* = *Gómez Farías* municipality; *Xicoténcatl* = *Xicoténcatl* municipality; *Llera* = *Llera* municipality; *Tamaulipas* = State of *Tamaulipas*.

Figure 1. Map of *El Cielo* Biosphere Reserve, *Tamaulipas*, Mexico.

Satellite image

GEE allows to select Sentinel-2 images with a processing level of 2A that represent surface reflectance. The criterion used for selection was images with less than 10 % cloud cover. The final image was a product of the median of images taken between February and March 2024, which was resampled to a resolution of 10 m and reprojected to the EPSG:32614 coordinate system (WGS84 Datum, UTM Zone 14 North projection).

Auxiliary variables

Auxiliary variables such as spectral and topographic indices were included, as they improve classification significantly (Phan et al., 2020). The elevations and slope of the terrain were obtained from the Digital Elevation Model (DEM) of the National Institute of Statistics, Geography, and Informatics (INEGI, 2013). The spectral indices were the Normalized Difference Vegetation Index (NDVI) $\left(\frac{NIR-RED}{NIR+RED}\right)$ (Rouse et al., 1974), the Soil Adjusted Vegetation Index (SAVI) $\left(\left(\frac{NIR-RED}{NIR+RED+L}\right) \times 1 + L\right)$ (Huete, 1988), and the Normalized Difference Moisture Index (NDMI) $\left(\frac{NIR-SWIR1}{NIR+SWIR1}\right)$ (Gao, 1996).

Methods

The workflow (Figure 2) consisted of capturing 2 300 training fields distributed throughout the study area to classify the Sentinel-2A image using different machine learning algorithms, which represented eight LCLU classes: agricultural areas, montane cloud forest, mixed pine and oak forest, tropical deciduous forest, submontane shrubland, bare soils, human settlements, and grasslands.

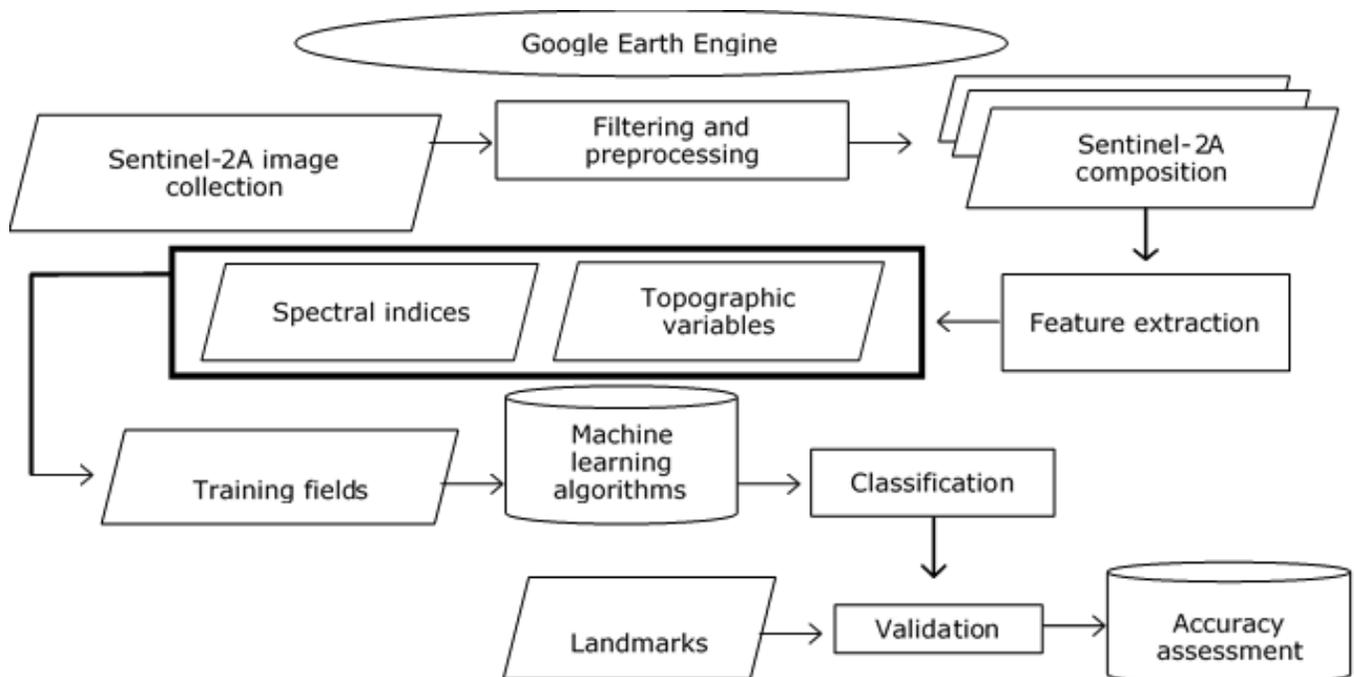


Figure 2. Workflow of the applied methods.

Machine learning algorithms

Table 1 shows the four machine learning algorithms used for classification.

Table 1. Description of the machine learning algorithms utilized.

No.	Machine learning algorithm	Description
1	Support Vector Machine (SVM)	A widely used nonparametric supervised method in classification and regression that seeks to optimize the margin of separation between classes. It stands out for its performance in large spaces, its ability to model nonlinear relationships using Kernel functions, and its computational efficiency, requiring only a fraction of the training data.
2	Classification and Regression Trees (CART)	Hierarchical model that recursively divides the attribute space using binary partitions, generating an interpretable structure. It is useful in both classification and regression, allows the relevance of variables to be identified, and provides models that are easy to interpret; however, its individual performance may be affected by data variability.
3	Random Forests (RF)	Set-based algorithms that use multiple decision trees independently, each trained with random subsets of features and data samples. By aggregating predictions through voting or averaging, it mitigates overfitting and improves the model's robustness.
4	Gradient-Boosted Trees (GBT)	An ensemble algorithm based on sequential optimization, which iteratively constructs decision trees to correct errors in the previous model. It is effective at capturing complex relationships in data, progressively improves predictive performance, and supports both numerical and categorical variables while maintaining good generalization capabilities.

Source: Khan et al. (2024).

Thematic accuracy

This study adopted the methodology of Olofsson *et al.* (2014) to accuracy assessment of classified maps. Thematic accuracy involves comparing classified results with reference data, and error matrices are used to assess the accuracy of the results. According to Olofsson *et al.* (2014): (a) the reference data must be of higher quality than the data utilized to create the map; (b) implement a probabilistic design indicating the sampling unit, sampling type, and sample size; (c) provide an adequate spatial and temporal representation, based on reference data, to accurately label each unit in the sample; (d) summarize the accuracy assessment in a confusion matrix in terms of proportion of area; (e) estimate overall accuracy, indicate errors of omission and commission; and (f) quantify the uncertainty of accuracy indices using confidence intervals.

The reference data were a high-resolution Google Earth mosaic downloaded from the SAS Planet application version 241111 (SAS Planet, 2024), with a visual interpretation performed by assigning the reference label to each sampling point. The sampling unit was one pixel per sample. Stratified random sampling is frequently used in thematic accuracy research (Leija *et al.*, 2020; Mas *et al.*, 2015), for which purpose Cochran (1977) and Olofsson *et al.* (2014) propose the equation 1 for sample size:

$$N = \left(\frac{\sum W_i S_i}{s(\hat{O})} \right)^2 \quad (1)$$

Where:

$s(\hat{O})$ = Standard error of the estimated overall precision to be achieved

W_i = Proportion of mapped area of the i class

S_i = Standard deviation of the stratum i , $S_i = \sqrt{U_i(1 - U_i)}$ (Cochran, 1977)

A target standard error was specified for the overall accuracy of 0.01.

In the study area, mixed and tropical deciduous forest predominate in total surface area, while other types are considered with small areas. Therefore, a mixed sample allocation approach was used to reconcile the need for accuracy in the overall and area estimates with the requirement to obtain accurate estimates for with small areas types (Olofsson et al., 2014). An allocation that is purely proportional to area would result in an insufficient number of samples for with small areas classes, leading to high standard errors for user accuracy (\hat{U}_i).

First, a fixed number of samples (50-100) was assigned to with small areas classes (Olofsson et al., 2014). This sample interval was based on the analysis of the estimated variance of user accuracy, with the aim of ensuring a sufficient sample size to achieve acceptable error in the estimation of the \hat{U}_i for these categories.

The information in Table 2 includes the proportions of mapped area (W_i), estimated user accuracy values (U_i), and standard deviations (S_i) for each class.

Table 2. Information needed for assigning sample size to classes.

Class	W_i	U_i	S_i	Alloc1
AA	0.100	0.480	0.500	50
MCF	0.005	0.903	0.296	50
MF	0.411	0.929	0.258	172
TDF	0.405	0.973	0.163	169
SS	0.035	0.771	0.420	50
BS	0.015	0.718	0.450	50
HS	0.004	0.921	0.270	50
GL	0.024	0.909	0.287	50

AA = Agricultural areas; MCF = Montane cloud forest; MF = Mixed forest; TDF = Tropical deciduous forest; SS = Submontane shrubland; BS = Bare soil; HS = Human settlement; GL = Grasslands. W_i = Percentage of area mapped; U_i = User accuracy; S_i = Standard deviation for each class.

“Alloc1” was chosen with 50 sampling units for six underrepresented classes. Accuracy was assessed using a confusion matrix, which consists of a cross-tabulation of the map classification (Sentinel-2) against reference data (SAS Planet). GEE allows calculating confusion matrices for each algorithm with the ‘ee.errorMatrix’ function. Card (1982) proposes a procedure for recording values in terms of estimated area proportion, which is essential for estimates to be unbiased under stratified random sampling. The proportion of mapped area for each class was estimated using equation 2 (Olofsson *et al.*, 2014):

$$\hat{P}_{ij} = W_i \frac{n_{ij}}{n_{i.}} \quad (2)$$

Where:

W_i = Proportion of the mapped area for the i class

n_{ij} = Number of pixels that were classified as class i

$n_{i.}$ = Total validation samples that were classified as class i

Overall accuracy (\hat{O}) represents the proportion of correctly classified area and is calculated by adding the diagonal elements of the adjusted matrix (equation 3) (Olofsson *et al.*, 2014).

$$\hat{O} = \sum_{j=1}^q \hat{P}_{jj} \quad (3)$$

Where:

\hat{P}_{jj} = Estimated proportion of the area that was correctly classified as class j

q = Total number of classes in the confusion matrix

Estimated user accuracy (\hat{U}_i) indicates the proportion of areas classified as i on the map that actually correspond to class i (equation 4) (Olofsson et al., 2014).

$$\hat{U}_i = \frac{\hat{P}_{ii}}{\hat{P}_{i.}} \quad (4)$$

Where:

\hat{P}_{ii} = Proportion of the area correctly classified as class i (value of the diagonal in the adjusted error matrix)

$\hat{P}_{i.}$ = Sum of all area proportions classified as class i

Estimated producer accuracy (\hat{P}_j) measures the accuracy of the algorithm from the perspective of the map producer and is calculated using equation 5 (Olofsson et al., 2014).

$$\hat{P}_j = \frac{\hat{P}_{jj}}{\hat{P}_{.j}} \quad (5)$$

Where:

\hat{P}_j = Estimated proportion of the area of class j

\hat{P}_{jj} = Proportion of the area correctly classified as class j (value on the diagonal of the adjusted error matrix)

$\hat{P}_{.j}$ = Sum of all area proportions that correspond to class j in the reference data

The variance of the overall accuracy of the map $\hat{V}(\hat{O})$, of the user's accuracy for class i $\hat{V}(\hat{U}_i)$, and the producer accuracy for class j $\hat{V}(\hat{P}_j)$ are estimated using equations 6, 7, and 8, respectively (Olofsson *et al.*, 2014):

$$\hat{V}(\hat{O}) = \sum_{i=1}^q \frac{w_i^2 \hat{U}_i (1 - \hat{U}_i)}{(n_i - 1)} \quad (6)$$

$$\hat{V}(\hat{U}_i) = \frac{\hat{U}_i (1 - \hat{U}_i)}{(n_i - 1)} \quad (7)$$

$$\hat{V}(\hat{P}_j) = \frac{1}{\hat{N}_{.j}^2} \left[\frac{N_{j.}^2 (1 - \hat{P}_j)^2 \hat{U}_j (1 - \hat{U}_j) + \hat{P}_j^2 \sum_{i \neq j}^q N_{i.}^2 \frac{n_{ij}}{n_i} \left(1 - \frac{n_{ij}}{n_i}\right)}{(n_i - 1)} \right] \quad (8)$$

Where:

$\hat{N}_{.j} = \sum_{i=1}^q \frac{N_{i.}}{n_i} n_{ij}$ = Estimated total marginal number of pixels in the reference class j

$N_{j.}$ = Total area mapped as class j

$N_{i.}$ = Total area mapped as class i

n_j = Total reference samples as class j

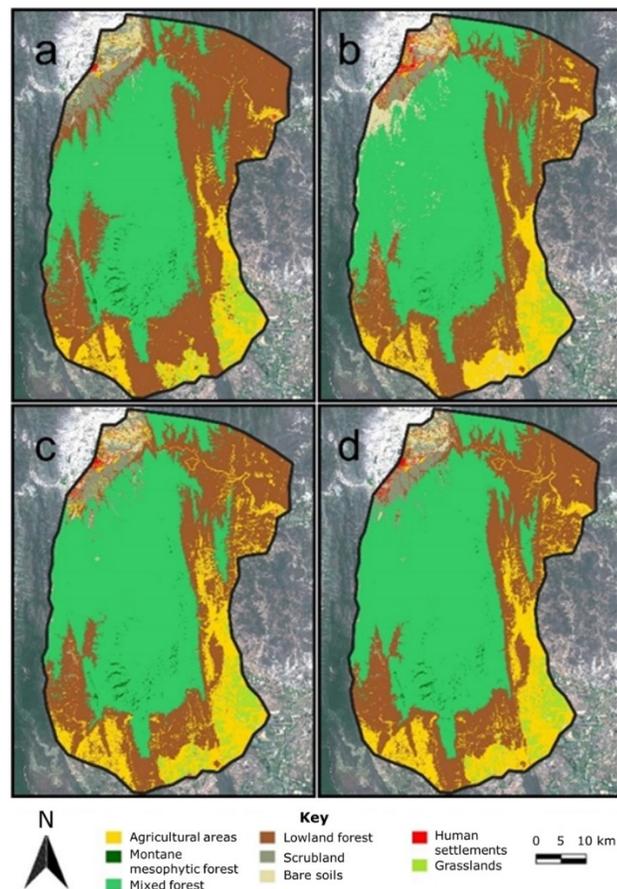
The 95 % confidence intervals are estimated as $\pm 1.96 \sqrt{\hat{V}(\hat{U}_i)}$ (\hat{U}_i is replaced with \hat{P}_j and \hat{O} for producer accuracy and overall reliability, respectively).

Statistical tests for comparing algorithms

The Kappa coefficient is a widely used measure for assessing thematic reliability in maps by considering the agreement attributable to chance. However, its use is not recommended for comparing two algorithms (Balha et al., 2021); since its validity depends on the independence between evaluators, the use of the same validation samples invalidates this assumption. In these cases, the McNemar test is more appropriate, as it is designed to evaluate differences between algorithms (Zar, 2009). McNemar's *chi*-square (χ^2) test, which allows for the evaluation of marginal homogeneity between predictions (McNemar, 1947), was applied. This homogeneity refers to equality in the marginal distributions of classifications made by two algorithms, *i. e.*, to the absence of statistically significant differences between them. This test is considered an efficient tool for making class-by-class comparisons, as it is a parametric procedure with a low risk of type I error and a simple formulation (Abdi, 2020). In addition, a *Z*-test for two proportions (Lachin, 1981) was used to compare the Correct Pixel Classification Ratio (CPR) between two algorithms at a time. The square of the *Z*-statistic generated by the test follows a χ^2 distribution with a degree of freedom (Abdi, 2020).

Results and Discussion

Figure 3 shows the classification maps generated with the four machine learning algorithms.



a = Classification and Regression Trees (CART); b = Support Vector Machine (SVM);
c = Random Trees (RF); d = Gradient-Boosted Trees (GBT).

Figure 3. Classified maps of the *El Cielo* Biosphere Reserve, *Tamaulipas*, Mexico.

Evaluation of the thematic accuracy of the algorithms

The overall accuracy results for the algorithms used are as follows: 88 % for SVM, 86 % for GBT, 84 % for RF, and 82 % for CART. Subsequently, user accuracy (*UA*) and producer accuracy (*PA*) metrics were examined for each of the eight LCLU classes, enabling determination of the strengths and limitations of each algorithm per class. Previous studies have highlighted RF as one of the most accurate algorithms in urban and homogeneous contexts, with accuracy rates exceeding 95 % (Zafar et al., 2024; Zhao et al., 2024). However, in this study, conducted in a heterogeneous landscape such as the ECBR, the SVM algorithm showed better performance, consistently with the findings of Abdi (2020), who estimated an accuracy rate of 76 % with SVM, followed by an accuracy rate of 74 % with RF.

Although the CART algorithm has been used successfully for Agricultural areas classification (Shelestov et al., 2017), in the present study, it was less accurate, which could be attributed to its sensitivity to overfitting in environments with high class diversity. In contrast, SVM and GBT achieved the highest overall accuracies (89 % and 86 %, respectively) and also demonstrated strong ability to distinguish classes with high spectral complexity.

GBT excelled in identifying agricultural areas (*PA*=95 %), while SVM showed greater user accuracy (*UA*=45 %), with fewer commission errors (Table 3). For the montane cloud forest, all algorithms had a low producer accuracy (*PA*= 13-19 %), although SVM obtained the highest *UA* (97 %), indicating less confusion with other classes.

Table 3. Assessment of the accuracy of land cover and land use, showing the proportion of the classified area (W_i).

Class	SVM			GBT			RF			CART		
	W_i	UA	PA									
AA	0.10	0.45	0.90	0.14	0.36	0.95	0.15	0.36	0.80	0.14	0.37	0.78
		±0.10	±0.06		±0.09	±0.06		±0.09	±0.07		±0.09	±0.06
MCF	0.01	0.97	0.19	0.01	0.93	0.17	0.01	0.96	0.13	0.01	0.88	0.19
		±0.06	±0.06		±0.10	±0.08		±0.07	±0.07		±0.11	±0.09
MF	0.41	0.95	1.00	0.48	0.92	1.00	0.47	0.93	1.00	0.51	0.89	1.00
		±0.03	±0.02		±0.04	±0.02		±0.04	±0.02		±0.05	±0.03
TDF	0.41	0.96	0.98	0.30	0.97	0.98	0.30	0.96	0.92	0.28	0.95	0.85
		±0.04	±0.06		±0.03	±0.11		±0.04	±0.10		±0.04	±0.11
SS	0.04	0.68	0.76	0.03	0.91	0.83	0.03	0.90	0.74	0.01	0.88	0.11
		±0.13	±0.14		±0.12	±0.36		±0.13	±0.41		±0.17	±0.12
BS	0.02	0.73	0.33	0.01	0.77	0.41	0.01	0.74	0.42	0.03	0.57	0.73
		±0.11	±0.09		±0.09	±0.07		±0.10	±0.08		±0.10	±0.08
HS	0.01	0.92	0.19	0.07	0.93	0.16	0.01	0.92	0.28	0.01	0.75	0.17
		±0.07	±0.06		±0.07	±0.06		±0.07	±0.06		±0.12	±0.10
GL	0.02	0.91	0.40	0.03	0.94	0.44	0.03	0.94	0.44	0.02	0.98	0.30
		±0.07	±0.09		±0.06	±0.09		±0.06	±0.09		±0.04	±0.08

AA = Agricultural areas; MCF = Montane cloud forest; MF = Mixed forest; TDF = Tropical deciduous forest; SS = Submontane shrubland; BS = Bare soil; HS = Human settlement; GL = Grasslands. W_i = Percentage of area mapped; UA = User accuracy; PA = Producer accuracy. SVM = Support Vector Machine; GBT = Gradient-Boosted Trees; RF = Random Forests; CART = Classification and Regression Trees.

In the submontane shrubland classification, decision trees outperformed SVM in user accuracy (>88 %), although CART exhibited a low producer accuracy. For bare soils, GBT and CART showed complementary performances, suggesting a possible future combination. Finally, for human settlements and grasslands, all algorithms achieved high UA (>90 %) but low PA values, indicating errors of omission.

Comparison between algorithms

The algorithms were analyzed from two perspectives: a detailed class-level comparison using McNemar's *chi*-square test, and a global-level comparison of classification accuracy using the two-proportion Z-test.

The results of McNemar's *chi*-square test (Table 4) showed that 31 % of comparisons within each class between paired algorithms were statistically significant ($P \leq 0.01$). Additionally, 10 % of comparisons showed marginal differences ($0.05 < p \leq 0.10$). In contrast, 44 % of the comparisons showed no statistically significant differences ($p > 0.10$).

Table 4. McNemar's *chi*-square (χ^2) test with its associated probability value (P).

Algorithm/ Class		AA	MCF	MF	TDF	SS	BS	HS	GL
SVM vs. RF	χ^2	4.90	3.00	3.00	6.76	25.14	6.72	0.53	3.00
	P	**	*	*	***	***	***	NS	*
SVM vs. GBT	χ^2	6.42	1.80	2.67	9.00	22.15	8.40	0.80	3.00
	P	**	NS	NS	***	***	***	NS	*
SVM vs. CART	χ^2	7.08	1.00	3.77	9.00	27.46	21.88	3.27	1.19
	P	***	NS	**	***	***	***	*	NS
RF vs. GBT	χ^2	0.36	0.00	0.33	0.33	1.80	0.40	4.46	0.00
	P	NS	NS	NS	NS	NS	NS	**	NS
RF vs. CART	χ^2	0.86	5.00	1.14	1.19	2.00	11.31	7.26	11.00
	P	NS	**	NS	NS	NS	***	***	***
GBT vs. CART	χ^2	0.10	5.00	0.60	0.39	4.46	10.31	2.46	11.00
	P	NS	**	NS	NS	**	***	NS	***

*** = $P \leq 0.01$, ** = $P \leq 0.05$, * = $P \leq 0.1$; NS = Not Significant. AA = Agricultural areas; MCF = Montane cloud forest; MF = Mixed forest; TDF = Tropical deciduous forest; SS = Submontane shrubland; BS = Bare soil; HS = Human settlement; GL = Grasslands. SVM = Support Vector Machine; GBT = Gradient-Boosted Trees; RF = Random Forests; CART = Classification and Regression Trees.

The smallest differences were observed between RF and GBT, which performed similarly across most of the classes analyzed. In particular, non-significant comparisons were identified between the pairs RF vs. GBT, RF vs. CART, and GBT vs. CART across the agricultural areas, mixed forest, and tropical deciduous forest classes. This suggests that these algorithms perform similarly in certain covers, particularly those with more defined spectral patterns, which could be exploited in future applications to optimize computational resources without compromising accuracy.

Among the significant cases in the McNemar test, comparisons with SVM showed the highest number of significant differences, especially in relation to CART and GBT, in the tropical deciduous forest, submontane shrubland, and bare soil classes ($P \leq 0.05$); this suggests that both these algorithms behave similarly across most classes. The most frequent differences occurred in cover types with high spectral complexity, such as submontane shrubland and bare soil. Comparisons between GBT and CART also showed differences at the level of 1 % for bare soil and grassland.

The results of the present study partially agree with Abdi (2020), who cites significant differences ($P \leq 0.05$) in 62 % of predictions per class when comparing SVM, RF, and Xgboost. In both studies, significant discrepancies are observed between SVM and RF for the tropical deciduous forest and bare soil classes ($P \leq 0.01$). However, differences were identified in agricultural areas, not recorded by Abdi (2020), and no discrepancies were found in human settlements, where this author did find them. This suggests that, despite using the same satellite collection, landscape composition and local spectral characteristics influence each algorithm's sensitivity to detecting specific classes.

The results of the two-proportion Z-test (Table 5) indicated statistically significant differences in CPR between the SVM and CART algorithms ($X^2=3.38$), between RF and CART ($X^2=2.73$), and between GBT and CART ($X^2=2.65$). On the other hand, no significant differences were observed between SVM and RF, SVM and GBT, or between RF and GBT. This indicates that CART performs significantly differently in pixel classification, whereas SVM, RF, and GBT show statistically similar accuracy.

Table 5. Two-proportion Z-test.

Algorithm pair	Two-proportion Z-test	P-value
SVM vs. RF	$\chi^2=0.66$	0.5100
SVM vs. GBT	$\chi^2=0.73$	0.4600
SVM vs. CART	$\chi^2=3.38$	0.0007
RF vs. GBT	$\chi^2=0.07$	0.9400
RF vs. CART	$\chi^2=2.73$	0.0064
GBT vs. CART	$\chi^2=2.65$	0.0080

The two-proportion Z-test was applied bilaterally to compare the correct pixel classification ratio (CPR) between algorithm pairs. The Z statistic is reported as its χ^2 equivalent with one degree of freedom. A significance level of $\alpha=0.05$ was adopted; values of $p \leq 0.05$ indicate statistically significant differences. SVM = Support Vector Machine; GBT = Gradient-Boosted Trees; RF = Random Forests; CART = Classification and Regression Trees.

Conclusions

The SVM and GBT algorithms proved to be generally the most effective for mapping land cover and land use in the ECBR, with overall accuracies of 88 % and 86 %, respectively. SVM excelled at classifying tropical deciduous and mixed forests, while GBT showed balanced performance across multiple classes. On the other hand, RF and CART proved useful for specific classes such as grasslands and bare soils. However, all algorithms have limitations in classes with low spatial extent, such as montane cloud forests, where high omission errors are recorded. These findings fulfill the objective of comparing machine learning algorithms and identify SVM and GBT as the most accurate. Likewise, accuracy assessment using confusion matrices enabled the determination of the strengths and limitations of each algorithm by class.

Acknowledgments

The authors are grateful to Ministry of Science, Humanities, Technology, and Innovation (Secihti) for the graduate studies scholarship awarded to the first author as well as the postgraduate program of the Faculty of Engineering and Sciences of the Autonomous University of Tamaulipas for carrying out the research

Conflict of interest

The authors declare no conflict of interest.

Contribution by author

Natalia Martínez de León: information search, data analysis, and drafting of the manuscript; Ignacio González Gutiérrez: data analysis and revision of the manuscript; X. Celeste Ramírez Campanur: information search and revision of the manuscript; Mario Rocandio Rodríguez: statistical analysis and revision of the manuscript; Arturo Medina Puente: revision of the manuscript.

References

- Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing*, 57(1), 1-20. <https://doi.org/10.1080/15481603.2019.1650447>
- Balha, A., Mallick, J., Pandey, S., Gupta, S., & Singh, C. K. (2021). A comparative analysis of different pixel and object-based classification algorithms using multi-source high spatial resolution satellite data for LULC mapping. *Earth Science Informatics*, 14, 2231-2247. <https://doi.org/10.1007/s12145-021-00685-4>
- Camacho-Sanabria, J. M., Juan-Pérez, J. I., Pineda-Jaimes, N. B., Cadena-Vargas, E. G., Bravo-Peña, L. C., & Sánchez-López, M. (2015). Cambios de cobertura/uso del suelo en una porción de la Zona de Transición Mexicana de Montaña. *Madera y Bosques*, 21(1), 93-112. <https://doi.org/10.21829/myb.2015.211435>
- Card, D. H. (1982). Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogrammetric Engineering and Remote Sensing*, 48, 431-439. <https://ntrs.nasa.gov/citations/19820041921>
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons. https://archive.org/details/samplingtechniqu0000coch_t4x6/page/438/mode/2up
- Escandón-Calderón, J., Ordóñez-Díaz, J. A. B., Nieto de Pascual-Pola, M. C. del C., & Ordóñez-Díaz, M. de J. (2018). Cambio en la cobertura vegetal y uso del suelo del 2000 al 2009 en Morelos, México. *Revista Mexicana de Ciencias Forestales*, 9(46), 27-51. <https://cienciasforestales.inifap.gob.mx/index.php/forestales/article/view/135>
- Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257-266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Ghatkar, J. G., Singh, R. K., & Shanmugam, P. (2019). Classification of algal bloom species from remote sensing data using an extreme gradient boosted decision tree model. *International Journal of Remote Sensing*, 40(24), 9412-9438. <https://doi.org/10.1080/01431161.2019.1633696>
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295-309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)

- Instituto Nacional de Estadística, Geografía e Informática. (2013). *Continuo de elevaciones mexicano y modelos digitales de elevación* [Base de datos TIFF]. Instituto Nacional de Estadística, Geografía e Informática. <https://www.inegi.org.mx/app/geo2/elevacionesmex/>
- Keskes, M. I., Mohamed, A. H., Borz, S. A., & Niță, M. D. (2025). Improving National Forest Mapping in Romania Using Machine Learning and Sentinel-2 Multispectral Imagery. *Remote Sensing*, 17(4), Article 715. <https://doi.org/10.3390/rs17040715>
- Khan, S., Bhardwaj, A., & Sakthivel, M. (2024). Accuracy assessment of land use land cover classification using machine learning classifiers in Google Earth Engine; A Case Study of Jammu District. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48(4), 263-268. <https://doi.org/10.5194/isprs-archives-XLVIII-4-2024-263-2024>
- Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials*, 2(2), 93-113. [https://doi.org/10.1016/0197-2456\(81\)90001-5](https://doi.org/10.1016/0197-2456(81)90001-5)
- Leija, E. G., Valenzuela-Ceballos, S. I., Valencia-Castro, M., Jiménez-González, G., Castañeda-Gaytán, G., Reyes-Hernández, H., & Mendoza, M. E. (2020). Análisis de cambio en la cobertura vegetal y uso del suelo en la región centro-norte de México. El caso de la cuenca baja del río Nazas. *Ecosistemas*, 29(1), Artículo 1826. <https://doi.org/10.7818/ECOS.1826>
- Mas, J.-F., Pérez-Vega, A., Ghilardi, A., Martínez, S., Loya-Carrillo, J. O., & Vega, E. (2015). Unas herramientas de uso libre para evaluar la fiabilidad temática de datos espaciales. En D. F. Marcolino-Gherardi & L. E. Oliveira e Cruz-de Aragão (Eds.), *XVII Simpósio Brasileiro de Sensoriamento Remoto* (pp. 1020-1026). Ministério da Ciência, Tecnologia e Inovação. https://www.ciga.unam.mx/wrappers/proyectoActual/modelacione/pdf/Mas2015_1020_sbsr.pdf
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2), 153-157. <https://doi.org/10.1007/BF02295996>

- Montenegro, D., & Díaz, M. (2021). Identificación de usos del suelo en los departamentos Simoca y Graneros, provincia de Tucumán, Argentina, mediante imágenes Sentinel 2. Año 2020. *Breves Contribuciones del Instituto de Estudios Geográficos*, (32), 54-72. <https://share.google/C0GBXY7QdmyPbK1kT>
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42-47. <https://doi.org/10.1016/j.rse.2014.02.015>
- Phan, T. N., Kuch, V., & Lehnert, L. W. (2020). Land cover classification using Google Earth Engine and Random Forest Classifier—The role of image composition. *Remote Sensing*, 12(15), Article 2411. <https://doi.org/10.3390/rs12152411>
- Rangel-Lucio, J. A. (2024). Geografía y regionalización. En A. Cruz-Angón, D. López-Higadera, E. D. Melgarejo & E. R. Rodríguez-Ruiz (Coords.), *La biodiversidad en Tamaulipas. Estudio de Estado. Volumen I* (pp. 27-39). Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. <https://bioteca.biodiversidad.gob.mx/janium/Documentos/17035.pdf>
- Rodríguez-González, K. D., Arista-Cázares, L. E., & Yépez-Rincón, F. D. (2024). Spatiotemporal land use land cover (LULC) change analysis of urban narrow river using Google Earth Engine and Machine learning algorithms in Monterrey, Mexico. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10(3), 371-375. <https://doi.org/10.5194/isprs-annals-X-3-2024-371-2024>
- Rodríguez-Rosales, J., González-Camacho, J. M., Macedo-Cruz, A., & Fernández-Ordoñez, Y. M. (2024). Estimation of land cover change using Landsat satellite imagery and the random forest classifier. *Agrociencia*, 2024, Article 2846. <https://doi.org/10.47163/agrociencia.v58i8.2846>
- Rouse Jr., J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. In S. C. Freden, E. P. Mercanti & M. A. Becker (Comps.), *Third Earth Resources Technology Satellite-1 Symposium. Volume I: Technical Presentations Section A* (pp. 309-317). National Aeronautics and Space Administration. <https://ntrs.nasa.gov/citations/19740022614>

- SAS Planet. (2024). SAS Planet: Software for viewing and downloading satellite imagery (Version 241111) [Software]. SAS Planet. <https://www.sasgis.org/>
- Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., & Skakun, S. (2017). Exploring google earth engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping. *Frontiers in Earth Science*, *5*, Article 17. <https://doi.org/10.3389/feart.2017.00017>
- Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308-6325. <https://doi.org/10.1109/JSTARS.2020.3026724>
- Vargas-Contreras, J. A., & Hernández-Huerta, A. (2001). Distribución altitudinal de la mastofauna en la Reserva de la Biosfera "El Cielo", Tamaulipas, México. *Acta Zoológica Mexicana (n. s.)*, *82*, 83-109. <https://www.scielo.org.mx/pdf/azm/n82/n82a5.pdf>
- Yang, L., Driscoll, J., Sarigai, S., Wu, Q., Chen, H., & Lippitt, C. D. (2022). Google Earth Engine and Artificial Intelligence (AI): A comprehensive review. *Remote Sensing*, *14*(14), Article 3253. <https://doi.org/10.3390/rs14143253>
- Zafar, Z., Zubair, M., Zha, Y., Fahd, S., & Nadeem, A. A. (2024). Performance assessment of machine learning algorithms for mapping of land use/land cover using remote sensing data. *The Egyptian Journal of Remote Sensing and Space Sciences*, *27*(2), 216-226. <https://doi.org/10.1016/j.ejrs.2024.03.003>
- Zar, J. H. (2009). *Biostatistical analysis* (5th ed.). Prentice Hall. https://books.google.com.mx/books/about/Biostatistical_Analysis.html?id=LCRFAQAAIAAJ&redir_esc=y
- Zhao, Q., Yu, L., Li, X., Peng, D., Zhang, Y., & Gong, P. (2021). Progress and trends in the application of Google Earth and Google Earth Engine. *Remote Sensing*, *13*(18), Article 3778. <https://doi.org/10.3390/rs13183778>
- Zhao, Q., Yu, L., Du, Z., Peng, D., Hao, P., Zhang, Y., & Gong, P. (2022). An overview of the applications of earth observation satellite data: impacts and future trends. *Remote Sensing*, *14*(8), Article 1863. <https://doi.org/10.3390/rs14081863>

Zhao, Z., Islam, F., Waseem, L. A., Tariq, A., Nawaz, M., Islam, I. U., Bibi, T., Rehman, N. U., Ahmad, W., Aslam, R. W., Raza, D., & Hatamleh, W. A. (2024). Comparison of three machine learning algorithms using google earth engine for land use land cover classification. *Rangeland Ecology & Management*, 92, 129-137. <https://doi.org/10.1016/j.rama.2023.10.007>



Todos los textos publicados por la **Revista Mexicana de Ciencias Forestales** –sin excepción– se distribuyen amparados bajo la licencia *Creative Commons 4.0 Atribución-No Comercial (CC BY-NC 4.0 Internacional)*, que permite a terceros utilizar lo publicado siempre que mencionen la autoría del trabajo y a la primera publicación en esta revista.