Ciencia e Ingeniería Neogranadina
ISSN: 0124-8170
ISSN: 1909-7735
Universidad Militar Nueva Granada

Garzón Barrero, Julián; Sánchez Pineda, Nancy Estela; Londoño Pinilla, Darío Fernando

Evaluación comparativa de los algoritmos de aprendizaje automático Support Vector Machine y Random Forest: efectos del tamaño del conjunto de entrenamiento*

Ciencia e Ingeniería Neogranadina, vol. 33, núm. 2, 2023, Julio-Diciembre, pp. 131-148
Universidad Militar Nueva Granada

DOI: https://doi.org/10.18359/rcin.6996

Disponible en: https://www.redalyc.org/articulo.oa?id=91178599010

Ciencia e Ingeniería Neogranadina
2023
Vol. 33(2)
julio-diciembre de 2023 ■ ISSN: 0124-8170 · ISSN-e: 1909-7735 ■ pp. 131-148

Editorial Neogranadina

DOI: https://doi.org/10.18359/rcin.6996

Evaluación comparativa de los algoritmos de aprendizaje automático Support Vector Machine y Random Forest: efectos del tamaño del conjunto de entrenamiento*

Julián Garzón Barrero[a] ■ Nancy Estela Sánchez Pineda[b] ■ Darío Fernando Londoño Pinilla[c]

Resumen: En el presente estudio se examinó el rendimiento de los algoritmos Support Vector Machine (SVM) y Random Forest (RF) utilizando un modelo de segmentación de imágenes basado en objetos (OBIA) en la zona metropolitana de Barranquilla, Colombia. El propósito fue investigar de qué manera los cambios en el tamaño de los conjuntos de entrenamiento y el desequilibrio en las clases de cobertura terrestre influyen en la precisión de los modelos clasificadores. Los valores del coeficiente Kappa y la precisión general revelaron que SVM superó consistentemente a RF. Además, la imposibilidad de calibrar ciertos parámetros de SVM en ArcGIS Pro planteó desafíos. La elección del número de árboles en RF mostró ser fundamental, con un número limitado de árboles (50) que afectó la adaptabilidad del modelo, especialmente en conjuntos de datos desequilibrados. Este estudio resalta la complejidad de elegir y configurar modelos de aprendizaje automático, que acentúan la importancia de considerar cuidadosamente las proporciones de clases y la homogeneidad en las distribuciones de datos para lograr predicciones precisas en la clasificación de uso del suelo y cobertura terrestre. Según los hallazgos, alcanzar precisiones de usuario superiores al 90 % en las clases de pastos limpios, bosques, red vial y agua continental, mediante el modelo SVM en ArcGIS Pro, requiere asignar muestras de entrenamiento que cubran respectivamente el 2 %, 1 %, 3 % y 8 % del área clasificada.

Palabras clave: Machine Learning (ML); Object-Based Image Analysis (OBIA); Support Vector Machine (SVM); Random Trees (RT); muestras de entrenamiento; clasificación de imágenes satelitales; ingeniería geomática; teledetección

* Artículo de investigación.

a Ph. D. en Ingeniería Geomática, magíster en Sistemas de Información Geográfica, especialista en Geomática. Universidad del Quindío, Programa de Ingeniería Topográfica y Geomática, Armenia, Colombia. Correo electrónico: juliangarzonb@uniquindio.edu.co ORCID: http://orcid.org/0000-0002-4871-3726

b Magíster en Ingeniería Hidráulica y Medio Ambiente, ingeniera civil. Universidad del Quindío, Programa de Ingeniería Topográfica y Geomática, Armenia, Colombia.
Correo electrónico: nesachez@uniquindio.edu.co ORCID: http://orcid.org/0009-0008-4259-9505

c Magíster en Ingeniería énfasis en Geomática. Licenciado en Matemáticas. Universidad del Quindío, Programa de Ingeniería Topográfica y Geomática, Armenia, Colombia.
Correo electrónico: dflondono@uniquindio.edu.co ORCID: http://orcid.org/0000-0002-6130-8071

Recibido: 17/10/2023 **Aceptado:** 05/12/2023 **Disponible en línea:** 27/12/2023

Cómo citar: J. Garzón Barrero, N. E. Sánchez Pineda, y D. F. Londoño Pinilla, «Evaluación comparativa de los algoritmos de aprendizaje automático Support Vector Machine y Random Forest: efectos del tamaño del conjunto de entrenamiento», Cien.Ing.Neogranadina, vol. 33, n.º 2, pp. 131–148. Diciembre 2023.

Comparative Evaluation of Support Vector Machine and Random Forest Machine Learning Algorithms: Effects of Training Set Size

Abstract: This study examined the performance of Support Vector Machine (SVM) and Random Forest (RF) algorithms using an Object-Based Image Analysis (OBIA) model in the metropolitan area of Barranquilla, Colombia. The purpose was to investigate how changes in training set size and imbalance in land cover classes influence the accuracy of classifier models. Kappa coefficient values and overall accuracy consistently revealed that SVM outperformed RF. Additionally, the inability to calibrate certain SVM parameters in ARCGIS Pro posed challenges. The choice of the number of trees in RF proved to be crucial, with a limited number of trees (50) affecting the model's adaptability, especially in imbalanced datasets. This study highlights the complexity of choosing and configuring machine learning models, emphasizing the importance of carefully considering class proportions and homogeneity in data distributions to achieve accurate predictions in land use and land cover classification. According to the findings, achieving user accuracies exceeding 90% in clean grass, forests, road networks, and continental water classes, using the SVM model in ARCGIS Pro, requires assigning training samples covering 2%, 1%, 3%, and 8% of the classified area, respectively.

Keywords: Machine Learning (ML); Object-Based Image Analysis (OBIA); Support Vector Machine (SVM); Random Trees (RT); Training Samples; Satellite Image Classification; Geomatic Engineering; Remote Sensing

Introducción

Existe una creciente demanda de mapas de uso y cobertura del suelo (LULC, por sus siglas en inglés) derivados de la investigación sobre el impacto del cambio climático, aplicaciones hidroecológicas y las intervenciones humanas en los ecosistemas [1]-[3]. El uso de la tierra (LU) se refiere a las acciones, que tienen lugar en la superficie, realizadas por los humanos como la urbanización o la industria; sin embargo, la cobertura terrestre (LC) se refiere a la descripción física de la superficie de la tierra como los bosques o cuerpos de agua [4]. Ambos conceptos son esenciales para comprender la dinámica del paisaje y su evolución temporal. Los mapas LULC desempeñan un papel fundamental en las políticas de planificación urbana, conservación y el monitoreo agrícola [5], [6].

Las imágenes capturadas por satelitales han adquirido un rol interesante para mapear el territorio debido a su amplia cobertura y eficiencia en términos de costos. El satélite Landsat-9 (L9) comenzó a operar desde finales del 2021, proporcionando imágenes multiespectrales de resolución moderada. Está equipado con un generador operativo de imágenes terrestres (OLI-2) y un sensor térmico infrarrojo (TIRS-2), que producen once bandas espectrales y resolución máxima de 15 m, con frecuencia periódica de 16 días [7]. Estas imágenes han ganado una considerable atención en la investigación gracias a su acceso libre y cobertura global [8].

La clasificación de imágenes en teledetección representa el método más común para identificar tipos de cobertura terrestre. En la última década, los algoritmos de aprendizaje automático (ML) se han utilizado con efectividad para la producción de mapas LULC debido a su capacidad de aumento de calidad, eficiencia y escalabilidad en comparación con los modelos tradicionales basados en discriminación de píxeles [9]. Estos algoritmos se pueden categorizar en paramétricos y no paramétricos según si es necesario o no hacer suposiciones específicas sobre la distribución de sus datos [10]. Los clasificadores paramétricos, como la regresión logística y el Naive Bayes (NB), son recomendados cuando se establecen relaciones lógicas claras y los datos se ajustan a los supuestos paramétricos subyacentes. Por otro lado, los clasificadores no paramétricos pueden adaptarse a relaciones no lineales complejas, que los hacen más adecuados para la clasificación de coberturas terrestres [11]. Según diversos estudios, los clasificadores no paramétricos más empleados en LULC son Random Forest (RF), k-Nearest Neighbor (KNN), Support Vector Machines (SVM) y Classification And Regression Trees (CART) [12]-[14]. De acuerdo con Ouma *et al.* [15], al comparar algoritmos como CART, RF, Gradient Tree Boosting (GTB) y SVM, la precisión del clasificador depende de la clase mapeada. El suelo desnudo se representa mejor utilizando RF y CART con una precisión del 98 % mientras que SVM y GTB fueron los más adecuados para cuerpos de agua. Los clasificadores de mejor rendimiento paras las cubiertas vegetales fueron RF, SVM y GTB. Por su parte, Deng *et al.* [16] señalan que la precisión de la clasificación de LULC está fuertemente influenciada por características del sensor y factores relacionados con los datos de la imagen, como la resolución espacial y temporal, además del *software* y *hardware* de procesamiento.

Los modelos de clasificación automática se basan en la obtención de firmas espectrales de clases predefinidas mediante datos de entrenamiento. Estos modelos se centran en diferenciar píxeles que corresponden a diferentes tipos de cobertura [17]. El análisis de imágenes basado en objetos (OBIA) ha superado a los métodos que se basan en la discriminación de píxeles al ofrecer una representación más precisa de la distribución de la cobertura del suelo [11].

La segmentación constituye el fundamento del modelo OBIA, la cual transforma datos complejos en unidades significativas. Los objetos segmentados proporcionan una representación más cercana a cómo los seres humanos perciben el mundo real. La segmentación implica dividir las imágenes en conjuntos de objetos espacialmente contiguos, donde cada uno está formado por un grupo de píxeles vecinos con homogeneidad o significado semántico [18].

Para lograr una buena precisión en la clasificación LULC, es esencial contar con conjuntos de datos de entrenamiento suficientemente amplios.

Sin embargo, surge un problema debido a que las diversas coberturas ocupan proporciones de área diferentes, lo que significa que algunos de estos conjuntos de datos son adecuados en tamaño mientras que otros resultan ser limitados [19].

Se han llevado a cabo varios estudios con el objetivo de encontrar el algoritmo de clasificación ideal para la creación de mapas LULC. Estos estudios implican comparar el rendimiento del algoritmo consigo mismo y en relación con otros métodos de clasificación; sin embargo, las conclusiones varían considerablemente [20], [21]. Yuh *et al*. [22] sostienen que los algoritmos ML se pueden entrenar utilizando conjuntos de datos equilibrados (con el mismo número de píxeles muestreados por clase) y desequilibrados (con diferente número de píxeles muestreados para cada clase) sin grandes incertidumbres de clasificación. Por otro lado, Azadbakht *et al*. [23] han indicado que los conjuntos de datos desequilibrados pueden plantear desafíos para los algoritmos de ML, especialmente en la categorización de las clases menos frecuentes. Esto sugiere que los conjuntos de datos desequilibrados todavía presentan desafíos significativos que deben abordarse adecuadamente en el entrenamiento de algoritmos ML. Existe poca investigación que compare RF y SVM en el contexto de imágenes Landsat-9, especialmente en Colombia. Si bien el diseño de muestreo está bien documentado en la literatura, quedan dudas sobre el número y tamaño de muestras requeridas, su calidad y el desequilibrio de clases.

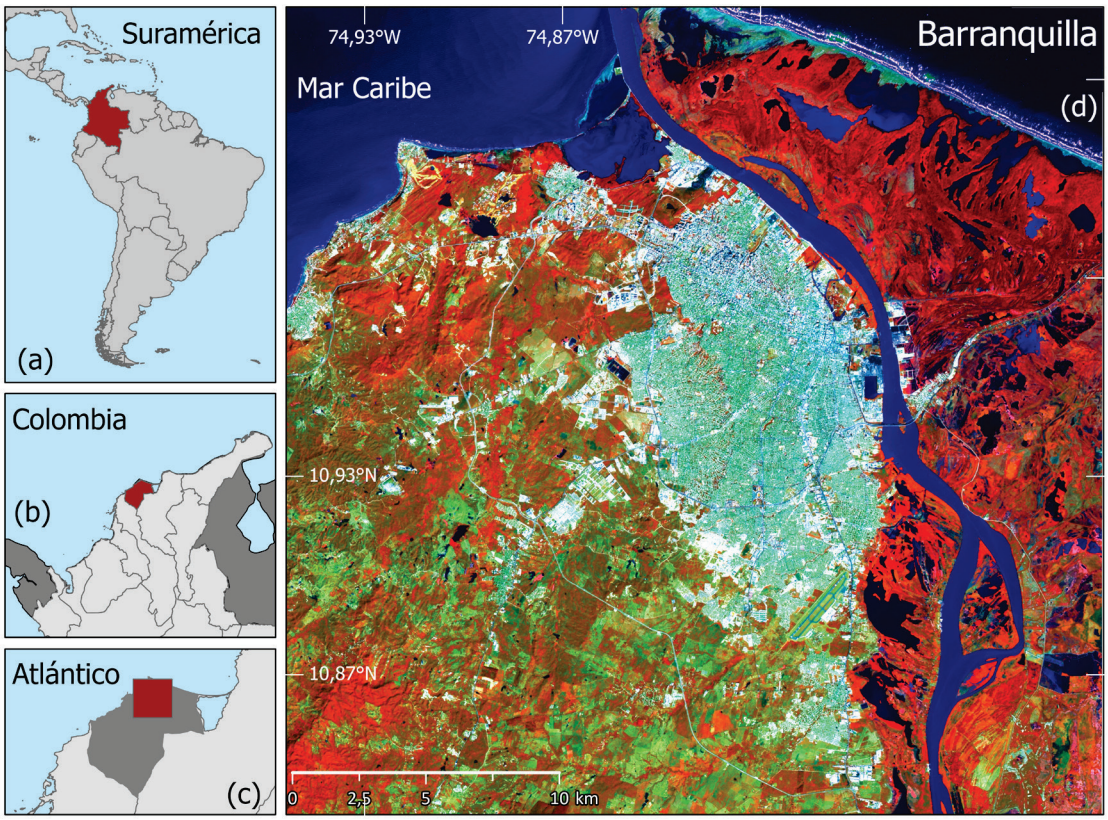
Este estudio se centra en la evaluación de dos algoritmos ampliamente utilizados en la clasificación LULC: Support Vector Machine (SVM) y Random Forest (RF) (referido como Random Trees (RT) en ArcGIS Pro). Estos algoritmos se seleccionaron debido a su capacidad para clasificar imágenes con robustez y abordar desafíos como el ruido y el sobreajuste. El objetivo de esta investigación es analizar cómo las fluctuaciones en la dimensión del grupo de entrenamiento (equilibrados y desequilibrados) influyen en la eficiencia de los algoritmos de aprendizaje automático supervisado bajo un modelo de segmentación OBIA. Este enfoque destaca la relevancia de la ingeniería geomática en la aplicación efectiva de modelos de clasificación de imágenes satelitales como SVM y RF en el ámbito de la teledetección. Estas técnicas son aplicables en el análisis multitemporal de LULC para la cartografía de diagnóstico en el ordenamiento territorial.

Materiales y métodos

Área de estudio

Con el objetivo de comparar el desempeño de los dos clasificadores mediante distintas estrategias para la obtención de las muestras de entrenamiento, se seleccionó una zona de 30 × 30 km en el Área Metropolitana de Barranquilla (AMB). Esta área está delimitada por las latitudes 10,82° N y 11,09° N, así como las longitudes 74,68° W y 74,96° W. Según informes oficiales, Barranquilla cuenta con una población de 2,7 millones de habitantes y se clasifica como la cuarta área metropolitana más grande de Colombia [24]. La ciudad se encuentra en el delta del río Magdalena, en la costa del mar Caribe (figura 1). Este río, que atraviesa Colombia de sur a norte, es considerado la principal arteria fluvial de la nación. Barranquilla es el epicentro del AMB y la principal ciudad del caribe colombiano, tanto en términos demográficos como económicos [25]. Los principales ecosistemas naturales en el área de estudio son el bosque tropical seco, el río Magdalena y los manglares estuarinos [26]. El uso del suelo es principalmente urbano y de conservación forestal. Por sus fronteras naturales, el crecimiento del AMB está restringido al suroeste de Barranquilla.

Figura 1. Localización del área de estudio



Nota: (a) Suramérica: Colombia resaltada. (b) Ubicación del departamento del Atlántico en Colombia. (c) Área de estudio 30 x 30 km (d) Área Metropolitana de Barranquilla Landsat-9 OLI-2 combinación de bandas (R: 5, G:6, B:4).

Fuente: elaboración propia

Datos

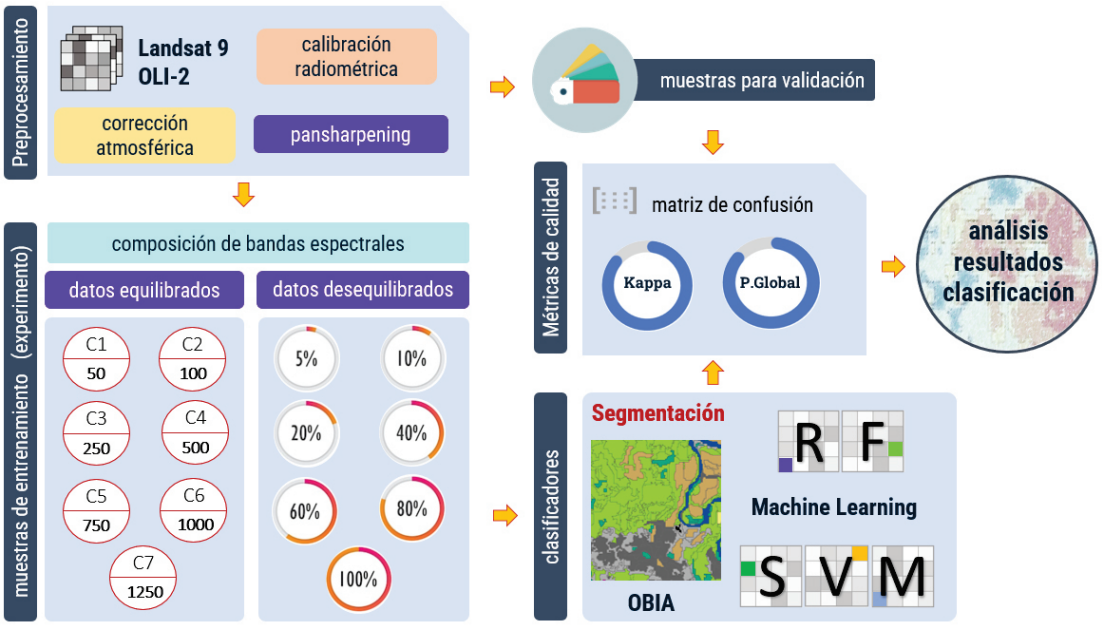
En este estudio, se utilizó una imagen satelital de la Tierra capturada por el instrumento OLI-2 a bordo del satélite L9, adquirida el 6 de enero del 2022. Este instrumento proporciona imágenes multiespectrales con una resolución espacial de 30 m, que abarcan cinco bandas visibles e infrarrojas cercanas (VNIR), dos bandas infrarrojas de onda corta (SWIR) y una banda cirrus, con una frecuencia de captura de datos cada 16 días. Además, L9 ofrece una banda pancromática de 15 m, que implica fusionar su información de alta resolución espacial con las bandas multiespectrales (que tienen mayor información espectral). Esta fusión permite crear imágenes compuestas con una alta resolución espacial y una información espectral detallada, que favorece la detección de características LULC más finas [27]. Esta técnica mejora significativamente la calidad del conjunto de datos procesados al combinar las bandas espectrales con un nivel de detalle superior.

La imagen se obtuvo del sitio web EarthExplorer del Servicio Geológico de los Estados Unidos (USGS) disponible en https://earthexplorer.usgs.gov/. La ruta y la fila de la imagen descargada fueron 009 y 052. Cada producto Landsat se entrega con bandas espectrales separadas en formato GeoTIFF y está georreferenciado al datum WGS84 en la proyección cartográfica UTM (18N). Las bandas utilizadas en este estudio incluyeron la banda 2 (0,45–0,51 µm), banda 3 (0,53–0,59 µm), banda 4 (0,64–0,67 µm), banda 5 (0,85–0,88 µm), banda 6 (1,57–1,65 µm), banda 7 (2,11–2,29 µm) y la banda 8 (0,53–0,68 µm). En el procesamiento de la imagen se emplearon los *softwares* ENVI 5.3 y ArcGIS Pro 2.8.0.

Métodos

La metodología propuesta está compuesta por cuatro etapas: (1) calibración de datos, (2) esquema de clasificación y áreas de entrenamiento, (3) segmentación y algoritmos de clasificación y (4) métricas de calidad. La figura 2 describe el flujo de datos aplicado en este experimento.

Figura 2. Diagrama de flujo del modelo experimental propuesto



Fuente: elaboración propia

Calibración de datos

La calibración radiométrica, que transformó los valores registrados por el sensor a unidades de radiancia aparente, se realizó siguiendo el procedimiento ampliamente descrito en el trabajo de Chander *et al.* [28]. Posteriormente, la imagen fue corregida por efectos atmosféricos para minimizar los errores causados por la dispersión y absorción de radiación debido al vapor de agua, partículas de polvo y aerosoles. Se empleó el método de substracción de objetos oscuros (DOS1) basado en el supuesto de que objetos oscuros como sombras, agua o bosques densos tienen una reflectancia cercana a cero, es decir, menor o igual al 1 %. Bajo esta suposición, los píxeles que representan estas características se consideran objetos oscuros y se utilizan para identificar la dispersión atmosférica [29]. Este método de corrección ha sido ampliamente utilizado en estudios previos de LULC [30], [31]. El procedimiento se aplicó empleando la calculadora ráster de ArcGIS Pro 2.8.0. La reflectancia de la superficie (ρ_s) se calculó mediante la siguiente expresión:

$$\boldsymbol{\rho_s = \frac{\pi \cdot (L_\lambda - L_p) \cdot d^2}{ESUN_\lambda \cdot cos\,\theta_s}} \quad (1)$$

donde L_λ es la radiancia espectral medida por el sensor, L_p es la radiancia de la atmósfera (efecto bruma), d es la distancia Tierra-Sol en unidades astronómicas, $ESUN\lambda$ es la irradiancia solar exoatmosférica y θ_s es el ángulo centital solar. El efecto bruma L_p está definido así:

$$\boldsymbol{L_p \quad L_{min} - L_{DO1}} \quad (2)$$

donde L_{min} es el valor mínimo de radiancia en la imagen, que se asume como radiancia del objeto oscuro en la sombra total y $L_{DO1\%}$) es el valor de radiancia correspondiente al 1 % de reflectancia, utilizado para identificar los píxeles oscuros en la imagen [32].

Finalmente, todas las bandas se recortaron para ajustarse a los límites del área de estudio.

Esquema de clasificación y áreas de entrenamiento

En este estudio las características superficiales terrestres fueron clasificadas en siete cubiertas referidas en el modelo Corine Land Cover (CLC) para Colombia. Estas son agua continental, agua marítima, tejido urbano, pastos limpios, mosaico de cultivos, bosque y red vial.

Se empleó un método de muestreo aleatorio simple para elegir las muestras del conjunto de datos tanto de entrenamiento como de validación, que busca asegurar la representación adecuada de las clases minoritarias [33].

Se seleccionaron cien segmentos para cada tipo de cobertura y se calculó la cantidad de píxeles en cada uno, como se indica en la tabla 1. Posteriormente, se calculó un número específico de píxeles para validar utilizando la ecuación de muestreo estratificado para poblaciones finitas propuesta en el estudio de Foody [34].

Tabla 1. Tamaños de muestras de entrenamiento y validación

Clase LULC	Definición	Entrenamiento (segmentos/ píxel)	Validación (píxeles)
Agua continental	Incluye masas de agua dulce, laguna costera, humedal, lagos y ciénagas.	100/ 71.190	383
Agua marítima	Se refiere a las masas de agua salada que cubren los mares, bahías y zonas como manglares.	100/ 101.575	384
Tejido urbano	Edificaciones y estructuras artificiales continuas y discontinuas creadas por el ser humano.	100/26.898	380
Pastos limpios	Áreas de cultivo de pasto para alimentación del ganado, incluye pastos arbolados, pastos enmalezados y praderas.	100/19.136	378
Mosaico de cultivos	Áreas de amplia variedad de cultivos agrícolas.	100/9.281	370
Bosque	Áreas cubiertas densamente por árboles perennes de gran altura.	100/16.569	376
Red vial	Áreas destinadas a infraestructuras de carreteras y redes de transporte.	100/2.411	332

Fuente: elaboración propia.

Posteriormente, los polígonos de entrenamiento de cada clase se dividieron en dos grupos: muestras equilibradas y muestras desequilibradas. En las muestras equilibradas, cada clase tuvo una cantidad similar de muestras de entrenamiento, que garantizó que todas las clases estuviesen representadas de manera equitativa en el conjunto de datos. En contraste, las muestras desequilibradas permitieron que algunas clases tuviesen más muestras de entrenamiento que otras, lo cual refleja la proporción original de píxeles en la estratificación inicial. Los datos de las muestras

de entrenamiento se dividieron en catorce conjuntos, siete equilibrados y siete desequilibrados de acuerdo con las cantidades expresadas en la tabla 2. Este enfoque adaptado de Thanh Noi y Kappas [35] permitió evaluar cómo el equilibrio de clases afectó el rendimiento de los algoritmos de clasificación en comparación con las muestras desequilibradas.

Tabla 2. Muestreo de datos para entrenar los clasificadores

Equilibrados	C1_e	C2_e	C3_e	C4_e	C5_e	C6_e	C7_e
Núm. píxeles	50	100	250	500	750	1000	1250

Desequilibrados	C1_d	C2_d	C3_d	C4_d	C5_d	C6_d	C7_d
Núm. píxeles	5 %	10 %	20 %	40 %	60 %	80 %	100 %

Nota. C1_e: conjunto 1 equilibrado; C1_d: conjunto 1 desequilibrado
Fuente: elaboración propia.

Segmentación y algoritmos de clasificación

La meta de la clasificación supervisada es asignar cada píxel de la imagen a clases particulares de cobertura del suelo. En este proceso, se utilizaron los algoritmos de Random Trees (RT) y Support Vector Machine (SVM) disponibles en ArcGIS Pro. El proceso de segmentación en ArcGIS Pro se fundamenta en el modelo Mean Shift. Este modelo emplea el reconocimiento de patrones a través de una ventana móvil para calcular valores promedio de píxeles, agrupándolos en segmentos. Una descripción detallada de Mean Shift se encuentra en Comaniciu y Meer [36]. Bajo el enfoque OBIA se categorizaron los segmentos constituidos por agrupaciones de píxeles con tipologías afines. Las características de los segmentos están determinadas por tres parámetros: detalle espectral, detalle espacial y tamaño mínimo del segmento. Debido a la falta de parámetros de referencia establecidos, se utilizó el método de prueba y error para encontrar la escala de segmentación adecuada [37]. En este estudio, se emplearon los siguientes valores para configurar los parámetros: detalle espectral: 20, detalle espacial: 20 y tamaño mínimo del segmento en píxeles: 15. En general, se buscó fragmentar la imagen en un número finito de regiones con la mayor escala posible, pero al mismo tiempo, asegurar la capacidad de distinguir entre ellas de manera efectiva. Estos segmentos fueron utilizados como muestras de entrenamiento en las que se probaron diversas combinaciones de bandas espectrales para lograr una mejor diferenciación de coberturas.

Support Vector Machine (SVM) se destaca como un algoritmo de aprendizaje automático supervisado de gran eficiencia en la segmentación de datos lineales y no lineales en el ámbito de los sensores remotos [38], [39]. Este modelo opera bajo la teoría del aprendizaje estadístico y la función de núcleo (Kernel). Su objetivo es encontrar de manera iterativa un hiperplano que maximice el margen libre de datos entre las clases de entrenamiento. El término "hiperplano" representa una regla de decisión lineal que emplea una función de mapeo derivada de las muestras de entrenamiento, que busca minimizar las clasificaciones erróneas [40]. Los puntos más cercanos al margen de clasificación se conocen como vectores de soporte. Luego, el modelo evalúa la estimación del hiperplano y hace predicciones con los datos de prueba. Si las predicciones son incorrectas, se produce una nueva selección de los vectores de soporte y se ajusta un nuevo hiperplano para mejorar la calidad del modelo. Las funciones de núcleo de base polinómica y radial se utilizan a menudo para proyectar clases no lineales en clases lineales separables en una dimensión superior [41]. Para llevar a cabo el experimento, se suministraron al clasificador SVM la imagen

segmentada, las muestras de entrenamiento y el esquema de clasificación. En ArcGIS Pro 2.8.0 solo es posible ajustar un parámetro dentro del modelo: la cantidad máxima de muestras por clase, que está limitada a 500. Este valor se mantuvo en su configuración predeterminada, luego, todas estas entradas se utilizaron conjuntamente para entrenar el clasificador SVM. La implementación de este algoritmo produjo un total de catorce modelos de clasificación SVM, divididos en siete para conjuntos de datos equilibrados y otros siete para conjuntos desequilibrados.

En ArcGIS Pro, el algoritmo Random Trees (RT) es un método de clasificación supervisado. Está fundamentado en el método estadístico de Random Forest (RF) [42]. Este emplea múltiples árboles de decisión integrados mediante la técnica de agregación *Bootstrap*, lo que implica entrenar cada árbol con diferentes subconjuntos de datos para obtener una clasificación similar. Al combinar los resultados de estos árboles, se compensan errores, lo que resulta en una decisión final basada en los votos de los árboles individuales. La característica destacada de este método es su habilidad para prevenir el sobreajuste del conjunto de entrenamiento, además de su eficiencia en el tiempo de clasificación [43]. El clasificador RT fue entrenado con el número máximo de árboles en 50 y su profundidad máxima en 30. Estos son los valores predeterminados sugeridos por ArcGIS Pro, que han sido probados también con resultados estables en un estudio de Wessel *et al*. [44]. Con esta configuración, se generaron catorce modelos de clasificación RT, distribuidos en siete conjuntos equilibrados y siete desequilibrados.

Métricas de calidad

Para evaluar el rendimiento de los clasificadores, se recolectaron muestras adicionales de segmentos en áreas distintas a las utilizadas para el entrenamiento. Cada segmento de muestra se asoció a una clase específica LULC mediante una imagen de Google Maps. Estas muestras se emplearon para realizar la evaluación cuantitativa de los clasificadores a partir del cálculo de matrices de confusión. Con ello se calculó la precisión global (OA), que ofrece una medida general de qué tan bien el clasificador ha asignado de manera correcta las clases en relación con todas las muestras de validación [45]. Así mismo, el coeficiente Kappa (K) fue calculado para evaluar la concordancia entre las clasificaciones observadas y las predicciones hechas por el clasificador, teniendo en cuenta las coincidencias que no podrían deberse al azar. Cuanto más cercano a 1 sea el valor de Kappa, mejor es el rendimiento del clasificador, lo que indica una alta concordancia entre las clasificaciones y las observaciones reales. El coeficiente Kappa (K) y la precisión global (OA) se calcularon a partir de las siguientes ecuaciones:

$$\boldsymbol{OA} = \frac{\sum_{i=1}^{r} x_{ii}}{N} \cdot \mathbf{100} \tag{3}$$

donde x_{ii} es el número de muestras clasificadas correctamente, N es el total de muestras extraídas y r es el número total de clases.

$$\boldsymbol{K} = \frac{N \cdot \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} \cdot x_{+i})} \tag{4}$$

x_{i+} es el número total de muestras clasificadas en la clase i, x_{+i} es el número total de muestras de la clase i en las muestras de referencia.

Resultados

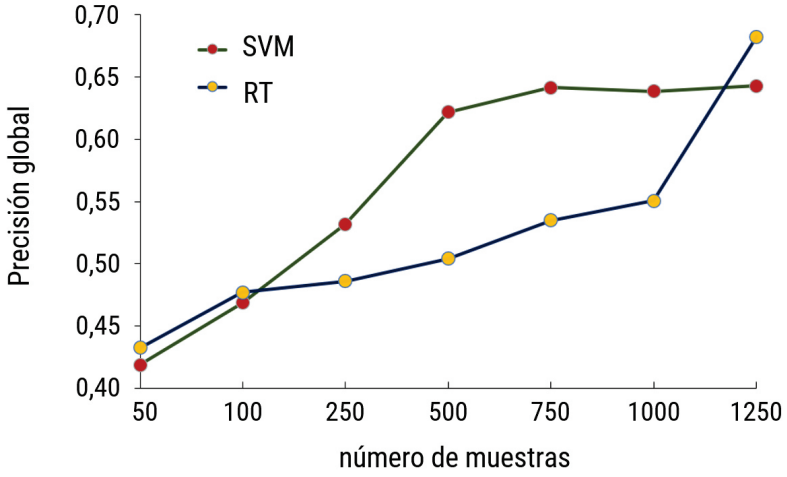
Rendimiento de los clasificadores SVM y RT en conjuntos de datos equilibrados

Este estudio analizó el desempeño de dos modelos de aprendizaje automático en la clasificación LULC basada en objetos utilizando una imagen L9 OLI-2. La figura 3 muestra la precisión general (OA) de los algoritmos evaluados en relación con el tamaño de la muestra para grupos equilibrados. En general, se observó un aumento en la precisión de ambos clasificadores a medida que aumentó la cantidad de píxeles de entrenamiento para la clasificación. Ambos métodos de clasificación mostraron su mayor precisión al emplear conjuntos de 1250 muestras mientras que la precisión más baja se registró con conjuntos de 50 muestras. Sin embargo, cada clasificador respondió de manera diferente al aumento del tamaño de la muestra. La clasificación de SVM alcanzó una precisión general del 64 % a partir de conjuntos con 750 muestras,

aunque, la estabilidad de este clasificador parece lograrse a partir de 500 muestras con una precisión global de 62 %. Desde 100 a 1000 muestras, el clasificador RT presentó un comportamiento casi lineal de incremento continuo, pero, con un crecimiento porcentual bajo del 7 %. No obstante, alcanzó su máxima precisión de 68 % con 1250 muestras que superan en 4 puntos porcentuales a SVM.

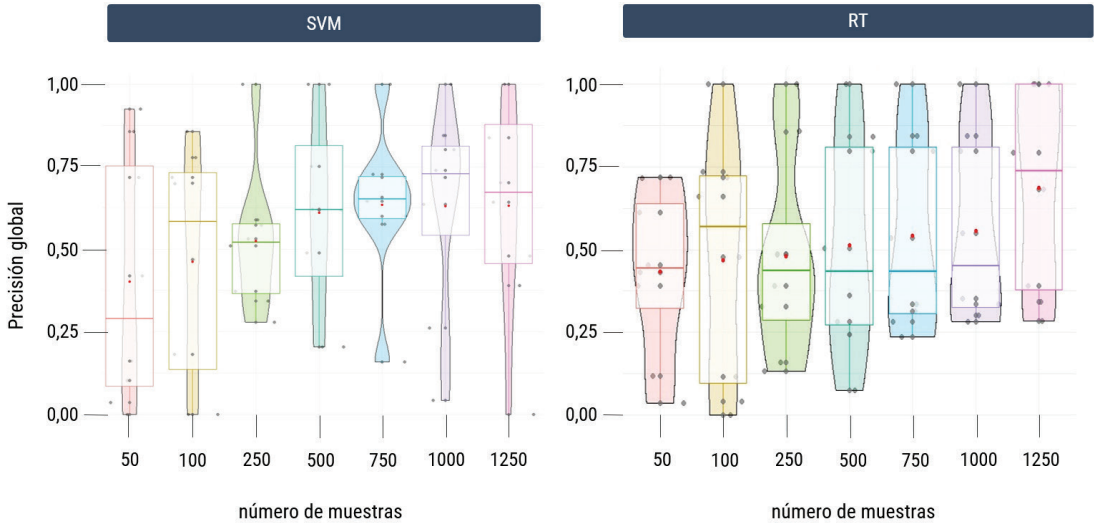
Las gráficas de violín en la figura 4 ilustran la distribución individual de los datos muestreados para cada clasificador y el número de píxeles seleccionado. La caja representa el rango intercuartílico (IQR) del conjunto de datos, es decir, el rango que contiene el 50 % central de los datos. El punto de color rojo indica la media, mientras que la línea en el interior de la caja representa la mediana.

Figura 3. Precisión global de las clasificaciones supervisadas y el tamaño del conjunto de entrenamiento en datos equilibrados



Fuente: elaboración propia.

Figura 4. Dispersión de datos muestrales de los clasificadores evaluados bajo conjuntos equilibrados



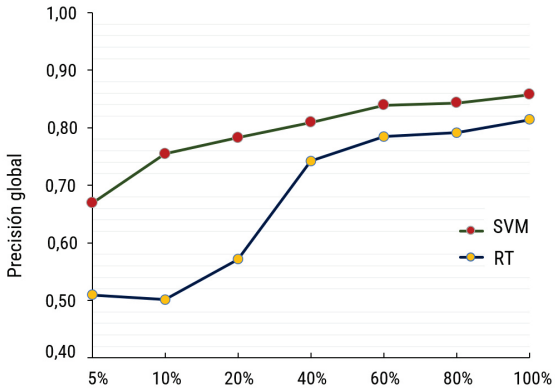
Fuente: elaboración propia.

Aunque la precisión de ambos clasificadores aumentó con la cantidad de píxeles de entrenamiento utilizados para la clasificación, la amplia variabilidad en las cajas IQR, especialmente para RT en comparación con SVM, indicó que hay una gran dispersión en los datos de entrenamiento. Esta variabilidad en RT sugiere inestabilidad en el rendimiento del modelo y la incapacidad para mejorar de manera consistente su precisión global, especialmente en comparación con SVM. La sensibilidad de RT a las variaciones en los datos de entrenamiento pudo conducir a resultados poco consistentes incluso cuando se utilizaron más datos para el entrenamiento, como se observa entre las 250 y 1000 muestras. Por otro lado, la menor variabilidad en SVM indicó una mayor estabilidad en el rendimiento del modelo, lo que sugiere que SVM fue capaz de manejar las variaciones en los datos de entrenamiento y producir resultados más consistentes con el aumento del tamaño del conjunto de entrenamiento.

Rendimiento de los clasificadores SVM y RT en conjuntos de datos desequilibrados

En la figura 5 se presenta el desempeño de los clasificadores con datos desequilibrados, evaluado mediante la precisión global (OA). En todos los niveles de remuestreo, se observó que SVM siempre produjo resultados de mayor precisión en comparación con RT. Sin embargo, las tres precisiones más altas obtenidas por ambos clasificadores variaron ligeramente entre sí. Los resultados de precisión del SVM no alcanzaron diferencias significativas entre los tamaños de muestra de entrenamiento del 60 %, 80 % y 100 % con valores de 0,84, 0,84 y 0,86 respectivamente. Por su parte, RT alcanzó valores de 0,78, 0,79 y 0,81 sobre el mismo muestreo. En los conjuntos de entrenamiento del clasificador SVM se evidenció un comportamiento casi lineal entre los tamaños de muestra del 10 % al 100 %. Esto significa que, al incrementar la cantidad de muestras de entrenamiento, el rendimiento del modelo mejora de manera predecible y constante. A través del comportamiento general de SVM, se interpreta que con un tamaño de muestra del 60 % se obtienen resultados consistentes de alta precisión, lo que indica que no es necesario aumentar el tamaño de las muestras para mantener un rendimiento óptimo.

Figura 5. Precisión global de las clasificaciones supervisadas y el tamaño del conjunto de entrenamiento en datos desequilibrados



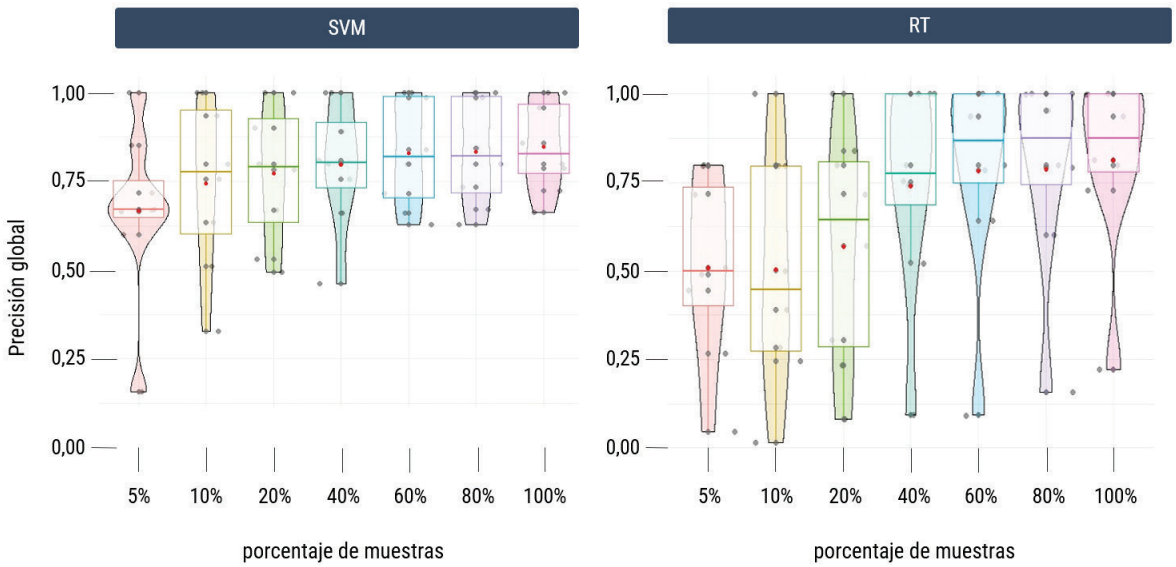
Fuente: elaboración propia.

La precisión del RT mostró diferencias significativas entre tamaños de muestra pequeños del 10 %, 20 % y 40 %. Se observó un aumento de 7 puntos porcentuales del 10 % al 20 % y un incremento de 17 puntos porcentuales entre el 20 % y 40 %, esto señala un aumento significativo en la precisión del modelo. Dicho aumento sugiere que el modelo logró aprender patrones más precisos entre el 20 % y el 40 % de las muestras de entrenamiento desequilibradas. A partir del 40 %, la figura mostró un comportamiento lineal con un aumento progresivo constante a medida que el tamaño de la muestra aumentó. Sin embargo, en las muestras del 60 %, 80 % y 100 % no se observó un aumento significativo, ya que sus precisiones fueron 0,78, 0,78 y 0,81 respectivamente. Esto indica que se alcanzó un punto de estabilidad en la precisión del algoritmo después de utilizar el 60 % de los datos de entrenamiento. A pesar de aumentar el tamaño de la muestra al 80 % y 100 %, la precisión se mantuvo en un nivel similar, lo que indica que no hubo mejoras significativas en la precisión del modelo con tamaños de muestra más grandes. Además, esto se interpreta como que el modelo RT logró su capacidad máxima de aprendizaje con el 60 % de los datos de entrenamiento empleados.

La figura 6 ilustra la distribución individual de los datos muestreados para cada clasificador y porcentaje de muestra mediante gráficas de violín. Resulta notable que en el muestreo del 5 % de los datos de SVM hubo una alta densidad alrededor de la media. Esto indica que espectralmente las muestras tuvieron valores cercanos entre sí, lo cual no es representativo de la población general y se evidenció a través del menor grado de precisión general obtenido. Las distribuciones de datos del 60 %, 80 % y 100 % presentaron un tamaño cercano al cuerpo del violín, lo cual indica homogeneidad en sus distribuciones y sugiere estabilidad en el rendimiento del modelo SVM. Las cajas del rango intercuartílico (IQR) del clasificador RT fueron significativamente más alargadas en comparación con las de SVM. La menor variabilidad en las predicciones del clasificador SVM, evidenciada por las cajas del IQR más cortas, señala una consistencia sobresaliente en las predicciones de este modelo. Ambos clasificadores parecen ser buenos generalizadores, ya que diferentes conjuntos de entrenamiento del mismo tamaño produjeron precisiones similares.

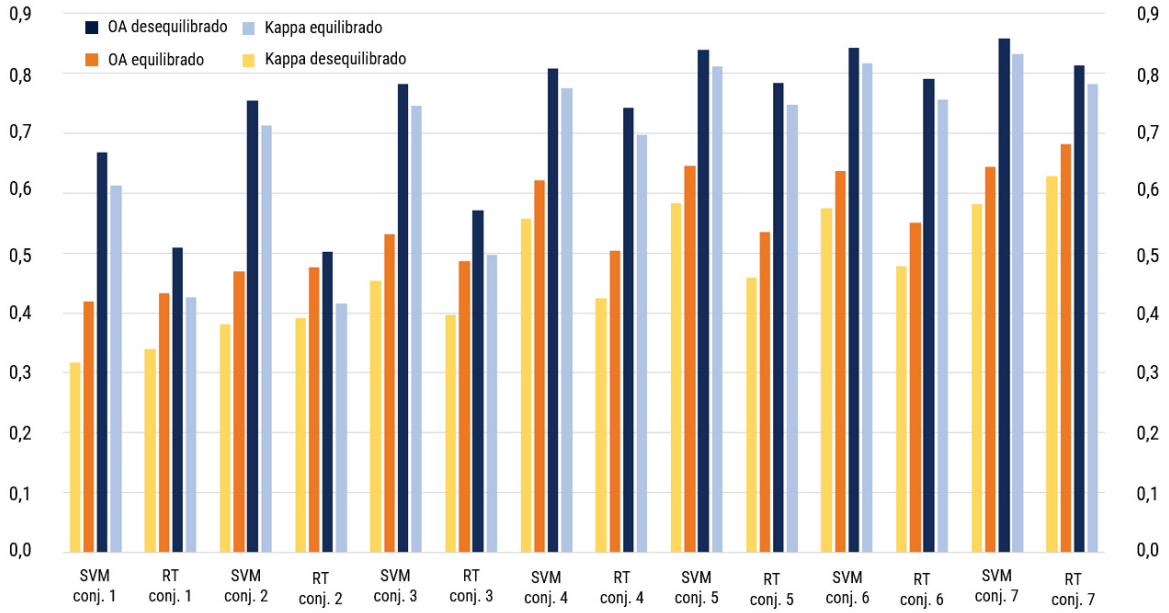
La figura 7 presenta el contraste de las diferencias obtenidas entre los conjuntos de datos equilibrados y desequilibrados de los catorce tamaños de muestra de entrenamiento para los clasificadores evaluados.

Figura 6. Dispersión de datos muestrales de los clasificadores evaluados en conjunto de datos de entrenamiento desequilibrados



Fuente: elaboración propia.

Figura 7. Diferencias de precisión entre conjuntos equilibrados y desequilibrados



Fuente: elaboración propia.

Discusión

En todos los casos, los valores del coeficiente Kappa (K) se encontraron cercanos a la precisión general. Esta cercanía indicó que las clasificaciones fueron consistentes, respaldando la calidad de los modelos y su capacidad para predecir con precisión las categorías sin depender del azar. Los conjuntos de entrenamiento desequilibrados (que muestrearon las clases en proporciones diferentes) mostraron una notable superioridad en precisión en comparación con los conjuntos equilibrados, donde por cada clase se muestreó la misma cantidad de píxeles que para las otras. Esto destaca la mayor capacidad de los clasificadores para aprender y predecir de manera más efectiva cuando se entrenan con datos desequilibrados. Los hallazgos mencionados destacan que los conjuntos de entrenamiento desequilibrados lograron capturar de manera más precisa la realidad del paisaje al reflejar las proporciones reales de las clases. En contraste con estos resultados, Thanh Noi y Kappas [35] argumentan que el rendimiento de los clasificadores en subconjuntos de datos equilibrados y desequilibrados fue comparable cuando el tamaño del conjunto de entrenamiento era convenientemente grande. Aunque los algoritmos y las técnicas de manejo del desequilibrio fueron equivalentes, los conjuntos de datos pudieron tener variaciones sutiles asociadas a la distribución y características de las clases. Otro factor que pudo influenciar estos resultados contradictorios fue la configuración de los parámetros del modelo SVM y RF.

En el caso de SVM, la función de penalización (C) y el valor de gamma (γ) son los dos parámetros esenciales que controlan el rendimiento de SVM cuando se utiliza la función de base radial (RBF) como núcleo central [46]. El parámetro de costo (C) controla el equilibrio entre la precisión de la clasificación y la simplicidad del modelo, mientras que el parámetro gamma (γ) regula la flexibilidad del límite de decisión en el modelo [47]. Thanh Noi y Kappas [35] realizaron pruebas con 10 valores para C y γ hasta encontrar su combinación óptima. Sin embargo, estos parámetros no están disponibles para la calibración del modelo SVM en ArcGIS Pro.

Nuestro estudio se suma al debate sobre el rendimiento del clasificador RF en diferentes tamaños de conjuntos de entrenamiento. RT siempre presentó mayor precisión general sobre su rendimiento en datos desequilibrados que sobre los

equilibrados. Este resultado respalda los hallazgos presentados por Mellor *et al*. [48] que señalaron que cuanto mayor fue el área de la clase de cobertura terrestre en el conjunto de datos, se necesitaron más muestras de entrenamiento para lograr una mejor precisión del clasificador. Por el contrario, Ramezan *et al*. [49] sostienen que RF es resistente a la variación en el tamaño de las muestras de entrenamiento, manteniendo una alta precisión, incluso con tamaños de muestra pequeños, y mostrando una disminución mínima en la precisión general a medida que aumentó el volumen de las muestras de entrenamiento. Resultados que fueron obtenidos con una configuración de número máximo de árboles en 500 mientras que en nuestro trabajo este parámetro fue de 50. Entrenar el clasificador RT con 50 árboles pudo limitar la capacidad del modelo para capturar la complejidad de los datos, especialmente cuando se enfrentó a tamaños pequeños de muestra.

Por otro lado, el estudio de Ramezan *et al*. [49], que utilizó 500 árboles, podría haber permitido que su modelo aprendiera patrones más sutiles y se adaptara mejor a los tamaños variados de muestra. Entrenar un clasificador RF con un número mayor de árboles aumenta generalmente la capacidad del modelo para aprender patrones complejos en los datos, lo que puede llevar a una mejor generalización. Sin embargo, después de un cierto punto, el beneficio de agregar más árboles disminuye y puede volverse computacionalmente costoso [50]. En este sentido, nuestros datos reportaron que RT en los conjuntos 5, 6 y 7 de la figura 7 (muestreos de 60 %, 80 % y 100 %) no presentaron diferencias significativas frente a la OA. Esto podría implicar que el modelo alcanzó su capacidad máxima de aprendizaje con el 60 % de los datos y agregar más datos no resultó en un aumento importante de su precisión general. A pesar de que se presentaron diferencias en el rendimiento entre conjuntos de datos de entrenamiento equilibrados y desequilibrados para ambos algoritmos, la precisión demostró ser consistentemente superior en el caso de SVM en comparación con RT. La homogeneidad tanto en las muestras de entrenamiento como en las de validación se identificó como un factor crítico para la estabilidad en el rendimiento de los modelos, esta tendencia estuvo respaldada por la precisión global (OA) y el índice Kappa.

El rendimiento del modelo SVM se estabilizó al emplear el 60 % del tamaño total de las muestras de datos de entrenamiento. Aumentar el tamaño de las muestras más allá de este punto no proporcionó mejoras significativas en la precisión del modelo. La tabla 3 presenta la superficie ocupada por las clases LULC, así como el área ocupada por las regiones de entrenamiento. Además, muestra el porcentaje que estas áreas representan en relación con el área total clasificada. También se detalla la precisión del usuario para cada clase correspondiente al modelo SVM con un 60 % del tamaño total de las muestras de datos de entrenamiento.

Tabla 3. Resultados de precisión de usuario para clases LULC (SVM, 60 % muestras de entrenamiento)

Clase LULC	Área clasificada (km^2)	Área de entrenamiento (km^2)	% área de entrenamiento sobre el área clasificada	Precisión de usuario
Agua continental	120	9,7	8,1	1,0
Red vial	12	0,3	2,6	1,0
Bosques	280	2,3	0,8	0,98
Pastos limpios	112	2,6	2,3	0,91
Agua marítima	127	13,9	10,9	0,85
Tejido urbano	151	3,7	2,4	0,69
Mosaico de cultivos	98	1,3	1,3	0,66

Fuente: elaboración propia.

La variación de los porcentajes del área de entrenamiento entre las cubiertas de agua continental (8,1 %), red vial (2,6 %) y bosques (0,8 %), todas con una precisión de usuario muy alta, resulta interesante y puede atribuirse a la complejidad inherente de estas clases. Es posible que existan variaciones naturales significativas en las características del agua, como profundidad, turbulencia y características del lecho, que requieren un área de entrenamiento más grande para capturar estas variaciones. Por otro lado, las características de las carreteras, como color, textura y forma, podrían ser más homogéneas y fáciles de aprender para el modelo, lo que explicaría por qué se necesita un área de entrenamiento menor. La clase bosque, a pesar de tener un porcentaje de área de entrenamiento baja, logró una alta precisión de usuario del 98 %. Las áreas de bosque a menudo presentan una mayor heterogeneidad espacial, lo que puede proporcionar una gama más amplia de firmas espectrales en un área relativamente pequeña, permitiendo así un mejor aprendizaje con menos datos de entrenamiento. El tejido urbano y los mosaicos de cultivos pueden ser muy heterogéneos en términos de estructuras, patrones y materiales. En áreas urbanas hay una gran variabilidad en materiales y tipologías constructivas. De manera similar, los mosaicos de cultivos pueden incluir diferentes tipos de cultivos con características espectrales disímiles. Estas clases necesitan considerar áreas más amplias de entrenamiento para representar mejor sus características.

Conclusiones

El propósito de esta investigación fue evaluar el desempeño de los algoritmos de aprendizaje automático supervisado Support Vector Machine (SVM) y Random Forest (RF), utilizando un modelo de segmentación OBIA y variaciones en el volumen de las muestras empleadas para el entrenamiento. Para evaluar la precisión de los clasificadores, se emplearon las métricas de precisión general (OA) y el coeficiente Kappa en una imagen multiespectral de Landsat-9 OLI-2 en la zona metropolitana de Barranquilla, Colombia. Los resultados demostraron que SVM logró mayor precisión general en datos desequilibrados. Se confirmó la importancia de que las muestras de entrenamiento sean proporcionales a la superficie ocupada por las diversas cubiertas sobre el paisaje. Los resultados indicaron que para lograr precisiones de usuario superiores al 90 % en las clases de pastos limpios, bosques, red vial y agua continental mediante el modelo SVM en ArcGIS Pro, se aconseja asignar muestras de entrenamiento que abarquen el 2 %, 1 %, 3 % y 8 % del área clasificada, respectivamente. Encontrar equilibrio entre la representación precisa del paisaje y la suficiente cantidad de datos para cada clase es fundamental para construir modelos de clasificación precisos y generalizables.

Referencias

[1] S. M. Oswald *et al.*, "Using urban climate modelling and improved land use classifications to support climate change adaptation in urban environments: A case study for the city of Klagenfurt, Austria", *Urban Clim.*, vol. 11, no. 10, p. 1692, mar., 2020, DOI: 10.1016/j.uclim.2020.100582

[2] S. Afrin, A. Gupta, B. Farjad, M. Razu Ahmed, G. Achari y Q. Hassan, "Development of land-use/land-cover maps using landsat-8 and MODIS data, and their integration for hydro-ecological applications", *Sensors*, vol. 19, no. 22, p. 4891, nov., 2019, DOI: 10.3390/s19224891.

[3] K. Vatitsi *et al.*, "LULC Change Effects on Environmental Quality and Ecosystem Services Using EO Data in Two Rural River Basins in Thrace, Greece", *Land*, vol. 12, no. 6, p. 1140, mayo, 2023, DOI: 10.3390/land12061140.

[4] C. Zhang y X. Li, "Land Use and Land Cover Mapping in the Era of Big Data", *Land*, vol. 11, no. 10, sept., 2022, DOI: 10.3390/land11101692.

[5] B. Rimal, L. Zhang, H. Keshtkar, B. N. Haack, S. Rijal y P. Zhang, "Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and markov chain", *ISPRS Int. J. Geo-Information*, vol. 7, no. 4, p. 154, abr., 2018, DOI: 10.3390/ijgi7040154.

[6] S. Dahhani, M. Raji, M. Hakdaoui y R. Lhissou, "Land Cover Mapping Using Sentinel-1 Time-Series Data and Machine-Learning Classifiers in Agricultural Sub-Saharan Landscape", *Remote Sens.*, vol. 15, no. 1, p. 65, dic., 2022, DOI: 10.3390/rs15010065.

[7] R. Showstack, "Landsat 9 Satellite Continues Half-Century of Earth Observations," *Bioscience*,

vol. 72, no. 3, pp. 226–232, mar., 2022, DOI: 10.1093/biosci/biab145.

[8] H. You, X. Tang, W. Deng, H. Song, Y. Wang y J. Chen, "A study on the difference of LULC classification results based on Landsat 8 and Landsat 9 data", *Sustainability*, vol. 14, no. 21, p. 13730, oct., 2022, DOI: 10.3390/su142113730.

[9] A. E. Maxwell, T. A. Warner y F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review", *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, feb., 2018, DOI: 10.1080/01431161.2018.1433343.

[10] D. Lu y Q. Weng, "A survey of image classification methods and techniques for improving classification performance", *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, mar., 2007, DOI: 10.1080/01431160600746456.

[11] N. Wu, L. G. T. Crusiol, G. Liu, D. Wuyun y G. Han, "Comparing Machine Learning Algorithms for Pixel/Object-Based Classifications of Semi-Arid Grassland in Northern China Using Multisource Medium Resolution Imageries", *Remote Sens.*, vol. 15, no. 3, p. 750, ene., 2023, DOI: 10.3390/rs15030750.

[12] E. Y. Boateng, J. Otoo y D. A. Abaye, "Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review", *J. Data Anal. Inf. Process.*, vol. 8, no. 4, pp. 341–357, nov., 2020, DOI: 10.4236/jdaip.2020.84020.

[13] C. Zhang, Y. Liu y N. Tie, "Forest Land Resource Information Acquisition with Sentinel-2 Image Utilizing Support Vector Machine, K-Nearest Neighbor, Random Forest, Decision Trees and Multi-Layer Perceptron", *Forests*, vol. 14, no. 2, p. 254, ene., 2023, DOI: 10.3390/f14020254

[14] T. K. Oo, N. Arunrat, S. Sereenonchai, A. Ussawarujikulchai, U. Chareonwong y W. Nutmagul, "Comparing Four Machine Learning Algorithms for Land Cover Classification in Gold Mining: A Case Study of Kyaukpahto Gold Mine, Northern Myanmar", *Sustainability*, vol. 14, no. 17, p. 10754, ago., 2022, DOI: 10.3390/su141710754

[15] Y. Ouma *et al.*, "Comparison of Machine Learning Classifiers for Multitemporal and Multisensor Mapping of Urban Lulc Features", *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. XLIII-B3-2, pp. 681–689, 2022, DOI: 10.5194/isprs-archives-XLIII-B3-2022-681-2022

[16] J. S. Deng, K. Wang, Y. H. Deng y G. J. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *Int. J. Remote Sens.*, vol. 29, no. 16, pp. 4823–4838, jul., 2008, DOI: 10.1080/01431160801950162

[17] M. Pfeifer, M. Disney, T. Quaife y R. Marchant, "Terrestrial ecosystems from space: A review of earth observation products for macroecology applications," *Glob. Ecol. Biogeogr.*, vol. 21, no. 6, pp. 603–624, oct., 2011, DOI: 10.1111/j.1466-8238.2011.00712.x

[18] P. Lourenço, A. C. Teodoro, J. A. Gonçalves, J. P. Honrado, M. Cunha y N. Sillero, "Assessing the performance of different OBIA software approaches for mapping invasive alien plants along roads with remote sensing data," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 95, p. 102263, mar., 2021, DOI: 10.1016/j.jag.2020.102263

[19] Q. Feng, Y. Li y B. Yang, "Modeling Land Seismic Exploration Random Noise in a Weakly Heterogeneous Medium and the Application to the Training Set," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 1–5, abr., 2020, DOI: 10.1109/LGRS.2019.2926756

[20] A. Jamali, "Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: a case study of the northern region of Iran," *SN Appl. Sci.*, vol. 1, p. 1448, oct., 2019, DOI: 10.1007/s42452-019-1527-8

[21] S. Basheer *et al.*, "Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques," *Remote Sens.*, vol. 14, no. 19, p. 4978, oct., 2022, DOI: 10.3390/rs14194978

[22] Y. G. Yuh, W. Tracz, H. D. Matthews y S. E. Turner, "Application of machine learning approaches for land cover monitoring in northern Cameroon," *Ecol. Inform.*, vol. 74, p. 101955, mayo, 2023, DOI: 10.1016/j.ecoinf.2022.101955

[23] M. Azadbakht, C. S. Fraser y K. Khoshelham, "Synergy of sampling techniques and ensemble classifiers for classification of urban environments using full-waveform LiDAR data," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 73, pp. 277–291, dic., 2018, DOI: 10.1016/j.jag.2018.06.009

[24] Alcaldía de Barranquilla, "Plan de Desarrollo. Soy Barranquilla 2020–2023," 2020. https://www.barranquilla.gov.co/transparencia/normatividad/normativa-de-la-entidad/politicas-lineamientos-y-manuales/plan-de-desarrollo

[25] J. Aldana Domínguez, I. Palomo, J. Gutiérrez-Angonese, C. Arnaiz-Schmitz, C. Montes y F. Narvaez, "Assessing the effects of past and future land cover changes in ecosystem services, disservices and biodiversity: A case study in Barranquilla Metropolitan Area (BMA), Colombia," *Ecosyst. Serv.*, vol. 37, p. 100915, jun., 2019, DOI: 10.1016/j.ecoser.2019.100915

[26] J. Aldana-Domínguez, C. Montes y J. A. González, "Understanding the past to envision a sustainable future: A social-ecological history of the Barranquilla Metropolitan Area (Colombia)," *Sustain.*, vol. 10, no. 7, p. 2247, jun., 2018, DOI: 10.3390/su10072247

[27] A. Tassi, D. Gigante, G. Modica, L. Di Martino y M. Vizzari, "Pixel-vs. Object-based landsat 8 data classification in google earth engine using random forest: The case study of maiella national park," *Remote Sens.*, vol. 13, no. 12, p. 2299, jun., 2021, DOI: 10.3390/rs13122299

[28] G. Chander, B. L. Markham y D. L. Helder, "Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors," *Remote Sens. Environ.*, vol. 113, no. 12, pp. 893–903, mayo, 2009, DOI: 10.1016/j.rse.2009.01.007

[29] P. S. J. Chavez, "An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data," *Remote Sens. Environ.*, vol. 24, no. 3, pp. 459–479, abr., 1988, DOI: 10.1016/0034-4257(88)90019-3

[30] C. Valdivieso-Ros, F. Alonso-Sarria y F. Gomariz-Castillo, "Effect of different atmospheric correction algorithms on sentinel-2 imagery classification accuracy in a semiarid mediterranean area," *Remote Sens.*, vol. 13, no. 9, p. 1770, mayo, 2021, DOI: 10.3390/rs13091770

[31] J. D. Revuelta-Acosta, E. S. Guerrero-Luis, J. E. Terrazas-Rodriguez, C. Gomez-Rodriguez y G. A. Perea, "Application of Remote Sensing Tools to Assess the Land Use and Land Cover Change in Coatzacoalcos, Veracruz, Mexico," *Appl. Sci.*, vol. 12, no. 4, p. 1882, feb., 2022, DOI: 10.3390/app12041882

[32] J. A. Sobrino, J. C. Jiménez-Muñoz y L. Paolini, "Land surface temperature retrieval from LANDSAT TM 5," *Remote Sens. Environ.*, vol. 90, no. 4, pp. 434–440, abr., 2004, DOI: 10.1016/j.rse.2004.02.003

[33] C. A. Ramezan, T. A. Warner y A. E. Maxwell, "Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification," *Remote Sens.*, vol. 11, no. 2, p. 185, ene., 2019, DOI: 10.3390/rs11020185

[34] G. M. Foody, "Sample size determination for image classification accuracy assessment and comparison," *Int. J. Remote Sens.*, vol. 30, no. 20, pp. 5273–5291, sep., 2009, DOI: 10.1080/01431160903130937

[35] P. Thanh Noi y M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery," *Sensors*, vol. 18, no. 1, p. 18, dic., 2017, DOI: 10.3390/s18010018

[36] D. Comaniciu y P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, mayo, 2002, DOI: 10.1109/34.1000236

[37] K. Luo, B. Li y J. P. Moiwo, "Monitoring land-use/land-cover changes at a provincial large scale using an object-oriented technique and medium-resolution remote-sensing images," *Remote Sens.*, vol. 10, no. 12, p. 2012, dic., 2018, DOI: 10.3390/rs10122012

[38] Y. Chabalala, E. Adam y K. A. Ali, "Machine Learning Classification of Fused Sentinel-1 and Sentinel-2 Image Data towards Mapping Fruit Plantations in Highly Heterogenous Landscapes," *Remote Sens.*, vol. 14, no. 11, p. 2621, mayo, 2022, DOI: 10.3390/rs14112621

[39] Y. Wei, W. Wang, X. Tang, H. Li, H. Hu y X. Wang, "Classification of Alpine Grasslands in Cold and High Altitudes Based on Multispectral Landsat-8 Images : A Case Study in Sanjiangyuan National Park , China," *Remote Sens.*, vol. 14, no. 15, p. 3714, ago., 2022, DOI: https://doi.org/10.3390/rs14153714

[40] G. De Luca *et al.*, "Object-based land cover classification of cork oak woodlands using UAV imagery and Orfeo Toolbox," *Remote Sens.*, vol. 11, no. 10, p. 1238, mayo, 2019, DOI: 10.3390/rs11101238

[41] S. Talukdar, P. Singha, S. Mahato, S. Pal, Y. A. Liou y A. Rahman, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," *Remote Sens.*, vol. 12, no. 7, p. 1135, abr., 2020, DOI: https://doi.org/10.3390/rs12071135

[42] G. R. Morgan, C. Wang, Z. Li, S. R. Schill y D. R. Morgan, "Deep Learning of High-Resolution Aerial Imagery for Coastal Marsh Change Detection: A Comparative Study," *ISPRS Int. J. Geo-Information*, vol. 11, no. 2, p. 100, feb., 2022, DOI: 10.3390/ijgi11020100

[43] A. Sabat-Tomala, E. Raczko y B. Zagajewski, "Comparison of support vector machine and random forest algorithms for invasive and expansive species classification using airborne hyperspectral data," *Remote Sens.*, vol. 12, no. 3, p. 516, feb., 2020, DOI: 10.3390/rs12030516

[44] M. Wessel, M. Brandmeier y D. Tiede, "Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data," *Remote Sens.*, vol. 10, no. 9, p. 1419, sept., 2018, DOI: 10.3390/rs10091419

[45] X. Li, R. Wang, X. Chen, Y. Li y Y. Duan, "Classification of Transmission Line Corridor Tree Species Based on Drone Data and Machine Learning," *Sustainability*, vol. 14, no. 14, p. 8273, jul., 2022, DOI: 10.3390/su14148273

[46] T. Adugna, W. Xu y J. Fan, "Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images," *Remote Sens.*, vol. 14, no. 3, p. 574, ene., 2022, DOI: 10.3390/rs14030574

[47] I. Potić *et al.*, "Improving Forest Detection Using Machine Learning and Remote Sensing: A Case Study in Southeastern Serbia," *Appl. Sci.*, vol. 13, no. 14, p. 8289, jul., 2023, DOI: 10.3390/app13148289

[48] A. Mellor, S. Boukir, A. Haywood y S. Jones, "Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 155–168, jul., 2015, DOI: 10.1016/j.isprsjprs.2015.03.014

[49] C. A. Ramezan, T. A. Warner, A. E. Maxwell y B. S. Price, "Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data," *Remote Sens.*, vol. 13, no. 3, p. 368, ene., 2021, DOI: 10.3390/rs13030368

[50] A. Zafari, R. Zurita-Milla y E. Izquierdo-Verdiguier, "Evaluating the performance of a Random Forest Kernel for land cover classification," *Remote Sens.*, vol. 11, no. 5, p. 575, mar., 2019, DOI: 10.3390/rs11050575