



CUANTIFICACIÓN DE CAMBIOS DE COBERTURA DE SUELO Y VEGETACIÓN CON BASE EN APRENDIZAJE AUTOMÁTICO E IMÁGENES LANDSAT

José Rodríguez Rosales¹, Juan Manuel González Camacho^{1*}

¹Colegio de Postgraduados Campus Montecillo. Posgrado en Socioeconomía, Estadística e Informática-Cómputo Aplicado. Carretera México-Texcoco km 36.5, Montecillo, Texcoco, Estado de México, México. C. P. 56264.

jmgc@colpos.mx (*Autor para correspondencia)

Resumen

La cuantificación de cambios de cobertura de suelo y vegetación se realiza en grandes extensiones a partir de imágenes de satélite. Esta cuantificación facilita la planeación y manejo sostenible de los recursos naturales. En esta investigación se clasificaron cuatro clases de cobertura: vegetación (V), agrícola (A), pastizal (P) y urbano-rural (U) en la región Huehuetla, de la sierra norte de Puebla con base en el clasificador bosque aleatorio (RF, random forest) e imágenes de satélite en 2002 y 2021. El clasificador se implementó en lenguaje Python y se evaluó en predicción. RF obtuvo buena precisión en predicción para clasificar los cuatro tipos de cobertura en ambas fechas. RF obtuvo una precisión global de clasificación de 92.5 %, en 2022 y de 92.3 % en 2021. Asimismo, RF obtuvo la mejor precisión global para clasificar la vegetación con puntajes F1 de 100 % y 98.2 % en 2002 y 2021. Durante el periodo analizado en un área de 39.5 km², la superficie agrícola disminuyó 6.6 %; la de pastizal disminuyó 46.4 %; la de vegetación aumentó 32.5 % y la urbano-rural aumentó 274.5 % con respecto a 2002. Las imágenes Landsat se transformaron al espacio de color RGB y se calcularon índices de vegetación para definir las variables de entrada de RF. La metodología propuesta en este estudio es una alternativa viable y de bajo costo para realizar la determinación de cambios de cobertura del suelo y vegetación en cuencas pequeñas.

Palabras claves: Aprendizaje supervisado, análisis de ensamble, árboles de decisión, clasificación multiclase.



Introducción

El manejo integral de los recursos representa uno de los problemas del desarrollo de las comunidades de la sierra norte del estado de Puebla; los cambios de uso de suelo y vegetación reflejan la transformación del entorno como resultado de la interacción de factores naturales y humanos (Guevara y Montalvo, 2015). El cultivo de café ha sido la principal fuente de ingresos de los pequeños productores, la crisis de este cultivo en los 90's favoreció el desarrollo de la ganadería y la apertura de potreros en áreas forestales y de cultivos perennes (Martínez et al., 2017). El uso de suelo y vegetación en la zona de Huehuetla, Puebla representa una superficie de 60 km²; dónde la agricultura cubre 52.6 %, zona urbana 1.6 %, pastizal 33.9 %, selva 11.5 % y bosque 0.38 % (INEGI, 2010).

Baeza et al. (2014) en Uruguay clasificaron cuatro coberturas (forrajes perennes, bosques, cultivos de verano y doble cultivo) con base en árboles de decisión e imágenes Landsat; estos autores reportaron una precisión global de clasificación de 89.6 %. Suárez et al. (2017) aplicaron los algoritmos de aprendizaje automático máquinas de soporte vectorial (SVM, support vector machine) y bosque aleatorio (RF, random forest) para estimar cambios de cobertura del suelo a partir de imágenes digitales, con una precisión global de clasificación de 86.5 % a 95.1 %.

Chucos y Vega (2022) realizaron un estudio en la Amazonia Peruana para evaluar algoritmos de aprendizaje automático a partir imágenes de satélite multiespectrales Landsat 8 y Sentinel 2; estos autores utilizaron cuatro clases: cuerpos de agua, cultivos, urbano y bosques; los clasificadores SVM y Bayes-naive obtuvieron una precisión global de clasificación de 90.9 % y los árboles de decisión 86.4 %.

El objetivo de este estudio fue evaluar los cambios de uso de suelo y vegetación (vegetación, agrícola, pastizal y urbano-rural) en la región de Huehuetla, Sierra norte de Puebla con base en el clasificador bosque aleatorio entrenado con imágenes de satélite Landsat 7 (de enero 2002) y Landsat 8 (de diciembre 2021).

Materiales y Métodos

Colecta de imágenes

Las imágenes Landsat 7 y Landsat 8 se obtuvieron del sitio web del departamento geológico de los Estados Unidos, (USGS, united state geological survey); <http://glovis.usgs.gov/>. La imagen Landsat 7 del 15/enero/2002 tiene una resolución por píxel, de 30 m × 30 m y la banda pancromática tiene una resolución de 15 m × 15 m (USGS, 2019). La imagen Landsat 8 del 29/diciembre/2021 tiene una resolución por píxel de 30 m × 30 m y la banda pancromática una resolución de 15 m × 15 m.



Las imágenes de la zona de estudio se seleccionaron en condiciones de no nubosidad. Los archivos de municipios se descargaron del portal de geo información del sistema nacional de información sobre biodiversidad (CONABIO, 2021).

Recursos computacionales

La creación del conjunto de datos de entrada con base en color y sus clases objetivo asociadas (tipos de cobertura de suelo) se realizó con el lenguaje de programación Python 3.10 y la biblioteca de programas Scikit-learn. El procesamiento y análisis de datos se realizaron en un pc Lenovo Ideapad 320 con sistema operativo Windows 11.

Preprocesamiento de imágenes

El procesamiento de las imágenes Landsat 7 y Landsat 8 se realizó con el programa QGIS 3.24.1. En este estudio se aplicaron las siguientes etapas de preprocesamiento de imágenes:

Una corrección geométrica se efectuó a cada imagen Landsat para asignar a cada pixel las coordenadas geográficas correspondientes a un sistema de georreferenciación, cada pixel o grupo de pixeles se le asoció un punto geográfico determinado. Las imágenes que proporciona el USGS se georreferenciaron con el software QGIS.

Una corrección atmosférica se realizó a cada imagen para eliminar el efecto de dispersión y absorción de la atmósfera y obtener la reflectancia de la superficie caracterizada con el método dark object subtraction de QGIS. Esto permitió mejorar la calidad visual de la imagen.

Pansharpening se aplicó para combinar la información de las bandas multiespectrales R (Red), G (Green) y B (Blue) con resolución de 30 m × 30 m de las imágenes Landsat 7 (bandas 1, 2, y 3), Landsat 8 (bandas 2, 3, y 4) y la banda pancromática 8 con resolución espacial 15 m × 15 m y se obtuvo una imagen multiespectral con la resolución de la banda pancromática.

La equalización del histograma se aplicó para obtener una distribución uniforme de los pixeles en la imagen y mejorar el contraste de las imágenes.

Extracción de muestras de imágenes

Las muestras de imágenes se obtuvieron para cada clase objetivo por medio del algoritmo de selección por región. Este método consiste en extraer los tres canales del espacio de color RGB de cada píxel y sus coordenadas de ubicación (x, y) en la imagen. El método consiste en definir un píxel semilla y por similitud compararlo con vecinos cercanos (Valdés *et al.*, 2015).



Con este procedimiento se crearon cuatro conjuntos de datos, uno por clase objetivo (agrícola, pastizal, vegetación y urbano-rural). El conjunto total de muestras de pixeles extraídos se procesó para eliminar observaciones repetidas dentro de clases y entre clases. El conjunto de datos de entrada consistió en tres canales de color R, G, B y su clase objetivo asociada.

Índices de vegetación

Los siete índices utilizados se calcularon con base en el espacio de color RGB de las imágenes Landsat (Cuadro 1).

Cuadro 1. Índices de vegetación que se calcularon a partir del espacio de color RGB.

Índice de vegetación	Fórmula
Exceso de verdes	$exG = 2G - R - B$
Exceso de rojos	$exR = 1.4R - G$
Diferencia de exceso de verde y exceso de rojo	$exGR = exG - exR$
Extracción del índice de vegetación	$CIVE = 0.44R - 0.81G + 0.39B + 18.79$
Vegetativo	$veG = \frac{G}{R^a B^{1-a}}$ con $a = 0.667$
Diferencia verde-rojo normalizada	$nGRdi = \frac{G - R}{G + R}$
Diferencia normalizada	$NDI = \left(\left(\frac{G - R}{G + R} \right) + 1 \right) * 128$

El conjunto total de características de entrada consistió en tres canales de color R, G, y B y siete índices de vegetación. El escenario de entrada E2002 consistió de diez características y 7370 muestras de datos distribuidas en clase A: 1 773, clase P: 1 653, clase V: 3 093 y clase U: 851 muestras; El escenario E2021 consistió de diez características y 10 143 muestras de datos distribuidas en clase A: 2 672, clase P: 2 246, clase V: 3 507 y clase U: 1 718 muestras.

Clasificador bosque aleatorio

El clasificador bosque aleatorio (RF, random forest) es un modelo de ensamble que consiste en calcular la respuesta promedio de múltiples árboles de decisión para reducir

la varianza de los árboles individuales para obtener un mejor desempeño en predicción y evitar el sobreajuste del modelo (Breiman, 2001). El algoritmo selecciona n muestras con remplazo del conjunto de datos de entrenamiento, luego hace crecer un árbol de decisión con un subconjunto de características que se seleccionan al azar. Después de entrenar los árboles de decisión, la predicción de una clase objetivo se realiza con el voto mayoritario. El modelo busca minimizar una medida de impureza, tales como el criterio de impureza de Gini.

El criterio de impureza de Gini $I_G(t)$, se define como:

$$I_G(t) = 1 - \sum_{i=1}^C p(i|t)^2 \quad (1)$$

donde $I_G(t)$ alcanza su valor máximo ($I_G(t) = 1$) cuando las muestras tienen una distribución uniforme dentro de las clases (Raschka y Mirjalili, 2019).

El mejor modelo RF se obtiene por medio de una búsqueda de hiperparámetros por retícula y un procedimiento de validación cruzada. Los hiperparámetros más comunes de RF son: número de estimadores o árboles de decisión (ne) del modelo, criterio de división de nodos (cr), profundidad máxima de los árboles de decisión (mp) y número máximo de características de entrada (mc) que se eligen al azar para dividir los nodos de un árbol de decisión.

Métricas de desempeño

Para evaluar el desempeño del clasificador se utilizaron las métricas precisión global de clasificación correcta (ACC , accuracy), precisión (P), sensibilidad (S), puntaje $F1$. Estas métricas se derivan de una matriz de confusión. Esta matriz contiene cuatro valores: verdaderos positivos (VP) son las respuestas verdaderas predichas correctamente; verdaderos negativos (VN), son las respuestas negativas predichas correctamente; falsos positivos (FP), son las respuestas negativas predichas como positivas, y falsos negativos (FN) son respuestas positivas predichas como negativas (Lipton *et al.*, 2014).

ACC se define como:

$$ACC = \frac{VP+VN}{VP+VN+FP+FN} \quad (2)$$

P se calcula por:

$$P = \frac{VP}{VP+FP} \quad (3)$$

S se calcula por:

$$S = \frac{VP}{VP + FN} \quad (4)$$

El puntaje $F1$ es la media armónica de P y S y se define como:

$$F1 = \frac{2P \times S}{P+S} \quad (5)$$

ACC , P , S y $F1$ toman valores entre 0 y 1. Un valor cercano a 0 indica un bajo desempeño del clasificador, mientras que un valor cercano a 1, indica un alto desempeño.

Para el diagnóstico de problemas de sesgo y varianza del clasificador se utilizaron curvas de aprendizaje y validación. Estas curvas se representan en una gráfica de ACC del entrenamiento y validación *versus* el número de muestras de los datos de entrada. Con esta gráfica se detecta si el modelo tiene alta varianza o alto sesgo, y si la recopilación de más muestras de datos ayuda a reducir estos problemas.

Entrenamiento del clasificador bosque aleatorio

Selección de características e hiperparámetros óptimos

El clasificador RF se entrenó con el conjunto total de diez características, luego se seleccionaron subconjuntos de características con base en la importancia relativa de las características que se obtuvo con el modelo bosque aleatorio (Raschka y Mirjalili, 2019).

La selección de hiperparámetros óptimos del clasificador RF se realizó por medio de una búsqueda por retícula y un procedimiento de validación cruzada (VC) con $k = 10$ particiones del subconjunto de entrenamiento. El entrenamiento y validación utilizó 90 % del conjunto de datos total. Para evaluar el desempeño del clasificador en el entrenamiento se generaron curvas de validación que muestran la variación de la métrica ACC de los conjuntos de entrenamiento y de validación, como una función del número de muestras en los conjuntos de datos de entrada.

La configuración final del modelo se obtuvo con la partición de los datos que generan el máximo valor de ACC después de probar las combinaciones posibles de los hiperparámetros (Cuadro 2).

Cuadro 2. Valores de los hiperparámetros que se combinaron en la búsqueda por retícula.

Modelo	Hiperparámetros	Intervalo
RF	ne	50, 100, 120, 150 y 200
	cr	Entropía o Gini
	mc	Raíz cuadrada o Log_2
	mp	8, 10, 11, 12, 13, 14, y 15

RF: modelo bosque aleatorio, ne : número de estimadores, cr : criterio de división, mc : número máximo de características, mp : profundidad máxima.

Evaluación del desempeño en predicción

La prueba de RF en predicción se realizó con el juego de hiperparámetros óptimos, con el conjunto total datos y una validación cruzada con $k = 10$ particiones. Además de ACC se obtuvieron todas métricas de desempeño promedio de RF para cada clase objetivo.

Cálculo de cambios de cobertura de suelo

La cuantificación de superficies y tasas de los cambios de cobertura de suelo durante el periodo (2002-2021) consistió en tres etapas: a) detección e identificación cartográfica digital del cambio; b) análisis de patrones de cambio, y c) análisis de las causas de los cambios de cobertura (Guevara y Montalvo, 2014). El análisis espacial se utilizó para evaluar los cambios de cobertura, la predicción con el modelo RF óptimo se realizó para cada pixel de cada imagen de satélite que delimita la zona de estudio, en los dos momentos de 2002 y 2021. La clase objetivo agrícola incluyó la superficie correspondiente a suelo desnudo en preparación; la clase pastizales incluyó la vegetación primaria y potreros; la clase vegetación incluyó bosques, vegetación secundaria arbustiva y cultivos bajo sombra; y la clase urbano-rural incluyó la superficie ocupada por construcciones e infraestructura de carreteras, caminos pavimentados, casas y calles de las principales agrupaciones poblacionales.

Resultados y Discusión

Imágenes Landsat preprocesadas

Las imágenes Landsat recortadas al tamaño del municipio de Huehuetla y procesadas se exportaron a un formato jpg para la extracción de características de color (Figura 1).

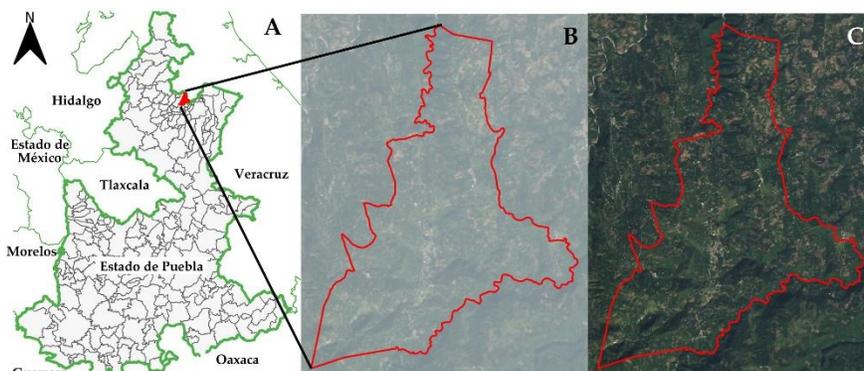


Figura 1. A: Ubicación del municipio de Huehuetla, en la sierra norte del estado de Puebla. Combinación de bandas RGB de la imagen de satélite Landsat 8 en 2021. B: imagen Landsat 8 sin correcciones, resolución de 30 × 30 m; C: imagen Landsat 8 con correcciones, resolución de 15 × 15 m.

Características seleccionadas

Con base en los resultados de la evaluación de la importancia de las características con RF, se seleccionaron las cinco características de mayor importancia para el escenario de entrada E2002; las cuales fueron R, G, B, ExG y VEG. Para el escenario de entrada E2021, se seleccionaron las características R, G, B, ExR y NGRDI (figura 2).

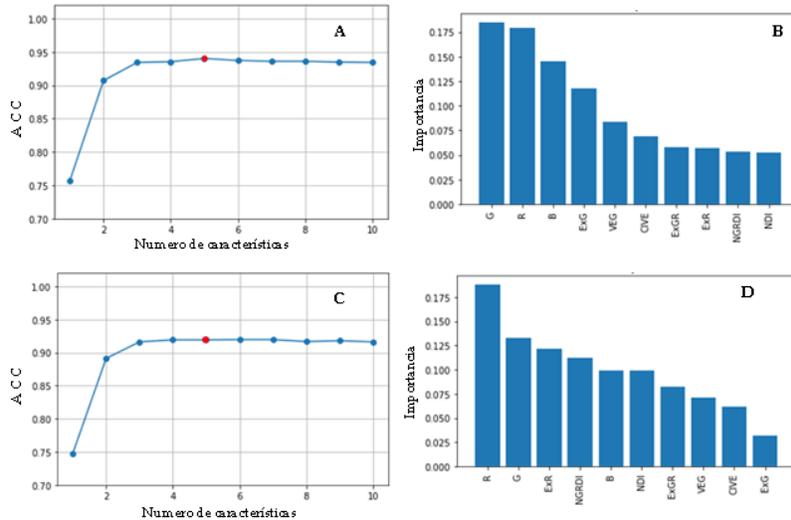


Figura 2. Número óptimo e importancia relativa de las características de entrada. A. curva de desempeño E2002; B. importancia relativa, escenario E2002; C. curva de desempeño E2021; D. importancia relativa, escenario E2021; ACC: precisión global de clasificación correcta.

Hiperparámetros óptimos y curvas de validación

Los hiperparámetros óptimos de RF seleccionados para el escenario E2002 fueron $ne = 120$, $cr =$ entropía, $mc =$ raíz cuadrada y $mp = 8$. Para E2021 fueron $ne = 200$, $cr =$ Gini, $mc = \log_2$ y $mp = 10$.

La curva de validación de RF para E2002 mostró un mejor ajuste de los datos con 6 000 muestras, donde las curvas de entrenamiento y validación convergen (figura 3-A). Para E2021 el modelo RF se ajustó mejor a partir de 6 500 muestras, donde se aprecia que las curvas entrenamiento y validación convergen y se estabilizan (figura 3-B).

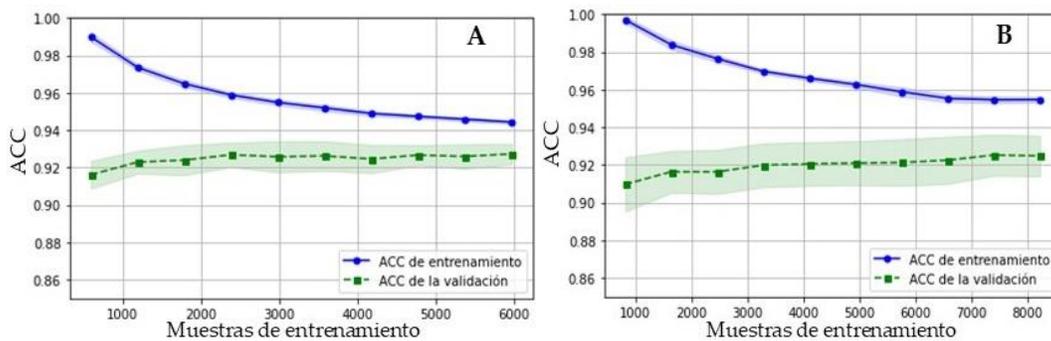


Figura 3. Curvas de aprendizaje del clasificador RF para predecir las clases objetivo A: agrícola, P: pastizal, V: vegetación y U: urbano-rural. A: escenario de entrada E2002; B: escenario de entrada E2021; ACC: precisión global de clasificación correcta.

Evaluación en predicción del clasificador bosque aleatorio

Para el escenario E2002, RF obtuvo una *ACC* en predicción de 92.5 % y para E2021 una *ACC* de 92.3 %; el desempeño de RF por clase objetivo mostró que la clase V fue la mejor clasificada en ambos escenarios de entrada y la clase U la peor clasificada, en términos de la métrica puntaje *F1* (cuadro 3).

Cuadro 3. Métricas de desempeño promedio en predicción del clasificador bosque aleatorio (RF) para los escenarios de entrada E2021 y E2022.

RF	ACC	Clase	P	S	F1
E2002	0.925 +/- 0.004	A	0.898	0.898	0.898
		P	0.906	0.933	0.919
		V	1.000	1.000	1.000
		U	0.738	0.694	0.715
E2021	0.924 +/- 0.005	A	0.906	0.849	0.877
		P	0.969	0.960	0.965
		V	0.980	0.983	0.981
		U	0.773	0.869	0.818

A: Agrícola, P: Pastizal, V: Vegetación, U: Urbano-rural, F1: Puntaje F1; P: Precisión; S: Sensibilidad.

Matrices de confusión

El desempeño en predicción de RF para clasificar las cuatro clases objetivo se analizó con las matrices de confusión normalizadas, escenarios E2002 y E2021. En ambos escenarios la clase V fue la mejor clasificada; sin embargo, RF en E2002 presentó problemas de confusión de la clase U con las coberturas agrícola y pastizal (clases A y P) y en E2021 presentó problemas con la clase A (Figura 4).

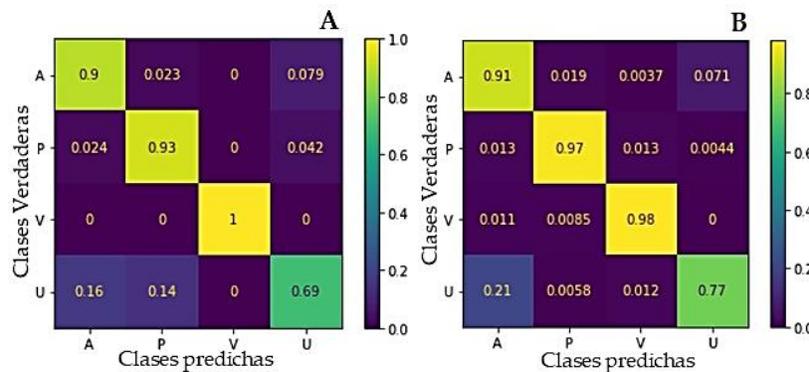


Figura 4. Matriz de confusión normalizada del clasificador bosque aleatorio para predecir las clases objetivo Agrícola (A), Pastizal (P), Vegetación (V) y Urbano-rural (U). A) Escenario de entrada E2002; B) Escenario de entrada E2021.

Cuantificación de superficies y tasas de cambio de coberturas de suelo

Con ambos modelos ajustados se realizó una predicción utilizando los pixeles que conforman el área correspondiente al municipio de Huehuetla, Puebla. Los resultados de ambas predicciones para los años 2002 y 2021 se muestran en la Figura 5. En el que se aprecia que la clase V no ha sufrido cambios tan drásticos a diferencia de la clase P, que muestra una disminución en superficie al igual que la clase A, mientras que la clase U visiblemente su superficie es mayor para el año 2021.

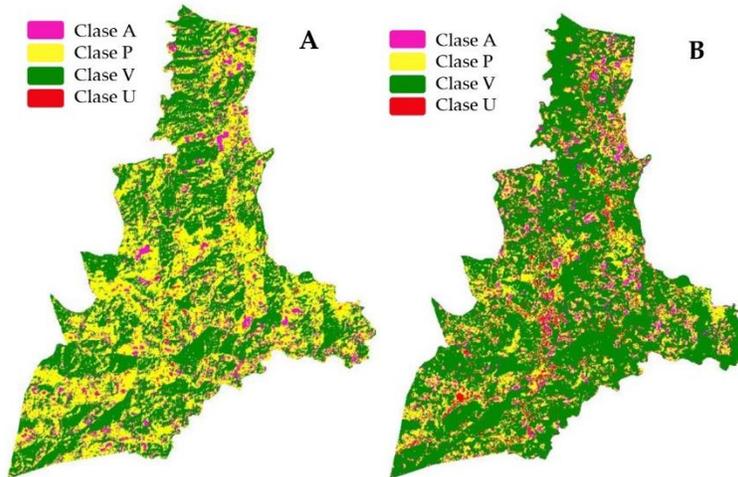


Figura 5. Predicciones de coberturas realizada con el clasificador bosque aleatorio: A: E2002 y B: E2021 de las clases objetivo (A: Agrícola, P: Pastizal, V: Vegetación y U: Urbano-rural).

En general, dos tipos de cobertura han aumentado en menor o en gran medida su superficie en el periodo analizado. La clase V incrementó 32.5 % de su superficie y la clase U incrementó 274.5 %; esto en términos de superficie representa un incremento de 617.3 ha en la cobertura de vegetación y un incremento de 186.3 ha en la cobertura urbano-rural. La cobertura agrícola (clase A) tuvo un decremento de superficie de 6.6 %; (19.3 ha), de forma similar la cobertura de pastizales (clase P) tuvo un decremento de 46.4 % (784.1 ha) (cuadro 4).

Cuadro 4. Cuantificación de los cambios de cobertura de suelo entre 2002 y 2021 en el municipio de Huehuetla, sierra norte de Puebla con el clasificador bosque aleatorio (RF).

Clase	2002		2021		Cambio	
	Núm. Pixeles	Área (ha.)	Núm. Pixeles	Área (ha.)	Área (ha.)	%
A	13 070	294.1	12 211	274.8	-19.3	-6.6%
P	75 064	1 688.9	40 214	904.8	-784.1	-46.4%
V	84 481	1 900.8	111 916	2 518.1	617.3	32.5%
U	3 015	67.8	11 291	254.0	186.2	274.5%



Clase A: Agrícola; P: Pastizal, V: Vegetación; U: Urbano-rural.

La disminución de superficies de coberturas agrícola y pastizal (clases A y P) se convirtieron a vegetación y zonas urbano-rurales (clases V y U). El incremento en la cobertura de vegetación no implica necesariamente que la superficie de bosques se ha recuperado, ya que esta clase incluye cultivos bajo sombra que se establecieron recientemente en paisajes agrícolas y pastizales debido al programa sembrando vida que el gobierno federal inició en 2019. Cotler *et al.*, (2020) reportaron que, en relación a la biodiversidad, las localidades sin el Programa Sembrando Vida presentan superficies significativamente mayores de sitios de atención prioritaria (paisajes agrícolas y pastizales alterados) para la conservación de la biodiversidad.

Conclusiones

El clasificador de aprendizaje automático bosque aleatorio (RF) entrenado a partir de imágenes de satélite Landsat 7 (2002) y Landsat 8 (2021) representativas de la región de Huehuetla, Sierra norte de Puebla, México alcanzó una precisión global de clasificación correcta (ACC) de 92.5 % en predicción para clasificar los cuatro tipos de cobertura de suelo analizadas Agrícola, Pastizal, Vegetación y Urbano-rural.

Para el escenario de entrada 2002, RF obtuvo una ACC de 92.5 %; mientras que, en el escenario de entrada 2021 RF alcanzó una ACC de 92.3 %. A nivel de clases objetivo, RF obtuvo mayor capacidad para discriminar la cobertura de vegetación, con un puntaje F1 de 100 % en 2002 y F1 de 98 % en 2021. En tanto que, con la cobertura Urbano-rural, RF el desempeño más bajo, un puntaje F1 de 71.5 % en 2002 y un puntaje F1 de 81.8 % en 2021.

La cuantificación de superficies de cobertura con las predicciones del modelo RF para toda la zona de estudio de Huehuetla, Puebla (3 951.7 ha) en 2002 las superficies estimadas fueron agrícola 294 ha, pastizal 1689 ha, vegetación 1901 ha y urbano-rural 68 ha; mientras que, en 2021 las superficies estimadas fueron agrícola 275 ha, pastizal 905 ha, vegetación 2518 ha y urbano-rural 254 ha. Los cambios de cobertura mostraron un incremento en vegetación y urbano-rural; mientras que se observó un decremento en las superficies agrícola y de pastizal de 2002 a 2021.

Referencias Bibliográficas

Baeza S, Baldassini P, Bagnato C, Pinto P, Paruelo J. 2014. Caracterización del uso/cobertura del suelo en Uruguay a partir de series temporales de imágenes MODIS. *Agrociencia* 18: 95–105. <https://doi.org/10.31285/AGRO.18.470>.



- Breiman L. 2001. Random forests. *Machine Learning* 45: 5–32. <https://doi.org/10.1023/A:1010933404324>.
- CONABIO (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad). 2021. Geoportal del Sistema Nacional de Información sobre Biodiversidad. <http://www.conabio.gob.mx/informacion/gis/> (Recuperado: mayo 2021).
- Cotler H, Manson R, Nava MJD. 2020. Evaluación de la focalización del Programa Sembrando Vida. *CentroGeo INECOL* pp: 26–43. <http://centrogeo.repositorioinstitucional.mx/jspui/handle/1012/294> (Recuperado: junio 2022).
- Chucos BN, Vega VEJ. 2022. Evaluación de algoritmos de machine learning en la clasificación de imágenes satelitales multiespectrales, caso: Amazonia Peruana. *Ciencia Latina Revista Científica Multidisciplinar* 6(1): 4946–4963. https://doi.org/10.37811/cl_rcm.v6i1.1843.
- Chuvienco ES. (2010). *Teledetección Ambiental: La observación de la Tierra desde el Espacio* Edición 2019. Digital Reasons, Ariel, España. pp: 413–416.
- Guevara RML, Montalvo VR. 2015. Cambio de uso de suelo y vegetación derivados de la dotación de infraestructura: Sierra norte del Estado de Puebla. *Nova Scientia* 7(13): 314–336. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-07052015000100017&lng=es&tlng=es (Recuperado: mayo 2022).
- INEGI (Instituto Nacional de Estadística y Geografía). 2020. México en cifras. <https://www.inegi.org.mx/app/areasgeograficas/#collapse-Resumen> (Recuperado: abril 2022).
- Lipton ZC, Elkan C, Naryanaswamy B. 2014. Optimal Thresholding of Classifiers to Maximize F1 Measure. In: *Machine Learning and Knowledge Discovery in Databases*. Calders T, Esposito F, Hüllermeier E, Meo R. (eds.) ECML PKDD 2014. Lecture Notes in Computer Science 8725. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44851-9_15.
- Martínez LJ, Jiménez SD, Robles SDE. 2017. Propuesta para mejorar la productividad y rentabilidad en el medio rural del Municipio de Huehuetla, Puebla. *ECORFAN* 2: 1–13. https://www.ecorfan.org/handbooks_cecoh2.php (Recuperado: abril 2022).
- Paredes GV, Del Blanco MV, Gutiérrez GA, Nafría GD. 2019. Seguimiento y evaluación de la capacidad de discriminación de cultivos herbáceos en regadío a partir de imágenes de satélite en el periodo 2016-2018. *Teledetección: hacia una visión global del cambio climático* 3: 15–18. <https://www.researchgate.net/publication/336364032> (Recuperado: abril 2022).
- Raschka S, Mirjalili V. (2019). *Python Machine Learning 3rd ed.* Packt Publishing Ltd. Birmingham, UK. pp: 90–96.
- Suárez A, Jiménez A, Castro FM, Cruz-Roa A. 2017. Clasificación y mapeo automático de coberturas del suelo en imágenes satelitales utilizando Redes Neuronales



Convolucionales. Un caso aplicado en Parques Nacionales Naturales de Colombia. ORINOQUIA 21: 64–75. <https://doi.org/10.22579/20112629.432>.

USGS. (United State Geological Survey). 2019. Landsat Satellite Missions. <https://www.usgs.gov/landsat-missions/landsat-satellite-missions> (Recuperado: mayo 2021).

Valdés SD, Mesejo LD, León MA. 2015. Selección de píxel semilla mediante wavelets para crecimiento por regiones difuso. Revista Internacional de Gestión del Conocimiento y la Tecnología 3: 2–4. <https://ssrn.com/abstract=2643159> (Recuperado: abril 2022).