TRABAJO DE GRADO
ANÁLISIS Y EXTRACCIÓN DE IMÁGENES SATELITALES EN LA NUBE POR MEDIO DE GOOGLE EARTH ENGINE

SOFÍA CARVAJAL GÓMEZ

ISABELLA RODRÍGUEZ FRANCO

ÓSCAR CARDONA MORALES

UNIVERSIDAD CATÓLICA DE

MANIZALES FACULTAD DE

INGENIERÍA Y ARQUITECTURA

INGENIERÍA DE

TELECOMUNICACIONES

FEBRERO 2020

1. CONTENIDO

2. Resumen 5

3. **Introducción** 5

4. **Marco Teórico** 5

4.1. Adquisición y Preparación de la Información 9

4.2. Análisis e Interpretación de Cobertura 10

4.3. Verificación de Campo 10

4.4. Control de Calidad 10

4.5. Generación de la Capa 1:100.00 11

4.6. Técnicas de Clasificación Terrestre 11

4.6.1. Naive Bayes 11

4.6.2. Random Forest 11

4.6.3. Support Vector Machine 12

4.6.4. Eo-Learn 12

4.6.5. Classification and Regressation Tree (CART) 13

4.6.6. GMO max entropy 13

4.7. Técnicas de Clasificación Empleadas en Google Earth Engine 14

4.8. Revisión de los Sensores y Datos Disponibles Para Usar en la Plataforma Google Earth Engine 15

4.9. Resolución de un Sistema Sensor 15

4.9.1. Resolución Espacial 15

4.9.2. Resolución Espectral 16

4.9.3. Resolución Radiométrica 16
4.9.4. Resolución Temporal 16

4.10. Tipos de Sensores 18
4.10.1. Sensores Activos 18
4.10.2. Sensores de Radar 18
4.10.3. Lidar 19
4.10.4. Sensores Pasivos 19
4.10.5. Sensores Fotográficos 19
4.10.6. Exploradores de Vidicon 19
4.10.7. Radiómetros de Microondas 20
4.10.8. Exploradores de Barrido y Empuje 20

4.11. Medidas de Error 20

4.12. Medidas Estadísticas 21
4.12.1. Matriz de Confusión 21
4.12.2. Precisión de Producto 21
4.12.3. Precisión del Usuario 22

4.13. Normalización Matricial 22

4.14. Análisis Multiderivado 23
4.14.1. Coeficiente Kappa 23

5. **Marco Experimental** 27
5.1. Zona de Estudio 27

6. **Metodología** 29

7. **Resultados** 34
7.1. Clasificación 1 34
7.2. Clasificación 2 40

8. **Conclusiones** 48

9. **Bibliografía** 49

ANÁLISIS Y EXTRACCIÓN DE IMÁGENES SATELITALES EN LA NUBE POR MEDIO DE GOOGLE EARTH ENGINE

2. RESUMEN

El proyecto de investigación está basado en el estudio comparativo entre métodos de clasificación. Con el fin de realizar un análisis multitemporal, a su vez implementar herramientas y plataformas tecnológicas que permitan el monitoreo de grandes superficies de tierra, donde los datos son provenientes de satélites, pero por la cantidad y flujo de los mismos, al ser obtenidos se hace dificultoso su análisis y extracción, dando origen a un problema de Big data. Es por ello que mediante Google Earth Engine, plataforma que contiene métodos de clasificación como Random Forest, GMO max entropy y CART se pretende conocer cuál es más eficiente para clasificar zonas con distintas categorías

3. INTRODUCCIÓN

Durante las últimas décadas, las tecnologías de observación para la visualización de tierras mediante satélites se han convertido en las más adecuadas para hacer más eficaz su seguimiento, al proporcionar imágenes de la superficie terrestre a intervalos regulares [1]. Debido a la falta de supervisión a grandes áreas de tierra, utilizadas por los seres humanos para diferentes actividades como cultivos, minería, ganadería, pesca, entre otras, las cuales se desarrollan en pro del crecimiento urbano y rural. Por tal razón surge la necesidad de implementar herramientas y plataformas tecnológicas quienes permiten el estudio y monitoreo de áreas extensas mediante imágenes satelitales almacenadas en la nube. Aunque con el aumento de la variedad y volúmenes de datos provenientes de satélites, en donde en algunos casos la resolución de la imagen no es muy buena y la disponibilidad de esta no es alta., va generando un problema de Big Data, haciendo compleja la extracción de información verdaderamente importante [2] para estudios e

investigaciones que realizan entidades como la Universidad Católica de Manizales y el Instituto Geográfico Agustín Codazzi (IGAC).

Google Earth Engine es una plataforma que almacena imágenes diariamente desde hace cuarenta años de datos geoespaciales a diferentes escalas [3]. A su vez proporciona métodos de clasificación basados en píxeles los cuales son utilizados para mapas y cultivos. A pesar de que se encuentra bien documentado, no hay estudios previos que comparen los métodos para clasificación y análisis multitemporal a grandes escalas [2]. Las tecnologías de observación de la tierra por medio de satélites han facilitado tareas como el seguimiento en el campo. Lo cual combinado con las tecnologías de información como el internet y la telefonía móvil, posibilitan que la información del estado del cultivo llegue de forma más rápida, fluida y personalizada al agricultor [1]. En Argentina y Chile se realizan múltiples estudios acerca de las estructuras térmicas superficiales de los océanos, hacia un análisis de la temperatura natural del mar (TSM) y logrando determinar qué fenómenos o eventos afectan la estructura del TSM en la franja costera [4].

El artículo busca explorar la eficiencia de dos métodos de clasificación (CART, GMO max entropy y Random Forest), en el uso de la plataforma Google Earth Engine, hacia una comparativa, conociendo cuál de éstos es más eficiente a la hora de clasificar una zona conformada por diversas categorías como lo son bosque, vegetación, zona urbana, suelo desnudo y agua.

4. MARCO TEÓRICO

Técnicas de clasificación en conbertura terrestre Corine Land Cover Colombia Corine(Coordination of information on the environment) es un proyecto de cobertura de

la tierra, que definió una metodología para realizar un inventario homogéneo de la cobertura de la tierra a partir de imágenes satelitales. Éste método se convirtió en un estándar para la generación de información sobre coberturas y cambios en el uso del suelo [5]. Colombia adaptó esta metodología en el 2005 [5], estandarizando un sistema de clasificación, el cual posee categorías jerárquicas y a su vez estas son dadas por medio de la información que proporcionan las imágenes. Corine Land Cover permite describir, caracterizar, clasificar y comparar las características de la tierra mediante técnicas de interpretación, e información auxiliar de otras fuentes provenientes de una base de datos que tiene en cuenta los procedimientos tecnológicos utilizados anteriormente y los métodos utilizados actualmente [7]. Durante este proceso se definen 5 fuentes de error:

- Material utilizado.
- Integración de los datos.
- Interpretación.
- Digitalización
- Transformación (Homogeneidad en sistemas de coordenadas y proyección).

El factor de error más crítico en la elaboración y recopilación de información para la base de datos CLC2000, es la exactitud espacial en la toma de imágenes satelitales [7]. El CLC2000 tiene una nomenclatura de tres niveles, donde se definen las coberturas y los usos del suelo, la unidad mínima de medida es de 25 hectáreas, exceptuando para los territorios artificializados donde la medida mínima es de 5 hectáreas y donde cualquier desplazamiento mayor a 100 metros se considera un error [8].

Colombia cuenta con dos mapas nacionales completos de coberturas, utilizando esta tecnología a una escala de 1:100.000. Corine Land Cover divide los territorios artificializados por coberturas como lo indica la tabla 1.

LEYENDA NACIONAL DE COBERTURAS DE LA TIERRA

1. TERRITORIOS ARTIFICIALIZADOS	3. BOSQUES Y ÁREAS SEMINATURALES
1.1 Zonas urbanizadas	**3.1 Bosques**
1.1.1 Tejido urbano continuo	3.1.1 Bosque denso
1.1.2 Tejido urbano discontinuo	3.1.2 Bosque abierto
1.2 Zonas industriales o comerciales y redes de comunicación	3.1.3 Bosque fragmentado
1.2.1 Zonas industriales o comerciales	**3.2 Áreas con vegetación herbácea y/o arbusiva**
1.2.2 Red vial, ferroviaria y terrenos asociados	3.2.1.1 Herbazal denso
1.2.3 Zonas portuarias	3.2.1.2 Herbazal abierto
1.4 Zonas verdes artificializadas no agrícolas	3.2.2.1 Arbustal denso
1.4.1 Zonas verdes urbanas	**3.3 Áreas abiertas, con o sin poca vegetación**
1.4.2 Instalaciones recreativas	3.3.1 Zonas arenosas naturales
	3.3.2 Afloramientos rocosos
2. TERRITORIOS AGRÍCOLAS	3.3.3 Tierras desnudas y degradadas
2.1 Cultivos transitorios	
2.1.1 Otros cultivos transitorios	4. ÁREAS HÚMEDAS
2.1.2 Cereales	**4.1 Áreas húmedas continentales**
2.1.3 Oleaginosas y leguminosas	4.1.1 Zonas pantanosas
2.2 Cultivos permanentes	4.1.2 Turberas
2.2.1 Cultivos permanentes herbáceos	4.1.3 Vegetación acuática sobre cuerpos de agua
2.2.1.1 Otros cultivos permanentes herbáceos	**4.2 Áreas húmedas costeras**
2.2.1.2 Caña	4.2.1 Pantanos costeros
2.2.1.3 Plátano y banano	4.2.2 Salitral
2.3 Pastos	4.2.3 Sedimientos expuestos en bajamar
2.3.1 Pastos limpios	
2.3.2 Pastos arbolados	5. SUPERFICIES DE AGUA
2.3.3 Pastos enmalezados	**5.1 Aguas continentales**
2.4 Áreas agrícolas heterogéneas	5.1.1 Ríos (50m)
2.4.1 Mosaico de cultivos	5.1.2 Lagunas, lagos y ciénagas naturales
2.4.2 Mosaico de pastos y cultivos	5.1.3 Canales
2.4.3 Mosaico de cultivos, pastos y espacios naturales	**5.2 Aguas marítimas**
	5.2.1 Lagunas costeras
	5.2.2 Mares y océanos
	5.2.3 Estanques para acuicultura marina

Tabla 1. Leyenda Nacional de coberturas de la tierra.[9]

El esquema metodológico de Corine Land está copuesto por seis etapas, ver ilustración 1.

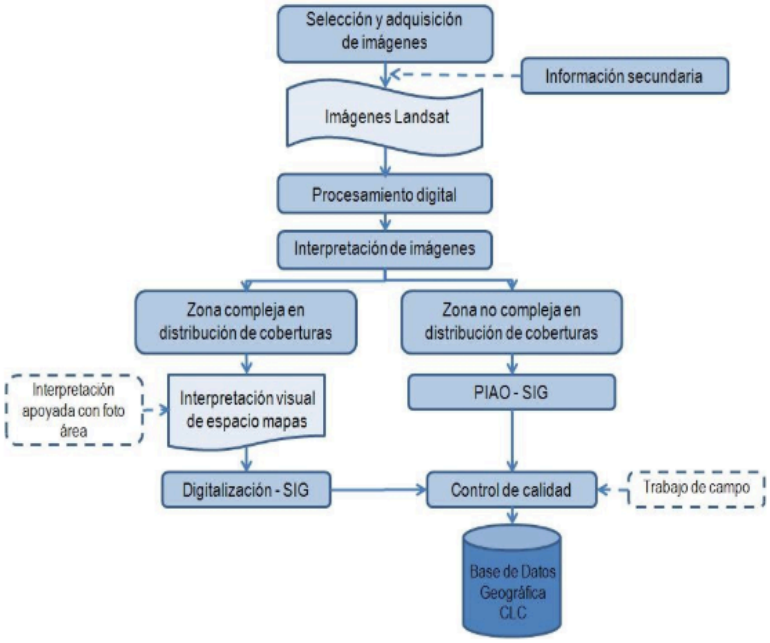

Ilustración 1. Metodología Corine Land Cover [10]

4.1 Adquisición y Preparación De La Información

La generación de las coberturas de la tierra, son mediante las imágenes Landsat o mediante el Banco Nacional de Imágenes. Éstas son utilizadas ya que cubren una gran parte de la superficie terrestre, contando con una frecuencia de adquisición de 16 días (35.000 Km 2 por escena) para Landsat y 26 días (3.600 Km2) para Spot. Al estar en formato digital, se hace necesario el uso de diferentes tipos de información, así comparar, completar y hacer la validación de las imágenes de referencia, sobre saltando imágenes satelitales, sensores, cartografía, topografía, mapas de cobertura del territorio y su ocupación. Finalmente recurriendo al proceso de transformación de la imagen satelital en proyección ortogonal, eliminando los efectos del relieve.

4.2 Análisis e Interpretación De Cobertura

Para iniciar la interpretación de las imágenes por satélite es necesario visualizar en una computadora la cual posea de Arcgis 9.3, para hacer una configuración e ir delineando las unidades del mapeo, seguido a ello, haciendo un mejoramiento espectral mediante cortes o proyecciones, mediante el software de Erdas 8.5.

4.3 Verificación De Campo

Se debe tener en cuenta la diversidad de cobertura de la tierra, por ende, se seleccionan zonas piloto de diferentes sectores del área de estudio. Para ello, se deben cumplir ciertas características, una de ellas es que sea de buen acceso para garantizar y tener control en el área de trabajo, seguido de que se puedan encontrar en la región biogeográfica todas las unidades de paisaje, para que al final se puedan aclarar las dudas que deja el proceso de interpretación de la cobertura, junto a su nomenclatura en determinada zona del estudio.

4.4 Control De Calidad

Proceso mediante el cual se van haciendo continuas correcciones al sistema para revisar si cada una de las etapas del mismo van teniendo un avance y garantizando la calidad de los datos. Comprendido en dos partes, la primera trata de una revisión a la plancha, en formato análogo

mediante la impresión de las fotografías para hacer observaciones y correcciones. El segundo paso es la plancha en formato digital,

quien facilita el proceso de control de calidad mediante los ajustes que se realizan en la capa de información de puntos.

4.5 Generación De La Capa Temática Escala 1:100.000

Mediante la información que cada intérprete genera, se van dando las coberturas, quienes obtienen atributos, códigos y nomenclaturas. Las Cuales se van anexando a una geodatabase, con el fin de estandarizar e ir articulando los objetos en un único esquema, con el fin de generar reportes de información.

4.6 TÉCNICAS DE CLASIFICACIÓN TERRESTRE

4.6.1 Naive Bayes: Se basa en el teorema de Bayes, el cual es fácil de implementar y no tiene hiper parámetros. Para los datos de entrada se asume la distribución Gaussiana.

4.6.2 Random Forest: Es un conjunto de métodos de aprendizaje que genera múltiples modelos débiles y unifica los resultados para obtener una buena predicción. Para realizar lo anterior, el método realiza una profusión de parámetros, los principales parámetros los ajusta en trios y el número de características los usa en cada punto de quiebre. Para estimar el valor de los tres números, se busca que éstos tengan un espacio entre 10 y 30. Lo anterior depende de la característica que se requiera analizar, entre más relevante sea ésta más pequeña deberá de ser el valor de diferencia, pero entre más irrelevante sea la característica el valor podrá ser mayor.

4.6.3 Support Vector Machine (SVM): Es un clasificador que separa las clases en dos maximizando el margen entre ellas. Éste se incorpora en tres kernels: lineal, polinomial y RBF. Kernel lineal el cual es usado esencialmente para problemas linealmente separables y trabaja bien cuando el número de características aumenta; En éste la regulación del parámetro c se encuentra optimizada por lo que resulta fácil calcularlo. Aunque cuando la regulación del parámetro c aumenta, la decisión del límite de clasificación puede ser más complicada de tolerar/evitar de forma correcta. Adicionalmente a la regulación del parámetro c, el grado (en el caso polinomial) y el ancho (en el caso de RBF) de la función Gaussiana γ debe de ser optimizada, la cual resulta más difícil ya que cuando γ es más sensible para crear el límite. En la ilustración 2 se puede observar un ejemplo de clasificación de imágenes utilizando SVM-RF, el proceso que representa la ilustración 2 se divide en tres partes: en la primera, la resolución multiespectral de la imagen y los datos de la geometría 3D son usados para extraer las características, un total de 24 diferentes características son tomadas de cada pixel. En la segunda, se usan éstos pixeles para entrenar el modelo de RF. Finalmente, los resultados del ruido de la clasificación son la entrada para el modelo FCCRF, y un un rango de interferencias de dependencias son usados para mejorar los resultados de la clasificación.

4.6.4 Eo-Learn: Procesamiento multitemporal para la clasificación de uso y cobertura de la tierra, permitiendo dividir en parches más pequeños, procesando los recursos computacionales limitados debido a que la opción de división depende de la cantidad de recursos disponibles, el cual se divide en dos partes para poder comprender de mejor manera los datos para que seguido a ello, continuar con el clasificador:

- La primera parte trata acerca del proceso de selección y división del área de interés, de ésta forma obtener información que corresponde hacia los datos de banda, Sentinel-2 y máscaras de nube con el fin de obtener una clasificación confiable.
- En la segunda parte se han de preparar los datos para el aprendizaje automático, haciéndole un muestreo al azar a un subconjunto específico de los píxeles, probando y filtrando escenas las cuales están nubladas, finalmente realizando una interpolación lineal en la dimensión temporal con el fin de ir llenando espacios y al estar preparados los datos, validarlos y mostrarlos en las gráficas.

Finalmente, al obtener los cuadros delimitados de los parches vacíos en su lugar, se permite una descarga automática de imágenes en Sentinel.

4.6.5 Classification and Regression Tree (CART): Un clasificador de árbol de decisión, se construye a partir de un conjunto de datos de entrenamiento utilizando el concepto de entropía de información. En cada nodo del árbol, se selecciona un atributo de los datos que divide más eficazmente su conjunto de muestras en subconjuntos enriquecidos en una clase u otra. Su criterio es la ganancia de información normalizada que resulta de elegir un atributo para dividir los datos. El atributo con la mayor ganancia de información normalizada se elige para tomar la decisión. El algoritmo se repite en las sub-listas más pequeñas. Una desventaja del clasificador DT es la considerable sensibilidad al conjunto de datos de entrenamiento, de modo que un pequeño cambio en los datos de entrenamiento puede resultar en un conjunto muy diferente de subconjuntos [2].

4.6.6 GMO Max Entropy: La regresión logística multinomial es entendida como la generalización de la regresión lineal, que utiliza la función de transformación softmax y la tarea principal es minimizar una función de error, lo hace tomando el logaritmo negativo de la probabilidad, lo que significa entropía cruzada. La principal diferencia con respecto a otros

modelos y algoritmos es el puntaje de resultado que podría considerarse como un valor de probabilidad [14].

4.7 TÉCNICAS DE CLASIFICACIÓN EMPLEADAS EN GOOGLE EARTH ENGINE

Google Earth Engine es una API, una interfaz gráfica que descarga imágenes correspondientes al estudio de áreas grandes. GGE combina más de 40 años en imágenes de satélites del planeta [8], también utiliza herramientas para analizar dicha información. Éste extrae la información de cientos de bases de datos sobre imágenes satelitales y geoespaciales, entre ellas la base de datos de CORINE, esta aplicación también permite cargar una base de datos externa para analizar. GGE utiliza el catálogo de bases de datos para analizar y comparar imágenes satelitales.

Para analizar una imagen se puede seleccionar las bandas en las que se desee trabajar para comenzar a filtrar y dejar solamente la información relevante.

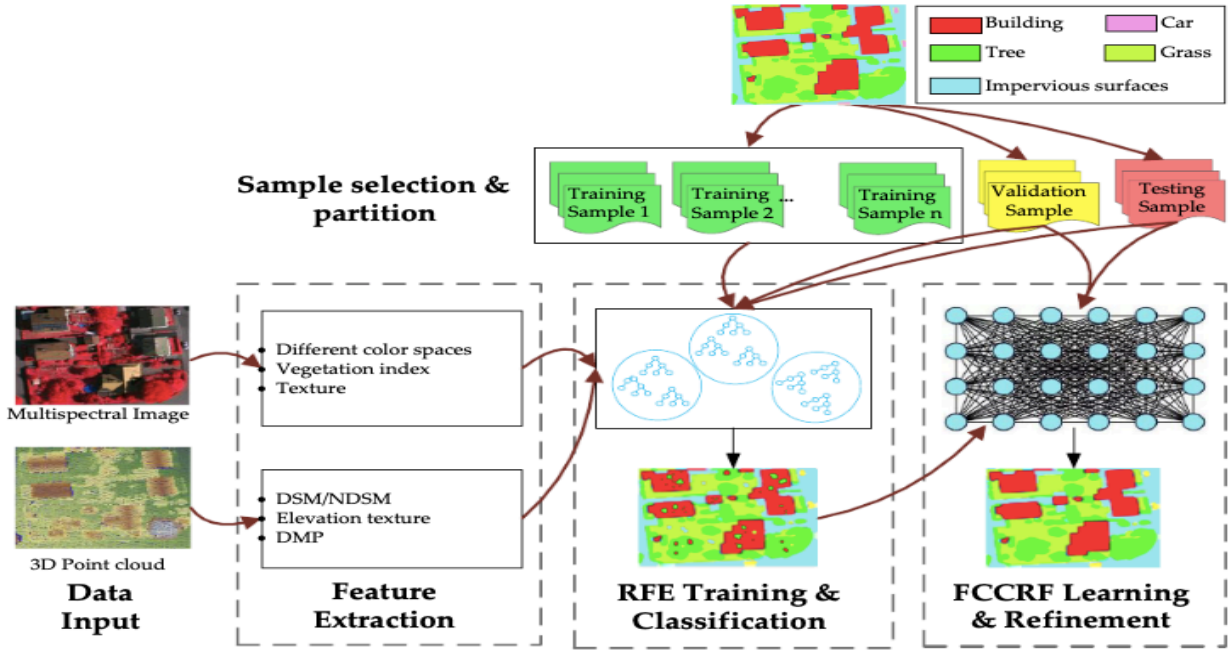

Ilustración 2. Ejemplo de clasificación SVM-RF

La API de Google Earth Engine utiliza dos lenguajes de programación, JavaScript y Python estos lenguajes son utilizados para realizar peticiones a los servidores de Earth Engine. Existen ejemplos almacenados en el repositorio de GitHub. Uno de los métodos más utilizados por Google Earth Engine es SVM (Support vector machine)

4.8 REVISIÓN DE LOS SENSORES Y DATOS DISPONIBLES PARA USARSE EN LA PLATAFORMA GOOGLE EARTH ENGINE (GEE)

Los sensores poseen ciertas características que por medio de ellas se permite analizar y concluir, dependiendo de cada satélite. Existen dos tipos de clasificar los sensores, pasivos y activos. El sensor pasivo trata de cuando existe un límite para recibir la energía que proviene de un foco externo, por otra parte los sensores activos son aquellos capaces de emitir su propio haz de energía, para que en el momento de diseñar un sensor se deba resaltar qué tipo de resolución es la que más se requiere

para su propósito, porque todas las resoluciones están relacionadas entre sí y por el incremento de alguna se pueden encontrar problemas a la hora de transmitir la imagen a la superficie, o sea que incrementa el volumen de datos lo que también crea un conflicto a la hora de procesar. Seguido a ello sobresalen algunos conceptos previos con los cuales se pueden comparar los distintos satélites.

4.9 Resolución de un Sistema Sensor: Trata del registro o de la información detallada, la cual implica el espacio, espectro, radiometría y tiempo.

4.9.1 Resolución Espacial: Su medida en unidades de longitud, a la separación mínima entre objetos separados de la fotografía. Dependiendo los factores de longitud como la altura de la

cámara y la superficie, se entiende que es la variación en función de la distancia del satélite y el ángulo que se observa, éstos son aspectos necesarios para una buena interpretación de los datos, debido a que en algunos casos según la actividad que se requiera son diferentes tipos de resoluciones, que a su vez corresponde a la información que se incluye en la imagen, designando al objeto más pequeño a distinguir en la imagen, la cual se determina por el tamaño del píxel, quien es medido en metros sobre el terreno (donde depende la altura, ángulo de visión y velocidad del sensor).

4.9.2 Resolución Espectral: Por medio de éste se conoce el número y el ancho de las bandas espectrales (número de canales espectrales), debido a que entre más bandas proporcione un sensor, se volverá más idóneo, a su vez permite la caracterización de las diferentes cubiertas, que son dadas de acuerdo del propósito del sensor. Por ende, a menor ancho espectral mayor es su resolución, pero se reduce la relación señal-ruido, por lo que es necesario encontrar una solución para equilibrar ambas magnitudes, ya que en algunos casos cuando el ancho de banda es incorrecto el resultado es muy pobre o inválido.

4.9.3 Resolución Radiométrica: Conocida como sensitividad radiométrica ó dinámica, por medio de ella se conocen los bits por píxel y niveles de grises en los que se divide la radiación para ser almacenada y procesada, siendo variable según la banda espectral, también existiendo una relación con la señal-ruido y entre lo que se quiere captar con la cantidad de información que se transmite con valores 64, 128, 256 y 1024 niveles.

4.9.4 Resolución Temporal: Entendido como el intérvalo de tiempo ó frecuencias del satélite por la misma zona, para obtener muestras de la zona de cobertura terrestre, las cuales dependen de

las características y diseño de la órbita terrestre y/o sensor, que a su vez varían según los objetivos que se fijan en el sensor.

	Landsat 8	Sentinel 2
Resolución espacial	ETM+ 30m en las bandas de visible y cercano medio, 60m en el inflarrojo o térmico y 15m en la panorámica	en las bandas 2, 3, 4 y 8 el tamaño del pixel es de 10m y en las bandas 5, 6, 7, 11 y 12 el pixel es de 20m. Finalmente las bandas 1, 9, 10 son de 60m
Resolución espectral	ETM+ 8 canales, tres en el espectro visible B1, B2 y B3, una en el inflarrojo cercano B4, dos en el IR medio B5 y B7, una en el IR térmico B6 y una banda panorámica B8	13 bandas espectrales
Resolución radiométrica	8 en cualquiera de sus 8 bandas	8 a 16 bits, aunque en MSI es de 12
Resolución temporal	16 días	5 días conjuntamente y 10 días individualmente

Tabla 2. Comparación resolución Vs satélite

Límites de la longitud de onda en Landsat 7:

Banda 1	Banda 2	Banda 3	Banda 4	Banda 5	Banda 6	Banda 7	Banda 8
0,45	0,53	0,63	0,78	1,55	10,4	2,09	0,52
0,52	0,61	0,69	0,9	1,75	12,5	2,35	0,9

Tabla 3. Límites de longitud Landsat 7.

Intervalos longitud de onda en Sentinel 2:

Banda 2	Banda 3	Banda 4	Banda 8	Banda 5	Banda 6	Banda 7	Banda 8	Banda 11	Banda 12
490 nm	560 nm	665 nm	842 nm	705 nm	740 nm	783 nm	865 nm	1610 nm	2190 nm

Banda 1	Banda 9	Banda 10
443 nm	940 nm	1375 nm

Tabla 4. intervalos de longitud de onda de Sentinel 2.

Comparación Landsat 7 y Landsat 8:

Landsat 7 ETM+			Landsat 8 OLI-TIRS		
Banda	Ancho (µm)	Resolución	Banda	Ancho (µm)	Resolución
			Banda 1 Coastal	0,435-0,451	30 m
Banda 1 Blue	0,441-0,514	30 m	Banda 2 Blue	0,452-0,512	30 m
Banda 2 Green	0,519-0,601	30 m	Banda 3 Green	0,533-0,590	30 m
Banda 3 Red	0,631-0,692	30 m	Banda 4 Red	0,636-0,673	30 m
Banda 4 NIR	0,772-0,898	30 m	Banda 5 NIR	0,851-0,879	30 m
Banda 5 SWIR 1	1,547-1,749	30 m	Banda 6 SWIR 1	1,566-1,651	30 m
Banda 5 SWIR 1	1,547-1,749	30 m	Banda 7 SWIR 2	2,107-2,294	30 m
Banda 8 Pan	0,515-0,896	15 m	Banda 8 Pan	0,503-0,676	15 m
			Banda 9 Cirus	1,363-1,384	30 m
Banda 6 TIR	10,31-12,36	60 m	Banda 10 TIRS 1	10,60-11,19	100 m
			Banda 11 TIRS 2	11,50-12,51	100 m

Tabla 5. Comparación banda, ancho y resolución Landsat 7 ETM+ Vs 8 OLI-TIRS

Para la clasificación del terreno, hay que determinar las unidades del mismo mediante la interpretación de las imágenes, resaltando factores como la geomorfología, material de la superficie, vegetación y red de drenaje. Aunque a medida que han pasado los años se han incluídos nuevos elementos, especialmente para la clasificación y estructura del paisaje a trabajar. Dando un cierto orden a la descripción en aspectos de litósfera (relieve, materiales entre otros), geomorfología (agua, clima) y fisiografía (descripción y clasificación de las geoformas), con el fin de hacer un inventario estructurado de los paisajes, para el reconocimiento de la zona, debido análisis y su aplicación.

4.10 TIPOS DE SENSORES

4.10.1 Sensores Activos

4.10.2 Sensores de Radar: Se denominan sensores de radar activos a los que por medio de la radiación electromagnética han de constituir la región de microonda del espectro electromagnético mediante longitudes de onda de 1 mm y 30 cm. Una de sus ventajas es que se pueden usar bajo condiciones atmosféricas, aunque a muy baja resolución espacial. Éstos a su vez pueden dividirse en 6 variantes (monofrecuencia, multipolarizada, polarizada, circular, multifrecuencia, pancromática y policromática)[8].

4.10.3 Lidar: Sistema activo el cual solamente ha de operar en un ambiente libre de nubes, que a su vez funciona entre el ultravioleta e infrarrojo, emitiendo radiación en forma contínua que se recoge mediante un sistema óptico a una resolución mayor que los sensores radar de microondas. Existen diferentes tipos de acuerdo a la dispersión (de Mie, Rayleight y Raman, de resonancia de absorción diferencial y fluorescencia inducida) [9]. Cada uno con aplicaciones específicas, unos ejemplos de ellos son las detección de aerosoles, partículas contaminantes en el aire, medición de humedad, temperatura, presión y velocidad del viento.

4.10.4 Sensores Pasivos

4.10.5 Sensores Fotográficos: Enfocados en la impresión de la imagen, en una especie de película fotosensible la cual depende de la longitud del lente o sistemas de lentes,

junto a su altura sobre la superficie terrestre. De ahí se pueden clasificar según el color e infrarrojo, números de objetivos y ángulos de observación y es cuando se encuentra su falla, debido a que solamente se pueden usar de día.

4.10.6 Exploradores de Vidicon: Sistema que trabaja de manera pancromática, lo cual significa que enfoca la imagen y la réplica de manera electrónica, manteniendo la copia hasta que se vuelven a escanear la superficie y se restaura el equilibrio de partida [8]. Teniendo mayor sensibilidad hacia los niveles bajos de luz y contando con mayor resolución, operando de manera simultánea en cada banda, proporcionando datos de manera real, especialmente para los satélites aplicados a la meteorología.

4.10.7 Radiómetros de Microondas: Algunas de sus características es que es de baja resolución espacial, el rango de emisividad superficial y la independencia de las condiciones atmosféricas, también constando de un elemento direccional y un receptor el cual amplifica y detecta la señal. Utilizados en cartografía de hielo, hidrología y oceanografía.

4.10.8 Exploradores de Barrido y de Empuje: Están formados por un espejo móvil el cual va oscilando de manera perpendicular de la trayectoria del satélite y a su vez va escaneando y trazando una línea a ambos lados del satélite, de ésta manera la radiación que se recibe se convierte en valor numérico a intervalos regulares para luego ser transmitidos y procesados en la tierra.

Por otra parte, los exploradores de empuje eliminan el espejo y a su vez incorporan una matriz lineal, correspondiendo cada línea a un detector, formando la imagen a medida que el satélite se va

moviendo y mejorando la relación señal-ruido.

4.11 MEDIDAS DE ERROR

Cuando se habla de medidas de error, la mejor forma de expresarlo es mediante la precisión de clasificación, para llegar a una matriz de confusión o contingencia, donde se muestra la tabulación de la clasificación de la tierra del sitio de la muestra, junto a diferentes medidas y porcentajes, basándose medidas estadíticas, análisis multiderivados y distribuciones binomiales. El diseño que se recomienda se indica en la tabla 6.

		DATOS DE REFERENCIA					
		B	I	U	A	∑	PRECISIÓN DE USUARIO
DATOS CLASIFICADOS	BOSQUE (B)	68	7	3	0	78	87.2%
	INDUSTRIAL (I)	12	112	15	10	149	75.2%
	URBANO (U)	3	9	89	0	101	88.1%
	AGUA (A)	0	2	5	56	63	88.9%
	∑	83	130	112	66	391	
	PRECISIÓN DE PRODUCTOR	81.9%	86.2%	79.5%	84.8%		
				Exactitud general		84%	

Tabla 6. Ejemplo medidas de error.[11]

4.12 MEDIDAS DE ESTADÍSTICAS

4.12.1 Matriz de Confusión: A través de ella se enumeran los valores para los diferentes tipos de coberturas que se conocen o trabajan, en las columnas se encuentran los datos de referencia, en las filas los datos clasificados y en la diagonal principal se enumeran los píxeles correctamente clasificados. Una de las medidas que son sacadas de allí es la precisión general

(calculada dividiendo los píxeles clasificados (que son la suma de los valores en la diagonal principal) entre el total de píxeles marcados) [10]. Por otro lado, la precisión de clasificación de clases individuales es calculada de dos maneras:

4.12.2 Precisión de Producto: Quien es obtenida dividiendo el número de píxeles correctos entre el número total de píxeles en los datos de referencia, para medir en un área determinada qué tan bien quedó su clasificación [11]. A su vez ha de incluir el error de omisión, que es la proporción de las características en el suelo y que no se clasifican, por ende, entre más errores existan, menos es la precisión del producto:

$$precisión\ del\ producto(\%) = 100\% - error\ de\ omisión(\%)$$

4.12.3 Precisión del Usuario: Resulta de la división de los píxeles clasificados correctamente por el número total de píxeles que se clasificaron. Informando al usuario si lo que aparece en el mapa representa verdaderamente lo que se encuentra en el suelo, hay dos tipos la clase correcta (es la misma clase de cobertura en el mapa y suelo) y la incorrecta, también vista como error de comisión (diferente cobertura del suelo y mapa) y entre más errores existan, disminuye la precisión del usuario.

$$precisión\ del\ usuario(\%) = 100\% - error\ de\ omisión(\%)$$

En la ilustración 3 se puede reconocer mejor la diferencia entre las medidas mencionadas anteriormente, en donde se observan cuatro clases diferentes de clases, coníferas (árboles que crecen en forma de cono), mixtas, arbustos y no forestales.

4.13 NORMALIZACIÓN MATRICIAL

Usado para comparar matrices y sus valores con diferentes enfoques de interpretación o clasificación. Mediante la fuerza de la suma de filas y columnas a una, quienes se encuentran equilibrados iterativamente, usado cuando no hay tantos valores de celda a cero ya que el algoritmo se encarga de cambiar dichos valores. La precisión se ha de calcular de forma análoga a la precisión general.

4.14 ANÁLISIS MULTIDERIVADO

4.14.1 Coeficiente Kappa: Medida general de una matriz, siendo la relación entre la suma de los valores diagonales

junto al número total de recuentos de células en la matriz, quien también tiene en cuenta elementos no diagonales. Evalúa si la clasificación actual es mejor que la basada al azar, a su vez, mide la proporción del contrato luego de eliminar los acuerdos casuales, cuando hay oportunidad aumenta en uno (acuerdo perfecto), al disminuir (acuerdo de menos oportunidad) es porque hay una oportunidad menor y al haber un cero (acuerdo de oportunidad) se entiende como la igualdad entre datos de verificación y clasificados. Volviéndose una medida para la precisión a la hora de la clasificación. Al mismo tiempo, el coeficiente Kappa ha de formar parte de la familia de coeficientes de acuerdo bivariado (en donde el propósito es averiguar si dos variables tienen relación entre sí), que al mismo tiempo agregan coeficientes como el índice fundamental de Turk,

índice de precisión media de Hellden y/o índice de precisión de mapeo de Short, utilizados para describir la precisión de la categoría [11].

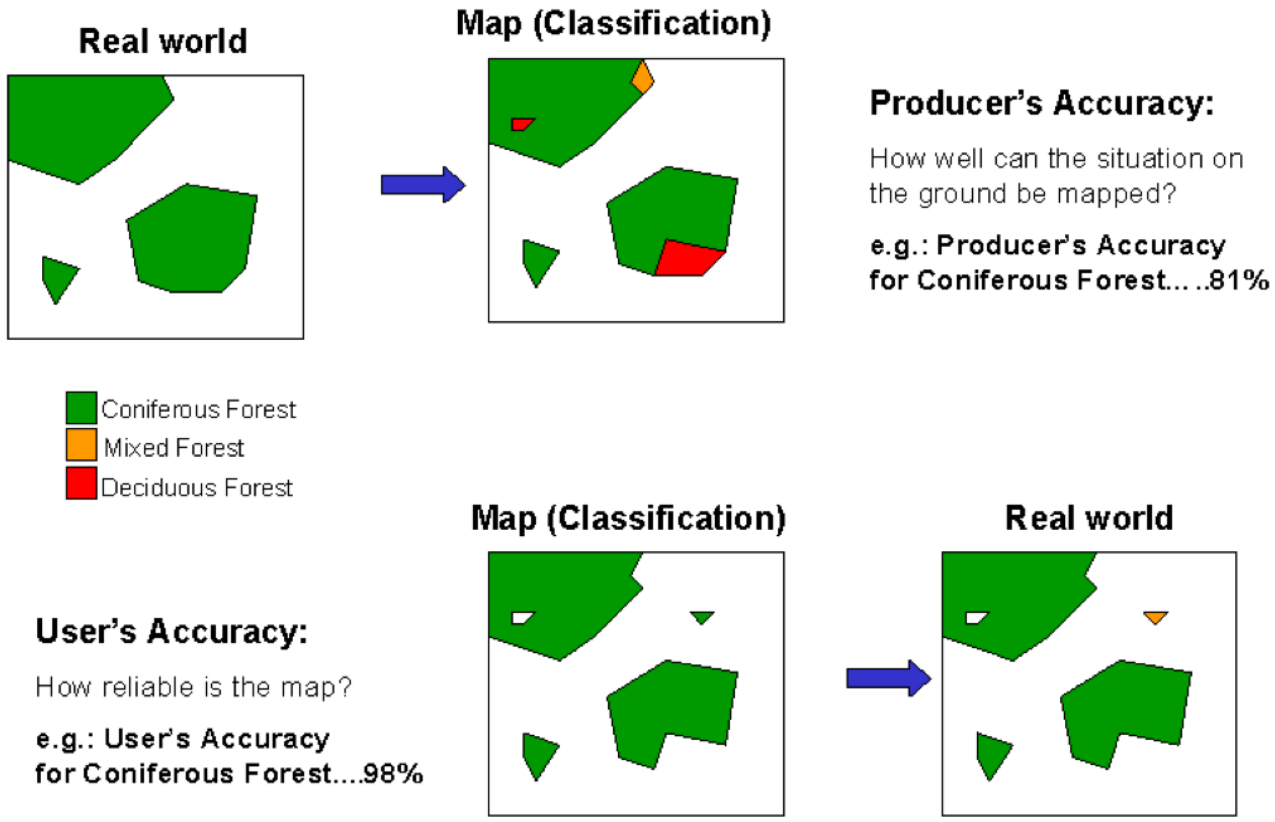

Ilustración 3. Ejemplo precisión de usuario y precisión de producto.[11]

El cálculo para el coeficiente Kappa, según Cohen:

$$K = \frac{P_O - P_e}{1 - P_e}$$

En donde:

$$P_o = Precisión\ del\ acuerdo\ observado$$

$$P_e = Estimado\ del\ acuerdo\ de\ oportunidad$$

Viéndose como un muestreo multinomial (un muestreo aleatorio con varios resultados posibles). Para observar la probabilidad en la coincidencia de probabilidad $\left(\frac{1}{ny}\right)$:

$$k = P_o - \frac{1}{n} \; que\ se\ divide\ en\ 1 - \frac{1}{n}$$

Para describir la prueba de significación entre dos Kappa independientes:

$$Z = \frac{\widehat{K_1} - \widehat{K_2}}{\sqrt{\widehat{\sigma_1} + \widehat{\sigma_2}}}$$

Es utilizada para probar si la precisión de clasificación difiere del convenio casual ó existe una diferencia entre varios enfoques de clasificación. Al momento de demostrar el uso del coeficiente Kappa, se requieren datos de matrices de error (recuentos de píxeles), para dos enfoques de clasificación que se enumeran en la tabla 7.

ENFOQUE SUPERVISADO					
		DATOS DE REFERENCIA			
		B	I	A	U
DATOS CLASIFICADOS	BOSQUE (B)	68	7	3	0
	INDUSTRIAL (I)	12	112	15	10
	URBANO (U)	3	9	89	0
	AGUA (A)	0	2	5	56

ENFOQUE NO SUPERVISADO					
		DATOS DE REFERENCIA			
		B	I	A	U
DATOS CLASIFICADOS	BOSQUE (B)	68	7	3	0
	INDUSTRIAL (I)	12	112	15	10
	URBANO (U)	3	9	89	0
	AGUA (A)	0	2	5	56

Tabla 7. Enofques de clasificación.

Para que a partir de ahí se pueda proporcionar una comparación entre la precisión normalizada, precisión general y coeficiente Kappa. Quienes son las medidas que deben coincidir en la

clasificación relativa de los enfoques de clasificación (supervisado y no supervisado), pero sus valores absolutos son diferentes por la razón que cada medida ha de incorporar distintos niveles de información como se observa en la tabla 8.

CLASIFICACIÓN	COEFICIENTE KAPPA	ESTADÍSTICA Z	RESULTADO a
CLASIFICACIÓN SUPERVISADA	0.7687%	29.41%	S^b
CLASIFICACIÓN NO SUPERVISADA	0.6956%	24.04%	S^b
SUPERVISADO vs NO SUPERVISADA		2%	NS^b

Tabla 8. Comparación entre la precisión normalizada, precisón general y coeficiente Kappa.[11]

De ésta manera, finalmente obtener los resultados del análisis de Kappa de significancia para matrices de error individual y comparación entre matrices, probando la precisión entre dos maneras (matriz clasificada y clasificación aleatoria) para obtener los enfoques de clasificación (Tabla 9).

CLASIFICACIÓN	COEFICIENTE KAPPA	ESTADÍSTICA Z	RESULTADO a
CLASIFICACIÓN SUPERVISADA	0.7687%	29.41%	S^b
CLASIFICACIÓN NO SUPERVISADA	0.6956%	24.04%	S^b
SUPERVISADO vs NO SUPERVISADA		2%	NS^b

Tabla 9 Precisión entre matriz clasificada y matriz aleatoria.[11]

Donde

$S^b = Significante$

$NS = No\ significante$

$a = 95\%\ de\ nivel\ de\ confidencia$

Se observa que la precisión de clasificación es del 77%, dándose a entender que la precisión de clasificación es mejor que la precisión que se daría de una posible asignación aleatoria. En diversos casos, la clasificación de las diferentes medidas de error, han de tener dependencia en la forma de

la matriz de errores, por las celdas diagonales con valores de cero son las que fuerzan que la precisión normalizada difiera.

5. MARCO EXPERIMENTAL

5.1 Zona de Estudio

La zona de estudio seleccionada se encuentra en el departamento de Cundinamarca, Colombia; abarcando principalmente la zona urbana de Zipaquirá y el embalse de Neusa. Utilizando un conjunto de imágenes de diciembre 27 del 2018 – 01 enero del 2019 mediante el satélite Landsat 8 como se aprecia en la ilustración 4; imagen en donde la nubosidad es mínima.

Ilustración 4. Zona de estudio Cundinamarca, Colombia.

6. METOLOGÍA

Se utiliza la herramienta de Google Earth Engine Explorer para la clasificación de la ilustración 4. Se selecciona el satélite Landsat 8, el cual contiene 11 bandas cuyas diferentes combinaciones permite resaltar multiples caracteríaticas del mapa y así poder discriminar fácilmente ciertos elementos que no son necesarios para el análisis, para más etalle ver tabla 10.

Luego de seleccionar la zona de estudio se procede a aplicar unas bandas de Landsat 8 (Agricultura. Combinación: B6, B5, B2) para mejorar la visualización de la imagen, teniendo como resultado la ilustración 5. La combinación de bandas escogida permite una mejor visión de la vegetación saludable siendo ésta de un color verde; como verde oscuro se observa una vegetación vieja y magenta las zonas de suelo desnudo (ilustración 6).

Banda en Landsat 8	Aplicación principal/Descripción
Banda 1 Coastal	Para mapeo de costas y estudios de aerosol
Banda 2 Blue	Útil para mapeo barimétrico, delimitar costas, diferenciar suelo de vegetación, diferenciar coníferas de latifoliadas, detección de rangos urbanos, vías y construcciones
Banda 3 Green	Empleada para discriminar sedimientos en suspensión, evaluar vigor de las plantas por la alta reflectancia de la vegetación verde y sana, delinear aguas poco profundas, rasgos urbanos y de infraestructuras
Banda 4 Red	Permite un mayor contraste de áreas con y sin vegetación, discriminar gradientes de vegetación, delimitar áreas urbanas y áreas agrícolas
Banda 5 NIR	Útil para el cálculo de biomasa de vegetación, delimitar costas, para diferenciar suelos-cultivos y suelos-agua, para geomorfología, suelos y geología
Banda 6 SWIR 1	Se alcanzan a penetrar nubes delgadas. Es útil para discriminar contenido de humedad en los suelos y la vegetación, diferenciar entre nubes, nieve y hielo
Banda 7 SWIR 2	Útil para discriminar tipos de rocas, para estudios de suelos y mejora la determinación de contenidos de humedad en suelos y vegetación
Banda 9 Pancromática	Es una imagen sensible a todo el espectro visible y más afinada en tanto su resolución es de 15 m
Banda 8 Cirrus	Ofrece una mejora en la detección de nubes cirrus
Banda 10 TIRS 1	Útil para el mapeo termal y estimación de humedad del suelo
Banda 11 TIRS 2 Térmica	Mapeo termal mejorado y estimación de humedad del suelo

Tabla 10. Descripción bandas Landsat 8.[12]

Ilustración 5. Zona de estudio con combinacion RGB 562, Agricultura.

Se procede a entrenar el clasificador indicándole puntos de referencia, asignando los colores en el mapa que representan vegetación joven (verde), bosque (verde oscuro), agua (azul), suelo desnudo (amarillo), urbanización (rojo) y cultivo cubierto (morado). En la tabla 9 se observa el numero de puntos y polígonos utilizados como muestra para entrenar al clasificador y en la ilustración 7 se aprecia la ubicación de éstos.

Clase	Punto	Polígono
Agua	28	12
Bosque	103	11
Vegetación	91	11
Suelo desnudo	107	18
Urbanización	32	3
Cultivos cubiertos	30	2

Tabla 11. Número de puntos y polígonos tomados.

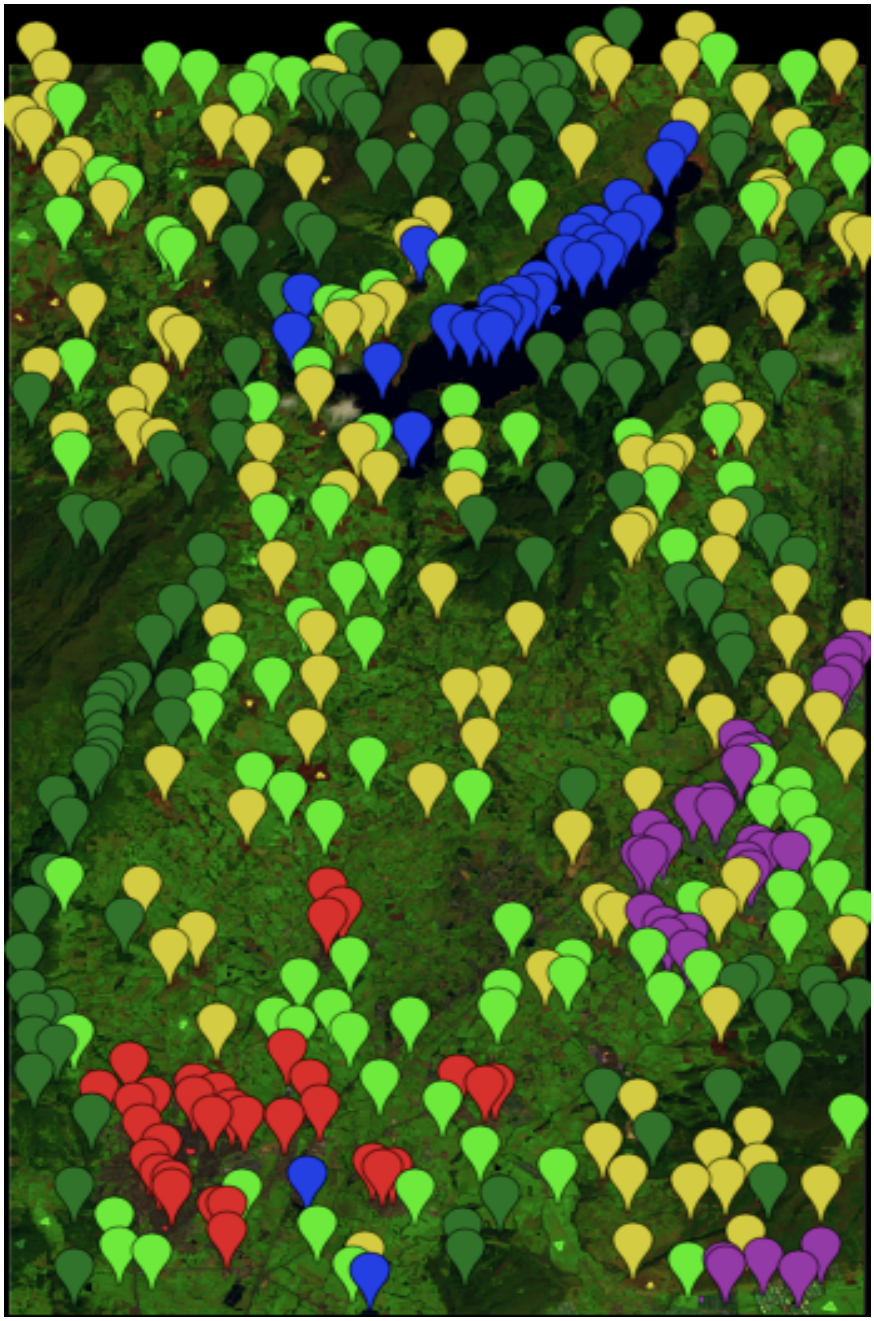

Ilustración 6. Puntos y polígonos para entrenar al clasificador. 391 puntos.

Color	Categoría
	Agua
	Bosque
	Urbano
	Vegetación
	Suelo desnudo
	Cultivo cubierto

Una vez se eligen los que serán los puntos de referencia, se procede a entrenar el clasificador y a elegir el método de clasificación. En este caso se van a comparar los métodos de clasificación Random Forest y CART, ambos con los mismos puntos de referencia.

7. RESULTADOS

7.1 Clasificacion 1

Random Forest

Validez general	96,67%						
	Puntos	Vegetación	Agua	Bosque	Urbano	Suelo desnudo	Cultivo cubierto
Vegetación	156	100%	0%	0%	0%	0%	0%
Agua	60	0%	100%	0%	0%	0%	0%
Bosque	158	0%	0%	100%	0%	0%	0%
Urbano	43	0%	0%	0%	97,67%	2,33%	0%
Suelo desnudo	166	0%	0%	0%	0,60%	99,40%	0%
Cultivo cubierto	31	0%	0%	0%	0%	0%	100%

Tabla 12. Matriz Random Forest. 391 puntos.

RANDOM FOREST	V	A	B	U	SD	CC	
Vegetación (V)	100	0	0	0	0	0	100
Agua (A)	0	100	0	0	0	0	100
Bosque (B)	0	0	100	0	0	0	100
Urbano (U)	0	0	0	97,67	2,33	0	100
Suelo Desnudo (SD)	0	0	0	0,6	99,4	0	100
Cultivo Cubierto (CC)	0	0	0	0	0	100	100
	100	100	100	98,27	101,73	100	**391**

PRECISIÓN GENERAL	CROKER Y ALGINA	ACUERDO ENTRE EVALUADORES	KAPPA
1,534526854	2,069053708	0,392462111	1,879824722

Tabla 13. Matriz de confusión Random Forest. 391 puntos.

En la ilustración 7 se observa unos errores alrededor del embalse de Neusa. La causa de éstos errores son una nubosidad que se concentra al sur del embalse o en en el sector montañoso la sombra proveniente del relieve de la montaña genera. Los errores son categorizados como zona urbana (rojo) o cultivos cubiertos (morado), en el sector de la laguna y en la zona montañosa los errores son categorizados como agua (azul).

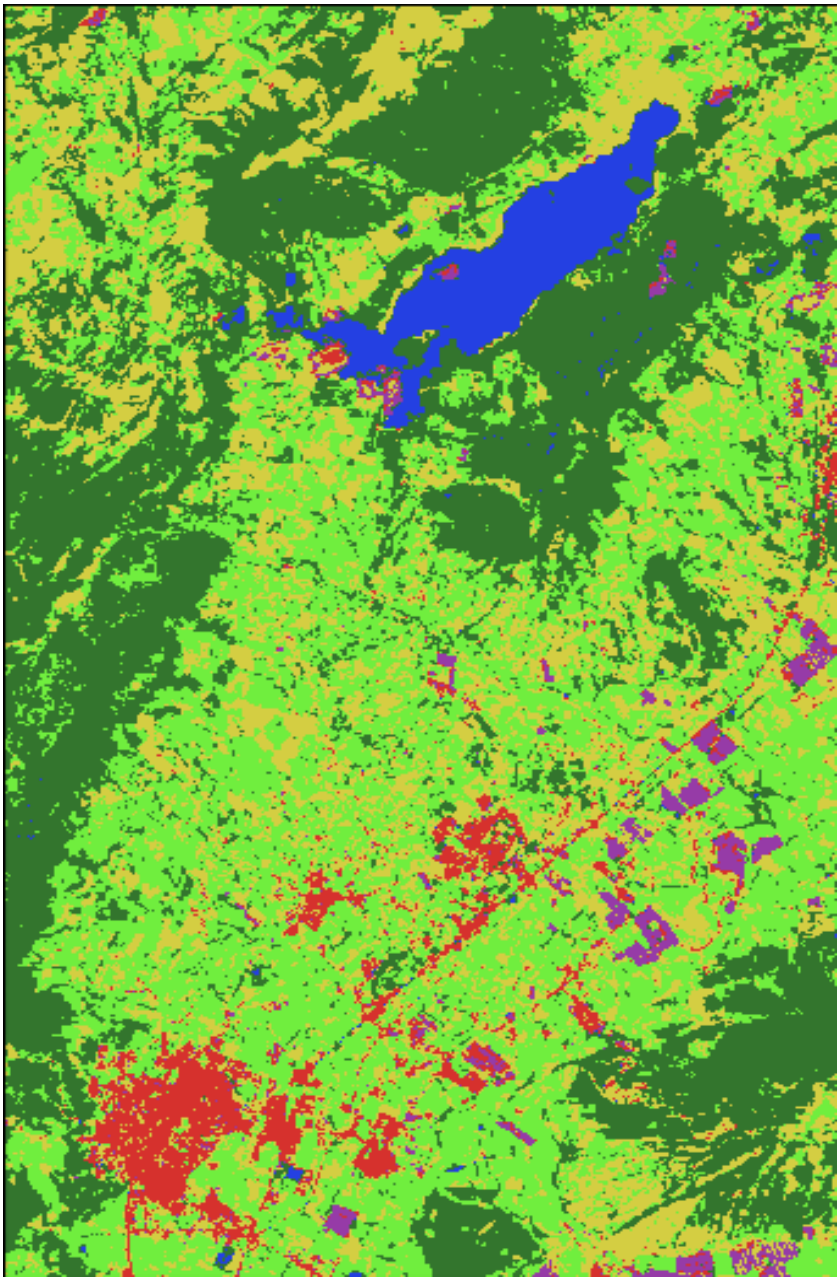

Ilustración 7. Clasificación Random Forest. 391 puntos.

Color	Categoría
	Agua
	Bosque
	Urbano
	Vegetación
	Suelo desnudo
	Cultivo cubierto

CART

Validez general	100%						
	Puntos	Vegetación	Agua	Bosque	Urbano	Suelo desnudo	Cultivo cubierto
Vegetación	156	100%	0%	0%	0%	0%	0%
Agua	60	0%	100%	0%	0%	0%	0%
Bosque	158	0%	0%	100%	0%	0%	0%
Urbano	43	0%	0%	0%	100%	0%	0%
Suelo desnudo	166	0%	0%	0%	0%	100%	0%
Cultivo cubierto	31	0%	0%	0%	0%	0%	100%

Tabla 14. Matriz CART. 391 punots.

CART	**V**	**A**	**B**	**U**	**SD**	**CC**	
Vegetación (V)	100	0	0	0	0	0	100
Agua (A)	0	100	0	0	0	0	100
Bosque (B)	0	0	100	0	0	0	100
Urbano (U)	0	0	0	100	0	0	100
Suelo Desnudo (SD)	0	0	0	0	100	0	100
Cultivo Cubierto (CC)	0	0	0	0	0	100	100
	100	100	100	100	100	100	**391**

PRECISIÓN GENERAL	CROKER Y ALGINA	ACUERDO ENTRE EVALUADORES	KAPPA
1,534526854	2,069053708	0,392462111	1,879824722

Tabla 15. Matriz de confusión CART. 391 puntos

No presenta coherencia entre la imagen y la matriz generada por la herramienta. La zona de nubosidad es categorizada como urbano (rojo) y cultivo cubierto (morado), siendo esta menor que en los otros clasificadores. En la parte superior de la imagen se observa errores entre la montaña y el suelo desnudo categorizados como cultivos cubiertos (morado) aunque éstos realmente son suelo desnudo (amarillo).

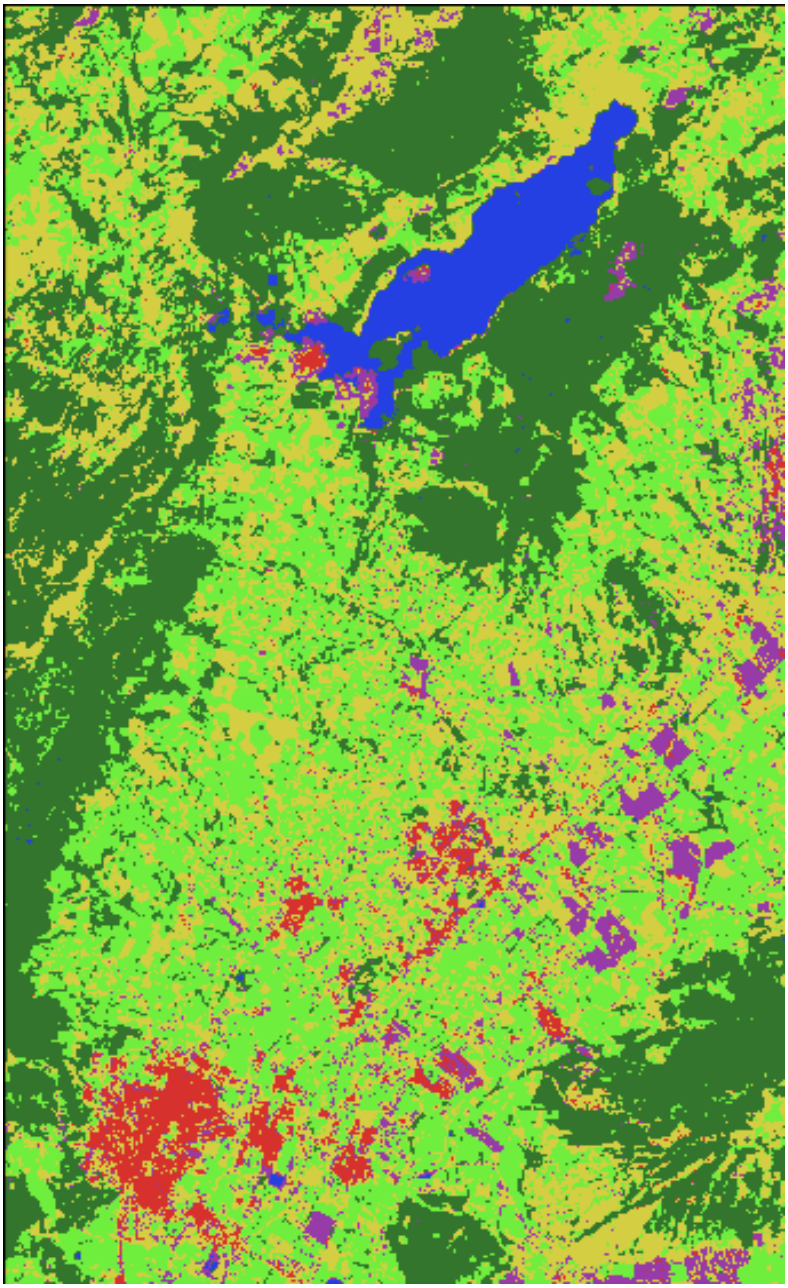

Ilustración 8. Clasificación CART.391 puntos.

Color	Categoría
	Agua
	Bosque
	Urbano
	Vegetación
	Suelo desnudo
	Cultivo cubierto

GMO max entropy

Validez general	97,56%						
	Puntos	Vegetación	Agua	Bosque	Urbano	Suelo desnudo	Cultivo cubierto
Vegetación	156	99,36%	0%	0,64%	0%	0%	0%
Agua	60	0%	100%	0%	0%	0%	0%
Bosque	158	0%	1,9%	98,01%	0%	0%	0%
Urbano	43	0%	0%	0%	93,02%	4,65%	2,33%
Suelo desnudo	166	0%	0%	0%	1,2%	98,8%	0%
Cultivo cubierto	31	0%	0%	19,35%	0%	0%	80,65%

Tabla 16. Matriz GMO max entropy. 391 puntos.

GMO MAX ENTROPY	**V**	**A**	**B**	**U**	**SD**	**CC**	
Vegetación (V)	99,36	0	0,64	0	0	0	100
Agua (A)	0	100	0	0	0	0	100
Bosque (B)	0	1,9	98,1	0	0	0	100
Urbano (U)	0	0	0	93,02	4,65	2,33	100
Suelo Desnudo (SD)	1,75	0	0	1,2	98,8	0	101,75
Cultivo Cubierto (CC)	0	0	19,35	0	0	80,65	100
	101,11	101,9	118,09	94,22	103,45	82,98	**391**

PRECISIÓN GENERAL	CROKER Y ALGINA	ACUERDO ENTRE EVALUADORES	KAPPA
1,539002558	2,078005115	0,394790965	1,890605603

Tabla 17. Matriz de confusión GMO max entropy. 391 puntos.

Presenta coherencia entre la imagen y la matriz generada por la herramienta. La zona de nubosidad es categorizada como urbano (rojo) y cultivo cubierto (morado). En la parte superior de la imagen se observa errores entre la montaña y el suelo desnudo categorizados como cultivos cubiertos (morado) aunque éstos realmente son suelo desnudo (amarillo).

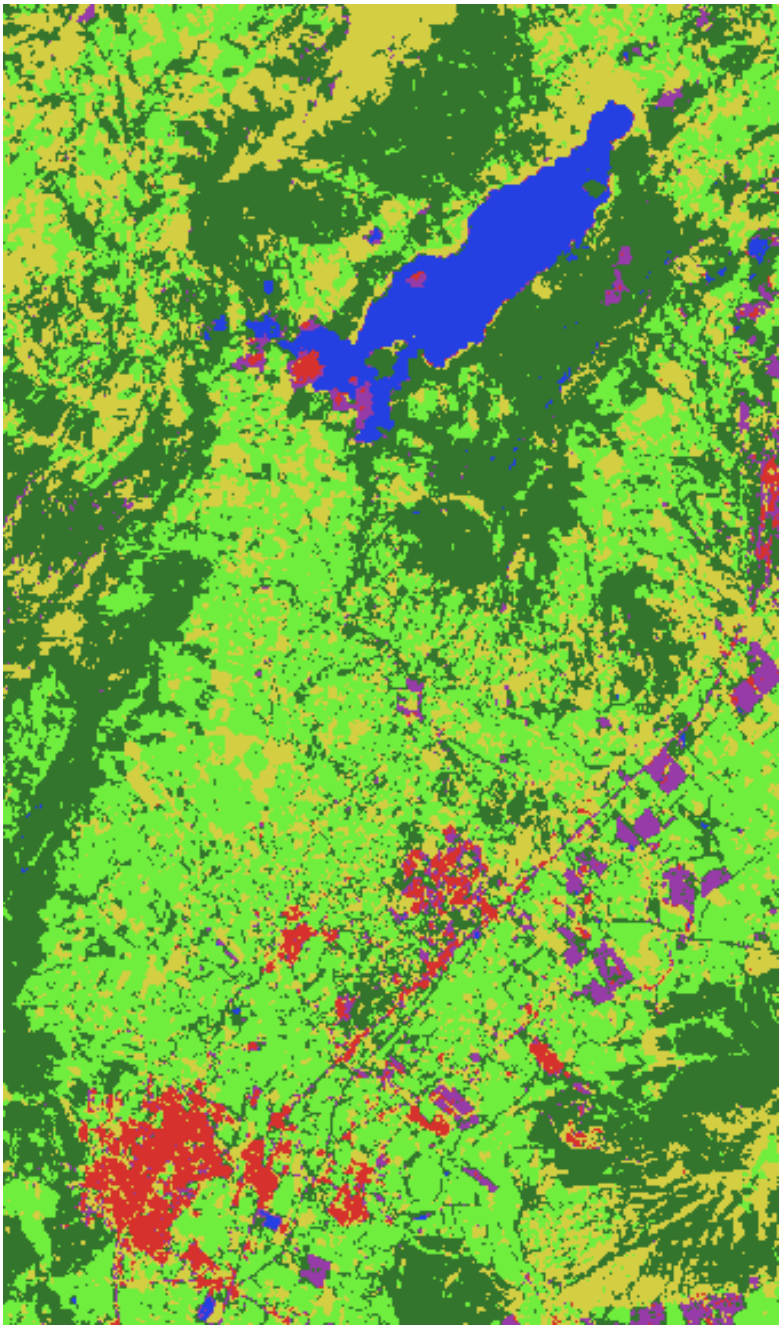

Ilustración 9. GMO max entropy. 391 puntos.

Color	Categoría
	Agua
	Bosque
	Urbano
	Vegetación
	Suelo desnudo
	Cultivo cubierto

7.2 Clasificación 2.

Con el fin de comparar la eficiencia del clasificador según el número de puntos de entrenamiento, se tomaron menos muestras que en la clasificación 1.

Clase	Punto
Agua	44
Bosque	7
Vegetación	44
Suelo desnudo	10
Urbanización	28
Cultivos cubiertos	9

Tabla 18. Numero de muestras por categoría para entrenar el clasificador

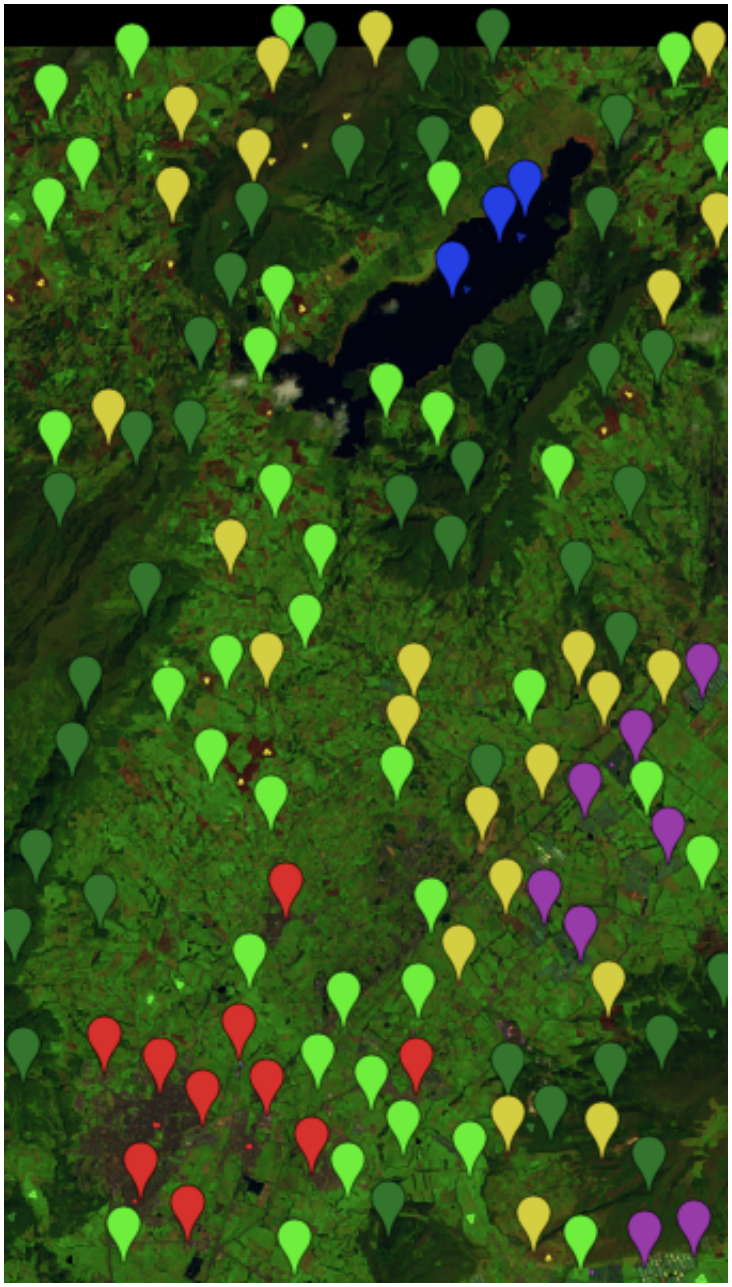

Ilustración 10. Puntos y polígonos de entrenamiento. 142 puntos

Color	Categoría
	Agua
	Bosque
	Urbano
	Vegetación
	Suelo desnudo
	Cultivo cubierto

Random Forest

Validez general	97,89%						
	Puntos	Vegetación	Agua	Bosque	Urbano	Suelo desnudo	Cultivo cubierto
Vegetación	44	100%	0%	0%	0%	0%	0%
Agua	7	0%	100%	0%	0%	0%	0%
Bosque	44	0%	0%	100%	0%	0%	0%
Urbano	10	0%	0%	0%	100%	0%	0%
Suelo desnudo	28	3,57%	0%	0%	7,14%	89,29%	0%
Cultivo cubierto	9	0%	0%	0%	0%	0%	100%

Tabla 19. Matriz Random Forest. 142 puntos

RANDOM FOREST	**V**	**A**	**B**	**U**	**SD**	**CC**	
Vegetación (V)	100	0	0	0	0	0	100
Agua (A)	0	100	0	0	0	0	100
Bosque (B)	0	0	100	0	0	0	100
Urbano (U)	0	0	0	100	0	0	100
Suelo Desnudo (SD)	3,57	0	0	7,14	89,28	0	99,99
Cultivo Cubierto (CC)	0	0	0	0	0	55,56	55,56
	103,57	100	100	107,14	89,28	55,56	**142**

PRECISIÓN GENERAL	CROKER Y ALGINA	ACUERDO ENTRE EVALUADORES	KAPPA
3,912323944	6,824647887	2,632663202	-0,8

Tabla 20. Matriz confusión Random Forest. 142 puntos

La clasifciación con 142 puntos presenta una menor exactitud que la clasificación con 391 puntos. La nubosidad es clasificada como zona urbana (rojo). Menos densidad en la zona urbana. Menos errores en la parte superior izquierda de la imagen comparada con la clasificación de 391 puntos.

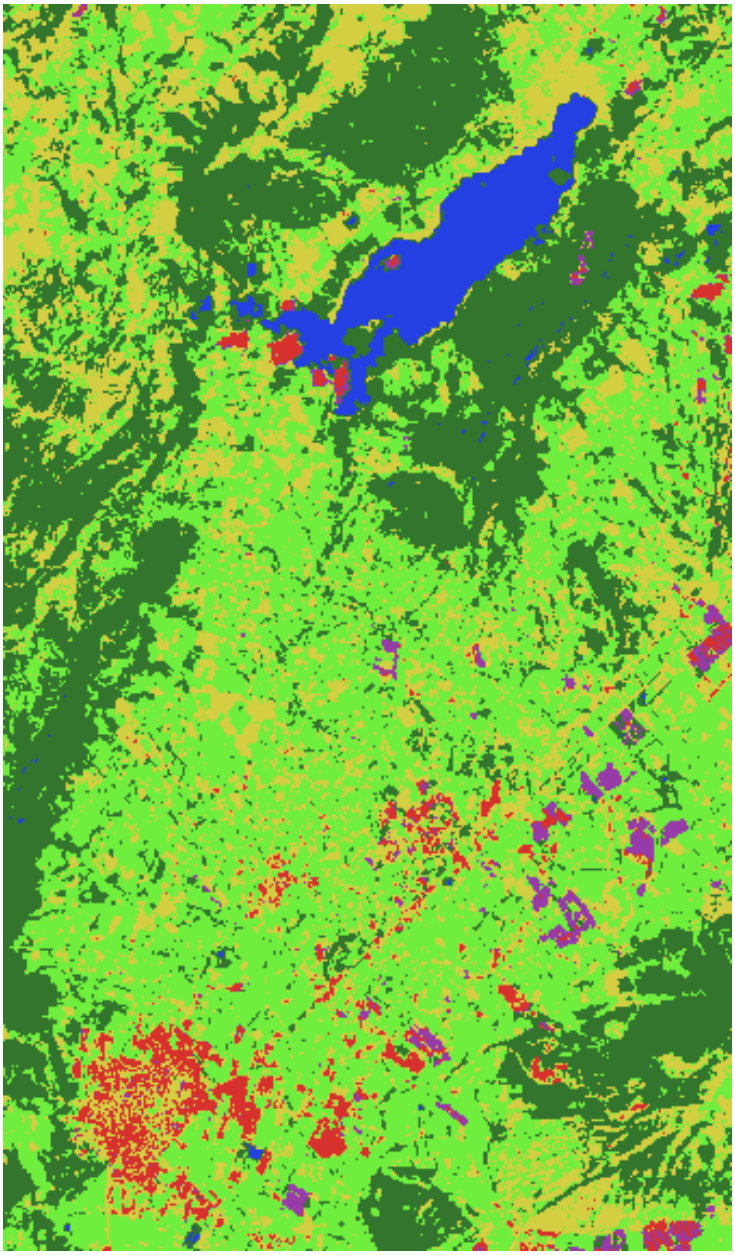

Ilustración 11. Random Forest. 142 puntos

Color	Categoría
	Agua
	Bosque
	Urbano
	Vegetación
	Suelo desnudo
	Cultivo cubierto

GMO max entropy

Validez general 94,37%

	Puntos	Vegetación	Agua	Bosque	Urbano	Suelo desnudo	Cultivo cubierto
Vegetación	44	97,73%	0%	0%	0%	2,27%	0%
Agua	7	0%	100	0%	0%	0%	0%
Bosque	44	0%	0%	97,73%	0%	2,27%	0%
Urbano	10	0%	0%	0%	100%	0%	0%
Suelo desnudo	28	3,57%	0%	0%	3,57%	92,86%	0%
Cultivo cubierto	9	0%	0%	44,44%	0%	0%	55,56%

Tabla 21. Matriz GMO max entropy. 142 puntos

GMO MAX ENTROPY	**V**	**A**	**B**	**U**	**SD**	**CC**	
Vegetación (V)	97,73	0	0	0	2,27	0	100
Agua (A)	0	100	0	0	0	0	100
Bosque (B)	0	0	97,73	0	2,27	0	100
Urbano (U)	0	0	0	100	0	0	100
Suelo Desnudo (SD)	3,57	0	0	3,57	92,86	0	100
Cultivo Cubierto (CC)	0	0	44,44	0	0	55,56	100
	101,3	100	142,17	103,57	97,4	55,56	**142**

PRECISIÓN GENERAL	CROKER Y ALGINA	ACUERDO ENTRE EVALUADORES	KAPPA
4,225352113	7,450704225	2,975600079	-0,6

Tabla 22. Matriz confusión GMO max entropy. 142 puntos

Disminuye la zona montañosa en el costado izquierdo inferior de la imagen. La nubosidad es categorizada como zona urbana (rojo) y cultivo cubierto (morado). Datos que se ven reflejados en la matriz (tabla 21).

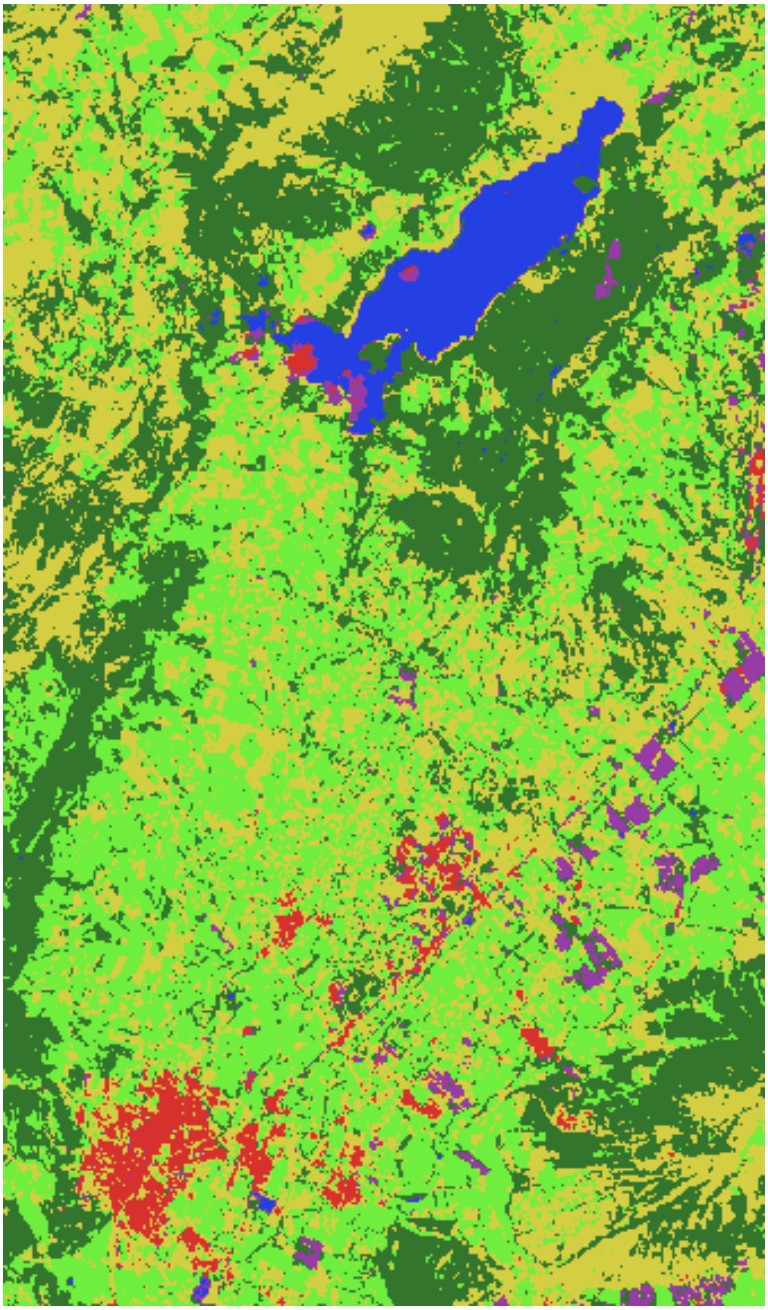

Ilustración 12. GMO max entropy. 142 puntos.

Color	Categoría
	Agua
	Bosque
	Urbano
	Vegetación
	Suelo desnudo
	Cultivo cubierto

CART

Validez general	100%						
	Puntos	Vegetación	Agua	Bosque	Urbano	Suelo desnudo	Cultivo cubierto
Vegetación	44	100%	0%	0%	0%	0%	0%
Agua	7	0%	100%	0%	0%	0%	0%
Bosque	44	0%	0%	100%	0%	0%	0%
Urbano	10	0%	0%	0%	100%	0%	0%
Suelo desnudo	28	0%	0%	0%	0%	100%	0%
Cultivo cubierto	9	0%	0%	0%	0%	0%	100%

Tabla 23. Matriz CART. 142 puntos

CART	**V**	**A**	**B**	**U**	**SD**	**CC**	
Vegetación (V)	100	0	0	0	2,82	0	102,82
Agua (A)	0	100	0	0	0	0	100
Bosque (B)	0	0	100	0	0	0	100
Urbano (U)	0	0	0	100	0	0	100
Suelo Desnudo (SD)	0	0	0	0	100	0	100
Cultivo Cubierto (CC)	0	0	0	0	0	100	100
	100	100	100	100	102,82	100	**142**

PRECISIÓN GENERAL	CROKER Y ALGINA	ACUERDO ENTRE EVALUADORES	KAPPA
4,245211268	7,490422535	3,00357072	-0,6

Tabla 24. Matriz confusión CART. 142 puntos

Al igual que en la clasificación de 391 puntos, CART de 142 puntos no tiene relación entre la matriz (tabla 23) y la clasificación (Ilustración 13), ya que los resultados no son ideales. La nubosidad es clasificada como zona urbana (roj) y cultivos cubiertos (morado).

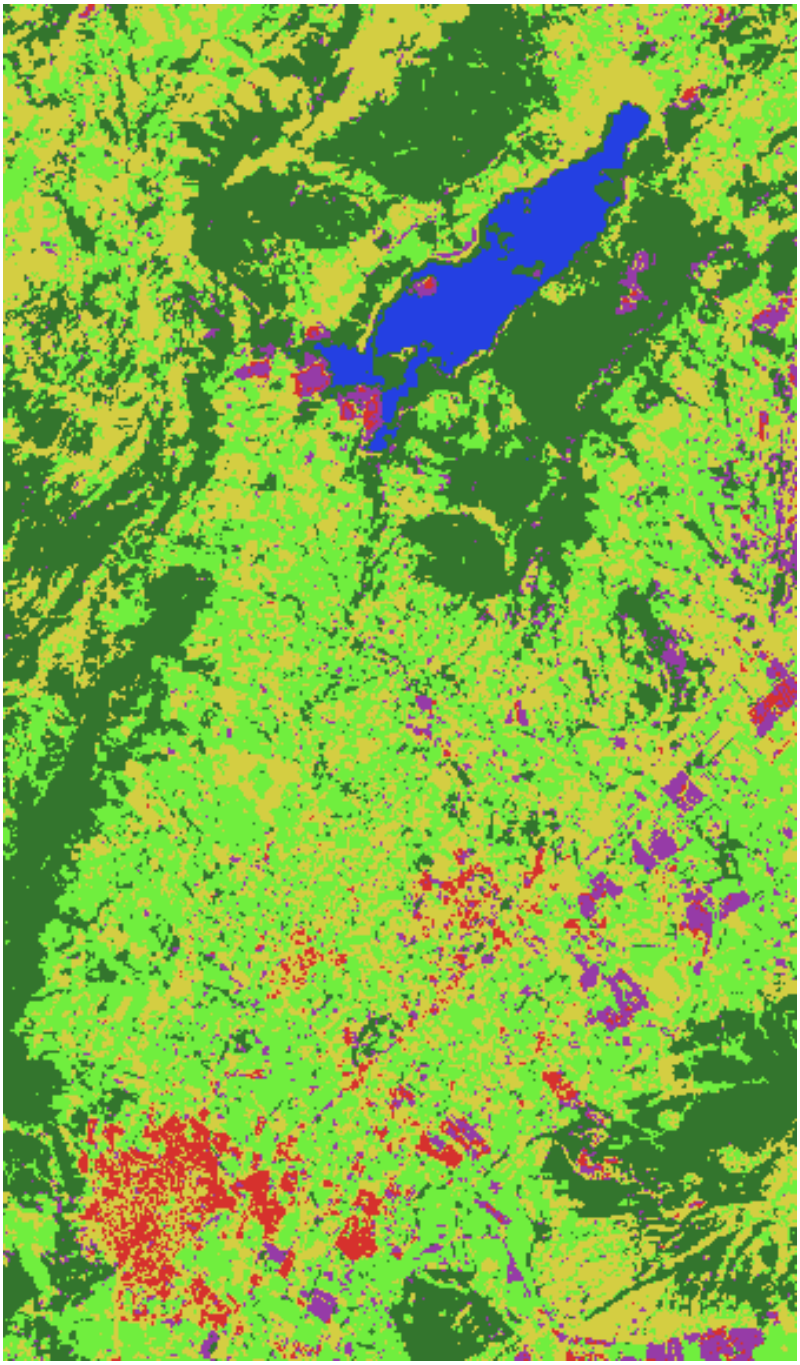

Ilustración 13. CART. 142 puntos

Color	Categoría
	Agua
	Bosque
	Urbano
	Vegetación
	Suelo desnudo
	Cultivo cubierto

8. CONCLUSIONES

- Según los resultados obtenidos se puede concluye que:
    - Los resultados obtenidos mediante el clasificador CART son incorrectos ya que los resultados son ideales cuando evidentemente no lo son.
    - Los tres clasificadores analizados en el estudio relacionan la nubosidad con zona urbana o cultivo cubierto.
    - Según los resultados de las imágenes y la tabla matricial el método de clasificación GMO max entropy obtiene los resultados mas coherentes.
    - Entre más puntos de entrenamiento tenga el clasificador, más precisa será la clasificación de éste.
    - Los métodos de clasificación están basados en pixeles.
- Google Earth Engine permite el acceso a información a través de la plataforma en la nube, dejando que el usuario acceda al procesamiento de la misma.
- El mapeo general, de bosques o cobertura del suelo, requiere de perfiles multitemporales (varias fechas) de crecimiento en los cultivos, en ocasiones se ve afectado por las nubes o sombras.
- Plataformas como Google Earth Engine contienen métodos de clasificación en donde se involucran grandes volúmenes de datos para áreas extensas.
- A la hora de visualizar las imágenes se muestra un mejor rendimiento en la precisión general.
- Por medio de los clasificadores es posible la discriminación de diferentes tipos categorías.

9. BIBLIOGRAFÍA

[1] E.E José Luis, P.V Enrique, T.C Leonardo, F.M Héctor, Q.N Abel, "Sistema de monitoreo satelital para el seguimiento y desarrollo de cultivos del Distrito de Riego 038", Tecnol. cienc. agua [online]. 2017, vol.8, n.1.

[2] Shelestov A, Lavreniuk M, Kussul N, Novikov A and Skakun S (2017) Exploring Google Earth Engine Platform for Big Data Processing: Classification of Multi-Temporal Satellite Imagery for Crop Mapping. Front. Earth Sci. 5:17. doi: 10.3389/feart.2017.00017.

[3] B Ramos, (enero 9 2018). ¿Qué es google earth engine? (Blog). Disponible: cursosgratis.com/que-es-google-earth-engine/

[4] Barbieri B, M.A, Bravo R, M, Farías S, M, González C, A, Pizarro A, O, & Yáñez R, E. (1995). Fenómenos asociados a la estructura térmica superficial del mar observados a través de imágenes satelitales en la zona norte de Chile. Investigaciones marinas, 23, 99-122. https://dx.doi.org/10.4067/S0717-71781995002300007

[5] Metodología Corine Land cover Melo y Camacho, 2005, en Mapa de Cobertura de la Tierra Cuenca Magdalena-Cauca, Metodología Corine Land Cover Adaptada para Colombia, escala 1:100.000, IDEAM, IGAC, CORMAGDALENA (2007)

[6] Informe final componente coberturas de la tierra en los páramos priorizados a escala 1:25.000 (Metodología corine land cover adaptada para Colombia) - Instituto Humbold - Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM).

[7] Implementación de la metodología corine land cover con imágenes IKONOS - Germán Mauricio Valencia, Jesús Adolfo Anaya.
G. Valencia, J. Anaya, "Implementación de la Tecnología Corine Land Cover con Imágenes Ikonos", Revista Ingeniería Universidad de Medellín, vol. 8, No.15, pp. 39-52 ISSN 1692-3324 - Julio-Diciembre de 2009/146 p. Medellín, Colombia.

[8] Leyenda nacional de Coberturas de la Tierra
Melo y Camacho, 2005, en Mapa de Cobertura de la Tierra Cuenca Magdalena-Cauca, Metodología Corine Land Cover Adaptada para Colombia, escala 1:100.000, IDEAM, IGAC, CORMAGDALENA (2007)

[9] 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, d. c., 72p.

[10] high resolution remote sensing data classification over urban areas using random forest esemble and fully connected conditional random field sun, xiaofeng & lin, xiangguo & shen, shuhan & hu, zhanyi. (2017). high-resolution remote sensing data classification over urban areas using random forest ensemble and fully connected conditional random field. international journal of geo-information. 6.

[11] Banko, Gebhard. (1998). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data and of Methods Including Remote Sensing Data in Forest Inventory.

[12] r. franco. (2017). "composiciones landsat en arcgis". disponible: https://mixdyr.files.wordpress.com/2017/06/composiciones_landsat_en-arcgis.pdf

[13] Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., & Skakun, S. (2017). Exploring Google earth engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping. frontiers in Earth Science, 5, 17.