

USO DE MACHINE LEARNING E IMÁGENES SATELITALES PARA LA CLASIFICACIÓN DEL USO DE SUELO EN EL CANTÓN RIOBAMBA

USE OF MACHINE LEARNING AND SATELLITE IMAGES FOR LAND USE CLASSIFICATION IN THE RIOBAMBA CANTON

Alfredo Rodrigo Colcha Ortiz¹, Raúl Alexis Salazar Flores², Hernán Vladimir Pazmiño Chiluiza³,
Marcelo David Guerra Valladares⁴

{alfredo.colcha@unach.edu.ec¹, raul.salazar@unach.edu.ec², hv pazmino@unach.edu.ec³, marcelo.guerra@unach.edu.ec⁴}

Fecha de recepción: 26/12/2025 / Fecha de aceptación: 05/01/2026 / Fecha de publicación: 06/01/2026

RESUMEN: El análisis del uso de suelo constituye un insumo fundamental para la planificación territorial y la gestión ambiental, especialmente en territorios con alta complejidad geográfica como las ciudades interandinas. En este contexto, el presente estudio tiene como objetivo aplicar técnicas de Machine Learning combinadas con imágenes satelitales Sentinel-2 para la clasificación del uso de suelo en el cantón Riobamba, Ecuador. El trabajo de investigación utiliza una metodología con un enfoque cuantitativo, de tipo aplicado, se utiliza imágenes de Sentinel-2 de reflectancia superficial correspondientes a agosto de 2025, descargadas con la plataforma de Google Earth Engine y procesadas mediante estadística multiespectral. Para la clasificación se definen siete clases de uso y cobertura del suelo: agua, suelo desnudo, bosque, vegetación baja/pasto, construcción, hielo y erosión. El proceso de clasificación se realizó mediante el algoritmo Support Vector Machines (SVM), el algoritmo fue creado en el lenguaje Python con el entorno Google Colab, incluye la optimización de parámetros mediante validación cruzada. La validación del modelo se efectuó a través de una matriz de confusión construida con 203 puntos de prueba, obteniéndose una exactitud global del 72,41 %, lo que evidencia un desempeño satisfactorio del clasificador. Los resultados muestran una predominancia de la vegetación baja y pastos, seguida por áreas de suelo desnudo, bosque y construcción, reflejando tanto las dinámicas agropecuarias como el proceso de expansión urbana del cantón. La identificación de las clases hielo y erosión permitió caracterizar zonas de alta montaña que están relacionadas al volcán Chimborazo y al

¹Docente Investigador, Facultad de Ingeniería, Universidad Nacional de Chimborazo (UNACH), <https://orcid.org/0009-0005-2280-5189>

²Docente Investigador, Facultad de Ingeniería, Universidad Nacional de Chimborazo (UNACH), <https://orcid.org/0000-0001-6483-2613>

³Docente Investigador, Facultad de Ingeniería, Universidad Nacional de Chimborazo (UNACH), <https://orcid.org/0000-0002-7169-7648>

⁴Docente Investigador, Facultad de Ingeniería, Universidad Nacional de Chimborazo (UNACH), <https://orcid.org/0000-0003-4119-4959>

nevado El Altar. En conjunto, el estudio demuestra que el uso de SVM constituye una herramienta robusta y confiable para la generación de cartografía temática de uso de suelo, aportando información clave para la planificación y gestión territorial a escala local.

Palabras clave: Clasificación del uso de suelo, Google Earth Engine, Imágenes satelitales, Matriz de confusión, Sentinel-2, Sistemas de Información Geográfica, Support Vector Machines

ABSTRACT: Land use analysis is a fundamental input for territorial planning and environmental management, especially in territories with high geographical complexity such as inter-Andean cities. In this context, the present study aims to apply machine learning techniques combined with Sentinel-2 satellite images for land use classification in the Riobamba canton, Ecuador. The research uses a quantitative, applied methodology, utilizing Sentinel-2 surface reflectance images from August 2025, downloaded using the Google Earth Engine platform and processed using multispectral statistics. Seven classes of land use and land cover are defined for classification: water, bare soil, forest, low vegetation/grass, construction, ice, and erosion. The classification process was performed using the Support Vector Machines (SVM) algorithm, which was created in Python with the Google Colab environment and includes parameter optimization through cross-validation. The model was validated using a confusion matrix constructed with 203 test points, obtaining an overall accuracy of 72.41%, which demonstrates the satisfactory performance of the classifier. The results show a predominance of low vegetation and pastures, followed by areas of bare soil, forest, and construction, reflecting both agricultural dynamics and the process of urban expansion in the canton. The identification of ice and erosion classes made it possible to characterize high mountain areas related to the Chimborazo volcano and the El Altar snow-capped mountain. Overall, the study demonstrates that the use of SVM is a robust and reliable tool for generating thematic land use mapping, providing key information for territorial planning and management at the local level.

Keywords: Land use classification, Google Earth Engine, satellite imagery, confusion matrix; Sentinel-2, geographic information systems, support vector machines

INTRODUCCIÓN

El análisis y la clasificación del uso de suelo constituyen herramientas fundamentales para la planificación territorial, la gestión ambiental y la toma de decisiones en el ámbito urbano y rural (1). El crecimiento poblacional acelerado, la expansión urbana no planificada y los cambios en las actividades productivas han generado transformaciones significativas en la cobertura y uso del suelo, especialmente en ciudades interandinas de países en desarrollo (2). Para las administraciones de todos los territorios es importante contar con información actualizada, confiable, precisa y espacial sobre el uso de suelo, ya que es indispensable para formular políticas públicas orientadas al ordenamiento territorial y al desarrollo sostenible (3).

El cantón Riobamba, que pertenece a la provincia de Chimborazo en la región central del Ecuador, experimenta en las últimas décadas procesos continuos de expansión urbana y cambios en el uso de su territorio, esto corresponde a factores demográficos, económicos y ambientales (4). Estas dinámicas han generado presiones sobre áreas agrícolas, zonas naturales y espacios periurbanos, lo que demanda el uso de metodologías eficientes que permitan identificar, clasificar y monitorear los distintos tipos de uso de suelo de manera objetiva y reproducible (5). En este escenario, las imágenes satelitales se han consolidado como una fuente de información clave, al ofrecer cobertura espacial amplia, periodicidad temporal y acceso a múltiples bandas espectrales que facilitan la discriminación de coberturas terrestres (6).

Las imágenes satelitales se han consolidado como una de las principales fuentes de información para la clasificación del uso de suelo, debido a su amplia cobertura espacial, disponibilidad temporal y capacidad para captar información espectral de la superficie terrestre (7). Sensores remotos de resolución media y alta permiten discriminar diferentes tipos de cobertura y uso del suelo mediante el análisis de firmas espectrales asociadas a áreas urbanas, agrícolas, forestales y cuerpos de agua (8). Estas características facilitan la generación de cartografía temática actualizada y objetiva, reduciendo costos y tiempos frente a métodos tradicionales de levantamiento en campo, en el contexto de la planificación territorial y urbana, el uso de imágenes satelitales posibilita el monitoreo continuo de las dinámicas espaciales, proporcionando insumos técnicos confiables para la toma de decisiones y la gestión sostenible del territorio (2).

En los últimos años, las técnicas de Machine Learning han adquirido un rol protagónico en el análisis de datos geoespaciales, debido a su capacidad para modelar relaciones complejas y no lineales entre variables espectrales y clases de uso de suelo, con modelos como SVM, Random Forest, Árboles de Decisión y k-NN (9). A diferencia de los métodos tradicionales de clasificación supervisada, los algoritmos de aprendizaje automático permiten mejorar la precisión de los resultados, reducir la subjetividad del analista y optimizar el procesamiento de grandes volúmenes de datos satelitales (10). Estas ventajas han impulsado su aplicación en estudios de clasificación del uso y cobertura del suelo a diferentes escalas territoriales.

El modelo Support vector machines (SVM) se han consolidado como una de las técnicas más robustas y eficientes para la clasificación del uso de suelo a partir de imágenes satelitales (11). La fortaleza de SVM radica en su capacidad para encontrar fronteras de decisión óptimas en espacios de alta dimensionalidad, lo que resulta especialmente adecuado para datos espectrales multibanda y no lineales. Mediante el uso de funciones kernel, SVM permite modelar relaciones complejas entre las firmas espectrales y las clases de uso de suelo, logrando altos niveles de precisión incluso con conjuntos de entrenamiento limitados (12). Diversos estudios han demostrado su desempeño superior frente a otros clasificadores tradicionales, particularmente en entornos heterogéneos y urbanos, lo que justifica su aplicación en el análisis territorial y la planificación del uso de suelo (13).

El lenguaje de programación Python se ha consolidado como una herramienta fundamental para la implementación de modelos de Machine Learning aplicados a la clasificación del uso de suelo, debido a su versatilidad, eficiencia y amplia disponibilidad de librerías especializadas (14).

Bibliotecas como *scikit-learn*, *TensorFlow* y *PyTorch* permiten desarrollar, entrenar y validar algoritmos de aprendizaje automático de manera eficiente, facilitando el procesamiento de grandes volúmenes de datos provenientes de imágenes satelitales (15). Además, Python ofrece una alta capacidad de integración con entornos de Sistemas de Información Geográfica, lo que posibilita automatizar flujos de trabajo, reproducir resultados y mejorar la precisión de los procesos de clasificación (16). Por sus características Python es un componente fundamental para el análisis geoespacial y la construcción de cartografía temática confiable.

La plataforma de Google Colab se ha consolidado como un entorno de desarrollo integrado (IDE) en la nube más utilizado para la implementación de modelos de Machine Learning en el lenguaje Python, especialmente para el análisis y clasificación del uso de suelo a partir de imágenes satelitales (17). La plataforma permite ejecutar código sin necesidad de configuraciones locales complejas, facilitando a los usuarios el acceso a librerías especializadas como *scikit-learn*, *TensorFlow* y *PyTorch* (18). Además, Google Colab ofrece capacidades de cómputo acelerado mediante CPU y GPU, lo que optimiza el entrenamiento de modelos y el procesamiento de grandes volúmenes de datos geoespaciales. Su carácter colaborativo y reproducible favorece la transparencia y validación de los resultados obtenidos (19).

Bajo este contexto, el presente estudio tiene como objetivo aplicar técnicas de Machine Learning combinadas con imágenes satelitales para la clasificación del uso de suelo en el cantón Riobamba, la investigación busca generar un producto cartográfico confiable que refleje la distribución espacial de las principales clases de uso de suelo, contribuyendo así al fortalecimiento de los procesos de análisis territorial y proporcionando insumos técnicos para la planificación y gestión del territorio a nivel local.

MATERIALES Y MÉTODOS

La presente investigación adopta un enfoque cuantitativo, sustentado en el análisis y modelado de datos geoespaciales para la clasificación del uso de suelo en el cantón Riobamba. El estudio es de tipo aplicado, con un alcance explicativo y descriptivo, ya que busca identificar y representar la distribución espacial de las principales clases de uso y cobertura del suelo mediante técnicas de Machine Learning. El diseño metodológico es no experimental y transversal, dado que se analizan datos satelitales correspondientes a un periodo específico sin manipulación directa de las variables de estudio.

Para la recolección de datos se utilizó la plataforma Google Earth Engine para la descarga de imágenes satelitales Sentinel-2 de reflectancia superficial (Sentinel-2 SR), considerando un compuesto multispectral de varias bandas correspondiente al mes de agosto de 2025. Para reducir la influencia de nubosidad y asegurar continuidad espacial, se generó un compuesto estadístico con la mediana de las escenas disponibles. Adicionalmente, se emplearon datos vectoriales que corresponden a los límites del cantón Riobamba y un conjunto de puntos de entrenamiento y validación georreferenciados y clasificados en campo.

En los instrumentos se incluyen herramientas como: Sistemas de Información Geográfica y programación en Python con librerías especializadas como scikit-learn, rasterio y geopandas para la extracción de firmas espectrales, el entrenamiento del modelo SVM, la optimización de parámetros mediante GridSearchCV y la validación de resultados. Asimismo, La plataforma Google Colab fue utilizado como entorno de desarrollo (IDE) por su capacidad de procesamiento y reproducibilidad en la nube.

La población de estudio corresponde al territorio total del cantón Riobamba, mientras que la muestra estuvo conformada por 203 puntos de validación, distribuidos de manera equilibrada entre las siete clases de uso de suelo definidas, lo que permitió evaluar la precisión del modelo mediante la matriz de confusión y garantizar la confiabilidad de los resultados obtenidos.

RESULTADOS

El área de estudio corresponde al Cantón Riobamba, ubicado en la Provincia de Chimborazo, en la región interandina central del Ecuador. El cantón se caracteriza por un relieve predominantemente andino, con altitudes que superan los 2 700 m s. n. m., y por una marcada heterogeneidad físico-ambiental que incluye áreas urbanas consolidadas, zonas periurbanas, superficies agrícolas, pastizales y remanentes de vegetación natural. Riobamba presenta un crecimiento urbano progresivo asociado al incremento poblacional, la expansión de infraestructura y el cambio en las dinámicas productivas, lo que ha generado transformaciones significativas en el uso del suelo durante las últimas décadas (20). Estas características convierten al cantón en un escenario adecuado para la aplicación de técnicas de clasificación del uso de suelo mediante imágenes satelitales y Machine Learning, ya que la diversidad de coberturas y la complejidad espacial permiten evaluar el desempeño de los modelos en contextos urbanos y rurales propios de ciudades interandinas del Ecuador.

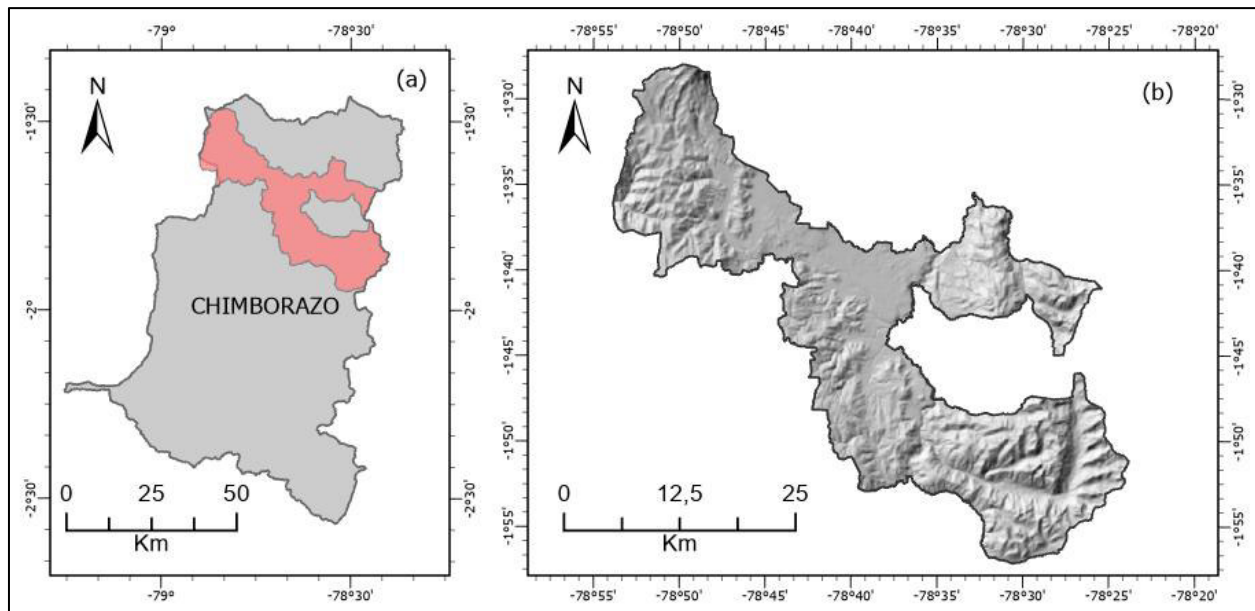


Figura 1. (a) Provincia de Chimborazo, (b) Cantón Riobamba.

La imagen satelital utilizada en este estudio corresponde a datos **Sentinel-2 de reflectancia superficial (Sentinel-2 SR)**, descargados desde la plataforma **Google Earth Engine**, considerando un **composite multispectral del mes de agosto de 2025** (21). Para optimizar el procesamiento y evitar pérdidas de cobertura espacial, se aplicó una simplificación geométrica del área de estudio y se generó un compuesto estadístico mediante la mediana de las escenas disponibles, reduciendo la influencia de la nubosidad. La imagen final integra bandas espectrales de **10 m y 20 m de resolución espacial**, re muestreadas a 10 m, incluyendo regiones del visible, infrarrojo cercano y de onda corta, lo que permite una adecuada discriminación espectral de las clases de uso y cobertura del suelo en los procesos de clasificación supervisada.

Código .js Parte 1

```
// =====  
// SENTINEL-2 OPTIMIZADO PARA DESCARGA LULC (rápido y estable)  
// - Simplifica geometría  
// - Usa composite median para evitar cortes y nubes  
// =====  
  
// 1) SHP del cantón (FeatureCollection)  
var canton = ee.FeatureCollection('users/tu_usuario/EC_Cantones_Riobamba');  
var geom = canton.geometry().dissolve().simplify(100);  
  
// 3) Para exportar sin cuelgues: usar rectángulo (bounds)  
var regionExport = geom.bounds();  
  
// verificar en mapa (muy ligero)  
Map.centerObject(geom, 10);  
Map.addLayer(ee.Image().paint(geom, 1, 2), {palette: 'red'}, 'Canton  
(simplificado)');  
Map.addLayer(ee.Image().paint(regionExport, 1, 2), {palette: 'blue'}, 'Bounds  
export (rectángulo)');  
  
// 4) Sentinel-2 SR (filtrado)  
var s2 = ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED')  
.filterBounds(geom)  
.filterDate('2025-08-07', '2025-08-09')  
.filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 30)); // 30% para tener más  
escenas  
  
print('Cantidad de imágenes:', s2.size());  
  
// 5) Composite (median) -> más estable que escoger una sola imagen  
// Nota: median reduce nubes residuales si hay varias escenas.  
var composite = s2.median().clip(geom);  
  
// 6) Bandas recomendadas para LULC (10 bandas)  
var bandsLULC = ['B2', 'B3', 'B4', 'B8', 'B5', 'B6', 'B7', 'B8A', 'B11', 'B12'];  
  
// Escalar reflectancia a [0..1]  
var s2_lulc = composite.select(bandsLULC).divide(10000);
```

Código .js Parte 2

```
// 7) Índices (opcionales, pero útiles)
var ndvi = s2_lulc.normalizedDifference(['B8', 'B4']).rename('NDVI');
var ndbi = s2_lulc.normalizedDifference(['B11', 'B8']).rename('NDBI');
var ndwi = s2_lulc.normalizedDifference(['B3', 'B8']).rename('NDWI');

// Stack final
var s2_stack = s2_lulc.addBands([ndvi, ndbi, ndwi]);

// 8) Visualización (solo RGB para no cargar demasiado)
Map.addLayer(composite, {bands:['B4','B3','B2'], min:0, max:3000}, 'RGB
composite');

// 9.1 Export LULC (10 bandas + índices)
Export.image.toDrive({
  image: s2_stack,
  description: 'S2_LULC_OPT_Riobamba_2025_08',
  folder: 'GEE_Exports',
  fileNamePrefix: 'S2_LULC_OPT_Riobamba_2025_08',
  region: regionExport,
  scale: 10,
  maxPixels: 1e13
});

// 9.2 Export SOLO 10 bandas (sin índices)
Export.image.toDrive({
  image: s2_lulc,
  description: 'S2_10B_OPT_Riobamba_2025_08',
  folder: 'GEE_Exports',
  fileNamePrefix: 'S2_10B_OPT_Riobamba_2025_08',
  region: regionExport,
  scale: 10,
  maxPixels: 1e13
});

// 9.3 Export RGB "foto"
Export.image.toDrive({
  image: composite.select(['B2', 'B3', 'B4']).divide(10000),
  description: 'S2_RGB_OPT_Riobamba_2025_08',
  folder: 'GEE_Exports',
  fileNamePrefix: 'S2_RGB_OPT_Riobamba_2025_08',
  region: regionExport,
  scale: 10,
  maxPixels: 1e13
});
```

El código implementado en Google Earth Engine permitió filtrar imágenes Sentinel-2 de reflectancia superficial para el mes de agosto de 2025, utilizando el área de estudio definida por el shapefile del cantón. Posteriormente, se generó un compuesto multispectral mediante la mediana de las escenas disponibles, se seleccionaron las bandas espectrales relevantes para clasificación y se exportó la imagen resultante en formato GeoTIFF para su análisis posterior.

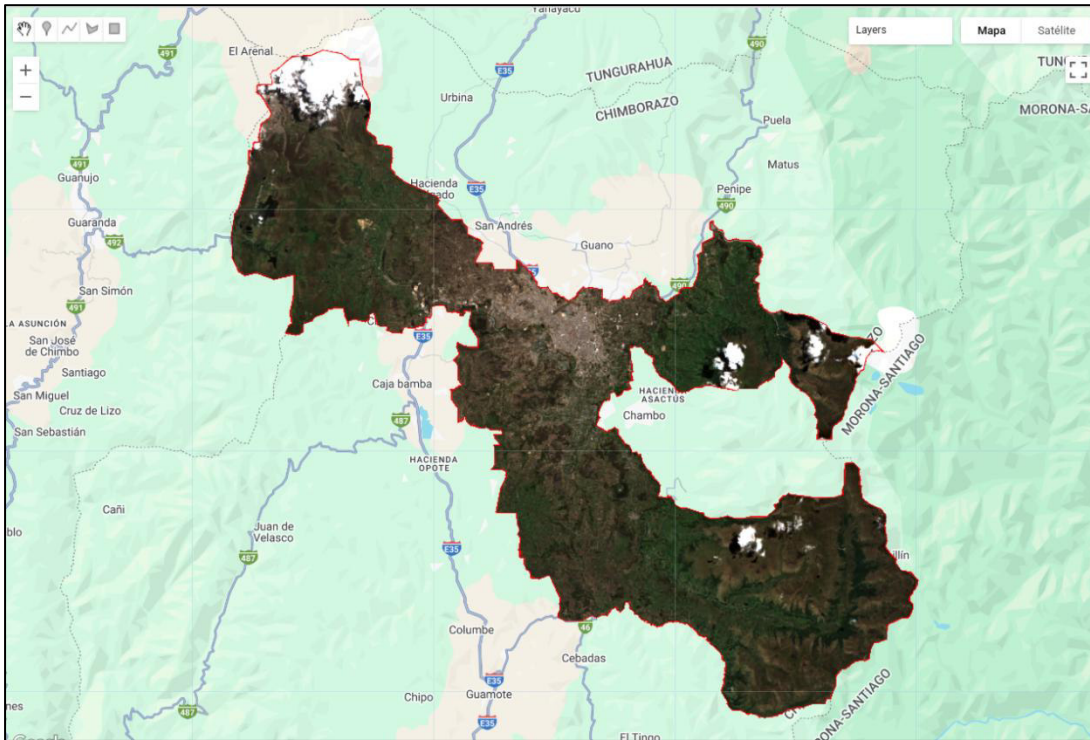


Figura 2. Resultado imágenes Sentinel-2 de reflectancia superficial para el mes de agosto de 2025.

Se define una tabla de clases de uso y cobertura del suelo para la clasificación de las categorías empleadas para el cantón Riobamba, considerando sus características físicas, ambientales y urbanas, estas clases permiten representar de manera diferenciada cuerpos de agua, vegetación, áreas urbanizadas y zonas degradadas, facilitando el análisis espacial y la interpretación de los resultados del modelo de clasificación.

Tabla 1. Clases de uso del suelo para la clasificación.

Clase	Nombre de la clase	Descripción general
1	Agua	Ríos, quebradas, lagunas, cuerpos de agua superficiales
2	Suelo desnudo	Zonas sin cobertura vegetal
3	Bosque	Cobertura densa de árboles, tanto bosque natural como plantaciones

Clase	Nombre de la clase	Descripción general
		forestales.
4	Vegetación baja / Pasto	Pastizales y cultivos de baja altura.
5	Construcción	Zonas consolidadas y urbanizadas con infraestructura como: edificaciones y vías
6	Hielo	Áreas con cobertura de nieve o hielo permanente específicamente en zonas de alta montaña.
7	Erosión	Zonas con procesos erosivos, cárcavas, taludes y áreas degradadas.

Utilizando las imágenes multiespectrales Sentinel-2 de diez bandas descargadas y preprocesadas, y el shapefile con los puntos de entrenamiento etiquetados con la clase a la que corresponde, se desarrolló en la plataforma Google Colab un modelo de clasificación supervisada mediante Máquinas de Vectores de Soporte (SVM). El código implementado en el lenguaje Python permite extraer firmas espectrales, entrenar el modelo y generar el mapa final de uso de suelo del cantón Riobamba con una precisión alta.

Configuración del modelo SVM

Código Python Parte 1

```
# === CONFIGURACIÓN GLOBAL ===
# Si quieres usar TODAS las bandas:
BAND_IDX = None

# Si quieres usar SOLO algunas bandas, escribe su número (1-based):
# BAND_IDX = [2, 4, 8] # Ejemplo: Blue (2), Red (4), NIR (8)

# === EXTRACCIÓN DE MUESTRAS ===

# Coordenadas de los puntos del shapefile
coords = [(x, y) for x, y in zip(gdf.geometry.x, gdf.geometry.y)]

# Extraer valores espectrales del raster en esas coordenadas
# Reopen the raster dataset
with rasterio.open(TIF_PATH) as src:
    samples = [list(src.sample([pt]))[0] for pt in coords]

X = np.array(samples)
```

Código Python Parte 2

```
# Número de bandas en el raster
n_bands = X.shape[1] # Get the number of bands from the sampled data
print("El raster tiene", n_bands, "bandas")

# Aplicar selección de bandas según configuración
if BAND_IDX is None:
    # X already contains all bands if src.sample was called without
    'indexes'
    print("👉 Usando todas las bandas")
else:
    # Select specific bands (adjusting for 0-based indexing)
    X = X[:, np.array(BAND_IDX)-1]
    print("👉 Usando solo las bandas:", BAND_IDX)
    n_bands = X.shape[1] # Update n_bands after band selection
    print("Número de bandas después de la selección:", n_bands)

# Etiquetas de clase
y = gdf[CLASS_COL_INT].values # Use CLASS_COL_INT here

print("Matriz de datos:", X.shape, "→ (n_muestras, n_bandas)")
print("Clases:", np.unique(y))
```

El código permite extraer firmas espectrales desde una imagen Sentinel-2 multibanda utilizando puntos de entrenamiento en formato shapefile, a partir de las coordenadas de los puntos, se muestrean los valores del raster con Rasterio, se organiza la matriz de características y se habilita la selección de todas o algunas bandas, preparando los datos para el entrenamiento del modelo SVM.

Código Python

```
# === GRID SEARCH SVM ===
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVC
# Definimos los hiperparámetros a probar
param_grid = {
    "C": [0.1, 1, 10],          # regularización
    "kernel": ["linear", "poly", "rbf"], # tipos de kernel
    "gamma": ["scale", "auto"]   # solo aplica a rbf/poly
}
# Configuramos la búsqueda
grid = GridSearchCV(
    SVC(),
    param_grid,
    cv=5,                      # validación cruzada en 3 particiones
    scoring="accuracy",        # métrica usada
    n_jobs=-1                   # usar todos los núcleos disponibles
)
# Entrenamos
grid.fit(Xtrain, ytrain)
```

Es importante definir los hiper parámetros, el código implementa un proceso de optimización del modelo SVM mediante GridSearchCV, evaluando distintas combinaciones de hiperparámetros como el coeficiente de regularización C, el tipo de kernel y el parámetro gamma, usando validación cruzada. Definir adecuadamente estos hiperparámetros es fundamental porque controlan la complejidad del modelo, el equilibrio entre sesgo y varianza y la capacidad de generalización, lo que permite maximizar la precisión y evitar el sobreajuste en la clasificación del uso de suelo.

Modelo SVM

En el presente estudio el modelo de clasificación se construyó mediante un flujo sistemático que integra preprocesamiento, optimización y selección del mejor clasificador SVM. Inicialmente, se gestionaron valores faltantes en los datos espectrales utilizando un imputador basado en la media, garantizando la consistencia entre los conjuntos de entrenamiento y prueba. Posteriormente, se definieron configuraciones específicas de hiperparámetros para distintos kernels de SVM (lineal, radial RBF y polinómico), considerando parámetros clave como el coeficiente de regularización C, el parámetro gamma y el grado del polinomio.

Para cada kernel, se aplicó un proceso de búsqueda en malla (GridSearchCV) con validación cruzada, permitiendo evaluar múltiples combinaciones y seleccionar automáticamente aquella con mejor desempeño. Los resultados obtenidos se almacenaron y compararon en función de la

precisión alcanzada en el conjunto de prueba, lo que permitió identificar el kernel con mayor capacidad de generalización.

El modelo SVM óptimo fue reentrenado utilizando todos los datos de entrenamiento y los hiper parámetros seleccionados, obteniéndose un clasificador robusto y listo para su aplicación en la clasificación del uso de suelo del cantón Riobamba a partir de imágenes multiespectrales Sentinel-2.

Código Python

```
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVC
import pandas as pd
from sklearn.impute import SimpleImputer # Import SimpleImputer
# Define the imputer
imputer = SimpleImputer(missing_values=np.nan, strategy='mean')
Xtrain_imputed = imputer.fit_transform(Xtrain)
Xtest_imputed = imputer.transform(Xtest)
# --- 1) Definir kernels y grids simples ---
param_grid = {
    'linear': {'C': [0.01, 0.1, 1, 10]},
    'rbf':     {'C': [0.1, 1, 10], 'gamma': ['scale', 0.01, 0.1, 1]},
    'poly':    {'C': [0.1, 1, 10], 'degree': [2, 3, 4], 'gamma': ['scale',
0.1]}
}
resultados = []
# --- 2) Probar cada kernel ---
for kernel, grid in param_grid.items():
    print(f"\n🌀 Probando kernel: {kernel}")
    # Definir modelo base
    svc = SVC(kernel=kernel, random_state=RANDOM_SEED)
    # Grid Search con CV=3 para rapidez
    clf = GridSearchCV(svc, grid, cv=3, scoring='accuracy', n_jobs=-1)
    # Use the imputed data for training
    clf.fit(Xtrain_imputed, ytrain)
    # Evaluar en TEST
    # Use the imputed data for testing
    test_acc = clf.best_estimator_.score(Xtest_imputed, ytest)
    resultados.append({
        "Kernel": kernel,
        "Mejores Parámetros": clf.best_params_,
        "Precisión CV": round(clf.best_score_, 3),
        "Precisión TEST": round(test_acc, 3)
    })
# --- 3) Convertir a DataFrame ordenado por TEST ---
df_resultados = pd.DataFrame(resultados).sort_values(by="Precisión TEST",
ascending=False)
print("\n=== RESULTADOS COMPARATIVOS ===")
```

Como resultado del proceso de clasificación supervisada con SVM se generaron tres archivos principales: El raster *clasificacion_svm.tif* contiene el mapa final de uso de suelo obtenido a partir de Sentinel-2, el archivo *comparacion_modelos.csv* con datos que resumen el desempeño de los kernels evaluados y su matriz de confusión, mientras que *svm_model.pkl* contiene el modelo SVM resultante entrenado para su utilización y validación.

En el archivo *clasificacion_svm.tif* se encuentra la representación espacial del uso de suelo del cantón Riobamba, este archivo es el resultado del proceso de clasificación supervisada mediante el algoritmo SVM. Este archivo en formato .tif es decir un ráster refleja la distribución de las diferentes clases de uso del suelo definidas en la **Tabla 1**, permitiendo visualizar su comportamiento espacial y su relación con el entorno urbano y rural. Los resultados de la clasificación pueden observarse de manera detallada en la **Figura 3**, donde se evidencia la diferenciación espectral entre las categorías establecidas.

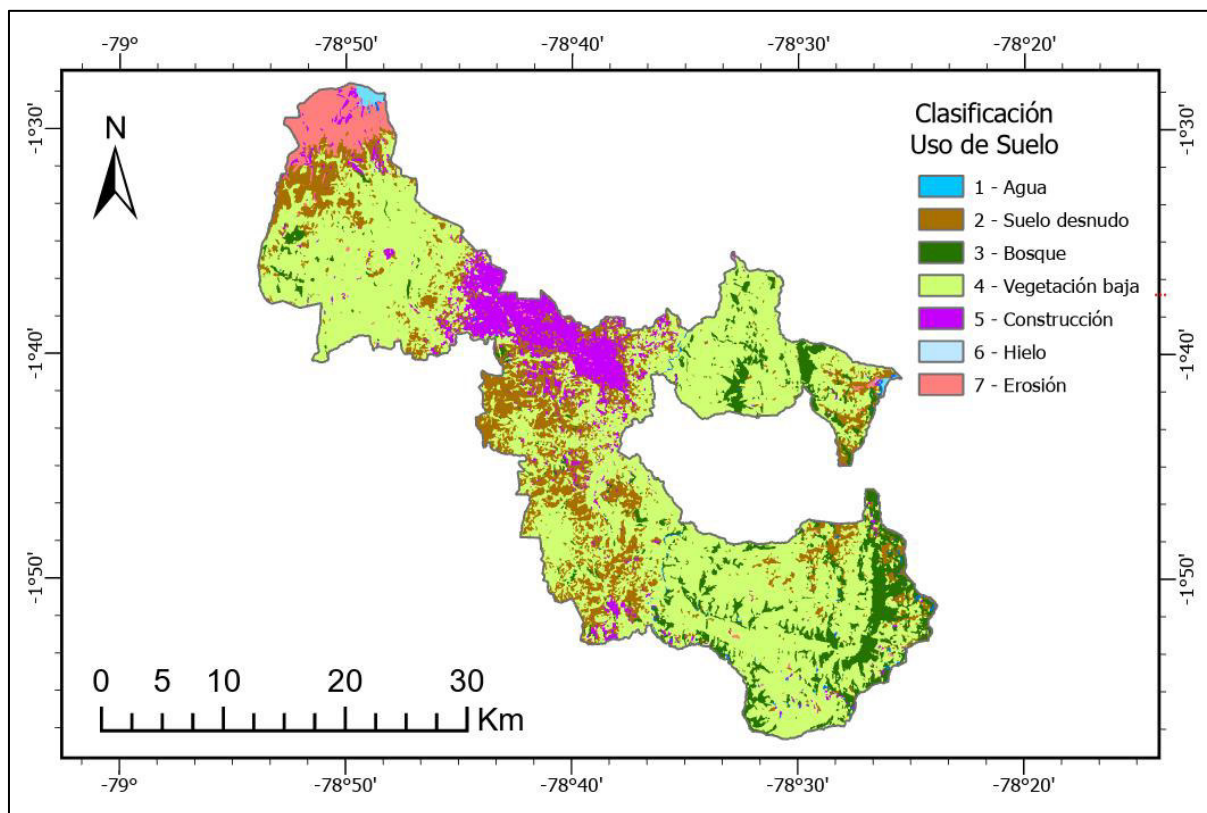


Figura 3. Clasificación de cobertura y uso de suelo del cantón Riobamba, modelo SVM.

La configuración del uso de suelo en el cantón Riobamba evidencia una marcada heterogeneidad espacial asociada a las condiciones topográficas, climáticas y antrópicas del territorio. Predominan las áreas de vegetación baja y bosque en las zonas rurales y de transición,

mientras que las superficies construidas se concentran principalmente en el núcleo urbano. La presencia de clases de hielo y erosión se localiza en los sectores de alta montaña, correspondientes al volcán Chimborazo y al Nevado El Altar, reflejando dinámicas naturales propias de ambientes andinos de gran altitud.

Validación

Tabla 2. Matriz de confusión del modelo SVM.

Clase real \ Clase predicha	Agua	Suelo desnudo	Bosque	Vegetación baja / Pasto	Construcción	Hielo	Erosión	Total	Exactitud del usuario
Agua	10	2	5	4	1	5	2	29	0,3448
Suelo desnudo	1	20	0	5	3	0	0	29	0,6897
Bosque	0	1	25	3	0	0	0	29	0,8621
Vegetación baja / Pasto	0	5	0	24	0	0	0	29	0,8276
Construcción	1	3	0	4	20	0	1	29	0,6897
Hielo	0	1	0	1	0	27	0	29	0,9310
Erosión	0	4	0	1	2	1	21	29	0,7241
Total	12	36	30	42	26	33	24	203	—
Exactitud del modelo SVM	0,8333	0,5556	0,8333	0,5714	0,7692	0,8182	0,8750	—	0,7241

La validación del modelo SVM constituye una etapa fundamental para evaluar la confiabilidad y el desempeño de la clasificación de uso de suelo, ya que permite contrastar los resultados obtenidos con datos independientes de prueba. En este estudio, la matriz de confusión representa una herramienta clave, pues cuantifica de manera detallada los aciertos y errores de clasificación para cada clase, permitiendo analizar tanto la exactitud del productor como la del usuario (22). A partir de los 203 puntos de validación, los resultados evidencian un nivel de desempeño global satisfactorio, con una exactitud general del 72,41 %, lo que confirma una adecuada capacidad del modelo para discriminar las diferentes categorías de uso de suelo. Este nivel de precisión respalda la validez del modelo SVM como una técnica robusta y confiable para el análisis territorial del cantón Riobamba.

DISCUSIÓN

Tabla 3. Distribución de las áreas de clasificación en kilómetros cuadrados y porcentajes.

**USO DE MACHINE LEARNING E IMÁGENES SATELITALES PARA LA CLASIFICACIÓN DEL USO DE SUELO EN EL CANTÓN
RIOBAMBA**

Código de clase	Clase	Área (km²)	Porcentaje (%)
1	Agua	1,38	0,14
2	Suelo desnudo	146,78	14,70
3	Bosque	94,13	9,43
4	Vegetación baja / Pasto	628,41	62,94
5	Construcción	87,23	8,74
6	Hielo	3,38	0,34
7	Erosión	37,16	3,72

Los resultados obtenidos del proceso de clasificación supervisada mediante el desarrollo del algoritmo SVM demuestra una apropiada representación del uso de suelo en el cantón Riobamba, reflejando las dinámicas naturales y las intervenciones antrópicas del territorio. Se observa que predomina la clase vegetación baja/pasto, que concentra el 62,94 % del área total, esto concuerda con la actividad agropecuario del cantón, donde amplias extensiones están destinadas a actividades agrícolas y ganaderas. Las clases bosque y suelo desnudo presentan una distribución significativa, asociada a zonas de forestación y reforestación que se impulsa en el cantón, además de la transición de altitud y a procesos de uso intensivo del suelo. Las áreas de construcción, con un 8,74 %, se encuentran principalmente en el núcleo urbano delimitado por la ciudad de Riobamba, evidenciando el proceso de expansión urbana característico de ciudades interandinas. Por otra parte, las clases hielo y erosión, con porcentajes reducidos, resultan fundamentales desde el punto de vista ambiental, ya que se localizan en sectores de alta montaña específicamente en el volcán Chimborazo y el Nevado El Altar, esto asegura la capacidad del modelo para distinguir coberturas espectralmente complejas. La validación del modelo se realiza mediante la matriz de confusión, con una exactitud global del 72,41 %, lo que respalda la confiabilidad de los resultados obtenidos, destacando altos niveles de precisión en clases como bosque, vegetación baja y hielo. En conjunto, estos resultados demuestran que el modelo SVM constituye una herramienta robusta para el análisis del uso de suelo, aportando información clave para la planificación territorial y la gestión sostenible del cantón Riobamba

CONCLUSIONES

El presente estudio confirma que la integración de técnicas de Machine Learning con imágenes satelitales Sentinel-2 lo que constituye una metodología sólida, eficiente y reproducible para la clasificación del uso de suelo en territorios con alta complejidad físico-ambiental y antrópica, como el cantón Riobamba. La aplicación del algoritmo Support Vector Machines (SVM), junto con el preprocesamiento multispectral y optimización de parámetros, permitió generar como resultado un mapa del uso de suelo del cantón Riobamba que representa de manera detallada la configuración espacial de su territorio.

Los resultados obtenidos reflejan un porcentaje mayor de la clase vegetación baja/pasto, que ocupa más del 60 % del área total, lo cual es coherente con el carácter agropecuario y periurbano del cantón. Asimismo, la distribución de las áreas construidas tiene patrones de concentración urbana definida, asociados a la consolidación de sectores urbanizados en virtud del crecimiento y expansión de la ciudad de Riobamba. Las clases de bosque y suelo desnudo presentan una presencia significativa en zonas rurales por la forestación y de transición, mientras que la identificación de hielo y erosión se encuentra en sectores de alta montaña, poniendo de manifiesto la capacidad del modelo para discriminar y distinguir coberturas complejas relacionadas con procesos naturales propios del entorno andino, especialmente en áreas asociadas con el volcán Chimborazo y el Nevado El Altar.

La validación del modelo se realizó mediante la construcción de la matriz de confusión con 203 puntos de prueba, dando como resultado una exactitud global del 72,41 %, lo que indica un desempeño adecuado y satisfactorio del clasificador SVM y respalda la confiabilidad de los resultados. Las métricas de exactitud del productor y del usuario evidencian un buen comportamiento en clases como bosque, vegetación baja y hielo, confirmando la estabilidad del modelo frente a la heterogeneidad espectral del área de estudio.

En conjunto, este trabajo demuestra que el uso de SVM, implementado en entornos computacionales abiertos como Python y Google Colab, representa una herramienta técnica eficaz para el análisis del uso de suelo, aportando información estratégica para la planificación territorial, la gestión ambiental y el diseño de políticas públicas orientadas al desarrollo sostenible del cantón Riobamba.

REFERENCIAS BIBLIOGRÁFICAS

1. Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, et al. Global Consequences of Land Use. *Science*. 22 de julio de 2005;309(5734):570-4.
2. Gong P, Wang J, Yu L, Zhao Y, Zhao Y, Liang L, et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *Int J Remote Sens*. 10 de abril de 2013;34(7):2607-54.
3. Liu W, Wu W, Thakuriah P, Wang J. The geography of human activity and land use: A big data approach. *Cities*. 1 de febrero de 2020;97:102523.
4. Heredia-R M, Torres B, Cabrera-Torres F, Torres E, Díaz-Ambrona CGH, Pappalardo SE. Land Use and Land Cover Changes in the Diversity and Life Zone for Uncontacted Indigenous People: Deforestation Hotspots in the Yasuní Biosphere Reserve, Ecuadorian Amazon. *Forests*. 8 de noviembre de 2021;12(11):1539.
5. Ayala-Izurieta J, Márquez C, García V, Recalde-Moreno C, Rodríguez-Llerena M, Damián-Carrión D. Land Cover Classification in an Ecuadorian Mountain Geosystem Using a Random Forest Classifier, Spectral Vegetation Indices, and Ancillary Geographic Data. *Geosciences*. 3 de mayo de 2017;7(2):34.
6. Universidad Técnica Estatal de Quevedo, Muñoz Marcillo JL, Cabrera Sinche B, Toulkeridis T. Analysis of land use change using remote sensing and GIS techniques in the Cazaderos forest, Loja province, Southern Ecuador. *Rev Geográfica Chile Terra Aust [Internet]*. 31 de

- diciembre de 2023 [citado 29 de diciembre de 2025];59. Disponible en: <https://www.revistaterraaustralis.cl/index.php/rgch/article/view/172>
7. Hwang T, Gholizadeh H, Sims DA, Novick KA, Brzostek ER, Phillips RP, et al. Capturing species-level drought responses in a temperate deciduous forest using ratios of photochemical reflectance indices between sunlit and shaded canopies. *Remote Sens Environ.* septiembre de 2017;199:350-9.
 8. Xie Y, Sha Z, Yu M. Remote sensing imagery in vegetation mapping: a review. *J Plant Ecol.* 1 de marzo de 2008;1(1):9-23.
 9. Maxwell AE, Warner TA, Fang F. Implementation of machine-learning classification in remote sensing: an applied review. *Int J Remote Sens.* 3 de mayo de 2018;39(9):2784-817.
 10. Wang M, Son S. VIIRS-derived chlorophyll-a using the ocean color index method. *Remote Sens Environ.* septiembre de 2016;182:141-9.
 11. Mountrakis G, Im J, Ogole C. Support vector machines in remote sensing: A review. *ISPRS J Photogramm Remote Sens.* mayo de 2011;66(3):247-59.
 12. Phinn S, Stanford M, Scarth P, Murray AT, Shyy PT. Monitoring the composition of urban environments based on the vegetation-impervious surface-soil (VIS) model by subpixel analysis techniques. *Int J Remote Sens.* enero de 2002;23(20):4131-53.
 13. Pal M, Mather PM. Support vector machines for classification in remote sensing. *Int J Remote Sens.* marzo de 2005;26(5):1007-11.
 14. Maxwell AE, Warner TA, Fang F. Implementation of machine-learning classification in remote sensing: an applied review. *Int J Remote Sens.* 3 de mayo de 2018;39(9):2784-817.
 15. Granger BE, Perez F. Jupyter: Thinking and Storytelling With Code and Data. *Comput Sci Eng.* 1 de marzo de 2021;23(2):7-14.
 16. Hall M. Mark Hall on Data Mining & Weka: CPython Integration in Weka [Internet]. Mark Hall on Data Mining & Weka. 2015 [citado 29 de diciembre de 2025]. Disponible en: <https://markahall.blogspot.com/2015/06/cpython-integration-in-weka.html>
 17. Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, et al. A view of cloud computing. *Commun ACM.* abril de 2010;53(4):50-8.
 18. Kluyver Thomas, Ragan-Kelley Benjamin, Pérez Fernando, Granger Brian, Bussonnier Matthias, Frederic Jonathan, et al. Jupyter Notebooks – a publishing format for reproducible computational workflows. En: *Positioning and Power in Academic Publishing: Players, Agents and Agendas* [Internet]. IOS Press; 2016 [citado 29 de diciembre de 2025]. Disponible en: <https://www.medra.org/servlet/aliasResolver?alias=iospressISBN&isbn=978-1-61499-648-4&spage=87&doi=10.3233/978-1-61499-649-1-87>
 19. Perkel JM. Why Jupyter is data scientists' computational notebook of choice. *Nature.* 1 de noviembre de 2018;563(7729):145-6.
 20. GoRaymi [Internet]. [citado 30 de diciembre de 2025]. Riobamba. Disponible en: <https://www.goraymi.com/es-ec/chimborazo/riobamba/guia-turismo-d894e5>
 21. Colecciones de Sentinel en Earth Engine | Earth Engine Data Catalog | Google for Developers [Internet]. [citado 30 de diciembre de 2025]. Disponible en: <https://developers.google.com/earth-engine/datasets/catalog/sentinel?hl=es-419>
 22. Congalton RG. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens Environ.* julio de 1991;37(1):35-46.

