



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

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

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

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 *chi*-cuadrada 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 *chi*-cuadrada 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*.

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.

Introducción

El uso de plataformas satelitales como *Sentinel*, *Landsat*, *MODIS*, *Gaofen* y *Worldview* ha crecido considerablemente, consolidándose como herramientas fundamentales para una amplia gama de aplicaciones, entre ellas el análisis de cambios en la cobertura y uso del suelo (CUS) (Zhao et al., 2022). En este contexto, los algoritmos de aprendizaje automático han demostrado ser eficaces para generar mapas de CUS precisos a partir de datos satelitales (Zhao et al., 2024), ya que responden a la necesidad de información confiable y actualizada. Por ejemplo, el Instituto Nacional de Geografía y Estadística (Inegi) proporciona mapas de CUS, pero con una periodicidad limitada, la última versión es del año 2017, con un nivel de detalle intermedio (escala 1:250 000). En cambio, las imágenes *Sentinel* permiten actualizaciones más frecuentes y detalladas, con escalas de hasta 1:50 000 o menores, un tamaño de pixel de 10 m.

En cuanto a la fiabilidad temática, los algoritmos tradicionales como mínima distancia alcanzan fiabilidades globales cercanas a 83 % (Montenegro & Díaz, 2021), mientras que con el método de máxima verosimilitud en imágenes

Landsat se han registrado valores entre 75 % y 90 % (Camacho-Sanabria et al., 2015; Escandón-Calderón et al., 2018).

Entre los métodos de aprendizaje automático, Máquina de Vectores de Soporte (SVM) se mantiene como una de las técnicas más robustas debido a su eficacia para manejar espacios de gran dimensión (Sheykhmousa et al., 2020). Bosques Aleatorios (RF) conserva su relevancia por su sencillez de implementación, dado que requiere un ajuste mínimo de hiperparámetros y por la fácil interpretación de sus datos (Ghatkar et al., 2019). A pesar del avance de los modelos de aprendizaje profundo, SVM y RF continúan siendo ampliamente utilizados en la comunidad científica, con fiabilidades globales superiores a 90 % (Zhao et al., 2024).

En México, los estudios que aplican estos algoritmos son escasos. Por ejemplo, Rodríguez-González et al. (2024) analizaron los cambios en las CUS en el norte del área metropolitana de Monterrey sobre imágenes *Planet* (4.7 de pixel) para el periodo 2016-2019; y concluyeron que RF tuvo el mejor desempeño con 91 %, seguido de SVM con 89 % y Árboles de Clasificación y Regresión (CART) con 83 % de fiabilidad global. De manera similar, Rodríguez-Rosales et al. (2024) evaluaron los cambios de CUS en Huehuetla, Puebla, con imágenes *Landsat* de 2002 y 2021, el método que utilizaron fue RF, y sus precisiones fueron de 92.5 % y 92.3 %, respectivamente. Keskes et al. (2025) mencionan que el algoritmo de Árboles de Gradiente Aumentado (GBT) es el más preciso visualmente para ambientes con alta variabilidad.

La pérdida de biodiversidad y el cambio climático representan los mayores desafíos ambientales actuales, por lo que tener clasificaciones precisas de CUS es esencial para la gestión óptima de los ecosistemas. Una de las plataformas que ha facilitado notablemente la aplicación de algoritmos de aprendizaje automático es *Google Earth Engine* (GEE), porque permite la visualización y el análisis de datos espaciales mediante una Interfaz de Programación de Aplicaciones (API) que brinda acceso a colecciones extensas de imágenes satelitales y posibilita la ejecución de análisis complejos en la nube (Zhao et al., 2021). La integración de algoritmos de aprendizaje automático como SVM o algoritmos de árboles de decisión en GEE, no siempre están disponibles en los *softwares* comerciales o de código abierto; por lo tanto, abren

nuevas oportunidades para la clasificación de imágenes, así como para el monitoreo automatizado en tiempo real de grandes áreas, con altos niveles de precisión a bajo costo (Yang et al., 2022).

La Reserva de la Biosfera El Cielo (RBEC) en Tamaulipas, México es ejemplo de una zona heterogénea, porque alberga una gran diversidad de ecosistemas que incluyen selvas bajas tropicales, bosques de pino y encino, bosque mesófilo de montaña y vegetación de matorrales submontanos. A partir de la clasificación con algoritmos de aprendizaje automático es factible analizar esta complejidad ecológica, útil para estudiar la adaptación de las especies, la interacción entre biomas y factores antrópicos, así como los efectos del gradiente altitudinal en la biodiversidad. Ante esta situación, el presente estudio tuvo como objetivo clasificar los usos del suelo y los tipos de vegetación en la RBEC, además de comparar diversos algoritmos de aprendizaje automático con el fin de evaluar la precisión de las clasificaciones que pueda adaptarse a paisajes heterogéneos. La implementación de algoritmos de aprendizaje automático en *Google Earth Engine* permite generar clasificaciones de CUS con altos niveles de precisión en áreas extensas y ecológicamente complejas.

Materiales y Métodos

Área de estudio

El área de estudio corresponde a la RBEC que fue declarada reserva de la biosfera en 1985 y es considerada como región prioritaria para la conservación de distintos tipos de vegetación; se ubica en el suroeste del estado de Tamaulipas, México (Figura 1), con una extensión de 2 695 km² (Rangel-Lucio, 2024). Los climas predominantes en los distintos tipos de vegetación son: semicálido-subhúmedo con lluvias en verano,

para selvas bajas tropicales; semicálido-húmedo, con abundantes precipitaciones en verano, en el bosque mesófilo de montaña; templado subhúmedo, con lluvias en verano, para el bosque de pino-encino; y semiseco-semicálido con precipitaciones en verano en el matorral submontano (Vargas-Contreras & Hernández-Huerta, 2001). Por otra parte, el área de estudio se localiza entre la zona transicional de las regiones biogeográficas Neártica y Neotropical. La RBEC se divide en dos zonas núcleo centrales que en conjunto cubren 376 km², una zona de amortiguamiento de 1 446 km² y una zona de influencia que comprende toda el área protegida.

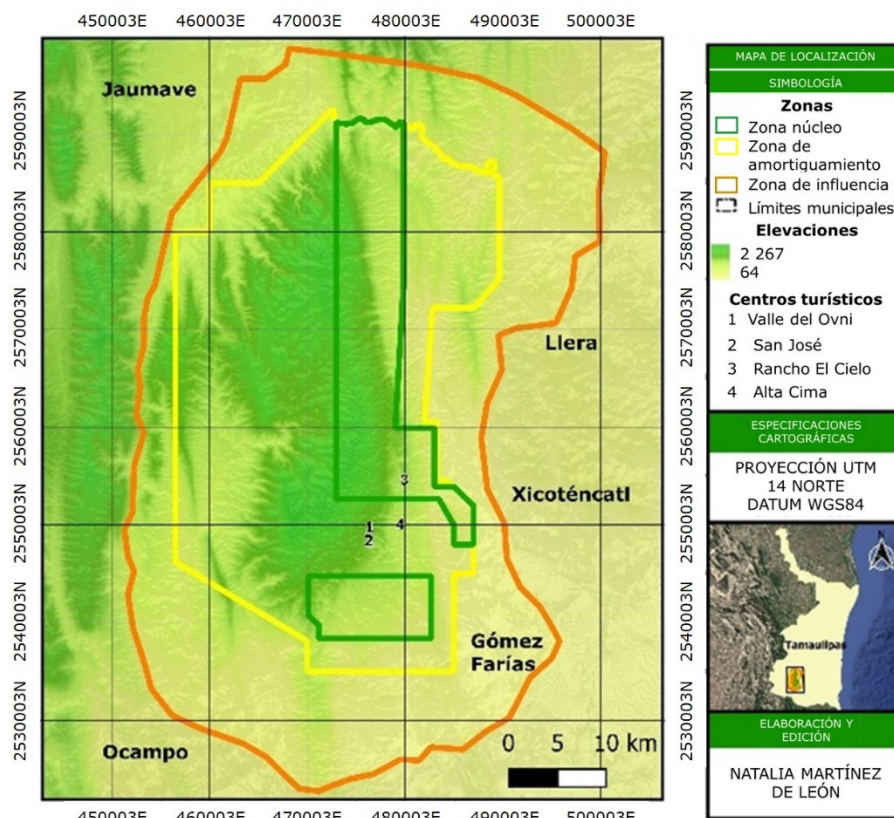


Figura 1. Mapa de la Reserva de la Biosfera El Cielo, Tamaulipas, México.

Imagen de satélite

GEE permite seleccionar imágenes *Sentinel-2* con un nivel de procesamiento 2A que representan la reflectancia superficial. El criterio utilizado para la elección fue aquellas imágenes con menos de 10 % de cobertura de nubes. La imagen final fue un producto de la mediana de imágenes tomadas entre febrero y marzo de 2024, la cual se remuestreó a una resolución de 10 m y reproyectó al sistema de coordenadas EPSG:32614 (*Datum* WGS84, proyección UTM Zona 14 norte).

Variables auxiliares

Se incluyeron variables auxiliares como índices espectrales y topográficas ya que está comprobado que mejoran significativamente la clasificación (Phan et al., 2020). Las alturas y la pendiente del terreno se obtuvieron a partir del Modelo Digital de Elevación (MDE) del Instituto Nacional de Estadística, Geografía e Informática (INEGI, 2013). Los índices espectrales utilizados fueron el Índice de Vegetación de Diferencia Normalizada (NDVI) $\left(\frac{NIR-RED}{NIR+RED}\right)$ (Rouse et al., 1974), el Índice de Vegetación Ajustado al Suelo (SAVI) $\left(\left(\frac{NIR-RED}{NIR+RED+L}\right) \times 1 + L\right)$ (Huete, 1988) y el Índice de Humedad de Diferencia Normalizada (NDMI) $\left(\frac{NIR-SWIR1}{NIR+SWIR1}\right)$ (Gao, 1996).

Métodos

El flujo de trabajo (Figura 2) consistió en la captura de 2 300 campos de entrenamiento distribuidos a lo largo del área de estudio para clasificar la imagen *Sentinel-2A* con los diferentes algoritmos de aprendizaje automático, que representaron ocho clases de CUS: áreas agrícolas, bosque mesófilo de montaña, bosque mixto de pino y encino, selva baja caducifolia, matorral submontano, suelos desnudos, asentamientos humanos y pastizales.

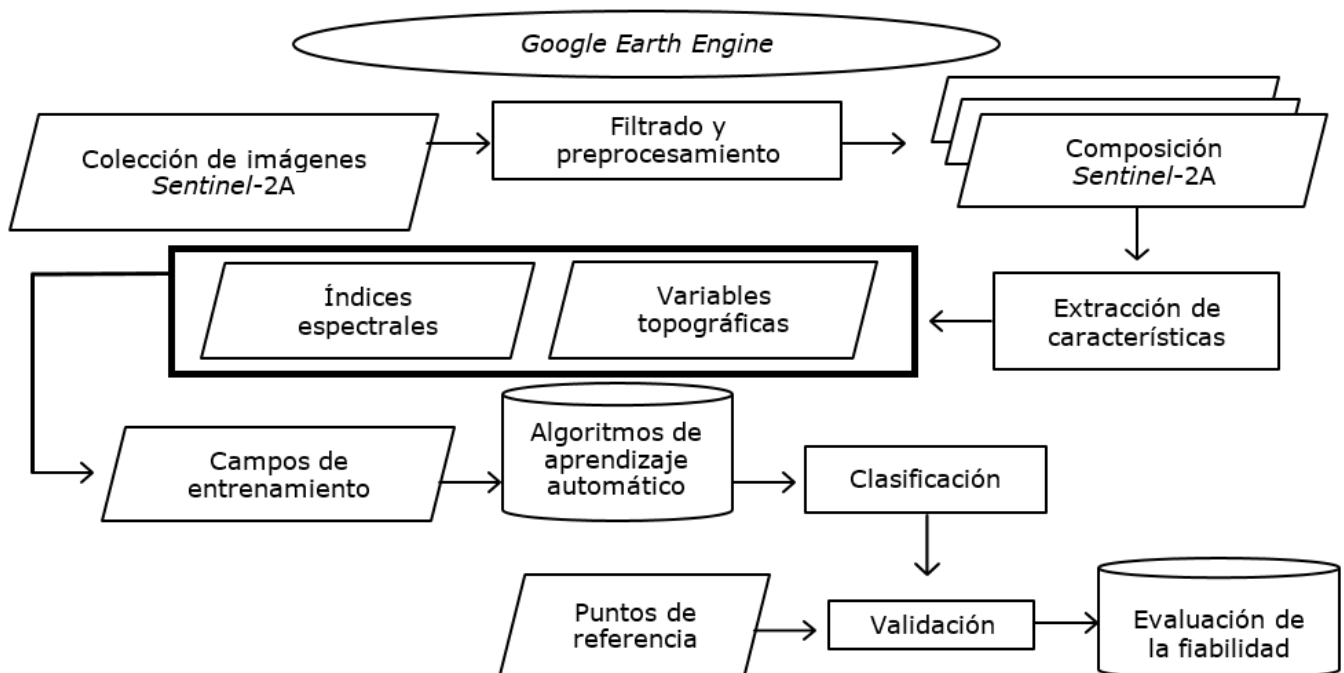


Figura 2. Diagrama de flujo de los métodos aplicados.

Algoritmos de aprendizaje automático

En el Cuadro 1 se muestran los cuatro algoritmos de aprendizaje automático empleados para la clasificación.

Cuadro 1. Descripción de los algoritmos de aprendizaje automático utilizados.

Núm.	Algoritmo de aprendizaje automático	Descripción
1	Máquina de Vectores de Soporte (SVM)	Método supervisado no paramétrico ampliamente utilizado en clasificación y regresión, que busca optimizar el margen de separación entre clases. Destaca por su desempeño en espacios de gran dimensión, su capacidad para modelar relaciones no lineales mediante funciones <i>Kernel</i> y su eficiencia computacional al requerir solo una fracción de los datos de entrenamiento.
2	Árboles de Clasificación y Regresión (CART)	Modelo jerárquico que divide recursivamente el espacio de atributos mediante particiones binarias, generando una estructura interpretable. Es útil tanto en clasificación como en regresión, permite identificar la relevancia de las variables y proporciona modelos de fácil interpretación; aunque su desempeño individual puede verse afectado por la variabilidad de los datos.
3	Bosques Aleatorios (RF)	Algoritmos basados en conjuntos que utilizan múltiples árboles de decisión de manera independiente, cada uno entrenado con subconjuntos aleatorios de características y muestras de datos. Al agregar las predicciones mediante votación o promediado, mitiga el sobreajuste y mejora la robustez del modelo.
4	Árboles de Gradiente Aumentado (GBT)	Algoritmo de ensamble basado en la optimización secuencial, que construye árboles de decisión de manera iterativa para corregir los errores del modelo previo. Es eficaz para capturar relaciones complejas en los datos, mejora progresivamente el rendimiento predictivo y admite variables tanto numéricas como categóricas, manteniendo buena capacidad generalizadora.

Fuente: Khan et al. (2024).

Fiabilidad temática

En este estudio se adoptó la metodología de Olofsson et al. (2014), orientada a evaluar la fiabilidad temática de los mapas clasificados. La fiabilidad temática consiste en comparar los resultados clasificados con datos de referencia, y se usan matrices de error. Según Olofsson et al. (2014): (a) los datos de referencia deben ser de mayor calidad que los datos utilizados para crear el mapa; (b) implementar un diseño probabilístico indicando la unidad de muestreo, tipo de muestreo y tamaño de la muestra; (c) proporcionar una representación espacial y temporal adecuada, a partir de los datos de referencia, para etiquetar cada unidad de la muestra con precisión; (d) resumir la evaluación de la fiabilidad en una matriz de confusión en términos de proporción de área; (e) estimar la fiabilidad global, indicar los errores de omisión y comisión; y (f) cuantificar la incertidumbre de los índices de fiabilidad mediante intervalos de confianza.

El dato de referencia fue un mosaico de alta resolución de *Google Earth* descargado de la aplicación *SAS Planet* versión 241111 (SAS Planet, 2024), donde se realizó una interpretación visual asignando la etiqueta de referencia a cada punto de muestreo. La unidad de muestreo fue de un pixel por muestra. El muestreo aleatorio estratificado se emplea con frecuencia en investigaciones de fiabilidad temática (Leija et al., 2020; Mas et al., 2015), para lo cual Cochran (1977) y Olofsson et al. (2014) proponen la Ecuación 1 para el tamaño de muestra:

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

Donde:

$s(\hat{O})$ = Error estándar de la precisión general estimada que se pretende alcanzar

W_i = Proporción de área mapeada de la clase i

S_i = Desviación estándar del estrato i , $S_i = \sqrt{U_i(1 - U_i)}$ (Cochran, 1977)

Se especificó un error estándar objetivo para la precisión general de 0.01.

En el área de estudio, las coberturas de bosque mixto y selva baja caducifolia predominan en la superficie total, las otras clases se consideran poco representadas por su superficie; por lo tanto, se empleó un enfoque de asignación de muestra mixto para conciliar la necesidad de precisión en la estimación global y de área, con la exigencia de obtener estimaciones precisas para las clases poco representadas (Olofsson et al., 2014). Una asignación puramente proporcional al área resultaría en un número insuficiente de muestras para las clases con poca superficie, lo que generaría errores estándar altos para su precisión del usuario ($\hat{U}_i$).

Primero se asignó un número fijo de muestras (entre 50 y 100) a las clases con poca superficie (Olofsson et al., 2014). Este intervalo de muestras se basó en el análisis de la varianza estimada de la precisión del usuario, con el objetivo de asegurar un tamaño de muestra suficiente para garantizar errores aceptables en la estimación de la $\hat{U}_i$ de esas categorías.

La información en el Cuadro 2 incluye las proporciones de área mapeada (W_i), los valores estimados de la precisión del usuario (U_i) y las desviaciones estándar (S_i) de cada clase.

Cuadro 2. Información necesaria para la asignación del tamaño de muestra a las clases.

Clase	W_i	U_i	S_i	Alloc1
AA	0.100	0.480	0.500	50
BMM	0.005	0.903	0.296	50
BM	0.411	0.929	0.258	172
SBC	0.405	0.973	0.163	169
MS	0.035	0.771	0.420	50
SD	0.015	0.718	0.450	50
AH	0.004	0.921	0.270	50
PZ	0.024	0.909	0.287	50

AA = Áreas agrícolas; BMM = Bosque mesófilo de montaña; BM = Bosque mixto, SBC = Selva baja caducifolia; MS = Matorral submontano; SD = Suelo desnudo; AH = Asentamiento humano; PZ = Pastizales. W_i = Proporción de área mapeada; U_i = Precisión del usuario; S_i = Desviación estándar de cada clase.

Se escogió “Alloc1” con 50 unidades de muestreo para seis clases con poca superficie. La precisión se evaluó mediante una matriz de confusión, que consiste en una tabulación cruzada de la clasificación del mapa (*Sentinel-2*) frente a los datos de referencia (*SAS Planet*). GEE permite calcular las matrices de confusión para cada algoritmo con la función ‘*ee.errorMatrix*’. Card (1982) propone un procedimiento para registrar los valores en términos de proporción de área estimada, lo cual es esencial para que las estimaciones sean insesgadas bajo el muestreo aleatorio estratificado. Se estimó la proporción de área mapeada para cada clase con la Ecuación 2 (Olofsson *et al.*, 2014):

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

Donde:

W_i = Proporción del área mapeada para la clase i

n_{ij} = Número de píxeles que fueron clasificados como clase i

$n_{i.}$ = Total de muestras de validación que fueron clasificadas como clase i

La fiabilidad global ($\hat{O}$) representa la proporción de área correctamente clasificada y se calcula sumando los elementos diagonales de la matriz ajustada (Ecuación 3) (Olofsson *et al.*, 2014).

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

Donde:

$\hat{P}_{jj}$ = Proporción estimada del área que fue correctamente clasificada como clase j

q = Número total de clases en la matriz de confusión

La precisión del usuario estimada ($\hat{U}_i$) indica la proporción de las áreas clasificadas como i en el mapa que realmente corresponden a la clase i (Ecuación 4) (Olofsson et al., 2014).

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

Donde:

$\hat{P}_{ii}$ = Proporción del área correctamente clasificada como clase i (valor de la diagonal en la matriz de error ajustada)

$\hat{P}_{i.}$ = Suma de todas las proporciones de área clasificadas como clase i

La precisión del productor estimada ($\hat{P}_j$) mide la exactitud del algoritmo desde la perspectiva del productor del mapa y se calcula con la Ecuación 5 (Olofsson et al., 2014).

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

Donde:

$\hat{P}_j$ = Proporción estimada del área de la clase j

$\hat{P}_{jj}$ = Proporción del área correctamente clasificada como clase j (valor en la diagonal de la matriz de error ajustada)

$\hat{P}_{.j}$ = Suma de todas las proporciones de área que en los datos de referencia corresponde a la clase j

La varianza de la precisión global del mapa $\hat{V}(\hat{O})$, de la precisión del usuario para la clase i $\hat{V}(\hat{U}_i)$, y de la precisión del productor para la clase j $\hat{V}(\hat{P}_j)$, se estiman utilizando las ecuaciones 6, 7 y 8, respectivamente (Olofsson *et al.*, 2014):

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

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

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

Donde:

$\hat{N}_{.j} = \sum_{i=1}^q \frac{N_{i.}}{n_{i.}} n_{ij}$ = Número total marginal estimado de píxeles de la clase de referencia j

$N_{j.}$ = Área total mapeada como clase j

$N_{i.}$ = Área total mapeada como clase i

$n_{j.}$ = Total de muestras de referencia como clase j

Los intervalos de confianza a 95 % se estiman como $\pm 1.96 \sqrt{\hat{V}(\hat{U}_i)}$ (se reemplaza $\hat{U}_i$ con $\hat{P}_j$ y $\hat{O}$ para la precisión del productor y la fiabilidad global, respectivamente).

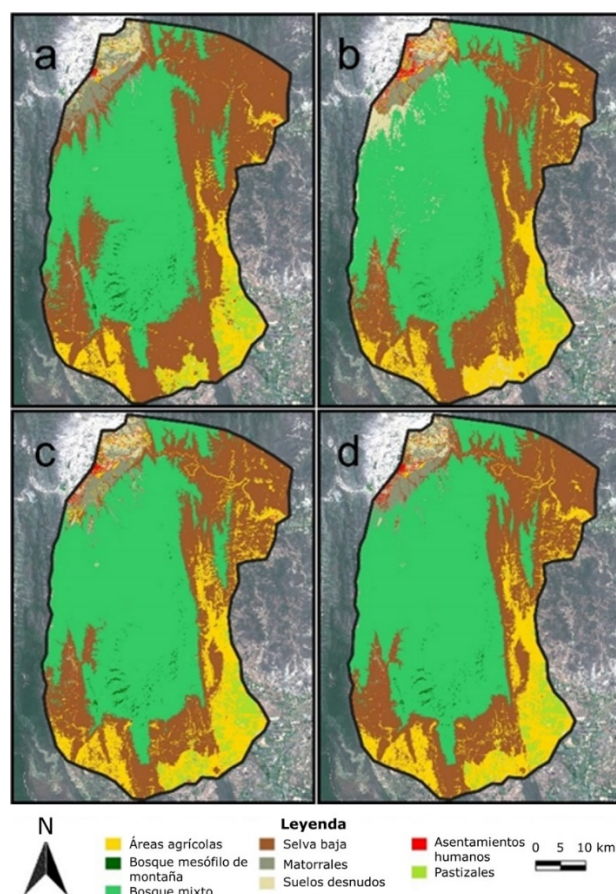
Pruebas estadísticas de comparación entre algoritmos

El Coeficiente *Kappa* es una medida ampliamente empleada para evaluar la fiabilidad temática en los mapas al considerar el acuerdo atribuible al azar. Sin embargo, no es recomendable su aplicación para comparar dos algoritmos (Balha et al., 2021); dado que su validez depende de la independencia entre evaluadores, el uso de las mismas muestras de validación invalida este supuesto. En estos casos, la prueba de *McNemar* resulta más adecuada, pues está diseñada para evaluar diferencias entre algoritmos (Zar, 2009).

Se aplicó la prueba *chi-cuadrada* (χ^2) de *McNemar* (McNemar, 1947), la cual permite evaluar la homogeneidad marginal entre las predicciones. Esa homogeneidad hace referencia a la igualdad en las distribuciones marginales de las clasificaciones realizadas por dos algoritmos distintos; es decir, a la ausencia de diferencias estadísticamente significativas entre ellos. Esta prueba se considera una herramienta eficiente para realizar comparaciones clase por clase, debido a que es un procedimiento paramétrico, con bajo riesgo de error tipo I y una formulación sencilla (Abdi, 2020). Además, se utilizó una prueba *Z* para dos proporciones (Lachin, 1981) para comparar las Proporciones de Píxeles Correctamente Clasificados (PCCP) entre dos algoritmos a la vez. El cuadrado del estadístico *Z* generado por la prueba sigue una distribución χ^2 con un grado de libertad (Abdi, 2020).

Resultados y Discusión

El resultado de aplicar los cuatro algoritmos de aprendizaje automático generó los mapas de clasificación que se muestran en la Figura 3.



a = Árboles de Clasificación y Regresión (CART); b = Máquina de Vectores de Soporte (SVM); c = Árboles Aleatorios (RF); d = Árboles de Gradiente Aumentado (GBT).

Figura 3. Mapas clasificados de Reserva de la Biosfera El Cielo, Tamaulipas, México.

Evaluación de la fiabilidad temática de los algoritmos

Los resultados de la fiabilidad global de los algoritmos utilizados son las siguientes: 88 % para SVM, 86 % para GBT, 84 % para RF y 82 % para CART. Posteriormente, se examinaron las métricas de precisión del usuario (*UA*) y precisión del productor (*PA*) para cada una de las ocho clases de CUS, lo cual permitió determinar las fortalezas y limitaciones de cada algoritmo por clase. Estudios previos han destacado a RF como uno de los algoritmos más precisos en contextos urbanos y homogéneos, con fiabilidades superiores a 95 % (Zafar et al., 2024; Zhao et al., 2024). Sin embargo, en este estudio desarrollado en un paisaje heterogéneo como la RBEC, el algoritmo SVM mostró un mejor desempeño que coincide con los hallazgos de Abdi (2020), quien obtuvo una precisión de 76 % con SVM, seguido de una precisión de 74 % con RF.

Aunque CART se ha utilizado con éxito para la clasificación de áreas agrícolas (Shelestov et al., 2017), en el presente trabajo tuvo menor precisión, lo que podría atribuirse a su sensibilidad al sobreajuste en ambientes con alta diversidad de clases. En contraste, SVM y GBT alcanzaron las mayores fiabilidades globales (89 % y 86 %, respectivamente), además destacaron por su capacidad para distinguir clases con alta complejidad espectral. GBT sobresalió en la identificación de áreas agrícolas (*PA*=95 %), mientras que SVM mostró mayor precisión del usuario (*UA*=45 %), con un menor número de errores de comisión (Cuadro 3). Para el bosque mesófilo de montaña, todos los algoritmos tuvieron baja precisión del productor (*PA*=13-19 %), aunque SVM obtuvo la *UA* más alta (97 %), lo que indica menor confusión con otras clases.

Cuadro 3. Evaluación de la fiabilidad temática de la cobertura y el uso del suelo, se muestra la proporción del área clasificada (W_i).

Clase	SVM			GBT			RF			CART		
	W_i	UA	PA	W_i	UA	PA	W_i	UA	PA	W_i	UA	PA
AA	0.10	0.45 ± 0.10	0.90 ± 0.06	0.14	0.36 ± 0.09	0.95 ± 0.06	0.15	0.36 ± 0.09	0.80 ± 0.07	0.14	0.37 ± 0.09	0.78 ± 0.06
BMM	0.01	0.97 ± 0.06	0.19 ± 0.06	0.01	0.93 ± 0.10	0.17 ± 0.08	0.01	0.96 ± 0.07	0.13 ± 0.07	0.01	0.88 ± 0.11	0.19 ± 0.09
BM	0.41	0.95 ± 0.03	1.00 ± 0.02	0.48	0.92 ± 0.04	1.00 ± 0.02	0.47	0.93 ± 0.04	1.00 ± 0.02	0.51	0.89 ± 0.05	1.00 ± 0.03
SBC	0.41	0.96 ± 0.04	0.98 ± 0.06	0.30	0.97 ± 0.03	0.98 ± 0.11	0.30	0.96 ± 0.04	0.92 ± 0.10	0.28	0.95 ± 0.04	0.85 ± 0.11
MS	0.04	0.68 ± 0.13	0.76 ± 0.14	0.03	0.91 ± 0.12	0.83 ± 0.36	0.03	0.90 ± 0.13	0.74 ± 0.41	0.01	0.88 ± 0.17	0.11 ± 0.12
SD	0.02	0.73 ± 0.11	0.33 ± 0.09	0.01	0.77 ± 0.09	0.41 ± 0.07	0.01	0.74 ± 0.10	0.42 ± 0.08	0.03	0.57 ± 0.10	0.73 ± 0.08
AH	0.01	0.92 ± 0.07	0.19 ± 0.06	0.07	0.93 ± 0.07	0.16 ± 0.06	0.01	0.92 ± 0.07	0.28 ± 0.06	0.01	0.75 ± 0.12	0.17 ± 0.10
PZ	0.02	0.91 ± 0.07	0.40 ± 0.09	0.03	0.94 ± 0.06	0.44 ± 0.09	0.03	0.94 ± 0.06	0.44 ± 0.09	0.02	0.98 ± 0.04	0.30 ± 0.08

AA = Áreas agrícolas; BMM = Bosque mesófilo de montaña; BM = Bosque mixto; SBC = Selva baja caducifolio; MS = Matorral submontano; SD = Suelo desnudo; AH = Asentamientos humanos; PZ = Pastizales. W_i = Proporción de área mapeada; UA = precisión del usuario; PA = precisión del productor. SVM = Máquina de Vectores de Soporte; GBT = Árboles de Gradiente Aumentado; RF = Bosques Aleatorios; CART = Árboles de Clasificación y Regresión.

En la clasificación de matorrales submontanos, los árboles de decisión superaron a SVM en precisión del usuario (>88 %), aunque CART presentó baja precisión del productor. Para suelos desnudos, GBT y CART evidenciaron desempeños complementarios, que sugieren una posible combinación futura. Finalmente, para asentamientos humanos y pastizales, todos los algoritmos alcanzaron UA altas (>90 %), pero con valores bajos de PA, lo que indica errores de omisión.

Comparación entre algoritmos

Los algoritmos se analizaron bajo dos perspectivas, la comparación detallada entre algoritmos a nivel clase por clase que se basó en la prueba *chi*-cuadrada de *McNemar* y la comparación de la precisión de la clasificación a nivel global basada en la prueba *Z* de dos proporciones.

Los resultados de la prueba *chi*-cuadrada de *McNemar* (Cuadro 4) mostraron que 31 % de las comparaciones por clase entre algoritmos pareados presentaron diferencias estadísticamente significativas ($P \leq 0.01$). Adicionalmente, 10 % de las comparaciones registraron diferencias marginales ($0.05 < p \leq 0.10$). En contraste, 44 % de las comparaciones no tuvieron diferencias estadísticamente significativas ($p > 0.10$).

Cuadro 4. Prueba *chi*-cuadrada (χ^2) de *McNemar* con su valor de probabilidad asociado (P).

Algoritmo/ Clase		AA	BMM	BM	SBC	MS	SD	AH	PZ
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 = No Significativo. AA = Áreas agrícolas; BMM = Bosque mesófilo de montaña; BM = Bosque mixto; SB = Selva

baja caducifolia; MS = Matorral submontano; SD = Suelo desnudo; AH = Asentamientos humanos; PZ = Pastizales. SVM = Máquina de Vectores de Soporte; GBT = Árboles de Gradiente Aumentado; RF = Bosques Aleatorios; CART = Árboles de Clasificación y Regresión.

Las menores diferencias se observaron entre RF y GBT, en los cuales se observaron desempeños estadísticamente similares en la mayoría de las clases analizadas. En particular, se identificaron comparaciones no significativas entre los pares RF vs. GBT, RF vs. CART y GBT vs. CART en las clases de áreas agrícolas, bosque mixto y selva baja caducifolia. Esto sugiere que esos algoritmos presentan desempeños similares en ciertas coberturas, particularmente en aquellas con patrones espectrales más definidos, lo que podría aprovecharse en futuras aplicaciones para optimizar recursos computacionales, sin comprometer la precisión.

Entre los casos significativos de la prueba de *McNemar* destacaron las comparaciones con SVM, en las que se obtuvo el mayor número de diferencias significativas, especialmente frente a CART y GBT, las clases de selva baja caducifolia, matorral submontano y suelo desnudo ($P \leq 0.05$); ello sugiere que ambos algoritmos tienen comportamientos similares en la mayoría de las clases. Las diferencias más frecuentes se presentaron en coberturas con alta complejidad espectral, como matorral submontano y suelo desnudo. Las comparaciones entre GBT y CART también evidenciaron diferencias al nivel de 1 % para suelos desnudos y pastizales.

Los resultados del presente estudio coinciden parcialmente con Abdi (2020), quien cita diferencias significativas ($P \leq 0.05$) en 62 % de las predicciones por clase al comparar SVM, RF y *Xgboost*. En ambos trabajos, se observan discrepancias relevantes entre SVM y RF para las clases de selva baja caducifolia y suelos desnudos ($P \leq 0.01$). No obstante, se identificaron diferencias en áreas agrícolas, no registradas por Abdi (2020), y ausencia de discrepancia en asentamientos humanos, donde dicho autor sí las obtuvo. Esto sugiere que, a pesar de trabajar con la misma colección satelital, la composición del paisaje y las características espectrales locales influyen en la sensibilidad de cada algoritmo para detectar ciertas clases.

Los resultados de la prueba Z (Cuadro 5) para dos proporciones indicaron diferencias estadísticamente significativas en las PCCP entre los algoritmos SVM y CART ($X^2=3.38$), entre RF y CART ($X^2=2.73$), y entre GBT y CART ($X^2=2.65$). Por otro lado, no se observaron diferencias significativas entre SVM y RF, SVM y GBT, ni entre RF y GBT. Esto indica que CART tiene un desempeño significativamente distinto en la clasificación de píxeles, mientras que SVM, RF y GBT muestran un comportamiento estadísticamente similar en precisión.

Cuadro 5. Prueba Z de dos proporciones.

Par de algoritmos	Prueba Z de dos proporciones	Valor P
SVM vs. RF	$X^2=0.66$	0.5100
SVM vs. GBT	$X^2=0.73$	0.4600
SVM vs. CART	$X^2=3.38$	0.0007
RF vs. GBT	$X^2=0.07$	0.9400
RF vs. CART	$X^2=2.73$	0.0064
GBT vs. CART	$X^2=2.65$	0.0080

La prueba Z de dos proporciones se aplicó de forma bilateral para comparar la proporción correcta de clasificación de píxeles (PCCP) entre pares de algoritmos. El estadístico Z se reporta como su equivalente X^2 con un grado de libertad. Se adoptó un nivel de significancia de $\alpha=0.05$; valores de $P \leq 0.05$ indican diferencias estadísticamente significativas. SVM = Máquina de Vectores de Soporte; GBT = Árboles de Gradiente Aumentado; RF = Bosques Aleatorios; CART = Árboles de Clasificación y Regresión.

Conclusiones

Los algoritmos SVM y GBT demostraron ser en general los más efectivos para el mapeo de coberturas y usos del suelo en la RBEC, con fiabilidades globales de 88 % y 86 %, respectivamente. SVM destacó en la clasificación de selva baja caducifolia y bosque mixto, mientras que GBT mostró un rendimiento equilibrado en múltiples clases. Por otro lado, RF y CART resultaron útiles para clases específicas como pastizales y suelos desnudos. Sin embargo, todos los algoritmos tienen limitaciones en clases con baja extensión espacial, como el bosque mesófilo de montaña, donde se registran altos errores de omisión. Estos hallazgos cumplen con el objetivo de comparar algoritmos de aprendizaje automático, al identificar a SVM y GBT como los más precisos. Asimismo, la evaluación de la fiabilidad temática mediante matrices de confusión permitió determinar las fortalezas y limitaciones de cada algoritmo por clase.

Agradecimientos

El primer autor agradece a la Secretaría de Ciencias, Humanidades, Tecnología e Innovación (Secihti) por la beca de maestría otorgada, así como al posgrado de la Facultad de Ingeniería y Ciencias de la Universidad Autónoma de Tamaulipas para llevar a cabo la investigación.

Conflicto de intereses

Los autores manifiestan no tener conflicto de intereses.

Contribución por autor

Natalia Martínez de León: búsqueda de información, análisis de los datos y redacción del manuscrito; Ignacio González Gutiérrez: análisis de los datos y revisión del manuscrito; X. Celeste Ramírez Campanur: búsqueda de información y revisión del manuscrito; Mario Rocandio Rodríguez: análisis estadístico y revisión del manuscrito; Arturo Medina Puente: revisión del manuscrito.

Referencias

- 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.