

Evaluación de métodos de clasificación supervisada para la estimación de cambios espacio-temporales de cobertura en los páramos de Merchán y Telecom, Cordillera Oriental de Colombia

Yoan Poveda-Sotelo^{1*} ; Mauricio A. Bermúdez-Cella² ;
Pablo Gil-Leguizamón³ 

¹Maestría en Ciencias de la Tierra, Universidad Pedagógica y Tecnológica de Colombia, Sogamoso, Colombia. (*) yoan.poveda@uptc.edu.co

²Escuela de Ingeniería Geológica, Universidad Pedagógica y Tecnológica de Colombia, Sogamoso, Colombia. mauricio.bermudez@uptc.edu.co

³Escuela de Ingeniería Ambiental, Universidad Pedagógica y Tecnológica de Colombia, Tunja, Colombia. pablo.gil@uptc.edu.co

Resumen

En los últimos años ha existido un avance significativo en los Sistemas de Información Geográfica (SIG) y en el desarrollo de métodos de clasificación supervisada, hasta ahora estos no habían sido utilizados para calcular con exactitud la extensión superficial de páramos en sectores de la Cordillera Oriental de Colombia y mucho menos para estimar la distancia entre los límites de esos páramos y los principales rasgos geológicos. Por esta razón, en la presente investigación se evaluaron cinco métodos de clasificación supervisada, con el propósito de determinar cuál de estos posee una mayor resolución para reproducir la extensión y distribución superficial de los páramos de Merchán y Telecom en Saboyá, Boyacá, pertenecientes al Complejo de Páramos Iguaque - Merchán en la Cordillera Oriental de Colombia. Con esta finalidad, se escogieron imágenes satelitales del área de estudio por medio de Landsat 8 para el año 2018 y se clasificaron utilizando algunos algoritmos basados en *Machine Learning* (SVM, RF, DT, BC y ANN). Para establecer la exactitud y confiabilidad de los datos de clasificación de las características del terreno se calculó el Índice Kappa, que permitió determinar que el método más preciso para este caso fue RF. Adicionalmente, dado que los límites de los páramos coinciden con estructuras geológicas o contactos entre formaciones, se estimó la distancia entre el borde de los páramos y esos rasgos. Los resultados obtenidos en esta investigación son considerados como insumo para futuros análisis multitemporales, y estimación de distintas sirven como herramienta para la elaboración y toma de decisiones en la gestión de recursos naturales, biodiversidad, prestación de servicios ecosistémicos, y ordenamiento territorial para el municipio de Saboyá-Boyacá.

Palabras clave: Clasificación supervisada; Índice Kappa; Machine Learning; Páramo; Teledetección.

Forma de citar: Poveda-Sotelo, Y.; Bermúdez-Cella, M.A.; Gil-Leguizamón, P. (2022). Evaluación de métodos de clasificación supervisada para la estimación de cambios espacio-temporales de cobertura en los páramos de Merchán y Telecom, Cordillera Oriental de Colombia. *Boletín de Geología*, 44(2), 51-72. <https://doi.org/10.18273/revbol.v44n2-2022002>

Evaluation of supervised classification methods for the estimation of spatiotemporal changes in the Merchán and Telecom paramos, Colombia

Abstract

In recent years there has been significant progress in Geographic Information Systems (GIS) and the development of supervised classification methods, but until now these had not been used to accurately calculate the surface extent of paramos in sectors of the Eastern Cordillera of Colombia. Furthermore, these methods had not been used to estimate the distance between the boundaries of these moors and major geological features. For this reason, in the present research, five different supervised classification methods were evaluated, with the purpose of determining which of them has a higher resolution in order to reproduce the extension and surface distribution of the paramos of Merchán and Telecom in Saboyá, Boyacá, belonging to the Merchán - Iguaque complex in the Eastern Cordillera of Colombia. For this purpose, satellite images of the study area by Landsat 8 for the year 2018 were chosen and classified into some algorithms based on Machine Learning (SVM, RF, DT, BC and ANN). To establish the accuracy and reliability of the classification data of the terrain features, the Kappa Index was calculated, which allowed determining that the most accurate method for this case was Random Forest. In addition, since the boundaries of the moorlands coincide with geological structures or contacts between formations, the distance between the edge of the moorlands and these features was estimated. The results obtained in this research are considered as an important input for future multitemporal analysis as in landscape metrics, which serve as a tool for the development and decision making in the management of natural resources, biodiversity, provision of ecosystem services, as in the land use planning for the municipality of Saboyá-Boyacá.

Keywords: Supervised classification; Kappa Index; Machine Learning; Paramo; Remote sensing.

Introducción

En el neotrópico, los ecosistemas de alta montaña (páramo y bosque altoandino) apoyan la regulación climática transformando el CO₂ en O₂ (Torres *et al.*, 2012); razón que determina la necesidad de su conservación. Estos ecosistemas recientes (Holoceno), presentan un interés por las comunidades locales para la realización de actividades como la agricultura itinerante, el pastoreo, la quema, las plantaciones exóticas (*Pinus patula* y *Eucalyptus globulus*) y la tala con rigor; estas actividades provocan alteraciones en la estructura ecológica de las comunidades vegetales (Cleef, 2008). Vale la pena mencionar que cerca del 99% de los páramos del planeta están ubicados en la Cordillera de los Andes (Garavito-Rincón, 2015).

Esta investigación se centra en dos páramos ubicados en la Cordillera Oriental de Colombia, específicamente, los páramos de Merchán y Telecom en el municipio de Saboyá (Complejo de Páramos Iguaque - Merchán, Boyacá), con el propósito de evaluar métodos de clasificación supervisada e identificar cuál de éstos permite una mejor discriminación de la distribución espacial del páramo.

Estos métodos de clasificación que hasta ahora no se habían implementado en el área de interés, sin embargo, se cuentan con aportes para otras localidades en el mundo, como reseñan Manna *et*

al. (2013), quienes evaluaron cambios de bosque de mangle en la isla Jharkhali (Sundarbans, India) entre el 2004 y 2010 con datos satelitales multitemporales y de vegetación a partir del Índice de Vegetación de Diferencia Transformada (TDVI por sus siglas en inglés). Este último determinó un resultado favorable con aumento de la cubierta vegetal gracias a los esfuerzos de conservación. En Egipto (Al-Kharj), Aly *et al.* (2016), con imágenes Landsat, determinaron un aumento de la cobertura vegetal (CV), en ecosistemas áridos (107,4% entre 1987 y 2000) y disminución de la misma (27,5% entre el 2000 y 2013), producto del deterioro y salinización del suelo y el agua.

En México, De León-Mata *et al.* (2014) estudiaron la composición de la cobertura vegetal entre los años 2008 y 2011 para el área prioritaria para la conservación de pastizales conocida como Cuchillas de la Zarca, estimaron una degradación lenta (-661 ha anuales de chaparral y -795,6 ha de bosques de encino - pino) debido a la existencia de grandes parches. En Sinaloa (México), Sanhouse-García *et al.* (2017) entre 1990 - 2000 identificaron una tasa de cambio durante ese periodo de 52% en la superficie total inicial y de 41% entre 2000 - 2014 por causa de la ocupación del territorio por el hombre. En Brasil (Paulo Assis Ribeiro, Colorado), dos Santos *et al.* (2015) determinaron entre 1991 - 2010, transiciones de bosque en pasto en un 70%. En Chile, Sepúlveda-Varas *et al.* (2019) estudiaron el

cambio de cobertura y uso del suelo entre 1994 y 2007 con el índice de Kappa, concluyeron mayor presión de uso en las categorías cultivos, praderas naturales y plantaciones forestales.

En Colombia, Muñoz-Guerrero *et al.* (2009) evaluaron la dinámica de cambio entre 1989 y 2008 en la microcuenca Las Minas (Pasto, Nariño), con imágenes Landsat10TM e IKONOS; los resultados permiten discriminar la reducción del bosque secundario (58,51 ha) con una tasa de deforestación de 3,08 ha/año, un aumento de áreas destinadas para la siembra (141,64 ha) por presiones de deforestación y un aumento de actividad bovina y pecuaria.

Por su parte, en algunos páramos de Colombia Arellano y Rangel (2008), identificaron áreas en la Serranía de Perijá y el Parque Nacional Natural Los Nevados (PNN) con favorable estado de conservación y conectividad.

Actualmente, los entes gubernamentales y la autoridad ambiental en el páramo de Merchán y Telecom no cuentan con instrumentos suficientes de planificación, vigilancia y control que permitan determinar los impactos naturales de la actividad humana en la zona. Por otra parte, la investigación biótica relacionada en las áreas del Complejo de Páramos Iguaque - Merchán es incipiente, y se requiere información sobre la dinámica espaciotemporal del páramo, el bosque altoandino y la influencia de la matriz agrícola sobre estos (Cortés-Duque y Sarmiento-Pinzón, 2013; Velasco-Linares *et al.*, 2015).

Por lo anterior, el presente estudio da respuesta a las siguientes interrogantes: ¿Qué método de clasificación supervisada permite estimar de forma confiable la dinámica de cambio de las coberturas vegetales presentes en el páramo de Merchán y Telecom? ¿Tienen algunos de estos métodos la capacidad para discriminar los rasgos geológicos que ejercen algún tipo de influencia sobre los páramos?

Marco teórico

Teledetección

Los estudios de dinámicas espaciales y temporales se apoyan en la geografía cuantitativa bajo un sistema lógico matemático y deductivo (Padrón, 2007). La investigación se enfoca en el estudio del territorio (González, 2011), como representación del espacio

sometido a una transformación continua, como resultado de la acción social del hombre (Llanos-Hernández, 2010).

El cambio se identifica con herramientas de teledetección que de acuerdo con Chuvieco (1995) y Di Somma *et al.* (2010) poseen las siguientes ventajas:

1. Permite adquirir información de la superficie terrestre a partir de sensores instalados en plataformas espaciales y la energía electromagnética.
2. Las imágenes de satélite permiten abordar análisis multitemporales, entendidos como cambios espaciales debido a presiones agrícolas y pecuarias, que acarrear modificaciones físicas (aumento o disminución de cobertura de estudio) y geográficas (aumento o disminución área de cada clase), en un periodo de tiempo.
3. Proporciona una visión de la evolución del territorio y ofrece instrumentos para una correcta lectura de los orígenes del paisaje actual y sus posibles desarrollos futuros.

Algoritmos de clasificación supervisada

Existen herramientas predictivas que clasifican datos basados en algoritmos que describen patrones. Estos algoritmos de clasificación supervisada tienen aplicaciones medioambientales, sociales, políticas, entre otras (Riquelme *et al.*, 2006).

Uno de ellos, es el algoritmo Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés), método muy utilizado debido al amplio rango de datos que permite procesar; además, identifica con exactitud las coberturas del suelo y permite visualizar la separación entre estas (Cervantes-Canales, 2009; Murillo-Castañeda, 2018; Parra, 2017).

Un segundo método son las Redes Neuronales Artificiales (ANN, por sus siglas en inglés), que utilizan muestras establecidas (entradas de entrenamiento) cuya cantidad y calidad de datos son indispensables para la red neuronal (Moya-Fuero, 2012). Las ANN, permiten determinar patrones que en varios casos se escapan de la observación humana (Becerra-Serial *et al.*, 2009). El método clasifica imágenes satelitales de forma automática por medio de muestras o puntos de control, dados por interpretación visual (Jaramillo y Antunes, 2018; Lanzarini, 2004).

El tercer método, el cual ha sido muy aceptado por la capacidad de integrar datos de diversas fuentes y calidad de los resultados, es Bosques Aleatorios (RF, por sus siglas en inglés) (Sánchez-Pellicer *et al.*, 2017). Este algoritmo es un clasificador de conjunto que produce múltiples árboles de decisión, conformado por un subconjunto de muestras al azar y variables de entrenamiento; una de sus ventajas es su alta precisión en la clasificación (Belgiu y Drăguț, 2016). En este último método, distintos árboles analizan distintas porciones de los datos; de esta manera hace que cada árbol se entrene de distintas muestras de datos para un mismo problema. Así, al combinar los resultados, unos errores serán compensados con otros, y se obtendrá un mejor resultado (Paredes-Inilupú, 2020).

Un cuarto método es Árboles de Decisión (DT, por sus siglas en inglés), tiene la bondad de que permite interpretar fácilmente las decisiones tomadas por el modelo, revela formas complejas de la estructura de datos que no se pueden detectar con métodos convencionales de regresión. A partir de una base de datos, crea diagramas de construcciones lógicas, técnica explicativa y descomposicional basada en procesos de división secuencial, iterativa y descendiente. La manera de obtener árboles de decisión binarios, es decir que cada nodo se divide en exactamente dos ramas, es por medio de la técnica CART (*Classification And Regression Trees*) diseñado por Breiman *et al.* (1984) (según Friedl y Brodley, 1997). Este modelo es considerado una herramienta que complementa técnicas estadísticas como la regresión múltiple, análisis de varianza, regresión logística, modelos log-lineales (De'Ath y Fabricius, 2000).

Finalmente, un quinto método corresponde a la clasificación de Bayes (BC, por sus siglas en inglés), conocido por su simplicidad y rapidez computacional. Permite predecir la probabilidad de diferentes resultados con base al Teorema BC (González y Rodríguez, 2006).

En síntesis, estos algoritmos de clasificación supervisada permiten medir el cambio de uso del suelo, direccionado a la interpretación de la pérdida o ganancia de cobertura vegetal, que se puede ver reflejado en la fragmentación de bosques, pérdida de biodiversidad, disminución de la conectividad ecológica, o, por el contrario, la recuperación a través de la sucesión y la resiliencia ecosistémica (Lozada, 2007).

Marco geológico

La zona de estudio se encuentra en la Cordillera Oriental de Colombia, específicamente en los límites entre los departamentos de Santander y Boyacá, esta corresponde a los páramos de Merchán y Telecom.

El área de interés se caracteriza por un sistema de anticlinales y sinclinales que pertenecen al llamado Sinclinorio de Chiquinquirá (Fuquen y Osorno, 2005), un bloque estructural que se encuentra delimitado al oriente por la Falla Pedro Gómez y al occidente por la Falla Carupa, pero que en el sector considerado con esta investigación es llamada Falla Peña Blanca (Figura 1), la cual ha sido reconocida en la plancha 170 (Fuquen y Osorno, 2005). En el sector sur de la zona de estudio, yace la Falla Garavito, una estructura que ha sido interpretada a partir de las fotografías aéreas de la zona. El área comprendida en la presente investigación presenta más plegamientos asociados a los procesos compresivos que edificaron la Cordillera Oriental de Colombia, que fallamientos. A lo largo de este bloque, las estructuras son asimétricas, y estrechas con ejes que poseen una dirección entre N30E a N40E. Los principales rasgos estructurales lo constituyen las fallas Carupa, Arboledas, El Roble, Peña Blanca, Garavito, y el sinclinal de Aposentos - Chiquinquirá (ver Figura 1).

La Falla de Carupa es considerada una falla de cabalgamiento con vergencia al occidente que afecta los estratos de las formaciones de edad Cenomaniano-Turoniano: Churuvita (predominantemente areniscas), Simijaca (unidad lutítica-limosa, en cuya base se encuentran delgadas capas de caliza) y Frontera (secuencia de limolitas silíceas). La Falla de Peña Blanca cartografiada por Ulloa y Rodríguez (1978) es considerada como una falla dextral con salto inverso con una dirección de N40-45°E, su trazo es rectilíneo y se infiere que el plano de esta falla es sub-vertical inclinado al NW con vergencia al SE. La falla deforma las rocas de la Formación Conejo de edad Coniciano-Santoniano, constituida principalmente por arcillolitas negras a grisáceas, con delgados niveles de areniscas y calizas, su prolongación hacia el sur es considerada la continuación de la Falla de Carupa y está cubierta por depósitos cuaternarios en casi toda su extensión y su traza queda cubierta por depósitos lagunares al noreste de Susa (Fuquen y Osorno, 2005).

La Falla Arboledas reconocida por Terraza (2004), tiene carácter sinistral extendiéndose con dirección N50°W, debido a su trazo rectilíneo se ha inferido (Fuquen y Osorno, 2005) que el plano de falla podría inclinarse subverticalmente al NE, sin embargo, no se descarta algún tipo de desplazamiento inverso local a lo largo de su extensión. Terraza (2004) estima un desplazamiento de 1 a 2 km de la falla, su expresión geomorfológica es de un escarpe prominente que se extiende hasta la Falla de Peña Blanca.

La Falla El Roble es considerada como una falla inversa convergencia al SE (Terraza, 2004) la cual controla el curso medio y alto de la quebrada El Roble. La dirección del plano de falla es al noreste con buzamiento hacia el noroeste. Posiblemente la estructura se haya generado por deslizamiento flexural entre rocas competentes e incompetentes durante el plegamiento que dio origen al sinclinal de Aposentos - Chiquinquirá tomando como superficie de despegue los niveles arcillosos por debajo de la Formación Churuvita.

A lo largo de la zona, existen otras estructuras como el sinclinal de Aposentos - Chiquinquirá, el cual fue definido por Ulloa y Rodríguez (1978) como sinclinal de Chiquinquirá - San José de Pare. Esta estructura de tipo simétrico ocupa toda la zona de estudio, y está delimitada por las rocas de la Formación Tablazo (conformada por una intercalación de lutitas, calizas y areniscas cuarzosas de edad Aptiano Superior a Albiano Inferior) en sus flancos y por rocas de la Formación Plaeners (arcillas grises oscuro a amarillas de edad Campaniano superior) en su núcleo. Según Fuquen y Osorno (2005) el eje de este sinclinal tiene una dirección general N35°E y se extiende desde el suroeste de Carupa conectando con otras estructuras. Hacia los flancos del sinclinal de Aposentos - Chiquinquirá se desarrollaron varios pliegues que son interpretados por Terraza (2004) como el resultado de la acomodación por deslizamiento flexural entre capas de distinta resistencia mecánica, resultando en el plegamiento de la Formación Churuvita y las unidades supra e infrayacentes, con superficies de charnela que presentan diferente orientación.

El páramo de Merchán se extiende entre ellos anticlinales de Río Piedras y Jesús María (Figura 1), y está limitado por un contraste litológico. En contraste,

el páramo de Telecom es parte del flanco oriental del Anticlinal de Tinjacá - Oiba (Figura 1).

El páramo de Telecom se encuentra sobre el bloque colgante generado por la Falla de Garavito. Por medio de la cartografía y cálculos realizados para el presente estudio, se observa que estos páramos juntos con otros más pequeños de Santander comprenden un área de 16.212 ha con alturas entre 3.150 y 3.850 msnm (Figura 2). En este complejo se distribuyen ecosistemas de páramo, subpáramo y bosques altoandinos (Morales-Rivas *et al.*, 2007). En la zona de estudio se originan los principales tributarios del río Suárez para este sector de la Cordillera Oriental (Figura 2).

Metodología

Basados en los métodos descritos en la segunda sección, se realizó una secuencia de pre y posprocesamiento a imágenes de satélite del sensor Landsat 8, apoyado en software ERDAS para ortorrectificación, QGIS para algunas herramientas de preprocesamiento como recorte, combinación de bandas, algunos complementos de código abierto y libre acceso como Algoritmo de Semiclasificación Automática (SCP, por sus siglas en inglés) creado por Luca Congedo (Bravo-Morales, 2020) y RStudio, para hacer las clasificaciones supervisadas a según los cinco métodos basados en *Machine Learning*.

Es importante mencionar, que esta investigación se realizó durante la emergencia sanitaria generada por la pandemia de COVID-19, por lo cual no fue posible tener acceso presencial a la zona de estudio, por esta razón la asignación de los puntos de control como insumos para todos los métodos de clasificación, se hizo con algunos puntos en campo tomados previamente a la emergencia sanitaria, imágenes satelitales de alta resolución espacial (geo portales) y *Pansharpening* en procesamiento de la imagen Landsat 8 bajo el SCP (para el año 2018 se aplicaron los dos últimos criterios); la resolución de 30 x 30 m se redujo a 15 x 15 m, de esta forma se permite una mayor visualización y mejor conocimiento de la zona de interés. De igual forma se tomó Landsat 8 por su resolución espacial y año 2018, debido a que años posteriores se presentaban dificultades de visualización de coberturas por la alta presencia de nubosidad.

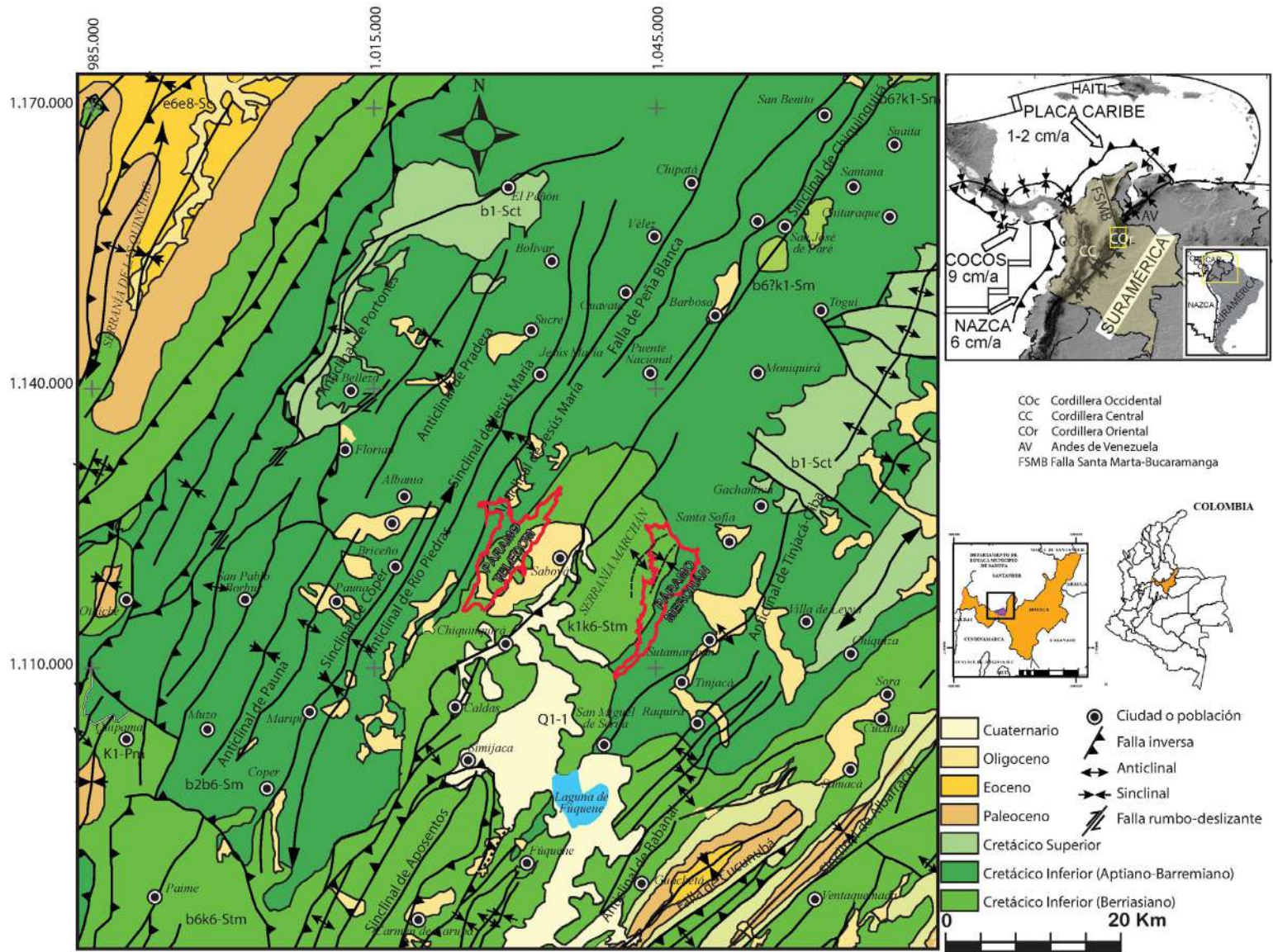


Figura 1. Mapa geológico de la zona de estudio (modificado de Ulloa y Rodríguez, 1978).

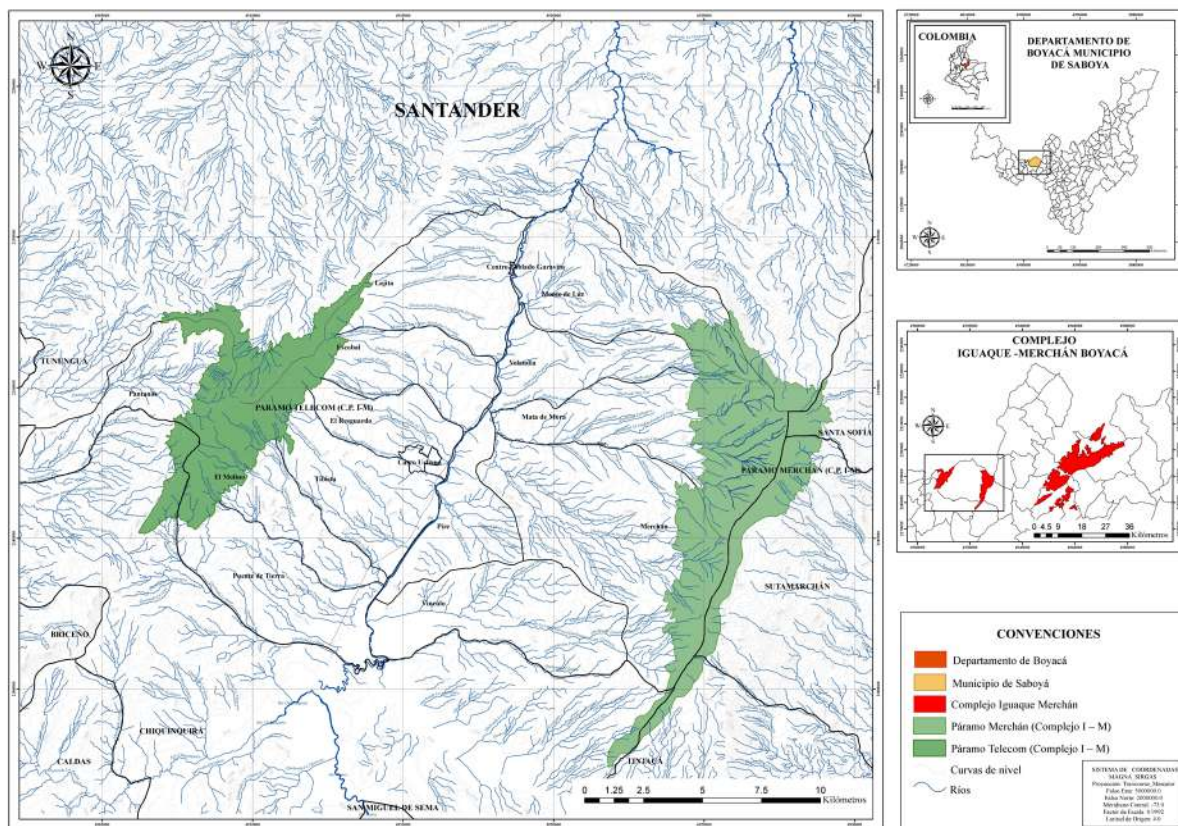


Figura 2. Ubicación geográfica de la zona de estudio y principales sistemas de drenajes, ambos páramos poseen tributarios del río Suárez.

Obtención de las imágenes satelitales de la zona de estudio

Se utilizaron imágenes satelitales del sensor Landsat 8, con una resolución espacial de 30 m y Path/row 08/56. Estas imágenes fueron obtenidas del visor Glovis del Servicio Geológico de los Estados Unidos (USGS),

que contaron con nubosidad <15% y cubrieron la extensión de la zona de estudio para el trimestre enero – marzo de 2018, algunas características de resolución son reflejadas en la [Tabla 1](#) (Chuvieco, 2002).

Tabla 1. Resoluciones del sensor Landsat 8.

Resolución espacial	Se asigna al objeto más pequeño que puede ser distinguido sobre una imagen, por otro lado, corresponde al tamaño de la mínima unidad de información presente en la imagen, denominada píxel.
Resolución espectral	Número y ancho de bandas que puede discriminar el sensor (cuanto más estrechas sean estas, mayor es la misma).
Resolución radiométrica	Sensibilidad del sensor, es decir, se enfoca en su capacidad para detectar variaciones en la radiancia espectral que percibe. Indica número de niveles de gris y varía según el sensor (Landsat 4 y 7 con 8 bits, 28=256 ND; Landsat 8 con 16 bits, 216 = 65536 ND).
Resolución temporal	Frecuencia de cobertura que proporciona el sensor, es decir, la periodicidad con la que éste adquiere imágenes en la superficie terrestre (Landsat 7 y 8 es de 16 días).
Resolución angular	Capacidad de un sensor para observar la misma zona de diferentes ángulos.

Preprocesamiento de imágenes satelitales

Combinación de bandas. Se realizó la combinación R:4, G:3, B:2 según especificaciones técnicas del sensor Landsat 8 con el fin de obtener una mejor visualización y diferencias de coberturas.

Recorte de la imagen a partir del polígono de referencia. Luego de la ortorrectificación se realizó recorte de cada imagen correspondiente al área de estudio de los páramos Merchán y Telecom (sistema de coordenadas WGS 84 UTM Zone 18N, código 32618), por medio del SCP de QGIS, durante el preprocesamiento se hicieron recortes de bandas, agregando el *shapefile* en primer caso para el páramo de Telecom y posteriormente páramo de Merchán.

Mejoramiento espacial

Para mejorar la resolución y la visualización de la zona de estudio se aplicó la herramienta SCP en QGIS y poder visualizar a mejor detalle la zona de estudio.

Categorización temática de imágenes

Para determinar las coberturas de las zonas de páramo se tomaron puntos de control, es decir, coordenadas de interés las cuales fueron la base para que a través de métodos de *Machine Learning* (SVM, RF, entre otros) se generara una clasificación temática en cada una de las imágenes multispectrales.

La clasificación digital de las imágenes Landsat 8, se realizó a través de una clasificación supervisada (lenguaje de programación R) de la siguiente manera:

1. Creación y delimitación de áreas de entrenamiento: donde se efectuó la definición digital de las categorías, es decir, se define la cantidad total de distintas cubiertas definidas según el IDEAM (2010).
2. Asignación: consistió en la categorización de los píxeles de la imagen, de esta manera se asignaron los píxeles específicos para cada tipo de cobertura.

Para que la clasificación sea satisfactoria por medio del lenguaje de programación R, la imagen satelital (preprocesada), debe presentarse de forma rectangular para evitar errores al correr los paquetes y librerías previamente instaladas.

Calibración de un modelo de clasificación supervisada

Se efectuó a través del lenguaje de programación R, este incorporó una serie de herramientas estadísticas robustas (regresión logística, análisis discriminante) y de inteligencia artificial (SVM, DT, BC) las cuales tienen la capacidad de realizar tareas propias de clasificación. Para realizar estas últimas se empleó un subconjunto de datos de prueba (entrenamiento) los cuales fueron usados para comprobar el comportamiento del modelo estimado.

Dentro del lenguaje de programación R, se utilizó el paquete *nnet* redes neuronales con capa oculta, trabajando con dos hiperplanos: uno es el número de neuronas en la capa oculta (*size*) y otro que controla la regulación durante el entrenamiento de la red (*decay*), arrojando una clasificación supervisada.

Para dicha clasificación, se debe hacer lo siguiente:

1. Instalación de algunos paquetes como *sp*, *raster*, *RStoolbox*, *ggplot2*, *reshape*, *rpart.plot*, *snow*, *rpart*, *sf*, *nnet*
2. Se activan las librerías de los anteriores paquetes, además de usar *suppressMessages* para evaluar las variables en un solo contexto, omitiendo mensajes de diagnóstico “simple”
3. Los valores seleccionados de la imagen satelital junto con el *shapefile* de los páramos se incorporan en el paquete *nnet* de R con los siguientes ítems:
 - a. *Size*: Número de unidades en la capa oculta (50)
 - b. *Data*: Marco de datos en este caso se tomó una muestra de 200 puntos de control, donde el 50% es tomado para ser validado por medio de la función *simple* en R
 - c. *Rang*: Pesos iniciales, se tomó de 0,1
 - d. *Decay*: Parámetro de caída de peso, se debe tomar un valor muy pequeño en este caso 40.0005
 - e. *Maxit*: Máximo número de iteraciones: 200

La Figura 3 resume la metodología utilizada en la presente investigación en forma de diagrama de flujo.

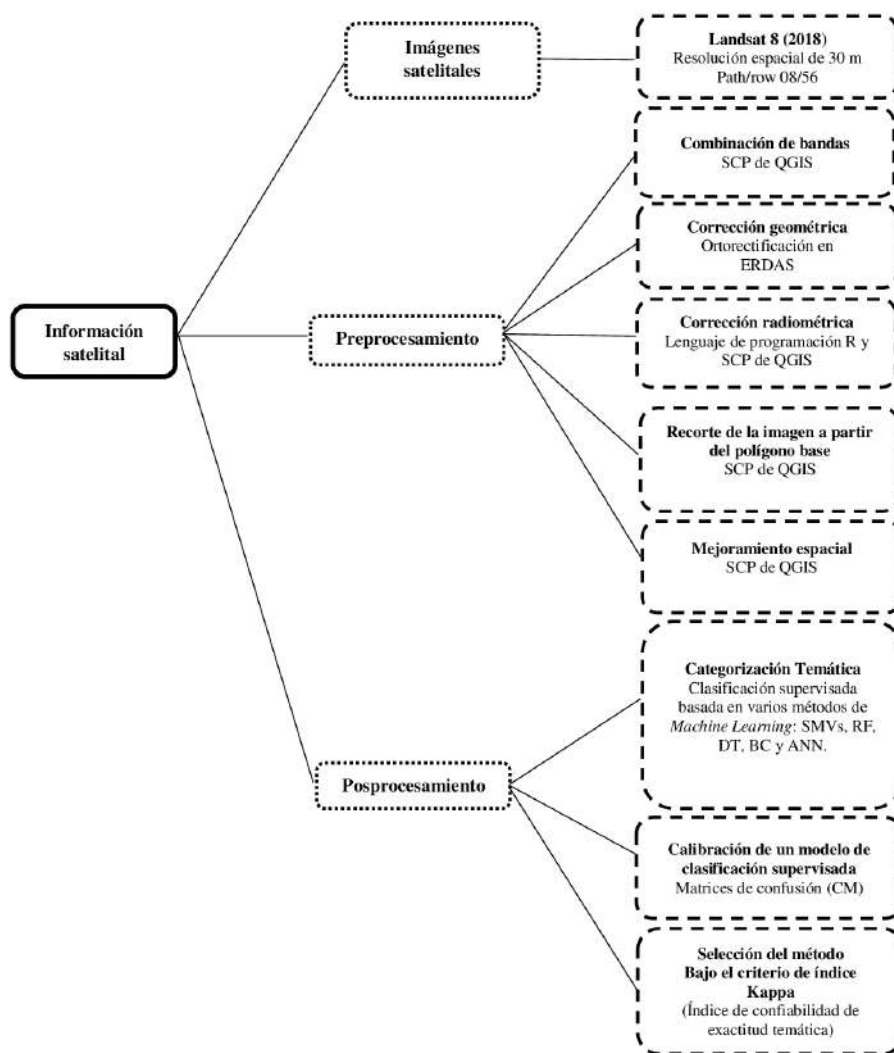


Figura 3. Diagrama de flujo que resume la metodología utilizada en la presente investigación.

Resultados

La [Figura 4](#) resume las distribuciones espaciales de los páramos Merchán y Telecom mediante los métodos de clasificación supervisada.

Discriminar visualmente cuál de los métodos de clasificación es más efectivo resulta una tarea compleja, por esta razón se calculó para cada uno de ellos el índice Kappa ([Tabla 2](#)), el cual se define como: $K = (Po - Pe) / (1 - Pe)$. Siendo: $Po = \text{Número de aciertos} / \text{Número total de clases}$, $Po = (A1 + B2 + C3) / M$, $Pe = \text{Cálculo generado desde la matriz de confusión}$, $Pe = (EP1 \times EU1) + (EP2 \times EU2) + (EP3 \times EU3)$ ([Llactayo et al., 2014](#)).

Con el propósito de discriminar la confiabilidad de cada uno de los métodos, se construyen entonces matrices de confusión (CM, por sus siglas en inglés), ya que este

ha sido uno de los métodos más utilizados para validar estos tipos de modelos ([Willington et al., 2013](#)). Así, para cada uno de los modelos se proporcionaron las matrices de confusión obtenidas.

Las [Tablas 3](#) a la [12](#) resumen los resultados obtenidos al aplicar los cinco métodos de clasificación supervisada.

Las [Tablas 5](#) y [6](#) resumen los resultados obtenidos al aplicar método RF.

Las [Tablas 7](#) y [8](#) resumen los resultados obtenidos al aplicar el método DT.

En las [Tablas 9](#) y [10](#) se presentan la CM del método de Bayes, así como las determinaciones de áreas y porcentaje de cada cobertura para los páramos Merchán y Telecom.

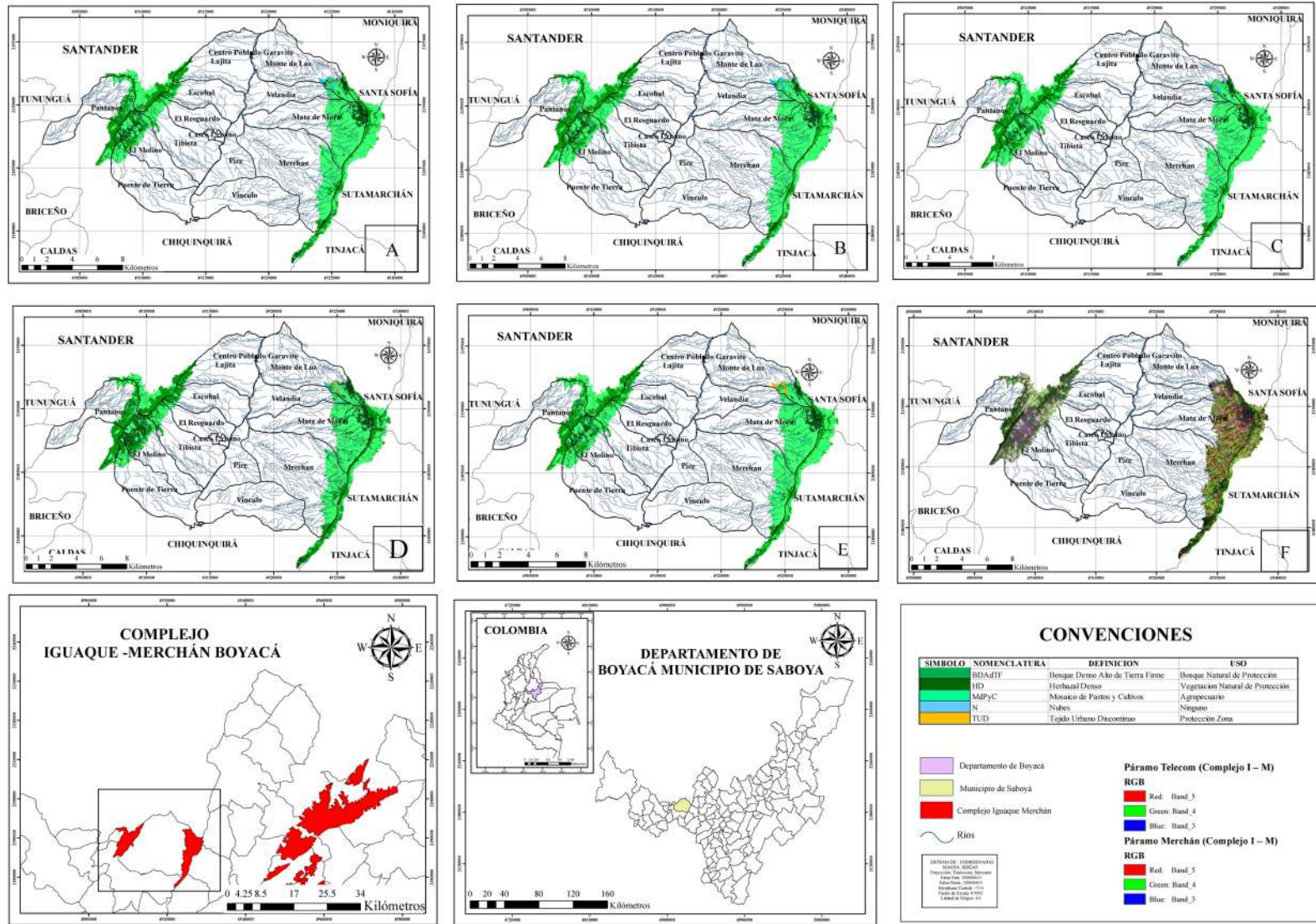


Figura 4. Clasificación supervisada basada en métodos de *Machine Learning*. **A.** Clasificación bajo el método Máquinas de Soporte Vectorial (SVM). **B.** Clasificación bajo el método Bosques Aleatorios (RF). **C.** Clasificación bajo el método Árboles de Decisión (DT). **D.** Clasificación bajo el método de Bayes (BC). **E.** Clasificación bajo el método Redes Neuronales Artificiales (ANN). **F.** Mejoramiento bajo SCP.

Tabla 2. Índice Kappa estimado por método de clasificación para el año 2018.

Método de clasificación	Índice Kappa (%)
Máquinas de Soporte Vectorial (SVM)	91,32
Bosque Aleatorios (RF)	92,53
Árboles de Decisión (DT)	80,07
Clasificación por Bayes (BC)	86,38
Redes Neuronales Artificiales (ANN)	91,51

Tabla 3. CM en píxeles de 15 metros mediante el método de SVM. Precisión general de la clasificación (Kappa) = 91,32%.

Clasificación	Datos de referencia del terreno					Total
	Bosque Denso Alto de Tierra Firme	Herbazal denso	Mosaico de Pastos y Cultivos	Nubes	Tejido Urbano Discontinuo	
Bosque Denso Alto de Tierra Firme	3197	52	230	0	0	3479
Herbazal denso	100	998	184	0	0	1282
Mosaico de Pastos y Cultivos	315	278	7589	1	7	8190
Nubes	0	0	0	414	1	415
Tejido Urbano Discontinuo	0	1	43	2	588	634
Total	3612	1329	0	0	596	14000

Tabla 4. Áreas y porcentaje de cobertura en los páramos analizados: en tonalidad gris se denotan los valores obtenidos del método SVM año 2018 para el páramo de Merchán, mientras que los valores en fondo blanco corresponden a los valores del páramo de Telecom.

Gridcode	Cobertura	Área (km ²)	Porcentaje (%)
1	BDAAdTF	4,41	13,12
		10,39	45,47
2	HD	3,14	9,34
		1,55	6,79
3	MdPyC	25,70	76,42
		10,89	47,66
4	N	0,34	1,02
		0,003	0,01
5	TUD	0,03	0,09
		0,01	0,05
	Total	33,63	100
		22,85	100

Tabla 5. CM en píxeles de 15 metros mediante el método RF. Precisión general de la clasificación (Kappa) = 92,53%.

Clasificación	Datos de referencia del terreno					Total
	Bosque Denso Alto de Tierra Firme	Herbazal denso	Mosaico de Pastos y Cultivos	Nubes	Tejido Urbano Discontinuo	
Bosque Denso Alto de Tierra Firme	3221	39	226	0	0	3486
Herbazal denso	68	944	248	0	0	1260
Mosaico de Pastos y Cultivos	219	211	7774	3	8	8215
Nubes	0	3	0	421	3	427
Tejido Urbano Discontinuo	0	4	10	3	595	612
Total	3508	1201	8258	427	606	14000

Tabla 6. Áreas y porcentaje de cobertura en los páramos analizados: en tonalidad gris se denotan los valores obtenidos del método RF año 2018 para el páramo de Merchán, mientras que los valores en fondo blanco corresponden a los valores del páramo de Telecom.

Gridcode	Cobertura	Área (km ²)	Porcentaje (%)
1	BDAdTF	4,37	13,01
		9,97	43,63
2	HD	2,75	8,19
		1,55	6,79
3	MdPyC	26,13	77,68
		11,29	49,41
4	N	0,02	0,06
		0,002	0,01
5	TUD	0,35	1,05
		0,03	0,14
Total	Total	33,63	100
		22,85	100

Tabla 7. CM en píxeles de 15 metros mediante el método DT. Precisión general de la clasificación (Kappa) = 80,07%.

Clasificación	Datos de referencia del terreno					Total
	Bosque Denso Alto de Tierra Firme	Herbazal denso	Mosaico de Pastos y Cultivos	Nubes	Tejido Urbano Discontinuo	
Bosque Denso Alto de Tierra Firme	3107	33	326	0	0	3466
Herbazal denso	234	792	222	0	0	1248
Mosaico de Pastos y Cultivos	316	290	7617	23	4	8250
Nubes	2	0	1	384	18	405
Tejido Urbano Discontinuo	2	1	42	15	571	631
Total	3661	1116	8208	422	593	14000

Tabla 8. Áreas y porcentaje de cobertura en los páramos analizados: en tonalidad gris se denotan los valores obtenidos del método DT año 2018 para el páramo de Merchán, mientras que los valores en fondo blanco corresponden a los valores del páramo de Telecom.

Gridcode	Cobertura	Área (km ²)	Porcentaje (%)
1	BDAdTF	4,55	13,52
		10,23	44,79
2	HD	2,72	8,09
		1,19	5,21
3	MdPyC	25,97	77,21
		11,41	49,93
4	N	0,31	0,94
		0,005	0,02
5	TUD	0,07	0,22
		0,008	0,03
Total		33,64	100
		22,85	100

Tabla 9. CM en píxeles de 15 metros mediante el método BC. Precisión general de la clasificación (Kappa) = 86,38%.

Clasificación	Datos de referencia del terreno					Total
	Bosque Denso Alto de Tierra Firme	Herbazal denso	Mosaico de Pastos y Cultivos	Nubes	Tejido Urbano Discontinuo	
Bosque Denso Alto de Tierra Firme	3272	150	126	0	0	3548
Herbazal denso	159	914	154	0	0	1227
Mosaico de Pastos y Cultivos	375	741	6957	7	87	8167
Nubes	0	0	4	344	60	408
Tejido Urbano Discontinuo	0	0	29	14	607	650
Total	3806	1805	7270	365	754	14000

Tabla 10. Áreas y porcentaje de cobertura en los páramos analizados: en tonalidad gris se denotan los valores obtenidos del método BC año 2018 para el páramo de Merchán, mientras que los valores en fondo blanco corresponden a los valores del páramo de Telecom.

Gridcode	Cobertura	Área (km ²)	Porcentaje (%)
1	BDAdTF	5,27	15,69
		10,26	44,90
2	HD	5,14	15,31
		3,79	16,59
3	MdPyC	22,73	67,58
		8,75	38,29
4	N	0,18	0,53
		0,004	0,18
5	TUD	0,30	0,89
		0,04	0,19
Total		33,63	100
		22,86	100

Análogamente que, en los casos anteriores, en las Tablas 11 y 12 se presentan la CM del método de redes neuronales artificiales, así como las determinaciones de áreas y porcentaje de cada cobertura para los páramos Merchán y Telecom.

Con el propósito de facilitar la discusión en la Tabla 13 se presenta un resumen de todos los métodos empleados para cada uno de los páramos estudiados.

Tabla 11. CM en píxeles de 15 metros mediante el método ANN. Precisión general de la clasificación (Kappa) = 91,51%.

Clasificación	Datos de referencia del terreno					Total
	Bosque Denso Alto de Tierra Firme	Herbazal denso	Mosaico de Pastos y Cultivos	Nubes	Tejido Urbano Discontinuo	
Bosque Denso Alto de Tierra Firme	3204	77	226	0	0	3507
Herbazal denso	85	982	197	0	0	1264
Mosaico de Pastos y Cultivos	333	233	7604	3	7	8180
Nubes	0	0	0	418	0	418
Tejido Urbano Discontinuo	0	1	26	0	604	631
Total	3622	1293	8053	421	611	14000

Tabla 12. Áreas y porcentaje de cobertura en los páramos analizados: en tonalidad gris se denotan los valores obtenidos del método ANN año 2018 para el páramo de Merchán, mientras que los valores en fondo blanco corresponden a los valores del páramo de Telecom.

Gridcode	Cobertura	Área (km ²)	Porcentaje (%)
1	BDAdTF	4,57	13,58
		10,47	45,67
2	HD	3,06	9,09
		1,55	6,76
3	MdPyC	25,61	76,16
		10,84	47,44
4	N	0,01	0,05
		0	0
5	TUD	0,38	1,12
		0,02	0,12
	Total	33,63	100
		22,89	100

Tabla 13. Síntesis de resultados de la determinación de áreas y porcentajes de coberturas en los páramos Telecom y Merchán año 2018, obtenidos por los cinco métodos de estudio: SVM, RF, DT, BC y ANN.

Páramo de Telecom											
Coberturas	SVM		RF		DT		BC		ANN		
	Área (km ²)	Porcentaje (%)	Área (km ²)	Porcentaje (%)	Área (km ²)	Porcentaje (%)	Área (km ²)	Porcentaje (%)	Área (km ²)	Porcentaje (%)	
BDAdTF	10,39	45,47	9,97	43,63	10,23	44,79	10,26	44,9	10,47	45,67	
HD	1,55	6,79	1,55	6,79	1,19	5,21	3,79	16,59	1,55	6,76	
MdPyC	10,89	47,66	11,29	49,41	11,41	49,93	8,75	38,29	10,84	47,44	
N	0,003	0,01	0,002	0,01	0,005	0,02	0,004	0,18	0	0	
TUD	0,01	0,05	0,03	0,14	0,008	0,03	0,04	0,19	0,02	0,12	
Páramo de Merchán											
Coberturas	SVM		RF		DT		BC		ANN		
	Área (km ²)	Porcentaje (%)	Área (km ²)	Porcentaje (%)	Área (km ²)	Porcentaje (%)	Área (km ²)	Porcentaje (%)	Área (km ²)	Porcentaje (%)	
BDAdTF	4,41	13,12	4,37	13,01	4,55	13,52	5,27	15,69	4,57	13,58	
HD	3,14	9,34	2,75	8,19	2,72	8,09	5,14	15,31	3,06	9,09	
MdPyC	25,7	76,42	26,13	77,68	25,97	77,21	22,73	67,58	25,61	76,16	
N	0,34	1,02	0,02	0,06	0,31	0,94	0,18	0,53	0,01	0,05	
TUD	0,03	0,09	0,35	1,05	0,07	0,22	0,3	0,89	0,38	1,12	

Discusión

El índice Kappa indica la confiabilidad y precisión general de la clasificación temática, de esta manera permite evaluar la concordancia de métodos cuyo resultado es categórico, con dos o más clases. Este índice representa la proporción de acuerdos observados respecto del máximo acuerdo posible más allá del azar (Borràs *et al.*, 2017). Según la clasificación establecida para el índice Kappa presentada en Llactayo *et al.* (2014), y con los resultados obtenidos de este índice en el presente estudio, se puede considerar que los cinco métodos de clasificación supervisada se encuentran en una concordancia casi perfecta (0,81 – 1) para calificar la exactitud obtenida.

El método que obtuvo mayor precisión en la clasificación supervisada para la imagen de Landsat 8 (2018) fue RF con un valor del 92,53% (Tabla 2). Los resultados afirman la precisión en la clasificación, pues este algoritmo permite resultados de calidad ya que genera múltiples árboles de decisión donde hay un subconjunto de muestras al azar y variables de entrenamiento (Belgiu y Drăguț, 2016); de igual forma como mencionan Sánchez-Pellicer *et al.* (2017), RF es uno de los algoritmos más usados en la actualidad para la teledetección, esto por su capacidad de integrar datos de diversas fuentes.

La investigación realizada por Fragoso-Campón y Quirós (2019), titulada *Sentinel Toolbox Application* (SNAP) aplicada a la clasificación supervisada de imágenes PNOA, en la cual la clasificación mediante el algoritmo RF alcanzó una exactitud global del 86%, ayudando a validar al software SNAP como herramienta de trabajo, no sólo de imágenes satelitales, sino también de imágenes aéreas y de alta resolución. De igual forma el valor de precisión y confiabilidad para RF (92,53%) obtenido en el presente estudio, tiene un valor cercano a una clasificación de cobertura terrestre realizada en un geo ecosistema montañoso ecuatoriano, realizado por Ayala-Izurieta *et al.* (2017), en el cual se clasificó la cobertura vegetal de las áreas montañosas dentro de un área de estudio de 1016 km², a una resolución espacial de 30 m, utilizando un método clasificador para RF que arrojó un mapa de cobertura terrestre con una precisión general del 95%.

El algoritmo de RF ha sido utilizado en diferentes estudios, entre los cuales se encuentra el de Espejo (2016), donde se concluyó que, para los objetos de deforestación generados a partir de los objetos de Bosque y No Bosque clasificados por el método de RF y SVM se presentan valores más altos en la similaridad temática (*Theme*), borde (*Edge*) y posición (*Position*).

En las CM, se observa que, para el método RF (Tabla 5), la cobertura vegetal Mosaico de Pastos y cultivos presentó un menor grado de precisión dado que obtuvo 7774 puntos correctos de 8215 puntos, de igual manera, Herbazal denso obtuvo 944 puntos de 1260; estos datos permiten verificar la exactitud temática respecto a la clasificación establecida con los datos de referencia del terreno. Es importante resaltar que los valores bajos con respecto a las otras coberturas se deben a la poca área ocupada en la zona, es decir la cobertura de mayor área es la que presenta el mayor número de puntos de control. Con esto es posible inferir que la clase con mayor precisión y agrupación espectral fue nubes, ya que, obtuvo un valor de 421 puntos correctos de 427, seguida de Tejido Urbano Discontinuo y Bosque Denso Alto de Tierra Firme.

Con base en la clasificación supervisada por el método de RF, en la Tabla 6, se puede observar que la mayor área de cobertura corresponde a Mosaico de Pastos y Cultivos (MdPyC), con 26,131 km² en el páramo de Merchán y 11,297 km² en el páramo de Telecom. La segunda cobertura vegetal mejor distribuida es el Bosque Denso Alto de Tierra Firme (BDAdTF), en un mayor porcentaje de área en el páramo de Telecom con un valor del 43,62% respecto al área total del páramo.

El segundo método más preciso en la clasificación supervisada para los páramos de Merchán y Telecom, fue ANN el cual obtuvo un índice Kappa de 91,51% (Tabla 2). La CM (Tabla 11), muestra que las coberturas de Bosque Denso Alto de Tierra Firme y Mosaico de Pastos y Cultivos fueron las menos precisas, dado que presentan la mayor diferencia entre puntos de control y los puntos correctos; la clase de nubes obtuvo un 100% de precisión tomando totalmente los puntos correctos.

Los valores de área para las clases de coberturas por ANN, son cercanos a las áreas presentadas por RF. En el caso del Herbazal Denso en el páramo de Merchán, por ejemplo, con RF se obtuvo un área de 2,757 km² (Tabla 6) y con ANN un valor de 3,060 km² (Tabla 12), esto se debe a que los dos métodos fueron los más altos en la precisión de la clasificación según el índice Kappa. Sin embargo, algunos valores de coberturas entre estos dos algoritmos presentan una mínima diferencia, por ejemplo: BDAdTF y MdPyC en el páramo de Telecom muestran porcentajes de área, respectivamente, de 45,66% y 47,44% con ANN (Tabla 12), y 43,62% y 49,41% con RF (Tabla 6).

El siguiente método más preciso en la clasificación, según el índice Kappa, fue SVM con un valor del

91,32%. Este algoritmo presenta una buena precisión de clasificación ya que, según Murillo-Castañeda, 2018, SVM tiene una alta resolución al momento de identificar coberturas presentes en el suelo, pues aparte de lograr disociación entre cada una de ellas, también separa los elementos a clasificar. Dentro de la CM (Tabla 3) se observa que la cobertura que presentó mayor asignación de puntos fue Mosaico de Pastos y Cultivos, sin embargo, obtuvo el menor grado de precisión ya que presentó una diferencia de 601 puntos respecto a los puntos totales de control; esta clase está seguida de Herbazal Denso (998 puntos correctos) y Bosque Denso Alto de Tierra Firme (3197).

Entre los algoritmos de SVM y BC, se observa que existe una diferencia significativa entre los valores de las áreas de coberturas. En la Tabla 10 del método BC, se observa que el porcentaje de área para Herbazal Denso en el páramo de Telecom es 16,59%, mientras en SVM presentan un valor de 6,79% para esta cobertura (Tabla 4). Lo mismo ocurre con MdPyC del páramo de Merchán, el cual alcanzó un valor de 67,57% de área con el método BC y de 76,42% de área con el método SVM. Con ello es posible observar la relación que existe con el índice Kappa, en el cual el método BC obtuvo un valor de 86,38%, uno de los valores de confiabilidad más bajos en comparación con los otros cuatro métodos evaluados.

El índice Kappa para los métodos BC y DT, fueron 86,38% y 80,07% respectivamente (Tabla 2). Estos valores, reflejan una buena precisión general de la clasificación, sin embargo, como se menciona anteriormente, fueron los valores más bajos en comparación con los demás métodos. Es de resaltar que el índice Kappa, como menciona Llactayo *et al.* (2014), es una medida de la diferencia entre la exactitud lograda en la clasificación con un software y la exactitud de lograr una clasificación correcta con una clasificación visual (en campo o con imágenes de alta resolución).

Las coberturas de Tejido Urbano Discontinuo (TUD) y Nubes, en los cinco métodos, presentaron áreas muy bajas en comparación con las demás. Por ejemplo, con el método de DT, en el páramo de Merchán se obtuvo un área de 0,0748 km² para TUD y 0,0082 km² de TUD en el páramo de Telecom. Sin embargo, es importante resaltar que para la imagen Landsat 8 tomada de estudio y/o referencia, el páramo de Telecom no presentaba ningún grado de nubosidad, contrario al de Merchán, de ahí el mínimo porcentaje de área; donde si hay presencia se debe a errores de precisión del algoritmo.

Respecto a la cobertura de Bosque Denso Alto de Tierra Firme, se observa que existe un mayor porcentaje de área 43,62% (método RF) en el páramo de Telecom, con lo cual se puede deducir que en el páramo de Merchán se ha presentado mayor intervención agrícola y pecuaria; el mayor valor de cobertura presente en este páramo es Mosaico de Pastos y Cultivos con un porcentaje de área de 77,68% respecto al área total del páramo.

Los valores obtenidos de las CM para los cinco métodos utilizados, muestran que las clases de Tejido Urbano Discontinuo y Nubes presentaron la mejor precisión en todos los algoritmos. Los pocos puntos presentados para Tejido Urbano Discontinuo se deben a un grado de error por el posible desacuerdo de la respuesta espectral de las nubes, dado que se puede presentar una confusión de los píxeles de este tipo de clase, clasificándolas como tejido urbano discontinuo; así mismo, en la zona de estudio del páramo de Telecom existe una base militar que interviene con la captura de los puntos de control.

De igual forma se observa dentro de los resultados obtenidos en las CM, que estas presentan una cantidad considerable de puntos de control lo cual se debe a que para la clasificación supervisada se tomó un polígono de tipo rectangular que abarco el área del páramo de Merchán junto con el de Telecom, ya que el lenguaje de programación R no considera polígonos irregulares como el límite perimetral de los páramos de estudio; por esta razón al establecerse los puntos de control dentro del polígono se visualizan varios de ellos.

Debido a que el contorno o límite de los páramos pareciera estar limitado en el caso del páramo de Merchán por el contacto entre las formaciones Cretácicas y Cuaternarias, y en el caso del páramo de Telecom por el eje del sinclinal, para cada uno de los métodos se realizaron 6 medidas de la distancia entre el borde del páramo dado por cada método y esos rasgos geológicos. Los resultados son resumidos en la [Figura 5](#). Se observa que desde un punto de vista geológico la distancia promedio predicha entre el borde del páramo y el eje del sinclinal es muy similar para los distintos métodos, y si se considera la desviación estándar prácticamente no hay diferencia entre ellos. En contraste, cuando se determina la distancia entre el borde del páramo y el límite de las formaciones cretácicas para el páramo de Merchán, los algoritmos SVM, RF, ANN y DT arrojan resultados similares en términos de la distancia promedio y la desviación estándar. Mientras que el método BC tiende a subestimar la distancia promedio. Para el páramo de Telecom, la distancia promedio entre el límite del páramo y el contacto con formaciones cuaternarias es

muy similar para cada uno de los métodos, siendo el algoritmo de RF, el que presenta la menor dispersión.

Al realizar una correcta clasificación supervisada de coberturas vegetales, fue posible obtener datos reales del terreno que se pueden considerar insumos precisos y confiables para la determinación y análisis de métricas del paisaje, ayudando a establecer y analizar la configuración, diversidad y forma del paisaje, en este caso para los páramos de Merchán y Telecom en el municipio de Saboyá-Boyacá. Estos estudios son importantes para la gestión de la estructura ecológica principal, entendiéndose esta según el PND 2010-2014 ([IDEAM, 2011](#)), como el conjunto de ecosistemas estratégicos que garantizan la integridad de la biodiversidad y la provisión de servicios ecosistémicos con el fin de satisfacer las necesidades básicas de la población. Estos páramos brindan diferentes servicios ecosistémicos, entre los cuales encontramos los de aprovisionamiento y regulación hídrica; motivo por el cual su cuidado y conservación es fundamental dentro del Plan de Ordenamiento Territorial para el desarrollo sostenible en el Municipio de Saboyá.

Importancia de la geología en la extensión de los páramos

En los resultados obtenidos ([Tabla 13](#)), independientemente del método utilizado, se observa una tendencia de disminución del área de la cobertura del páramo de Merchán, en comparación con el de Telecom, diversos factores asociados a la geología pudieran justificar la diferencia entre ambos páramos, el primero de ellos: la litología, al producirse los procesos de erosión por efecto del agua, se crean los suelos que servirán de soporte a los herbazales densos, el páramo de Merchán tiene una densidad de drenajes mucho mayor que el de Telecom, esto hace que rápidamente la roca se erosione y genere distintos tipos de suelos. En el páramo de Telecom existe una importante acumulación de sedimentos no consolidados de edad cuaternaria, las cuales limitan la extensión del páramo hacia el sector oriental ([Figura 1](#)). Segundo: en el páramo de Merchán se observa un sistema local de plegamientos, un sinclinal y un anticlinal son indicativos de que la zona está sufriendo un proceso importante de compresión, la cual pudiera causar un aumento en la rugosidad del terreno. Ambos páramos están separados por un sistema de fallas de tipo inverso, donde el páramo de Telecom cabalga sobre la Serranía de Merchán, este proceso podría generar basculamientos de bloques que inciden en la pendiente del terreno y por lo tanto los métodos usados pudieran ser sensibles a estos fenómenos de incremento de la pendiente a lo largo del tiempo.

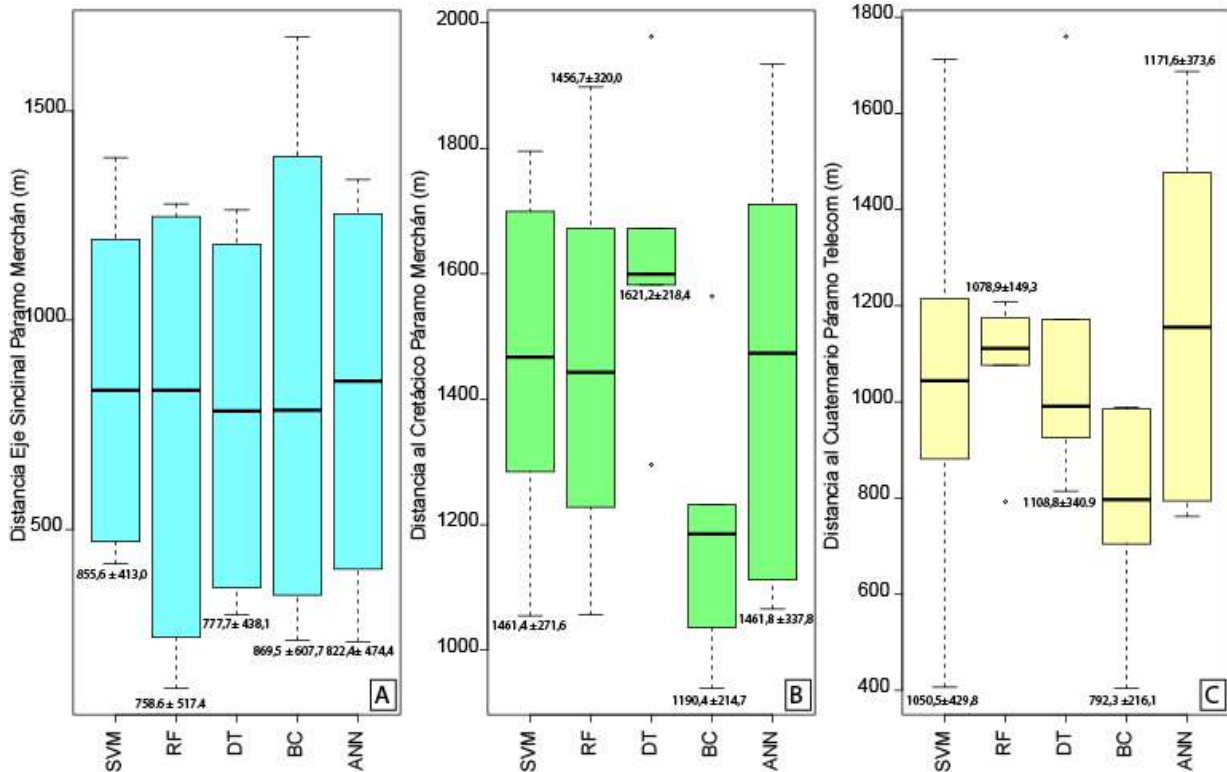


Figura 5. Diagramas Whisker-Plot o boxplots mostrando: **A.** Distancias del borde del páramo de Merchán al eje del sinclinal. **B.** Distancia del borde del páramo de Mechán con el limite de las formaciones cretácicas. **C.** Distancia del borde del páramo de Telecom al contacto con las formaciones cuaternarias para cada uno de los métodos.

Conclusiones

Con base en los resultados obtenidos en las matrices de confusión y el índice Kappa, es posible confirmar que el método con mayor precisión en confiabilidad para la clasificación supervisada es RF con un valor del 92,53% del índice Kappa, durante la aplicación de este método se observó que identificaba con mayor facilidad el contacto entre rocas cretácicas y cuaternarias. Este algoritmo, permite representar de forma clara y precisa las características del terreno, pues por medio de un proceso de múltiples árboles de decisión, logra obtener resultados de alta precisión y calidad.

Por medio de esta investigación, se buscó evaluar la eficacia y precisión de los métodos de *Machine Learning* para los procesos de clasificación supervisada, con los resultados obtenidos se pudo observar que los cinco métodos son efectivos para la identificación de coberturas; para zonas de estudio como páramos se sugiere en principio utilizar los dos métodos que por naturaleza de su algoritmos obtuvieron los mejores índices de confiabilidad y precisión, siendo estos RF seguido de ANN.

El índice Kappa es una medida estadística que permite ver la exactitud y precisión de los algoritmos o métodos de clasificación, con lo cual se puede elegir el método que mejor representa la realidad de terreno, por esta razón se sugiere la inclusión del índice en futuras investigaciones.

Por medio de la clasificación supervisada para los páramos Merchán y Telecom en Saboyá en el Complejo de Páramos Iguaque - Merchán, Boyacá, realizada con Landsat 8 (2018), se observa, en los cinco métodos, que la cobertura de mayor presencia es Mosaico de Pastos y Cultivos; esto demuestra que en la zona de páramo han existido cambios en el tipo y uso de suelo, con presencia de actividades de tipo agrícola o pecuario que afectan el cambio de cobertura reciente del páramo y la conectividad ecológica, biodiversidad y prestación de servicios ecosistémicos.

Los diferentes algoritmos de teledetección son una herramienta útil para establecer los tipos de coberturas vegetales en un terreno y los cambios que se han presentado en estas a lo largo del tiempo. Implementar estos métodos en análisis multitemporales en zonas

de importancia ecológica como los páramos, permite identificar los impactos y problemáticas ambientales que se han generado allí, ayudando así a establecer estrategias y medidas de gestión ambiental para los recursos naturales.

Esta investigación es un insumo importante para la continuidad de estudios de métricas del paisaje y análisis multitemporales para los páramos de Merchán y Telecom en el municipio de Saboyá-Boyacá. El grado de exactitud de los métodos evaluados en este estudio, permite que los posibles análisis postclasificación tengan una mayor confiabilidad y certeza, de esta manera servirán como base para la gestión ambiental del territorio y ordenamiento territorial.

Agradecimientos

La presente investigación es parte de la tesis de maestría en Ciencias de la Tierra de la Universidad Pedagógica y Tecnológica de Colombia del primer autor, razón por la cual agradecemos a la Corporación Autónoma regional de Cundinamarca CAR. Al ingeniero Nino Bravo de la Universidad Nacional Agraria de la Selva en Perú por apoyo para el uso de los códigos en R, al Instituto Científico del Perú (ICIP) por sus bases y servicios. A Luca Congedo por su tiempo para la elaboración del código SCP usado en la presente investigación, de libre acceso en QGIS. A Jana Poveda, Adriana Montoya y Daniel Ricardo Mahecha por su soporte en el desarrollo de la presente investigación. Al financiamiento proporcionado por la UPTC bajo el Proyecto SGI-UPTC 3104. A los revisores Héctor Fonseca y Brigitte Baptiste por los comentarios y sugerencias realizados que ayudaron a mejorar el presente documento.

Referencias

Aly, A.A.; Al-Omran, A.M.; Sallam, A.S.; Al-Wabel, M.I.; Al-Shayaa, M.S. (2016). Vegetation cover change detection and assessment in arid environment using multi-temporal remote sensing images and ecosystem management approach. *Solid Earth*, 7(2), 713-725. <https://doi.org/10.5194/se-7-713-2016>

Arellano, H.; Rangel, J.O. (2008). Patrones en la distribución de la vegetación en áreas de páramo de Colombia: heterogeneidad y dependencia espacial. *Caldasia*, 30(2), 355-411.

Ayala-Izurieta, J.; Márquez, C.; García, V.; Recalde-Moreno, C.; Rodríguez-Llerena, M.; Damián-Carrión, D. (2017). Land Cover Classification in an Ecuadorian Mountain Geosystem Using a Random Forest Classifier, Spectral Vegetation Indices, and Ancillary Geographic Data. *Geosciences*, 7(2). <https://doi.org/10.3390/geosciences7020034>

Becerra-Serial, R.M.; Czibener, D.; Nabel, P.E. (2009). Teledetección y redes neuronales aplicadas al mapeo de coberturas del suelo de la cuenca del Matanza-Riachuelo, Buenos Aires, Argentina. *Revista Geográfica*, 146, 125-152.

Belgiu, M.; Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>

Borràs, J.; Delegido, J.; Pezzola, A.; Pereira, M.; Morassi, G.; Camps-Valls, G. (2017). Clasificación de usos del suelo a partir de imágenes Sentinel-2. *Revista de Teledetección*, 48, 55-66. <https://doi.org/10.4995/raet.2017.7133>

Bravo-Morales, N.F. (2020). *Teledetección y procesamiento de imágenes satelitales*. GEOMATICA AMBIENTAL S.R.L.

Cervantes-Canales, J. (2009). Clasificación de grandes conjuntos de datos vía máquinas de vectores soporte y aplicaciones en sistemas biológicos. Tesis de PhD, Instituto Politécnico Nacional, México.

Chuvienco, E. (1995). *Fundamentos de teledetección espacial*. RIALP, S.A.

Chuvienco, E. (2002). *Teledetección ambiental. La observación de la Tierra desde el Espacio*. Ariel, S.A.

Cleef, A.M. (2008). Influencia humana en los páramos. En: J.P. Castañeda (ed.). *Panorama y perspectivas sobre la gestión ambiental de los ecosistemas de páramo: memorias* (pp. 26-33). Procuraduría Delegada para Asunto Ambientales y Agrarios.

Cortés-Duque, J.; Sarmiento-Pinzón, C. (2013). *Visión socioecosistémica de los páramos y la alta montaña colombiana: memorias del proceso de*

definición de criterios para la delimitación de páramos. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt.

- De'Ath, G.; Fabricius, K.E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, 81(11), 3178-3192. [https://doi.org/10.1890/0012-9658\(2000\)081\[3178:CARTAP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2)
- De León-Mata, G.D.; Pinedo-Álvarez, A.; Martínez-Guerrero, J.H. (2014). Aplicación de sensores remotos en el análisis de la fragmentación del paisaje en Cuchillas de la Zarca, México. *Investigaciones Geográficas*, 84, 42-53. <https://doi.org/10.14350/rig.36568>
- Di Somma, A.; Ferrari, V.; Ramos-de-Las-Heras, N. (2010). El uso del suelo y el análisis multitemporal – Modificaciones del tejido urbano en la Provincia de Roma (Italia). *Colóquio Ibérico de Geografía*, XVII, Porto, Italia.
- dos Santos, D.I.; Araújo, É.; Pagani, P.C.; Pagani, C.H.; de Araújo, M.E.; Negrão, M. (2015). Análise multitemporal de uso e ocupação do solo do núcleo inicial do projeto integrado de colonização Paulo de Assis Ribeiro no município de Colorado do Oeste - RO. *Caderno de Geografia*, 25(43), 34-51. <https://doi.org/10.5752/p.2318-2962.2015v25n43p34>
- Espejo, O. (2016). Desarrollo de una metodología para estimación de la deforestación mediante el análisis multitemporal de imágenes multiespectrales en un entorno de análisis basado en objetos geográficos (GEOBIA). Tesis de Maestría. Universidad Distrital Francisco José de Caldas, Colombia.
- Fragoso-Campón, L.; Quirós, E. (2019). Sentinel Toolbox Application (SNAP) aplicado a la clasificación supervisada de imágenes PNOA. *VXIII Congreso de la Asociación Española de Teledetección*. Cáceres, España.
- Friedl, M.A.; Brodley, C.E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61(3), 399-409. [https://doi.org/10.1016/S0034-4257\(97\)00049-7](https://doi.org/10.1016/S0034-4257(97)00049-7)
- Fuquen, J.A.; Osorno, J.F. (2005). Geología de la plancha 190-Chiquinquirá. Informe No. I - 1794, INGEOMINAS, 272 p.
- Garavito-Rincón, L.N. (2015). Los páramos en Colombia, un ecosistema en riesgo. *Ingeniare*, 19, 127-136.
- González, L.; Rodríguez, J. (2006). Prototipo de software para la clasificación de datos mediante el método bayesiano TAN - “udTAN”. *Vínculos*, 3(1), 35-54.
- González, A.R. (2011). Nuevas percepciones del territorio, espacio social y el tiempo. Un estudio desde los conceptos tradicionales (o clásicos) hasta su concepción en el siglo XXI (Facultad de Humanidades, Artes y Ciencias Sociales). *VI Jornadas de Jóvenes Investigadores*. Buenos Aires. Argentina.
- IDEAM (2010). Leyenda Nacional de Coberturas de la Tierra. Metodología CORINE Land Cover adaptada para Colombia Escala 1:100.000. Instituto de Hidrología, Meteorología y Estudios Ambientales. Bogotá.
- IDEAM (2011). Aportes del IDEAM para la definición y aplicación de la Estructura Ecológica Nacional. Proceso metodológico y aplicación para la definición de la estructura ecológica nacional: énfasis en servicios ecosistémicos Escala 1:500.000. Bogotá. Instituto de Hidrología, Meteorología y Estudios Ambientales.
- Jaramillo, L.V.; Antunes, A.F. (2018). Detección de cambios en la cobertura vegetal mediante interpretación de imágenes Landsat por redes neuronales artificiales (RNA). Caso de estudio: Región Amazónica Ecuatoriana. *Revista de Teledetección*, 51, 33-46. <https://doi.org/10.4995/ract.2018.8995>
- Lanzarini, L.C. (2004). Redes Neuronales Artificiales. Un Enfoque Práctico. *Journal of Computer Science and Technology*, 4(2), 122-123.
- Llactayo, W.; Salcedo, K.; Victoria, E. (2014). Protocolo de evaluación de la exactitud temática del mapa de deforestación. Dirección General de Ordenamiento Territorial (DGOT). Perú.

- Llanos-Hernández, L. (2010). El concepto del territorio y la investigación en las ciencias sociales. *Agricultura, Sociedad y Desarrollo*, 7(3), 207-220.
- Lozada, J.R. (2007). Situación actual y perspectivas del manejo de recursos forestales en Venezuela. *Revista Forestal Venezolana*, 51(1), 195-218.
- Manna, S.; Mondal, P.P.; Mukhopadhyay, A.; Akhand, A.; Hazra, S.; Mitra, D. (2013). Vegetation cover change analysis from multi-temporal satellite data in Jharkhali Island, Sundarbans, India. *Indian Journal of Geo-Marine Sciences*, 42(3), 331-342.
- Morales-Rivas, M.; Otero-García, J.; Van-der-Hammen, T.; Torres-Perdigón, A.; Cadena-Vargas, C.; Pedraza-Peñaloza, C.; Rodríguez-Eraso, N.; Franco-Aguilera, C.; Betancourth-Suárez, J.; Olaya-Ospina, É.; Posada-Gilede, E.; Cárdenas-Valencia, L. (2007). *Atlas de páramos de Colombia*. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt.
- Moya-Fuero, A. (2012). Detección automática de nuevas construcciones a partir de ortofotos del Instituto Cartográfico Valenciano. Tesis de Maestría, Universidad de Valencia, España.
- Muñoz-Guerrero, D.; Rodríguez-Montenegro, M.; Romero-Hernández, M. (2009). Análisis multitemporal de cambios de uso del suelo y coberturas, en la microcuenca las minas, corregimiento de la laguna, Municipio de Pasto, departamento de Nariño. *Revista de Ciencias Agrícolas*, 26(1), 11-24.
- Murillo-Castañeda, R.A. (2018). Implementación del método máquinas de soporte vectorial en bases de datos espaciales para análisis de clasificación supervisada en imágenes de sensores remotos. Tesis de Maestría, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.
- Padrón, J. (2007). Tendencias epistemológicas de la investigación científica en el siglo XXI. *Cinta de Moebio*, 28, 1-28.
- Paredes-Inilup, D. (2020). *Data Science con R. Análisis de datos y algoritmos de predicción con R*. Editorial Leanpub.
- Parra, F. (2017). *Estadística y Machine Learning con R: Ejercicios resueltos con R*. Editorial Académica Española.
- Riquelme, J.C.; Ruiz, R.; Gilbert, K. (2006). Minería de datos: conceptos y tendencias. *Inteligencia Artificial*, 10(29), 11-18.
- Sánchez-Pellicer, T.; Martín-Alcón, S.; Tomé-Morán, J.L.; Navarro, J.A.; Fernández-Landa, A. (2017). ForestCO2: monitoring of carbon sinks in Pinus halepensis stands in the Region of Murcia. *XVII Congreso Nacional de Teledetección*, Valencia, España.
- Sanhouse-García, A.J.; Bustos-Terrones, Y.; Rangel-Peraza, J.G.; Quevedo-Castro, A.; Pacheco, C. (2017). Multi-temporal analysis for land use and land cover changes in an agricultural region using open source tools. *Remote Sensing Applications: Society and Environment*, 8, 278-290. <https://doi.org/10.1016/j.rsase.2016.11.002>
- Sepúlveda-Varas, A.; Saavedra-Briones, P.; Esse, C. (2019). Análisis de cambio de cobertura y uso de suelo en una subcuenca preandina chilena. Herramienta para la sustentabilidad productiva de un territorio. *Revista de Geografía Norte Grande*, 72, 9-25. <https://doi.org/10.4067/S0718-34022019000100009>
- Terraza, R. (2004). Significado facial y cartografía geológica de la Arenisca de Chiquinquirá en alrededores de la localidad tipo. Tesis de Maestría, Universidad Nacional de Colombia, Bogotá, Colombia.
- Torres, A.M.; Peña, E.J.; Zúñiga, O.; Peña, J.A. (2012). Evaluación del impacto de actividades antrópicas en el almacenamiento de carbono en biomasa vegetal en ecosistemas de alta montaña de Colombia. *Boletín Científico. Centro de Museos. Museo de Historia Natural*, 16(1), 132-142.
- Ulloa, C.; Rodríguez, E. (1978). Mapa geológico preliminar de la Plancha 170-Vélez. Publicada en 1984. Escala 1:100.000. INGEOMINAS. Bogotá
- Velasco-Linares, P.; Ávila, L.A.; Manosalva-Moreno, L.; Parra, S.L.; Caro-Roa, A.M.; García, H.F.; Cifuentes, Y.; Díaz, J.E.; Moreno, J.; Caro-

Cruz, A.; González-Zárate, A.; Caguazango, Á.P.; Chaparro-Herrera, S.; Montealegre-Talero, C. (2015). Patrones ecológicos en la zona de transición bosque - páramo. Complejo de Páramos Iguaque-Merchán: Transectos en Municipios Arcabuco, Sotaquirá, Cómbita (Boyacá) y Gámbita (Santander). Instituto de Investigación de Recursos Biológicos Alexander von Humboldt – Bosques & Semillas SAS.

Willington, E.; Nolasco, M.; Bocco, M. (2013). Clasificación supervisada de suelos de uso agrícola en la zona central de Córdoba (Argentina): comparación de distintos algoritmos sobre imágenes Landsat. *Congreso Argentino de AgroInformatica*, Argentina.

Fecha de recepción: octubre 12 de 2021

Fecha de aprobado: abril 08 de 2022
