DOI: 10.24850/j-tyca-14-04-03

Artículos

Algoritmos de aprendizaje automático para clasificar zonas de inundación a partir de imágenes de radar de apertura sintética

Machine learning algorithms for classifying flood areas from synthetic aperture radar images

Juan Pablo Ambrosio-Ambrosio[1], ORCID: https://orcid.org/0000-0002-7032-6976

Juan Manuel González-Camacho[2], ORCID: https://orcid.org/0000-0001-5479-7316

[1]Colegio de Postgraduados Campus Montecillo, Montecillo, México, ambrosio.juan@colpos.mx

[2]Colegio de Postgraduados Campus Montecillo, Montecillo, México, jmgc@colpos.mx

Autor para correspondencia: Juan Manuel González-Camacho, jmgc@colpos.mx

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Resumen

El uso de imágenes de radar de apertura sintética (SAR) representa una fuente valiosa de información para caracterizar regiones geográficas susceptibles de inundaciones, como en el sureste de México, ya que éstas no son sensibles a condiciones de nubosidad y/u oscuridad. En esta investigación se presenta una metodología para identificar cuerpos de agua en una región del sureste de México. Se aplicaron tres algoritmos de aprendizaje automático: bosque aleatorio (RF), potenciación del gradiente (GB) y máquina de soporte vectorial (SVM) para clasificar las tres clases objetivo A: agua, áreas inundadas y cuerpos de agua; I: infraestructura urbana y/o suelo desnudo, y V: vegetación a partir de imágenes SAR. La imagen SAR utilizada cubre una zona geográfica proyectada UTM Zona 15 Norte WGS84, localizada en los estados de Tabasco y Chiapas, la cual fue preprocesada para disminuir errores en la imagen. Los modelos RF, GB y SVM se implementaron en lenguaje Python, que fueron entrenados y probados en predicción a partir de una base de datos de 12 000 muestras, con valores de amplitud de la imagen SAR. El modelo RF obtuvo una precisión global de clasificación (PG) de 0.979(+/-0.003); GB obtuvo PG = 0.979(+/-0.003), y SVM PG = 0.974(+/-0.005). Los tres modelos obtuvieron un valor de F1_*score* superior a 0.99 para predecir la clase A; el clasificador RF obtuvo valores de AUC = 1 para las tres clases objetivo evaluadas. Este estudio permite mostrar el uso potencial de las imágenes satelitales SAR y el alto desempeño de los modelos de aprendizaje automático RF, GB y SVM para clasificar e

identificar los cuerpos de agua, así como resaltar su importancia en estudios de los posibles impactos de las inundaciones.

Palabras clave: aprendizaje supervisado, modelos de predicción, imágenes satelitales, árboles de decisión, cuenca hidrológica, sensores remotos, curvas ROC.

Abstract

The use of synthetic aperture radar (SAR) images represents a valuable source of information to characterize geographic regions susceptible to flooding, such as southeastern Mexico, they are not sensitive to cloudy and / or dark conditions. This research presents a methodology to identify bodies of water in a region of southeastern Mexico. Three machine learning algorithms were implemented: Random forests (RF), Gradient Boosting (GB) and Support Vector Machines (SVM) to classify three target classes: Class A (water, flooded areas, and bodies of water); class I (urban infrastructure and / or bare soil), and class V (vegetation) from SAR images. The SAR image used covers a projected geographical area UTM Zona 15 Norte WGS84 located in the states of Tabasco and Chiapas; this was pre-processed to reduce errors in the image. The RF, GB and SVM models were implemented in Python language. These were trained and tested in prediction from a database of 12 000 samples with amplitude values of the SAR image. The RF model obtained an overall classification accuracy (PG) of 97.9 (+/- 0.003) %; GB obtained PG = 97.9 (+/- 0.003) %, and SVM PG = 97.4 (+/- 0.005). The three models

obtained an $F1\_s$ value higher than 0.99 to predict class A; RF obtained $AUC = 1$ for the three target classes. This study shows the potential use of SAR satellite images and the high performance of RF, GB and SVM machine learning models to classify and identify water bodies as well as highlighting its importance in studies of possible impacts of floods.

Keywords: Supervised learning, prediction models, satellite images, decision trees, watershed, remote sensing, ROC curves.

Recibido: 22/12/2020

Aceptado: 27/12/2021

Introducción

En la actualidad es posible disponer de imágenes de radar de apertura sintética obtenidas por medio de sensores remotos. A diferencia de los sensores ópticos, las imágenes de radar no dependen de la radiación solar reflejada o la radiación térmica emitida por la Tierra, sino que emiten su propia radiación electromagnética para realizar sondeos. Por ello, una imagen de radar no es afectada por condiciones meteorológicas o de oscuridad (Sami & Abdulmunem, 2020; Fernández-Ordoñez & Soria-Ruiz,

2015). Las imágenes de satélite abarcan grandes superficies e incluso zonas de difícil acceso y a partir de su análisis es posible extraer información de interés de manera indirecta. El procesamiento de imágenes de radar implica un reto para tratar deformaciones y ruido causados por la inclinación del sensor. Por ello, la precisión global de un clasificador está condicionada por el nivel de procesamiento de los datos para reducir el moteado en imágenes simples y multitemporales; las características de la imagen empleadas como predictoras son textura, color, objeto a clasificar, zonas agrícolas, cuerpos de agua, zonas urbanas y el número de clases objetivo (Gomarasca *et al.*, 2019).

Existen diferentes enfoques para clasificar una imagen, por ende, la elección de un enfoque específico depende de la naturaleza del objeto a detectar, de la cantidad y calidad de los datos. Los algoritmos de aprendizaje automático poseen buena capacidad predictiva de clasificación de acuerdo con Avendaño-Pérez, Parra-Plazas y Fredy-Bayona (2014), quienes reportan el uso de máquinas de soporte vectorial (SVM) con kernel Gaussiano y un modelo Bayesiano para clasificar agua, tierra y población, y señalan que el desempeño del modelo SVM fue superior al Bayesiano. Pulella, Aragão-Santos, Sica, Posovszky y Rizzoli (2020) realizaron un mapeo de áreas forestales en el estado de Rondonia, Brasil, con un modelo bosque aleatorio (RF) entrenado con imágenes de radar multitemporal sentinel-1 para clasificar áreas artificiales, áreas forestales, y no forestales, con una precisión global de clasificación superior al 80 %.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Las imágenes de radar se han utilizado en diversas áreas de estudio, por ejemplo, en la identificación de derrame de petróleo, monitoreo del hielo marino, detección de vehículos marítimos, clasificación de cobertura del suelo, monitoreo de la humedad del suelo y detección de áreas inundadas. Shen, Wang, Mao, Anagnostou y Hong (2019) presentan una revisión amplia acerca de las ventajas y desventajas que implica el uso de estas imágenes para mapear extensiones de inundaciones. Tales autores reportan una precisión aceptable en el mapeo automático de inundaciones sin obstrucciones; sin embargo, la detección de inundaciones debajo de vegetación en áreas urbanas fue menos satisfactoria. Lin, Yun, Bhardwaj y Hill (2019) reportan un estudio para detectar inundaciones en áreas urbanas a partir de imágenes multitemporales del satélite sentinel-1, originadas por el huracán Matthew en 2016, en Carolina del Norte, EUA, y obtuvieron baja precisión. De manera similar, Zhang *et al.* (2018) reportan el uso de datos multitemporales Sentinel-1A y Sentinel-1B para zonificar las inundaciones inducidas por el huracán Irma en Florida, EUA, en 2017.

En el sureste de México se presentan con frecuencia intensos periodos de lluvias en los meses de agosto, septiembre, octubre y noviembre que originan inundaciones y provocan pérdidas económicas en diferentes sectores de la población, así como, en ocasiones, pérdidas de vidas humanas (Sánchez, Salcedo, Florido, & Mendoza, 2015). Por ello es de interés contar con métodos indirectos de identificación confiables para identificar superficies afectadas por inundaciones, pues una identificación

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

in situ puede resultar costosa y tardada. En estas situaciones es de gran utilidad el uso de imágenes de satélite y modelos de aprendizaje automático supervisado para clasificar zonas afectadas por inundaciones.

El estado de Tabasco, por su situación geográfica, se confronta a inundaciones debido a factores como la presencia de dos de los ríos más caudalosos de México (el Usumacinta y el Grijalva); una precipitación promedio anual de 2426 mm; falta de ordenamiento territorial; deforestación de las partes altas de la cuenca; cambio climático, y factores antrópicos (Arreguín-Cortés & Rubio-Gutiérrez, 2014; Perevochtchikova & Lezama-de-la-Torre, 2010).

Por lo anterior, el objetivo en este estudio consistió en la aplicación de tres algoritmos de aprendizaje automático: potenciación del gradiente, bosque aleatorio, y máquinas de soporte vectorial para zonificar una región geográfica comprendida en los estados de Tabasco y Chiapas dentro de tres clases objetivo A: agua, áreas inundadas y cuerpos de agua; I: infraestructura urbana y/o suelo desnudo; y V: vegetación a partir de imágenes de radar de apertura sintética; así como ilustrar el uso potencial de imágenes de radar para detectar cuerpos de agua. La imagen de radar que se analizó corresponde a un escenario causado por la inundación ocurrida en la zona de estudio el 8 de octubre de 2017 en el estado de Tabasco, México.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Materiales y métodos

Área de estudio

El área de estudio se encuentra delimitada por el polígono rojo descrito en la Figura 1, con referencia al sistema geodésico de coordenadas geográficas World Geodetic System 1984 (WGS84), con una proyección al sistema de coordenadas UTM (Universal Transverse Mercator). La imagen de radar de apertura sintética (SAR, synthetic aperture radar) abarca gran parte del estado de Tabasco y una pequeña parte del estado de Chiapas. La zona de estudio comprende una superficie de 408 687 ha.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

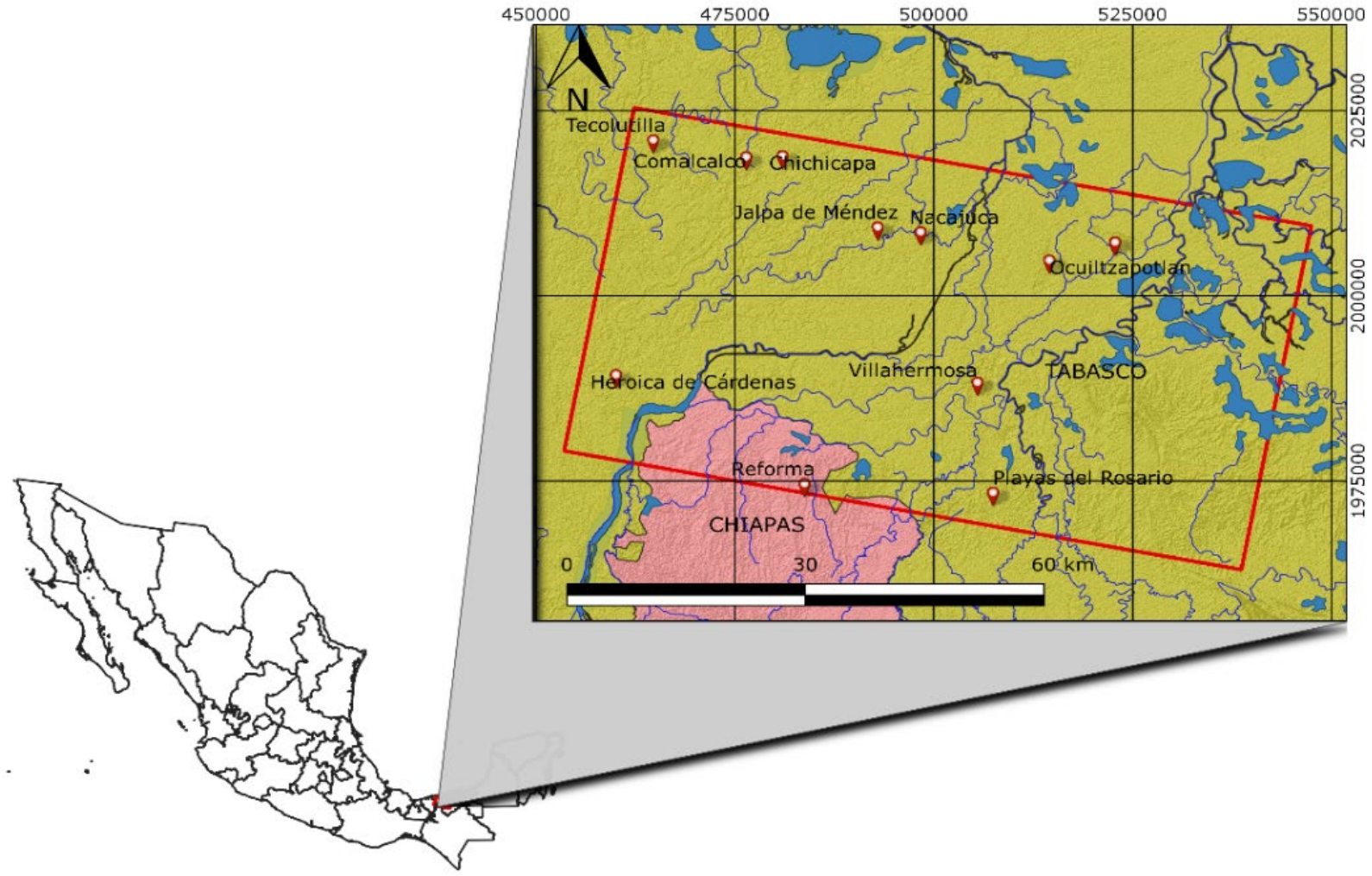

Figura 1. Descripción de la zona de estudio, Tabasco y Chiapas, SRC UTM Zona 15 Norte WGS84.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Datos de entrada

Para realizar este estudio se obtuvieron imágenes SAR del satélite Sentinel-1A. La descarga se hizo a través del sitio web del Instituto de Geofísica de la Universidad de Alaska, Fairbanks (ASF, 2020). En la Tabla 1 se describen las características específicas de la imagen SAR.

Tabla 1. Características de la imagen SAR.

Concepto	Descripción
Producto	S1A_IW_GRDH_1SDV_20171008T
Fecha de adquisición	8-OCT-2017 12:00:16.264908
Nivel del producto	1, Producto estándar georreferenciado
Modo de adquisición	IW (Interferometric Wide)
Ancho de banda Azimuth	327 Hz
Tipo de producto	GRD (Grand Range Deteccion)
Polarización	Dual VV + VH
Frecuencia	Banda C
Paso	Descendente

Fuente: Copernicus Sentinel Data (2017).

El procesamiento digital de las imágenes se realizó con el *software* de uso libre SNAP (Sentinel Application Platform); la caja de herramientas Sentinel-1 facilita el procesamiento de imágenes SAR (ESA & SEOM, 2019). La transformación de imágenes SAR a formato matricial se realizó

con el *software* Matlab (MathWorks Inc., 2016), y el procesamiento geoespacial de las imágenes SAR se realizó con el *software* QGIS (QGIS.org, 2020). Los modelos de aprendizaje automático para clasificación se implementaron en lenguaje Python con la biblioteca Scikit-learn (Pedregosa *et al.*, 2011). El procesamiento de imágenes, datos y modelos se realizó con un sistema de cómputo bajo ambiente Windows 10 de 64 bits, procesador Intel Core i5 @2.50 GHz y memoria RAM de 8 GB.

Preprocesamiento de imágenes SAR

Las imágenes SAR, previo a su análisis, se preprocesaron con SNAP. En la Figura 2 se describen las etapas de preprocesamiento de la imagen.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

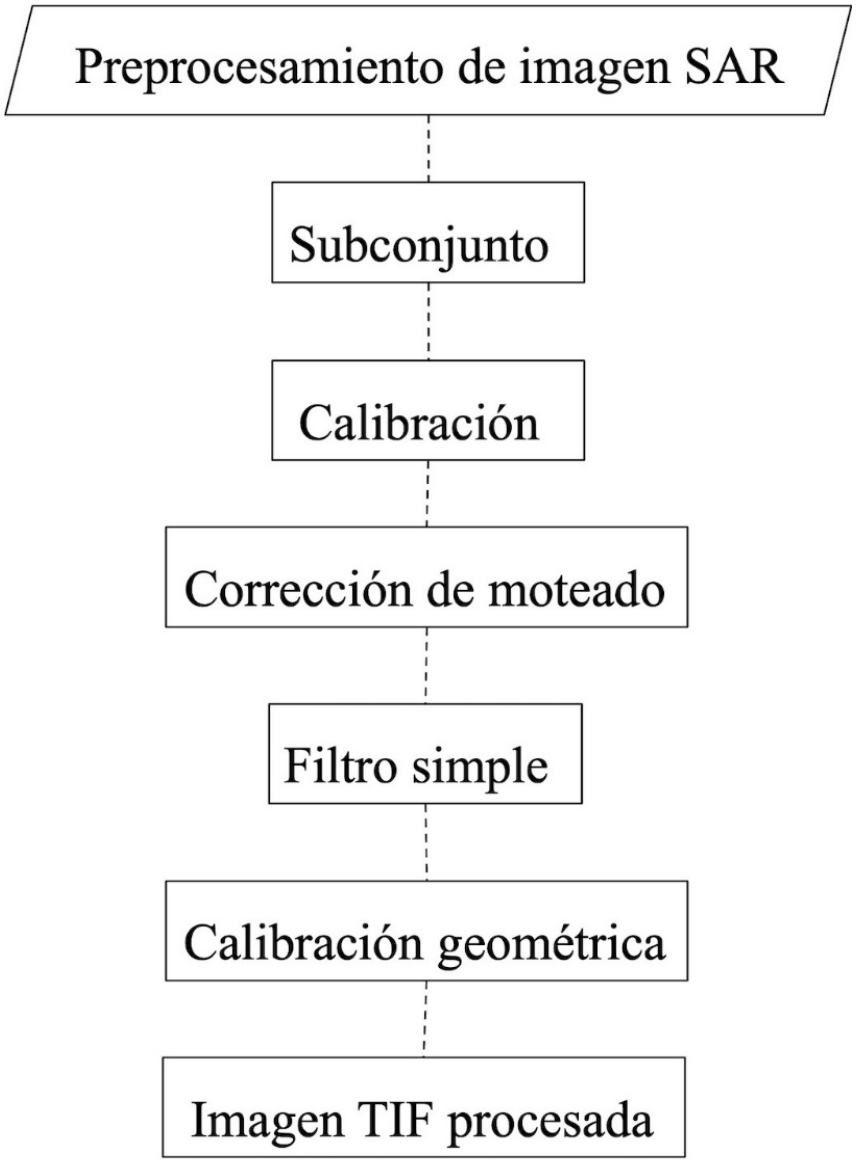

Figura 2. Etapas de preprocesamiento de la imagen de radar de apertura sintética (SAR).

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

El proceso de corrección consiste en definir una subsección de la imagen original que depende del área de interés y la capacidad de procesamiento computacional disponible. A esta subsección se le aplica una corrección radiométrica para eliminar los moteados más visibles de la imagen; se emplea un filtro Lee con parámetros $look = 1$ y tamaño de ventana 3×3. Luego, se aplica una corrección geométrica para georreferenciar la imagen y facilitar su visualización geográfica. Finalmente, las imágenes se exportan a un formato gráfico TIF (Abdurahman-Bayanudin & Heru-Jatmiko, 2016; Podest, 2018; UN-SPIDER, 2020).

Selección de muestras

La base de datos se creó con base en muestras de la imagen SAR asociadas con tres clases objetivo A: agua, áreas inundadas y cuerpos de agua; I: infraestructura urbana y/o suelo desnudo, y V: vegetación. Las muestras de la imagen resaltan las características de interés, tales como: ríos, lagunas, pantanos, ciudades, carreteras, vegetación, y suelo desnudo.

La clase A se integra por los cuerpos de agua que se definen por ríos perennes, ríos intermitentes, lagos, lagunas y zonas agrícolas inundadas. Los pixeles de esta clase se seleccionaron con base en un

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

muestreo de conveniencia. Los límites de los cuerpos de agua se detectaron por medio de dos métodos de segmentación: el método de crecimiento por región, que consiste en aplicar un parámetro de disimilaridad ($d = 0.035$) a partir de un pixel semilla para obtener una imagen binaria con valores iguales a 1 para los pixeles (x, y) similares a la semilla y 0 de otro modo; y el método de umbral, que consiste en analizar el histograma de la imagen para definir los valores de los umbrales superior e inferior que segmentan los objetos de la mejor manera posible; este método genera una imagen binaria con pixeles (x, y) iguales a 1 si éstos se encuentran dentro del intervalo de umbrales definidos previamente y 0 de otra forma.

La clase A se creó con los valores de amplitud de las bandas sigma0_VH y sigma0_VV para cada pixel (x, y); luego, estos pixeles se exportaron a un archivo de texto con las coordenadas (x, y) de cada pixel y sus valores de sigma0_VH y sigma0_VV.

La clase I representa la infraestructura física que consiste en carreteras, edificios, casas, comercios, fábricas, y suelo con poca o nula vegetación, que son visibles en la imagen auxiliar híbrida satelital de Google. Para definir esta clase se utilizó un muestreo de conveniencia; a cada muestra de la imagen se le realizó una segmentación binaria por medio de una definición manual del umbral. La muestra de grandes ciudades que se ilustra en formato *uint*8 se segmentó con el intervalo de umbrales $0.35 - 255$. La exportación de las bandas sigma0_VH y

sigma0_VV se realizó para los pixeles (x, y) con valores iguales a 1 que resultan de la segmentación.

La clase V representa la vegetación: bosque, pastizal, zona agrícola y selva. Esta clase se identificó por medio de un muestreo de conveniencia, que consiste en definir pequeñas ventanas, que representan en su totalidad un tipo de vegetación; luego, por medio de una segmentación mediante crecimiento por región ($d = 0.5$) se accede a la matriz de datos de la imagen para extraer los valores de las bandas sigma0_VH y sigma0_VV.

Algoritmos de aprendizaje automático

Tres algoritmos de aprendizaje automático se seleccionaron para realizar la clasificación multiclase a partir de las características de las bandas de la imagen SAR.

Modelo bosque aleatorio

El modelo bosque aleatorio (RF, random forest) fue propuesto por Breiman (2001); RF consiste en una combinación de predictores de árboles de decisión, en la cual cada árbol depende de los valores de un vector aleatorio independiente y con igual distribución para todos los árboles del bosque; luego se genera una gran cantidad de árboles de decisión y se elige la clase predicha más frecuente. RF se define como un clasificador que consta de un conjunto de clasificadores árboles-estructurados $\{h(\mathbf{x}, \Theta k), k = 1, \ldots\}$m donde Θk son vectores aleatorios independientes e idénticamente distribuidos, y cada árbol emite un voto unitario para la clase más frecuente, dada una entrada $\mathbf{x}$. Geurts, Ernst y Wehenkel (2006) proponen que en la construcción de un árbol de decisión se utilice una réplica de arranque de la muestra de aprendizaje y el algoritmo CART (Classification and Regression Trees) junto con la modificación que se utiliza en el método subespacial. En cada nodo de prueba, la división óptima se obtiene mediante la búsqueda de un subconjunto de tamaño k de variables de entrada candidatas (seleccionadas sin remplazo).

En la división de un nodo en un árbol de decisión (en problemas de clasificación), la función de pérdida más empleada es el índice de Gini ($i_G(t)$), la cual se define por (Gini, 1912):

$$L = i_G(t) = \sum_{k=1}^{J} p(C_k|t)\big(1 - p(C_k|t)\big) \qquad (1)$$

donde $p(C_k|t)$ es la proporción de muestras que pertenecen a la clase k; y J representa el número total de clases para un nodo particular t.

Modelo potenciación del gradiente

El modelo potenciación del gradiente (GB, Gradient Boosting) se define por un conjunto de árboles de decisión que se entrenan secuencialmente por medio de un procedimiento computacional iterativo, donde cada modelo subsiguiente corrige los errores de su predecesor; como resultado, se obtiene un modelo más robusto (Friedman, 2001). La función de pérdida a optimizar se expresa por:

$$L(\{y_k, F_k(\mathbf{x})\}1^K) = -\sum_{k=1}^{K} y_k \log P_k(\mathbf{x}) \qquad (2)$$

donde $y_k \in \{0,1\}$; $k = 1, \ldots, K$ clases; y $P_k(\mathbf{x}) = Pr(y_k = 1|\mathbf{x})$. Con frecuencia se emplea la regresión logística (función sigmoide) para encontrar la probabilidad de que la entrada $\mathbf{x}$ genere una salida $y = 1$.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

Máquina de soporte vectorial

Una máquina de soporte vectorial (SVM, Support Vector Machine) se considera como una extensión del modelo perceptrón (Raschka & Mirjalili, 2017). SVM maximiza el margen; esto es, la distancia entre el hiperplano de separación (límite de decisión) y las muestras de entrenamiento próximas a este hiperplano, que se denominan vectores soporte (Figura 3).

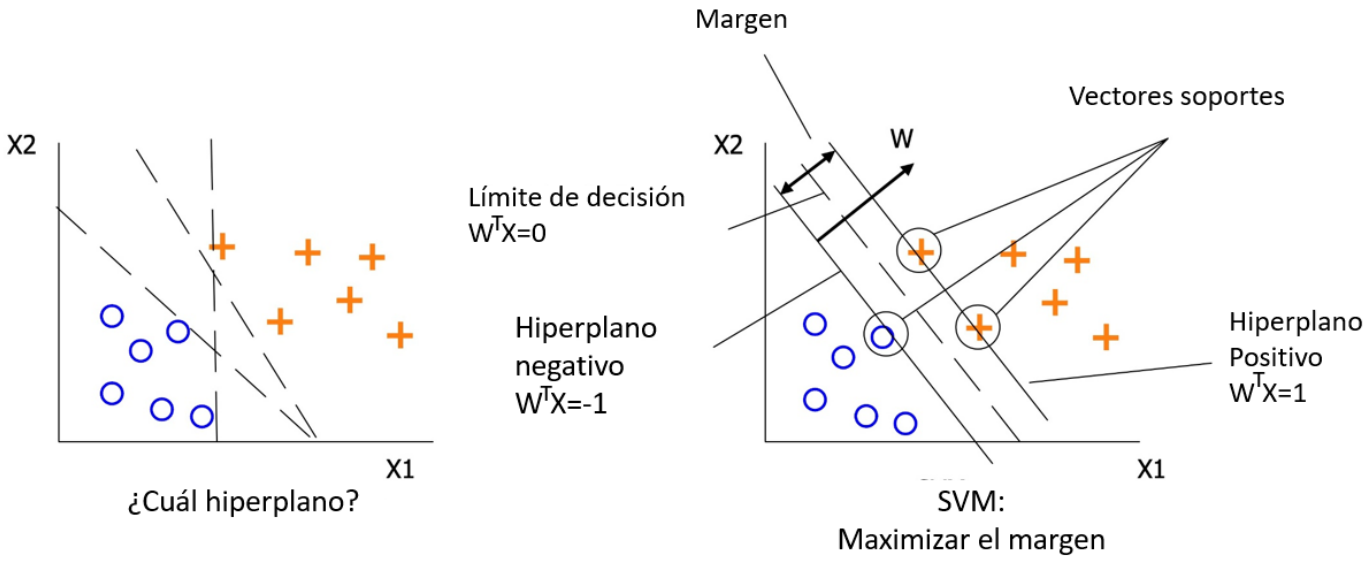

Figura 3. Descripción de la búsqueda del hiperplano máximo de separación entre clases en una máquina de soporte vectorial (SVM). Fuente: Raschka y Mirjalili (2017).

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Vapnik (1995) introdujo el concepto de variables de holgura (ξ, ξ^*) para suavizar las restricciones lineales y alcanzar convergencia en la optimización de problemas que son no linealmente separables. La función de pérdida consiste en minimizar la función L (Ecuación (3)) sujeta a las restricciones (ecuaciones (4), (5) y (6)). El hiperparámetro C controla el ancho del margen de separación entre clases (Smola & Schölkopf, 2004).

$$\mathrm{L} = \min\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\left(\xi_i + \xi^*{}_i\right)\right) \quad (3)$$

$$y_i - \langle w, \mathrm{x}_i\rangle - b \leq \varepsilon + \xi_i \quad (4)$$

$$\langle w, \mathrm{x}_i\rangle + b - y_i \leq \varepsilon + \xi^*{}_i \quad (5)$$

$$\varepsilon_i, \xi^*{}_i \geq 0 \quad (6)$$

donde $\{(\mathrm{x}_1, y_1), \ldots, (\mathrm{x}_l, y_l)\}$ es el conjunto de datos de entrenamiento contenido en el espacio X; $\langle w, \mathrm{x}_i\rangle$ denota el producto punto entre el vector de pesos w y las entradas x_i en X, y b es una variable que representa el sesgo.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

El objetivo es obtener una función que tenga como máxima desviación ε de los objetivos y_i obtenidos para todos los datos de entrenamiento y sea lo más plana posible; plana significa que se busca la w más pequeña. Para garantizar esta condición se obtiene el mínimo de la norma $\|w\|^2 = \langle w, w \rangle$; esto se transforma en un problema de optimización convexa (Ecuación (3)).

Entrenamiento de los clasificadores

Para realizar el entrenamiento de los clasificadores RF, GB y SVM se realizó un procedimiento de validación cruzada (CV), que consiste en subdividir el conjunto de datos original en k subconjuntos disjuntos. Para $k = 10$ se generan 10 combinaciones de conjuntos de datos de entrenamiento y prueba en proporción 9:1 (MathWorks Inc., 2021). Con estos subconjuntos se evalúa la capacidad predictiva de los clasificadores; esta técnica permite evitar problemas de sobreajuste de los algoritmos a los datos de entrenamiento. Asimismo, se realiza una estratificación para obtener las mismas proporciones de clase en cada subconjunto de entrenamiento y prueba seleccionado. Al finalizar el entrenamiento se obtienen k valores de desempeño (PG) del modelo y PG promedio (Raschka, 2018).

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Selección de hiperparámetros de los clasificadores

La selección óptima de los hiperparámetros (párametros de ajuste que no dependen de los datos de entrada) de cada clasificador se realizó en dos etapas: la primera consiste en seleccionar una gráfica de validación que describe la variación del error de clasificación, en función de los valores de cada hiperparámetro seleccionado; luego, en la segunda etapa, la selección se afina por medio de una búsqueda por retícula que consiste en definir un intervalo de búsqueda para cada hiperparámetro (Tabla 2). Esto permite seleccionar la combinación de valores de los hiperparámetros que obtiene el mejor desempeño del algoritmo.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Tabla 2. Intervalos de búsqueda de hiperparámetros para los modelos; bosque aleatorio (RF), potenciación del gradiente (GB), y máquina de soporte vectorial (SVM).

Modelo	Hiperparámetro	Intervalo
RF	Número de estimadores (N)	(1, 500)
	Máxima profundidad (MP)	(1, 46)
	Mínimo de muestras por nodo (MMD)	(2, 10)
	Mínimo de muestras por hoja (MMH)	(1, 10)
GB	Tasa de aprendizaje (TA)	(0.0001, 10)
	Número de estimadores (N)	(1, 800)
	Mínimo de muestras por hoja (MMH)	(1, 10)
	Máxima profundidad (MP)	(1, 15)
SVM	*Kernel* (K)	*rbf*, *poly*, *s*, *l*
	Parámetro de penalización (C)	(0.1, 100)
	Gamma (G)	(0.1, 1)

rbf: Gaussiana

poly: polynomial

s: sigmoide

l: lineal.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Los modelos de ensamble RF y GB se basan en el algoritmo estándar de árboles de decisión, cuya estructura se puede controlar mediante la definición de hiperparámetros: número de árboles o estimadores (N); número mínimo de muestras que debe tener una división para ser considerado un nodo (MMD); número mínimo de muestras para ser considerado una hoja (MMH); profundidad límite máximo permisible (MP); es decir, no se permitirá ninguna división adicional a este valor (Swamynathan, 2017). La tasa de aprendizaje (TA) se refiere a la magnitud de las actualizaciones de los parámetros del algoritmo GB; un valor bajo demora el tiempo de convergencia con alta probabilidad de seleccionar un valor mínimo local y un valor alto incrementa la posibilidad de saltar la solución mínima global. La selección de este hiperparámetro es experimental y con conocimiento previo de problemas similares (Raschka & Mirjalili, 2017). En SVM, C es el hiperparámetro que equilibra entre minimizar las variables de holgura y maximizar el margen. Por otra parte, SVM implementa funciones de mapeo o kernels para aumentar la dimensión del problema y encontrar un plano de separación, que sin esta transformación no es posible visualizar; el kernel más frecuente es el Gaussiano (*rbf*, radial basis function) con su hiperparámetro *Gamma* (Patle & Chouhan, 2013).

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Métricas de evaluación del desempeño

Para evaluar el desempeño de los clasificadores RF, GB y SVM se utilizaron las métricas siguientes: precisión global de clasificación correcta (PG); precisión (P); sensibilidad (S); F1_score ($F1s$), y el área bajo la curva ROC (Receiver Operating Characteristic Curve) (AUC, Area Under the Curve). Estas métricas se obtienen de una matriz de confusión que describe en un cuadro de doble entrada los resultados del clasificador, donde se muestran los valores predichos (columnas) y observados (hileras) de cada clase (Figura 4). La matriz de confusión permite visualizar los aciertos y fallas del clasificador (Barrero-Ortiz, 2019).

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

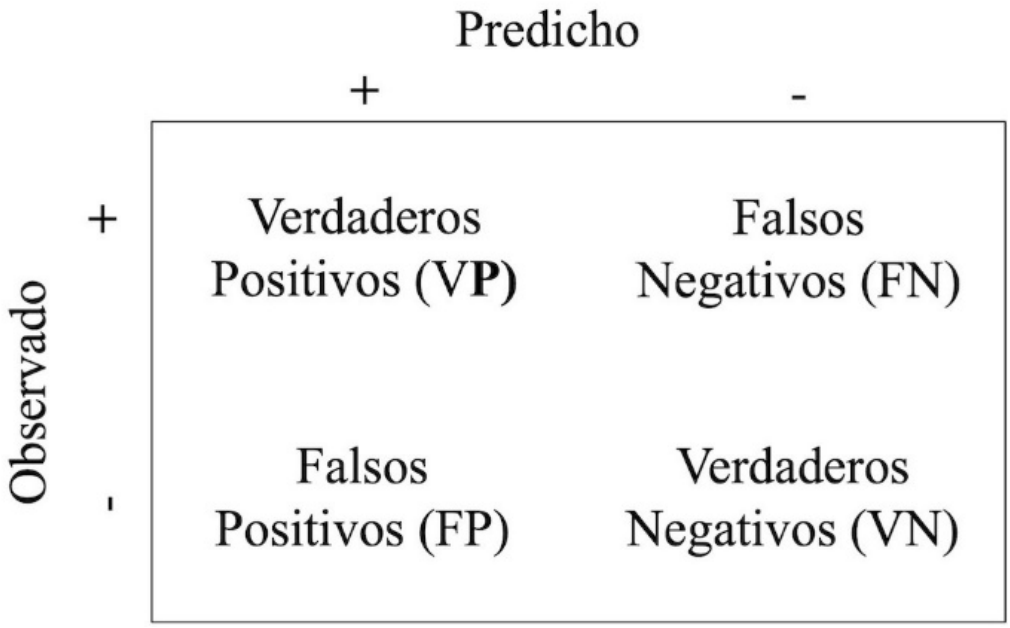

Figura 4. Descripción de una matriz de confusión para el caso de una clasificación binaria.

$PG = 1 - Ec$ mide la proporción global de clasificaciones correctas; Ec es el error de clasificación; PG se calcula por:

$$PG = \frac{VP+VN}{VP+VN+FP+FN} \quad (7)$$

Donde VP es el número de muestras positivas que se clasifican correctamente; VN es el número de muestras negativas que se clasifican correctamente; FP es el número de muestras negativas que se clasifican

incorrectamente como positivas; FN es el número de muestras positivas que se clasifican incorrectamente como negativas. S es la proporción de casos que se predicen correctamente como positivos con respecto al total de valores positivos observados y se define por:

$$S = \frac{VP}{VP+FN} \tag{8}$$

P es la proporción de VP con respecto al total de valores positivos predichos y se calcula como (Bradley, 1997):

$$P = \frac{VP}{VP+FP} \tag{9}$$

$F1s$ es la media armónica de S y P, y se expresa por:

$$F1s = \frac{2*S*P}{S+P} \tag{10}$$

donde $F1s$ varía en el intervalo (0, 1).

La curva ROC mide la capacidad discriminante del modelo para diferenciar las clases, y compara la capacidad discriminante de dos o más modelos. En una gráfica ROC típica, cada punto de la curva corresponde a un posible punto de corte del modelo y muestra sus respectivos valores

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

de S (eje Y) y $TFP = FP/(FP + VN)$ (tasa de falsos positivos en el eje X). AUC mide qué tan bien el modelo discrimina muestras que pertenecen a una clase A, de otra no A a lo largo de todo el intervalo de puntos de corte posibles (Parikh, Mathai, Parikh, Chandra-Sekhar, & Thomas, 2008; Fawcett, 2006).

Resultados y discusión

Procesamiento de imágenes de radar

En la Figura 5 se muestra el resultado del procesamiento de la imagen de radar del área de estudio. La calibración radiométrica para las bandas VH (Figura 5A) y VV (Figura 5C) estandariza los valores de retrodispersión. Por otra parte, las correcciones de ruido "sal y pimienta" en las imágenes se realizaron con filtros *Lee;* luego, se efectuó una corrección geométrica al formato WGS84 UTM zona 15 norte de las bandas VH (Figura 5B) y VV (Figura 5D).

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

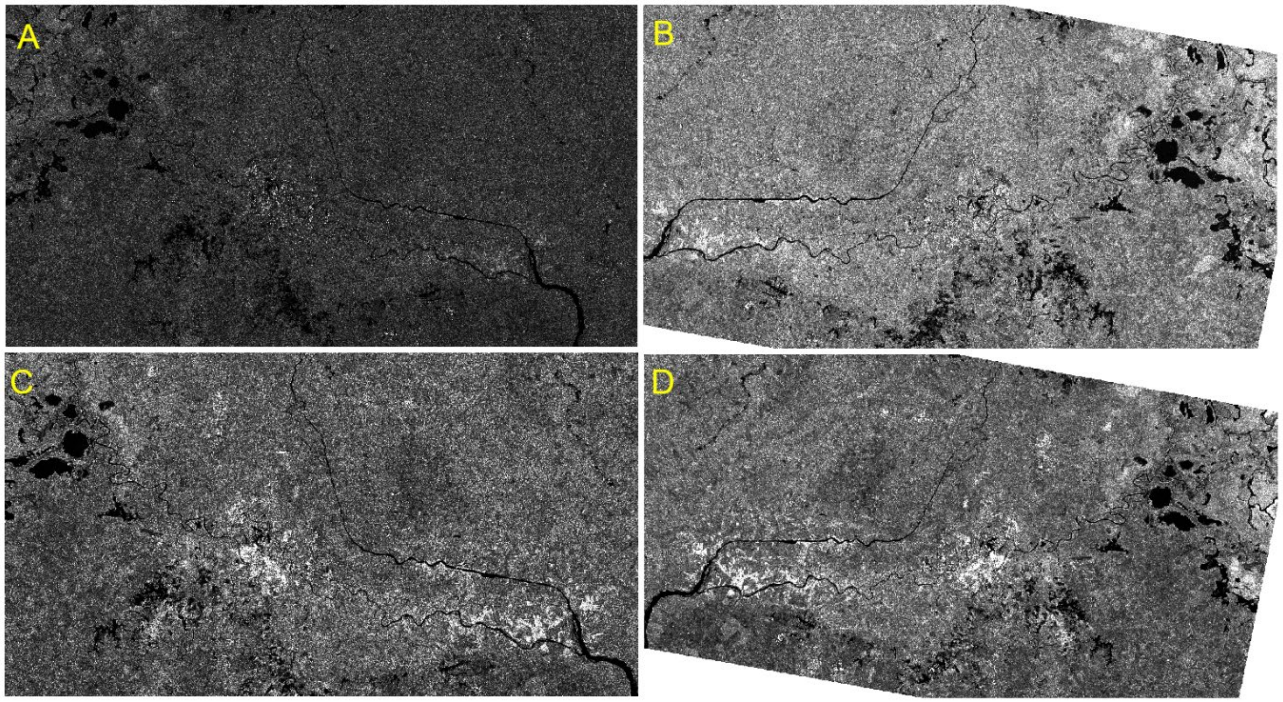

Figura 5. Procesamiento de imágenes de radar, polarización VH: A) calibración radiométrica; B) corrección de moteado y calibración geométrica; polarización VV; C) calibración radiométrica, y D) corrección de moteado y calibración geométrica.

Muestras seleccionadas y etiquetadas

Como resultado del procesamiento y análisis de las imágenes SAR se obtuvo una base de datos íntegra, depurada de muestras repetidas e

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

inconsistencias con 12 000 muestras, 4 000 en cada clase A, I y V, respectivamente.

Hiperparámetros de los clasificadores

Análisis gráfico

Los hiperparámetros óptimos de los modelos RF, GB y SVM seleccionados en la primera etapa con base en el análisis gráfico de curvas de error de clasificación con validación cruzada (k = 10) se describen en la Tabla 3.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Tabla 3. Valores óptimos de los hiperparámetros en la primera etapa de los modelos bosque aleatorio (RF), potenciación del gradiente (GB), y máquina de soporte vectorial (SVM); error promedio de clasificación (Ec), y desviación estándar (Std).

Modelo	Hiperparámetro	Óptimo	Ec	Std
RF	N	200	0.024	0.004
	MP	9	0.022	0.003
	MMD	10	0.021	0.003
	MMH	1	0.021	0.003
GB	TA	0.08	0.022	0.003
	N	200	0.022	0.003
	MMH	1	0.022	0.003
	MP	1	0.022	0.004
SVM	K	rbf	0.027	0.004
	C	120	0.027	0.004
	G	1	0.026	0.005

En la Figura 6 se describe el comportamiento de los hiperparámetros N, MP, MMD y MMH del clasificador RF. En la cual se observa el efecto individual de cada hiperparámetro en el error de clasificación promedio (Ec). El análisis de las gráficas (Figura 6) Ec permite

visualizar el comportamiento del desempeño y seleccionar el valor del hiperparámetro con Ec mínimo.

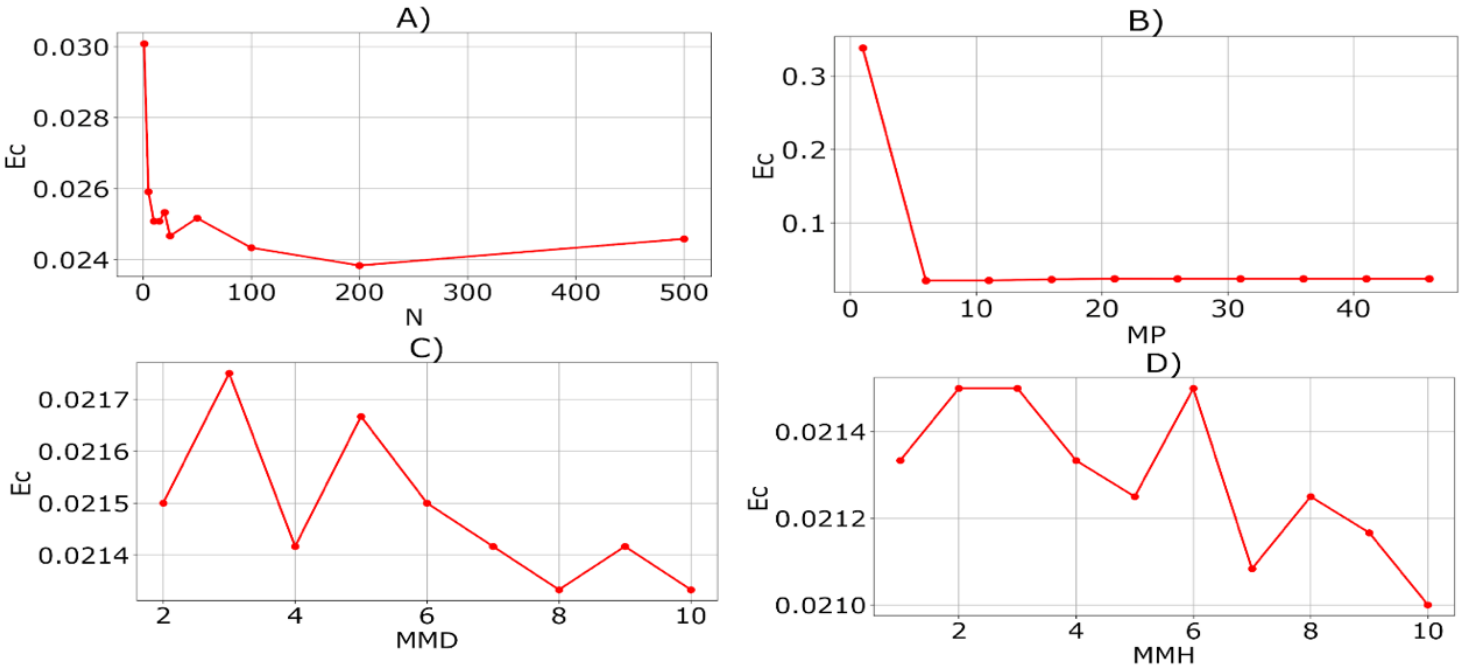

Figura 6. Comportamiento del error de clasificación del modelo bosque aleatorio (RF) en función de los hiperparámetros A) N; B) MP; C) MMD, y D) MMH.

Búsqueda por retícula

Para afinar la búsqueda de valores óptimos de los hiperparámetros de cada modelo se efectuó una búsqueda por retícula con validación cruzada (k = 10), y se seleccionaron las mejores combinaciones con el menor error de clasificación.

El modelo RF con valores $N = 200$, $MP = 9$, $MMD = 10$ y $MMH = 1$ obtuvo una $PG = 0.979\ (+/- 0.003)$; GB con valores $TA = 0.08$, $N = 200$, $MMH = 1$ y $MP = 1$ obtuvo una $PG = 0.979\ (+/- 0.003)$; y SVM con valores $K = rbf$, $C = 120$ y $G = 1$ obtuvo una $PG = 0.974\ (+/- 0.005)$.

Con base en la búsqueda por retícula, los valores óptimos de los hiperparámetros resultaron diferentes a los valores seleccionados con el análisis de las gráficas de Ec y se mejoró el desempeño PG de los tres clasificadores. Estas combinaciones de hiperparámetros óptimos se utilizaron en la evaluación, comparación y predicción final de los clasificadores RF, GB y SVM.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Evaluación del desempeño

Las métricas de evaluación del desempeño de los clasificadores obtenidas con una validación cruzada se ilustran en la Tabla 4. La evaluación del desempeño se realizó con un subconjunto de 1 200 muestras (400 por clase). Los tres clasificadores obtuvieron mejor desempeño para identificar los cuerpos de agua (clase A); RF y GB superaron ligeramente a SVM con $F1_S$ de 99.3 %; en cambio, AUC = 1 para los tres clasificadores significa que son excelentes clasificadores con la adecuada selección del umbral de balance entre S y P. El alto desempeño que se alcanzó para la clase A se debe a su separabilidad del resto, con valores de retrodispersión definidos en el intervalo (-24.0, -22.0) (Fernández-Ordoñez *et al.*, 2020).

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Tabla 4. Métricas de evaluación del desempeño por clase de los clasificadores bosque aleatorio (RF), potenciación del gradiente (GB) y máquina de soporte vectorial (SVM) con base en el conjunto de prueba (10 % del total de datos) de la partición *k* = 10 de la validación cruzada.

Modelo	S	P	$F1s$	AUC
Clase A				
RF	0.998	0.988	0.993	1.00
GB	0.995	0.990	0.993	1.00
SVM	0.995	0.988	0.991	1.00
Clase I				
RF	0.963	0.980	0.971	1.00
GB	0.965	0.977	0.971	1.00
SVM	0.950	0.987	0.968	1.00
Clase V				
RF	0.980	0.973	0.976	1.00
GB	0.980	0.973	0.976	0.99
SVM	0.990	0.961	0.975	1.000

Para identificar la infraestructura física (clase I), RF y GB obtuvieron mejor desempeño que SVM, con $F1s$ = 97.1 % *versus* 96.8 %. Respecto a la identificación de la vegetación (clase V), RF y GB obtuvieron mejor

desempeño que SVM, con $F1s$ = 97.6 % *versus* 97.5 %; en términos de AUC, los tres clasificadores alcanzaron el mismo desempeño. El desempeño de los algoritmos analizados disminuye debido a la confusión que existe entre la clase I y V, provocada por la similitud de valores de retro dispersión, que finalmente se traduce en predicción de falsos negativos para ambas clases (Figura 7).

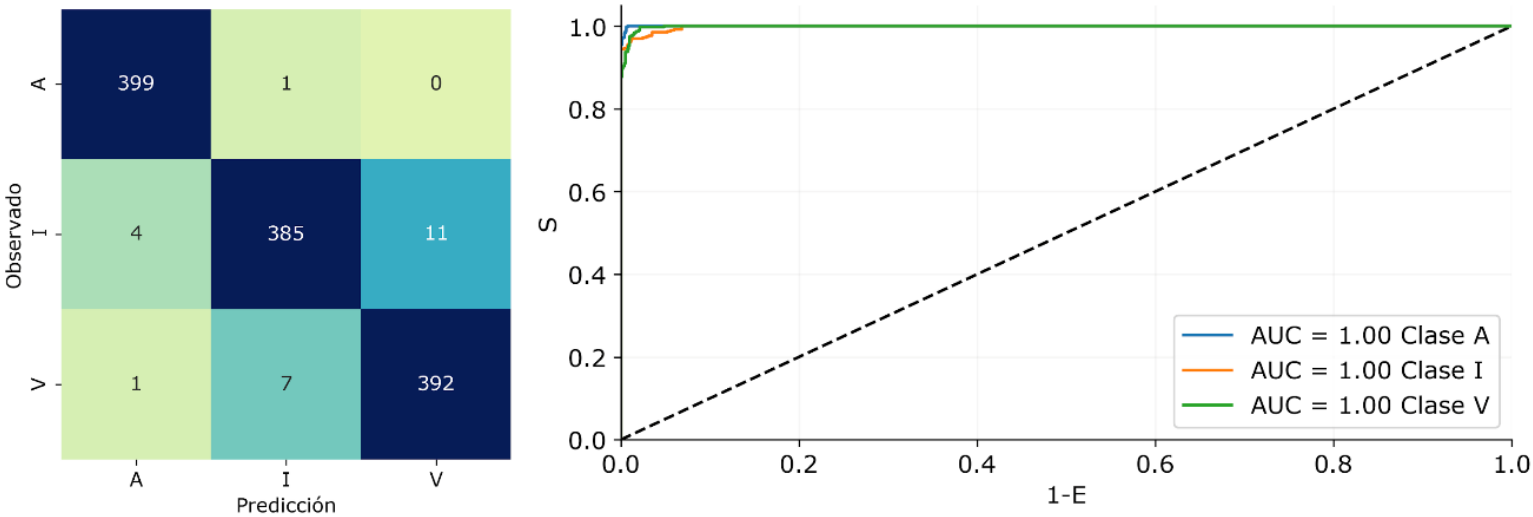

Figura 7. Matriz de confusión y curva ROC del desempeño en predicción del clasificador bosque aleatorio (RF) con base en el conjunto de prueba (10 % del total de datos) de la partición *k* = 10 de la validación cruzada.

Para obtener un mejor estimador de PG se realizó una validación cruzada (VC) con *k* = 10 particiones. En la Figura 8 se describe el comportamiento de PG de los clasificadores RF, GB y SVM. PG promedio

se representa por el punto negro dentro de cada caja. RF obtuvo PG = 0.979 (+/-0.003); GB obtuvo PG = 0.979 (+/-0.003), y SVM obtuvo PG = 0.974(+/-0.005). Estos resultados muestran que el desempeño promedio de los tres clasificadores es muy similar al desempeño obtenido con un solo conjunto de prueba. Asimismo, se observa que los clasificadores de ensamble RF y GB obtuvieron un desempeño ligeramente superior a SVM. Para propósitos prácticos, los tres clasificadores obtuvieron una precisión global de clasificación alta en predicción.

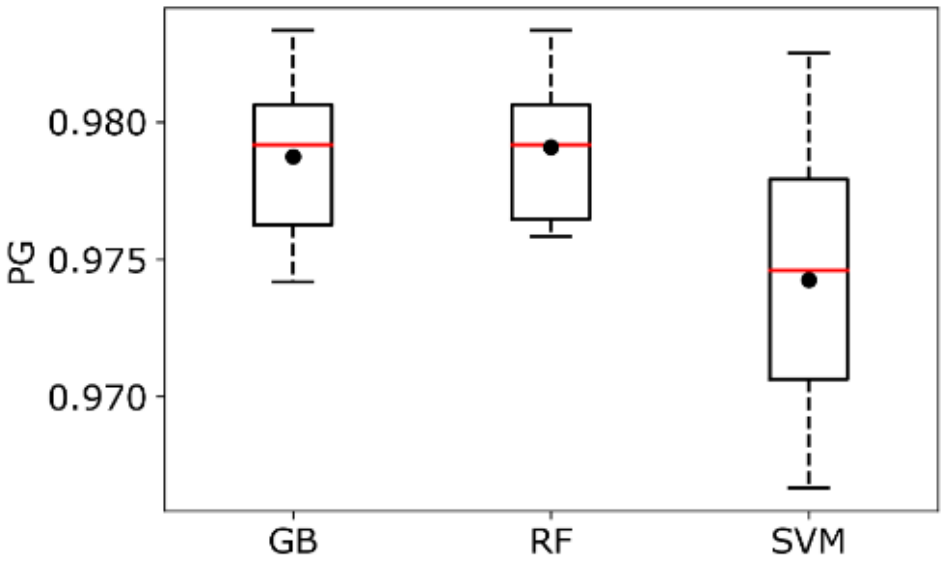

Figura 8. Diagrama de caja para comparar la precisión global (PG) obtenida por validación cruzada (k = 10 particiones) de los clasificadores potenciación del gradiente (GB), bosque aleatorio (RF) y máquina de soporte vectorial (SVM).

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Zonificación del área bajo estudio

Con los clasificadores RF, GB y SVM entrenados, y con sus hiperparámetros óptimos se procedió a la clasificación de pixeles. Las imágenes TIF de entrada y salida están georreferenciadas con el sistema geodésico mundial (WGS84, World Geodetic System 1984) proyectado en UTM zona 15 norte. En la Figura 9 se ilustra la clasificación de una subsección del área de estudio que corresponde a 46 842.8 ha (11.5 % del área total). Para cada clase A, I y V; RF clasificó 3.6, 8.3 y 88.1 %; GB clasificó 3.7, 8.0 y 88.3 %, y SVM 3.7, 15.6 y 80.7 %, respectivamente. Para la clase A, los porcentajes de área calculados por los tres modelos fueron similares; sin embargo, para la clase I, SVM duplicó el área estimada con respecto a RF y GB. En la Figura 10 se presenta la clasificación del área de estudio en su totalidad (408 687.1 ha). El modelo RF obtuvo A = 15 139.2 ha (3.7 %), I = 30 318.8 ha (7.4 %) ha y V = 363 229.1 ha (88.9 %).

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

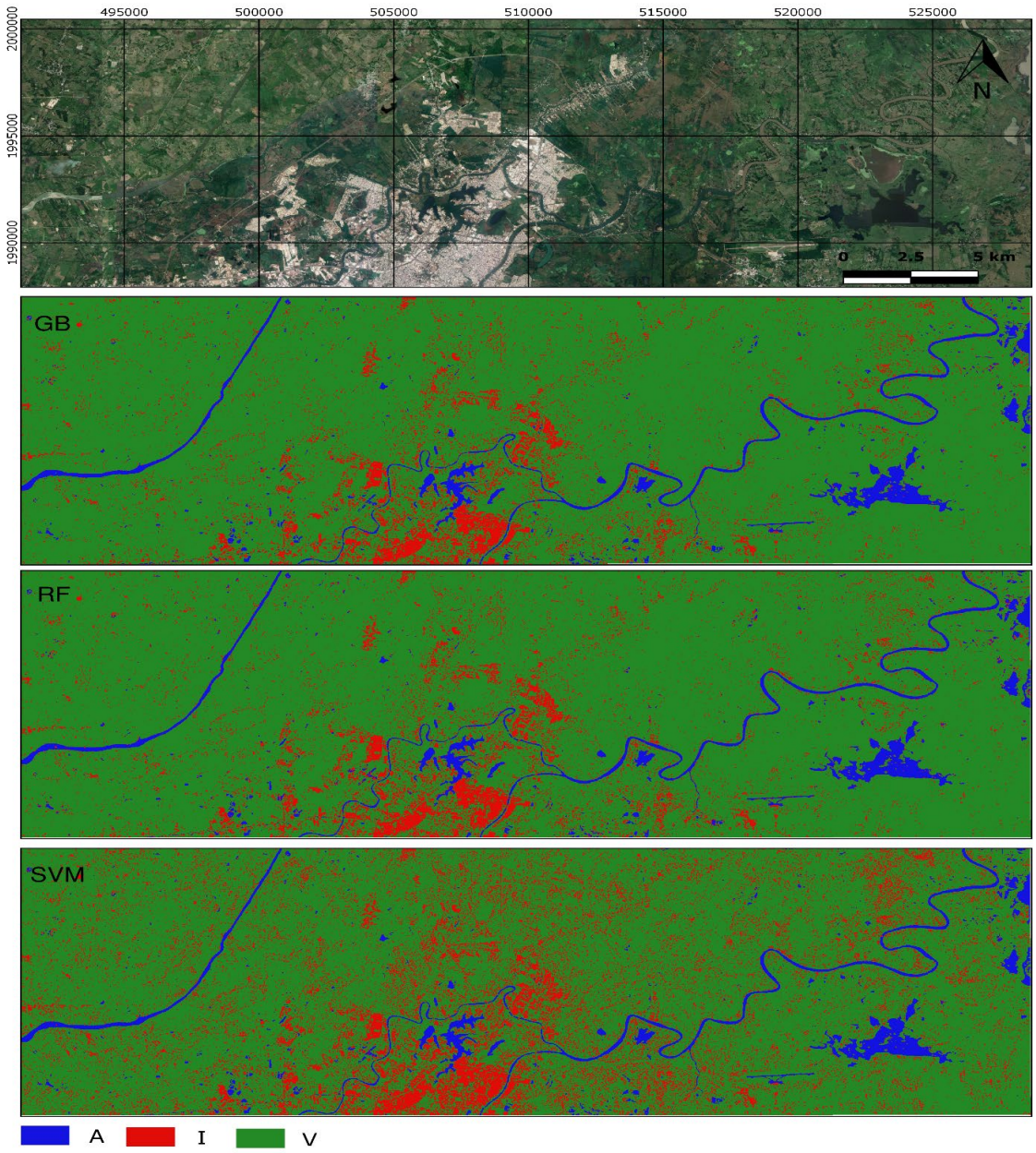

Figura 9. Clasificación de cuerpos de agua (A); infraestructura y suelo desnudo (I), y vegetación (V) con los modelos bosque aleatorio (RF), potenciación del gradiente (GB), y SVM con kernel Gaussiano en una sección de la zona de estudio localizada entre Tabasco y Chiapas, México.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

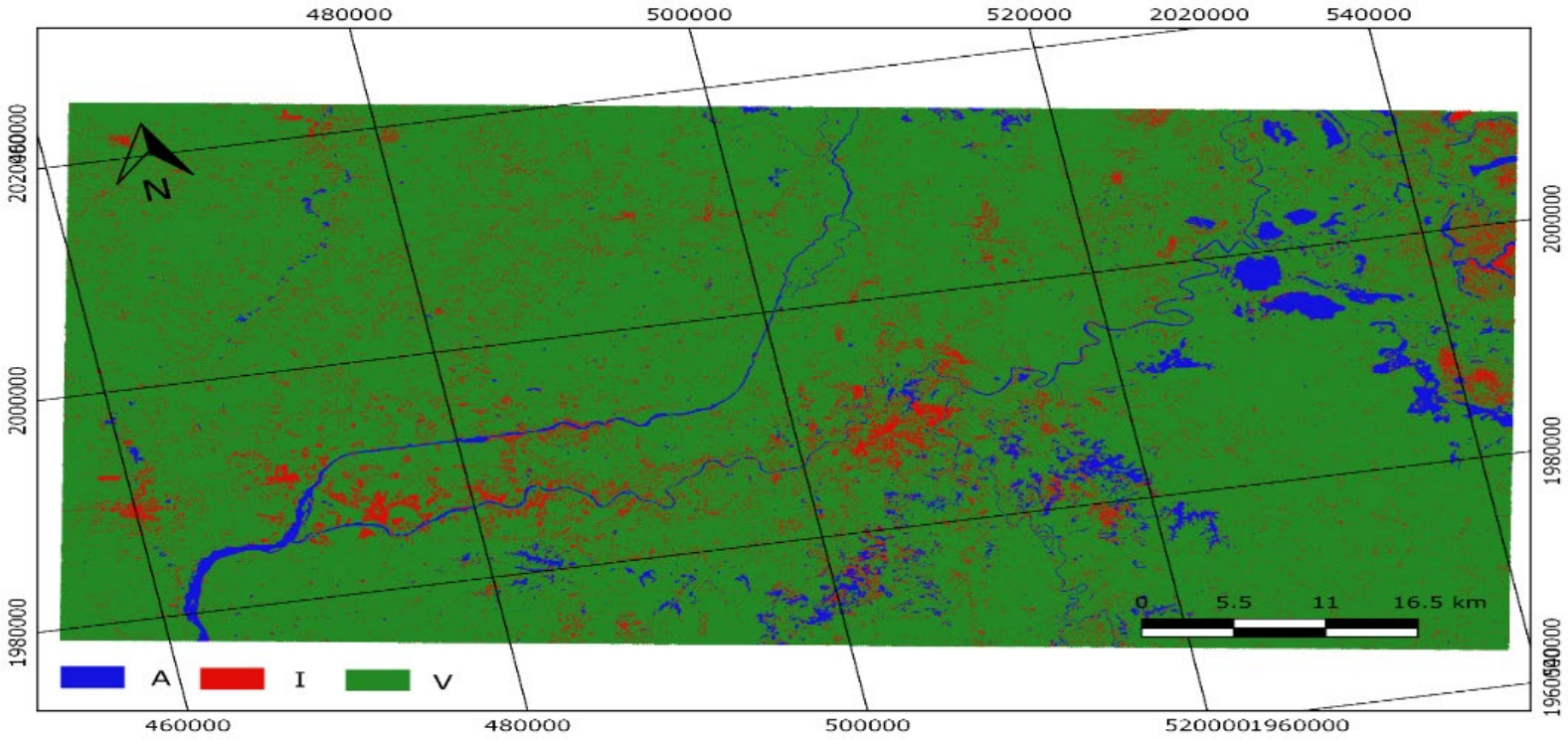

Figura 10. Clasificación de cuerpos de agua (A); infraestructura y suelo desnudo (I), y Vegetación (V) mediante el clasificador bosque aleatorio (RF) en la zona de estudio entre Tabasco y Chiapas, México.

En la Figura 9 se puede confirmar el alto desempeño de los modelos (RF, GB y SVM) para identificar cuerpos de agua con superficie libre y con ligera vegetación acuática. Se determinan perfectamente los límites de los ríos perennes Grijalva y Usumacinta, además de las lagunas, con su respectivo incremento en extensión. Los modelos de aprendizaje

automático permitieron identificar las zonas urbanas inundadas de Villahermosa originada por las intensas lluvias del 6, 7 y 8 de octubre de 2017, que provocaron el desbordamiento de los ríos. El desempeño de los clasificadores utilizados en este estudio son ligeramente superiores a los obtenidos por Chen, Huang, Chen y Feng (2021), dichos autores emplearon un método estándar de umbral adaptativo para automatizar la identificación de zonas inundadas con una PG de 95-97 %. De manera similar al trabajo de Hlaváčová, Kačmařík, Lazecký, Struhár y Rapant (2021), que reporta PG = 83 % debido principalmente al tamaño del área de estudio y enfoque automatizado que emplearon estos autores. Los tres clasificadores identificaron de manera correcta la clase I, que corresponde al asentamiento urbano de Villahermosa. No obstante, predicen una tasa alta de falsos positivos que corresponden a la clase V. Este error se duplica con el modelo SVM (Figura 9). El número de errores de identificación aumenta cuando la predicción se realiza en la totalidad del área de estudio (Figura 10).

Conclusiones

Los clasificadores de aprendizaje automático bosque aleatorio (RF) y potenciación del gradiente (GB) obtuvieron mejor desempeño en predicción que el clasificador máquina de soporte vectorial (SVM) para

identificar cuerpos de agua a partir de imágenes de radar de apertura sintética (SAR). RF y GB obtuvieron una precisión global de clasificación promedio (PG) de 97.9 %; y SVM PG = 97.4 %. Los tres modelos obtuvieron un valor de F1_score superior a 99.3 % para predecir la clase A: agua; 97.6 % para la clase V: vegetación; y 97.1 % para la clase I: infraestructura.

El alto desempeño en predicción de los tres clasificadores se debe, entre otras razones, a que las clases objetivo agua, vegetación y suelo fueron balanceadas; que las muestras de imagen SAR asociadas con cada clase fueron separables con los valores de las bandas de radar, y que la búsqueda por retícula con validación cruzada de los hiperparámetros de cada modelo permitió reducir el error de clasificación.

La predicción de superficies cubiertas por agua con una precisión del 99.2 % de los clasificadores RF y GB a partir de imágenes SAR muestran el uso potencial de estas imágenes para realizar estudios relacionados con la detección de cuerpos de agua, monitoreo y evaluación de daños por inundaciones. Los modelos presentaron ligeramente menor precisión para distinguir las clases de infraestructura y vegetación; sin embargo, la precisión es alta.

Tecnología y ciencias del agua, ISSN 2007-2422, 14(4), 107-154. DOI: 10.24850/j-tyca-14-04-03

Referencias

Abdurahman-Bayanudin, A., & Heru-Jatmiko, R. (2016). Orthorectification of Sentinel-1 SAR (Synthetic Aperture Radar) data in some parts of south-eastern Sulawesi using Sentinel-1 Toolbox. 2nd International Conference of Indonesian Society for Remote Sensing (ICOIRS), IOP Conference Series: Earth and Environmental Science, 47(012007). DOI: 10.1088/1755-1315/47/1/012007

ASF, Alaska Satellite Facility. (2020). Distributed active archive center. Recuperado de https://asf.alaska.edu/

Arreguín-Cortés, F. I., & Rubio-Gutiérrez, H. (2014). Análisis de las inundaciones en la planicie tabasqueña en el periodo 1995-2010. Tecnología y ciencias del agua, 5(3), 5-32.

Avendaño-Pérez, J., Parra-Plazas, J., & Fredy-Bayona, J. (2014). Segmentación y clasificación de imágenes SAR en zonas de inundación en Colombia, una herramienta computacional para prevención de desastres. Universidad Antonio Nariño, Revista Facultades de Ingeniería, 4(8), 24-38.

Barrero-Ortiz, G. (2019). Evaluación de la eficiencia de los modelos machine learning para la predicción de la calidad del software desarrollado en IBM RPG usando la matriz de confusión y las curvas ROC. Maestría en Gestión de Tecnologías de Información de la escuela de Postgrado de la Universidad Cesar Vallejo de Lima Perú. ResearchGate. DOI: 10.13140/RG.2.2.31760.87040

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145-1159. DOI: 10.1016/S0031-3203(96)00142-2

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32.

Chen, S., Huang, W., Chen, Y., & Feng, M. (2021). An adaptive thresholding approach toward rapid flood coverage extraction from Sentinel-1 SAR imagery. Remote Sensing, 13(23). DOI: https://doi.org/10.3390/rs13234899

Copernicus Sentinel Data. (2017). Open access to Sentinel-1 user products. Recuperado de https://scihub.copernicus.eu/

ESA & SEOM, European Spatial Agency & Scientific Exploitation of Operational Missions (2019). Sentinel Application Platform (SNAP 7.0). Programa para el análisis y procesamiento de imágenes satelitales (software). Recuperado de http://step.esa.int/main/download/snap-download/

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27, 861-874. DOI: 10.1016/j.patrec.2005.10.010

Fernández-Ordoñez, Y. M., Soria-Ruiz, J., Leblon, B., Macedo-Cruz, A., Ramírez-Guzmán, M. E., & Escalona-Maurice, M. (2020). Imágenes de radar para estudios territoriales, caso: inundaciones en Tabasco con el uso de imágenes SAR Sentinel-1A y Radarsat-2. Realidad, Datos y Espacio, Revista Internacional de Estadística y Geografía, 11(1).

Fernández-Ordoñez, Y. M., & Soria-Ruiz, J. (2015). Imágenes de radar de apertura sintética y conceptos básicos de polarimetría. En: Fernández-Ordoñez, Y., Escalona-Maurice, M. J., & Valdez-Lazalde, J. R. (eds.). Avances y perspectivas de geomática con aplicaciones ambientales, agrícolas y urbanas (pp. 37-66). Montecillo, México: Colegio de Postgraduados.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Mach Learn, 63, 3-42. DOI: 10.1007/s10994-006-6226-1

Gini, C. (1912). Variabilità e mutabilità. Reprinted in: Pizetti, E., & Salvemini, T (eds.). Memorie di metodologica statistica. Rome, Italy: Libreria Eredi Virgilio Veschi.

Gomarasca, M. A., Tornato, A., Spizzichino, D., Valentini, E., Taramelli, A., Satalino, G., Vincini, M., Boschetti, M., Colombo, R., Rossi, L., Borgogno-Mondino, E., Perotti, L., Alberto, W., & Villa, F. (2019). Sentinel for applications in agriculture. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-3/W6, 91-98.

Hlaváčová, I., Kačmařík, M., Lazecký, M., Struhár, J., & Rapant, P. (2021). Automatic open water flood detection from Sentinel-1 multi-temporal imagery. Water, 13(23), 3392. DOI: https://doi.org/10.3390/w13233392

Lin, Y. N., Yun, S., Bhardwaj, A., & Hill, E. M. (2019). Urban flood detection with Sentinel-1 multi-temporal Synthetic Aperture Radar (SAR) observations in a Bayesian framework: A case study for hurricane Matthew. Remote Sensing, 11(15). DOI: 10.3390/rs11151778

MathWorks Inc. (2021). Documentación: cvpartition. Recuperado de https://la.mathworks.com/help/stats/cvpartition.html

MathWorks Inc. (2016). MATrix LABoratory (MATLAB R2016a). Plataforma matemática sumamente potente en la manipulación de matrices y análisis numérico (software). Natick, USA: MathWorks Inc.

Parikh, R., Mathai, A., Parikh, S., Chandra-Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. Indian Journal of Ophthalmology, 56(1), 45-50. DOI: 10.4103/0301-4738.37595

Patle, A., & Chouhan, D. S. (2013). SVM kernel functions for classification. 2013 International Conference on Advances in Technology and Engineering (ICATE), 2013, 1-9. DOI: 10.1109/ICAdTE.2013.6524743

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Perevochtchikova, M., & Lezama-de-la-Torre, J. L. (2010). Causas de un desastre: inundaciones del 2007 en Tabasco, México. Journal of Latin American Geography, 9(2), 73-98.

Podest, E. (16 de abril, 2018). Imágenes Radar (SAR) Seminario NASA - Preprocesamiento, clasificación agua, tierra, Sentinel 1 (2/4) (video). Recuperado de https://youtu.be/pVeh9ChATwA

Pulella, A., Aragão-Santos, R., Sica, F., Posovszky, P., & Rizzoli, P. (2020). Multi-temporal Sentinel-1 backscatter and coherence for rainforest mapping. Remote Sensing, 12(847). DOI: 10.3390/rs12050847

QGIS.org. (2020). QuantumGis (QGIS versión 3.10.7-A Coruña). Sistema de Información Geográfica, proyecto de fundación geoespacial de código abierto (software). Recuperado de https://qgis.org/es/site/

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. Madison, USA: University of Wisconsin.

Raschka, S., & Mirjalili, V. (2017). Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-learn, and TensorFlow (2nd ed.). Birmingham, UK: Packt Publishing Ltd.

Sami, A., & Abdulmunem, M. E. (2020). Synthetic aperture radar image classification: A survey. Iraqi Journal of Science, 61(5), 1223-1232. DOI: 10.24996/ijs.2020.61.5.29.

Sánchez, A. J., Salcedo, M. A., Florido, R., & Mendoza, J. D. (2015). Ciclos de inundación y conservación de servicios ambientales en la cuenca baja de los ríos Grijalva-Usumacinta. Contactos, 97, 5-14.

Shen, X., Wang, D., Mao, K., Anagnostou, E., & Hong, Y. (2019). Inundation extent mapping by synthetic aperture radar: A review. Remote Sensing, 11(879). DOI: 10.3390/rs11070879

Smola, A., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14, 199-222.

Swamynathan, M. (2017). Mastering Machine Learning with Python in Six Steps. Apress. DOI: 10.1007/978-1-4842-2866-1

UN-SPIDER, United Nations Platform for Space-based Information for Disaster Management and Emergency Response. (2020). Step-by-step: Mudslides and associated flood detection using Sentinel-1 data. Recuperado de https://un-spider.org/advisory-support/recommended-practices/mudslides-flood-sentinel-1/step-by-step

Vapnik, V. (1995). The nature of statistical learning theory. New York, USA: Springer.

Zhang, B., Wdowinski, S., Oliver-Cabrera, T., Koirala, R., Jo, M. J., & Osmanoglu, B. (2018). Mapping the extent and magnitude of sever flooding induced by hurricane Irma with multi-temporal sentinel-1 SAR and InSAR observations. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-3(3), 2237-2244. DOI: 10.5194/isprs-archives-XLII-3-2237-2018