



Cobertura urbana y expansión territorial en el AMGR: análisis multitemporal con Random Forest y SVM

Urban Land Cover and Territorial Expansion in the AMGR: Multitemporal Analysis Using Random Forest and SVM

 Laura Fabiana Gómez

Instituto de Investigación para el Desarrollo Territorial y el Hábitat Humano (UNNE-CONICET), Universidad Nacional del Nordeste, Argentina
 agrimlauragomez@gmail.com

Recibido: 29 Enero 2025

Aprobado: 13 Marzo 2025

Publicado: 01 noviembre 2025

Cita sugerida: Gómez, L. F. (2025). Cobertura urbana y expansión territorial en el AMGR: análisis multitemporal con Random Forest y SVM. *Geograficando*, 21(2), e183.
<https://doi.org/10.24215/2346898Xe183>

Resumen: La generación de información geográfica de coberturas urbanas es fundamental para la planificación y gestión del territorio. En este contexto, el objetivo de este trabajo fue comparar la cobertura urbana del Área Metropolitana del Gran Resistencia (AMGR), considerando su área de influencia y las localidades incluidas en ella, de los años 1991 y 2022, mediante la generación de mapas de áreas edificadas y no edificadas utilizando clasificaciones con algoritmos de aprendizaje automático (*Random Forest - RF* y *Support Vector Machine - SVM*). Los resultados muestran fiabilidades globales aceptables para ambas clasificaciones, siendo superior la clasificación generada con SVM. Por un lado, se observó un incremento superior al 200% en la superficie edificada; por otro, la distribución espacial de la superficie edificada ocurrió en torno a las superficies edificadas existentes en 1991, así como en espacios interiores a las mismas, aportando a la expansión de la ciudad por un lado, y a la consolidación de esta por otro. Además, se identificaron nuevos espacios edificadas de manera discontinua en sectores intermedios de las localidades que conforman el área de estudio.

Palabras clave: Expansión Urbana, Aprendizaje Automático, Bosques Aleatorios, Máquinas de Vectores Soporte

Abstract: The generation of geographic data on urban land cover is essential for territorial planning and management. In this context, the objective of this study was to compare urban land cover in Área Metropolitana del Gran Resistencia (AMGR), including its zone of influence and the localities within it, between 1991 and 2022. This was achieved by generating maps of built and unbuilt areas using machine learning classification algorithms (Random Forest - RF and Support Vector Machine - SVM). The results show acceptable overall accuracy for both classifications, with the SVM classification showing superior results. On the one hand, built-up areas increased by over 200%; on the other, spatial distribution of this expansion occurred both around pre-existing built-up areas (1991) and within their interiors, contributing to urban sprawl in some sectors and urban consolidation in others. Additionally, new discontinuous built-up areas were identified in intermediate zones of the localities comprising the study area.

Keywords: Urban Expansion, Machine Learning, Random Forest, Support Vector Machines



Introducción

El fenómeno de expansión urbana es un proceso desafiante que se produce de forma acelerada y que afecta ciudades de diferentes escalas. En las ciudades de tamaño intermedio, cuya definición según un criterio cuantitativo varía dependiendo del país o la región en la que se encuentre siendo en Argentina las que cuentan entre 50.000 y 1.000.000 de habitantes (Bellet y Llop, 2004); se ha verificado que este proceso se ha dado a un mayor ritmo a partir de la década de 1970 (CAF Banco de Desarrollo de América Latina, 2017; Ministerio de Planificación Federal, Inversión Pública y Servicios, 2011). Particularmente en Argentina se detectó que, entre los años 2003 y 2016, los grandes aglomerados urbanos (GAU), “ciudades con más de 100.000 habitantes y donde vive más del 66% de la población” (Lanfranchi, 2019) y entre las que se encuentra el Gran Resistencia, crecieron en promedio el doble en territorio que en habitantes, con picos de hasta seis veces (Lanfranchi, Duarte y Granero Realini, 2018).

Dado su tamaño poblacional y sus dimensiones espaciales, en las ciudades intermedias es posible lograr (planificación urbana de por medio) desarrollos territoriales equilibrados y sustentables, detener el crecimiento incontrolado en las grandes aglomeraciones urbanas, proveer de bienes y servicios a las poblaciones que albergan municipios cercanos y descentralizar administrativa y gubernamentalmente a las grandes ciudades generando más participación ciudadana. Asimismo, la expansión de una ciudad y las actividades humanas que van de la mano, ejercen presión sobre los ecosistemas afectando las dinámicas ambientales, por lo que contar con información actualizada y precisa relativa al desarrollo espacial de la cobertura urbana, su evolución y cambios en el tiempo, se configura como información esencial para gestionar el territorio y establecer líneas de acción que persigan la mejora en la calidad de vida de los habitantes, ya sea en lo relativo al ambiente, cuestiones sociales o económicas, aportando de esta manera a la generación de una ciudad sustentable.

En este punto es importante introducir la diferencia entre cobertura de la tierra y uso de la tierra. Por un lado, al hablar de cobertura terrestre, se hace referencia a la cubierta observable sobre la superficie terrestre. En este sentido, Di Gregorio y Jansen (2005) incluyen como cobertura terrestre a la vegetación natural, los cursos y cuerpos de agua, el suelo desnudo y las rocas, y elementos antrópicos propios de la actividad humana. Por otro lado, el uso de la tierra se refiere a una interpretación sobre la forma en que se manejan las coberturas terrestres, y generalmente tienen una connotación económica, por ejemplo: uso residencial, industrial, agrícola, forestal, etc. (Meyer y Turner, 1994). Es claro entonces que es la cobertura terrestre la que interacciona con la energía electromagnética y, a partir de la respuesta espectral obtenida, permite la obtención de información relativa a la cobertura presente en el territorio, no así del uso que se le dé a dicha cobertura (Marlenko, 2011).

Dadas sus características en torno a disponibilidad, gratuidad, periodicidad, resolución espacial, alcance, etc., las imágenes multiespectrales son un insumo de gran importancia para el monitoreo espacio temporal permanente de las coberturas terrestres. En cuanto a su clasificación para la obtención de dichas coberturas, el gran volumen de datos disponibles, junto con avances tecnológicos como la computación en la nube y el desarrollo de técnicas de aprendizaje automático posibilitan la generación de información precisa con menores tiempos de procesamiento y a diferentes escalas espacio temporales (Yuan et al., 2020; Beron de la Puente, Montico, Revollo, Gil y Zapperi, 2024). Gracias a estos avances, la actualización periódica de estos datos se ha vuelto factible y accesible.

Gil Leiva, Díaz Ortuño y Rodríguez Muñoz (2019) señalan que es posible distinguir tres tipos de procesos en el análisis de imágenes: procesos de bajo nivel, que incluyen aspectos tales como la adquisición y preparación de las imágenes, y el uso de diferentes combinaciones de bandas espectrales asignadas como ejes en el espacio de color RGB; procesos de nivel medio, en el que se encuentran la segmentación y el marcado de partes de imágenes; y procesos de nivel alto, en el que se incluye el reconocimiento de patrones y la interpretación, entre los que se encuentran Random Forest - RF y Support Vector Machine - SVM. Por su parte, Dabija et al. (2021), revisan una década de publicaciones científicas y observan que la mayoría de las metodologías empleadas para clasificar imágenes son de tipo supervisado basadas en la aplicación de RF, SVM y redes neuronales.

Diversos estudios han evaluado el desempeño comparativo de algoritmos supervisados como RF y SVM en clasificación de imágenes satelitales, demostrando consistentemente su alta precisión estadística, con ligeras ventajas para SVM. A modo de ejemplo, Chucos Baquerizo y Vega Ventocilla (2022) analizaron clasificaciones de coberturas boscosas, agrícolas, urbanas y acuáticas en la Amazonía peruana utilizando imágenes Landsat 8. Compararon el método no supervisado K-means con enfoques supervisados (RF y SVM), concluyendo que estos últimos superan significativamente en precisión al primero, y obteniendo mayor precisión para SVM respecto de RF. Los autores destacaron algunas limitaciones críticas: la necesidad de conjuntos de entrenamiento suficientemente grandes para evitar sobreajuste en algoritmos supervisados; la disminución en la exactitud de la clasificación de las áreas urbanas (aun contando con una precisión global del 91%), lo que atribuyen a la complejidad espectral de las zonas urbanas, donde materiales heterogéneos (techos, asfalto, vegetación urbana) generan firmas espectrales similares a otras clases (suelo desnudo o cultivos); y exponen que el uso de imágenes Landsat 8 (resolución de 30m) dificulta la delimitación precisa de estructuras urbanas pequeñas o dispersas por lo sugieren el uso de imágenes de mayor resolución espacial (Sentinel-2 a 10 m) para mejorar la delimitación de características detalladas y aumentar la precisión.

En un enfoque multiescala, Dabija et al. (2021) evaluaron RF (500 árboles) y SVM (con núcleos lineal, polinomial, radial y sigmoidal) en tres regiones europeas, empleando datos de Landsat 8 y Sentinel-2 (resampleados a 60 m) bajo el sistema Corine Land Cover. Entre sus resultados evidenciaron que el incremento en el número de píxeles de entrenamiento por clase mejora la precisión general, y que SVM con núcleo radial mostró superioridad al capturar patrones espaciales complejos en paisajes heterogéneos, especialmente con datos Sentinel-2.

Abdi (2019) amplió esta comparativa al incluir cuatro algoritmos (RF, SVM, XGBoost y redes neuronales) para clasificar coberturas mixtas en Suecia mediante imágenes multitemporales Sentinel-2. SVM nuevamente destacó con la mayor precisión global (75.8%), aunque se identificaron desafíos persistentes en la clasificación de superficies artificiales, coincidiendo con observaciones previas (Chucos Baquerizo y Vega Ventocilla, 2022).

En una perspectiva temporal más amplia, no centrada en áreas urbanas, Nery de Figueiredo, Sadler, Solis-Aulestia, White, Polyakov y Chalak Haghighi (2016) compararon los algoritmos supervisados SVM, K- Nearest Neighbors, RF, Learning Vector Quantization, Recursive Partitioning y Gradient Boosting Machine en la clasificación de series temporales de imágenes Landsat (de sensores MSS, TM, ETM+ y OLI) entre 1979 y 2014 en una región australiana, utilizando clases como agricultura, agua, bosque nativo, dunas, plantación forestal y bosque nativo talado. Sus resultados afirmaron la mayor precisión de SVM y RF, que ocuparon el primer y segundo lugar respectivamente. Los autores además encontraron que las precisiones disminuyeron en imágenes más recientes y atribuyeron esto a cambios en resoluciones espectrales y ruidos en los sensores. Entre las limitaciones expuestas, comentan por un lado, la necesidad de aumentar el número de píxeles de entrenamiento cuando el objetivo es discriminar clases complejas con similitudes espectrales; por otro lado, al no incluir áreas urbanizadas (por ser de una superficie despreciable en el área de estudio, menos del 0,5%) no podrían generalizar sus resultados en áreas con mayores intervenciones humanas.

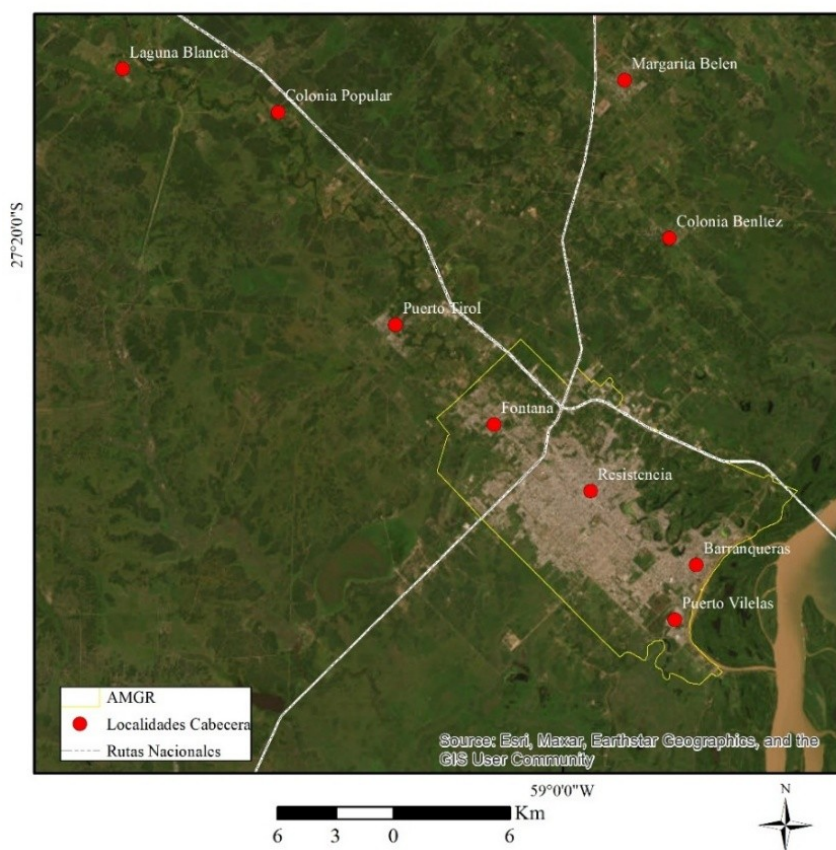
El objetivo del presente trabajo es generar evidencia que permita cuantificar la superficie con cobertura urbana y describir la distribución espacial de la expansión urbana experimentada por el Área Metropolitana del Gran Resistencia (AMGR) con relación a sus localidades de influencia, entre los años 1991 y 2022, a partir de la aplicación de algoritmos de aprendizaje automático. Para ello se generaron dos pares de mapas de coberturas de la tierra con la leyenda edificado y no edificado para los años 1991 y 2022, utilizando los algoritmos de aprendizaje automático *Random Forest* (RF) y *Support Vector Machine* (SVM); y se compararon las confiabilidades globales de las clasificaciones obtenidas para avanzar con el resultado más fiable sobre el análisis de la expansión de la mancha urbana.

Materiales y Métodos

Área de estudio

El análisis se llevó a cabo al Noreste de la República Argentina, en el Área Metropolitana del Gran Resistencia (Provincia del Chaco) que incluye a los municipios de Resistencia, Barranqueras, Puerto Vilelas y Fontana, y las localidades Puerto Tirol, Colonia Popular, Laguna Blanca, Colonia Benítez y Margarita Belén por estar en su área de influencia directa y que dependen funcionalmente de Resistencia. (Scornik y Godoy, 2008) (Figura N° 1). En los últimos años el AMGR experimentó el proceso de expansión urbana (de forma discontinua) vinculando estas localidades debido a las dinámicas habituales en torno a diversas actividades (comerciales, laborales, de salud, ocio, etc.) y a la movilidad residencial existente. En palabras de Zalazar (2024) en relación a la periferia del AMGR, “en las zonas periféricas se advierte el denominado fenómeno de turistificación que en la actualidad comienza a acentuarse cada vez más, acompañado de grupos poblacionales que buscan o desean “alejarse” del mundo convulsionado en que se vive en las grandes ciudades”.

Figura 1. AMGR y localidades en su área de influencia



Fuente: Elaboración propia sobre la base de archivos del Instituto Geográfico Nacional (2025)

Insumos y su tratamiento

Para analizar la expansión de la mancha urbana se utilizó como insumo primario imágenes satelitales de las Misiones Landsat 5 (para el año 1991) y Landsat 8 (para 2022) descargadas desde la página web del Servicio Geológico de los Estados Unidos (USGS). Ambas imágenes poseen la misma resolución espacial (30m) en las bandas utilizadas.

A partir de las mismas se extrajo el *subset* que incluye al AMGR y las localidades que se encuentran a 20 Km a la redonda. Esta extracción se realizó a partir de un polígono construido a tal fin a partir de la interpretación visual de imágenes de alta resolución espacial como son las provistas por *Google Earth*®.

Para la mejor identificación de la mancha urbana se utilizó, por un lado, la composición color verdadero y, por otro, composiciones falso color que posibilitaran una mejor discriminación entre la superficie edificada y la no edificada. La composición falso color utilizada fue la de Infrarrojo cercano, rojo y verde (correspondientes a las bandas 4-3-2 de la imagen Landsat 5 para el año 1991 y a las bandas 5-4-3 de la imagen Landsat 8 para el año 2022). A partir de allí se recogieron regiones de interés (ROI) procurando que la elección de estas sea sobre áreas homogéneas a su interior y heterogéneas entre ellas, y que además estén distribuidas en toda el área de estudio de forma equilibrada evitando concentraciones en áreas específicas. Cabe mencionar en este punto, que la clase denominada superficie edificada corresponde a áreas donde la superficie ha sido transformada por actividades humanas para construir infraestructuras de tipo vial, edilicia, y otras estructuras urbanas relacionadas con actividades ciudadanas; los píxeles que no respondían a estas características fueron asignados a la clase no edificado (cursos de agua, distintos tipos de vegetación, suelo desnudo). En este proceso, a cada ROI se le asignó un código según la clase a la que perteneció generando de esta forma un archivo utilizado para entrenar los algoritmos RF y SVM y generar mapas de cobertura edificada y no edificada para cada año. En ambos casos el modelo toma el 50% de los píxeles considerados para entrenar los modelos y el restante 50% para validarlo. Los mapas resultantes son acompañados por mapas de confianza que permite obtener el complemento DZetsaka y que indican mayor confianza a mayor valor indicado en la leyenda. A efectos de simplificar la interpretación visual de los mapas de confianza, su leyenda es presentada en rangos construidos a partir de la división en rangos naturales, que diferencia las clases de la forma más pronunciada posible manteniendo mínima la variación dentro cada una.

Tras las clasificaciones realizadas se procedió a validar las correspondientes al año 2022 mediante una matriz de tabulación cruzada, considerando como verdad de terreno imágenes de mayor resolución espacial disponibles en *Google Earth*® para dicho año. Se realizó un muestreo aleatorio estratificado de puntos distribuidos de forma proporcional a la cantidad de píxeles en cada categoría, determinando previamente el tamaño de la muestra de 665 píxeles con un nivel de confianza del 99% y una tolerancia de error del 5%. Los puntos para cada clase fueron creados de forma aleatoria con la herramienta destinada a tal fin disponible en software SIG de escritorio, para luego ser superpuestos sobre las imágenes de *Google Earth*® y de esta manera determinar, mediante apreciación visual la coincidencia o no entre la clase a la que pertenece cada punto generado y la presente en la verdad terreno en la misma localización. Con esta información se calcularon métricas específicas para evaluar la confiabilidad global. Cabe mencionar en este punto que para la clasificación del año 1991 no fue posible realizar una validación del resultado dada la inexistencia de información que pudiera ser considerada como verdad de terreno.

A partir de la cartografía del año 2022 validada con mayor porcentaje de confiabilidad global, se realizó la corrección de aquellos píxeles que fueron clasificados como “edificados” en el año 1991 y posteriormente fueron clasificados como “no edificados” en el año 2022, partiendo del supuesto de que, en términos generales, la superficie edificada o intervenida por el hombre no retrocede (no desaparece/no cambia) con el avance del tiempo. Este último proceso se realizó a partir de la metodología propuesta por Linares, Iturrioz y Di Nucci (2022), quienes a partir de operaciones algebraicas entre los rásters resultantes logran eliminar lo que resultó “erróneamente clasificado” en el pasado.

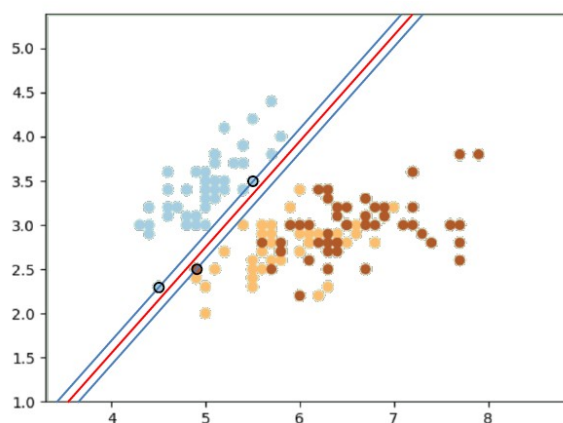
Con relación al software, se utilizó el QGIS versión 3.36.3 para el procesamiento de las imágenes y el levantamiento de regiones de interés (ROI), mientras que, para la etapa de clasificación, se utilizó el complemento *DZetsaka classification tool*. Esta herramienta permite evaluar la calidad de la clasificación mediante la generación de un mapa de confianza por píxel que se crea en función de los datos de entrenamiento. Este mapa se puede usar para mostrar dónde el algoritmo obtuvo el mejor ajuste entre el modelo y los datos de entrenamiento (Parr, 2024). Para la etapa de validación, se generaron puntos aleatorios con la herramienta *create random points* disponible en el software ArcGIS versión 10.5.

Clasificación con Máquinas de Vector Soportes (SVM)

La clasificación realizada fue de tipo supervisada que se apoya en el algoritmo SVM, que persigue encontrar un hiperplano óptimo que divida las clases. El proceso de selección de dicho hiperplano responde a un proceso iterativo que construye márgenes óptimos entre los vectores de soporte (conformados por los píxeles que tocan esos límites), es decir, el margen mayor entre las clases (Campo León, s/f). Se entiende que, a mayor margen (distancia entre clases), mejor será la segregación entre las mismas. Asimismo, dado que el entrenamiento se realiza a partir de una muestra de elementos, obtener un margen amplio ofrece mayor “garantía” de realizar una mejor clasificación cuando se tomen todos los píxeles. Sin embargo, esta “mejora” generará la posibilidad de clasificar de forma incorrecta valores atípicos.

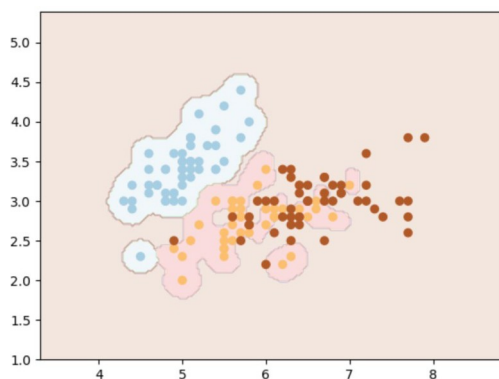
Sin perder generalidad, la figura a continuación muestra esquemáticamente y de forma simplificada el modo en que opera el algoritmo (en este diagrama la distribución de los datos es linealmente separable). La línea roja representa el hiperplano óptimo que separa los puntos pertenecientes a cada clase; líneas azules corresponden a los vectores soporte (los que pasan por puntos o píxeles de la imagen) (Figura N°2). Cuando existe mayor confusión entre las clases, el algoritmo se vale de transformaciones Kernel que permiten la distribución de los datos en un espacio de dimensión superior, con el objetivo de separarlos de una forma más sencilla. La figura N°3 esquematiza de forma sencilla la utilidad de usar Kernels para distinguir elementos pertenecientes a cada clase. Para esta contribución se utilizó un Kernel de base radial (RBF) o gaussiano.

Figura 2. Diagrama de SVM



Fuente: Clark Labs, 2020, p. 165.

Figura 3. Diagrama de SVM usando una transformación Kernel para separar puntos

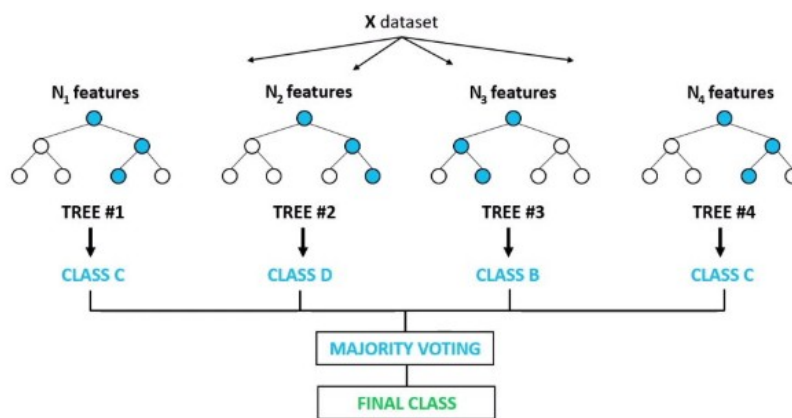


Fuente: Clark Labs, 2020, p. 166.

Clasificación con Bosque Aleatorio (*Random Forest*)

Los bosques aleatorios se encuentran dentro de los algoritmos de aprendizaje automático y basan su funcionamiento en el trabajo de un conjunto de árboles de decisión que se producen a partir de datos divididos de forma aleatoria (muestra de arranque de la cual un tercio se reserva para validar el modelo) (Jehad, Rehanullah, Nasir y Imran, 2012). Su inicio es a partir de un nodo (raíz) desde el cual salen ramas a un siguiente nodo (interno o de decisión) donde se evalúa la presencia o no de determinada característica para dividir los datos en conjuntos homogéneos (nodos hoja o terminales). Las ramas definen un rango de valores. En términos generales, como describe Breiman (2001) *Random Forest* cultiva muchos árboles de clasificación a partir de nuevas muestras (que pueden ser con reemplazo) del conjunto de datos de entrenamiento (Biau, 2012). Cada árbol genera una clasificación a partir de estos valores de entrada (“un voto”). El bosque elige la clasificación que tiene más votos (sobre todos los árboles del bosque) (Figura N°4).

Figura 4. Diagrama de Bosque Aleatorio

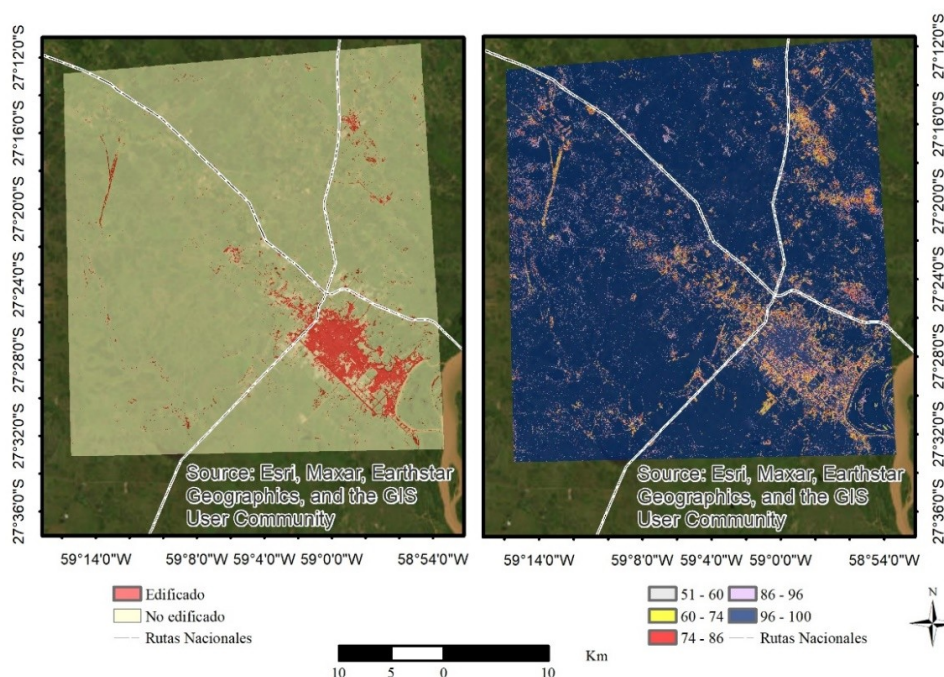


Fuente: Davis, 2020.

Resultados

La apreciación visual de los mapas de confianza generados en cada caso evidencia un mayor grado de incertidumbre en la clasificación realizada mediante RF (Figura N°5), particularmente en los sectores próximos a Colonia Benítez y Margarita Belén y al suroeste del AMGR (valores entre 60 y 86). Asimismo, el algoritmo RF clasifica gran cantidad de píxeles con mayor grado de incertidumbre al interior del AMGR. Por otro lado, es posible apreciar que la superficie no edificada ha sido clasificada con el mayor nivel de confianza (entre 96-100) en la mayor parte del área de estudio.

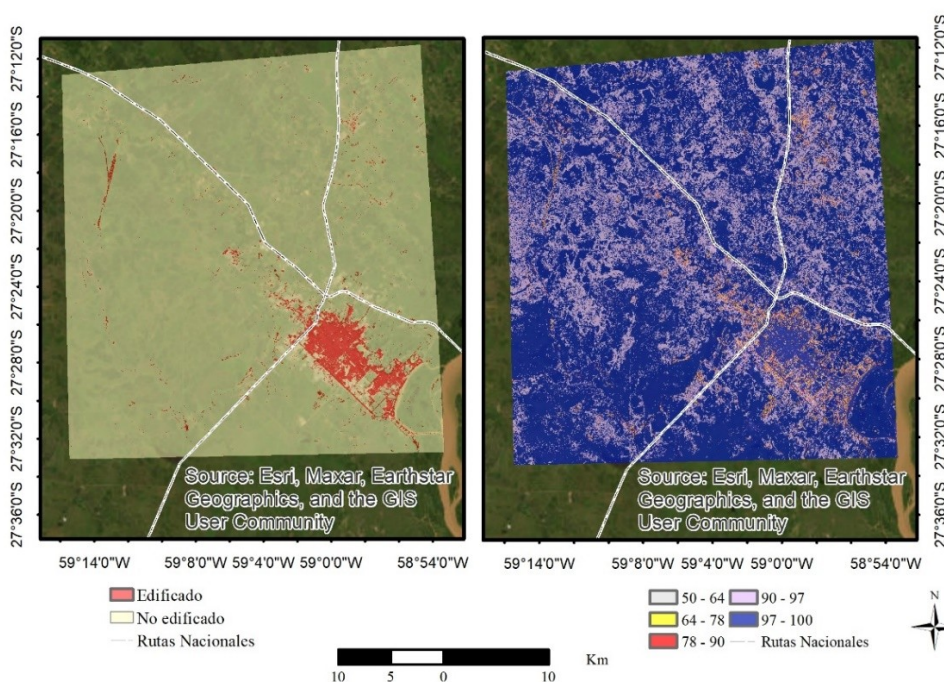
Figura 5. Clasificación supervisada año 1991.
Algoritmo *Random Forest* (izquierda). Mapa de confianza (derecha)



Fuente: Elaboración propia sobre la base de archivos vectoriales del Instituto Geográfico Nacional (2025)

Por su parte, el mapa de confianza de la clasificación generada a partir de SVM (Figura N° 6), permite afirmar que existe mayor certeza en la asignación de los píxeles en cada categoría al mostrarse un nivel de confianza alto, aunque discriminado en dos rangos (entre 90 y 97 y entre 97 y 100). Sin embargo, en los mismos sectores, aunque en menor proporción también se presentan píxeles clasificados con mayor incerteza en este modelo aunque con valores mayores que en el caso anterior (entre 64 y 90).

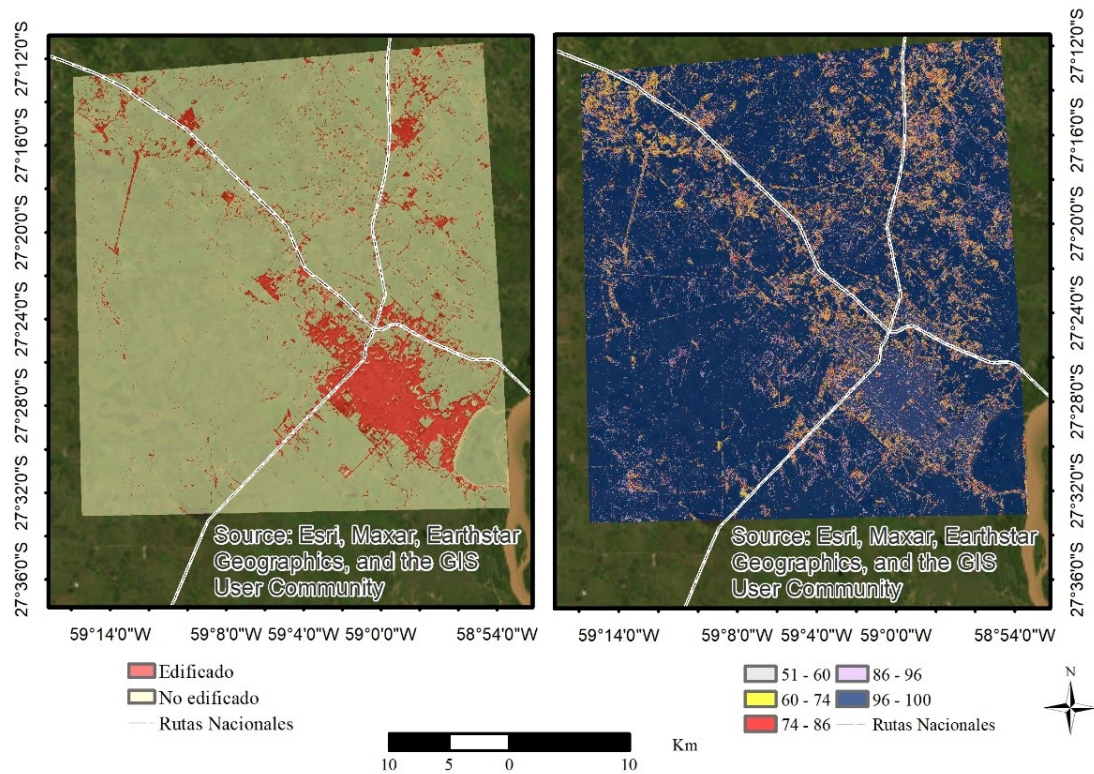
Figura 6. Clasificación supervisada año 1991. Algoritmo SVM (izquierda). Mapa de confianza (derecha)



Fuente: Elaboración propia sobre la base de archivos vectoriales del Instituto Geográfico Nacional (2025)

Una situación similar se observa en las clasificaciones para el año 2022, aunque en este caso, la incertidumbre de las clasificaciones realizadas se distribuye en otros sectores del área de estudio. En este sentido, es posible observar (Figura N°7) que, existe mayor cantidad de píxeles clasificados con un alto nivel de incertidumbre (entre 60 y 86) para la clasificación realizada con RF. Asimismo, es posible apreciar un nivel de confianza alto para la clase no edificada (entre 96 y 100).

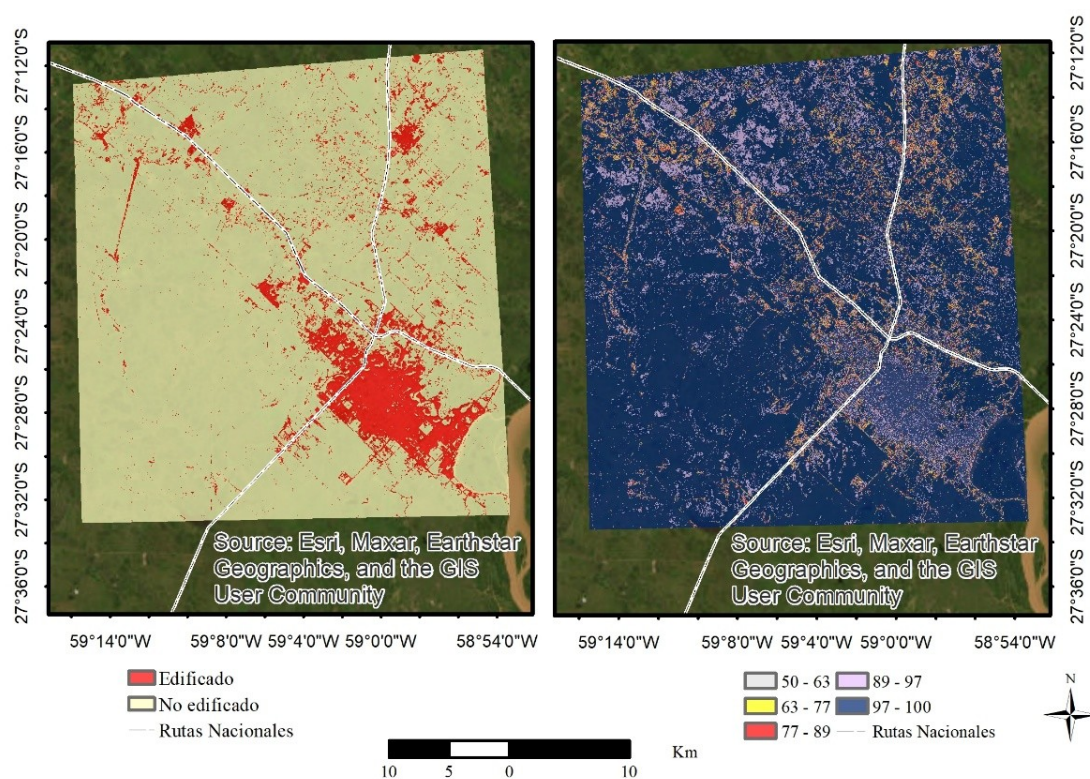
Figura 7. Clasificación supervisada año 2022. Algoritmo *Random Forest* (izquierda). Mapa de confianza (derecha)



Fuente: Elaboración propia sobre la base de archivos vectoriales del Instituto Geográfico Nacional (2025)

De manera similar que para el año 1991, la clasificación realizada para el año 2022 con SVM (Figura N°8) muestra mayor proporción de píxeles con mayor confianza en la clasificación que lo observable para la clasificación realizada con RF (entre 89 y 100); y la distribución de los píxeles menos fiables en la clasificación ocurre en los mismos sectores que aquellos menos fiables clasificados con RF aunque con valores de certeza mayores que en el primer caso (entre 63 y 89).

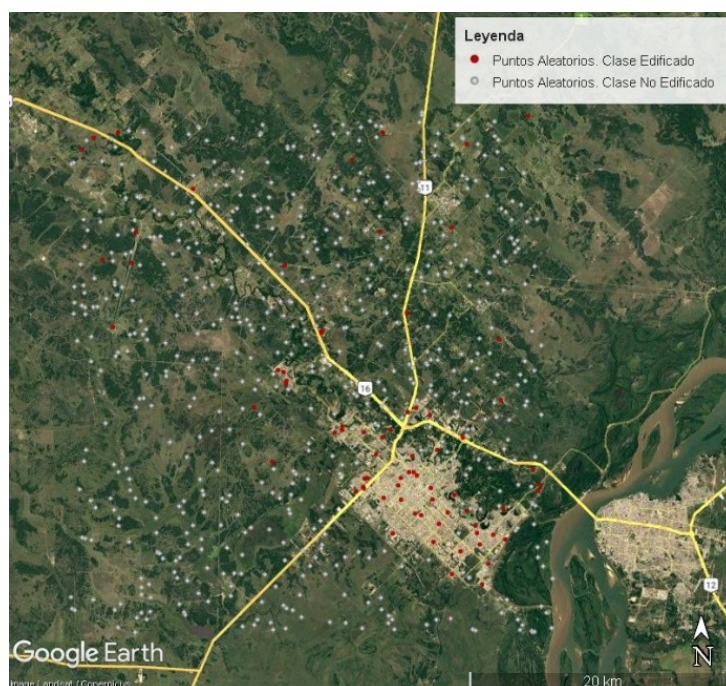
Figura 8. Clasificación supervisada año 2022. Algoritmo SVM (izquierda). Mapa de confianza (derecha)



Fuente: Elaboración propia sobre la base de archivos vectoriales del Instituto Geográfico Nacional (2025)

En otro orden, las Figuras N°9 y N°10 muestran la distribución aleatoria de los puntos generados para los muestreos usados para validar las cartografías generadas con los algoritmos RF y SVM respectivamente para el año 2022; y las tablas N°1 y N°2 resumen la comparación de estos puntos con lo presente en la imagen, en función de la definición de las clases edificado y no edificado expuestas en el apartado metodológico, y presentan los valores de confiabilidad calculados. Al respecto, es posible comentar que con el algoritmo RF se obtuvo una precisión por clase (del usuario) de 86,7% al clasificar la categoría Edificado, mientras que con el algoritmo SVM este valor asciende al 94,7%; un resultado similar ocurre al evaluar la precisión para la clase No Edificado, con el algoritmo RF su valor es de 88,8%, mientras que con SVM este valor aumenta a 95,9%. Con relación a la sensibilidad (exactitud el productor), para el algoritmo RF es posible afirmar que identifica correctamente como Edificado el 49,6% de los píxeles que efectivamente son Edificados y como No Edificado el 98,1% de los píxeles que en la realidad son No Edificado; mientras que para el algoritmo SVM, los valores que reflejan su sensibilidad son del 74,7% y del 99,3% para las clases Edificado y No Edificado respectivamente. Por último, se expone el porcentaje de exactitud global para ambas clasificaciones, siendo 88,57% para la generada con el algoritmo RF y 95,78% para el resultado generado por el algoritmo SVM.

Figura 9. Distribución espacial de puntos generados aleatoriamente para validar la clasificación realizada para el año 2022 con el algoritmo RF



Fuente: Elaboración propia sobre la base de Google Earth

Tabla 1. Matriz de Tabulación cruzada para validar clasificación realizada para el año 2022 con RF

	Edificado	No edificado	Total	Exactitud usuario	Error comisión
Edificado	65	10	75	86.7%	13.3%
No edificado	66	524	590	88.8%	11.2%
Total	131	534	665		
Exactitud productor	49.6%	98.1%			
Error omisión	50.4%	1.9%			

Exactitud global 88.57%

Fuente: Elaboración propia

Figura 10. Distribución espacial de puntos generados aleatoriamente para validar la clasificación realizada para el año 2022 con el algoritmo SVM



Fuente: Elaboración propia sobre la base de Google Earth

Tabla 2. Matriz de Tabulación cruzada para validar clasificación realizada para el año 2022 con RF

	Edificado	No edificado	Total	Exactitud usuario	Error comisión
Edificado	71	4	75	94.7%	5.3%
No edificado	24	565	589	95.9%	4.1%
Total	95	569	664		
Exactitud productor	74.7%	99.3%			
Error omisión	25.3%	0.7%			

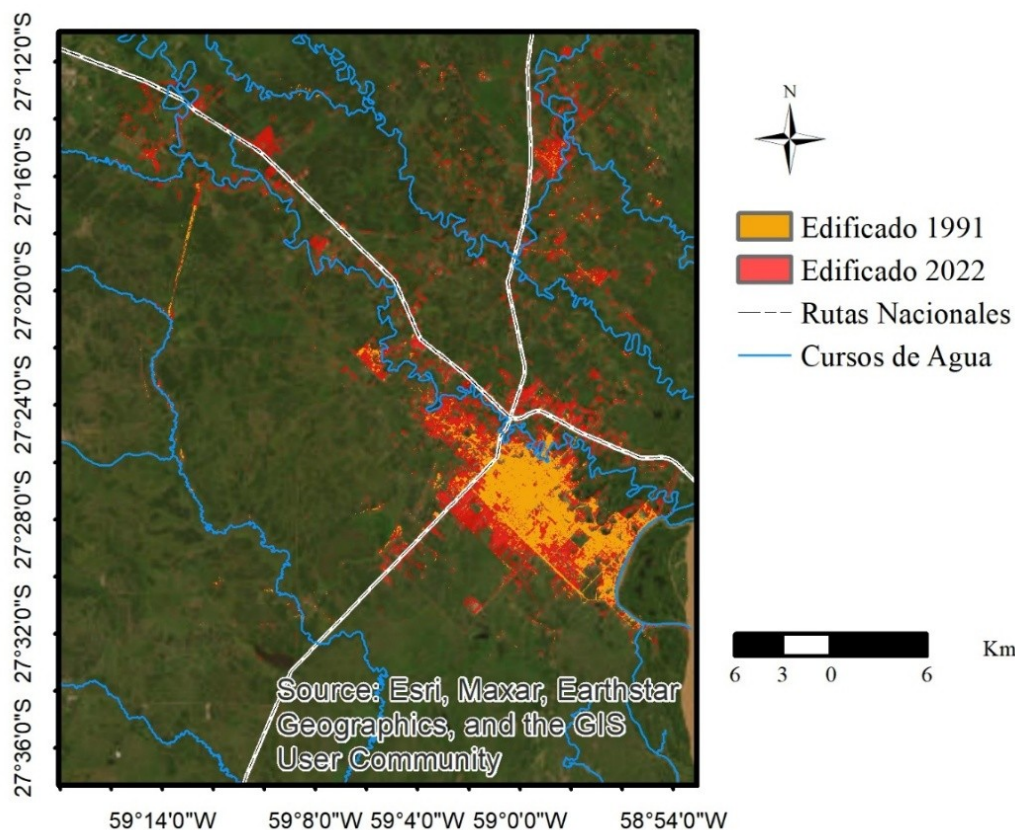
Exactitud global 95.78%

Fuente: Elaboración Propia

Dada la mayor fiabilidad global de la clasificación obtenida por SVM, la comparación de los mapas se realizó a partir del mapa del año 1991 corregido (según la metodología propuesta por Linares, *et al.*, 2022, *up supra*) y el mapa del año 2022, ambos generados a partir de la aplicación de este algoritmo. La figura N°11 permite apreciar una clara expansión del AMGR y del resto de las localidades en los sectores más cercanos a lo clasificado como edificado en el año 1991, en todas las direcciones que le es posible. Incluso, es observable que los cursos de agua existentes en los alrededores no se configuran como una barrera para la expansión de la superficie edificada en el área de estudio. Asimismo, se evidencia un fuerte crecimiento de lo edificado en proximidades de las rutas nacionales, de forma discontinua como por ejemplos urbanizaciones de dimensiones importantes en cercanías de Colonia Popular y Laguna Blanca. Como expresa Zalazar (2024), “la gente busca satisfacer nuevas “necesidades”, como tener una casa en un lugar tranquilo fuera de la ciudad. Esto está cambiando el paisaje, convirtiendo campos en barrios nuevos, cerrados y privados. Estos cambios responden a las acciones de diferentes actores sociales que van

moldeando cómo se usa la tierra y cómo crece la ciudad”. Por último, también es evidente la consolidación que experimentan tanto el AMGR como el resto de las localidades al completarse espacios al interior de lo edificado para el año 1991, en el año 2022.

Figura 11. Expansión de la superficie edificada en el AMGR entre el año 1991 y 2022. Obtenido a partir de clasificación con SVM.



Fuente: Elaboración Propia.

Finalmente, en términos porcentuales, el aumento en 31 años de la superficie edificada en el área de estudio fue de más 239% (11.186,81 Ha en términos absolutos), pasando de representar 4.665,87 Ha en 1991 a 15.852,68 Ha en 2022.

Discusión de resultados

En primer lugar, a partir de los materiales y metodología abordados en el trabajo fue posible cuantificar la superficie con cobertura edificada en los años 1991 y 2022 en el área de estudio y describir su distribución espacial, por lo que se afirma la concreción del objetivo propuesto en esta contribución. Para ello, se usaron los algoritmos RF y SVM para clasificar imágenes satelitales y evaluar la confiabilidad de ambos resultados en el caso de la clasificación obtenida para el año 2022, concluyendo que SVM supera a RF en precisión global, lo que coincide con hallazgos previos (Dabija et al., 2021; Chucos Baquerizo y Vega Ventocilla, 2022). Sin embargo, persisten desafíos en la clasificación de zonas urbanas, donde la heterogeneidad espectral (techos, asfalto, vegetación urbana, etc.) reduce la precisión, un problema común en estudios similares (Abdi, 2019). Al respecto, se presenta un error de omisión del 25,3%, es decir que un 25,3% de píxeles edificados no fueron clasificados como tales, lo que podría deberse a un posible desbalance en el conjunto de datos de entrenamiento dado el mayor tamaño de la superficie no edificada, y que afecta los resultados en contra de la clase edificada.

Otros resultados muestran un incremento significativo de la superficie edificada entre los momentos estudiados (más del 200%), con patrones de expansión tanto en la periferia de las áreas ya edificadas en 1991, como en sectores intermedios entre localidades de forma dispersa y discontinua destacándose la

importancia de la cercanía a las rutas nacionales, lo que podría generar un consumo ineficiente del suelo, dependencia del automóvil, mayores costos en infraestructura y provisión de servicios básicos, además de plantear riesgos socioambientales.

Entre las limitaciones presentes en la investigación es posible mencionar el uso de imágenes Landsat (30m) dada la inexistencia de imágenes de más fina resolución disponibles de forma gratuita para el año 1991, lo que condiciona la detección de estructuras urbanas pequeñas o dispersas como señalan Chucos Baquerizo y Vega Ventocilla (2022). Sin embargo, se destaca que la resolución utilizada permite la detección de patrones de expansión urbana. Teniendo en cuenta esto, sería interesante incorporar en futuras investigaciones datos de imágenes de mayor resolución espacial como Sentinel-2 para mejorar la precisión.

Otra limitación existente se vincula con la imposibilidad de validar los resultados correspondientes al año 1991 dada la falta de insumos que pudieran considerarse como “verdad de terreno” para ese año. Si bien la corrección aplicada a las clasificaciones basada en la metodología propuesta por Linares et al. (2022) contribuye a disminuir los errores, la misma lo hace bajo el supuesto de que la superficie edificada no desaparece o no cambia con el avance del tiempo. No obstante, permanece cierto grado de incertidumbre al no contar con los datos de referencia iniciales.

Por último, se debe mencionar que con relación al posible desbalance entre las clases, el mismo pudo haber afectado el rendimiento de los algoritmos, por lo que en futuros trabajos podría ser útil considerar métricas como F1-Score o realizar un muestreo estratificado para determinar con mayor precisión el tamaño de las muestras por clase (aunque se deberá considerar información de referencia que permita conocer el tamaño, al menos de forma aproximada, del estrato).

Es sabido, que el crecimiento de las ciudades, particularmente las de tamaño intermedio como se presenta en este estudio, es un fenómeno global que responde a acciones y decisiones de diferentes actores. La falta de ordenamiento y planificación integral del territorio abre camino a la generación de una ciudad poco sustentable, por lo tanto, la información generada en esta contribución podría resultar útil para orientar decisiones futuras con el objetivo de alcanzar un equilibrio entre el desarrollo urbano y el cuidado del ambiente. El estudio proporciona evidencia que podría resultar valiosa para la planificación del territorio, destacando la importancia de monitorear la expansión de una ciudad con insumos y herramientas accesibles como los son las imágenes satelitales y los sistemas de información Geográfica.

Conclusiones

A partir del análisis realizado es posible concluir que las clasificaciones obtenidas por ambos métodos (RF o SVM) configurados en sus parámetros por defecto según el complemento Dzetsaka tuvieron resultados aceptables en términos estadísticos, superando el 88% la fiabilidad global en la etapa de validación. Para nuestro caso, en estos términos, la clasificación con SVM fue superadora al obtenerse una fiabilidad global del 95,78% al ser comparada con una imagen de más fina resolución espacial como verdad de terreno, proceso que requirió la intervención de un intérprete visual para determinar la correspondencia entre clases. A partir de ello, los resultados de este estudio muestran por un lado la distribución de la clase edificada en el año 2022 en proximidades al área edificada del año 1991, dando continuidad de esta forma a la expansión de la ciudad. Por otro lado, se detectaron nuevos espacios clasificados como edificados, de forma discontinua en proximidades a las Rutas Nacionales, en sectores intermedios entre las localidades del área de estudio, lo que podría promover un consumo ineficiente del suelo, la degradación de recursos naturales, un modelo de ciudad basado en el uso del automóvil, mayores costos de inversión de recursos públicos, entre otros factores que van en contra de crecimiento de una ciudad de forma sustentable.

Por último, dado que las ciudades se configuran como espacios complejos donde convergen factores ambientales, sociales, económicos, culturales, etc.; e intereses de diferentes actores, se destaca la importancia de contar con información fiable y actualizada en torno al crecimiento de las superficies construidas de una ciudad, tanto en términos de cantidad como de localización, que sirva de herramienta para la toma de decisiones en torno a la gestión y planificación eficaz del territorio.

Referencias bibliográficas

- Abdi, A. (2019). Rendimiento de la clasificación de la cobertura y el uso del suelo de algoritmos de aprendizaje automático en un paisaje boreal utilizando datos de Sentinel-2. *GISciencia y teledetección*, 57(1), 1-20. <https://doi.org/10.1080/15481603.2019.1650447>
- Bellet, C. y Llop, J. (2004). Miradas a otros espacios urbanos: las ciudades intermedias. *Geo Crítica Scripta Nova* (en línea), 8(165). <https://www.ub.edu/geocrit/sn/sn-165.htm>
- Beron de la Puente, F., Montico, A., Revollo, N., Gil, V. y Zapperi, P. (12 al 16 de agosto de 2024). *Clasificación de uso y cobertura del suelo basada en modelos de Aprendizaje de Máquina Supervisados* [Comunicación]. Memorias de las 53 JAIIO - CAI - IS.
- Biau, G. (2012). Analysis of a Random Forests Model. *Journal of Machine Learning Research*, 13, 1063-1095.
- Breiman, L. (2001). Random Forest. *Machine Learning*, 45, 5-32.
- CAF Banco de Desarrollo de América Latina (2017). Crecimiento urbano y acceso a oportunidades: un desafío para América Latina. *Reporte de Economía y Desarrollo*. <https://www.caf.com>
- Campo León, E. (s.f). *Introducción a las máquinas de vector soporte (SVM) en aprendizaje supervisado*. [Trabajo de Fin de Grado en Matemáticas]. Facultad de Ciencias. Universidad de Zaragoza. <https://zaguan.unizar.es/record/59156/files/TAZ-TFG-2016-2057.pdf>
- Chucos Baquerizo, N. y Vega Ventocilla, E. (2022). Evaluación de algoritmos de machine learning en la clasificación de imágenes satelitales multispectrales, caso: Amazonia Peruana. *Ciencia Latina Revista Científica Multidisciplinar*, 6(1). https://doi.org/10.37811/cl_rcm.v6i1.1843 p 4963
- Clark Labs. (2020). *Manual de TerrSet 2020: Monitoreo y modelado geoespacial*. Clark University.
- Dabija, A., Kluczek, M., Zagajewski, B., Raczko, E., Kycko, M., Al-Sulttani, A. H., Tardà, A., Pineda, L. y Corbera, J. (2021). Comparison of Support Vector Machines and Random Forests for Corine Land Cover Mapping. *Remote Sens*, 13, 777. <https://doi.org/10.3390/rs13040777>
- Davis, D. (6 de agosto de 2020). *Random Forest Classifier Tutorial: How to Use Tree-Based Algorithms for Machine Learning*. <https://www.freecodecamp.org/news/how-to-use-the-tree-based-algorithm-for-machine-learning/>
- Di Gregorio, A. y Jansen, L. J. M. (2005). *Land Cover Classification System (LCCS): Classification Concepts and User Manual* (8va. Ed.). Food and Agriculture Organization of the United Nations.
- Gil Leiva, I., Díaz Ortuño, P. y Rodríguez Muñoz, J. (10 al 11 de Julio de 2019). *Técnicas y usos en la clasificación automática de imágenes* [Comunicación]. 14º Congreso ISKO España (4º ISKO España-Portugal) (Barcelona).
- Jehad, A., Rehanullah, K., Nasir, A. y Imran, M. (2012). Random Forest and Decision Trees. *IJCSI International Journal of Computer Science Issues*, 9(5), 3.
- Lanfranchi, G., Duarte, J. I. y Granero Realini, G. (2018). *La expansión de los Grandes Aglomerados Urbanos argentinos*. CIPPEC (Documento de Políticas Públicas/Recomendación N°197).
- Lanfranchi, G. (29 de Mayo de 2019). *Hay que volver a hacer ciudad como nuestros abuelos, con más equidad y mejor desarrollo social*. (A. Torres, Entrevistador). <https://www.lanacion.com.ar/buenos-aires/nombre-apellidoelis-am-siquat-amsedfg-commy-ullaorpero-enis-enisi-orperae-strud-eu-feuguercin-utat-velit-venim-9700gabriel-lanfranchicolumna-ipis-dolore-moluptatx-it-nulputet-vel-et-re-jhasgd-4600-nid2251780/>

- Linares, S., Iturrioz, I. y Di Nucci, J. (2022). La Expansión Urbana de la ciudad de Tandil 1991-2020 mediante el uso de Landscape Expansion Index (LEI). *Pleamar. Revista del Departamento de Geografía*, 2(2).
- Marlenko, N. (2011). Aplicaciones en el uso de la tierra. En S. M. Navone (Coord.), *Sensores Remotos aplicados al estudio de los Recursos Naturales* (pp. 119-143). Editorial Facultad de Agronomía.
- Meyer, W. B. y Turner, B. (1994). Changes in land use and land cover: a global perspective. In *Changes in land use and land cover: a global perspective*. Cambridge University Press.
- Ministerio de Planificación Federal, Inversión Pública y Servicios (2011). *Plan Estratégico Territorial Avance II: Argentina Urbana*. Ministerio de Planificación Federal, Inversión Pública y Servicios.
- Nery de Figueiredo, T., Sadler, R., Solis-Aulestia, M., White, B., Polyakov, M. y Chalak Haghighi, M. (2016). Comparing supervised algorithms in Land Use and Land Cover classification of a Landsat time-series. In J. Wu, & Y. Jin (Eds.), *International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, Instituto de Ingenieros Eléctricos y Electrónicos. <https://doi.org/10.1109/IGARSS.2016.7730346>
- Parr, D. (1 de marzo de 2024). *Machine Learning Classification in QGIS*. <https://daveparr.com/2024/03/01/machine-learning-classification-in-qgis/>
- Scornik, M. y Godoy, S. (2008). Amenazas y Vulnerabilidad. Lineamientos para el ordenamiento territorial de Barranqueras. En N. Bono, O. Bragos y C. Scornik (Eds.), *Riberas bajo amenaza. Desafíos de la gestión territorial en localidades metropolitanas* (pp. 11-25). UNR Editora.
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J. y Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241, 111716. <https://doi.org/10.1016/j.rse.2020.111716>.
- Zalazar, M. (29 de agosto de 2024). *Expansión Urbana: un estudio pone bajo la lupa el crecimiento del Gran Resistencia*. UNNE Medios. <https://medios.unne.edu.ar/2024/08/29/expansion-urbana-un-estudio-pone-bajo-la-lupa-el-crecimiento-del-gran-resistencia/>