

Revista Difusiones, ISSN 2314-1662, Núm. 25, 2(2) julio-diciembre 2023, pp.116-132
Fecha de recepción: 23-06-2023. Fecha de aceptación: 13-12-2023

La Caracterización Espacial de Coberturas del Suelo a partir de Información Satelital, Caso de Estudio: Zona Agroproductiva de Santa Fe, Argentina

Spatial Characterization of Land Cover from Satellite Information, Case Study: Agro-productive Zone of Santa Fe, Argentina

Cristina Alejandra Castillo¹ cristinadv3582@gmail.com
Universidad Nacional de Jujuy, Facultad de Ingeniería, Jujuy, Argentina

Brenda Noelia Veramendi² brendaveramendi@gmail.com
ORCID: <http://orcid.org/0009-0000-5495-4246>
Universidad Nacional de Jujuy, Facultad de Ingeniería, Jujuy, Argentina

Gisela Noelia Revollo Sarmiento³ grevollo@fi.unju.edu.ar
ORCID: <https://orcid.org/0000-0002-1532-5428>
Universidad Nacional de Jujuy, Facultad de Ingeniería, Instituto de Ecorregiones
Andinas CONICET, Jujuy, Argentina

¹ Estudiante avanzada de la carrera Ingeniería Informática, en proceso de desarrollo de tesis de grado. A lo largo de mi formación, he adquirido habilidades en inteligencia artificial, sensores remotos. Durante el periodo 01/09/2022 hasta 21/12/2022 realice mis Práctica Profesional Supervisada en el Instituto de Ecorregiones Andinas (INECOA) CONICET. Tuve participación en las Jornadas Argentinas de Informática del presente año (JAIIO). Actualmente soy becaria de las Becas EVC-CIN convocatoria 2022 cuyo tema es Predicción de Rendimiento de Cultivos usando Técnicas de Data Mining.

² Estudiante avanzada de la carrera Ingeniería en Informática, en desarrollo de la tesis. Becaria de la beca EVC-CIN, el tema de investigación es Caracterización de Cultivos Invernales con Sensores Remotos, Datos de Campo y Técnicas de Inteligencia Artificial: Aplicado a la zona Sud Oeste Bonaerense y zona de influencia del Puerto de Bahía Blanca. Desde el 01/09/2022 hasta 21/12/2022 realicé prácticas profesionales en el INECOA, con el tema de desarrollo de métodos de clasificación de cultivos y al análisis de sensores remotos, vinculados al sector agrícola. Entre 2016-2019 incursioné en el ámbito educativo mediante la creación de aplicaciones con Realidad Aumentada destinadas a la enseñanza primaria, con modelos 3D. Integrante del Grupo de Investigación GIDIS.

³ Doctora en Ingeniería por la Universidad Nacional del Sur, Departamento de Ingeniería Eléctrica y de Computadoras. Actualmente desarrolla sus actividades en el Instituto de Ecorregiones Andinas del Consejo Nacional de Investigaciones Científicas y Técnicas (INECOA - CONICET). Es Investigadora Asistente del CONICET y docente de la Facultad de Ingeniería en la Cátedra Introducción al Procesamiento Digital de Imágenes e Introducción a la Programación. En cuanto a desarrollos productivos es la responsable de la Plataforma de Gestión Integral de Información Agropecuaria (AgroInfinito) y varias aplicaciones móviles para el relevamiento y procesamiento de datos de campo. Ha participado en más de 18 proyectos de transferencia tecnológica e I+D. Tiene más de 12 publicaciones científicas en revistas indexadas, congresos y conferencias, también ha publicado capítulos de libros e informes técnicos. Inclusive, trabaja activamente en la formación de recursos humanos y demás actividades de docencia.

Resumen

El cambio de cobertura del suelo siempre ha tenido un papel central en la ciencia del cambio terrestre. Este enfoque es en gran medida el resultado de las posibilidades de mapear y caracterizar la cobertura del suelo basándose en observaciones y sensores remotos. Considerando la importancia de las funciones del uso del suelo que proporcionan una amplia gama de bienes y servicios, es esencial realizar evaluaciones más integradas del cambio de la tierra. Los métodos para mapear y cuantificar la dinámica de las coberturas de la tierra mejoran nuestra capacidad para comprender y modelar los cambios en el sistema de la tierra e informar adecuadamente las políticas y la planificación territorial. El objetivo del presente trabajo es desarrollar un modelo que caracteriza espacialmente las coberturas del suelo, implementando técnicas de Machine Learning e imágenes satelitales. La región de estudio es una zona agroproductiva del departamento General López, provincia de Buenos Aires, Argentina. Se implementaron dos modelos: Random Forest y Support Vector Machine para la clasificación de coberturas de suelo utilizando la plataforma de Google Earth Engine (GEE). Ambos modelos superan una precisión global de 90%, clasificando con gran fiabilidad las distintas clases de suelo. La implementación de GEE facilitó el procesamiento de los datos, reduciendo el tiempo de trabajo.

Palabras clave

Cobertura de suelo, Sensado Remoto, Inteligencia Artificial, Google Earth Engine

Abstract

Land cover change has always had a key role in land change science thanks to the possibilities for mapping and typifying land cover based on remote sensing and observations. When considering the importance of the functions of land use that provide a wide range of goods and services, it is essential to carry out more integrated land change assessments. Methods for mapping and quantifying land cover dynamics improve our ability to understand and model land system change and adequately inform policy and land planning. This work aims to develop a model that spatially characterises land covers, implementing Machine Learning techniques and satellite imagery. The region of study is an agro-productive area of the General López department, province of Buenos Aires, Argentina. We implemented two models, Random Forest and Support Vector Machine, for land cover classification using the Google Earth Engine (GEE) platform. Both models exceed an overall accuracy of 90% and classify, with high reliability, the different soil classes. The implementation of GEE facilitated data processing and reduced work time.

Key Words

Land Cover, Remote Sensing, Artificial Intelligence, Google Earth Engine.

Introducción

La distribución espacial del uso y cobertura de suelo es fundamental para la planificación, gestión del territorio y manejo de los recursos naturales, como así también para la prevención e intervención en materia de Riesgo de Desastres (Trujillo-Jiménez et al., 2021). Posee también una importancia central en la identificación, descripción y manejo de los recursos agrícolas a nivel nacional. La generación de un sistema de estimación de la productividad del sector agrícola en Argentina requiere conocer, con un margen de error bajo y conocido, su ubicación en el espacio. A pesar de la importancia de cuantificar los cambios en el uso del suelo, las caracterizaciones de la cobertura del suelo no suelen ser exhaustivas en nuestro país (Ferrelli et al., 2021).

El avance de la tecnología y las técnicas asociadas a la percepción remota, permiten desarrollar modelos que identifiquen cambios en la cobertura y uso del suelo a corto plazo y con niveles de precisión aceptables (da Silva et al., 2019). A nivel mundial existen numerosos antecedentes de clasificación del uso y cobertura del suelo en diversas escalas espacio temporales (Alhassan et al., 2019). Los sensores a bordo de satélites son una gran fuente de información que permiten mapear a nivel local o global la superficie terrestre (Veneros et al., 2020). Para ello, existen distintas misiones satelitales que proporcionan imágenes multi o hiper espectrales de alta, mediana y baja resolución. Por ejemplo, Landsat-8 proporciona datos con una resolución espacial de 30 metros, Sentinel-2 de 10 metros, Spot 6 hasta 1,5 metros. Estas imágenes registran la radiación reflejada por la superficie del planeta, guardando la información en distintas bandas espectrales. Las distintas misiones satelitales tienen características en común: un periodo de revisita, hora de la adquisición de imágenes, cobertura global, ángulo de visión, resolución espacial, ancho de barrido y siguen una política de datos libres y abiertos. En este contexto, el sensado remoto es la alternativa más viable para monitorear y clasificar las coberturas y usos del suelo en extensas áreas (Guerschman et al., 2003; Bernal y Baquero, 2021).

La sinergia de combinar el sensado remoto con el avance de la informática, permite potenciar el análisis espacial de datos a gran escala, bajo costo y en tiempo y forma (Nolasco et al., 2015). La inteligencia artificial (IA), en el contexto de las ciencias de la computación, es una disciplina que abarca una gran variedad de subcampos (aprendizaje y percepción, reconocimiento de voz, juegos, teoremas matemáticos, diagnóstico de enfermedades, etc.) (Curioso y Brunette, 2020; Ramírez Álvarez et al., 2020; Cardiel et al., 2022). La IA demanda recursos como el acceso a grandes volúmenes de datos para procesar y aprender y procesamiento en la nube utilizando tecnología GPU (graphics processing unit o procesador gráfico). Google Earth Engine (GEE) es una plataforma en la nube que reúne los recursos necesarios para implementar algoritmos de IA (Coca-Castro et al., 2021). Esta plataforma gratuita permite el análisis y visualización de datos geoespaciales a escala global, dispone de un gran catálogo de datos con petabytes de imágenes históricas de

distintas Misiones satelitales (Landsat, Modis, Sentinel), además de datos climáticos, meteorológicos y modelos digitales de elevación (DEMs), entre otros. GEE tiene una Interfaz de programación de aplicaciones (API) amigable en la que se pueden generar los algoritmos para el procesamiento de la información geoespacial.

El objetivo del presente trabajo es realizar un análisis espacial de la cobertura del suelo en una zona agroproductiva de la provincia de Santa Fe. Para ello, se utilizarán datos de campo, imágenes satelitales Sentinel-2, algoritmos de machine learning (ML) y el procesamiento se realizará utilizando la plataforma GEE.

Materiales y métodos

Área de estudio y adquisición de Datos

El departamento de General López se encuentra en el extremo sudoeste de la provincia de Santa Fe (Figura 1) y se caracteriza por ser uno de los más grandes, con un área de 11.558 km² y una población de 191.042 habitantes. Esta región geomorfológica, denominada Modelado Eólico Pos-Pampeano tiene un relieve con lomadas y zonas deprimidas, cubiertas por cuencas lagunares. La principal actividad económica de la región es la agropecuaria, principalmente cultivos de cereales y ganadería.

Los datos de las distintas coberturas fueron relevados in-situ por la Bolsa de Cereales de Buenos Aires y se corresponden a las campañas 2018/2019 y 2019/2020 (<https://www.bolsadecereales.com>). La información obtenida está compuesta por 849 puntos (Fig. 1), etiquetados con diferentes coberturas de suelo, cantidad y por campaña (Tabla 1).

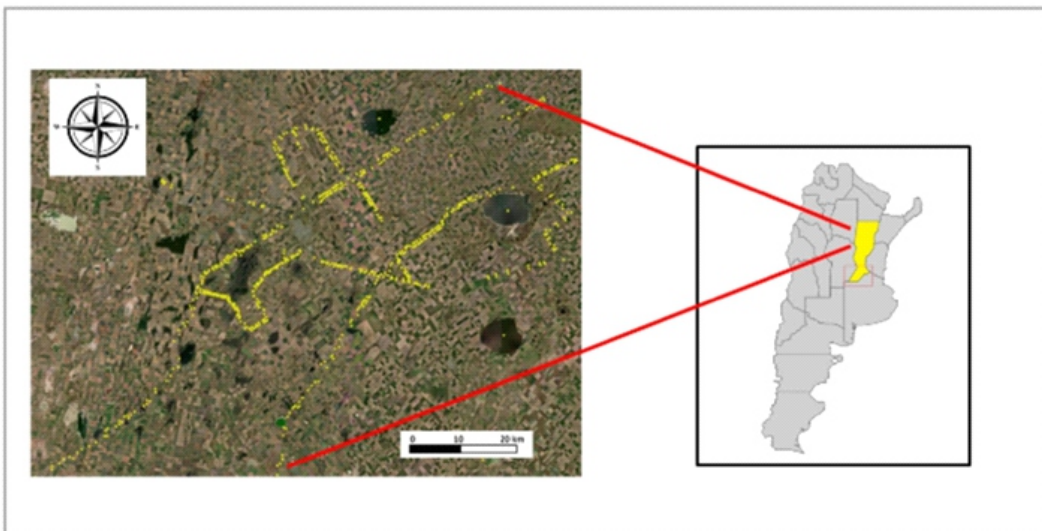


Figura 1: Mapeo de las muestras de coberturas de suelo (puntos amarillos) en el departamento General López.

Cultivo	2018/2019	2019/2020
Soja de 1°	155	189
Soja de 2°	1	88
Maíz temprano	82	128
Maíz tardío	1	3
Trigo	0	2
Girasol	1	0
Sorgo	66	0
Forrajes	10	45
Campo Natural	25	57
Barbecho	0	6
Agua	1	1
Alfalfa	2	0
No sabe	9	25
Urbano	0	12
Total	293	556

Tabla 1: Tipo y cantidad de cultivos que forman el dataset inicial

Las coberturas de suelo a identificar se corresponden con las clases mayoritarias, soja, maíz, campo natural, forrajes, cuerpos de agua y zonas urbanas. Se excluye la clase sorgo en el análisis debido a la ausencia representativa para la campaña agrícola 2019-2020. Dado el desbalance de datos que se observa en los datos originales (Tabla 1) se consideraron los datos de ambas campañas, y se realizaron procedimientos para la aumentación de los mismos, como la creación de polígonos para cada clase de cobertura. La identificación de coberturas de agua y zonas urbanas se realizaron por medio de la fotointerpretación llevada a cabo sobre un mosaico RGB de imágenes Sentinel-2. Los cultivos Maíz temprano, Maíz tardío, Soja 1° y Soja 2° fueron agrupados en una sola categoría (Maíz y Soja). El dataset resultante está formado por seis coberturas de suelo y una cantidad total de 945 muestras (Tabla 2) entre puntos y polígonos. La adición de polígonos busca mejorar la representación de las distintas coberturas de suelo, los mismos se crearon a partir de los puntos georeferenciados.

Clase	Cobertura	Cantidad de Puntos	Cantidad de Polígonos
1	Maíz	204	10
2	Soja	423	10
3	Campo Natural	77	5
4	Forrajes	54	1
5	Agua	69	2
6	Urbano	87	3
Total		914	31

Tabla 2: Categorías y cantidad de cultivos del dataset

Procesamiento de datos

El procesamiento de imágenes se realizó utilizando los servicios de cloud computing de la plataforma Google Earth Engine (<https://earthengine.google.com>). Se utilizaron las colecciones de imágenes multiespectrales Sentinel-2 con nivel 2A de pre procesado (COPERNICUS/S2_SR) correspondientes al periodo 2018-01-19 / 2020-04-30, con un porcentaje de nubosidad menor al 10%, a las cuales se accedió a través de la plataforma GEE. Estas imágenes están corregidas atmosféricamente y con valores de reflectancia a niveles inferiores de la atmósfera (Botton Of Atmosphere). Cada imagen de la colección está compuesta por 13 bandas espectrales de 10, 20 y 60 m de resolución. Se utilizó todo el espectro de reflectancia registrado por los sensores del satélite, tanto en el espectro visible, como en el infrarrojo medio, siendo las bandas utilizadas la B2, B3, B4, B5, B6, B7, B8 y B11 (Tabla 3). Se analizaron cinco índices espectrales para destacar aspectos de las diferentes coberturas. Los mismo fueron: NDVI, SAVI, NDWI, NDBI y PVR (Eq 1, 2, 3, 4 y 5) que se añadieron como bandas a la colección de imágenes. Para cada uno de los cuales, se calculó la media, el máximo, la mínima y desviación estándar. Toda esta información se agrupó en una imagen stack, donde cada banda e índices se incorporaron como una banda independiente. Sobre este dataset de imágenes se procedió a la superposición de la capa vectorial generada con los puntos y a la extracción de grupos de píxeles prototipos de cada clase.

	SENTINEL 2	
	Longitud de onda (µm)	Resolución (m)
Banda 1 - Aerosol	0,43 - 0,45	60
Banda 2 - Blue	0,45 - 0,52	10
Banda 3 - Green	0,54 - 0,57	10
Banda 4 - Red	0,65 - 0,68	10
Banda 5 - Red edge 1	0,69 - 0,71	20
Banda 6 - Red edge 2	0,73 - 0,74	20
Banda 7 - Red edge 3	0,77 - 0,79	20
Banda 8 - Near Infrared (NIR) 1	0,78 - 0,90	10
Banda 11 - SWIR 1	1,56 - 1,65	20

Tabla 3: Bandas Sentinel-2.

$$NDVI = \frac{NIR1 - RED}{NIR1 + RED} \quad (1)$$

$$SAVI = \frac{NIR1 - RED}{NIR1 + RED + L} (1 + L) \quad (2)$$

$$NDWI = \frac{GREEN - NIR1}{GREEN + NIR1} \quad (3)$$

$$NDBI = \frac{SWIR - NIR1}{SWIR + NIR1} \quad (4)$$

$$PVR = \frac{GREEN - RED}{REEN + RED} \quad (5)$$

El factor L, es encargado de amortiguar la presencia del suelo a través de valores comprendidos entre 0 (para zonas con gran densidad vegetal) y 1 (para zonas con escasa densidad vegetal), por defecto se consideró a L con un valor de 0,5.

Modelos de Clasificación y Métricas de evaluación

Dos algoritmos de aprendizaje supervisado fueron implementados para realizar la clasificación de coberturas de suelo: Random Forest (RF) y Support Vector Machine (SVM). El modelo "Random Forest" está formado por un conjunto de árboles de decisión individuales, cada uno entrenado con una muestra ligeramente distinta de los datos de entrenamiento generados mediante bootstrapping. La predicción de una nueva observación se obtiene agregando las predicciones de todos los árboles individuales que forman el modelo. En este caso, la salida es la moda de las predicciones de los árboles (<https://www.cienciadedatos.net>).

El algoritmo Support Vector Machine está basado en teoría estadística y encuentra el límite entre clases, utilizando planos de soporte generados a partir del set de datos de entrenamiento. El límite entre dos clases se encontrará equidistante a los dos planos generados por los vectores más cercanos a los de la clase vecina (vectores de soporte). En la literatura, SVM ha demostrado una alta precisión en predicción y capacidad de modelado en una amplia gama de problemas de clasificación real, reconocimiento de patrones y regresión. El rendimiento de un SVM, depende fuertemente del ajuste de sus hiper parámetros, los cuales inciden en la eficiencia de la clasificación (Rojas et al., 2020).

Para realizar la clasificación, el dataset se dividió en dos conjuntos de datos: entrenamiento

y validación. Para el entrenamiento se seleccionó en forma aleatoria el 80 % de los datos y un 20 % fue usado para la validación. Durante el proceso de entrenamiento del modelo RF, se llevaron a cabo diversas pruebas, donde se realizaron ajustes en la cantidad de árboles y semillas utilizadas. Con el fin de determinar la configuración óptima para lograr un clasificador con mayor precisión global, los valores de los parámetros del Random forest (Tabla 4) fueron los siguientes:

```
ee.Classifier.smileRandomForest(numberOfTrees,variablesPerSplit,minLeafPopulation,
bagFraction,maxNodes,seed)
```

Parámetro	Tipo	Descripción
number Of Trees	Entero	Número de árboles de decisión que se van a crear
variables Per Split	Entero, por defecto: null	El número de variables por división. Si no se especifica, se utiliza la raíz cuadrada del número de variables
mindLeaf Population	Entero, por defecto: 1	Crea únicamente nodos cuyo conjunto de entrenamiento contenga al menos esta cantidad de puntos.
bagFraction	Real, por defecto: 0,5	La fracción de entrada a bolsa por árbol.
maxNodes	Entero, por defecto: null	El número máximo de nodos de hojas en cada árbol. Si no se especifica, el valor predeterminado es sin limite
seed	Entero, por defecto: 0	La semilla de la aleatorización

Tabla 4: Descripción de parámetros RF

Donde el parámetro denominado número de árboles (number Of Trees) se estableció en 100 y la cantidad de semillas (seed) fueron 10.

En el modelo de SVM, se utilizaron los parámetros estándares (Tabla 5):

```
ee.Classifier.libsvm(decisionProcedure, svmType, kernelType, shrinking, degree, gamma,
coef0, cost, nu, terminationEpsilon, lossEpsilon, oneClass)
```

Parámetro	Tipo	Descripción
decisionProcedure	Cadena, por defecto: 'Voting'	El procedimiento de decisión a utilizar para la clasificación. Ya sea 'Votación' o 'Margen'
svmType	Cadena, por defecto: 'C_SVC'	El tipo de SVM. 'C_SVC', 'NU_SVC', 'ONE_CLASS', 'EPSILON_SVR' o 'NU_SVR'
kernelType	Cadena, por defecto: 'LINEAR'	Tipo de kernel: LINEAR, POLY, RBF o SIGMOID.
shrinking	Boleano, por defecto: true.	Si se debe utilizar heurística de contracción.
degree	Entero, por defecto: null	El grado de polinomio. Válido para núcleos POLY.
gamma	Real, por defecto: 0	El valor de gamma en la función del núcleo. El valor predeterminado es el recíproco del número de características. Válido para los kernels POLY, RBF y SIGMOID.
coef0	Real, por defecto: null	El valor predeterminado es 0. Válido para núcleos POLY y SIGMOID.
cost	Real, por defecto: null	El valor predeterminado es 1. Solo válido para C-SVC, epsilon-SVR y nu-SVR
nu	Real, por defecto: null	El valor predeterminado es 0,5. Solo válido para un-SVC, SVM de una clase un-SVR.
terminationEpsilon	Real, por defecto: null	El valor predeterminado es 0,001. Solo válido para epsilon-SVR.
lossEpsilon	Real, por defecto: null	El valor predeterminado es 0,1. Solo válido para epsilon-SVR
oneClass	Entero, por defecto: null	La clase de los datos de entrenamiento sobre los cuales entrenar en una SVM de una clase. El valor predeterminado es 0. Solo es válido para SVM de una clase.

Tabla 5: Descripción de parámetros SVM

Los algoritmos de clasificación se realizaron utilizando el Code Editor de GEE, el cual está basado en una IDE que usa el lenguaje para la API JavaScript.

El desempeño de la clasificación, se evaluó mediante la matriz de confusión (Tabla 6), en donde cada columna de la matriz representa el número de predicciones de cada clase, y cada fila representa a las instancias en la clase real:

1. Verdadero positivo (Tp) si la instancia es clasificada correctamente y su clase pertenece a la positiva.
2. Verdadero negativo (Tn) si la instancia es correctamente clasificada con la clase de valor negativo.
3. Falso positivo (Fp) cuando la instancia es clasificada incorrectamente como positiva cuando en realidad es negativa.
4. Falso negativo (Fn) se presenta cuando la instancia fue clasificada incorrectamente como negativa cuando en realidad es positiva.

	Negativo	Positivo	Total
Negativo	Tn	Fp	An=Tn+Fp
Positivo	Fn	Tp	Ap=Fn+Tp
Total	Pn =Tn+Fn	Pp = Fp+Tp	T

Tabla 6 Matriz de Confusión

Esta matriz permite computar la precisión global (Ag) del clasificador, como el porcentaje de clases correctamente identificadas sobre el total, además de ver los tipos de aciertos y errores que está teniendo nuestro modelo a la hora de ajustarse a los datos observados (Godoy, 2021). De esta matriz se calculan dos tipos de errores: de 'omisión' (O) y 'comisión' (C). Un error tipo (O) es cuando se excluye un valor de la categoría que está siendo evaluada y un error tipo (C), cuando se incluye un valor incorrectamente en la categoría que está siendo evaluada. Las mediciones de precisión, expresadas como el porcentaje de clasificaciones correctas, se conocen como Precisión del Usuario y Precisión del Productor. La Precisión del Usuario mide por complemento los errores de comisión, mientras que la Precisión del Productor evalúa por complemento los errores de omisión. Otra métrica normalmente utilizada es el coeficiente Kappa (K). Este índice representa la proporción de acuerdos observados respecto del máximo acuerdo posible más allá del azar y no tiene en cuenta el desbalance de clases (Borras et al., 2017). K toma valores entre -1 y +1 (Tabla 7). Valores de K cercanos a +1 indican un mayor grado de concordancia interobservador y, más cercano a -1 mayor grado de discordancia (Marini, 2013).

Kappa	Grado de Acuerdo
< 0.00	Sin acuerdo
0.01-0.20	Insignificante
0.20-0.40	Bajo
0.40-0.70	Bueno
0.70-1.00	Óptimo

Tabla 7. Grado de Concordancia del Índice Kappa

Resultados

Dada la cantidad de datos de muestreo, que consiste en 945 muestras de puntos, clasificadas en categorías como Soja, Maíz, Cuerpos de Agua, Urbano, Campo Natural y Forraje, la precisión global (Tabla 11) inicial utilizando los algoritmos de clasificación Random Forest y SVM fue de aproximadamente el 71% para ambos modelos. Con el objetivo de mejorar la precisión en la clasificación, se decidió tomar polígonos de los datos de muestreo, lo que permitió aumentar de manera significativa la cantidad de píxeles disponibles para la clasificación. Como resultado los modelos de clasificación implementando RF y SVM tienen una precisión global del 97% y el 93% respectivamente, el coeficiente Kappa fue del 0,96 para el modelo Random Forest y de 0.91 para SVM, mostrando que existe un óptimo grado de concordancia entre las clases predichas y la verdad de campo. El aumento de datos mediante la inclusión de polígonos demostró ser efectivo para mejorar la clasificación en comparación con la cantidad limitada de datos de muestreo. Observando las matrices de confusión Tabla 8 y 9 y la Figura 2, se extraen algunas conclusiones que para la clase Cuerpos de Agua tiene una alta precisión en ambos modelos, mientras que para las clases Campo Natural y Forraje ambos modelos presentan mayor confusión en la clasificación. Otra métrica que se usó para la evaluación de los modelos fue precisión de comisión y omisión (Tabla 10a y 10b), se observan en las tablas que los valores son superiores a 0.85 asegurando una alta probabilidad de aciertos, salvo para el caso de forraje del clasificador SVM. En la figura 2 se muestra el resultado final de la clasificación para el área de estudio.

	Soja	Maíz	Agua	Urbano	Camp o	Forraje	Total
Soja	2107	34	0	3	7	1	2146
Maíz	39	1964	0	0	2	0	2005
Agua	0	0	1094	0	1	0	1095
Urbano	0	1	0	711	2	0	714
Campo	6	4	0	11	633	0	654
Forraje	1	3	0	5	4	53	66
Total	2153	2006	1094	730	649	54	6686

Tabla 8: Matriz de confusión RF. Predicción

	Soja	Maíz	Agua	Urbano	Camp o	Forraje	Total
Soja	2025	109	0	1	16	1	2152
Maíz	168	1828	0	1	7	1	2005
Agua	0	0	1095	0	0	0	1095
Urbano	2	1	0	701	10	0	714
Campo	49	23	1	20	560	1	654
Forraje	14	18	0	1	4	29	66
Total	2258	1979	1096	724	597	32	6686

Tabla 9: Matriz de confusión SVM. Predicción

a)	Omisión	Comisión	b)	Omisión	Comisión
Soja	0.98	0.97	Soja	0.94	0.89
Maíz	0.98	0.98	Maíz	0.91	0.92
Agua	0.99	0.99	Agua	0.99	0.99
Urbano	0.99	0.99	Urbano	0.98	0.96
Campo Nat.	0.97	0.97	Campo Nat.	0.85	0.93
Forraje	0.82	0.97	Forraje	0.43	0.90

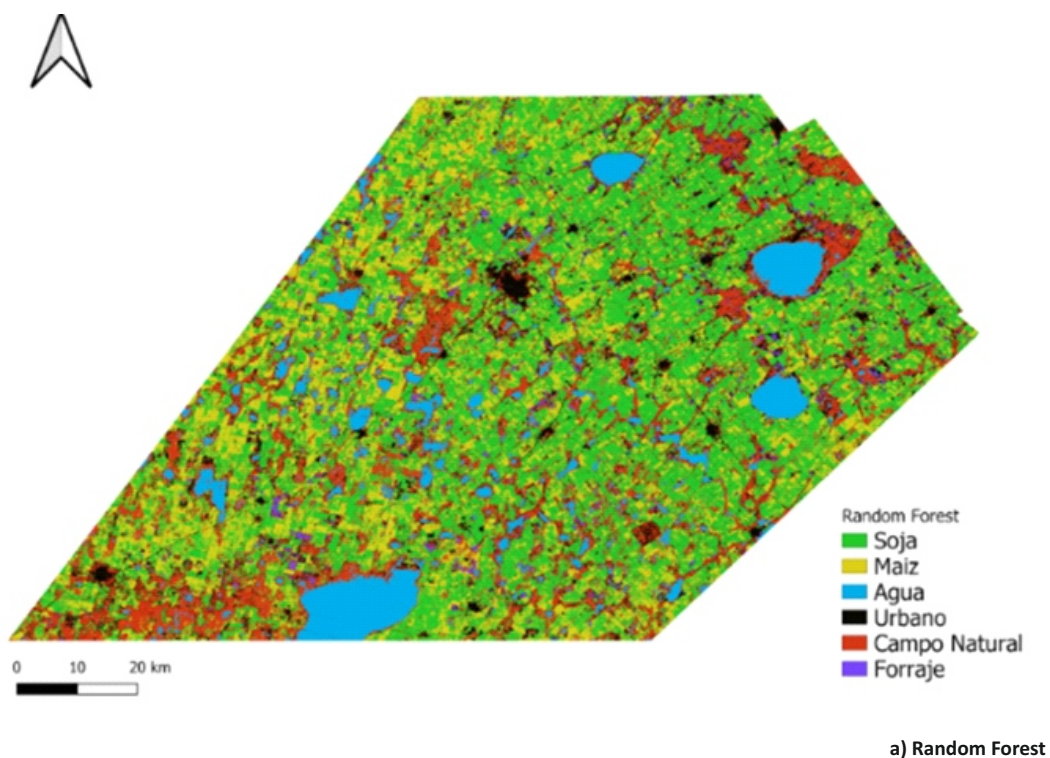
Tabla 10: Precisión Comisión y Omisión: a) RF vs b) SVM

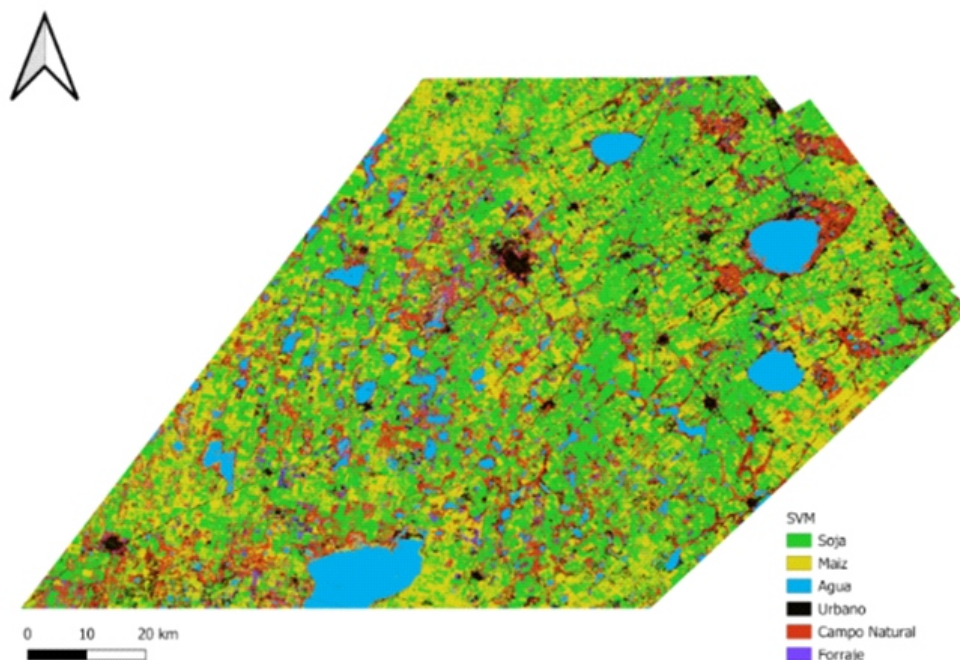
Resultados

Parámetro	SVM	Random Forest
Precisión global (%)	0.93	0.97
Índice Kappa	0.91	0.96

Tabla 11: Métricas de evaluación

En la Figura 2 se observa el mapeo de las seis clases con ambos métodos de clasificación. En ambas imágenes, se puede ver la correspondencia visual existente.





b) Support Vector Machine

Figura 2: Clasificación de cultivos, cuerpos de agua y zonas urbanas.

Conclusión

En este estudio, se realizó una clasificación de coberturas de suelo representando seis clases prioritarias dada la actividad económica de la región. El análisis se desarrolló con datos de las campañas 2018-2019 y 2019-2020. Dos modelos de clasificación se implementaron para clasificar las diferentes coberturas de suelo en el departamento General Lopez (Random Forest y Suport Vector Machine). Ambos modelos mostraron una buena exactitud y precisión global superior al 90%, siendo Random Forest el modelo que mejor ajuste tiene a los datos, con un Ag de 0.97, mientras que el modelo Support Vector Machine tuvo un Ag de 0.93. Los resultados obtenidos a partir de esta investigación, proporcionan una visión general de la distribución espacial de coberturas de suelo, lo que permite cuantificar el uso del suelo para una mejor planificación, gestión del territorio y manejo de los recursos naturales. El trabajo a futuro es mejorar el modelo introduciendo más clases y separando los cultivos de primera y segunda, inclusive extrapolar el modelo a otras áreas y analizar su comportamiento.

Referencias bibliográficas

1. Alhassan, V., Henry, C., Ramanna, S., & Storie, C. (2020). A deep learning framework for land-use/land-cover mapping and analysis using multispectral satellite imagery. *Neural Computing and Applications*, 32, 8529-8544.
2. Álvarez, A. R., Gálvez, L. A. S., García, M. A., Gálvez, S. S., & Gómez, M. L. (2020). Sistema de reconocimiento de voz basado en un método de aprendizaje supervisado y la correlación de Pearson (K-NN Algorithm and Pearson correlation-based a voice recognition system). *Pistas Educativas*, 42(136).
3. Bolsa de Cereales Homepage, <https://www.bolsadecereales.com>
4. Borràs, J., Delegido, J., Pezzola, A., Pereira-Sandoval, M., Morassi, G., & Camps-Valls, G. (2017). Clasificación de usos del suelo a partir de imágenes Sentinel-2. *Revista de Teledetección*, (48), 55-66.
5. Cardiel, J. P., Rodrigo, S. G., & Moreno, L. M. (2022). Operadores matemáticos en Física simulados mediante Inteligencia Artificial. Universidad de Zaragoza, CIEN, Departamento de Física de la Materia Condensada, Área de Física de la Materia Condensada.
6. Ciencia de datos, <https://www.cienciadedatos.net>
7. Coca-Castro, A., Zaraza-Aguilera, M. A., Benavides-Miranda, Y. T., Montilla-Montilla, Y. M., Posada-Fandiño, H. B., Avendaño-Gómez, A. L., & Franco-Prieto, C. A. (2021). Evaluación de algoritmos de clasificación en la plataforma Google Earth Engine para la identificación y detección de cambios de construcciones rurales y periurbanas a partir de imágenes de alta resolución. *Revista de Teledetección*, (58), 71-88.
8. Curioso, W. H., & Brunette, M. J. (2020). Artificial intelligence and innovation to optimize the tuberculosis diagnostic process. *Revista Peruana de Medicina Experimental y Salud Pública*, 37(3), 554-558.
9. da Silva, V. S., Salami, G., da Silva, M. I. O., Silva, E. A., Monteiro Junior, J. J., & Alba, E. (2020). Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification. *Geology, Ecology, and Landscapes*, 4(2), 159-169.
10. Ferrelli, F., Brendel, A., Perillo, G. M. E., & Piccolo, M. C. (2022). Dinámica estacional de las coberturas del suelo en una región semiárida de Argentina mediante teledetección. *Revista Geográfica Venezolana*, 63 (1), 64-79.
11. Trujillo-Jiménez, M. A., Liberoff, A. L., Pessacq, N., Pacheco, C., & Flaherty, S. (2021). Metodología de clasificación automática de uso y cobertura de suelo. In XIII Congreso de AgroInformática (CAI 2021)-JAIIO 50 (Modalidad virtual).
12. Godoy, F. E. (2021). Métodos clásicos de clasificación: comparación y aplicación (Bachelor's thesis). Recuperado de: <http://hdl.handle.net/11086/19768>.
13. Google Earth Engine, <https://earthengine.google.com>
14. Guerschman J. P., Paruelo J. M., C. Di Bella, M. C. Giallorenzi & F. Pacin (2003): Land cover classification in the Argentine Pampas using multi-temporal Landsat TM data,

International Journal of Remote Sensing, 24:17, 3381-3402.

15. Marini, M. F. (2013). Discriminación de cultivos de distinto desarrollo utilizando imágenes satelitales MODIS. *GeoFocus. International Review of Geographical Information Science and Technology*, (13_1), 48-60.

16. Martín Bernal, R. C. & Rodríguez Baquero, W. C. (2021). Clasificación de imágenes de radar de apertura sintética aplicando Corine Land Cover adaptada para Colombia mediante redes neuronales convolucionales. Recuperado de: <http://hdl.handle.net/11349/29098>.

17. Nolasco, M., Willington, E., & Bocco, M. (2016). Uso del suelo agrícola: comparación entre series temporales e imágenes satelitales individuales para su clasificación. *Revista de Investigaciones de la Facultad de Ciencias Agrarias-UNR*, (26), 017-021.

18. Rojas, M. G., Carballido, J. A., Olivera, A. C., & Vidal, P. J. (2020). Optimización de Support Vector Machine mediante metaheurísticas para clasificación de retinopatía diabética. XXI Simposio Argentino de Inteligencia Artificial (ASAI 2020) - JAIIO 49 (Modalidad virtual); Buenos Aires; Argentina; 2020; 73-86

19. Veneros, J., García, L., Morales, E., Gómez, V., Torres, M., López-Morales, F. (2020). Aplicación de sensores remotos para el análisis de cobertura vegetal y cuerpos de agua. *Idesia (Arica)*, 38(4), 99-107