

Mapeo de cobertura terrestre utilizando aprendizaje máquina

Land cover mapping using machine learning

Cristian Muñoz Ordóñez  Apolinar Figueroa  Leonairo Pencue Fierro 
Universidad del Cauca, Colombia

Julián Muñoz Ordóñez 
Corporación Universitaria Comfacaucá, Colombia

OPEN  ACCESS

Recibido: 22/09/2020

Aceptado: 22/10/2020

Publicado: 23/11/2020

Correspondencia de autores:
camilomu@unicauca.edu.co



Copyright 2020
by Investigación e
Innovación en Ingenierías

Resumen

Objetivo: Determinar las dinámicas de evolución de la cobertura terrestre sobre la Cuenca Alta del Río Cauca. **Metodología:** Esta investigación, presenta un proceso de clasificación de coberturas sobre la CARC utilizando imágenes Landsat-8 y Sentinel-1. Este análisis se realizó mediante la clasificación de características de imágenes satelitales ópticas y de radar, utilizando algoritmos de aprendizaje máquina. **Resultados y conclusiones:** se identificaron regiones que presentan afectación en sus coberturas vegetales mostrando la importancia de la utilización conjugada de satélites ópticos y de radar. La clasificación tuvo una precisión del 88.9% y un coeficiente kappa de 0.86.

Palabras clave: Clasificación de imágenes satelitales, percepción remota satelital, aprendizaje máquina, sistemas de producción combinada.

Abstract

Objective: Make an evolution dynamics analysis of covers on the Upper Cauca River Basin. **Methodology:** This research presents a landcover classification process on the UCB using Landsat-8 and Sentinel-1 data. This analysis was carried out by means of the features classification of optical and radar satellite data using machine learning algorithms. **Results and conclusions:** regions that present affectation in their vegetation cover were identified, showing the importance of the conjugated use of these two types of sources. The classification had an accuracy of 88.9% and a kappa coefficient of 0.86.

Keywords: Satellite imagery, remote sensing, machine learning, combined production systems.

Como citar (IEEE): C. Muñoz-Ordóñez., J. Muñoz-Ordóñez., L. Pencue-Fierro., y Apolinar-Figueroa. Mapeo de cobertura terrestre utilizando aprendizaje máquina, vol. 8, n°3, pp. 85-101, 2020. DOI: <https://doi.org/10.17081/invinno.8.3.4706>

Introducción

El monitoreo de la evolución temporal y espacial de la tierra es de suma importancia para la administración de los recursos terrestres e hídricos [1], especialmente en países como Colombia, que presenta a lo largo de su territorio una notable biodiversidad agrícola a causa de sus diferentes pisos térmicos [2]. La percepción remota sobre la cuenca Alta del río Cauca (CARC) cobra mayor importancia, ya que abarca casi todas las cadenas de producción priorizadas por el Ministerio de Agricultura de Colombia [3].

La Cuenca Alta del Cauca abarca gran parte de la agricultura técnica y de alto valor en el país, se extiende desde el Macizo Colombiano hasta los límites de los departamentos de Caldas y Antioquia, entre las cimas de las cordilleras occidental y central. Su altitud varía desde los 5.400 m.s.n.m en las nieves perpetuas de la cordillera central y los 950m.s.n.m en el valle aluvial del Cauca. El área de estudio aloja 191 subcuencas que tributan al río Cauca, comprende 99 municipios y alberga una población de aproximadamente 7.056.000 habitantes. A momento es el corredor agrícola e industrial más importante de Colombia donde se desarrolla agricultura tradicional, cultivo de café, de caña de azúcar, frijol, plátano, cítricos, cacao, maíz, aguacate, entre otros. Además, aproximadamente el 20% de la población nacional se beneficia de la oferta ambiental de los recursos de esta cuenca.

Actualmente la CARC presenta un alto deterioro, debido principalmente al uso inadecuado del suelo, deforestación, contaminación hídrica de origen antrópico y desecamiento de ciénagas y humedales, para la expansión de la frontera agrícola especialmente en la zona plana y afectación constante y continua de sus coberturas vegetales, todos estos factores son determinantes para la sostenibilidad ambiental de la Cuenca.

Sus condiciones atmosféricas, su cercanía a la ZCIT donde se tiene la convergencia de aire cálido y húmedo de latitudes por encima y por debajo del ecuador, es un rasgo atmosférico que favorece la distribución del calor y humedad, que es transportado verticalmente hacia arriba por la actividad convectiva de las tormentas, haciendo que las regiones situadas en esta área reciben precipitación más de 200 días al año. Esta condición genera alto grado de nubosidad sobre la CARC lo cual reduce drásticamente la resolución temporal de imágenes satelitales ópticas dificultando el análisis cuantitativo y cualitativo relacionado con el proceso de evolución de coberturas tales como: cultivos, cuerpos de agua, bosques naturales, paramos, esta condición demanda el uso de tecnologías alternativas para hacer el seguimiento y monitoreo de cambios de estas coberturas en los ecosistemas de la cuenca.

En este sentido las imágenes capturadas por un radar de apertura sintética trabajan en longitudes de onda superiores a las ópticas, evitando que la onda sea reflejada por las nubes y proporcionando características biofísicas adicionales y complementarias para los sistemas de clasificación. Logrando, de esta forma, complementar el estudio hecho únicamente con imágenes ópticas, incrementando el conjunto de herramientas para el desarrollo de estrategias de conservación y mitigación frente a posibles cambios que estén sucediendo en la CARC.

Este artículo se organiza de la siguiente manera: en la sección 2 se presentan los trabajos relacionados. Sección 3 presenta el clasificador Random Forest y la forma de extracción de características de ambos satélites. Sección 4 muestra los resultados experimentales sobre el dataset con los valores espectrales de imágenes Landsat 8 y Sentinel-1, mostrando los resultados de medidas de contraste entre coberturas de la región presentando una inspección visual de la clasificación obtenida. Finalmente, se exponen algunas conclusiones y sugerencias para trabajos futuros.

Trabajos relacionados

La percepción remota ha adquirido mayor importancia en los últimos años debido al surgimiento de algoritmos avanzados para la adquisición y procesamiento de imágenes [4], estos avances permiten generar herramientas de apoyo a la toma de decisiones, que incluyen el monitoreo operacional de los cultivos [5], lo cual se basa principalmente en índices de vegetación provenientes de datos ópticos, combinados con métricas agrometeorológicas como precipitación, temperatura, radiación solar y humedad del suelo [6]. Estas aplicaciones se han centrado en estudiar las interacciones sociedad naturaleza, analizando la relación existente entre el ser humano y la deforestación, la pérdida de la biodiversidad, el cambio climático [7] y otros procesos de transformación ocasionados por el desarrollo sociocultural y las actividades de apropiación y producción.

Tradicionalmente, los esfuerzos por determinar y conocer los cambios generados sobre la superficie terrestre se basaban en imágenes ópticas de alta resolución temporal que facilitaban el mapeo de las diferentes coberturas a escala regional y/o continental [8,9,10] discrepancies of LCCSs have widely existed in various LCDSs, which have caused negative impacts on comprehensive comparison and integrated utilization of multiple LCDSs. This study attempted to summarize the independent diagnostic criteria hidden in the existing LCCSs based on the induction method, and to synchronously discover the roles of data sources and classification methods in designing LCCSs. A total of 13 existing regional- or global-scale LCDSs were chosen. The analysis results show that phenology, coverage rate, vertical structure, and leaf type were the most frequently adopted criteria in the LCCSs of existing LCDSs. The decision of whether to adopt a diagnostic criterion in the LCCS of LCDS depended on the availability, effectiveness, and quality of the relevant data sources and classification methods. Currently, optical remote sensing images are still the prominent data source for regional- or global-scale LCDSs, and the potential of each diagnostic criterion could not be fully played. Multi-source and heterogeneous spatial data, ARD (Analysis Ready Data. Un enfoque popular utiliza índices ópticos de alta resolución temporal de sensores como MODIS, para extraer información fenológica y aprovechar la relación dinámica entre los índices de vegetación combinados con máscaras de la superficie. Por ejemplo, índice de agua superficial terrestre (LSWI), índice de vegetación mejorado (EVI), e índice de vegetación de diferencia normalizada (NDVI) generados cada 8 días a partir de la combinación de información obtenida por parte del satélite MODIS con las máscaras de uso de la tierra y pendientes [11,12,13]covering the time period 2000–2019. GUD was calculated for ~170 000 pixels with deciduous broadleaf forest cover at 500 m spatial resolution. The GUD exhibited large interannual and elevation-dependent variability where the latter likely indicates the distribution of the different species. The longest mean GUD occurred in 2017 (32.7 days).

El seguimiento del patrón de estos índices en el tiempo permite lograr una acertada o factible caracterización de las coberturas presentes en la superficie terrestre [14]. Para el mapeo de coberturas de grandes áreas se utiliza técnicas de correspondencia espectral (SMT), Árboles de decisión y enfoques basados en la transformada de wavelet [8,9,12,13]discrepancies of LCCSs have widely existed in various LCDSs, which have caused negative impacts on comprehensive comparison and integrated utilization of multiple LCDSs. This study attempted to summarize the independent diagnostic criteria hidden in the existing LCCSs based on the induction method, and to synchronously discover the roles of data sources and classification methods in designing LCCSs. A total of 13 existing regional- or global-scale LCDSs were chosen. The analysis results show that phenology, coverage rate, vertical structure, and leaf type were the most frequently adopted criteria in the LCCSs of existing LCDSs. The decision of whether to adopt a diagnostic criterion in the LCCS of LCDS depended on the availability, effectiveness, and quality of the relevant data sources and classification methods. Currently, optical remote sensing images are still the prominent data source for regional- or

global-scale LCDSS, and the potential of each diagnostic criterion could not be fully played. Multi-source and heterogeneous spatial data, ARD (Analysis Ready Data). Las principales limitaciones al utilizar estas técnicas de percepción son: la cobertura de nubes, la configuración espacial, la yuxtaposición de algunos cultivos y la baja resolución espacial de los sensores ópticos de alta resolución temporal [6].

Landsat y sus índices derivados tienen bandas espectrales sensibles a las diferentes coberturas [15,16] the change rules of normalized vegetation index (NDVI; sin embargo, los cambios fenológicos entre las escenas y la baja resolución temporal en los archivos históricos han limitado el mapeo de estas coberturas. En las regiones propensas a nubes como la CARC, es frecuente tener meses sin una escena Landsat sin nubosidad [7]. Las investigaciones recientes han aprovechado los sistemas de detección activa como el radar de apertura sintética (SAR), que además de su alta sensibilidad a las características superficiales y subsuperficiales funciona independientemente de la iluminación solar[17], lo que facilita cartografiar las coberturas presentes sobre las diferentes regiones. Se han desarrollado muchas técnicas de mapeo y la utilización de la información temporal ha sido un enfoque exitoso [18,19]green vegetation cover, leaf area index, and total dry biomass. A paddy rice field in Miyazaki, Japan was studied from April to July of 2018, which is the rice cultivation season. The SAR backscattering coefficients were provided by Sentinel-1 satellite. Backscattering coefficients of two polarization settings-VH (vertical transmitting, horizontal receiving. Esto hace que SAR sea particularmente útil para mapear la extensión de las coberturas considerando el rango dinámico y los ciclos de vida de las especies asociadas a cada cobertura.

Metodología

Se seleccionaron dos imágenes Landsat 8, orto-rectificadas, correspondientes a los path/row 009/057 y 009/058 del 29 de diciembre del 2015, así como, una imagen de radar del satélite Sentinel-1A del 9 de diciembre del 2015. Las escenas de Landsat 8 y Sentinel-1 abarcan la mayor parte de la CARC.

Sentinel-1 brinda una imagen en banda C a 5.405 GHz con un ángulo de incidencia entre 20° y 45° con una órbita circular, a una altura de 693 km, la plataforma 1A tiene una resolución temporal de 12 días respecto al ecuador y tiene cuatro modos de adquisición de imágenes: Stripmap (SM), Inter-ferometric Wide Swath (IW), Extra-Wide (EW) y Wave, con diferentes resoluciones, amplitudes, ángulos de incidencia y polarizaciones.

Este estudio se enfocó en el uso de datos del modo IW en polarización simple VV (transmisión vertical, recepción vertical) para la zona de la CARC. Todos los datos están disponibles en ESA Data Hub, el formato descargado para las imágenes fue Standard Archive Format for Europe (SAFE), que contiene información general del producto en XML para toda la región. Los metadatos contienen la información de las principales características: adquisición, propiedades de la imagen, polarización, fusión de bandas, calibración y ubicación geográfica. Los productos SAFE se obtuvieron como Single Look Complex (SLC) decidiendo trabajar con esta imagen, puesto que los productos GRD Sentinel-1 se obtienen a partir del sistema “slant range” después de aplicar procesos como: multi-looked y projected to ground range usando un modelo elipsoidal de la tierra, lo que supone que la Tierra es igual a una elipsoide, esto es aproximadamente verdadero donde no hay montañas, para el caso de la CARC esto no se cumple, por tal razón la imagen SLC es seleccionada, un proceso de corrección del terreno es realizado sobre la imagen SAR para corregir los datos teniendo en cuenta la topografía del terreno.

Los algoritmos de preprocesamiento aplicados en la imagen SLC fueron: corrección de órbita, calibración, deburst, multilooking y finalmente terrain correction. Posteriormente, se extraen las siguientes características: desviación estándar, promedio, filtros: frost, Boxcar, gammaMap, IDAN, Lee, LeeSigma,

median; texturas como: texturaU, texturaVA, texturaVI, texturaVL, RefinedLee, también el conjunto de características Grey Level Co-occurrence Matrix (GLCM) que abarca atributos de contraste como Disimilaridad, Homogeneidad, Contraste, características de orden como: Momento angular (ASM), Probabilidad máxima (MAX), Entropía (ENT) y características de estadísticas: GLCM Mean, GLCM Variance, GLCM Correlation [20,21], además, se generaron mapas de sombra-escala para ubicar zonas con pérdida de información en la imagen Sentinel-1, sobre estas zonas no se extraen puntos de muestras de coberturas. Para la base de datos óptica se trabajó con un total de 2 escenas Landsat 8, las cuales se usaron para obtener información de reflectancia de la superficie y los índices de vegetación para caracterizar el área de estudio. Todos los datos del Landsat-8 se obtuvieron del Earth Explorer de United States Geological Survey (USGS).

Los índices de vegetación son menos sensibles al ruido de la imagen, a la geometría de visualización y a la atenuación atmosférica, haciéndolos particularmente ventajosos sobre los productos que usan únicamente la información de reflectancia. De las imágenes Landsat-8 se extraen los siguientes índices de vegetación: Normalized difference water index (NDVI), Soil adjusted vegetation index (SAVI), Enhanced vegetation index (EVI), Normalized difference water index (NDWI), Simple ratio (SR), Transformed vegetation index (TVI), Atmospheric resistance vegetation index (ARVI), Structure insensitive pigment index (SIPI), Carotenoid reflectance index (CRI) además, de las bandas 2,3,4,5,6,7 del satélite los cuales han demostrado utilidad en la cartografía de la biomasa (coberturas vegetales) y condiciones superficiales, compensando el brillo del suelo y los artefactos de fondo [7].

Tres procesos de clasificación fueron realizados con las características ópticas y de radar extraídas, a saber: clasificación de características ópticas, clasificación de características de radar y clasificación con características ópticas y de radar, el algoritmo de clasificación utilizado fue Random Forest, este fue escogido porque dio mejores resultados frente a otros métodos de aprendizaje máquina en [6], sin embargo, se realiza una prueba de la clasificación de coberturas utilizando una arquitectura de red neuronal profunda demostrando que Random Forest logra obtener resultados competitivos y computacionalmente menos costosos. Posteriormente, se realiza un proceso de selección de atributos a todas las características, la finalidad es disminuir el espacio de soluciones, el coste computacional e incrementar el porcentaje de acierto en la clasificación.

La cantidad de muestras extraídas para cada clase, la etiqueta y color asignado para cada una de las coberturas se presentan en la Tabla 1.

Tabla 1. Caracterización de las coberturas clasificadas

Cobertura	Cant. muestras	Etiqueta de clase	Color de clase
Agua	2041	0	
Zonas urbanas	2704	1	
Suelo desnudo	2146	2	
Cultivos	3582	3	
Bosque natural	2649	4	
Pastizales	3143	5	
Nubes	--	6	
Sombras de nubes	--	7	

Fuente: Elaboración Propia

En la Tabla 1, se presenta la descripción del conjunto de datos trabajado en esta investigación, las muestras fueron extraídas teniendo en cuenta la técnica de la fotointerpretación relacionada con la respuesta espectral de los índices de vegetación de las imágenes del satélite Landsat-8, por otra parte, la respuesta de la retrodispersión del satélite Sentinel-1 permitió anotar sobre las imágenes los cuerpos de agua.

Resultados

Del análisis de la imagen óptica son extraídas 15 características, 16625 registros o muestras y con el algoritmo de clasificación Random Forest se procede a clasificar las coberturas presentes sobre la CARC con una resolución espacial de 15m, el dataset se divide en 70% para entrenamiento, 30% para prueba, los resultados fueron los siguientes: Instancias correctamente clasificadas igual a 77.22%, coeficiente Kappa igual a 0.72, error medio absoluto de 0.10 y recuerdo de 0.77.

Del análisis de la imagen de radar se obtienen 25 características con una resolución espacial de 15m, los resultados fueron los siguientes: Instancias correctamente clasificadas igual a 60.41%, coeficiente Kappa igual a 0.51, error medio absoluto de 0.16 y recuerdo 0.60.

Hasta el momento los mejores resultados se presentan para la clasificación con la base de datos óptica superando en cada indicador de evaluación a la base de datos de radar, sin embargo, se plantea fusionar las fuentes de información satelital para conseguir mejorar la precisión en la clasificación. Un dataset con un total de 40 atributos es creado, almacenando todas las características extraídas de las imágenes sentinel-1 y Landsat 8, el dataset contiene 16625 muestras, un 70% de ellas se utiliza para entrenamiento y un 30% para prueba. Los resultados obtenidos de la clasificación fueron los siguientes: porcentaje de instancias correctamente clasificadas (ICC) igual a 88.18%, Kappa igual a 0.85, error medio absoluto (EMA) de 0.07 y recuerdo (Recall) de 0.88 respectivamente. La precisión y recuerdo de cada cobertura se presenta en la Tabla 2.

Tabla 2. Datos de precisión y recuerdo para cada cobertura

Cobertura	Etiqueta	Precisión	Recall
Cuerpos de Agua	0	0.912	0.902
Zonas Urbanas	1	0.926	0.948
Suelo Desnudo	2	0.779	0.716
Cultivos	3	0.884	0.904
Bosques	4	0.847	0.920
Pastizales	5	0.914	0.862

Fuente: Elaboración Propia

La Tabla 3. presenta la matriz de confusión del proceso de clasificación basado en el conjunto de características ópticas y de radar.

El proceso de clasificación logra detectar con mayor precisión la cobertura correspondiente a zonas urbanas (clase 1), siendo el valor más alto alcanzado en todo el proceso y para todas las clases (0.926), el recuerdo también es el más alto 0.948 lo que hace que las instancias incorrectamente clasificadas no se distribuyan entre todas las coberturas, este caso es similar a lo que ocurre con las clases correspondientes a pastizales (clase 5) y cuerpos de agua (clase 0) que ocupan el segundo y tercer lugar respectivamente con los valores más altos de precisión. La cobertura de suelo desnudo (clase 2) tiene una precisión y recuerdo aceptable, pero puede ser mejorada enriqueciendo el dataset

Tabla 3. Matriz de confusión para dataset con todas las características

Clases reales	Clasificadas como						Total	Precisión (%)
	0	1	2	3	4	5		
0	645	9	14	0	23	24	715	91.2
1	6	787	16	0	4	17	830	92.6
2	22	21	451	104	22	10	630	77.9
3	0	0	89	1002	17	0	1108	88.4
4	3	0	6	28	709	25	771	84.7
5	31	33	3	0	62	804	993	91.4
Total	707	850	579	1134	837	880	4987	88.2
Recall(%)	90.2	94.8	71.6	90.4	92.0	86.2	88.2	
Kappa	0.857							

Fuente: Elaboración Propia

Se observa que los resultados conseguidos con la unión de los dos dataset óptico y radar superan en cada instancia a la clasificación obtenida individualmente para cada uno de ellos. Con la finalidad de incrementar aún más la precisión y el recuerdo de los resultados y reducir el costo computacional que implica entrenar un dataset de 40 características y 16625 registros, se procede a realizar una selección de atributos, como resultado de este proceso las características que se deben tener en cuenta son, por parte del satélite Sentinel-1: Idan, textura VL, GLCMCorrelation y GLCMMean y por parte del satélite Landsat-8: banda 2,3,4,6 y 7, índice de vegetación CRI y NDWI. Las nuevas métricas de evaluación son: instancias correctamente clasificadas 88.91%, Kappa igual a 0.86, error absoluto medio 0.06, con una precisión y recuerdo general de 0.88 respectivamente.

La precisión y recuerdo de cada cobertura obtenida mediante el clasificador se presenta en la Tabla 4.

Tabla 4. Datos de precisión y recuerdo usando selección de atributos,

Cobertura	Etiqueta	Precisión	Recall
Cuerpos de Agua	0	0.911	0.917
Zonas Urbanas	1	0.928	0.946
Suelo Desnudo	2	0.786	0.737
Cultivos	3	0.896	0.904
Bosques	4	0.861	0.925
Pastizales	5	0.920	0.872

Fuente: Elaboración Propia

En la Tabla 5, se observa la matriz de confusión con selección de atributos, los estadísticos de precisión para cada clase también son mostrados:

Tabla 5. Matriz de confusión para el dataset con selección de atributos,

Clases reales	Clasificadas como						Total	Precisión (%)
	0	1	2	3	4	5		
0	656	6	13	0	21	19	715	91.1
1	11	785	15	0	2	17	830	92.8
2	20	23	464	93	18	12	630	78.6
3	0	0	89	1002	17	0	1108	89.6
4	5	2	5	23	713	23	771	86.1
5	28	30	4	0	57	814	933	92.0
Total	720	846	590	1118	828	885	4987	88.9
Recall(%)	91.7	94.6	73.7	90.4	92.5	87.2	88.9	
Kappa	0.866							

Fuente: Elaboración Propia

Se observa que la selección de atributos mejora las estadísticas generales de la clasificación (ver Tabla 5) incrementando la precisión general y el coeficiente Kappa, además, disminuye el costo computacional (dataset con 12 características). Las zonas urbanas (clase1) es la clase con mayor precisión en la detección. En la Tabla 6, se presenta en resumen todos los resultados de las estadísticas generales de cada proceso: solo óptico, solo radar y unión de las dos fuentes: óptico y radar. Además, se incluye las métricas de evaluación obtenidas para una red neuronal profunda la cual es un algoritmo actual en el proceso de clasificación de las imágenes.

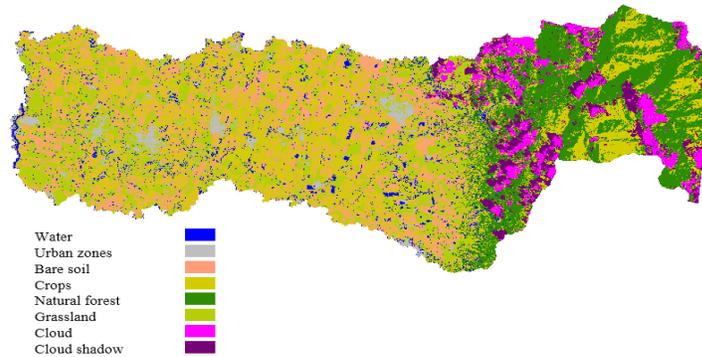
Tabla 6. Datos estadísticos para los distintos dataset (óptico, radar, radar-óptico con y sin selección de atributos),

Fuentes satelitales	ICC (%)	Kappa	EMA	Precisión	Recall
Óptica	77.22	0.72	0.10	0.77	0.77
Radar	60.41	0.51	0.16	0.59	0.60
Óptica y radar sin selección de atributos	88.18	0.85	0.07	0.88	0.88
Óptica y radar con selección de atributos	88.91	0.86	0.06	0.88	0.88
Óptica y radar con selección de atributos con red neuronal profunda	96.96	0.96	0.01	0.96	0.94

Fuente: Elaboración Propia

Aunque el modelo de red neuronal profunda obtuvo las mejores métricas de evaluación, el coste computacional en tiempo fue aproximadamente 6 veces superior. Considerando que la Cuenca Alta del Río Cauca abarca una zona bastante extensa de la geografía colombiana y que los satélites trabajados tienen una resolución temporal de 15 días, la cantidad de información para reconstruir y realizar nuevos entrenamientos es bastante amplia, razón por la cual se decide continuar con el algoritmo de random forest privilegiando el tiempo de respuesta con una precisión que puede ser considerada aceptable. Sin embargo, queda como trabajo a futuro realizar nuevas clasificaciones con marcos de trabajo que exploten la capacidad del cómputo GPU en aprendizaje profundo. La aplicación del modelo entrenado se realiza sobre una subcuenca piloto: cuenca del río Guachal, la reconstrucción de la clasificación se observa en la Figura 1, el estudio se enfoca en gran medida en los cultivos, puesto que estos presentan un alto nivel de variabilidad espacial, y están asociados con la principal actividad social y económica de la región como es la agricultura.

Figura 1. Clasificación de coberturas en la Cuenca del Río Guachal perteneciente a CARC



Fuente: Elaboración Propia

La clasificación que se observa en la Figura 1, tiene una resolución espacial de 15m con lo cual se logra detectar de una mejor forma los límites de separación entre los distintos cultivos, además, se observa visualmente que las clases pastizales, suelo desnudo y cultivos superan en tamaño a los bosques, esta información se puede corroborar calculando el área y porcentaje de ocupación sobre la clasificación final, los datos se muestran a continuación:

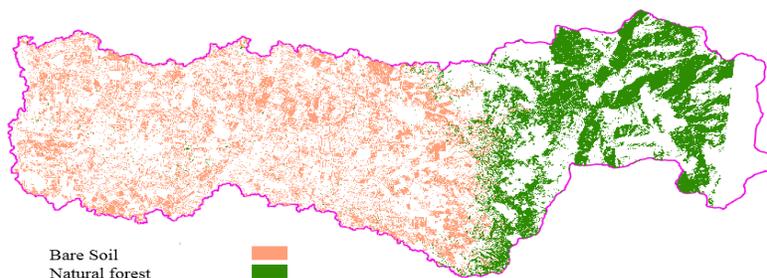
Tabla 7. Área de ocupación de las coberturas sobre la cuenca del río Guachal

Cobertura	Área (ha)	Ocupación efectiva (%)
Cuerpos de agua	2336.65	4.33
Zonas urbanas	1304.69	2.42
Suelo desnudo	11217.5	20.03
Pastizales	14940.3	26.03
Bosque natural	10045.1	18.64
Cultivos	14024.6	27.73
Nubes	2793.71	N. A
Sombra de nubes	2885.08	N. A
Total	59547.5	100.00
Área real	53868.7	

Fuente: Elaboración Propia

Los Bosques solo ocupan el 18.64% del total de la cuenca Guachal, un resultado que inquieta puesto que una cobertura tan importante para los diferentes procesos ambientales es menor que los altos niveles de fragmentación de los corredores con bosque natural. Es apreciable la diferencia entre las áreas de expansión agrícola, como se puede observar en la Figura 2, los suelos desnudos tienden a ocupar más área, haciendo evidente la diferencia en uso del suelo entre la zona plana y la zona de pendientes (ver Figura 4).

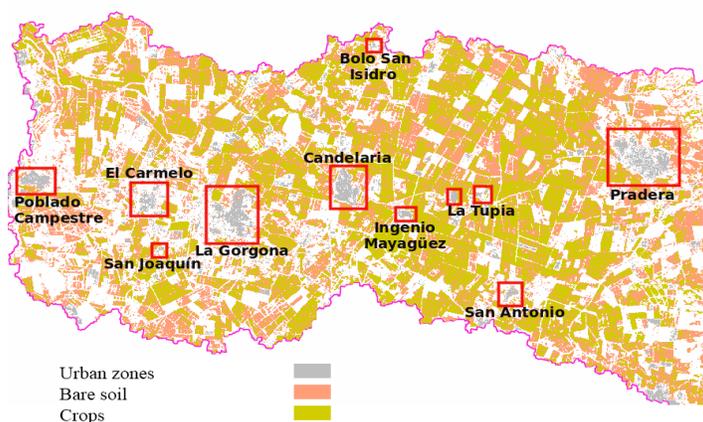
Figura 2. Suelo desnudo y bosques naturales



Fuente: Elaboración Propia

Los suelos desnudos tienen una presencia importante, lo cual potencia los procesos erosivos y es una respuesta a la dinámica de cambios de cultivos en la cuenca del Río Guachal, en la Figura 3 se resaltan los suelos desnudos, cultivos y zonas urbanas, el cálculo del área de las zonas urbanas se presenta en la Tabla 8.

Figura 3. Ciudades, suelo desnudo y cultivos en la zona de la Cuenca del río Guachal



Fuente: Elaboración Propia

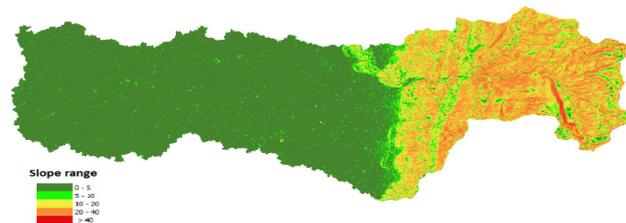
Tabla 8. Área de ocupación de las zonas urbanas sobre la cuenca del río Guachal

Zonas Urbanas seleccionadas	Área (ha)
Pradera	154.395
Ingenio Mayagüez	14.535
La Tupia	23.152
Candelaria	87.885
La Gorgona	156.285
El Carmelo	57.465
San Joaquín	7.042
Poblado Campestre	56.79
Bolo San Isidro	3.712

Fuente: Elaboración Propia

En la Figura 4 un mapa de pendientes se anexa, donde se observa que la zona de cultivos y zonas urbanas se encuentran en los rangos de 0 a 5% que corresponden a una zona plana.

Figura 4. Mapa de pendientes en la Cuenca del río Guachal



Fuente: Elaboración Propia

Basados en la información de la Tabla 5 aproximadamente 14000 hectáreas están dedicadas a actividades de siembra, manutención y cosecha de cultivos que representan la dinámica económica de la cuenca. Esta condición de uso y apropiación también denota la existencia de conflictos socioambientales por las actividades humanas y plantean una presión ambiental sobre los recursos naturales de estos ecosistemas específicamente en la oferta y demanda de recurso hídrico. En este sentido, el territorio donde se cultiva la mayor parte de la caña de azúcar del país está ubicado en la CARC, específicamente en el Valle del Cauca, con una disponibilidad anual promedio de 467 m³/s, siendo aportados 272 m³/s por los diferentes ríos del departamento del Valle del Cauca. Esta zona se caracteriza por tener una elevada demanda de agua, la caña de azúcar tiene un uso intensivo del agua (10.300 m³/ha Vs. 3.600 m³/ha y 3.400 m³/ha respectivamente para el maíz y la soja) [22], concentra sobre ella la mayor parte de la población y la actividad económica del departamento. Así, de los 4,6 millones de habitantes del departamento del Valle del Cauca, 3,6 millones vivían en la zona plana. En este escenario, el incremento del consumo de agua por el cultivo de caña, al pasar este de 0,7 a 1,9 Gm³. Representa un crecimiento promedio anual de 6,9%, muy superior a la dinámica de consumo de agua agrícola total. Esta situación genera una creciente concentración del recurso hídrico (y

del suelo) en manos de un solo sector productivo: la caña de azúcar. Esta condición se refleja en la pérdida de coberturas de esta cuenca y se verifica con el uso intensivo del suelo, lo cual es verificable a partir del análisis espacial y computacional realizado.

Conclusiones

Un proceso de clasificación basado en el algoritmo Random Forest, utilizando imágenes satelitales ópticas y de radar fue realizado, uniendo las dos fuentes de información junto con una selección de atributos, la precisión lograda fue del 88.9%, obteniendo precisiones superiores al 91% en 3 de las 6 clases evaluadas. Con respecto a la selección de atributos se comprueba que las características superiores al 70% de importancia son las responsables del incremento en las métricas de evaluación y disminución del coste computacional. La investigación logra con éxito vincular el aprendizaje máquina, la percepción remota satelital y el componente ecológico para generar una aplicación con impacto ambiental para Colombia a través del estudio de las coberturas vegetales de la CARC. La cuenca del río Guachal es caracterizada por tener un clima propicio para la agricultura, por lo tanto, esta evaluación de las coberturas permite ver la diferenciación entre las áreas con geomorfología montañosa donde las pendientes determinan la utilización de procesos mecanizados de siembra, lo cual es evidente por el tipo de cultivo: caña de azúcar.

Como trabajo a futuro, se plantea extraer nuevas características que involucren las dos polarizaciones del satélite Sentinel-1, así mismo, enriquecer el conjunto de datos de entrenamiento para balancear el dataset y mejorar el poder de clasificación, además, seguir aprovechando las ventajas del aprendizaje máquina para implementar un análisis temporal con la finalidad de ver la evolución de las coberturas vegetales. Finalmente, teniendo como base los resultados relacionados con el aprendizaje profundo, se plantea realizar las clasificaciones con librerías como Tensorflow y Keras que permiten aprovechar el computo paralelo sobre GPU.

Agradecimientos

Los autores agradecen a la Universidad del Cauca y al proyecto “Water Security and Sustainable Development Hub” financiado por UKRI GCRF por el apoyo en la realización del proyecto.

Referencias bibliográficas

1. M. A. Munnaf, G. Haesaert, M. Van Meirvenne, and A. M. Mouazen, “Site-specific seeding using multi-sensor and data fusion techniques: A review,” in *Advances in Agronomy*, vol. 161, Academic Press Inc., 2020, pp. 241–323.
2. G. Vargas, N. León, and Y. Hernández, “Agricultural Socio-economic Effects in Colombia due to Degradation of Soils,” in *Sustainable Management of Soil and Environment*, Springer Singapore, 2019, pp. 289–337.
3. C. Martinez, A. Jarvis, C. Navarrete, N. Riaño, and A. Figueroa-Casas, “Sector Agrícola de la Cuenca Alta del Río Cauca Altamente Vulnerable al Cambio Climático,” 2014. [Online]. Available: https://cdkn.org/wp-content/uploads/2012/08/Brief_General-Cuenca-alta-rio-cauca.pdf.
4. J. Muñoz, C. Cobos, and M. Mendoza, “Vegetation index based on genetic programming for bare ground detection in the Amazon,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Oct. 2018, vol. 10633 LNAI, pp. 259–271, DOI: 10.1007/978-3-030-02840-4_21.

5. J. Muñoz, L. Pencue, A. Figueroa, and C. Guzmán, "Crop monitoring in high andean ecosystems of the upper basin of the palacé river using planet images," in *Advances in Intelligent Systems and Computing*, Nov. 2018, vol. 687, pp. 155–169, DOI: 10.1007/978-3-319-70187-5_12.
6. T. Talema and B. T. Hailu, "Mapping rice crop using sentinels (1 SAR and 2 MSI) images in tropical area: A case study in Fogera wereda, Ethiopia," *Remote Sens. Appl. Soc. Environ.*, vol. 18, Apr. 2020, DOI: 10.1016/j.rsase.2020.100290.
7. E. L. Pencue-Fierro, Y. T. Solano-Correa, J. C. Corrales-Muñoz, and A. Figueroa-Casas, "A Semi-Supervised Hybrid Approach for Multitemporal Multi-Region Multisensor Landsat Data Classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 12, pp. 5424–5435, Dec. 2016, DOI: 10.1109/JSTARS.2016.2623567.
8. G. Lei, A. Li, J. Bian, and Z. Zhang, "The roles of criteria, data and classification methods in designing land cover classification systems: evidence from existing land cover data sets," *International Journal of Remote Sensing*, vol. 41, no. 14. Taylor and Francis Ltd., pp. 5062–5082, Jul. 17, 2020, DOI: 10.1080/01431161.2020.1724349.
9. S. Ji, Z. Zhang, C. Zhang, S. Wei, M. Lu, and Y. Duan, "Learning discriminative spatiotemporal features for precise crop classification from multi-temporal satellite images," *Int. J. Remote Sens.*, vol. 41, no. 8, pp. 3162–3174, Apr. 2020, DOI: 10.1080/01431161.2019.1699973.
10. C. Huang et al., "Land Cover Mapping in Cloud-Prone Tropical Areas Using Sentinel-2 Data: Integrating Spectral Features with Ndvi Temporal Dynamics," *Remote Sens.*, vol. 12, no. 7, p. 1163, Apr. 2020, DOI: 10.3390/rs12071163.
11. A. Kern, H. Marjanović, and Z. Barcza, "Spring vegetation green-up dynamics in Central Europe based on 20-year long MODIS NDVI data," *Agric. For. Meteorol.*, vol. 287, p. 107969, Jun. 2020, DOI: 10.1016/j.agrformet.2020.107969.
12. M. K. Gumma et al., "Mapping cropland fallow areas in myanmar to scale up sustainable intensification of pulse crops in the farming system," *GIScience Remote Sens.*, vol. 55, no. 6, pp. 926–949, Nov. 2018, DOI: 10.1080/15481603.2018.1482855.
13. R. Simoes et al., "Land use and cover maps for Mato Grosso State in Brazil from 2001 to 2017," *Sci. Data*, vol. 7, no. 1, pp. 1–10, Dec. 2020, DOI: 10.1038/s41597-020-0371-4.
14. M. C. A. Picoli et al., "Big earth observation time series analysis for monitoring Brazilian agriculture," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 328–339, Nov. 2018, doi:10.1016/j.isprsjprs.2018.08.007.
15. Y. Bai, J. Gao, and B. Zhang, "Monitoring of Crops Growth Based on NDVI and EVI," *Nongye Jixie Xuebao/Transactions Chinese Soc. Agric. Mach.*, vol. 50, no. 9, pp. 153–161, Sep. 2019, DOI: 10.6041/j.issn.1000-1298.2019.09.017.
16. Y. Lin, X. Hu, R. Qiu, Z. Zhang, Q. Lin, and J. Lin, "Responses of Landsat-based NDVI to Interaction of Vegetation and Influencing Factors," *Nongye Jixie Xuebao/Transactions Chinese Soc. Agric. Mach.*, vol. 49, no. 10, pp. 212–219, Oct. 2018, DOI: 10.6041/j.issn.1000-1298.2018.10.024.
17. A. Ferro, "Advanced methods for the analysis of radar sounder and VHR SAR signals," University of Trento, 2011.
18. E. Wali, M. Tasumi, and M. Moriyama, "Combination of linear regression lines to understand the response of sentinel-1 dual polarization SAR data with crop phenology-case study in Miyazaki, Japan," *Remote Sens.*, vol. 12, no. 1, 2020, DOI: 10.3390/rs12010189.

19. M. A. Merchant, R. K. Warren, R. Edwards, and J. K. Kenyon, "An Object-Based Assessment of Multi-Wavelength SAR, Optical Imagery and Topographical Datasets for Operational Wetland Mapping in Boreal Yukon, Canada," *Can. J. Remote Sens.*, vol. 45, no. 3–4, pp. 308–332, Jul. 2019, DOI: 10.1080/07038992.2019.1605500.
20. R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural Features for Image Classification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973, DOI: 10.1109/TSMC.1973.4309314.
21. A. Darzi, I. Lang, A. Taklikar, H. Averbuch-Elor, and S. Avidan, "Co-occurrence Based Texture Synthesis," May 2020, Accessed: May 25, 2020. [Online]. Available: <http://arxiv.org/abs/2005.08186>.
22. D. I. V. Danny Ibarra Vega, C. Peña Rincón, J. Valencia Calvo, J. M. Redondo, and G. Olivar Tost, "Water resources planning in bioethanol production from sugarcane," *Memorias*, no. 1, pp. 30–35, Nov. 2018, DOI: 10.22490/25904779.3055.