

DOI: 10.24850/j-tyca-2025-06-02

Artículos

Estimación de volumen de una presa usando percepción remota y Machine Learning

Estimation of dam volume using remote sensing and Machine Learning

Fernando Flores-Escobar¹, ORCID: <https://orcid.org/0000-0002-6572-1058>

Enrique Palacios-Vélez², ORCID: <https://orcid.org/0000-0002-1716-9377>

Adolfo Antenor Exebio-García³, ORCID: <https://orcid.org/0000-0002-3280-0278>

Demetrio Salvador Fernández-Reynoso⁴, ORCID: <https://orcid.org/0000-0002-1734-3152>

Ramón Valdivia-Alcalá⁵, ORCID: <https://orcid.org/0000-0003-0434-3169>

¹Colegio de Postgraduados, Estado de México, México, fernando231082@gmail.com

²Colegio de Postgraduados, Estado de México, México, epalaciospave80@gmail.com

³Colegio de Postgraduados, Estado de México, México, adolfoantenor@hotmail.com

⁴Colegio de Postgraduados, Estado de México, México, demetrio@colpos.mx

⁵Universidad Autónoma Chapingo, Estado de México, México,
ramvaldi@gmail.com

Autor para correspondencia: Fernando Flores-Escobar,
fernando231082@gmail.com

Resumen

En México, donde el sector más importante de consumo de agua es la agricultura y debido a que es cada vez la mayor la demanda de alimentos, es necesario llevar un adecuado balance entre la producción agrícola y el consumo de agua. En este trabajo, con la ayuda de imágenes Sentinel, se generaron modelos de Machine Learning para estimar la superficie y el volumen de la presa Manuel Ávila Camacho (Valsequillo) del Distrito de Riego 030, que recibe aguas residuales, lo que provocó la proliferación de lirio acuático. Existen índices de agua como NDWI, $NDWI_{Mcfeeter}$, $NDWI_{Gao}$, $NDWI_{Xu}$, $AWEI_{non-shadow}$, $AWEI_{shadow}$ e ICEDEX, que permiten la discriminación de los espejos de agua mediante el uso de imágenes satelitales. Sin embargo, en este caso, el uso de los índices confunde el lirio con la vegetación natural. El entrenamiento se realizó con los valores de los índices mencionados anteriormente y de las capas RGB, NIR y SWIR a nivel pixel, y los resultados obtenidos de la precisión fueron los siguientes: modelo de discriminación lineal, 98.1%; árboles de decisión, 99.2 %; y regresión logística, 98.5 %; con estos modelos fue posible estimar el área del espejo de agua de la presa. Con la gráfica de áreas capacidades de la presa se generó una función cuadrática con una r^2 de 0.9988, en la que el volumen está en función de área y se estimó el

volumen de la presa donde la diferencia promedio es de un 8.5 % con el modelo de discriminación lineal.

Palabras clave: Machine Learning, lirio acuático, estimación de volumen presa, índices de agua, imagen Sentinel 2.

Abstract

In Mexico, where the most significant water consumption sector is agriculture, and due to the increasing demand for food, it is necessary to maintain a proper balance between agricultural production and water consumption. In this study, with the assistance of Sentinel 2 images, Machine Learning models were developed to estimate the surface area and volume of the Manuel Ávila Camacho (Valsequillo) dam in Irrigation District 030, which receives wastewater, leading to the proliferation of aquatic lily. There are water indices such as NDWI, $NDWI_{Mcfeeter}$, $NDWI_{Gao}$, $NDWI_{Xu}$, $AWEI_{non-shadow}$, $AWEI_{shadow}$, and ICEDEX, which enable the discrimination of water bodies through the use of satellite images. However, in this case, the use of these indices results in confusion between the lilies and natural vegetation. The training was conducted using the values of the aforementioned indices and the RGB, NIR, and SWIR layers at the pixel level, and the accuracy results obtained were as follows: Linear discrimination model at 98.1 %, decision trees at 99.2 %, and logistic regression at 98.5 %. With these models, it was possible to estimate the surface area of the dam's water body. Using the graph of the dam's capacity areas, a quadratic function with an R-squared value of 0.9988 was generated, where volume is a function of area, and the dam's volume was estimated, with an average difference of 8.5 % compared to the linear discrimination model.

Keywords: Machine Learning, aquatic lily, dam volume estimation, water indices, Sentinel 2 image.

Recibido: 19/12/2023

Aceptado: 26/09/2024

Publicado *ahead of print*: 08/10/2024

Versión final: 01/11/2025

Introducción

El módulo monitor de presas presenta el volumen almacenado al día de 210 presas del país; cuenta con históricos desde el año 2007.

Existen presas que reciben aportaciones de aguas residuales, lo que implica que el agua sea de mala calidad, con alta concentración de contaminantes, como es el caso de la presa Manuel Ávila Camacho (Valsequillo), ubicada en el Distrito de Riego 030, Puebla. El agua de la presa Valsequillo está clasificada como C3S1, agua altamente salina con bajo contenido de sodio; con estas propiedades, el agua debe tener uso restringido en cultivos poco tolerantes a la salinidad y en suelos con drenaje deficiente (Bonilla, Silva, Cabrera, & Sánchez, 2013); tener alto grado de contaminación ha implicado que la escala de medición de la presa se borre, motivo por el cual no es posible medir de manera precisa el volumen almacenado de agua en la presa; así, se busca una alternativa confiable para cuantificar los volúmenes de agua en la presa usando imágenes satelitales.

En el presente trabajo se propone desarrollar un modelo de clasificación supervisada para cuerpos de agua con lirio, mediante el uso de imágenes captadas por el satélite Sentinel 2 y algoritmos de Machine Learning (M.L.), y de esa forma poder realizar una estimación del volumen de la presa Valsequillo del Distrito de Riego 030. Esto es importante dado que permite ver la disponibilidad del recurso agua en las fuentes de abastecimiento superficiales para planificar adecuadamente los riegos agrícolas.

Antecedentes

La agricultura en México y en el mundo consume aproximadamente el 70% del agua dulce (Conagua, 2018), y se registra cada vez más una fuerte competencia por el uso con el consumo urbano e industrial.

Las presas regulan el flujo del agua en los ríos y pueden tener uno o varios usos a la vez; proporcionan energía hidroeléctrica, además de volúmenes de agua para el riego y el abastecimiento público, o para controlar avenidas. Dentro del uso agrícola es importante conocer la disponibilidad de agua para la planificación adecuada del riego.

México cuenta con el Sistema Nacional de Información del Agua (SINA), un instrumento de gestión de la información estratégica de los recursos hídricos en México a cargo de la Gerencia de Planificación Hídrica de la Dirección General de la Comisión Nacional del Agua (Conagua).

El SINA integra, analiza y brinda información estadística y geográfica del sector hídrico proveniente de diversas áreas de la Conagua y de otras instituciones para apoyar la toma de decisiones en el diseño, implementación y seguimiento de políticas públicas para lograr la

sustentabilidad y seguridad hídrica en las cuencas y acuíferos del país (Conagua, 2013).

La percepción remota o teledetección es una herramienta que permite el estudio de la cubierta terrestre desde las superficies continentales hasta los cuerpos de agua como mares, ríos, etcétera.

La extracción de información útil a partir de imágenes de satélite, por medio de la clasificación, es uno de los problemas técnicos más importantes de la teledetección. El objetivo de la teledetección es la determinación de las características y de los fenómenos que tienen lugar en la superficie de la Tierra a través de su firma espectral, los estudios de teledetección, las variaciones temporales, espaciales y espectrales de la superficie del planeta (Chuvieco, 2002).

Existen índices de vegetación de gran importancia que utilizan la banda roja, el infrarrojo cercano (NIR) e infrarrojo medio (SWIR), como son el índice de vegetación de la diferencia normalizada (NDVI) (Rouse, Hass, Deering, & Schell, 1974); el índice de vegetación ajustado al suelo (SAVI), y el índice de estrés de humedad (MSI), que ayudan en la discriminación de vegetación, pero también ayudan a discriminar los cuerpos de agua en imágenes satelitales. Sin embargo, en los estudios de teledetección, la mayoría de los algoritmos de detección de cuerpos de agua se basan solo en los índices de agua.

Estos índices se han usado para la discriminación de cuerpos de agua. McFeeters (1996) diseñó un índice de agua popular: el índice de agua de diferencia normalizada (NDWI) para el mapeo de agua a partir de imágenes satelitales. Se ha utilizado el infrarrojo cercano (NIR) y la banda verde del Landsat Thematic Mapper (TM) para representar las características del agua. Xu (2006) modificó el NDWI reemplazando la

banda NIR por infrarrojos de onda corta (SWIR) y lo denominó índice de agua de diferencia normalizada modificada (MNDWI). MNDWI redujo parcialmente la tasa de error generada por el suelo, la vegetación y las áreas urbanizadas. Feyisa, Meilby, Fensholt y Proud (2014) presentaron el índice de extracción de agua automatizada (AWEI) para atender la clasificación errónea de la sombra como agua mediante el uso de bandas multiespectrales. Se creó un nuevo índice de agua con análisis discriminante lineal de Fisher (Fisher, Flood, & Danaher, 2016) revisado utilizando cinco bandas de reflectancia superficial (SR) de Landsat. También proporcionó una comparación exhaustiva de los índices de agua para las imágenes de Landsat.

Fisher *et al.* (2016) evaluaron la precisión de siete métodos de índice de agua para clasificar el agua en imágenes Landsat TM/ETM+/OLI de resolución de 30 m del este de Australia. Estos índices fueron el índice de extracción de agua automatizado para imágenes con sombras (AWEI_{shadow}) y sin sombras (AWEI_{non-shadow}); el índice de humedad *tasseled cap* (TCW_{Crist}); dos variaciones del índice de agua de diferencia normalizada (NDWI_{McFeeters} y NDWI_{Xu}); un índice de agua creado mediante el análisis de variables canónicas a partir de datos de la parte superior de la atmósfera WI₂₀₀₆, y un nuevo índice de agua creado con análisis discriminante lineal a partir de datos procesados hasta la reflectancia de la superficie WI₂₀₁₅. En general, el WI₂₀₁₅, WI₂₀₀₆ y WEI_{shadow} se desempeñaron mejor; todos los índices lograron precisiones generales de 95 a 99 % para píxeles puros, y de 73 a 75 % para píxeles mixtos.

Otro índice es el ICEDEX (Índice del Centro de Estudios y Experimentación de Obras Públicas) de España. Está destinado al mapeo de aguas continentales, en que se busca reducir la influencia de

vegetación y la presencia de material en suspensión (sedimentos, algas) de acuerdo con Ariza, García, Rojas y Ramírez (2013).

Chul, Hun y Yeal (2015) proponen un nuevo método de clasificación de masas de agua utilizando la reflectancia de la parte superior de la atmósfera (TOA) y los índices de agua (WI) del sensor Landsat 8 Operational Land Imager (OLI), y sus correspondientes clasificadores de bosques aleatorios potenciados (BRF); demostraron que el rendimiento de la combinación propuesta de dos clasificadores BRF muestra resultados sólidos de clasificación de cuerpos de agua, independientemente de la topología, las propiedades del río y el entorno de fondo.

La teledetección y la inteligencia artificial en los últimos años están desarrollando técnicas para el procesamiento de imágenes satelitales y su interpretación en diferentes ámbitos, como el estudio de coberturas, la agricultura de precisión y problemas ambientales. Esto también es posible con el uso de tecnologías y procesadores de algoritmos de Machine Learning (ML) con eficiencia computacional y el acceso a datos accesibles de plataformas satelitales como Landsat y Sentinel para fines de investigación (Chucos & Vega, 2022).

Los índices de agua de manera individual discriminan de forma muy exacta los cuerpos de agua. Sin embargo, cuando existen aguas residuales que permiten la proliferación de lirio acuático u otro tipo de vegetación, dicha discriminación suele confundirse con la vegetación natural, es por eso que para esta investigación se utilizó el entrenamiento algoritmos de Machine Learning mediante el uso de la inteligencia artificial para disminuir dicho error.

La automatización de clasificación de cuerpos de agua mediante el uso de imágenes para identificar las fuentes de abastecimiento (presas)

de los distritos de riego (DR) es un apoyo en la operación del DR, que permitirá tener un estimado de la disponibilidad del recurso hídrico.

Materiales y métodos

El Distrito de Riego 030, Valsequillo, se ubica en el centro-sureste del estado de Puebla (Figura 1), entre los paralelos 18° 28' 37" y 18° 53' 49" latitud norte y 97° 27' 19" y 97° 55' 57" longitud oeste.



Figura 1. Ubicación del Distrito de Riego 030 y presa Valsequillo.

La presa de almacenamiento Manuel Ávila Camacho (Valsequillo) está situada en las coordenadas 18° 54' 45.23" latitud norte y 98° 06' 30.08" longitud oeste de Greenwich, sobre el río Atoyac (Figura 1); se localiza en el municipio de Puebla, Puebla. La construcción fue terminada el 13 de diciembre de 1946 para almacenar los escurrimientos naturales del río Atoyac. El área de la cuenca es de 3 954 km², con un escurrimiento medio anual de 296.448 Millones de m³.

La presa de almacenamiento es de tipo mixta, con una altura de 83 m; tiene una longitud de corona de 425 m, construida con materiales de tierra compacta y roca. Su vaso tiene una capacidad total de 304.246 Millones de m³, una capacidad útil 294.502 Millones de m³, capacidad de azolves de 9.744 Millones de m³. La obra de toma es de tipo torre, con una carga máxima de 29.31 m, un gasto máximo de 50 m³/s y una superficie regable de 25.942 ha. Tiene un vertedor tipo Creager con longitud de cresta de 170 m y una capacidad máxima de descarga de 2 000 m³/s.

La última información proporcionada por el distrito de riego de batimetría de la presa Valsequillo es del 2008, de un levantamiento topobatimétrico que se presenta en la Figura 2; la capacidad de la presa en la actualidad ha disminuido debido a los azolves.

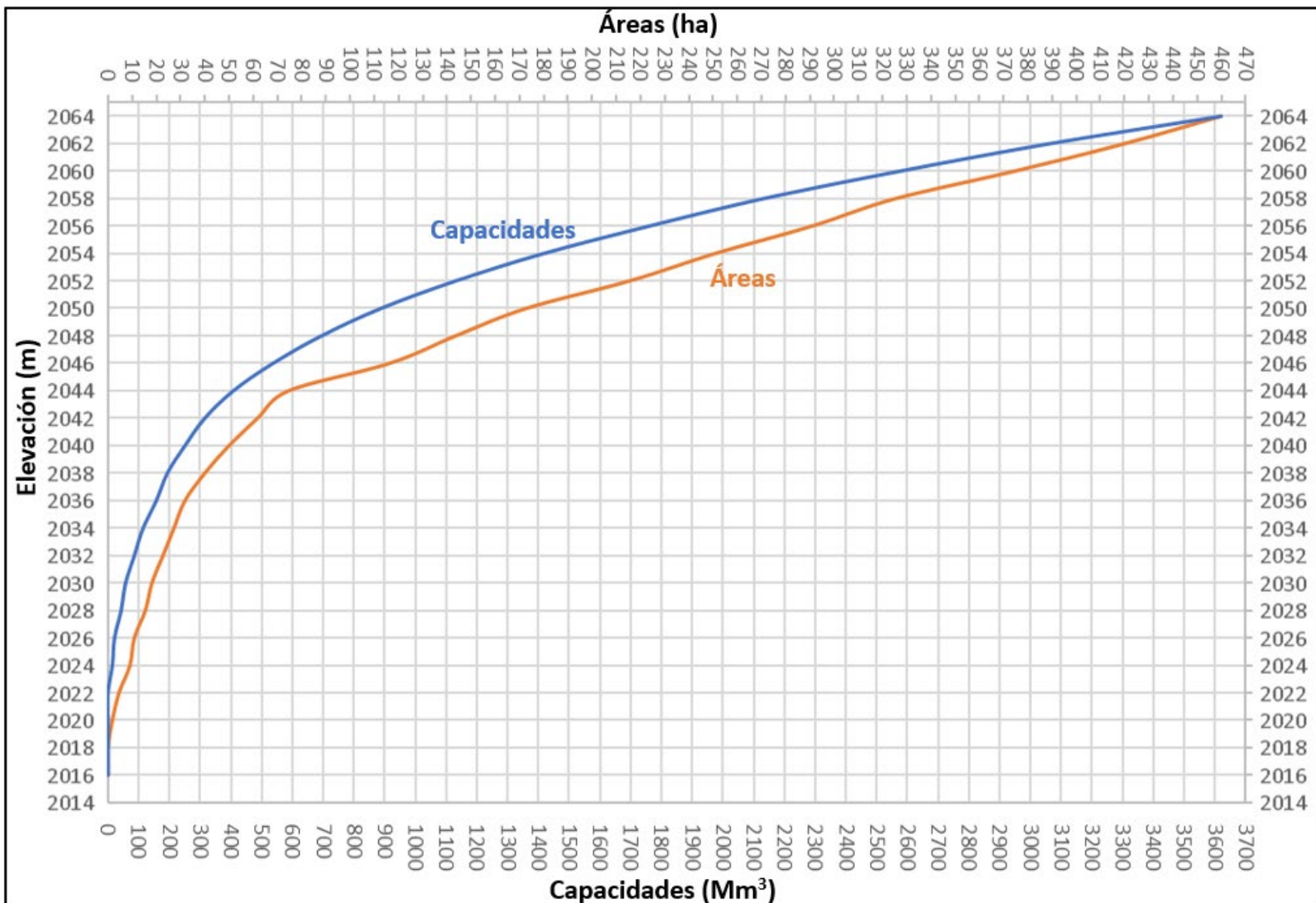


Figura 2. Información topobatimétrica de la presa Manuel Ávila Camacho (Valsequillo) en 2008.

Para el presente trabajo se descargaron imágenes del Sentinel 2 que estuvieran libres de nubosidad dentro del área donde se realizó el trabajo. Actualmente la escala de medición de la presa se ha borrado debido a la mala calidad del agua, por lo que se utilizaron imágenes del

2015-2016, para poder comparar los datos obtenidos del modelo con los datos reportados por la Comisión Nacional del Agua.

Fechas de las imágenes Sentinel obtenidas:

- 06/08/2015
- 05/10/2015
- 04/12/2015
- 03/01/2016
- 28/11/2016
- 28/12/2016

Para las imágenes descargadas se hizo corrección atmosférica y transformación de datos utilizando el método DOS (Dark Object Substraction) con la ayuda del *software* Qgis y el complemento Semi-Automatic Classification Plugin v6 (SCP6), y se recortó el área de trabajo (Figura 3, Figura 4, Figura 5, Figura 6, Figura 7 y Figura 8).



Figura 3. Imagen Sentinel 2 del área de estudio de fecha 06/08/2015.

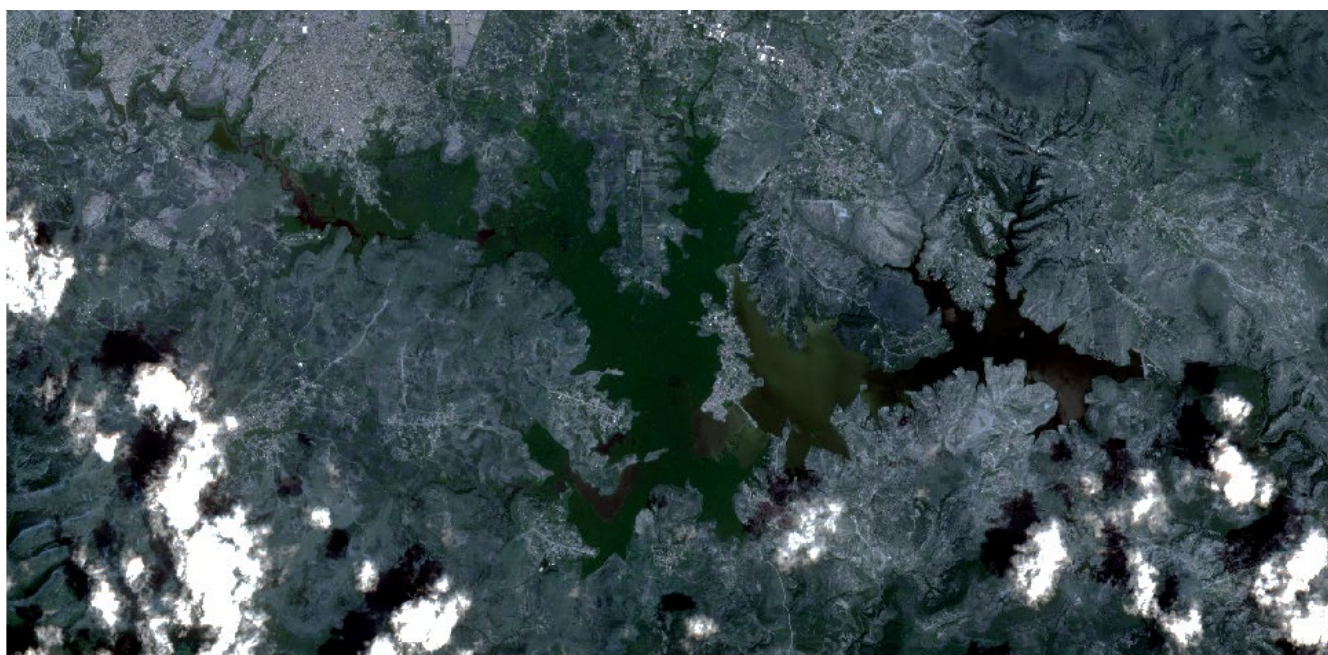


Figura 4. Imagen Sentinel 2 del área de estudio de fecha 05/10/2015.



Figura 5. Imagen Sentinel 2 del área de estudio de fecha 04/12/2015.

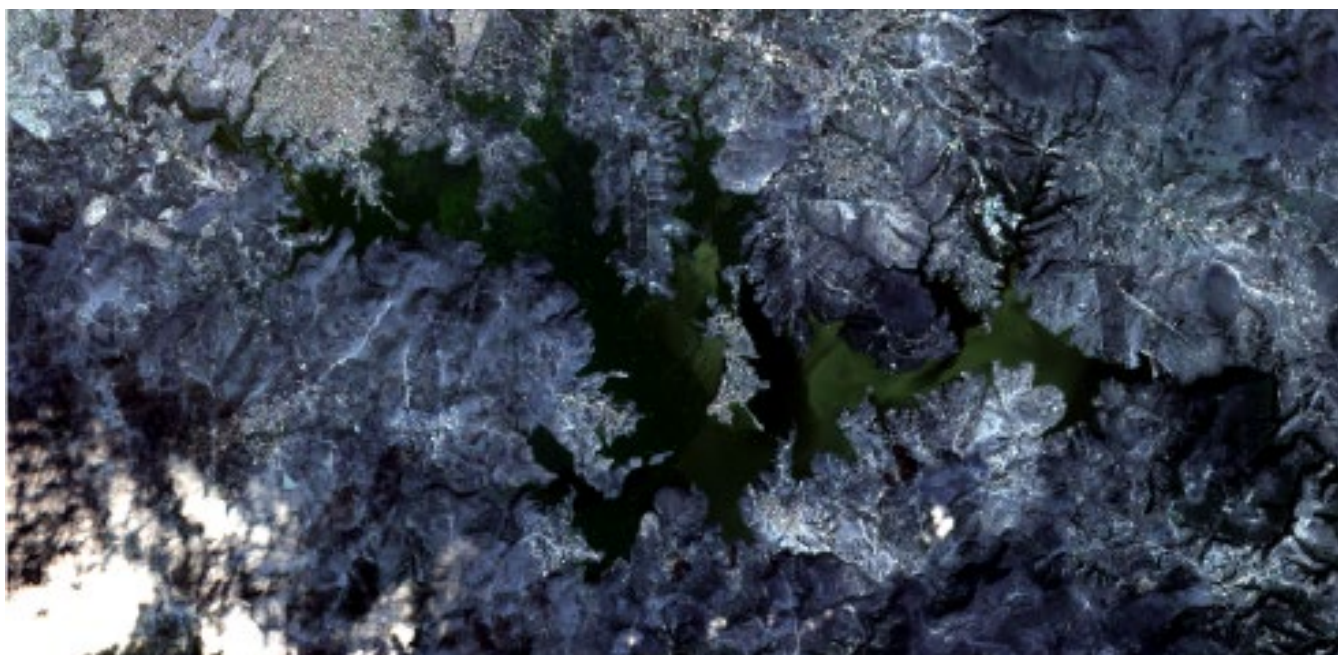


Figura 6. Imagen Sentinel 2 del área de estudio de fecha 03/01/2016.



Figura 7. Imagen Sentinel 2 del área de estudio de fecha 28/11/2016.

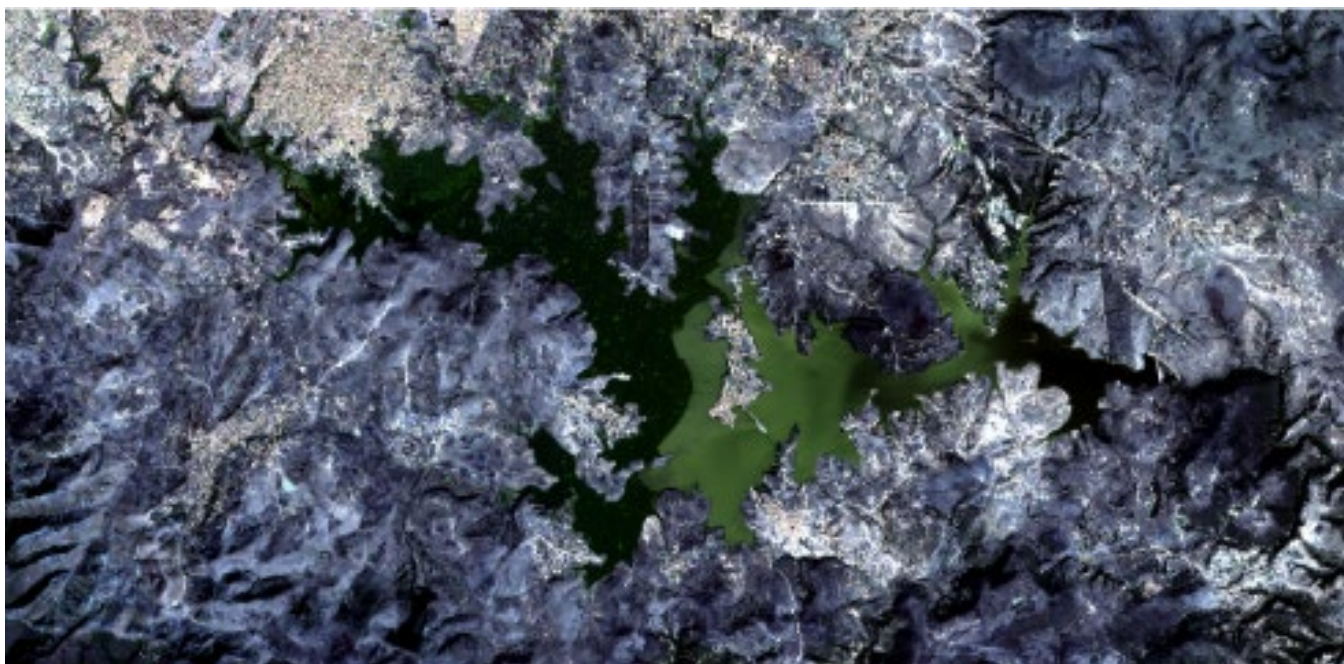


Figura 8. Imagen Sentinel 2 del área de estudio de fecha 28/12/2016.

Para el entrenamiento de los modelos de ML se realizó la extracción de características de las imágenes; se dividieron en dos clases: 1) clase no agua y 2) clase agua (Figura 9).

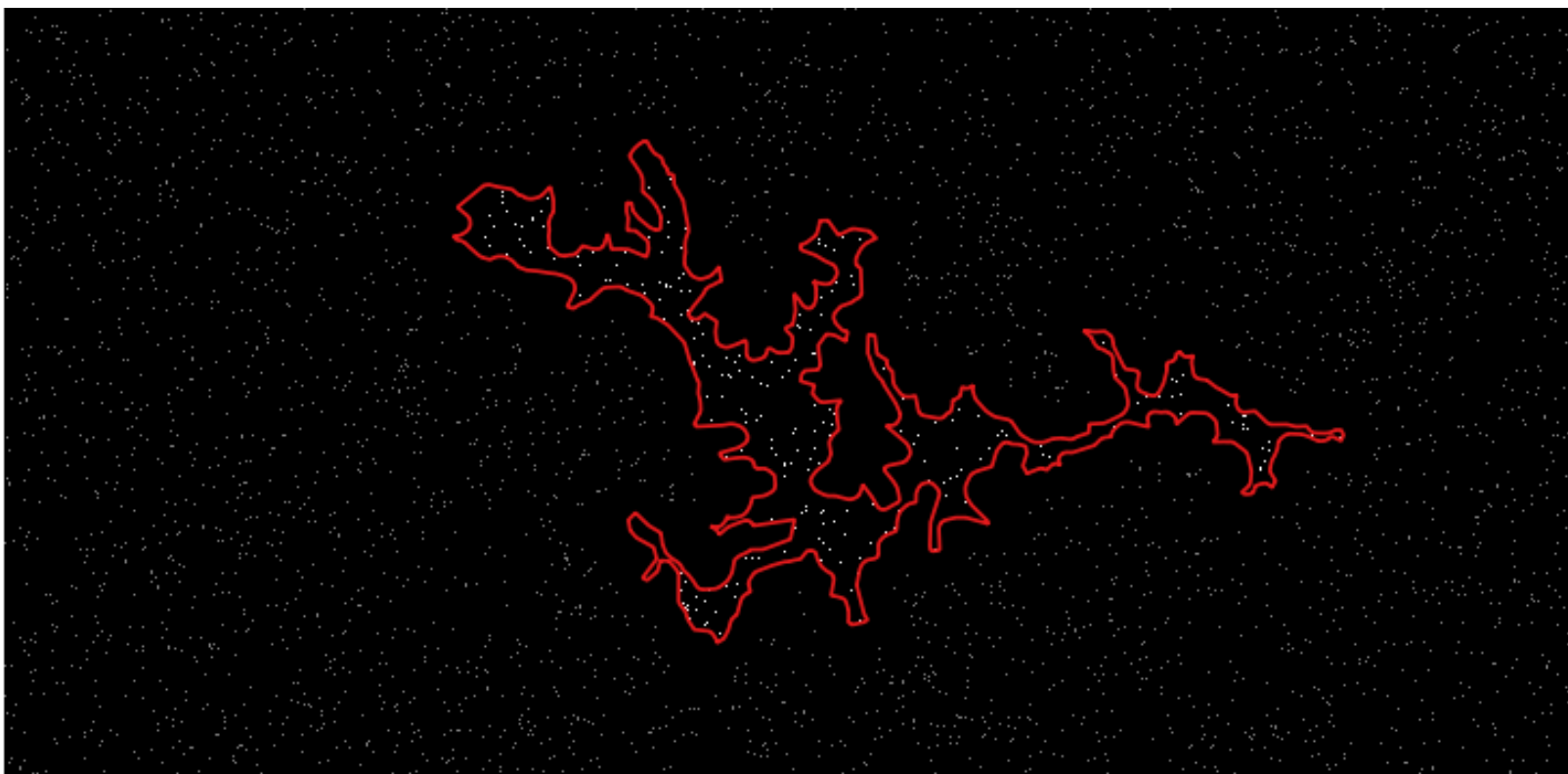


Figura 9. Píxeles muestra para el modelo de clasificación del área de la presa.

La extracción de características se realizó a nivel píxel para cada una de las seis imágenes, y se utilizaron los valores de las bandas y los índices de agua encontrados en la literatura para entrenar los modelos de Machine Learning. Para la extracción de características se programó un script en Matlab y poder trabajar con los datos de los pixeles seleccionados como muestras.

Se extrajeron para cada clase los valores por pixel de las siguientes bandas:

- Banda roja (RED).
- Banda verde (GREEN).
- Banda azul (BLUE).
- Banda infrarroja cercana (NIR)
- Banda infrarroja de onda corta 1 (SWIR-1)
- Banda infrarroja de onda corta 2 (SWIR-2)

Se extrajeron los valores por pixel de los siguientes índices de agua para cada clase:

- Índice de vegetación de diferencia normalizada (NDVI) (Rouse *et al.*, 1974):

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

- Índice de agua de diferencia normalizada (NDWI_{McFeeter}) (McFeeters, 1996):

$$NDWI_{Mcfeeter} = \frac{GREEN-NIR}{GREEN+NIR} \quad (2)$$

- Índice de agua de diferencia normalizada (NDWI_{Gao}) (Gao, 1996):

$$NDWI_{Gao} = \frac{NIR - SWIR1}{NIR + SWIR1} \quad (3)$$

- Índice de agua de diferencia normalizada modificada (MNDWI_{Xu}) (Xu, 2006):

$$MNDWI_{Xu} = \frac{GREEN - SWIR1}{GREEN + SWIR1} \quad (4)$$

- Índice de agua de diferencia normalizada modificada (MNDWI) (Xu, 2006):

$$MNDWI = \frac{GREEN - SWIR2}{GREEN + SWIR2} \quad (5)$$

- Índice de extracción de agua automatizada (AWEI_{non-shadow}) (Fisher *et al.*, 2016):

$$AWEI_{non-shadow} = 4 \times (GREEN - SWIR1) - 0.25 \times NIR - 2.75 \times SWIR1 \quad (6)$$

- Índice de extracción de agua automatizada (AWEI_{shadow}) (Fisher *et al.*, 2016):

$$AWEI_{shadow} = BLUE + 2.5 \times GREEN - 1.5 \times (NIR - SWIR1) - 0.25 \times SWIR2 \quad (7)$$

- Índice de agua ICEDEX (Ariza *et al.*, 2013):

$$ICEDEX = \frac{NIR}{RED} - \frac{NIR}{SWIR1} \quad (8)$$

Para entrenar y validar los modelos se utilizó APP Classification Learner de Matlab, con 14 variables, para cada imagen.

Para el entrenamiento del modelo se utilizaron tres métodos:

- Discriminación lineal
- Árboles de decisión
- Regresión logística

Para la validación se eligió la opción de validación cruzada con cinco conjuntos de datos para validación.

Con los modelos generados se estimaron las superficies del espejo de agua de la presa para las fechas de las imágenes.

Se calcularon los índices de vegetación de manera individual y se hizo una comparación de los resultados obtenidos mediante los modelos de ML

Se graficaron los datos de las áreas y las capacidades obtenidos de la Figura 2, y se obtuvo una función cuadrática con una r^2 de 0.9988, como se muestra en la Figura 10.

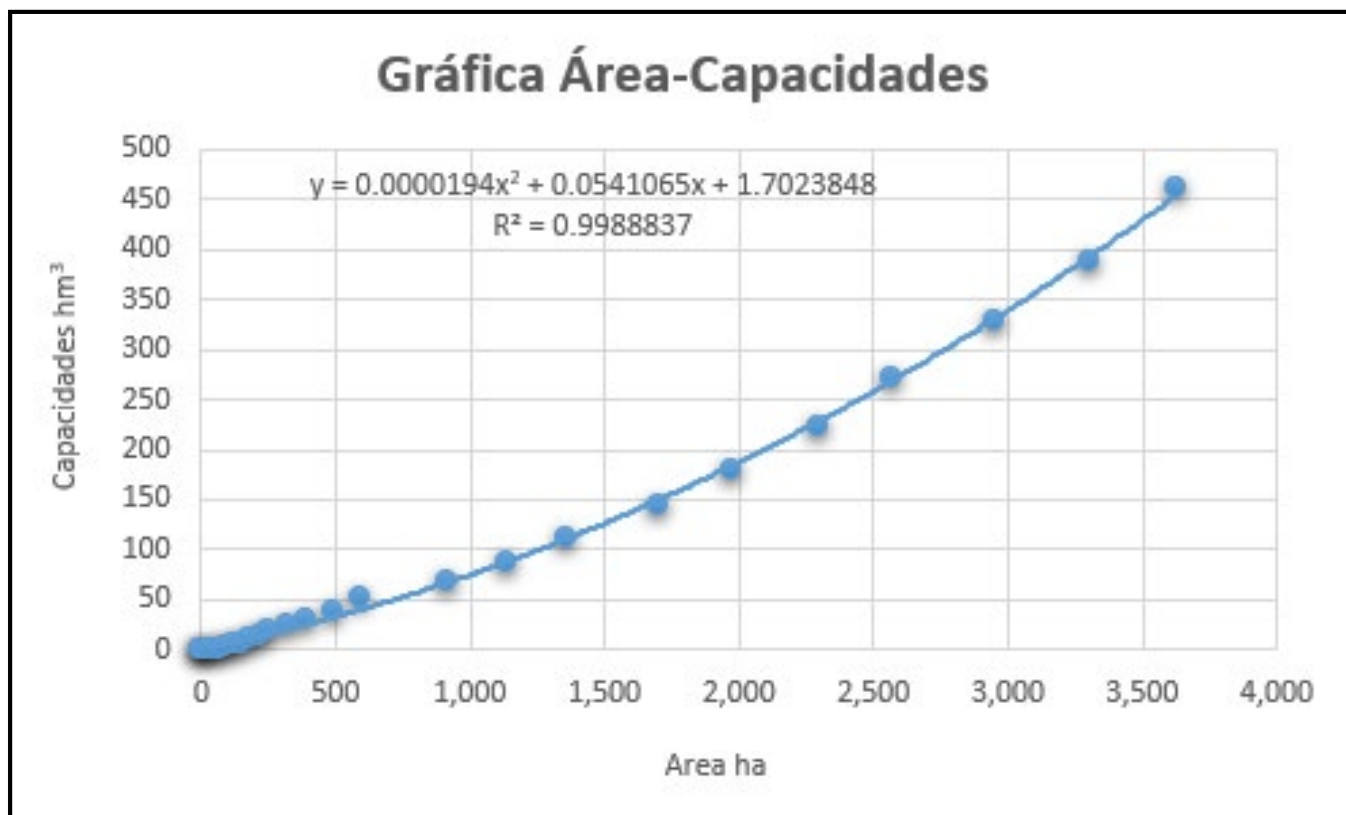


Figura 10. Gráfica área, capacidades de la presa Valsequillo.

Con la función obtenida y el área estimada mediante los modelos de ML se estimó el volumen almacenado para las fechas de las imágenes seleccionadas y se realizó un comparativo con las publicadas por la Comisión Nacional del Agua.

En resumen, se presenta un esquema de la metodología empleada en la Figura 11.

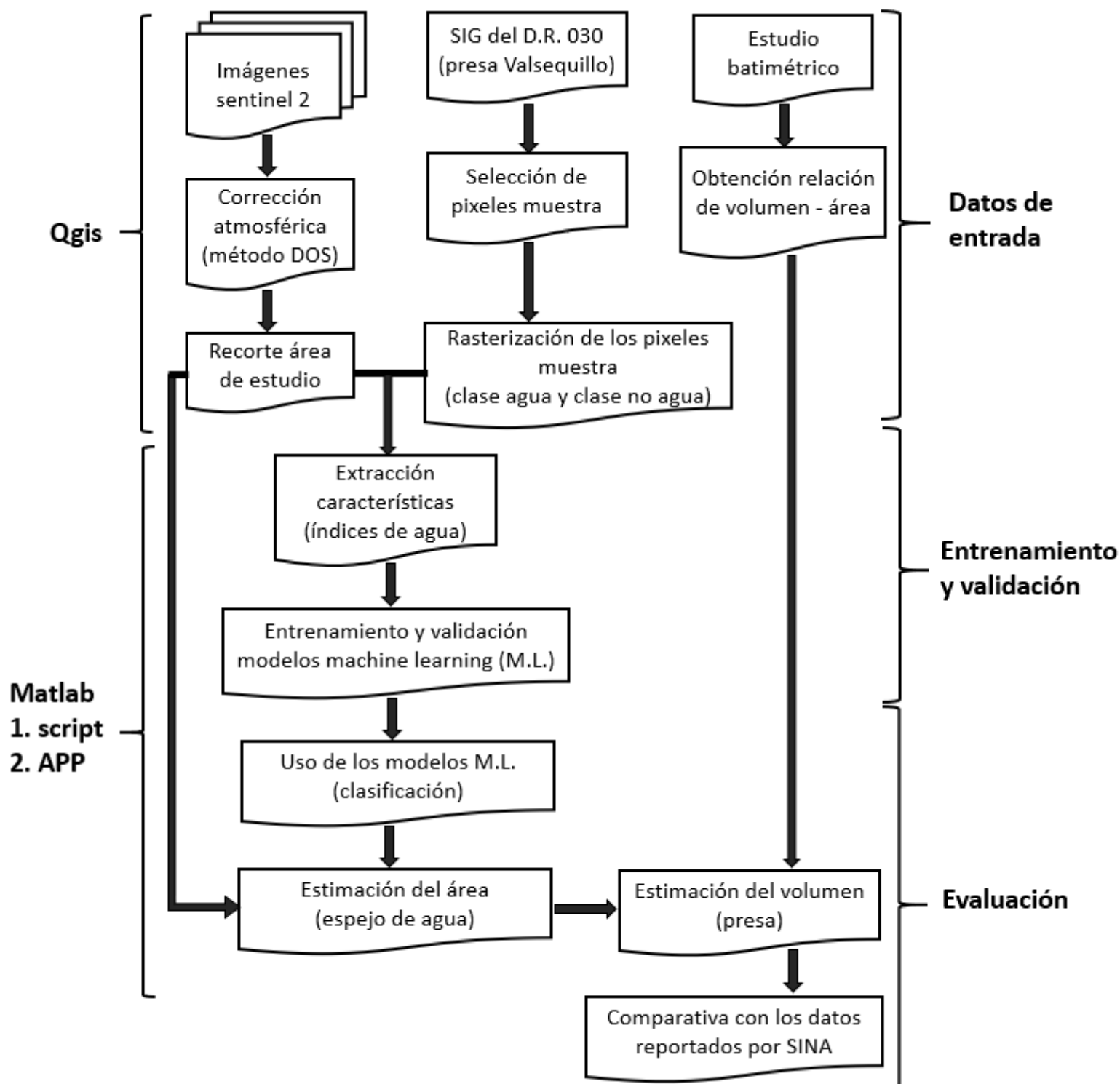


Figura 11. Metodología empleada para la estimación de volumen de la presa Valsequillo del D.R. 030.

Esta metodología de estimación de volumen mediante el uso de modelos ML e imágenes satelitales se aplicó también para la presa Solís ubicada en el cauce del río Lerma en el municipio de Acámbaro, Guanajuato, perteneciente a la infraestructura del Distrito de Riego 011 y con una capacidad útil de almacenamiento de 728 Millones de m³, con la finalidad de tener otro ejemplo para su validación. Para la estimación se usaron imágenes Sentinel 2 de tres fechas diferentes (Figura 12):

- 15/12/2021
- 30/03/2022
- 15/12/2022



Imagen 15/12/2021

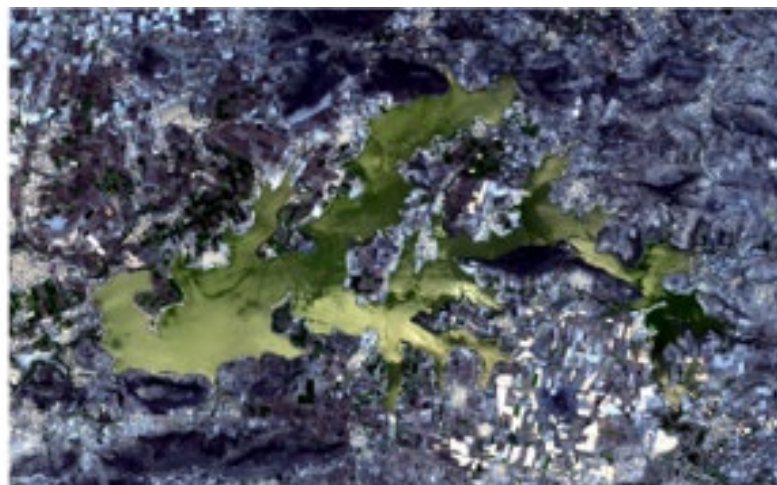


Imagen 30/03/2022



Imagen 15/12/2022

Figura 12. Imagen Sentinel 2 del área de la presa Solís.

Resultados

Se calcularon los índices de NDWI, MNDWI y NDVI de manera individual para discriminar los cuerpos de agua; sin embargo, en este caso, con aguas residuales y lirio acuático, la discriminación se confunde con la vegetación natural (Figura 13).

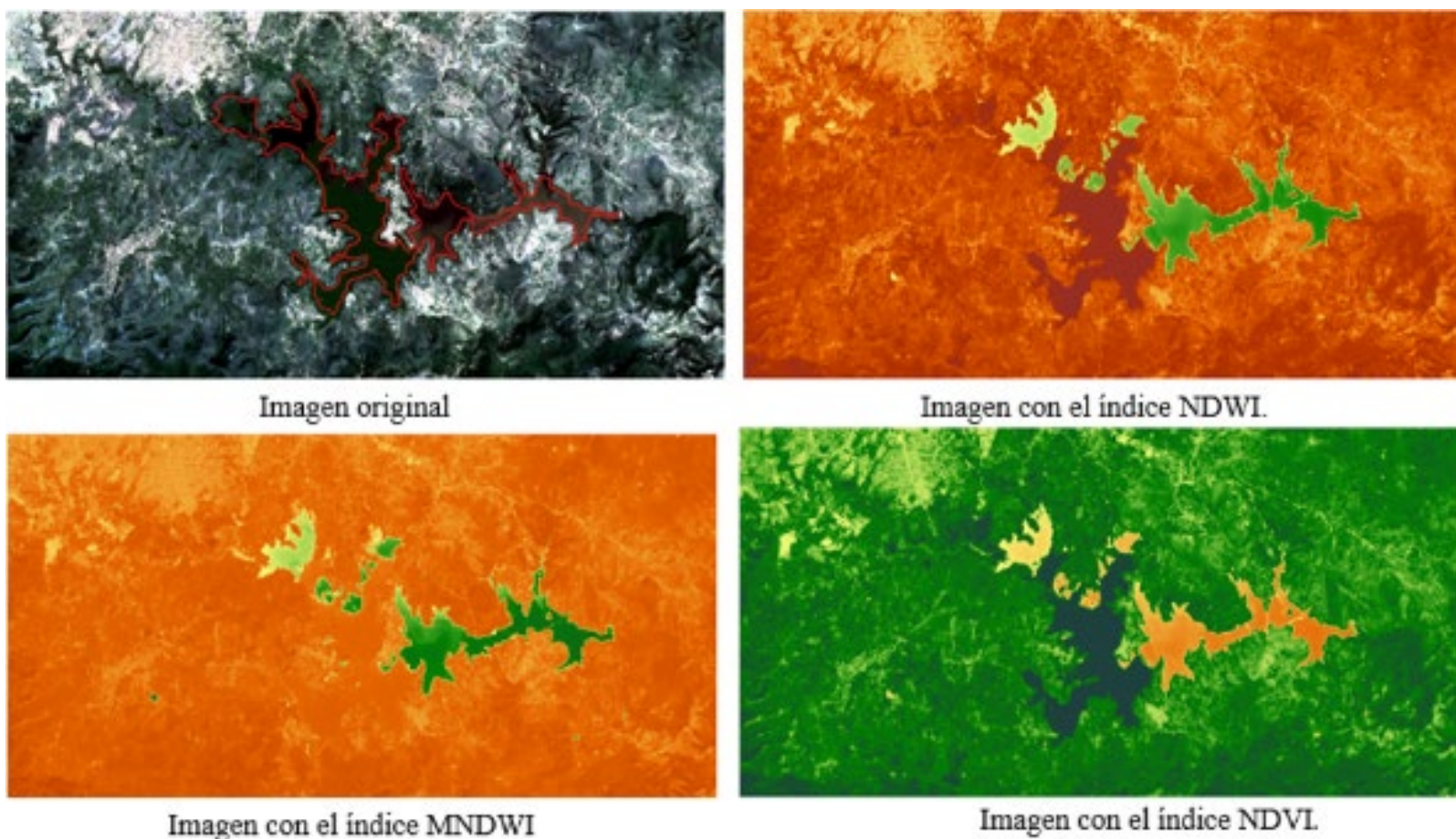


Figura 13. Índices calculados para discriminar la presa Valsequillo.

De los resultados obtenidos de los modelos generados mediante Machine Learning para la discriminar cuerpos de agua, el modelo que mejor precisión tuvo fue el de árboles de decisión con un 99.2 %, como se muestra en la Tabla 1 y en la Figura 14, Figura 15, y Figura 16.

Tabla 1. Precisión de los modelos generados con la APP Classification Learner de Matlab (presa Valsequillo).

Modelo	Precisión (%)
Discriminación lineal (DL)	98.1
Árboles de decisión (AD)	99.2
Regresión logística (RL)	98.5

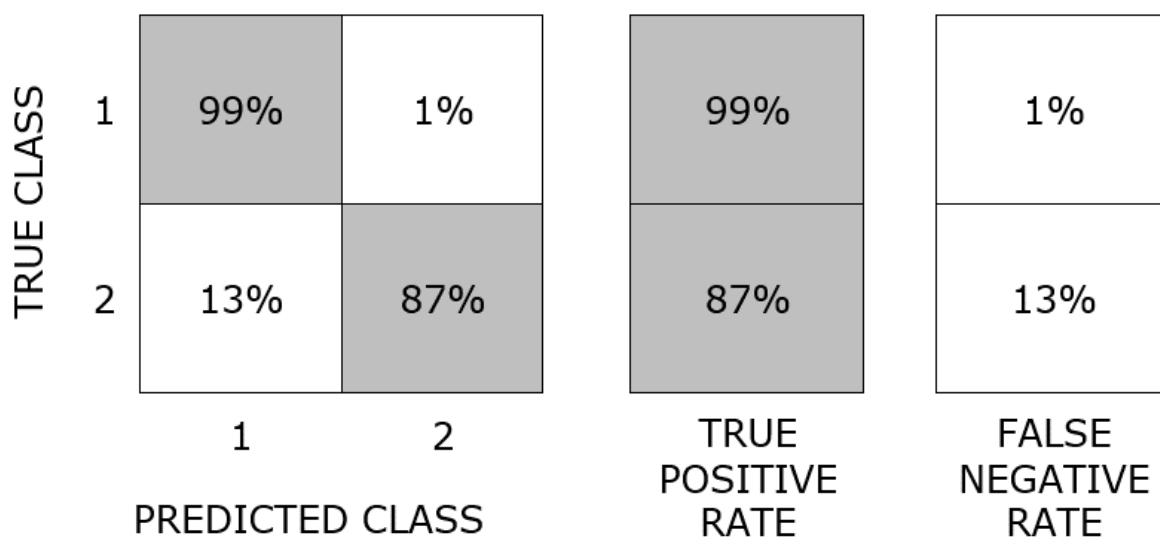


Figura 14. Gráfica de *true positive-false negative* de discriminación lineal.

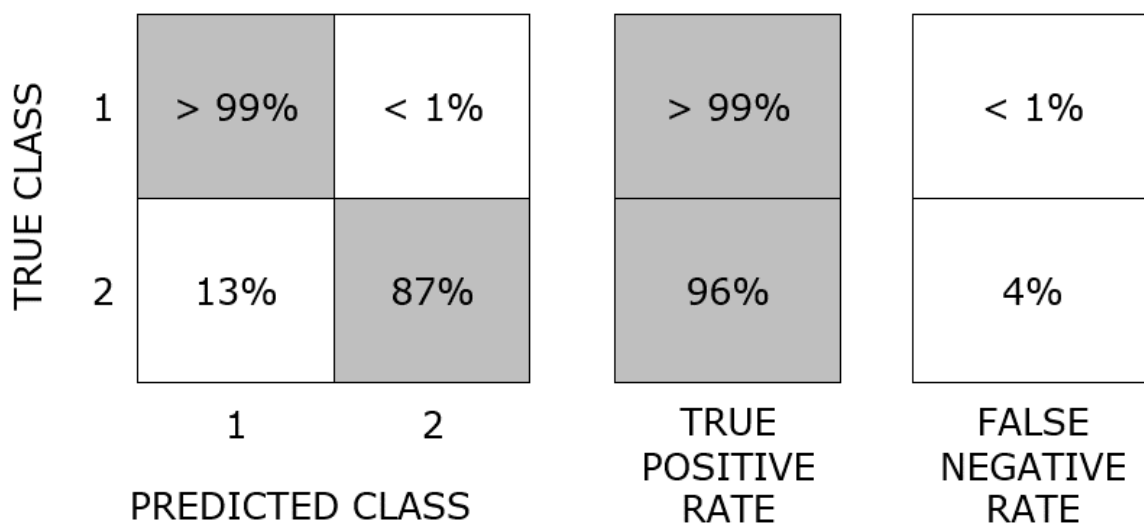


Figura 15. Gráfica de *true positive-false negative* de árboles de decisión.



Figura 16. Gráfica de *true positive-false negative* de regresión logística.

Los resultados son parecidos a los encontrados en la literatura, con la diferencia de que no incluían vegetación en los cuerpos de agua como en este caso la presa Valsequillo, que presenta lirio acuático. Y que,

utilizando los mismos índices de manera individual, en este caso en particular la discriminación no es buena.

Se utilizaron las imágenes de la presa Valsequillo para implementar el modelo de árboles de decisión para realizar una clasificación y se obtuvieron los resultados que se muestran en la Figura 17.

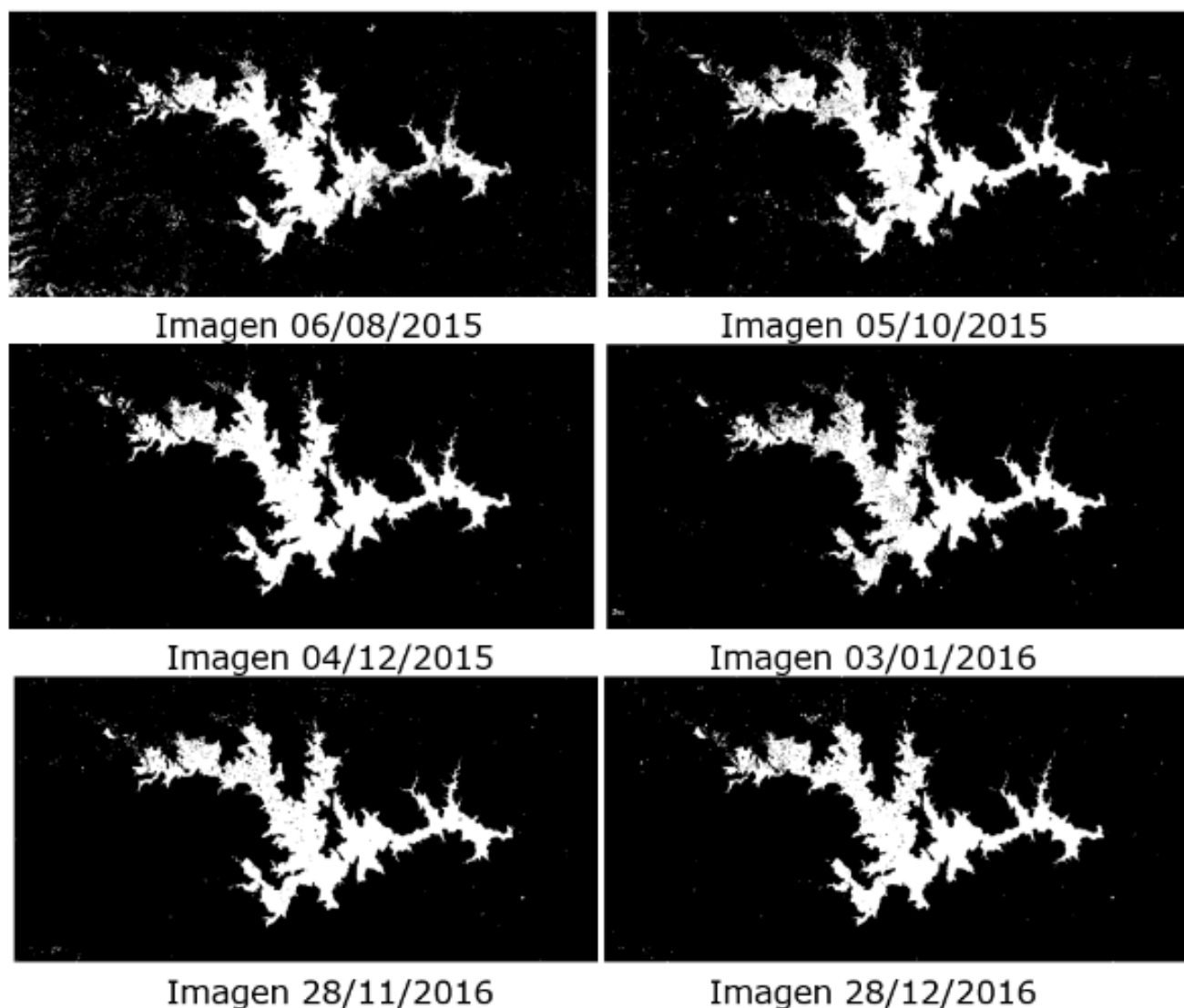


Figura 17. Clasificación de imágenes con el modelo obtenido para la presa Valsequillo (árboles de decisión).

Con los datos obtenidos de los modelos de clasificación supervisada, el tamaño de pixel (10 x 10 m) y el uso de un *script*, se estimó el área de la superficie libre del agua (Tabla 2). Con la ecuación obtenida a partir de la gráfica de área-capacidades (Figura 10) y el área estimada se estimó el volumen de la presa para cada imagen (Tabla 3):

$$Volumen = 0.0000194 * Area^2 + 0.0541065 * Area + 1.7023848$$

Los datos son área estimada mediante ML en ha y volumen en hm³.

Tabla 2. Área estimada de la presa Valsequillo con los modelos Machine Learning.

Imagen	Área estimada (DL) ha	Área estimada (AD) ha	Área estimada (RL) ha
06/08/2015	2 416.1	2 589.9	2 830.6
05/10/2015	2 501.9	2 773.2	2 684.0
04/12/2015	2 541.8	2 674.4	2 628.2
28/11/2016	2 492.1	2 577.8	2 593.5
28/12/2016	2 270.7	2 557.1	2 511.4

Tabla 3. Volumen estimado de la presa Valsequillo con la ecuación de área capacidades.

Imagen	Vol. estimado (DL) hm ³	Vol. estimado (AD) hm ³	Vol. estimado (RL) hm ³
06/08/2015	245.68	265.65	310.29
05/10/2015	258.50	300.94	286.68
04/12/2015	264.57	285.15	277.90
28/11/2016	257.02	270.09	272.51
28/12/2016	224.58	266.91	259.94

Se aplicó el modelo para realizar una clasificación y estimar el volumen almacenado en la presa Solís, donde se obtuvieron los resultados que se muestran en la Figura 18 y Tabla 4, el volumen de la presa Solís de acuerdo con su estudio batimétrico se calculó con la siguiente ecuación, con una r^2 de 0.9995:

$$Volumen = 0.0000197 * Area^2 + 0.0120685 * Area + 4.3967922$$

Los datos son área estimada mediante ML en ha y volumen en hm³.



Imagen 15/12/2021



Imagen 30/03/2022



Imagen 15/12/2022

Figura 18. Clasificación de imágenes con el modelo obtenido para la presa Solís.

Tabla 4. Volumen estimado de la presa Solís por algoritmo ML y el reportado en el Sistema Nacional de Información del Agua por la Comisión Nacional del Agua.

Imagen	Área estimada ha	Volumen estimado hm ³	Volumen reportado hm ³	Diferencia %
06/08/2015	5 971.21	778.87	809.00	-3.72
05/10/2015	5 217.60	603.67	653.06	-7.42
28/12/2016	5 081.64	574.44	653.00	-12.05

Discusión

Se puede observar que la superficie de agua de la presa se discrimina con mayor exactitud con los modelos generados de Machine Learning con precisiones mayores al 99 % (Figura 17), que usando individualmente los índices de agua y el NDVI (Figura 13); esto, debido a la mala calidad del agua, donde el lirio en la presa se confunde y clasifica como vegetación.

En los volúmenes estimados para la presa Valsequillo comparados con las estadísticas reportadas de la Conagua en la página del SINA (Tabla 5) se observa que la diferencias respecto a la información reportada oficial tiene una variación promedio en la imágenes analizadas del 16.12 % para el área estimada mediante discriminación lineal; 6.89 % para el área estimada mediante árboles de decisión, y el 10.99 % para el área estimada mediante regresión logística, lo que indica que el modelo de árboles de decisión es con el que se tiene una mejor estimación del volumen almacenado en la presa. Para los volúmenes estimados de la presa Solís, la variación fue de un 7.13 % en promedio por debajo del reportado en el SINA; estas diferencias pueden ser por el lirio que se

encuentra en las orillas del espejo de agua, por errores reportados en la medición del volumen almacenado en la presa, pero principalmente por el tamaño de pixel, dado que, como se puede observar, la presa Solís no cuenta con problemas de lirio acuático y el error es muy parecido al de la presa Valsequillo. Para mejores resultados se recomienda realizar el análisis con imágenes de mayor resolución espacial, y mediciones en campo con un estudio batimétrico actualizado y nueva escala.

Tabla 5. Volumen almacenado de la presa Valsequillo reportado en el Sistema Nacional de Información del Agua por la Comisión Nacional del Agua.

Imagen	Vol. reportado hm ³	Diferencia (DL) (%)	Diferencia (AD) (%)	Diferencia (RL) (%)
06/08/2015	271.66	-9.56	-2.21	14.22
05/10/2015	306.93	-15.78	-1.95	-6.60
04/12/2015	306.39	-13.65	-6.93	-9.30
28/11/2016	270.09	-15.53	-11.23	-10.44
28/12/2016	266.91	-26.06	-12.12	-14.41

El modelo que mejor precisión tuvo para la clasificación supervisada de cuerpos de agua fue el de árboles de decisión, con una precisión del 99.2 %, por lo que podemos definir que las variaciones de las áreas y los volúmenes de agua estimados se deben principalmente al tamaño de pixel, que en este caso en particular es de 10 x 10 m.

Estos resultados podrían implementarse como una alternativa de medición en los procesos de operación para todo tipo de presa. Como se observó, los resultados obtenidos en la presa Valsequillo son parecidos a

los obtenidos en la presa Solís. Estos modelos que usan algoritmos ML, imágenes satélites e incluso podrían usar las imágenes de drones, también pueden ser utilizados para el seguimiento geoespacial y multitemporal de las variaciones del área de otros cuerpos de agua, como son lagos o lagunas que cuenten con algún tipo de vegetación, como el lirio acuático, ya que son de gran importancias para los ecosistemas y el desarrollo socioeconómico local y regional, pues permiten, entre otras cosas: equilibrio ecológico; protección de las sequías e inundaciones; hábitat único para muchas plantas y animales; agricultura; pesca, e industria, incluso para realizar estudios donde se requiera información de años pasados.

Las imágenes satelitales proporcionan datos visuales actualizados y detallados sobre la superficie del agua en las presas. Estas imágenes, combinadas con algoritmos de Machine Learning, pueden ser procesadas para calcular el área superficial del embalse y, con información topográfica y/o batimétrica, estimar de manera más precisa el volumen de agua almacenado. Como alternativa al estudio batimétrico se pueden realizar levantamientos topográficos LIDAR (Light Detection and Ranging) mediante el uso de drones.

La precisión de estas estimaciones será con base en la resolución de las imágenes y el estudio de batimetría actualizado.

Las imágenes satelitales también proporcionan información valiosa sobre los cambios en la cuenca hidrográfica, como la deforestación o urbanización, que pueden afectar el volumen de agua que llega a la presa. Machine Learning puede integrar estos datos en modelos de simulación para prever los impactos a largo plazo de estas alteraciones, ayudando a los operadores a planificar de manera más eficiente la gestión del recurso hídrico.

Conclusiones

La identificación de los cuerpos de agua es muy importante para diversas estimaciones científicas, así como para la resolución de problemas sociales. Se han desarrollado muchos métodos y se están explorando nuevos enfoques, sin embargo, existen casos donde la resolución espectral y espacial son determinantes para la exactitud del área de cuerpos de agua identificados. En el presente trabajo se obtuvieron buenos resultados en el entrenamiento de los modelos de Machine Learning, con precisiones del 98.1 % para el modelo de discriminación lineal, 99.2 % para el de árboles de decisión y del 98.5 % para el de regresión logística, para una clasificación supervisada de cuerpos de agua con lirio acuático. El error fue mayor para obtener estimaciones en el volumen de agua en una presa, donde obtuvimos variaciones en promedio del 16.12, 6.89 y 10.99 %, respectivamente, para los modelos utilizados; estos resultados podrían ser más precisos si se realizaran con imágenes de mayor resolución espacial y con mediciones actualizadas en la presa. Es recomendable actualizar el estudio batimétrico para obtener datos más apegados a la situación actual de la presa.

El Machine Learning ha logrado con éxito plantear modelos en varias ramas, incluyendo el de los recursos hídricos que ayudan en la optimización de recursos y procesos; estos modelos, basados en ML, representan un ahorro sustancial en tiempo y medios económicos, de tal manera que automatizando toda la metodología implementada en este trabajo y con el uso de imágenes satelitales se puede obtener información que permita estimar como una opción alterna los volúmenes de agua en una presa sin importar las condiciones de la calidad del agua.

Para mejores resultados se recomienda utilizar imágenes satelitales de mayor resolución y actualizar los estudios batimétricos, o el uso drones para captar imágenes de mayor resolución, e implementar levantamientos topográficos LIDAR.

Agradecimientos

Agradecimiento al Consejo Nacional de Humanidades, Ciencia y Tecnología (Conahcyt) y al Colegio de Posgraduados por el apoyo otorgado.

Referencias

- Ariza, A., García, J. S., Rojas, B. S., & Ramírez, D. M. (2013). *Desarrollo de un modelo de corrección de imágenes de satélite para inundaciones: (CAIN, corrección atmosférica e índices de inundación)*. Bogotá, Colombia: Centro de Investigación y Desarrollo en Información Geográfica del IGAC, UNSPIDER IGAC. Recuperado de https://www.researchgate.net/publication/313346186_Desarrollo_de_un_modelo_de_correccion_de_imagenes_de_satelite_para_inundaciones_CAIN_-_Correccion_Atmosferica_e_Indices_de_Inundacion
- Bonilla, F. M., Silva, G. S., Cabrera, M. C., & Sánchez, T. R. (2013). Calidad del agua residual no entubada vertida por dos parques industriales en la ciudad de Puebla, México. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 4(7), 1-36. DOI: 10.23913/ride.v4i7.83

- Chucos, B. N., & Vega, V. E. (2022). Evaluación de algoritmos de machine learning en la clasificación de imágenes satelitales multiespectrales, caso: amazonia peruana. *Ciencia Latina, Revista Multidisciplinar*, 6(1). DOI: 10.37811/cl_rcm.v6i1.1843
- Chul, K. B., Hun, K. H., & Yeal, N. J. (2015). Classification of potential water bodies using Landsat 8 OLI and a combination of two boosted random forest classifiers. *Journal Sensors*, 15(6), 13763-13777. DOI: 10.3390/s150613763
- Chuvieco, S. (2002). *Teledetección ambiental: la observación de la Tierra desde el espacio*. Barcelona, España: Ariel.
- Conagua, Comisión Nacional del Agua. (2013). *Sistema Nacional de Información del Agua. Monitoreo de las principales presas de México*. Recuperado de <https://sinav30.conagua.gob.mx:8080/SINA/?opcion=monitoreo>
- Conagua, Comisión Nacional del Agua. (2018). *Atlas del agua en México, edición 2018*. Recuperado de <https://files.conagua.gob.mx/conagua/publicaciones/Publicaciones/AAM2018.pdf>
- Feyisa, G., Meilby, H., Fensholt, R., & Proud, S. (2014). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, 140, January 2014, 23-35.
- Fisher, A., Flood, N., & Danaher, T. (2016). Comparing Landsat water index methods for automated water classification in eastern Australia. *Remote Sensing of Environment*, 175, 167-182. DOI: 10.1016/j.rse.2015.12.055

- Gao, B. C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58, 257-266. DOI: 10.1016/S0034-4257(96)00067-3
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425-1432. DOI: 10.1080/01431169608948714
- Rouse, J. W., Hass, R. H., Deering, D. W., & Schell, J. A. (1974). *Monitoring the vernal advancement and retrogradation (Green wave effect) of natural vegetation*. Monterey, USA: Remote Sensing Center.
- Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27, 3025-3033. DOI: 10.1080/01431160600589179