

Proyección topológica del agua para la toma de decisiones informadas mediante algoritmos de forecasting aplicados al NDWI en imágenes satelitales

Tania Deisy Pinchao Muñoz y Andres Javier Vallejo Jaramillo

Programa de Ingeniería de Sistemas, Facultad de Ingeniería, Universidad CESMAG

Nota del autor

El presente Trabajo de Grado tiene como propósito cumplir el requisito exigido para optar al título de pregrado como Ingenieros de Sistemas en la Universidad CESMAG.

La correspondencia referente a este trabajo debe dirigirse al programa de ingeniería de sistemas de la universidad CESMAG. Correo electrónico:
ingenieriadesistemas@unicesmag.edu.co

Proyección topológica del agua para la toma de decisiones informadas mediante algoritmos de forecasting aplicados al NDWI en imágenes satelitales

Tania Deisy Pinchao Muñoz y Andres Javier Vallejo Jaramillo

Programa de Ingeniería de Sistemas, Facultad de Ingeniería, Universidad CESMAG

Asesora: Mg. Dalila Pachajoa Pachajoa

10 de agosto de 2026

Nota de aceptación

“Va la aceptación definida desde la universidad”.

Nota de exclusión

DEDICATORIA

A Dios, por darme la fortaleza en aquellos días en los que sentí que ya no podía continuar, por sostenerme cuando mis fuerzas se agotaban, por guiar cada uno de mis pasos en medio de la incertidumbre y por recordarme, incluso en los momentos más oscuros, que siempre existe una razón para seguir adelante.

A mi madre Yolanda, quien es mi mayor inspiración y el motor de mi vida. Cada esfuerzo, cada noche de cansancio, cada lagrima, todo ha valido la pena por ti. Gracias por enseñarme a ser fuerte, a no rendirme, a creer en mí incluso cuando yo no podía hacerlo. Eres el reflejo más puro de amor, lucha y de resiliencia que tengo en mi vida. Todo lo que soy te lo debo a ti.

A mi hermano Wilson, que, aunque ya no está físicamente conmigo, vive en mi corazón y en cada uno de mis logros. Aunque ha dolido, aquí estoy, con el corazón en la mano, honrando tu recuerdo en cada paso que doy.

A mi hermano Wilfred, quien, aunque a veces logre sacarme de quicio, siempre ha estado ahí para mí, apoyándome a su manera, cuidándome y queriendo lo mejor para mi vida.

A mi novio Wiston, por ser mi refugio en medio del caos, por creer en mí cuando ni siquiera yo lo hacía y por no soltar mi mano. Gracias por ser mi calma, mi fuerza y mi luz en los momentos más difíciles, por acompañarme, sostenerme y recordarme siempre que soy capaz y que no estoy sola.

A mi perrhija Luna, mi fiel compañera, mi alegría constante y mi refugio en los días difíciles. Eres esa compañía silenciosa que nunca falla, ese amor puro que me ha dado fuerzas incluso sin decir una palabra.

Este logro no es solo mío, es de cada persona que ha sido parte de mi camino, de cada palabra de aliento, de cada abrazo en los momentos difíciles y de cada lágrima que se transformó en fuerza. Hoy culmino una etapa, pero también honro cada esfuerzo, cada sacrificio y cada historia que me trajo hasta aquí.

Tania Deisy Pinchao Muñoz

DEDICATORIA

En primer lugar, dedico este trabajo de grado a Dios, por darme la vida, la salud y la fortaleza necesaria para seguir adelante cada día, por acompañarme en los momentos difíciles y por iluminar mi camino en cada paso de este proceso.

A mi familia, quienes han sido el pilar fundamental de mi vida, brindándome su apoyo incondicional, amor y motivación constante para no rendirme.

A mi madre, por creer en mí desde el inicio, por impulsarme a perseguir esta carrera que siempre soñé, por su amor infinito, su paciencia inagotable y sus sabios consejos que han guiado cada una de mis decisiones.

Y de manera muy especial, a Tania, mi compañera de investigación y amiga, por su apoyo sincero, su confianza en mí y por caminar a mi lado durante todo este proceso. Gracias por el trabajo en equipo, la dedicación compartida y por haber construido juntos este proyecto que hoy se hace realidad.

Andres Javier Vallejo Jaramillo

AGRADECIMIENTOS

Los autores expresan su más sincero agradecimiento a todas las personas e instituciones que contribuyeron de manera significativa al desarrollo y culminación de este trabajo de investigación.

A la Universidad Centro de Estudios Superiores María Goretti (CESMAG), por brindarnos la oportunidad de formarnos como profesionales en Ingeniería de Sistemas, por su acompañamiento constante y por fomentar en nosotros el compromiso, la responsabilidad.

A nuestra asesora, la Mg. Dalila Pachajoa Pachajoa, por su invaluable orientación, dedicación y compromiso durante todo el proceso investigativo. Su experiencia, conocimientos y acompañamiento permanente fueron fundamentales para la correcta estructuración y desarrollo de este proyecto.

Al Mg. Héctor Mora Paz, quien con sus conocimientos, aportes y disposición contribuyó de manera significativa a que este proyecto se hiciera realidad, fortaleciendo cada una de sus etapas.

A nuestros docentes y mentores, quienes no solo nos apoyaron en el ámbito académico, sino que también nos brindaron la oportunidad de explorar diferentes áreas del conocimiento. Su guía y compromiso fueron clave para nuestro crecimiento profesional, despertando el interés, la motivación y el gusto por nuestra carrera.

Finalmente, a todas las personas que de una u otra manera hicieron parte de este proceso, aportando ideas, apoyo y motivación, les extendemos nuestro más profundo agradecimiento por contribuir al logro de esta meta tan importante en nuestras vidas.

TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN.....	16
I. PROBLEMA DE INVESTIGACIÓN	18
A. OBJETO O TEMA DE ESTUDIO	18
B. LÍNEA DE INVESTIGACIÓN	18
1) Sub línea de investigación	18
C. PLANTEAMIENTO DEL PROBLEMA.....	19
D. FORMULACIÓN DEL PROBLEMA	20
E. OBJETIVOS.....	20
1) Objetivo general.....	20
2) Objetivos específicos	20
F. JUSTIFICACIÓN.....	21
G. DELIMITACIÓN.....	23
II. MARCO TEÓRICO	24
A. ANTECEDENTES.....	24
1) Internacionales	24
2) Nacionales.....	25
3) Locales	26
B. SUPUESTOS TEÓRICOS.....	27
C. VARIABLES DEL ESTUDIO.....	37
D. FORMULACIÓN DE HIPOTESIS	43
1) Hipótesis de investigación	43
2) Hipótesis nula.....	43
3) Hipótesis alterna.....	44
III. METODOLOGÍA	45
A. PARADIGMA.....	45
B. ENFOQUE	45
C. MÉTODO.....	45
D. TIPO DE INVESTIGACIÓN.....	46
E. DISEÑO DE LA INVESTIGACIÓN	46
F. POBLACIÓN	47
G. MUESTRA.....	48
H. TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN	48

I. INSTRUMENTO DE RECOLECCIÓN DE DATOS	49
IV. RESULTADOS DE LA INVESTIGACIÓN.....	62
Resultado del objetivo 1	62
Resultado del objetivo 2.....	73
Resultado del objetivo 3.....	79
V. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS	86
CONCLUSIONES	88
RECOMENDACIONES	89
REFERENCIAS BIBLIOGRÁFICAS	91

LISTA DE FIGURAS

	Pág.
Fig. 1. Topología del agua.....	28
Fig. 2. Calidad de agua.....	28
Fig. 3. NDWI (índice de agua de diferencia normalizada)	29
Fig. 4. Proyección topológica.....	30
Fig. 5. Gestión hídrica	30
Fig. 6. Inteligencia Artificial	31
Fig. 7. Diferencia Deep y Machine	32
Fig. 8. Series temporales	33
Fig. 9 Redes Neuronales.....	34
Fig. 10 Red neural convolucional	34
Fig. 11 Imagen multiespectral	36
Fig. 12 Reflectancia.....	37
Fig. 13 Flujo de metodología de investigación	50
Fig. 14 Flujo de datos	51
Fig. 15 Flujo de Pipeline para generación del dataset NDWI	54
Fig. 16 Flujo para la predicción de NDWI.....	56
Fig. 17 Flujo de evaluación y selección de modelos de predicción del NDWI	57
Fig. 18 Flujo de procesamiento para la visualización del NDWI	60
Fig. 19 Registros del dataset	65
Fig. 20 Shapefile del departamento de Nariño.	66
Fig. 21 NDWI agregado por latitud y longitud: (a) mínimo, (b) promedio, (c) máximo.....	67
Fig. 22 Modelos entrenado almacenados en formato .pkl.....	69
Fig. 23 Variables temporales con codificación cíclica.....	71
Fig. 24 Interpolación geoestadística con Kriging.....	72
Fig. 25 Archivos CSV generados para cada fecha del periodo de estudio.....	73
Fig. 26 Curva de pérdida del modelo ConvLSTM2D durante el entrenamiento	76
Fig. 27 Comparación espacial del NDWI real y predicho mediante el modelo ConvLSTM2D....	77
Fig. 28 Mapa de NDWI real y predicho.....	78
Fig. 29 Comportamiento de la interpolación del NDWI mediante Kriging.....	78
Fig. 30 Mapa continuo del NDWI para el departamento de Nariño	79
Fig. 31 Interfaz de la aplicación web para la consulta de predicciones del NDWI.....	80
Fig. 32 Visualización de la predicción del NDWI sobre el mapa base	81
Fig. 33 Estructura de almacenamiento de resultados generados por fecha	82
Fig. 34 Comparación de predicciones del NDWI en dos fechas diferentes.....	83
Fig. 35 Visualización de estadísticas asociadas a la predicción del NDWI	84
Fig. 36 Restricción de fechas	85

LISTA DE ECUACIONES

	Pág.
Ecuación 1 NDWI	41
Ecuación 2. MSE	42
Ecuación 3. R^2	42
Ecuación 4 MAE	43
Ecuación 5 Partición de la reflectancia por rango de fechas	62
Ecuación 6 segmentación por rango de fecha	62
Ecuación 7 Cruce de los segmentos	63

LISTA DE TABLAS

	Pág.
TABLA I. METODOLOGÍA PRELIMINAR.....	46
TABLA II. DESCRIPCIÓN DE LA BASE DATOS.....	51
TABLA III. CONSULTAS SQL PARA OBTENCIÓN DE SEGMENTOS DE LANDSAT Y MODIS	52
TABLA IV. CONSULTA DE ALMACENAMIENTO.....	53
TABLA V. HERRAMIENTAS UTILIZADAS EN EL CALCULO DEL NDWI	55
TABLA VI. CARACTERÍSTICAS DEL ENTORNO DE EJECUCIÓN EN GOOGLE COLAB	58
TABLA VII. ELEMENTOS DE LAS ECUACIONES DE SEGMENTACIÓN	63
TABLA VIII. RESULTADOS DE LA SEGMENTACIÓN TEMPORAL	64
TABLA IX. DATASETS CONSOLIDADOS OBTENIDOS DEL PROCESAMIENTO	64
TABLA X. ESTRUCTURA DE LOS DATOS SATELITALES OBTENIDOS	65
TABLA XI. VARIABLES PARA EL ENTRENAMIENTO DE LOS MODELOS	67
TABLA XII. MODELOS UTILIZADOS PARA LA ESTIMACION DE BANDAS ESPECTRALES	68
TABLA XIII. DESEMPEÑO (R^2) DE LOS MODELOS DE ESTIMACIÓN DE BANDAS	68
TABLA XIV. ARCHIVOS UTILIZADOS PARA LA GENERACIÓN DEL DATASET NDWI	70
TABLA XV. VARIABLES TEMPORALES GENERADAS.....	70
TABLA XVI. VARIABLES Y BANDAS UTILIZADAS EN EL CÁLCULO DEL NDWI.....	71
TABLA XVII. CONTENIDO DE LOS ARCHIVOS GENERADOS	72
TABLA XVIII. ESTRUCTURA DEL DATASET NDWI GENERADO.....	73
TABLA XIX. CONFIGURACION DE LOS MODELOS EVALUADOS	74
TABLA XX. DATOS DE ENTRENAMIENTO, PRUEBA Y VALIDACIÓN.....	74
TABLA XXI. RESULTADOS DE LA SINTONIZACIÓN DE MODELOS DE FORECASTING PARA NDWI.....	75

GLOSARIO

Este glosario define los conceptos clave de la investigación, enfocada en el análisis de datos satelitales para generar series temporales del NDWI, integrando preprocesamiento, análisis geoespacial y aprendizaje automático, con el fin de facilitar la comprensión del dataset, el cálculo del índice y su predicción.

APRENDIZAJE AUTOMÁTICO: Rama de la inteligencia artificial que permite a los sistemas aprender a partir de datos y realizar predicciones sin ser programados explícitamente.

CODIFICACIÓN CÍCLICA (SENO/COSENO): Técnica de transformación de variables temporales (como meses o días) en funciones trigonométricas seno y coseno, con el fin de representar su naturaleza periódica. Esta codificación evita discontinuidades entre valores extremos (por ejemplo, diciembre y enero), mejorando el desempeño de los modelos de aprendizaje automático.

CONVLSTM2D: Arquitectura de red neuronal que combina redes convolucionales (CNN) y redes neuronales recurrentes tipo LSTM. Está diseñada para modelar datos espacio temporales, permitiendo capturar tanto patrones espaciales (imágenes) como su evolución en el tiempo.

DATASET: Conjunto de datos organizados en forma de tabla, donde cada fila representa una observación y cada columna una variable, utilizado para el análisis y modelado.

DATOS MULTIESPECTRALES: Información obtenida en diferentes bandas del espectro electromagnético, utilizada para analizar características de la superficie terrestre.

GEOREFERENCIACIÓN: Proceso mediante el cual se asignan coordenadas geográficas a los datos para ubicarlos en el espacio.

INTERPOLACIÓN KRIGING: Método geoestadístico utilizado para estimar valores en ubicaciones donde no se tienen datos medidos, a partir de valores conocidos en puntos cercanos. Permite generar superficies continuas considerando la correlación espacial de los datos, siendo especialmente útil en el análisis de variables ambientales como el NDWI.

IMÁGENES SATELITALES: Representaciones de la superficie terrestre captadas por sensores en satélites, utilizadas para el análisis ambiental.

MODELOS DE PREDICCIÓN: Algoritmos que utilizan datos históricos para estimar valores futuros de una variable

NDWI: Índice de Agua de Diferencia Normalizada que permite identificar la presencia de agua en la superficie terrestre mediante el uso de bandas espectrales.

PREPROCESAMIENTO DE DATOS: Conjunto de técnicas aplicadas para limpiar, organizar y transformar los datos antes de su análisis.

SEGMENTACIÓN TEMPORAL: Proceso de dividir los datos en intervalos de tiempo para facilitar su análisis.

SERIES TEMPORALES: Secuencia de datos ordenados en el tiempo que permite analizar el comportamiento de una variable.

RESUMEN

El análisis de los recursos hídricos es fundamental para comprender la dinámica ambiental y los efectos del cambio climático en la superficie terrestre. En este contexto, las imágenes satelitales multiespectrales permiten identificar y monitorear la presencia de cuerpos de agua mediante el Índice de Agua de Diferencia Normalizada (NDWI), facilitando el estudio de su variación en el tiempo.

La presente investigación tuvo como objetivo generar y predecir series temporales del NDWI en el departamento de Nariño, integrando técnicas de procesamiento de datos satelitales y aprendizaje automático. Metodológicamente, se empleó el enfoque científico junto con la metodología KDD, que permitió la selección, limpieza, transformación y estructuración de grandes volúmenes de datos provenientes de sensores Landsat y MODIS. A partir de estos datos, se construyó un dataset organizado mediante segmentación temporal y cálculo del NDWI.

Se implementaron modelos de predicción de series temporales, destacándose el modelo ConvLSTM2D por su capacidad para capturar relaciones espacio-temporales. Los resultados evidenciaron un buen desempeño en métricas como R^2 , MAE y MSE, lo que valida la efectividad del enfoque propuesto.

Y, se concluye que la integración de datos satelitales y técnicas de aprendizaje automático permite mejorar la comprensión y predicción del comportamiento de los cuerpos de agua, aportando herramientas relevantes para la gestión ambiental y el monitoreo de recursos hídricos en la región.

Palabras Clave: Forecasting, imágenes satelitales, inteligencia artificial, NDWI, recurso hídrico.

ABSTRACT

The analysis of water resources is fundamental to understanding environmental dynamics and the effects of climate change on the Earth's surface. In this context, multispectral satellite imagery allows for the identification and monitoring of water bodies using the Normalized Difference Water Index (NDWI), facilitating the study of their variation over time.

This research aimed to generate and predict NDWI time series for the department of Nariño, integrating satellite data processing techniques and machine learning. Methodologically, the scientific approach was employed along with the KDD methodology, which allowed for the selection, cleaning, transformation, and structuring of large volumes of data from Landsat and MODIS sensors. From this data, an organized dataset was constructed using temporal segmentation and NDWI calculation.

Time series prediction models were implemented, with the ConvLSTM2D model standing out for its ability to capture spatiotemporal relationships. The results showed good performance in metrics such as R^2 , MAE, and MSE, validating the effectiveness of the proposed approach.

It is concluded that the integration of satellite data and machine learning techniques improves the understanding and prediction of water body behavior, providing relevant tools for environmental management and water resource monitoring in the region.

Keywords: Forecasting, satellite imagery, artificial intelligence, NDWI, water resources.

INTRODUCCIÓN

Actualmente, el monitoreo de las fuentes hídricas enfrenta grandes desafíos debido al impacto del cambio climático, el incremento de los gases de efecto invernadero y la alteración de los ecosistemas como consecuencia de la actividad humana. Estas condiciones han generado una mayor variabilidad en los patrones climáticos, afectando directamente la distribución y comportamiento del recurso hídrico, lo que dificulta su gestión y planificación en distintos territorios. En el departamento de Nariño se evidencian variaciones climáticas constantes e impredecibles, caracterizadas por la presencia de lluvias intensas, sol radiante y nubosidad persistente en intervalos cortos de tiempo. Esto no solo complica las actividades del monitoreo de los satélites, sino que también incrementa el riesgo de desastres naturales, como inundaciones repentinas, sequías, deslizamientos de tierra, poniendo en peligro tanto a las comunidades como a los ecosistemas locales.

El uso de NDWI (Normalized Difference Water Index) se convierte en una herramienta fundamental, ya que permite identificar y cuantificar la presencia de agua en la superficie terrestre a partir del análisis de imágenes satelitales multiespectrales. Este índice facilita el monitoreo de cuerpos de agua, la evaluación de la humedad del suelo y el análisis de cambios en el comportamiento hídrico a lo largo del tiempo. Su aplicación en el departamento de Nariño representa una oportunidad para optimizar el monitoreo ambiental, permitiendo la detección temprana de fenómenos anteriormente mencionados.

Dado lo anterior, la presente investigación propone el desarrollo de un modelo basado en técnicas de inteligencia artificial, orientado a la predicción del comportamiento del NDWI mediante el uso de algoritmos de Forecasting. Este enfoque permite analizar grandes volúmenes de datos históricos provenientes de imágenes satelitales multiespectrales, las cuales, gracias a su alta resolución espacial y temporal, capturan de manera detallada las dinámicas del recurso hídrico en la superficie terrestre. A partir de esta información, es posible identificar patrones, tendencias y relaciones complejas que contribuyen a una mejor comprensión del comportamiento hídrico en el territorio.

Ante este panorama, la presente investigación propone la integración de técnicas de procesamiento de datos satelitales y algoritmos de aprendizaje automático para la generación y predicción de series

temporales del NDWI. Para ello, se emplean datos provenientes de sensores Landsat y MODIS, los cuales son sometidos a procesos de preprocesamiento, transformación y modelado, con el fin de construir un dataset estructurado y apto para el análisis. Posteriormente, se implementan modelos de forecasting, destacándose arquitecturas de aprendizaje profundo como ConvLSTM2D, capaces de capturar patrones espacio temporales complejos. Por último, se desarrolla una aplicación web que permite visualizar las predicciones del NDWI en un entorno geoespacial, facilitando su interpretación y uso en escenarios reales.

De esta manera, el estudio no solo contribuye al desarrollo de modelos predictivos aplicados al análisis del recurso hídrico, sino que también propone una herramienta tecnológica que apoya la toma de decisiones en el contexto ambiental del departamento de Nariño.

I. PROBLEMA DE INVESTIGACIÓN

A. OBJETO O TEMA DE ESTUDIO

La presente investigación se centró en la proyección topológica de la disponibilidad del agua en el departamento de Nariño, mediante el uso de imágenes satelitales multiespectrales y algoritmos de Forecasting. Se analizó la aplicabilidad del índice de agua de diferencia normalizada (NDWI)[1] como herramienta fundamental para modelar la cantidad y distribución espacial temporal de este recurso hídrico, teniendo en cuenta datos espaciales y temporales[2]. Este estudio se desarrolló integrando técnicas de preprocesamiento de imágenes y aprendizaje automático de diferentes algoritmos[3] Esta aproximación permitió generar predicciones que apoyan a la toma de decisiones informadas en la gestión y planificación hídrica regional.[4].

B. LÍNEA DE INVESTIGACIÓN

Este proyecto se enmarca en la línea de investigación “**Tecnologías de la Información y la Comunicación (TIC)**”[5], al integrar el desarrollo de modelos de aprendizaje automático y el procesamiento de imágenes satelitales multiespectrales para el pronóstico forecasting de recursos hídricos. Asimismo, se orienta hacia la gestión eficiente de la información mediante la implementación de herramientas digitales que optimizan la planificación del recurso hídrico y fortalecen la toma de decisiones estratégicas[6]. La investigación se articula directamente con el grupo de investigación Tecnofilia de la Universidad CESMAG y se fundamenta en la línea y sublínea establecidas en el documento oficial proporcionado por el programa de Ingeniería de Sistemas, el cual respalda el desarrollo de soluciones basadas en TIC aplicadas a problemáticas reales del entorno regional. la línea de investigación que soporta el desarrollo del proyecto[7].

1) Sub línea de investigación

Este proyecto se enmarca en la sublínea de investigación “**Inteligencia Artificial**”, dado que incorpora algoritmos de pronóstico forecasting y el procesamiento de imágenes satelitales multiespectrales. A través de este enfoque, se aplican técnicas de aprendizaje automático y modelos predictivos orientados a evaluar el Índice de Agua de Diferencia Normalizada (NDWI). Para ello, se emplean procesos de análisis y tratamiento de datos, modelado de series temporales y

optimización de modelos[8], con el propósito de desarrollar una propuesta basada en inteligencia artificial que contribuya a mejorar la gestión y planificación del recurso hídrico[9]. La línea y sublínea de investigación se fundamentan en el documento oficial proporcionado por el programa de Ingeniería de Sistemas, donde se establecen y articulan los campos de estudio. Asimismo, el proyecto se vincula directamente con la línea definida por el grupo de investigación Tecnofilia de la Universidad CESMAG, el cual respalda el desarrollo de investigaciones en inteligencia artificial aplicadas a problemáticas reales del entorno[7].

C. PLANTEAMIENTO DEL PROBLEMA

En el departamento de Nariño, la gestión del recurso hídrico enfrenta diversos desafíos debido a la falta de información precisa y actualizada sobre la disponibilidad de agua. Actualmente, muchos procesos de planificación y abastecimiento se basan en datos históricos o en mediciones del presente, sin considerar posibles variaciones futuras del recurso. Esta situación limita la capacidad de las entidades encargadas para tomar decisiones oportunas y genera improvisación presupuestal, así como planes de contingencia poco efectivos frente a fenómenos ambientales extremos como incendios, sequías e inundaciones. Además, la limitada disponibilidad de información confiable dificulta el desarrollo de estrategias de monitoreo y gestión que permitan anticipar cambios en la disponibilidad del agua y mejorar la planificación territorial. Los estudios que se han encontrado priorizan la generación de mapas de irradiancia en función de la reflectancia utilizando métodos de clasificación y regresión estos no se han utilizado para predecir cambios a futuro. Igualmente, no se han enfocado en la predicción y pronóstico del Índice de Agua de Diferencia Normalizada (NDWI)[10].

La principal causa es la falta de información precisa y segura para determinar sobre cómo se encuentra la calidad del agua en el territorio, en las búsquedas realizadas no se ha encontrado un dataset que integre adecuadamente criterios espaciales y temporales relevante para modelar la viabilidad, calidad y disponibilidad del recurso hídrico, a estas barreras se suman condiciones atmosféricas como la nubosidad que dificulta el proceso del muestreo, afectando la calidad de los datos obtenidos. Por lo tanto, las averías en sistemas de monitoreo disminuyen la frecuencia y confiabilidad de las mediciones lo que dificulta disponer de datos constantes y valiosos para análisis predictivos, por otro lado, se identificó que los estudios previos se han enfocado

principalmente en generar mapas de irradiancia basados en funciones de la reflectancia como la regresión de Cabrera y Mora[11], [12] y clasificación de Pachajoa[13]. No obstante, no incorporan modelos predictivos ni el Índice de Agua de Diferencia Normalizada (NDWI) para proyecciones temporales. Esta brecha se agrava por la ausencia de datasets integrados espacial y temporalmente, nubosidad que afecta el muestreo satelital, y averías en sistemas de monitoreo, reduciendo la confiabilidad de los datos, lo que constituye una desigualdad significativa en el contexto actual de gestión del recurso[14].

La falta de una planificación adecuada en el abastecimiento y cuidado del agua, por la escasez de datos precisos y la ausencia de índices como el NDWI (Índice de Agua de Diferenciada Normalizada), conduce a la improvisación presupuestal y a la ineficacia de los planes de contingencia para enfrentar los efectos ambientales. Esta situación aumenta el riesgo de incendios forestales y dificulta la adaptación al cambio climático, poniendo en peligro el cumplimiento de acuerdos internacionales como el de París y el de políticas nacionales del Ministerio de Ambiente y Desarrollo Sostenible de Colombia[15].

D. FORMULACIÓN DEL PROBLEMA

¿Cuál es el grado de predictibilidad de distintos algoritmos de aprendizaje automático aplicados al NDWI, derivados de imágenes satelitales multiespectrales, para la proyección topológica de la disponibilidad hídrica en el departamento de Nariño?

E. OBJETIVOS

1) Objetivo general

Evaluar el grado de predictibilidad de diferentes algoritmos de aprendizaje automático para la estimación del Índice de Agua de Diferencia Normalizada (NDWI) a partir de imágenes satelitales multiespectrales.

2) Objetivos específicos

- Preprocesar un conjunto de base de datos (Imágenes satelitales de LANDSAT Y MODIS) para generar series temporales de NDWI.

- Componer scripts para la sintonización de algoritmos de series temporales para predecir el NDWI, identificando el modelo con mayor compromiso en sus métricas de calidad (R^2 , MAE, MSE) y tiempo de ejecución.
- Implementar una aplicación web que permita visualizar las predicciones del mejor modelo sobre un visor de mapas.

F. JUSTIFICACIÓN

Los pronósticos que se presentan en el estado de oferta del agua, en un territorio determinado toman mayor relevancia para la toma de decisiones en la gestión del recurso hídrico ya que permite conocer la distribución del agua en una región y su calidad, de igual forma esto permite que se realicen planificaciones más acertadas en cuanto a su almacenamiento y repartición lo cual repercute en el crecimiento urbano asociado al ordenamiento territorial, de igual manera estos pronósticos ayudan a una prevención más acertada sobre posibles desastres naturales como inundaciones o sequías[10]. Otro aspecto importante a resaltar es que hay una correlación directamente proporcional de la oferta del agua con los costos del servicio ya que de no contar con una oferta suficiente el agua tendría que ser transportada hacia los centros poblados que la necesiten añadiéndosele al costo natural unos costos indirectos. También es importante destacar que para pronosticar la distribución del agua de una zona geográfica se necesita de una cantidad importante de datos y las imágenes satelitales cumplen con este requerimiento ya que sensores como Landsat y MODIS pasan frecuentemente capturando imágenes multiespectrales de todo el planeta[12], por lo tanto aquí hay una oportunidad para hallar una correlación entre las bandas de luz capturadas en el tiempo y el índice de agua de diferencia normalizada NDWI que se encarga de determinar y monitorear los cambios en el contenido de agua superficial, observar el estado hídrico de los cultivos, identificar la saturación de humedad en el suelo entre otros[14]. Una de las formas para hallar dicha correlación es la utilización de inteligencia artificial y más específicamente algoritmos de Forecasting, y debido a que para el departamento de Nariño en las búsquedas realizadas en el presente estudio aún no se ha realizado experimentaciones asociadas a este aspecto con imágenes satelitales hay una oportunidad para indagar sobre el compromiso en sus métricas de predictibilidad y coste computacional de diferentes algoritmos, significando un aporte epistemológico y técnico[15], que sirva como base a otros estudios.

El uso del NDWI en imágenes satelitales es común, pero su combinación con algoritmos de Forecasting y proyección topológica es innovadora ya que aún hay algunos vacíos experimentales que se pueden cubrir como por ejemplo averiguar la correlación del NDWI con imágenes satelitales como señal en dos dimensiones y su contraparte algorítmica utilizando redes neuronales convolucionales de memoria corta y de largo plazo CONVLSTM o CONVGRU. De igual manera este estudio se concentra en determinar pronósticos para el departamento de Nariño a partir del conjunto de datos de Cabrera [11], aportando en el preprocesamiento y transformación de datos. Finalmente, el presente estudio se concentrará en generar mapas de NDWI a futuro y no sobre el presente como generalmente se realiza.

Esta investigación aportará al programa de ingeniería de sistemas y al grupo de investigación Tecnofilia en la línea de inteligencia artificial con el tratamiento de imágenes satelitales multiespectrales, en cada una de las etapas de descubrimiento de conocimiento en base de datos KDD con algoritmos de preprocesamiento, transformación, minería de datos y validación para experimentos con series temporales hallando patrones mediante algoritmos de Forecasting que conlleve en la mejora de análisis de datos geoespaciales y específicamente en el NDWI[3], [16]. Igualmente, y no menos importante, también beneficiará a la línea de investigación en desarrollo de software tanto en la construcción de los scripts experimentales como también en la implementación de una aplicación web la cuál facilitará la visualización de dichas predicciones.

Finalmente, es muy importante mencionar que este proyecto se articula con los compromisos adquiridos en el acuerdo de París al colaborar con la adaptación y mitigación del cambio climático mediante la gestión sostenible del recurso hídrico. Igualmente, responde a las políticas nacionales del Ministerio de Ambiente y Desarrollo Sostenible de Colombia, en particular a los lineamientos del Plan Nacional de Gestión Integral del Recurso Hídrico (PNGIRH), que impulsa el uso de tecnologías para el monitoreo, modelación y análisis de variables hidrológicas [15]. La creación de herramientas de análisis predictivo con imágenes satelitales multiespectrales se permitirá tener mayor información ambiental y por lo mismo generar estrategias basadas en evidencia científica para la planificación del recurso hídrico en territorios del departamento de Nariño [11], [17], donde puede haber altas vulnerabilidades.

G. DELIMITACIÓN

Esta investigación se enfocó en la evaluación comparativa de algoritmos de aprendizaje automático para la estimación predictiva del NDWI a partir de imágenes satelitales multiespectrales del departamento de Nariño, donde el objetivo central fue determinar el grado de eficacia predictiva de modelos de aprendizaje automático (tradicionales y de Deep Learning) en la estimación del Índice de Diferencia Normalizada de Agua (NDWI)[16], con aplicación en el monitoreo de la oferta hídrica regional. Las fuentes de datos que se utilizaron son imágenes multiespectrales de: Landsat (2003-2014) y MODIS (2005-2015) y variables complementarias: series temporales climáticas. En cuanto al procesamiento, se tiene en cuenta la fusión de datos espaciotemporales, la validación cruzada estratificada y métricas de evaluación: MAE, RMSE y R^2 ajustado. En el ámbito geográfico la zona de estudio es el Departamento de Nariño (Colombia) y en cuanto a la cobertura espacial, está el área total capturada por los sensores Landsat y MODIS (resolución 30m y 250m respectivamente). La innovación radica en un protocolo replicable para monitoreo hídrico regional, integrando forecasting temporal y análisis comparativo de algoritmos.

II. MARCO TEÓRICO

A. ANTECEDENTES

1) *Internacionales*

Meléndez Salazar (2024) llevo a cabo un análisis multitemporal de la dinámica de coberturas en el manglar del estero de Puntarenas durante el período 1974-2024. Hizo uso de imágenes satelitales de los sensores Landsat 1, 2, 5 y 8, aplicando índices espectrales como NDVI, SAVI, NDWI, NDWI Modificado y MVI para identificar y analizar la evolución del bosque de manglar. La clasificación de las coberturas del suelo se llevó a cabo mediante la técnica Support Vector Machine (SVM) en ArcGIS Pro, validando los resultados con matrices de confusión y el índice Kappa. por lo tanto, se desarrolló un visor cartográfico en ArcGIS Online para visualizar los cambios en la cobertura del manglar a lo largo de 50 años. Este estudio se centró en la evolución de los ecosistemas de manglar, sin abordar específicamente la evolución detallada del NDWI[17]. Este antecedente respalda la viabilidad técnica de usar NDWI y clasificación supervisada para el monitoreo ambiental, aportando bases metodológicas que se retoman y adaptan en la investigación actual para centrarse de forma más detallada en la evolución y predicción del índice.

Cunza Robles (2024) estimó el retroceso y predijo el tiempo de permanencia del nevado Ausangate en la región de Cusco entre los años 1985 y 2020[18]. Utilizo imágenes satelitales obtenidas del Servicio Geológico de Estados Unidos (USGS), procesadas con el software ArcGIS 10.5, y realizó análisis estadísticos utilizando Excel. La investigación cuantificó la pérdida de área y la variación del perímetro glaciar, además de estimar el tiempo de permanencia de la cobertura glaciar. Este trabajo se enfocó en el retroceso glaciar como indicador del cambio climático, sin centrarse en la evolución detallada del NDWI. Este estudio muestra cómo se puede aplicar análisis temporal y modelado predictivo a variables ambientales, lo que resulta útil como referente para la investigación actual, que traslada ese enfoque predictivo al estudio de cuerpos de agua mediante el NDWI.

Delgado Pascual (2024) analizó la evolución espaciotemporal de las áreas de regadío en Castilla y León durante las últimas dos décadas. Desarrolló una metodología basada en series multitemporales de imágenes satelitales Landsat para la cartografía del regadío, con el objetivo de mejorar la gestión sostenible del agua en el sector agrícola. Este estudio aplicó técnicas de

teledetección para identificar cambios en las superficies de regadío, sin enfocarse específicamente en la evolución detallada del NDWI[19]. Este antecedente aporta referentes metodológicos en el uso de teledetección para el análisis de cambios en superficies asociadas al recurso hídrico. La presente investigación retoma esta base para enfocarse en el análisis del agua superficial mediante NDWI y su modelado predictivo.

2) Nacionales

Herrera Posada y Aristizábal (2022) desarrollaron un modelo de aprendizaje automático para predecir espacial y temporalmente acontecimientos de sequía en el departamento del Magdalena[8]. Emplearon algoritmos como Random Forest y Árboles de Decisión, integrando variables como NDVI, NDWI, temperatura superficial, humedad del suelo. Este trabajo resalta el potencial del uso de imágenes satelitales e inteligencia artificial para apoyar la toma de decisiones frente a fenómenos extremos. Por lo tanto, se enfocaron en predicción de sequías y no en la evolución detallada del NDW. Este estudio respalda la aplicación de inteligencia artificial con datos satelitales y NDWI como insumo, lo cual es base directa para la investigación actual que aplica algoritmos predictivos al NDWI como variable principal de análisis.

Cuellar et al. (2023) utilizaron redes neuronales LSTM y NARX junto con filtros de Kalman para predecir el nivel del agua en el río San Juan (Chocó)[20]. Consiguieron incrementar la exactitud en comparación con modelos tradicionales, lo que demuestra el potencial del aprendizaje profundo en predicción hidrológica. El análisis se centró en niveles de río, no en índices espectrales. Este trabajo evidencia la eficacia de técnicas de aprendizaje profundo para predicción hidrológica, lo cual refuerza la pertinencia de usar algoritmos avanzados como los propuestos en la investigación actual, pero aplicados al análisis del NDWI.

Garzón et al. (2021) analizaron la transformación del bosque seco tropical en Agua de Dios[21], Cundinamarca, mediante imágenes Landsat y los índices NDVI y NDWI para los años 1991, 2015 y 2020. Descubrieron pérdida de cobertura vegetal y humedad del suelo. Si bien no usaron predicción ni IA, demostraron el valor del NDWI para monitoreo ambiental. Este antecedente justifica la elección del NDWI como índice relevante para el monitoreo ambiental, base sobre la cual la investigación actual propone profundizar en su evolución temporal y predicción futura.

López y Rodríguez (2021) emplearon los índices NDVI, NDWI y SAVI para estudiar los cambios en la cobertura vegetal y la humedad del suelo en humedales de Bogotá[22]. Usaron imágenes multitemporales y series históricas. Aunque se enfocaron en análisis descriptivo, el estudio aporta a la gestión ambiental urbana. Este estudio fundamenta el valor del NDWI como herramienta de análisis en contextos urbanos, lo cual es útil para la investigación actual, que amplía este enfoque incorporando series temporales y técnicas predictivas.

3) Locales

Cabrera et al. (2016) recopilaron y analizaron imágenes satelitales Landsat y MODIS para el departamento de Nariño entre los años 2000 y 2015, con el objetivo de construir modelos predictivos relacionados con la oferta de energía solar[11]. Aunque su investigación no se centró en el índice NDWI ni en la gestión del agua, emplearon técnicas avanzadas de teledetección para caracterizar variables ambientales en el territorio. Este trabajo demuestra la aplicabilidad de la teledetección para estudios energéticos en contextos regionales. Por lo tanto, se enfocaron en el recurso solar y no en predicción ambiental hídrica ni cambio en cuerpos de agua. Este antecedente local muestra el uso de teledetección para modelado predictivo en Nariño, validando la aplicabilidad de estas técnicas en el territorio. La investigación actual se diferencia al centrarse en cuerpos de agua y usar el NDWI como variable central.

Mora Paz (2019) desarrolló una investigación orientada a la identificación de zonas potenciales para el aprovechamiento de energía solar en Nariño, apoyándose en datos satelitales e información climática. Utilizó herramientas de sistemas de información geográfica (SIG) y modelos de radiación solar, integrando variables como nubosidad y cobertura del terreno[12]. Su estudio, aunque no centrado en NDWI, evidencia cómo las tecnologías geoespaciales pueden aportar a la planificación territorial y la gestión sostenible del recurso energético. Por lo tanto, su enfoque fue energético y no ambiental desde el componente hídrico. Este trabajo evidencia el uso de SIG y datos climáticos para planificación territorial en Nariño. Aporta indirectamente al estudio actual al validar la utilidad de estas herramientas en estudios ambientales en el mismo contexto geográfico.

Pachajoa , D.-M. , Mora-Peace , H. , & Majorca-Torres , D. (2021).desarrollaron una investigación centrada en la clasificación de zonas con alta irradiancia solar en el departamento de Nariño, utilizando imágenes satelitales multiespectrales del satélite Landsat. Su estudio empleó funciones

Kernel alternativas en combinación con máquinas de vectores de soporte (SVM) para evaluar el rendimiento de distintos modelos de clasificación[13]. Aunque su enfoque principal no fue el índice NDWI ni el componente hídrico, su investigación evidencia el potencial del análisis satelital y de los algoritmos de predicción para la toma de decisiones territoriales. Este trabajo aporta una base metodológica valiosa para el desarrollo de modelos predictivos aplicables a diferentes variables ambientales, entre ellas, el recurso hídrico. Este antecedente contribuye metodológicamente al estudio actual al demostrar la efectividad de técnicas como SVM en la clasificación de imágenes, lo cual se adapta en este trabajo al análisis y predicción del comportamiento del NDWI.

B. SUPUESTOS TEÓRICOS

TOPOLOGÍA H2O

Topología

La topología es una rama fundamental de las matemáticas que se enfoca en el estudio de las propiedades espaciales de los objetos que se conservan incluso cuando estos se deforman, estiran o encogen continuamente, sin importar su tamaño o forma precisa[23]. En el contexto del análisis geoespacial, esta capacidad de abstraer las dimensiones exactas permite representar y analizar la conectividad y las relaciones espaciales entre elementos como cuerpos de agua o flujos hidrológicos, facilitando la construcción de modelos complejos que reflejan patrones de distribución hídrica, esto es importante para la investigación en curso.

Topología del agua

En la idea de topología del agua representa las conexiones y relaciones espaciales entre cuerpos de agua, redes de distribución y flujos, ya sean naturales o artificiales[24], [25]. En el análisis geoespacial, esta perspectiva permite abstraer las dimensiones exactas para examinar la estructura del sistema hídrico como una red, evaluando su eficiencia, vulnerabilidades y comportamiento ante cambios climáticos o geográficos. Por lo tanto, permite construir modelos complicados que muestren patrones de distribución del agua. Este proyecto va a utilizar algoritmos de Forecasting aplicados a datos NDWI (Índice De Agua De Diferencia Normalizada) para proyectar la topología del agua en el futuro, lo que contribuye a anticipar dinámicas espaciales y optimizar la gestión del recurso.



Fig. 1. Topología del agua

Fuente: CERO MÁS TRES CEROS S.A. DE C.V.

Calidad del agua

La calidad del agua abarca sus características físicas, químicas y biológicas, determinando si es apta para usos como el consumo humano, la agricultura o el mantenimiento de ecosistemas. Factores como la temperatura, el pH, la presencia de contaminantes y microorganismos influyen directamente en esta calidad, la cual puede verse amenazada por la contaminación, el uso del suelo y el cambio climático[26], [27]. Aunque en esta investigación el NDWI (Índice De Agua De Diferencia Normalizada), está enfocado en detectar la presencia y cambios en superficies de agua, ofrece información indirecta clave para monitorear la calidad del recurso. Al estudiar su dinámica espacial, es posible identificar zonas en riesgo como áreas con sequías, pérdida de humedad o alteraciones rápidas en cuerpos de agua que, a su vez, pueden indicar problemas potenciales en la calidad hídrica, asociados a contaminación, estrés ambiental o cambios climáticos.



Fig. 2. Calidad de agua

Fuente: El Instituto de Estudios Peruanos (IEP)

NDWI (Índice De Agua De Diferencia Normalizada)

El NDWI es una herramienta que se usa en la observación de la Tierra desde el espacio para encontrar y seguir la pista del agua que vemos en la superficie. Se calcula usando la información de los colores verde e infrarrojo cercano que captan los satélites en sus imágenes. Con este índice, es posible reconocer la presencia de cuerpos lacustres, fluviales, zonas anegadas y el nivel de humedad del terreno. Su escala de valores oscila entre -1 y 1, siendo los valores positivos generalmente indicativos de la existencia de agua[28], [29]. Para este estudio, el NDWI es lo más importante que vamos a analizar y tratar de predecir. Para hacer esto, usaremos algoritmos de aprendizaje automático imágenes tomadas por satélites como Landsat y MODIS en diferentes momentos.

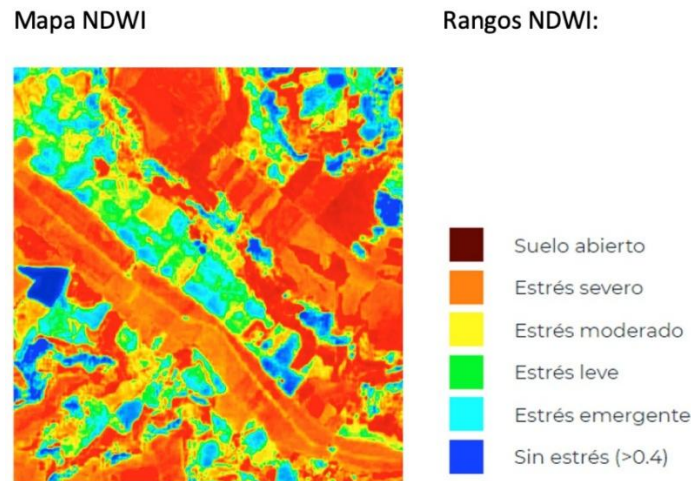


Fig. 3. NDWI (índice de agua de diferencia normalizada)

Fuente: VISUAL

Proyección topológica

La idea de la proyección topológica es mostrar en un mapa cómo se relacionan diferentes elementos dentro de un sistema, manteniendo sus conexiones y cómo siguen unidos, sin importar su forma precisa o qué tan grandes sean[30]. Cuando se habla de la gestión del agua, esta forma de representarla ayuda a ver cómo están distribuidos los ríos, lagos, etc. en un área y cómo podrían cambiar en el futuro según los modelos de predicción. En esta investigación, se propone usar la topología para mostrar dónde hay agua, utilizando un índice llamado NDWI y modelos de aprendizaje automático que ayude a prevenir cómo se comportará el agua con el tiempo y en diferentes lugares.

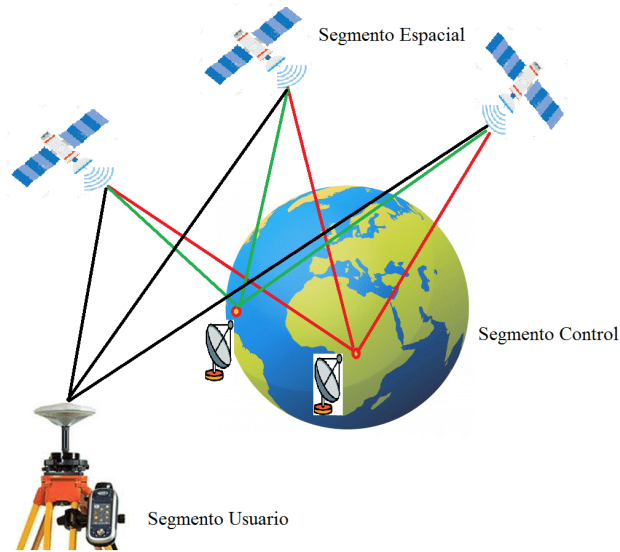


Fig. 4. Proyección topológica

Fuente: Paratopografía

Gestión hídrica

La gestión hídrica es el arte y la ciencia de cuidar el agua, implica desde políticas públicas hasta soluciones tecnológicas para garantizar que este recurso vital llegue a quienes lo necesitan, incluso en épocas de sequía o contaminación[31]. Con este propósito, se propone prestar especial atención a la información proporcionada por satélites en órbita terrestre y utilizar estos datos en complejos cálculos predictivos para comprender los patrones de movimiento y comportamiento del agua superficial visible[8]. De esta manera, este proyecto utiliza modelos predictivos basados en el NDWI, que actúan como un "pronóstico del tiempo" para los recursos hídricos. Fig. 5



Fig. 5. Gestión hídrica

Fuente: Técnicaña

Medio ambiente

El medio ambiente es el entorno natural que rodea a los seres vivos, incluyendo elementos bióticos (seres vivos) y abióticos (elementos no vivos como agua, aire, suelo). Es el espacio donde se desarrolla la vida, donde se interactúan los organismos con su entorno, y es estudiado por la ecología[32]. Para los propósitos de este estudio, consideramos el ambiente natural como una red compleja donde la forma en que funciona la naturaleza y cómo la tecnología avanza se conectan, influyendo directamente en cuánta agua hay, cómo se comporta y proporciona.

INTELIGENCIA ARTIFICIAL

En el concepto de la inteligencia artificial esta una disciplina que desarrolla sistemas aptos de simular capacidades cognitivas humanas a través del uso de algoritmos y modelos matemáticos bastante complejos, busca replicar habilidades humanas como el reconocimiento de patrones, el aprendizaje a partir de experiencias, el razonar, tomar decisiones por sí mismo y entender el lenguaje natural. Su objetivo es lograr que las computadoras muestren un comportamiento inteligente y se adapten a nuevas situaciones resolviendo problemas de manera similar a como lo haría un humano[33].

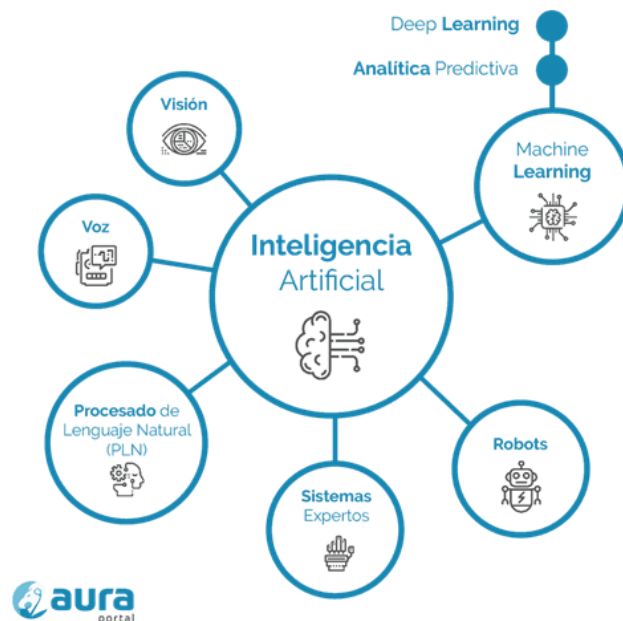


Fig. 6. Inteligencia Artificial

Fuente: Crehana

Deep Learning

El aprendizaje profundo, un método avanzado dentro del aprendizaje automático enfocado en aprender a interpretar datos, se basa en redes neuronales jerárquicas. En el nivel inicial, la red identifica características simples de la información otorgada, las cuales son transmitidas a la siguiente capa. Allí, esta información se combina y se compone para formar representaciones un poco más complejas, un proceso que se repite sucesivamente a través de las capas[34]. Este análisis progresivo, desde patrones elementales hasta relaciones complejas en sus distintas capas, facilita la extracción de conocimientos valiosos para la gestión y conservación de los recursos hídricos.

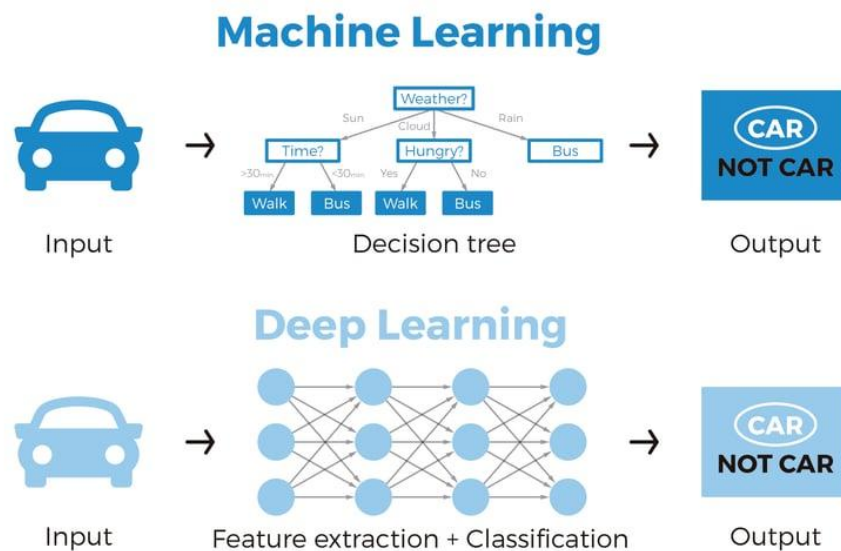


Fig. 7. Diferencia Deep y Machine

Fuente: Bismart

Forecasting

El concepto de Forecasting es el proceso de llevar a cabo una predicción hacia el futuro teniendo en cuenta datos históricos, como objetivo es disminuir las inseguridades a la hora de tomar decisiones, asimismo tener una planificación más efectiva en campos como la economía, la producción, las ventas o el clima. También puede apoyarse de métodos como lo son los cualitativos (opinión de expertos) o cuantitativos (modelos estadísticos o series temporales), es una herramienta muy fundamental en la gestión empresarial y en varias disciplinas científicas. Con la aplicación de este permite anticipar cambios y prepararse para distintos escenarios futuros[35].

Series temporales

La idea de serie temporal es un conjunto de datos ordenados que se registran a lo largo del tiempo para observar cómo evoluciona una variable. Su análisis permite identificar tendencias, patrones y anomalías en diferentes datos los cuales ayudan a entender y predecir comportamientos futuros. Se utiliza bastante en áreas como la economía, la meteorología, las finanzas y la producción. El estudio de series temporales permite tomar decisiones más informadas y a su vez planificar estrategias a mediano y largo plazo[36].

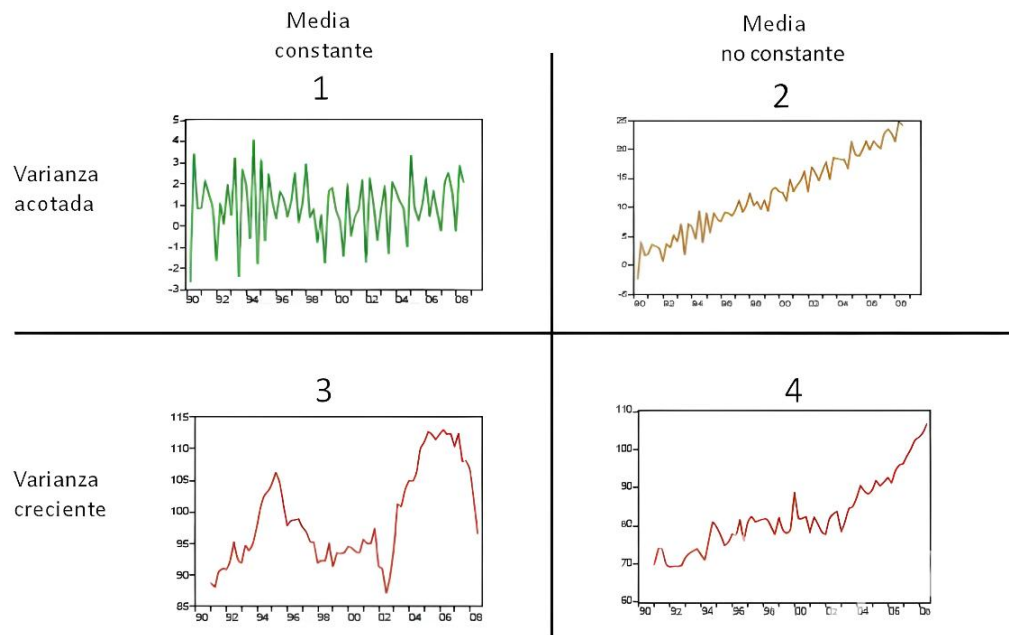


Fig. 8. Series temporales

Fuente: Economipedia

Redes neuronales

Redes Neuronales son modelos computacionales que inicialmente fueron inspirados en el cerebro humano para procesar información, también diseñados para identificar patrones y realizar predicciones. Están formadas por varias capas de neuronas interconectadas que procesan y aprenden a partir de datos con información clave, estas redes ajustan sus conexiones internas durante el proceso de entrenamiento mejorando su comportamiento para realizar tareas como la clasificación o predicción. Es utilizada para clasificación de imágenes, el procesamiento de lenguaje natural y toma de decisiones complejas[37].

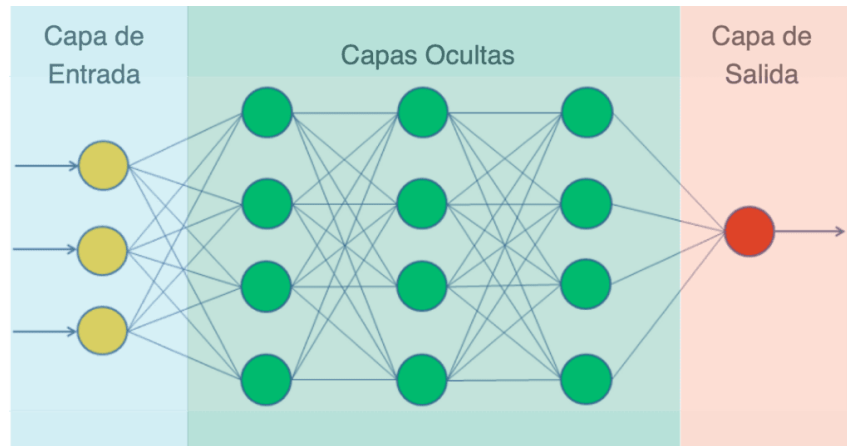


Fig. 9 Redes Neuronales

Fuente: AprendeIA

Red neural convolucional

Son una serie de redes que son ampliamente utilizadas para tratar con imágenes, están formadas por una o más capas convolucionales. Estas capas tienen la ventaja de extraer características de las imágenes con lo cual se puede aprender características dentro de unas pequeñas ventanas de la imagen y no características globales de la imagen como lo que haría una red totalmente conectada. Estas redes son capaces de aprender en los diferentes niveles de abstracción: en la primera capa se diferencian formas simples, colores o bordes; en la siguiente se pueden distinguir combinaciones de bordes y colores; mientras que la última capa se fija en la forma con el fin de conseguir averiguar qué es exactamente[38].

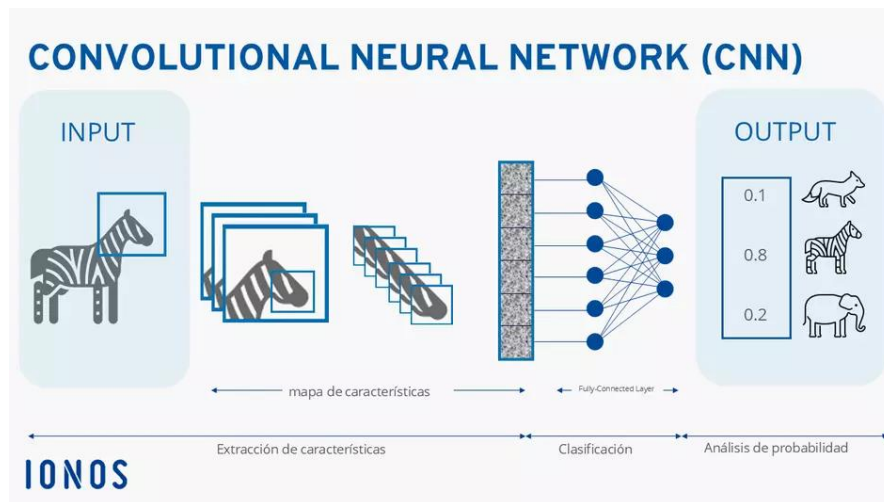


Fig. 10 Red neural convolucional

Fuente: IONOS

Herramientas de visualización

El concepto de herramientas de visualización se creó con el fin de modificar datos numéricos complejos, esto permite representarlos de manera gráfica o visual, fáciles de entender. Su objetivo principal identificar patrones ocultos, comprender información compleja y ayuda a la toma de decisiones. Son ampliamente utilizadas en áreas de negocios, ciencia y tecnología[39]. El uso de los sistemas de información geográfica ha crecido significativamente en los últimos años, y gracias a ello, hoy en día existen una gran cantidad de herramientas de georreferenciación que permiten visualizar imágenes mediante drones o directamente desde imágenes de satélite mediante coordenadas espaciales GPS. Actualmente existen diferentes tipos de licencias de software, hay licencias privadas y licencias libres[40].

Flask

El concepto de flask es microframework escrito en Python, que permite desarrollar de una manera muy sencilla y fácil desde pequeños sitios web, aplicaciones, prototipos hasta APIs interfaces para que diferentes programas se comuniquen entre sí. Siendo así una herramienta muy eficaz con el fin de facilitar la construcción rápida y sencilla. Esta ofrece solo lo básico pero permite agregar extensiones según las necesidades no solo eso sino que también es muy fácil de usar ofrece buena documentación y rapidez para crear proyectos[41].

IMÁGENES SATELITALES

La idea de contexto de las imágenes satelitales es capturas de la superficie terrestre por satélites en el espacio. Con estas imágenes permiten observar y analizar fenómenos naturales, cambios climáticos, ver como se encuentra el suelo, estructuras humanas y entender como está cambiando el planeta. Son bastante utilizadas en entornos como la meteorología, la agricultura, la cartografía y la gestión de desastres naturales, pueden captar diferentes tipos de información, como luz visible, infrarroja o de radar, son muy importantes para el monitoreo y estudio de la tierra o de otros planetas[42].

Imagen satelital multispectral

El concepto de imágenes satelitales multispectrales es capturas de la superficie terrestre tomadas por satélites que capturan esta información en varias bandas de espectro electromagnético como la luz visible y el infrarrojo, entre otras. cada banda resalta diferentes características del terreno, dando así a conocer detalles como tipos de vegetación, cuerpos de agua, suelos o áreas urbanas, esto no sería visible con una imagen convencional[43], [44]. Aquellas imágenes representan un

mismo objeto a través de diferentes longitudes de onda, formadas por bandas que no necesariamente son contiguas entre sí. Esto tiene como objetivo facilitar su uso y comprensión, permitiendo procesar las bandas espectrales de manera que se obtenga una imagen apropiada para los objetivos de la investigación.

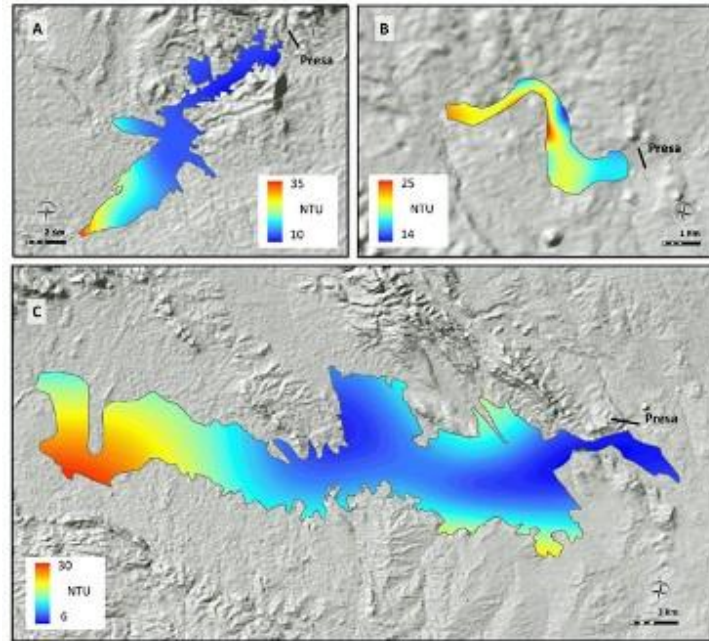


Fig. 11 Imagen multiespectral

Fuente: Universidad Nacional De Costa Rica

Satélite

Si se habla de satélite es un objeto que órbita alrededor de un cuerpo celeste, como la tierra. Existen diferentes satélites naturales, como la luna, y artificiales, que son realizados por el hombre. Estos satélites artificiales cumplen varias funciones importantes, que facilitan las comunicaciones, observar y monitorear el medio ambiente, se puede estudiar el clima, apoyar la navegación GPS y a científicos con investigaciones científicas en el espacio[45]. Por medio de los satélites, es viable obtener datos muy valiosos para el desarrollo de la tecnología, la ciencia y en la vida diaria.

Reflectancia

El concepto de reflectancia es una medida física de la proporción de energía electromagnética reflejada por una superficie después de la exposición a la energía electromagnética incidente. Se define por la relación entre la potencia electromagnética incidente y reflejada. Cuando se trata de luz visible, la reflectancia se expresa como un número entre 0 y 1. Un valor de 0 significa que la superficie no refleja luz, mientras que un valor de 1 significa que la superficie refleja toda la luz

incidente[46]. La reflectancia se puede utilizar para evaluar la calidad del agua. Un estudio realizado por María Isabel González[47] encontró que los cambios en la reflectancia del agua se pueden utilizar para monitorear la calidad del agua. En el caso de un aumento en la reflectancia del agua puede indicar una disminución en la calidad del agua debido a la contaminación o la sedimentación.



Fig. 12 Reflectancia

Fuente: Continuemos estudiando.

C. VARIABLES DEL ESTUDIO

VARIABLES INDEPENDIENTES

- Algoritmo de Forecasting
- Algoritmo de preprocesamiento
- Imágenes satelitales

VARIABLES DEPENDIENTES

- NDWI (Índice De Agua De Diferencia Normalizada)
- MSE (El error cuadrático medio)
- R^2 (Coeficiente de determinación)
- MAE (El error absoluto de medición)
- Tiempo de predicción

Definición nominal de variables

VARIABLES INDEPENDIENTES

Algoritmos de Forecasting: son técnicas computacionales que combinan modelos matemáticos y estadísticos para estimar o predecir patrones basándose en datos históricos o del presente[35]. Estos algoritmos, son útiles en el análisis de series temporales, permiten calcular tendencias, técnicas o comportamientos de variables futuras específicas. Herramientas más comunes como GRU para series con tendencia y estacionalidad, promedio móvil, regresión lineal para relaciones lineales simples, redes neuronales y modelos LSTM redes neuronales para datos secuenciales complejos son aplicados en campos de economía, logística para anticipar la demanda, precio o comportamientos futuros, teniendo una mejor toma de decisiones y planificar estrategias en contextos dinámicos e inciertos[48], como la gestión de recurso hídrico mediante la predicción de indicadores como el NDWI.

Algoritmos de preprocesamiento: son técnicas computacionales esenciales para preparar y transformar datos obtenidos mejorando su calidad, consistencia y utilidad de los datos, para aplicar modelo de análisis, predicción, Machine Learning o Deep Learning, con el propósito de obtener resultados más precisos. Requieren métodos de limpieza de datos, manejo de valores faltantes, normalización, transformación de variables y codificación de datos categóricos. En imágenes satelitales, aplican correcciones radiométricas y filtrado, mientras que en series temporales realizan suavizado o interpolación[49], [50], [51]. Los algoritmos son muy importantes porque los modelos predictivos requieren datos bien estructurados y sin errores para funcionar correctamente, un buen preprocesamiento puede marcar la diferencia entre un modelo exitoso y uno ineficaz.

Imágenes satelitales: son resultados generados por sensores instalados en satélites que orbitan la tierra, que recolectan información de diferentes bandas del espectro electromagnético tales como el visible, infrarrojo cercano e infrarrojo medio. Las imágenes ofrecen información multiespectral y multitemporal lo que facilita observar la superficie terrestre y a su vez analizar fenómenos ambientales, agrícolas, hídricos y urbanos[42]. En esta investigación, las imágenes satelitales obtenidas de los sensores Landsat y MODIS representan la base de datos fundamental de la cual se obtuvieron los valores del NDWI, permitiendo así estudiar la dinámica hídrica a través del tiempo y entrenar modelos de predicción para anticipar su comportamiento en el futuro.

VARIABLES DEPENDIENTES

NDWI (Índice De Agua De Diferencia Normalizada): es un índice espectral utilizado para identificar la presencia de agua superficial y evaluar la humedad en la vegetación. Se calcula a partir de la diferencia entre las bandas del infrarrojo cercano y el verde de imágenes satelitales donde valores positivos indican mayor contenido de agua, proporcionando información clave sobre el estado hídrico de áreas específicas[28].

MSE (El error cuadrático medio): es un indicador estadístico que evalúa la diferencia entre los valores predichos y los valores reales. Se obtiene calculando la raíz cuadrada del promedio de los errores elevados al cuadrado[52]. En problemas de regresión se los utiliza bastante para evaluar la exactitud de predicción de los modelos. Un RMSE menor indica una mejor capacidad de predicción del modelo[53]. Por lo tanto, en este estudio se utilizó para medir el grado de error entre el NDWI predicho por los algoritmos de Forecasting y los valores obtenidos de las imágenes satelitales.

R² (Coeficiente de determinación): es una medida estadística que indica qué proporción de la variación de una variable dependiente es explicada por el modelo. Su valor va de 0 a 1, donde 1 representa una predicción perfecta. Un R² alto implica que el modelo captura bien la tendencia de los datos[54]. En esta investigación, se lo utilizó para medir la capacidad del modelo de Forecasting para explicar la variación en los valores de NDWI.

MAE (El error absoluto de medición): El MAE es una métrica utilizada para evaluar la precisión de un modelo de predicción, especialmente en problemas de regresión. Mide el promedio de las diferencias absolutas entre los valores reales y los valores predichos por el modelo. Al tomar el valor absoluto, evita que los errores positivos y negativos se compensen entre sí, proporcionando una interpretación clara del error promedio en las mismas unidades de la variable evaluada. Un valor de MAE más cercano a cero indica un mejor desempeño del modelo[55].

Tiempo de predicción: hace referencia a la cantidad de tiempo que un modelo de aprendizaje automático tarda en realizar una predicción de un algoritmo para procesar los datos de entrada y generar una salida predictiva. Casi siempre se miden en unidades de tiempo como segundos, milisegundos, o incluso hasta macrosegundos, esto depende de la rapidez del algoritmo y la complejidad de los datos. Esta medida es vital para evaluar la eficiencia computacional de un modelo, especialmente en aplicaciones donde se requiere una respuesta inmediata, como en la gestión de recursos hídricos o la prevención de desastres naturales[56].

Definición operativa de variables

VARIABLES INDEPENDIENTES

Algoritmos de Forecasting: es una variable que define los parámetros utilizados para categorizar los métodos de predicción. Estos parámetros es posible que sean de una naturaleza técnica, como el modelo de matemática utilizado, o de una naturaleza más amplias, como el propósito del algoritmo o la clase de datos que se tratan.

Condiciones técnicas

- Los principios técnicos se fundan en las particularidades de la tecnología de predicción. Por ejemplo, los métodos de predicción se pueden categorizar en base al modelo de matemática utilizado, que puede ser lineal, no lineal o probabilístico[57].

Estos algoritmos fueron implementados para predecir los valores futuros del NDWI, utilizando como base los datos históricos extraídos de las imágenes satelitales de Landsat y MODIS. Aquellos algoritmos integran modelos de aprendizaje automático y profundo, como lo son, LSTM, GRU, MLP. Su implementación se realizó mediante el lenguaje de programación de Python, integrando bibliotecas especializadas como Scikit-learn, TensorFlow y Keras, en un entorno de desarrollo como Google Colab. Cada modelo se entrenó utilizando series temporales construidas a partir de imágenes multiespectrales preprocesadas, organizadas en una base de datos PostgreSQL. El desempeño de estos algoritmos se midió mediante indicadores como MSE, R^2 , MAE y tiempo de predicción, que fueron fundamentales para comparar su capacidad de calcular con exactitud los valores futuros del NDWI teniendo una buena predicción.

Algoritmos de preprocesamiento: es una variable que especifica los pasos tomados para transformar datos sin procesar a un formato que sea más adecuado y comprensible para los algoritmos de aprendizaje automático.

Los pasos de pre procesamiento pueden incluir

- Limpieza de datos: elimine errores, valores atípicos y datos faltantes.
- Normalización de datos: escalar los datos para que tengan rangos similares.
- Codificación de datos: Representar datos de una manera que el algoritmo pueda entender.
- Reducción de dimensionalidad: reducir la cantidad de características de datos sin perder información importante[49], [51], [58].

Los algoritmos de preprocesamiento fueron aplicados como etapa inicial del flujo de trabajo para preparar los datos provenientes de las imágenes satelitales (Landsat y MODIS) antes de utilizar los modelos de Forecasting. Este proceso incluyó técnicas de limpieza, normalización, interpolación de datos faltantes, correcciones radiométricas y filtrado espacial, con el objetivo de mejorar la calidad de los datos y asegurar que cumplan con los requisitos de entrada de los modelos de aprendizaje automático.

Imágenes satelitales: son visualizaciones de la superficie terrestre capturadas por sensores en satélites, usadas para observar, analizar y supervisar fenómenos naturales o actividades humanas desde el espacio[42]. Las imágenes satelitales utilizadas en esta investigación corresponden a registros multiespectrales de los sensores Landsat (2003–2014) y MODIS (2005–2015), previamente descargadas y almacenadas por la Universidad de Nariño desde fuentes oficiales como la NASA y el USGS. Estas imágenes contienen información en diferentes bandas del espectro electromagnético, especialmente la banda verde y la del infrarrojo cercano (NIR), necesarias para calcular el Índice de Agua de Diferencia Normalizada (NDWI). El conjunto final de datos estructurados sirvió como base para la predicción del NDWI, permitiendo analizar la evolución hídrica de zonas específicas del departamento de Nariño.

VARIABLES DEPENDIENTES

NDWI (Índice De Agua De Diferencia Normalizada): es una característica que define los métodos que se emplean para calcular el NDWI. Estos parámetros incluyen los límites de longitud de onda de los espectros utilizados, la forma en que se normaliza y se interpreta el número.

El NDWI se calcula utilizando la siguiente fórmula:

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (1)$$

Ecuación 1 NDWI

- Green es el valor reflectante de la banda verde.
- NIR es el valor que corresponde al infrarrojo cercano.
- NDWI se puede equiparar para hacer más comparable las diferentes imágenes. La normalización se lleva a cabo mediante la división del NDWI por su mayor valor posible, que es 1.

- El NDWI es un parámetro de H2O que tiene como objetivo categorizar y calcular la calidad del H2O. Los índices altos de NDWI representan una mayor calidad del agua, en tanto que los índices bajos representan una menor calidad del agua[59], [2], [60].

MSE (El error cuadrático medio): indicador numérico que mide el promedio de los errores al cuadrado entre valores reales y valores predichos, utilizado para evaluar la precisión de modelos de predicción[61]. Se utilizó como una métrica de evaluación del rendimiento de los modelos de Forecasting aplicados en la predicción del NDWI. Esta métrica se calculó a partir de la diferencia entre los valores reales del NDWI, obtenidos directamente de las imágenes satelitales, y los valores predichos por cada algoritmo evaluado.

El MSE se calcula utilizando la siguiente fórmula:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

Ecuación 2. MSE

- y_i representa el valor inicial del NDWI.
- $\hat{y}_i$ es el valor predicho del NDWI.
- n la cantidad de observaciones

R² (Coeficiente de determinación): medida que indica qué porcentaje de la variación de una variable dependiente es explicado por un modelo; cuanto más cerca esté de 1, mejor es el ajuste del modelo[54]. Será utilizado como métrica estadística para evaluar el grado en que los modelos de Forecasting explican la variabilidad de los valores reales del NDWI a partir de las predicciones generadas. Esta medida compara la proporción de varianza explicada por el modelo frente a la varianza total observada en los datos reales[36].

El R² se calcula utilizando la siguiente fórmula:

$$R^2 = 1 - \left(\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \right) \quad (3)$$

Ecuación 3. R²

- y_i representa el valor real.
- $\hat{y}_i$ es el valor predicho.
- $\bar{y}$ es la medida de los valores reales, este varía entre 0 y 1.

MAE (Error Absoluto Medio): es una métrica estadística utilizada para evaluar la precisión de modelos de predicción en series temporales. Mide el promedio de las diferencias absolutas entre los valores reales observados y los valores estimados por el modelo, sin considerar la dirección del error. Esta característica permite interpretar directamente el error promedio en las mismas unidades de la variable analizada. Un valor de MAE cercano a cero indica mayor exactitud del modelo[55], [62]. Se utilizó para medir el error promedio entre la serie real del NDWI y la serie predicha, permitiendo evaluar la precisión de los modelos de Forecasting.

El MAE se calcula utilizando la siguiente fórmula:

$$MSE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

Ecuación 4 MAE

Donde:

1. n = número total de observaciones
2. y_i = valor real observado
3. $\hat{y}_i$ = valor predicho por el modelo

Tiempo de predicción: es el intervalo que transcurre desde que se ingresa un dato en un modelo hasta que este entregue una predicción; es clave para evaluar la rapidez de un sistema de inteligencia artificial o de análisis de datos. Esta métrica evaluó la eficiencia computacional de los modelos, especialmente en contextos donde la rapidez en la respuesta es crucial para la toma de decisiones informadas en la gestión hídrica[56]. Se calculó el intervalo que tarda cada algoritmo de Forecasting en generar una salida predicha del NDWI, una vez completado el entrenamiento del modelo.

D. FORMULACIÓN DE HIPOTESIS

1) Hipótesis de investigación

La eficiencia medida a través de métricas como: el Error Cuadrático Medio (MSE), Coeficiente de Determinación (R^2), Error Absoluto Medio (MAE) y tiempo de predicción es superior a un modelo aleatorio (0.5).

2) Hipótesis nula

La eficiencia medida a través de métricas como: el Error Cuadrático Medio (MSE), Coeficiente de Determinación (R^2), Error Absoluto Medio (MAE) y tiempo de predicción no es superior a un modelo aleatorio (0.5).

3) Hipótesis alterna

Existe una diferencia significativa en la capacidad predictiva de los algoritmos de aprendizaje automático al estimar los valores futuros del NDWI a partir de imágenes satelitales, mediante métricas como MSE, R^2 , MAE y tiempo de predicción.

III. METODOLOGÍA

A. PARADIGMA

La presente investigación se enmarco dentro del paradigma positivista, ya que se fundamenta en el análisis de datos cuantitativos objetivos obtenidos mediante imágenes satelitales y el uso de algoritmos de aprendizaje automático. Esta perspectiva busco identificar patrones, medir el desempeño predictivo de diferentes modelos y establecer relaciones verificables entre variables, con el propósito de generar conocimiento reproducible y comprobable que contribuya a una gestión informada del recurso hídrico. El paradigma positivista considera que la realidad es objetiva y medible, y que el conocimiento se obtiene mediante métodos empíricos y cuantificables[63].

B. ENFOQUE

El enfoque de la presente investigación es cuantitativo, debido a que se fundamenta en la observación y análisis de hechos reales, medibles y verificables, mediante la recolección y procesamiento de datos numéricos. Este enfoque permitió abordar el estudio del comportamiento de NDWI mediante el estudio de series temporales extraídas de imágenes satelitales como lo son los sensores de Landsat 21GB y MODIS 79GB, utilizando un total de 446 registros obtenidos mediante un script SQL aplicados a una base de datos estructurada. El conjunto de datos utiliza variables numéricas reales derivadas de bandas espectrales, como las bandas del sensor Landsat 7, que abarcan desde la luz visible azul, verde, roja hasta el infrarrojo cercano, medio y térmico. Estas variables permitieron construir y validar modelos predictivos mediante técnicas de aprendizaje automático, con el fin de generar estimaciones del comportamiento hídrico en áreas de estudio específicas. La investigación cuantitativa considera que el conocimiento debe ser objetivo, y que este se genera a partir de un proceso deductivo en el que, a través de la medicación numérica y el análisis estadístico inferencial, se prueban hipótesis previamente formuladas.[64]

C. MÉTODO

El método de la investigación se desarrolló bajo el enfoque científico, el cual permitió abordar el problema de investigación de manera estructurada mediante: análisis y validación de los resultados obtenidos, para el preprocesamiento de la información se empleó la metodología KDD (Knowledge Discovery in Databases), la cual permitió extraer conocimientos útiles a partir de grandes volúmenes de datos mediante un conjunto de etapas definidas. Esta metodología fue seleccionada

debido a su pertinencia para el tratamiento de datos satelitales y la generación de series temporales del NDWI. Para determinar la fase de despliegue del modelo se tomó como referencia los principios de la metodología CRISP-DM, con el fin de estructurar la implantación del sistema en un entorno aplicado.

D. TIPO DE INVESTIGACIÓN

La investigación es de tipo experimental, ya que se buscó comprobar que algoritmo de Forecasting predice con mayor precisión la variación del NDWI utilizando imágenes satelitales multiespectrales. Para esto, se llevaron a cabo varios experimentos aplicando diferentes modelos de predicción, evaluando tanto la precisión de sus resultados como el tiempo que requieren para generar las predicciones. Lo cual permitió identificar cuál modelo ofrece el mejor desempeño en términos de exactitud y eficiencia, aportando herramientas efectivas para el monitoreo y análisis del recurso hídrico.

E. DISEÑO DE LA INVESTIGACIÓN

El diseño de investigación se fundamentó en una metodología experimental, ya que se realizaron diferentes pruebas de rendimiento y precisión sobre un conjunto de algoritmos de Forecasting aplicados a la predicción de la variación del NDWI a partir de imágenes satelitales multiespectrales. El proyecto se estructuro a través de un proceso ordenado que incluye desde la recolección y procesamiento de las imágenes satelitales, el cálculo del NDWI, la generación de series temporales, hasta la aplicación y evaluación comparativa de los modelos predictivos. Finalmente, se visualizarán espacialmente los resultados obtenidos para facilitar su análisis e interpretación. Las fases del proceso son las siguientes.

TABLA I. METODOLOGÍA PRELIMINAR

FASE	ACTIVIDAD	RESULTADOS	RECURSOS
Preprocesar datos	Adquirir datos de imágenes satelitales (Landsat y MODIS) en Dataframes.	Conjunto de datos de series temporales de NDWI, organizado espacialmente y temporalmente.	Python, bibliotecas (pandas, Geopandas, Rasterio), Google Colab, datos de imágenes satelitales (Landsat, MODIS)
	Transformar y unir segmentos de imágenes para el área de estudio (Nariño).		
	Construir un conjunto de datos unificado espacial y temporalmente con valores de NDWI.		

Entrenar algoritmos de Forecasting	Entrenar modelos de Forecasting (MLP, LSTM, GRU) con el conjunto de datos de series temporales de NDWI.	Modelos de Forecasting entrenados. Predicciones de NDWI con sus métricas de evaluación (MAE, MSE, R ²).	Python, bibliotecas (TensorFlow, Scikit-Learn), Google Colab
	Optimizar hiperparámetros de los modelos.		
	Generar predicciones de NDWI a futuro.		
Validar y clasificar resultados	Obtener métricas de evaluación de las predicciones de NDWI.	Gráficos de series temporales de NDWI reales vs. predichos. tablas comparativas de métricas de los modelos. mapas de NDWI predicho y real.	Python, bibliotecas (Matplotlib, Seaborn), Google Colab, herramientas de visualización (QGIS)
	Comparar el rendimiento de los diferentes algoritmos de Forecasting.		
	Visualizar las predicciones y los valores reales de NDWI.		
Implementar aplicación web con visualización topológica	Diseñar e implementar una aplicación web para visualizar las predicciones de NDWI.	Aplicación web funcional con visor de mapas interactivo. interfaz para visualizar la topología del agua predicha.	Python (Flask), JavaScript, bibliotecas de mapas (Leaflet, OpenLayers), base de datos (PostgreSQL)
	Integrar un visor de mapas para la representación geoespacial de la topología del agua.		
	Permitir la interacción del usuario con las predicciones.		

F. POBLACIÓN

Dado que la naturaleza del estudio es netamente computacional y basada en datos, no se necesitó una población en el sentido tradicional. En lugar de ello, el enfoque se centró en el uso de información geoespacial como insumo principal, recurriendo al análisis de series temporales y datos satelitales ya existentes. Este enfoque permitió evaluar el desempeño de distintos algoritmos de Forecasting aplicados a bases de datos previamente recopiladas, sin necesidad de recurrir a una muestra poblacional convencional. En esta investigación la población estuvo conformada por el conjunto de datos multiespectrales provenientes de sensores satelitales Landsat (2003–2014) y (MODIS 2005–2015), previamente recolectados por la Universidad de Nariño. Estas imágenes representan mediciones espaciales y temporales del comportamiento de cuerpos de agua superficial

en diferentes zonas del departamento de Nariño, constituyendo la base para el análisis y predicción del NDWI

G. MUESTRA

La muestra que se utilizó en esta investigación estuvo compuesta por conjuntos seleccionados de datos multiespectrales provenientes de sensores satelitales Landsat (2003–2014) y MODIS (2005–2015), proporcionados por la Universidad de Nariño. Estos datos correspondieron a áreas geográficas del departamento de Nariño que tuvieron cuerpos de agua superficial y presentaron una adecuada continuidad temporal y calidad espectral para el análisis. La selección de las muestras se realizó aplicando criterios técnicos como la baja nubosidad, la disponibilidad de bandas espectrales necesarias para calcular el NDWI, y la relevancia geográfica para el estudio. Asimismo, se aplicó técnicas de limpieza, interpolación y suavizado de datos con el fin de construir series temporales consistentes que permitieron evaluar el desempeño de distintos algoritmos de Forecasting. Estas muestras permitieron obtener la base para el entrenamiento, validación y prueba de los modelos predictivos propuestos.

H. TÉCNICAS DE RECOLECCIÓN DE INFORMACIÓN

Este proyecto cuenta con un dataset de imágenes satelitales multiespectrales por tanto no necesito técnicas de recolección, esta información se obtuvo a partir de bases de datos compartidas por la Universidad de Nariño, que contienen imágenes satelitales descargadas desde plataformas oficiales como la NASA. Estas imágenes fueron tratadas con técnicas de preprocesamiento geoespacial para la extracción de bandas específicas, cálculo del NDWI y estructuración de series temporales que alimentarán los modelos predictivos.

Validez de la técnica

La validez de esta investigación se respaldó en el uso de información de los sensores satelitales Landsat y MODIS, reconocidas internacionalmente por su calidad y confiabilidad. Landsat proporciona imágenes multiespectrales de alta resolución espacial y pasa cada 16 días, mientras que MODIS ofrece datos con alta frecuencia temporal, permitiendo el seguimiento continuo de cuerpos de agua y este pasa cada 1 o 2 días. Estas características fueron fundamentales para calcular el NDWI con precisión y consistencia. Además, la accesibilidad y uso generalizado de estas fuentes en investigaciones científicas refuerzan su validez como insumo para el desarrollo y evaluación.

Confiabilidad de la técnica

En esta investigación se utilizó información que proviene de bases de datos altamente conocidas, como las imágenes satelitales generadas por los sensores Landsat (2003–2014) y MODIS (2005–2015), gestionadas por instituciones de gran trayectoria como la NASA y el Servicio Geológico de los Estados Unidos (USGS). Estas entidades son ejemplares internacionales en el monitoreo de la Tierra, lo que concede un respaldo a la calidad y precisión de los datos utilizados. Además, estos datos han sido aplicados en varios estudios científicos, siendo validados mediante comparaciones con mediciones reales tomadas en el área, lo que aporta un nivel adicional de confianza sobre su fiabilidad. Este acontecimiento hace que estas bases estén disponibles de forma abierta para la comunidad científica fomenta la transparencia, permite la replicación de estudios y fortalece la credibilidad de los resultados obtenidos. El uso de información proveniente de estos sensores no solo garantiza confiabilidad, sino que también permitió avanzar de manera responsable en el análisis y predicción del comportamiento de cuerpos de agua en el departamento de Nariño, dentro del marco del NDWI y los métodos de Forecasting.

I. INSTRUMENTO DE RECOLECCIÓN DE DATOS

Este estudio no utilizó instrumentos de recolección de datos si no que utilizó los datos recolectados y suministrados por la Universidad de Nariño, no se emplearon instrumentos directos de recolección en el área ni plataformas de descarga satelital. Los instrumentos utilizados se limitan a herramientas computacionales aplicadas al tratamiento y preprocesamiento de los datos entregados, tales como programas de análisis estadístico, lenguajes de programación y software especializado para estructurar y procesar series temporales del NDWI. Estas herramientas permitieron preparar la información para el entrenamiento y validación de modelos de predicción, sin modificar la integridad de los datos originales.

METODOLOGÍA DE DESARROLLO

Se realizó recolección, procesamiento, análisis y validación de los datos. Se adoptó como base la metodología KDD (Knowledge Discovery in Databases). Para la fase de implementación del modelo predictivo y su despliegue, se tomaron principios de la metodología CRISP-DM.

En la Fig. 13, se describe el procedimiento metodológico desarrollado para cada uno de los objetivos planteados del proyecto.

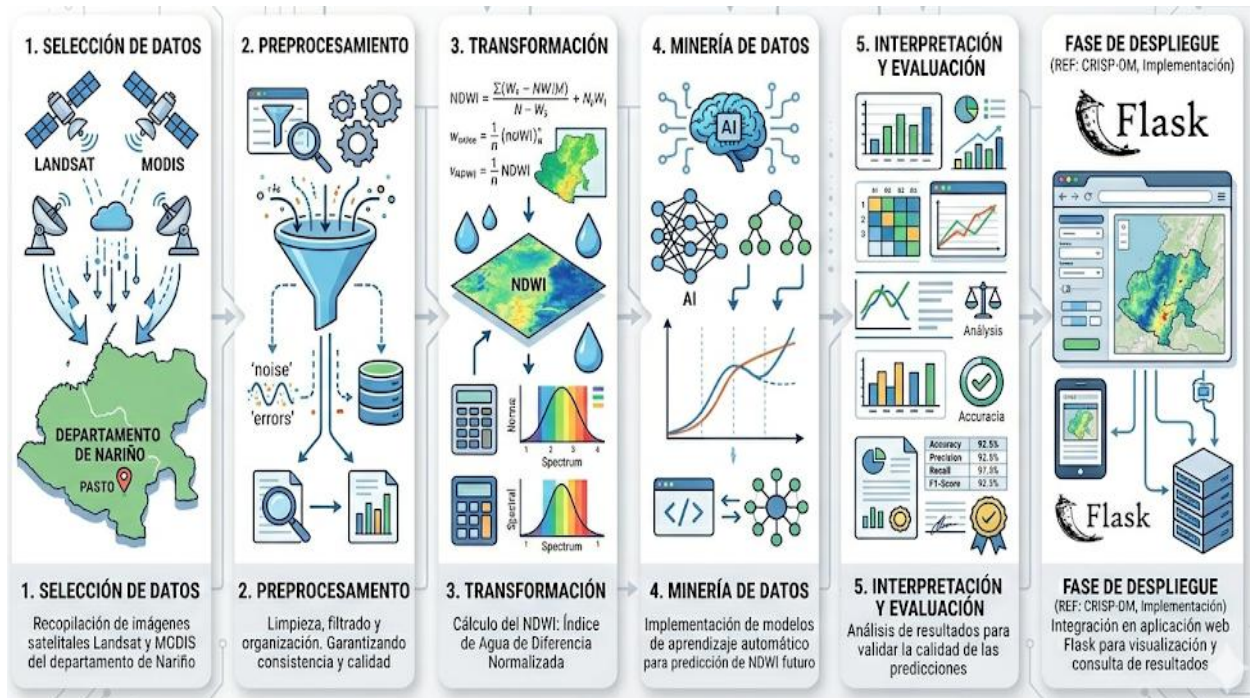


Fig. 13 Flujo de metodología de investigación

Fuente: Elaboración propia mediante herramientas de inteligencia artificial.

Las fases se describen a continuación:

FASE 1. SELECCIÓN Y RECOLECCIÓN DE DATOS

En esta etapa se realizó la recuperación de los conjuntos de datos satelitales correspondientes a los sensores Landsat y MODIS, los cuales fueron proporcionados por el grupo de investigación GIEEE de la Universidad de Nariño.

Para obtener esta información fue necesario restaurar localmente la base de datos mediante PostgreSQL, lo que permitió facilitar su consulta, segmentación y posterior procesamiento. Los datos recuperados corresponden a imágenes satelitales multiespectrales asociadas al departamento de Nariño, donde cada registro contiene información espacial, temporal y espectral, representada por las coordenadas geográficas (latitud y longitud), la fecha de captura y los valores de reflectancia de las diferentes bandas del sensor satelital.

En la TABLA II se presenta una descripción general de las características principales de la base de datos utilizada en la investigación.

Segmentación temporal de los datos

Debido al gran volumen de información satelital disponible, se implementó un proceso de segmentación temporal de los datos con el fin de facilitar su almacenamiento, consulta y posterior procesamiento.

Para ello se utilizó una estrategia de partición basada en rangos de fechas, donde cada subconjunto de datos corresponde a un periodo temporal específico. La organización de la información se realizó mediante segmentos mensuales, considerando como llave primaria la combinación de los atributos espaciales latitud, longitud y el atributo temporal datecapture, lo que permite identificar de manera única cada observación dentro del conjunto de datos.

Generación de tablas mediante consultas SQL

La segmentación temporal fue implementada mediante consultas SQL ejecutadas directamente sobre la base de datos restaurada. Estas consultas permitieron extraer subconjuntos de información correspondientes a intervalos específicos de tiempo y generar nuevas tablas con los atributos requeridos para el procesamiento posterior.

En la TABLA III se presenta un ejemplo de las consultas SQL utilizadas para la obtención de un segmento temporal

TABLA III. CONSULTAS SQL PARA OBTENCIÓN DE SEGMENTOS DE LANDSAT Y MODIS

Dataset	Rango de fecha	Consulta SQL
Landsat	D1=2010-03-01	<i>CREATE table dataset_2010_3 AS</i>
	D2=2010-03-31	<i>SELECT R.datecapture, R.latitude, R.longitude, band1, band2, band3, band4, band5, band6, band7, I.value</i>
		<i>FROM (SELECT datecapture, latitude, longitude, band1, band2, band3, band4, band5, band6, band7</i>
		<i>FROM public.reflectance where datecapture between '2012-01-01' and '2010-03-31') R, (SELECT datecapture, latitude, longitude, value</i>

		<i>FROM public.irradiaceGrid450 where datecapture between '2012-01-01' and '2010-03-31') I WHERE R.datecapture=I.datecapture and R.latitude=I.latitude and R.longitude=I.longitude;</i>
MODIS	D1=2012-01-01	<i>CREATE table dataset_2012_1 AS</i>
	D2=2012-01-31	<i>SELECT R.datecapture, R.latitude, R.longitude, band1, band2, band3, band4, band5, band6, band7,I.value</i>
		<i>FROM (SELECT datecapture, latitude, longitude, band1, band2, band3, band4, band5, band6, band7</i>
		<i>FROM public.bands1_7 where datecapture between '2012-01-01' and '2012-01-31') R,</i>
		<i>(SELECT datecapture, latitude, longitude, value</i>
		<i>FROM public.irradiation where datecapture between '2012-01-01' and '2012-01-31') I</i>
		<i>WHERE R.datecapture=I.datecapture and R.latitude=I.latitude and R.longitude=I.longitude;</i>

Exportación y consolidación de segmentos en archivos CSV

Una vez generadas las tablas correspondientes a cada segmento temporal, se procedió a su exportación en formato CSV con el fin de facilitar su almacenamiento, trazabilidad y posterior uso en herramientas de análisis de datos. Este procedimiento permitió conservar cada segmento como una unidad independiente y preparar la información para su consolidación posterior.

En **TABLA IV** se representa un ejemplo de la consulta utilizada para exportar los segmentos generados

TABLA IV. CONSULTA DE ALMACENAMIENTO

Consulta

COPY

```
(SELECT datecapture, latitude, longitude, band1, band2, band3, band4, band5, band6, band7, value
  FROM public.dataset_2012_1)
```

TO

```
'D:/DB/dataset_2012_1.csv'
```

WITH CSV HEADER;

FASE 3. TRANSFORMACIÓN DE LOS DATOS

Construcción del dataset NDWI para el departamento de Nariño

Este flujo integra diferentes etapas que incluyen la carga de información espacial, la generación de variables temporales, la predicción de bandas espectrales mediante modelos previamente entrenados y el cálculo del índice NDWI. En la Fig. 15 se presenta el flujo general del pipeline utilizado para la generación del dataset.

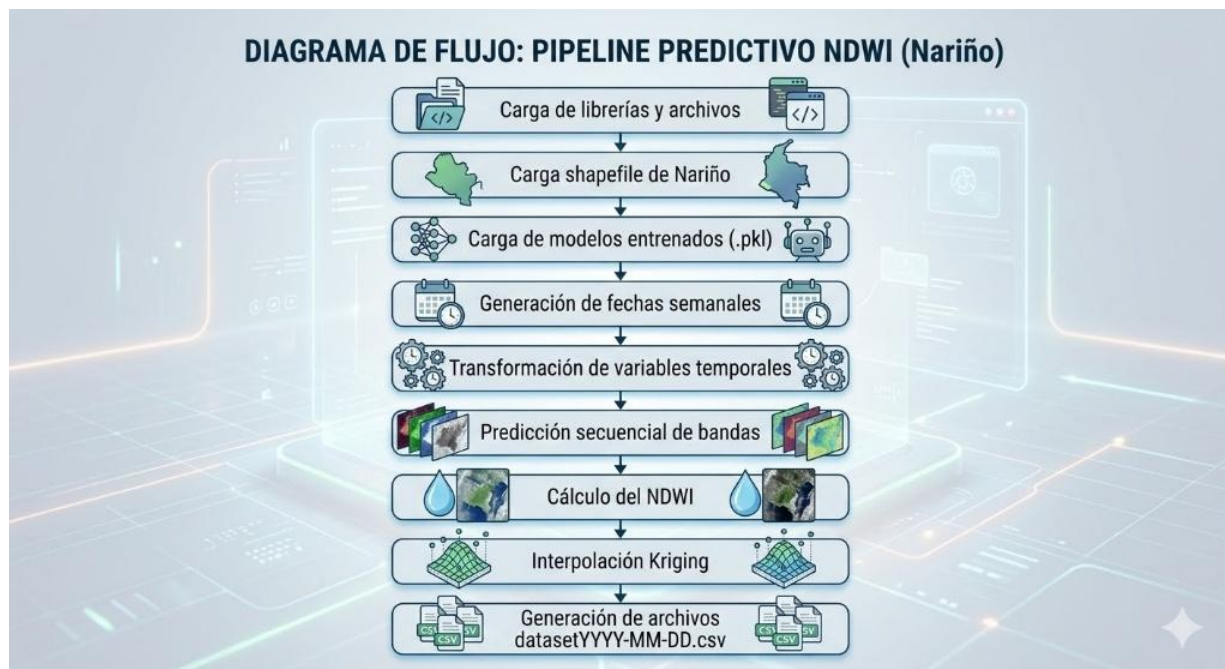


Fig. 15 Flujo de Pipeline para generación del dataset NDWI

Fuente: Elaboración propia mediante herramientas de inteligencia artificial

Modelado y validación de bandas espectrales para el cálculo del NDWI

Con el propósito de garantizar la consistencia de las bandas espectrales y la integridad espacial de los datos asociados al departamento de Nariño, se implementó un proceso de modelado orientado a la estimación de dichas bandas.

Esta etapa se implementó en un entorno en Google Colab, integrando librerías especializadas para análisis de datos, procesamiento geoespacial y aprendizaje automático, con el fin de estructurar un pipeline reproducible de procesamiento.

En el proceso se emplearon diferentes herramientas orientadas al manejo de datos satelitales y al entrenamiento de modelos de estimación de bandas espectrales. En la TABLA V se presentan las principales herramientas utilizadas.

TABLA V. HERRAMIENTAS UTILIZADAS EN EL CALCULO DEL NDWI

Herramienta	Propósito en el proceso
GeoPandas	Lectura del shapefile y manejo geoespacial de Nariño
PyKrige	Interpolación geoestadística (Kriging) para completar vacíos
Pandas / NumPy	Limpieza, transformación y cálculo del NDWI
Matplotlib	Visualización de mapas NDWI y análisis exploratorio
Scikit-learn / sklearnkernels	Entrenamiento y validación de modelos para estimación de bandas
Joblib	Persistencia de modelos (.pkl)

Para la estimación de las bandas espectrales faltantes o inconsistentes, se emplearon modelos de regresión como Support Vector Regression (SVR) y K-Nearest Neighbors (KNN), seleccionados por su capacidad para modelar relaciones no lineales entre variables espectrales y su buen desempeño en conjuntos de datos de alta dimensionalidad. A diferencia de los modelos de forecasting utilizados en el objetivo 2, cuyo propósito es predecir la evolución temporal del NDWI, estos modelos de regresión se enfocan en la reconstrucción puntual de valores dentro de un mismo instante temporal, permitiendo mejorar la calidad e integridad del dataset previo al cálculo del índice.

FASE 4. MINERÍA DE DATOS

Con el propósito de predecir el comportamiento del Índice de Agua de Diferencia Normalizada (NDWI) a partir de series temporales, se desarrolló una fase de minería de datos orientada a la

construcción, sintonización y evaluación de algoritmos de forecasting. Como se presenta en el flujo mostrado en la Fig. 16.

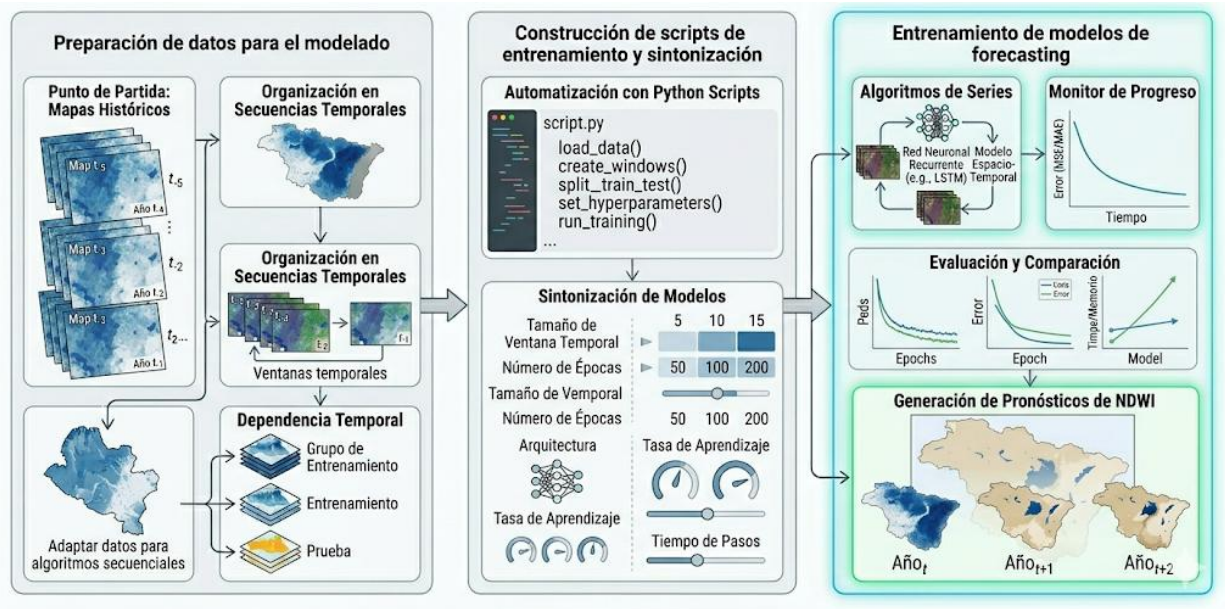


Fig. 16 Flujo para la predicción de NDWI

Fuente: Elaboración propia mediante herramientas de inteligencia artificial

Para ello, se tomaron como base las series temporales generadas en la fase anterior, las cuales contienen la evolución del NDWI en función del tiempo y del espacio dentro del departamento de Nariño. A partir de este conjunto de datos, se estructuraron scripts en Python para automatizar el entrenamiento, ajuste y comparación de diferentes modelos predictivos.

Preparación de datos para el modelado

Como etapa inicial, se organizó el dataset en secuencias temporales, con el fin de adaptar la información a los requerimientos de los algoritmos de series de tiempo. Para ello, se construyeron ventanas temporales de distinta longitud, donde cada secuencia de entrada contiene observaciones históricas del NDWI utilizadas para predecir valores futuros. Este procedimiento permitió representar la dependencia temporal de la variable de estudio y preparar los datos para el entrenamiento de modelos con capacidad de aprendizaje secuencial.

Construcción de scripts de entrenamiento y sintonización

Se desarrollaron scripts orientados a la ejecución de experimentos con distintos algoritmos de forecasting. Estos scripts permitieron automatizar tareas como la carga del dataset, la conformación de ventanas temporales, la división de datos en entrenamiento y prueba, la configuración de

hiperparámetros y la ejecución de los entrenamientos. La sintonización de los modelos se realizó mediante la variación controlada de parámetros asociados a cada algoritmo, tales como el tamaño de la ventana temporal, el número de épocas de entrenamiento, la arquitectura del modelo y otros parámetros propios de cada enfoque evaluado.

Entrenamiento de modelos de forecasting

A partir de los scripts desarrollados, se entrenaron diferentes algoritmos de series temporales orientados a la predicción del NDWI. Entre los modelos evaluados se incluyeron enfoques basados en redes neuronales recurrentes y modelos espacio-temporales, seleccionados por su capacidad para capturar patrones secuenciales y dependencias en los datos.

Cada modelo fue ejecutado bajo distintas configuraciones experimentales, con el fin de analizar su comportamiento frente a las métricas de calidad y su eficiencia computacional, estableciendo así una base para la comparación de los modelos.

FASE 5. INTERPRETACIÓN Y EVALUACIÓN DE LOS MODELOS

Con el propósito de analizar el desempeño de los modelos de series temporales entrenados para la predicción del Índice de Agua de Diferencia Normalizada (NDWI), se desarrolló una fase de interpretación y evaluación orientada a la comparación objetiva de los resultados obtenidos.



Fig. 17 Flujo de evaluación y selección de modelos de predicción del NDWI

Fuente: Elaboración propia mediante herramientas de inteligencia artificial

El flujo presentado en la Fig. 17 sintetiza el proceso metodológico seguido para la evaluación y selección del modelo óptimo.

Construcción de ventanas

En una primera etapa se construyó el conjunto de datos de entrada para los modelos de forecasting a partir de las series temporales generadas en el objetivo anterior. Para ello, se organizaron las observaciones del NDWI en ventanas temporales de distinto tamaño, con el fin de suministrar a los algoritmos secuencias históricas capaces de representar la dinámica temporal del índice.

Evaluación mediante métricas de calidad

Para evaluar el rendimiento de los modelos, se emplearon métricas de regresión ampliamente utilizadas en problemas de predicción continua, tales como el coeficiente de determinación (R^2), el error absoluto medio (MAE) y el error cuadrático medio (MSE).

Estas métricas permitieron cuantificar el grado de ajuste entre los valores reales y los valores predichos del NDWI, facilitando la medición de la precisión de cada modelo.

Medición del tiempo de ejecución

De manera complementaria, se registró el tiempo de ejecución de cada modelo durante el proceso de entrenamiento y evaluación. Este criterio permitió incorporar un análisis de eficiencia computacional, considerando no solo la calidad de las predicciones, sino también el costo temporal asociado a cada algoritmo.

En la TABLA VI se presentan las características del entorno de ejecución utilizadas durante el entrenamiento de los modelos en Google Colab y Google Colab Pro.

TABLA VI. CARACTERÍSTICAS DEL ENTORNO DE EJECUCIÓN EN GOOGLE COLAB

RECURSO	ESPECIFICACIÓN
Plataforma	Google Colab Pro
GPU	NVIDIA Tesla T4 y NVIDIA A100 (según disponibilidad)
Memoria RAM	Entre 12 GB y 25 GB
CPU	Procesadores virtuales de alto rendimiento
Tipo de ejecución	Entorno en la nube con aceleración por hardware

Las características del entorno de ejecución influyen directamente en los tiempos de entrenamiento de los modelos, los cuales pueden variar en función de la disponibilidad de recursos computacionales y la complejidad de cada algoritmo evaluado.

Comparación de modelos

A partir de las métricas obtenidas, se realizó una comparación entre los diferentes modelos evaluados. Este análisis permitió identificar las fortalezas y debilidades de cada enfoque en términos de precisión y eficiencia.

La comparación se llevó a cabo considerando de manera conjunta los valores de R^2 , MAE, MSE y tiempo de ejecución, con el fin de establecer un criterio integral de evaluación.

Selección del modelo óptimo

La selección del modelo se realizó priorizando configuraciones que presentaran altos valores de R^2 , bajos errores (MAE y MSE) y tiempos de ejecución razonables, en concordancia con los objetivos de la investigación.

Con el desarrollo de las fases de minería de datos, interpretación y evaluación, se dio cumplimiento al objetivo específico 2 de la investigación, permitiendo la construcción, sintonización y comparación de modelos de series temporales para la predicción del NDWI, así como la identificación del modelo con mejor compromiso entre métricas de calidad y tiempo de ejecución.

FASE 6. DESPLIEGUE DEL MODELO

Para representar de manera más clara el proceso de desarrollo en la fase de despliegue del modelo predictivo se presenta la Fig. 18 en la cual se muestra el flujo realizado para obtener la visualización del NDWI.

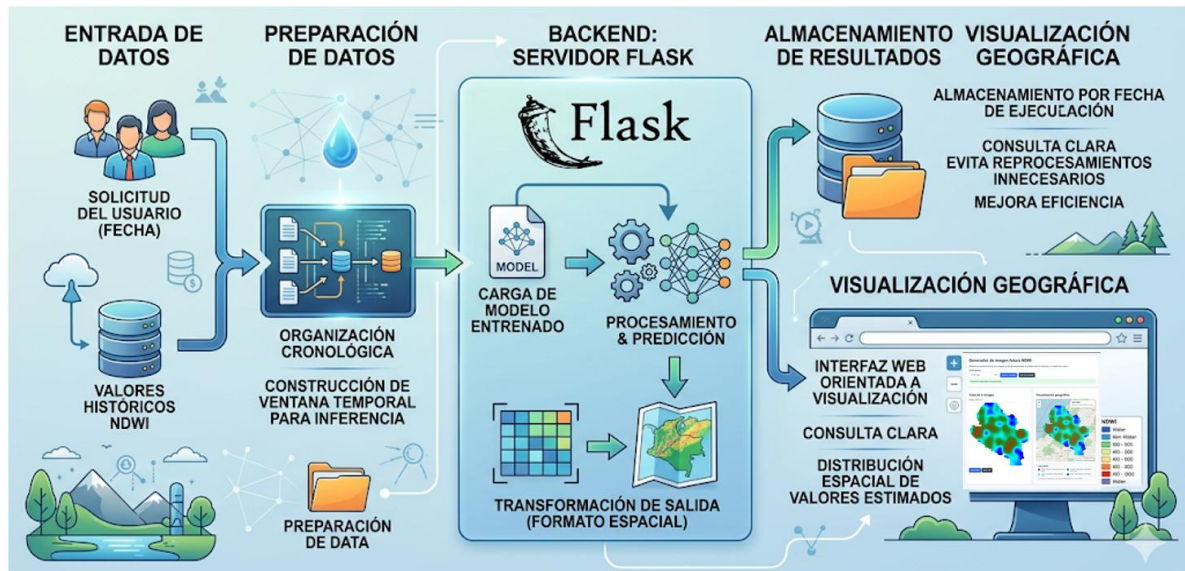


Fig. 18 Flujo de procesamiento para la visualización del NDWI

Fuente: Elaboración propia mediante herramientas de inteligencia artificial

Como se observa en la Fig. 18 Para el cumplimiento de este objetivo, se desarrolló una aplicación web utilizando Flask como framework backend, con el fin de integrar el modelo de predicción previamente entrenado en un entorno funcional de consulta y visualización.

Inicialmente, se integró el modelo predictivo previamente entrenado, correspondiente a una arquitectura ConvLSTM2D, el cual es cargado desde un archivo guardado (model09.keras). Este modelo recibe como entrada una ventana temporal de 7 días con dimensiones espaciales de 52x49 y 8 canales, incluyendo el canal correspondiente al índice NDWI.

Posteriormente, se implementó un mecanismo de validación de fechas, considerando que la última fecha real disponible en el conjunto de datos (Ventana) es el 14 de diciembre de 2014. En este sentido, el sistema solo permite generar predicciones para fechas posteriores a este valor, estableciendo como primera fecha válida el 15 de diciembre de 2014. Se diseñó un proceso de predicción recursiva, mediante el cual el sistema utiliza la última ventana temporal disponible (last_window.npy) para generar la primera predicción, y posteriormente reutiliza las salidas generadas como entrada para estimar fechas futuras más lejanas.

En cuanto a la gestión de resultados, se implementó un mecanismo de almacenamiento estructurado por fechas, donde cada predicción es guardada en formato GeoTIFF (.tif) y como imagen (.png) dentro de directorios organizados. Antes de generar una nueva predicción, el sistema verifica si el resultado ya existe; en caso afirmativo, se reutiliza la información almacenada, optimizando el rendimiento y evitando reprocesamientos innecesarios.

Para la visualización, se desarrolló una interfaz interactiva que permite superponer las predicciones NDWI sobre un mapa base de OpenStreetMap mediante el uso de `L.imageOverlay()` en Leaflet, facilitando la interpretación espacial de los resultados. Asimismo, se incorporó la visualización de estadísticas descriptivas como valores mínimos, medianos y máximos del índice. La aplicación permite la comparación simultánea de dos fechas seleccionadas por el usuario, mostrando los resultados de manera paralela, lo cual fortalece el análisis temporal del comportamiento del índice NDWI.

IV. RESULTADOS DE LA INVESTIGACIÓN

Resultado del objetivo 1

Preprocesar un conjunto de base de datos (Imágenes satelitales de LANSAT Y MODIS) para generar series temporales de NDWI.

De acuerdo a la metodología de desarrollo a continuación se detallan las fases para el objetivo 1:

Fase 1. Selección y recolección: con la restauración en PostgreSQL y las características de las bases de datos Landsat (20GB) y MODIS (70GB)

Fase 2. Preprocesamiento: segmentación temporal por rangos de fecha y la exportación a CSV.

Fase 3. Transformación: generación del dataset, incluyendo el modelado de bandas faltantes mediante modelos de regresión (SVR/KNN) en Google Colab.

Como resultado de estas fases se obtiene el preprocesamiento, la información es almacenada en la base de datos y se organizó en tablas estructuradas, archivos consolidados en formato CSV y datasets organizados espacial y temporalmente, los cuales sirvieron como base para la construcción de las series temporales del NDWI.

Segmentación temporal de los datos satelitales

La segmentación de los datos se formalizó mediante operaciones de álgebra relacional presentadas en las Ecuaciones 5, 6 y 7, donde R representa la relación que contiene los valores de reflectancia espectral y I representa la relación que almacena el valor del índice NDWI calculado para cada registro.

La partición temporal de los datos de reflectancia se define mediante la Ecuación 5

$$R_i = \pi_{p_k, band_1, band_2, \dots, band_7} \sigma_{datecapture \text{ between } D_1 \text{ and } D_2} (R)$$

Ecuación 5 Partición de la reflectancia por rango de fechas

Para definir la segmentación del valor de NDWI asociado a cada registro del dataset se utiliza la Ecuación 6

$$I_i = \pi_{p_k, NDWI} \sigma_{datecapture \text{ between } D_1 \text{ and } D_2} (I)$$

Ecuación 6 segmentación por rango de fecha

Los segmentos generados a partir de las bandas espectrales y del valor del NDWI se integran para conformar el conjunto segmentado como lo muestra la Ecuación 7

$$Ds_i = \pi p_k, band_1, \dots, NDWI \sigma \text{ datecapture between } D_1 \text{ and } D_2 (R_i \times I_i)$$

Ecuación 7 Cruce de los segmentos

En la **TABLA VII** se presentan los principales elementos utilizados en la formulación de las ecuaciones de segmentación temporal.

TABLA VII. ELEMENTOS DE LAS ECUACIONES DE SEGMENTACIÓN

Símbolo	Descripción
<i>PK</i>	Llave primaria compuesta por los atributos espaciales latitude, longitude y la dimensión temporal datecapture.
<i>D₁</i>	Fecha inicial del intervalo temporal considerado en la partición del dataset.
<i>D₂</i>	Fecha final del intervalo temporal considerado en la partición del dataset.
<i>D_{Si}</i>	Conjunto de datos correspondiente al i-ésimo segmento temporal generado.

A partir de esta operación se generaron subconjuntos de datos asociados a periodos mensuales, lo que permitió organizar la información en unidades manejables y mantener la continuidad espacio-temporal de los registros.

Generación de tablas mediante consultas SQL

Como resultado del proceso de segmentación temporal implementado mediante consultas SQL, fue posible dividir el conjunto de datos satelitales en múltiples segmentos organizados por periodos de tiempo específicos. Este procedimiento permitió manejar de forma más eficiente el gran volumen de información proveniente de los sensores Landsat y MODIS, facilitando su almacenamiento y posterior procesamiento. A partir de esta segmentación se generaron tablas correspondientes a los diferentes intervalos temporales analizados.

TABLA VIII. RESULTADOS DE LA SEGMENTACIÓN TEMPORAL

Sensor satelital	Número de tablas generadas	Periodo cubierto	Tiempo de procesamiento
Landsat	120	2004 – 2014	3 horas 20 minutos
MODIS	60	2005, 2007, 2010, 2012 y 2015	6 horas 40 minutos

En la **TABLA VIII** se evidencia la magnitud del volumen de datos procesado durante esta etapa, así como la necesidad de implementar estrategias de segmentación temporal para facilitar la manipulación y organización de la información satelital utilizada en la investigación.

Exportación y consolidación de segmentos en archivos CSV

Una vez generadas las tablas correspondientes a cada segmento temporal, se procedió a exportarlas a archivos en formato CSV con el fin de facilitar su almacenamiento, trazabilidad y manipulación mediante herramientas de análisis de datos.

Posteriormente, los archivos exportados fueron integrados en conjuntos de datos consolidados de mayor tamaño, los cuales preservan la estructura espacial y temporal de la información satelital. Como resultado de este proceso se obtuvieron los archivos Landsat_all.csv y Modis_week_all.csv, los cuales constituyen la base principal para la construcción del dataset orientado al cálculo del NDWI se evidencia en la **TABLA IX**.

TABLA IX. DATASETS CONSOLIDADOS OBTENIDOS DEL PROCESAMIENTO

Dataset	Tamaño	Preprocesamiento
Landsat.csv Modis.csv	65.7Kb 67.6Kb	Unión de segmentos y promedio por latitud y longitud
Landsat_all.csv	690.6Mb	Unión de segmentos
Modis_week_all.csv	2.19Gb	Unión de segmentos suavizados por promedio agrupados por latitud, longitud y semanales

Organización y validación de los datasets iniciales

A partir de los archivos consolidados se realizó una etapa de organización orientada a validar la estructura de los registros, verificar la consistencia de las coordenadas geográficas y conservar la información cronológica. Cada registro quedó conformado por la fecha de captura, la latitud, la longitud y los valores de reflectancia correspondientes a las siete bandas espectrales.

TABLA X. ESTRUCTURA DE LOS DATOS SATELITALES OBTENIDOS

Variable	Descripción
latitude	Coordenada geográfica de latitud
longitude	Coordenada geográfica de longitud
band1	Banda espectral 1
band2	Banda espectral 2
band3	Banda espectral 3
band4	Banda espectral 4
band5	Banda espectral 5
band6	Banda espectral 6
band7	Banda espectral 7
Value	Valor

La estructura presentada en la **TABLA X** describe las variables que conforman el dataset generado a partir del preprocesamiento de las imágenes satelitales. Cada registro contiene información espacial correspondiente a la latitud y longitud del punto observado, junto con los valores de reflectancia de las siete bandas espectrales y el valor asociado del índice NDWI. Un ejemplo de los registros obtenidos del dataset inicial se muestra en la Fig. 19, donde se puede observar la organización tabular de los datos generados.

	A	B	C	D	E	F	G	H	I	J
1	latitude	longitude	band1	band2	band3	band4	band5	band6	band7	value
2	-8789850	184950	0.1051125905	0.08271564244	0.05275554907	0.3155197183	0.1326408839	296.1786069	0.04458619665	203.4
3	-8784900	174600	0.1048404213	0.08573163966	0.05352727115	0.2714632767	0.09024746761	295.9697098	0.02729183302	198.3
4	-8775000	164700	0.1034676758	0.0791469091	0.04893191661	0.2486185447	0.07859331321	296.8426895	0.02209766349	203.1
5	-8775000	169650	0.108848644	0.0906645707	0.05571492217	0.338889248	0.1282274576	296.1239273	0.03979753696	199.1
6	-8775000	174800	0.09652626522	0.07102843495	0.04224698498	0.1947884317	0.05446750733	294.4074912	0.01771206647	196.3
7	-8775000	194850	0.1147502456	0.09446627902	0.06208231261	0.3389840959	0.1331359444	296.8987684	0.04275377112	201.7
8	-8769600	164700	0.1060903768	0.08960354563	0.05748188957	0.3608179474	0.156995398	296.7202432	0.05395090965	199.4
9	-8769600	174600	0.109988866	0.09101283304	0.05940039255	0.3429740757	0.1552977377	296.4705482	0.05494264934	196.3
10	-8769600	184950	0.112143067	0.09338760201	0.05916869817	0.3338078995	0.1551032405	296.4747906	0.05534492327	196.4
11	-8769600	189900	0.107985264	0.08625670443	0.06454178054	0.2773909273	0.1370125929	297.818269	0.05217630288	196.7
12	-8769600	194850	0.1036768505	0.08979321575	0.05372435145	0.2759818265	0.09993877713	296.8413022	0.0302681538	198.5
13	-8764650	154800	0.1066772756	0.08476108172	0.05419595755	0.2744306307	0.1305969463	296.4884722	0.04570082924	198.2
14	-8764650	164700	0.1119534554	0.09625550102	0.06220420392	0.3628123198	0.1828757688	296.2969027	0.06422799926	198.3
15	-8764650	169650	0.112206715	0.09389568833	0.06226469248	0.3504757678	0.1518425228	295.3547427	0.05157909327	197.7
16	-8764650	174600	0.1153344945	0.09622097141	0.06280524034	0.3820510367	0.1867983931	297.0160779	0.0684296131	196.5
17	-8764650	184950	0.1117671188	0.09589087355	0.06139941887	0.3532152207	0.1693304707	296.8444691	0.05893423771	195.8

Fig. 19 Registros del dataset

Construcción del dataset NDWI para el departamento de Nariño

El pipeline de procesamiento permitió la generación del dataset utilizado para el análisis del índice NDWI en el departamento de Nariño.

Modelado y validación de bandas espectrales para el cálculo del NDWI

Con el objetivo de delimitar el área de estudio, se cargó el shapefile del departamento de Nariño, permitiendo extraer los puntos correspondientes a su geometría. Este procedimiento aseguró que el procesamiento posterior se realizará únicamente dentro del territorio analizado, evitando inconsistencias espaciales como lo muestra Fig. 20.

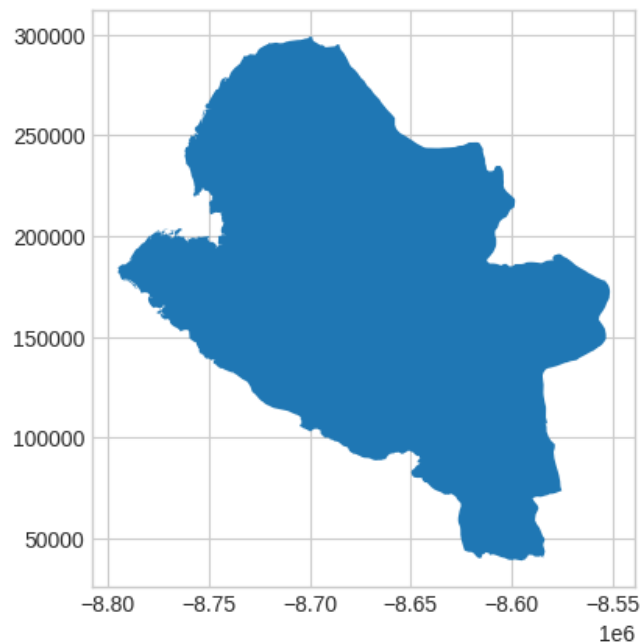


Fig. 20 Shapefile del departamento de Nariño.

Posteriormente, se realizaron procesos de agregación espacial agrupando observaciones por latitud y longitud, lo que permitió generar estadísticas representativas del comportamiento espectral en el territorio.

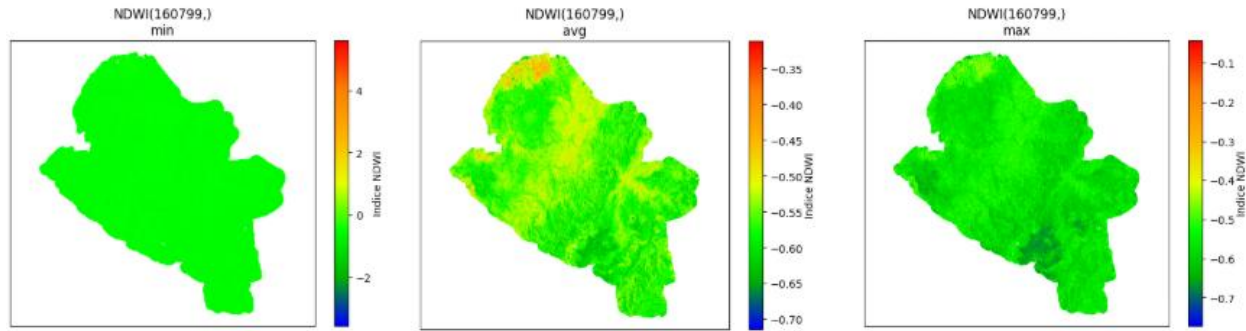


Fig. 21 NDWI agregado por latitud y longitud: (a) mínimo, (b) promedio, (c) máximo

A partir de esta agregación se calcularon valores representativos del NDWI, tales como el valor mínimo, promedio y máximo para cada ubicación espacial. Estos resultados permiten analizar la variabilidad espacial del índice hídrico en el departamento de Nariño, como se muestra en la Fig. 21.

Entrenamiento y evaluación de modelos para estimación de bandas

Para reconstruir la información espectral necesaria en cada instante temporal, se entrenaron modelos de aprendizaje automático capaces de estimar las bandas espectrales a partir de variables espaciales y temporales. Estas variables permiten representar tanto la ubicación geográfica de los registros como su comportamiento a lo largo del tiempo dentro de la serie temporal utilizada en la investigación como se muestra en la **TABLA XI**.

TABLA XI. VARIABLES PARA EL ENTRENAMIENTO DE LOS MODELOS

Variable	Descripción
latitude	Coordenada geográfica de latitud
longitude	Coordenada geográfica de longitud
year	Año correspondiente a la fecha de captura
month	Mes correspondiente a la fecha de captura
day	Día correspondiente a la fecha de captura

Se emplearon diferentes modelos de aprendizaje automático orientados a tareas de regresión con el fin de estimar los valores de las bandas espectrales. Los algoritmos utilizados se presentan en la **TABLA XII**.

TABLA XII. MODELOS UTILIZADOS PARA LA ESTIMACION DE BANDAS ESPECTRALES

Modelo	Descripción
SVR (Support Vector Regression)	Modelo de regresión basado en máquinas de soporte vectorial que permite capturar relaciones no lineales entre variables.
KNN (K-Nearest Neighbors)	Algoritmo de aprendizaje supervisado que estima valores a partir de la proximidad con observaciones similares dentro del conjunto de datos.

Los modelos fueron entrenados utilizando combinaciones progresivas de bandas espectrales como variables de entrada. De esta manera, la predicción de cada banda se realizó de forma secuencial, donde el resultado de una banda sirve como entrada para la estimación de la siguiente.

Para evaluar el desempeño de los modelos se utilizó el coeficiente de determinación (R^2), el cual permite medir el grado de ajuste entre los valores observados y los valores predichos por cada modelo. Los resultados obtenidos se representan en la **TABLA XIII**.

TABLA XIII. DESEMPEÑO (R^2) DE LOS MODELOS DE ESTIMACIÓN DE BANDAS

Banda objetivo	Variables de entrada	Algoritmo	R^2
Band1	lat, lon, año, mes, día (features)	SVR (Pipeline)	0.7688
Band2	Band1	KNN	0.7651
Band3	Band1, Band2	KNN	0.8733
Band4	Band1, Band2, Band3	KNN	0.6853
Band5	Band1–Band4	KNN	0.8041
Band6	Band1–Band5, Band7	SVR	0.5069
Band7	Band1–Band5	KNN	0.9467

Los resultados muestran que los modelos presentan diferentes niveles de desempeño dependiendo de la banda espectral estimada. En particular, el modelo KNN obtuvo el mejor resultado para la banda 7, alcanzando un coeficiente de determinación de $R^2 = 0.9467$, lo que indica una alta capacidad de predicción.

Es importante resaltar que la banda 6 presenta un desempeño inferior en comparación con las demás, con un valor de $R^2 = 0.5069$. Este comportamiento puede estar asociado a la naturaleza espectral de esta banda, la cual es más sensible a condiciones atmosféricas como la nubosidad y

variaciones térmicas, lo que introduce mayor ruido en los datos. Sin embargo, su impacto en el cálculo del NDWI es limitado, dado que este índice se basa principalmente en la relación entre la banda verde y la banda del infrarrojo cercano (NIR), por lo que la menor precisión en esta banda no afecta de manera significativa el resultado final del índice.

Para garantizar la reutilización de los modelos entrenados en etapas posteriores del pipeline de procesamiento, estos fueron almacenados en archivos con extensión .pkl utilizando la librería Joblib. Este procedimiento permite conservar los parámetros aprendidos por cada modelo y facilita su posterior carga durante el proceso de generación del dataset NDWI en la Fig. 22 se muestra los modelos guardados.

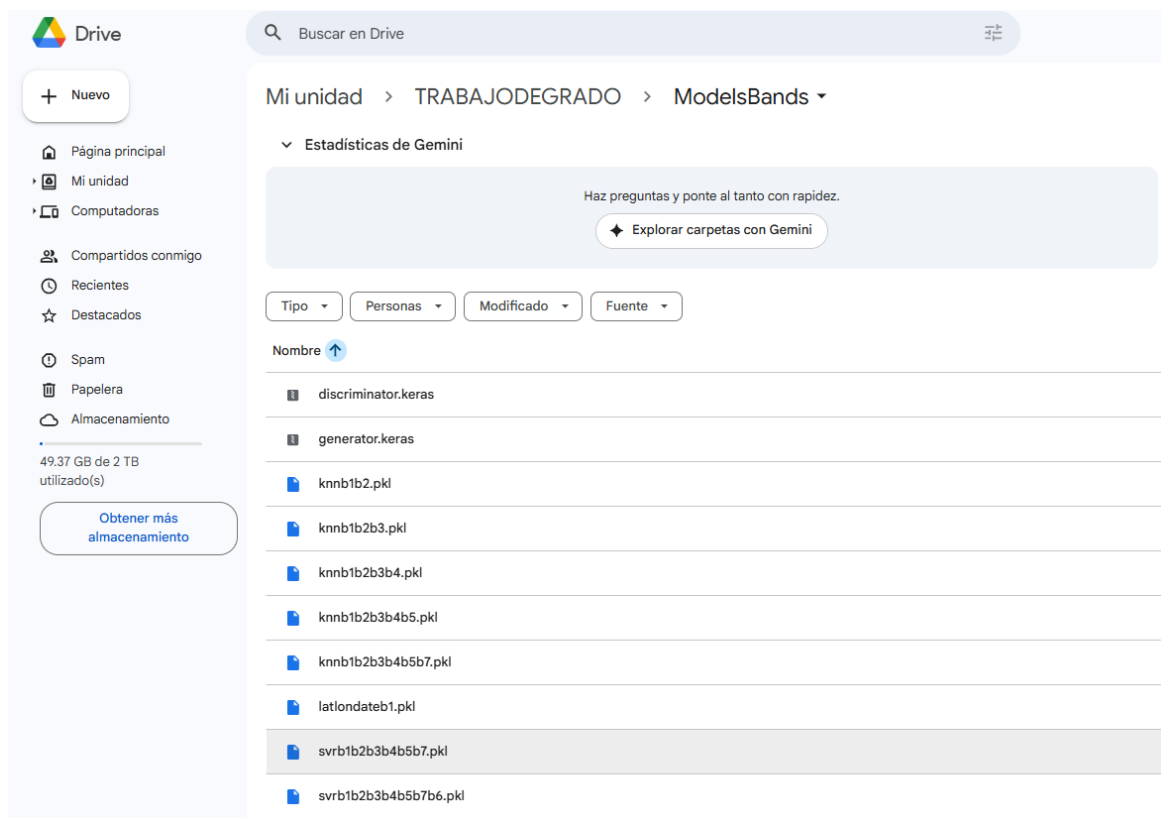


Fig. 22 Modelos entrenado almacenados en formato .pkl

Construcción del dataset NDWI

Una vez completadas las etapas de validación y modelado de bandas espectrales, se obtuvo la construcción automatizada del dataset del NDWI para el departamento de Nariño. Inicialmente se cargaron las bibliotecas necesarias y los archivos base del proyecto, los cuales constituyen los

insumos fundamentales para la generación del dataset. En la **TABLA XIV** se presentan los principales archivos utilizados durante esta etapa.

TABLA XIV. ARCHIVOS UTILIZADOS PARA LA GENERACIÓN DEL DATASET NDWI

Archivo	Descripción
Shapefile de Nariño	Archivo geoespacial que delimita el área de estudio
Datasets base	Conjunto inicial de datos satelitales procesados
Grilla espacial	Conjunto de coordenadas geográficas utilizadas para el análisis
Modelos entrenados (.pkl)	Modelos de aprendizaje automático utilizados para estimar las bandas espectrales

A partir del archivo **landsat_model.csv** se generó una grilla base de coordenadas espaciales, sobre la cual se aplicó una transformación temporal semanal para el periodo comprendido entre **2003** y **2014**. Posteriormente se generaron variables temporales derivadas que permiten representar la evolución de los datos dentro de la serie temporal. Estas variables se presentan en la **TABLA XV**.

TABLA XV. VARIABLES TEMPORALES GENERADAS

Variable	Descripción
year	Año correspondiente a la fecha de análisis
month	Mes correspondiente a la fecha de análisis
day	Día correspondiente a la fecha de análisis
month_sin	Transformación seno del mes
month_cos	Transformación coseno del mes
day_sin	Transformación seno del día
day_cos	Transformación coseno del día

Para representar adecuadamente la periodicidad temporal de los datos, las variables temporales fueron transformadas mediante codificación cíclica utilizando funciones seno y coseno. Esta transformación permite evitar discontinuidades entre valores consecutivos, como ocurre entre los meses de diciembre y enero o entre los días finales e iniciales del mes. Como resultado se generó un conjunto de variables temporales transformadas que permiten representar la periodicidad dentro de la serie temporal Fig. 23.

	date	month	year	day	month_sin	month_cos	day_sin	day_cos	year_norm
0	2003-01-05	1	2003	5	5.000000e-01	0.866025	0.848644	0.528964	0.0
1	2003-01-12	1	2003	12	5.000000e-01	0.866025	0.651372	-0.758758	0.0
2	2003-01-19	1	2003	19	5.000000e-01	0.866025	-0.651372	-0.758758	0.0
3	2003-01-26	1	2003	26	5.000000e-01	0.866025	-0.848644	0.528964	0.0
4	2003-02-02	2	2003	2	8.660254e-01	0.500000	0.394356	0.918958	0.0
...	...	...	...	...	...	...	...	...	...
621	2014-11-30	11	2014	30	-5.000000e-01	0.866025	-0.201299	0.979530	1.0
622	2014-12-07	12	2014	7	-2.449294e-16	1.000000	0.988468	0.151428	1.0
623	2014-12-14	12	2014	14	-2.449294e-16	1.000000	0.299363	-0.954139	1.0
624	2014-12-21	12	2014	21	-2.449294e-16	1.000000	-0.897805	-0.440394	1.0
625	2014-12-28	12	2014	28	-2.449294e-16	1.000000	-0.571268	0.820763	1.0
626 rows x 9 columns									

Fig. 23 Variables temporales con codificación cíclica.

Cálculo del NDWI e interpolación espacial

Una vez obtenidas las bandas espectrales predichas, se obtuvo el cálculo del NDWI a partir de la relación entre la banda verde y la banda del infrarrojo cercano. Ya que nos permitió identificar la presencia de agua en la superficie terrestre como Ecuación 1. Y en la **TABLA XVI** se observa los valores utilizados.

TABLA XVI. VARIABLES Y BANDAS UTILIZADAS EN EL CÁLCULO DEL NDWI

Elemento	Descripción
Green	Banda verde utilizada en el cálculo del NDWI
NIR	Banda del infrarrojo cercano utilizada en el cálculo del NDWI
Landsat	Green = Band 3, NIR = Band 5
MODIS	Green = Band 4, NIR = Band 2

Posteriormente al aplicar interpolación geoestadística mediante Kriging Ordinario, utilizando un modelo de variograma esférico. Como resultado, se obtuvo estimar valores en ubicaciones donde no existían observaciones directas, generando una superficie continua del índice NDWI sobre el departamento de Nariño Fig. 24.

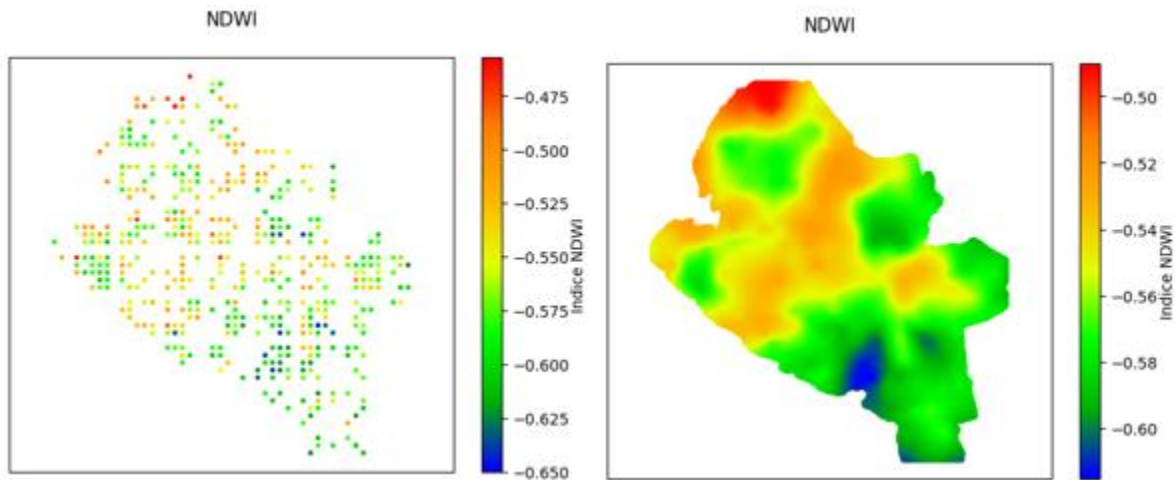


Fig. 24 Interpolación geoestadística con Kriging

Para cada fecha procesada se generó un archivo independiente en formato CSV, el cual contiene la información espacial y espectral necesaria para el análisis del NDWI. Cada archivo corresponde a un instante temporal específico dentro de la serie de estudio y fue almacenado automáticamente siguiendo una nomenclatura basada en la fecha de generación. En la **TABLA XVII** se presentan los elementos que componen cada archivo generado.

TABLA XVII. CONTENIDO DE LOS ARCHIVOS GENERADOS

Elemento	Descripción
Coordenadas espaciales	Latitud y longitud correspondientes a cada punto geográfico
Bandas espectrales	Valores estimados de las siete bandas espectrales
Índice NDWI	Valor calculado del índice de agua de diferencia normalizada

En total, el script procesó y permitió la generación de 626 instantes temporales semanales en el periodo comprendido entre 2003 y 2014, conformando un conjunto de series temporales continuas que representan la evolución del NDWI en el territorio analizado. Los resultados obtenidos evidencian que el proceso desarrollado permitió transformar datos satelitales crudos en un dataset estructurado, consistente y apto para el modelado de series temporales del NDWI, cumpliendo así con el objetivo planteado, como lo muestra en la **TABLA XVIII**.

TABLA XVIII. ESTRUCTURA DEL DATASET NDWI GENERADO

Variable	Descripción
lat	Latitud
lon	Longitud
band1 – band7	Bandas espectrales estimadas
value	Índice NDWI

En la Fig. 25 se representa los archivos CSV generados automáticamente durante el proceso de construcción del dataset NDWI

Nombre	Propietario	Fecha de modificación	Tamaño del archivo	Ordenar
dataset2003-01-05.csv	yo	18 oct 2025 yo	75 KB	
dataset2003-01-12.csv	yo	18 oct 2025 yo	75 KB	
dataset2003-01-19.csv	yo	18 oct 2025 yo	75 KB	
dataset2003-01-26.csv	yo	18 oct 2025 yo	75 KB	
dataset2003-02-02.csv	yo	18 oct 2025 yo	75 KB	
dataset2003-02-09.csv	yo	18 oct 2025 yo	75 KB	

Fig. 25 Archivos CSV generados para cada fecha del periodo de estudio

Resultado del objetivo 2

Componer scripts para la sintonización de algoritmos de series temporales para predecir el NDWI, identificando el modelo con mayor compromiso en sus métricas de calidad (R^2 , MAE, MSE) y tiempo de ejecución.

Para el cumplimiento del objetivo 2, se realizaron las siguientes fases según la metodología de desarrollo.

Fase 4. Minería de datos: construcción de ventanas temporales (3, 7 y 15 días) y la automatización de scripts para modelos como ConvLSTM2D, GRU y MLP.

Fase 5. Interpretación y evaluación: uso de métricas R^2 , MAE y MSE, así como el registro de tiempos de ejecución en entornos con GPU NVIDIA Tesla T4

Con el propósito de predecir el comportamiento del Índice de Agua de Diferencia Normalizada (NDWI) a partir de series temporales, se desarrollaron y ejecutaron scripts en Python orientados al entrenamiento, sintonización y comparación de diferentes algoritmos de forecasting.

Construcción de ventanas temporales

Como resultado de este procedimiento permitió implementar un esquema de sintonización basado principalmente en el tamaño de la ventana temporal y en el número de épocas de entrenamiento, generando diferentes configuraciones experimentales para los modelos evaluados como se muestra en la **TABLA XIX**.

TABLA XIX. CONFIGURACION DE LOS MODELOS EVALUADOS

Modelo	Ventana temporal (días)	Épocas	Descripción
ConvLSTM2D	3	1000	Modelo espacio-temporal con ventana corta
ConvLSTM2D	7	1000	Modelo espacio-temporal con ventana media
ConvLSTM2D	15	1000	Modelo espacio-temporal con ventana larga
GRU	7	100	Modelo recurrente enfocado en relaciones temporales
MLP	7	100	Modelo denso tradicional sin estructura temporal explícita

Distribución de conjunto de datos

Posterior a la aplicación de la técnica de ventanas temporales, se obtuvo un total de 619 muestras. Estas fueron divididas en conjuntos de entrenamiento, validación y prueba, con el fin de garantizar una adecuada evaluación del modelo y evitar problemas de sobreajuste se evidencia en **TABLA XX**.

TABLA XX. DATOS DE ENTRENAMIENTO, PRUEBA Y VALIDACIÓN

CONJUNTO DE DATOS	NÚMERO DE MUESTRAS	PORCENTAJE	DESCRIPCIÓN
Total de muestras	619	100%	Total de registros generados después de aplicar la función de ventanas temporales
Conjunto de entrenamiento (X_train, y_train)	495	80%	Datos utilizados para el entrenamiento del modelo mediante la función model.fit

Conjunto de validación (X_val, y_val)	99	20% del entrenamiento	Subconjunto utilizado durante el entrenamiento para validar el desempeño del modelo
Conjunto de prueba final (X_test, y_test)	124	20%	Datos utilizados para la evaluación final del modelo mediante model.evaluate y cálculo de R ²

Resultados obtenidos de los experimentos

Con el fin de valorar el desempeño de cada configuración, se calcularon las métricas coeficiente de determinación R², error cuadrático medio MSE y error absoluto medio MAE, junto con el tiempo de ejecución requerido en cada experimento. Los resultados obtenidos se presentan en la **TABLA XXI**.

TABLA XXI. RESULTADOS DE LA SINTONIZACIÓN DE MODELOS DE FORECASTING PARA NDWI

Modelo	Configuración	R ²	MSE	MAE	Tiempo de ejecución
ConvLSTM2D	Ventana 3 días, 1000 épocas	0.9885	0.001084	0.02337	4 horas
ConvLSTM2D	Ventana 7 días, 1000 épocas	0.9925	0.001462	0.02480	20 minutos
ConvLSTM2D	Ventana 15 días, 1000 épocas	0.9802	0.001820	0.02650	2 horas
MLP	Ventana 7 días, 100 épocas	-0.0133	0.000247	0.01203	3 horas
GRU	Ventana 7 días, 100 épocas	0.9464	0.005049	0.06790	1 hora 30 minutos

En el modelo ConvLSTM2D se puede evidenciar que se obtuvo el mayor compromiso entre las métricas evaluadas, alcanzando el valor más alto de R² **0.9925** en la configuración de ventana de 7 días, lo que indica una alta capacidad para representar la variabilidad del NDWI. Igualmente, mantiene valores bajos de MSE **0.001462** y MAE **0.02480**, junto con un tiempo de ejecución significativamente menor **20 minutos** en comparación con otras configuraciones del mismo modelo.

Por otro lado, el modelo MLP presentó un coeficiente de determinación negativo (R² = -0.0133), lo cual indica que no logró capturar la tendencia de los datos y su desempeño es inferior incluso a

un modelo basado en la media. Este resultado evidencia que los modelos densos tradicionales no son adecuados para este tipo de problema, ya que no incorporan de manera explícita la dependencia temporal ni la estructura espacial de los datos. En contraste, modelos como ConvLSTM2D integran operaciones convolucionales y memoria temporal, lo que les permite capturar patrones espacio-temporales complejos presentes en las series del NDWI, justificando así su superioridad en términos de precisión y capacidad de generalización.

Con el fin de profundizar en el análisis del modelo seleccionado, se presentan a continuación representaciones gráficas de su comportamiento durante el entrenamiento y la predicción.

Curva de entrenamiento

La Fig. 26 representa la curva del ConvLSTM2D durante el proceso de entrenamiento. En esta se observa la evolución del error tanto en el conjunto de entrenamiento como en el conjunto de validación a lo largo de las épocas.

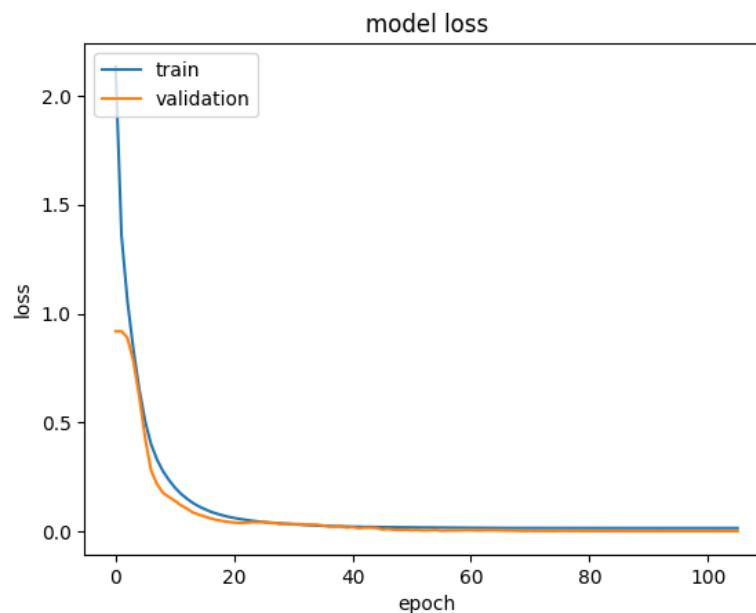


Fig. 26 Curva de pérdida del modelo ConvLSTM2D durante el entrenamiento

La pérdida de entrenamiento, representada por la línea azul, muestra una disminución progresiva a lo largo de las épocas, lo que indica que el modelo logra aprender adecuadamente a partir de los datos de entrenamiento. Por su parte, la pérdida de validación, representada por la línea naranja, sigue una tendencia similar, lo que evidencia una adecuada capacidad de generalización del modelo sobre datos no vistos. Igualmente, no se observa una separación significativa entre ambas curvas,

lo que indica la ausencia de sobreajuste (overfitting). Esto confirma que el modelo seleccionado presenta un comportamiento estable, un aprendizaje eficiente y una adecuada capacidad de predicción.

Comparación entre valores reales y predichos del NDWI

Con el fin de validar visualmente el desempeño del modelo seleccionado, se presenta en la Fig. 27 una comparación entre los valores reales y predichos del NDWI.

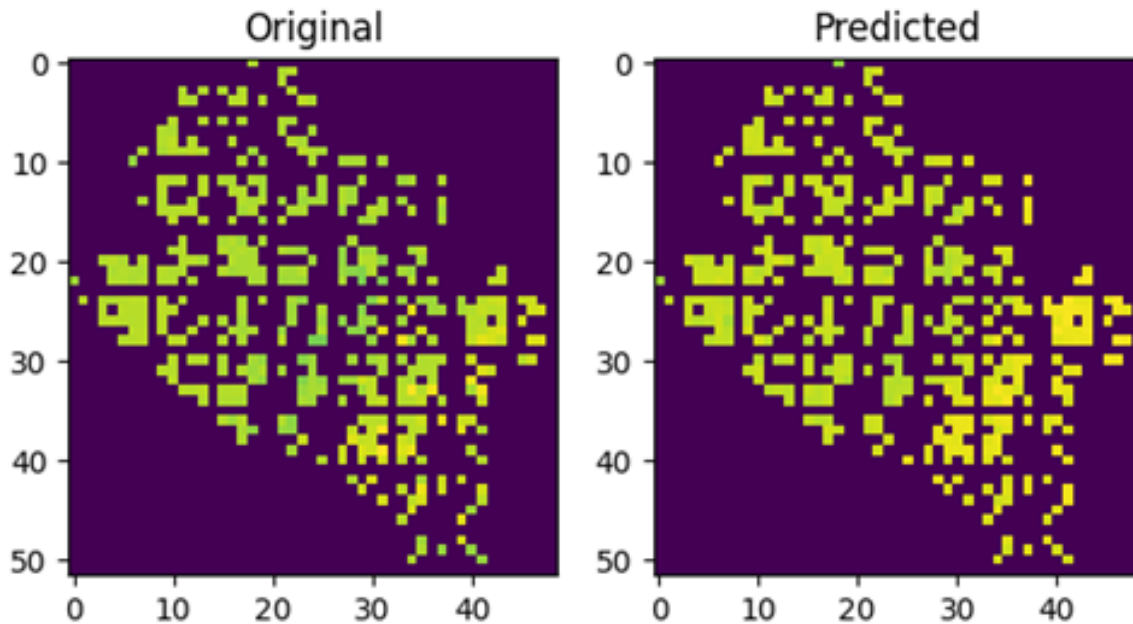


Fig. 27 Comparación espacial del NDWI real y predicho mediante el modelo ConvLSTM2D

Como se observa en la figura, ambos mapas presentan una distribución espacial similar, donde los valores más altos del NDWI corresponden a zonas con mayor presencia de agua se mantienen consistentes entre los valores reales y los predichos, lo que evidencia que el modelo logra capturar adecuadamente los patrones del índice en el territorio analizado.

Generación del mapa del NDWI

A partir de las predicciones obtenidas por el modelo, se generaron representaciones espaciales del Índice de Agua de Diferencia Normalizada (NDWI), las cuales fueron visualizadas mediante puntos geográficos codificados por color como lo muestra la Fig. 28.

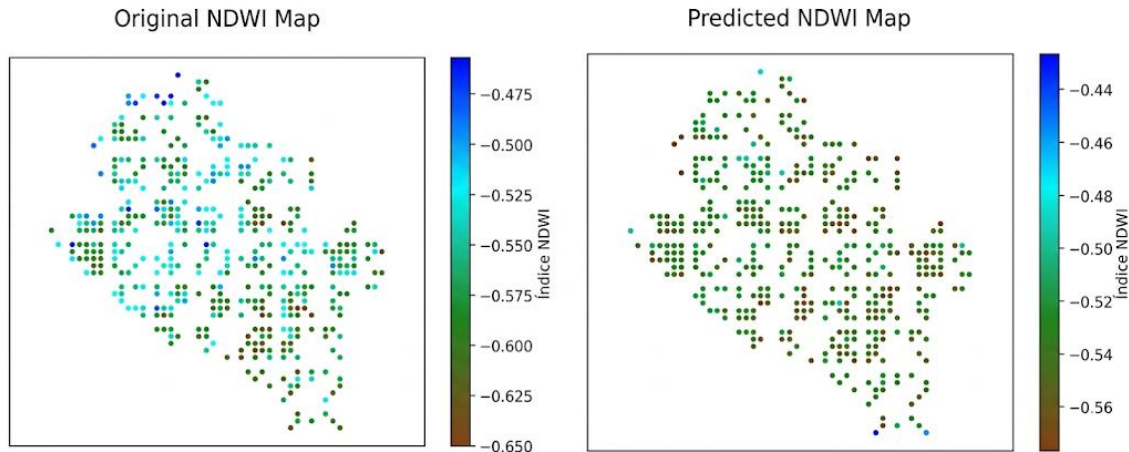


Fig. 28 Mapa de NDWI real y predicho

En esta representación, cada punto corresponde a una ubicación geográfica específica, donde la intensidad del color indica el valor del NDWI, permitiendo identificar zonas con mayor o menor presencia de agua en el área de estudio.

Interpolación espacial mediante Kriging

Para obtener una representación continua del NDWI, se aplicó un proceso de interpolación geoestadística utilizando el método de Kriging Fig. 29.

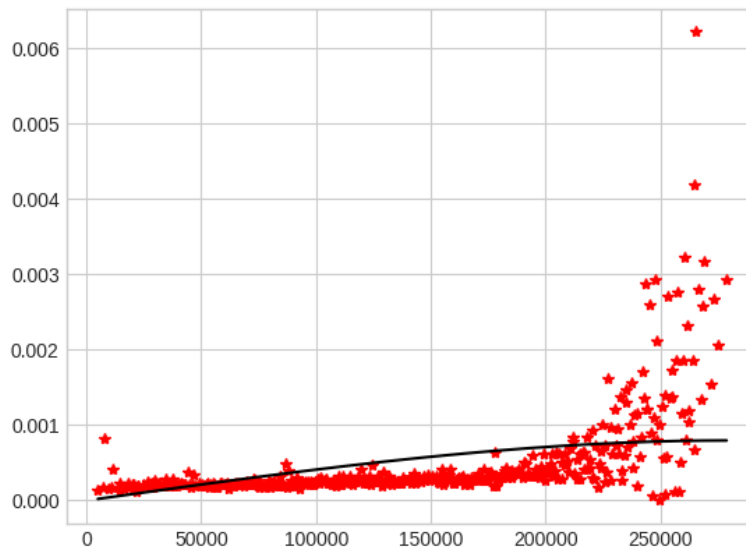


Fig. 29 Comportamiento de la interpolación del NDWI mediante Kriging

Este procedimiento permitió estimar valores en ubicaciones donde no se disponía de datos directos, generando una superficie continua a partir de los puntos discretos. La interpolación facilita la visualización del comportamiento espacial del NDWI y mejora la interpretación de los resultados.

Mapa final del NDWI

Como resultado del proceso de interpolación, se obtuvo el mapa final del NDWI para el departamento de Nariño, el cual representa de manera continua la distribución espacial de la presencia de agua en el territorio.

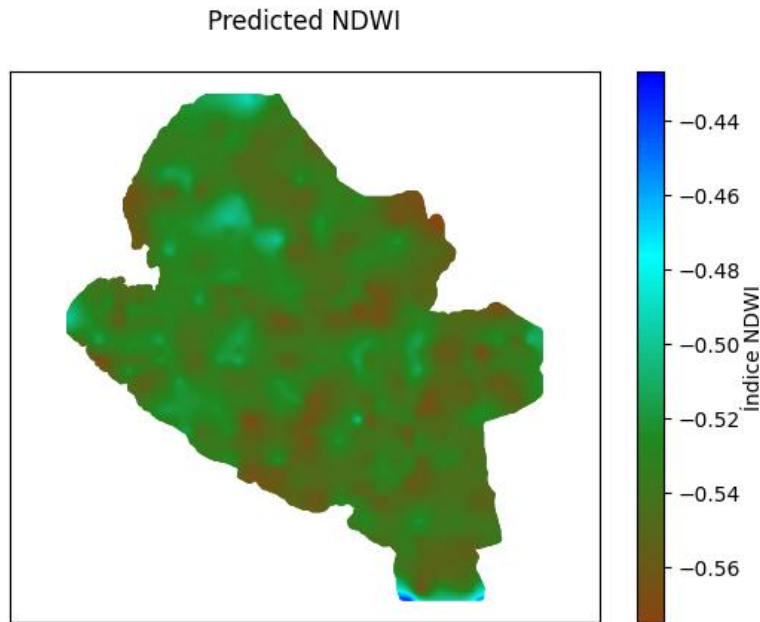


Fig. 30 Mapa continuo del NDWI para el departamento de Nariño

En Fig. 30 este mapa se observa una distribución más uniforme del índice, donde las variaciones espaciales del NDWI permiten identificar zonas con mayor concentración de agua. Esta representación facilita el análisis territorial y la interpretación de los patrones espaciales del fenómeno estudiado.

los resultados obtenidos demuestran la efectividad del proceso de sintonización y selección de modelos, permitiendo identificar una solución óptima para la predicción del NDWI. La combinación de análisis numérico y validación visual confirma la capacidad del modelo para capturar tanto la dinámica temporal como la distribución espacial del índice en el área de estudio.

Resultado del objetivo 3

Implementar una aplicación web que permita visualizar las predicciones del mejor modelo sobre un visor de mapas.

Para este objetivo la fase que se realizó fue:

Fase 6. Despliegue del modelo: uso de Flask como backend, la carga de modelos .keras, el mecanismo de predicción recursiva y el almacenamiento de resultados en formato GeoTIFF e imágenes PNG.

Como resultado de la implementación del sistema, se obtuvo una aplicación web funcional que permite visualizar las predicciones del índice NDWI generadas por el mejor modelo seleccionado, correspondiente a una arquitectura ConvLSTM2D, previamente entrenado, el cual es ejecutado en el backend para generar predicciones cuando el usuario solicita una fecha específica. Aunque el modelo no es visible directamente en la interfaz, su funcionamiento se evidencia en los resultados generados, los cuales corresponden a las salidas producidas por el modelo.

Generación de predicciones

El sistema permite generar predicciones del índice NDWI para fechas posteriores a la última fecha real disponible (14 de diciembre de 2014), garantizando la coherencia temporal del proceso. Cuando el usuario ingresa una fecha válida, el sistema verifica si ya existe una predicción almacenada; en caso contrario, ejecuta el modelo para generar el resultado correspondiente.

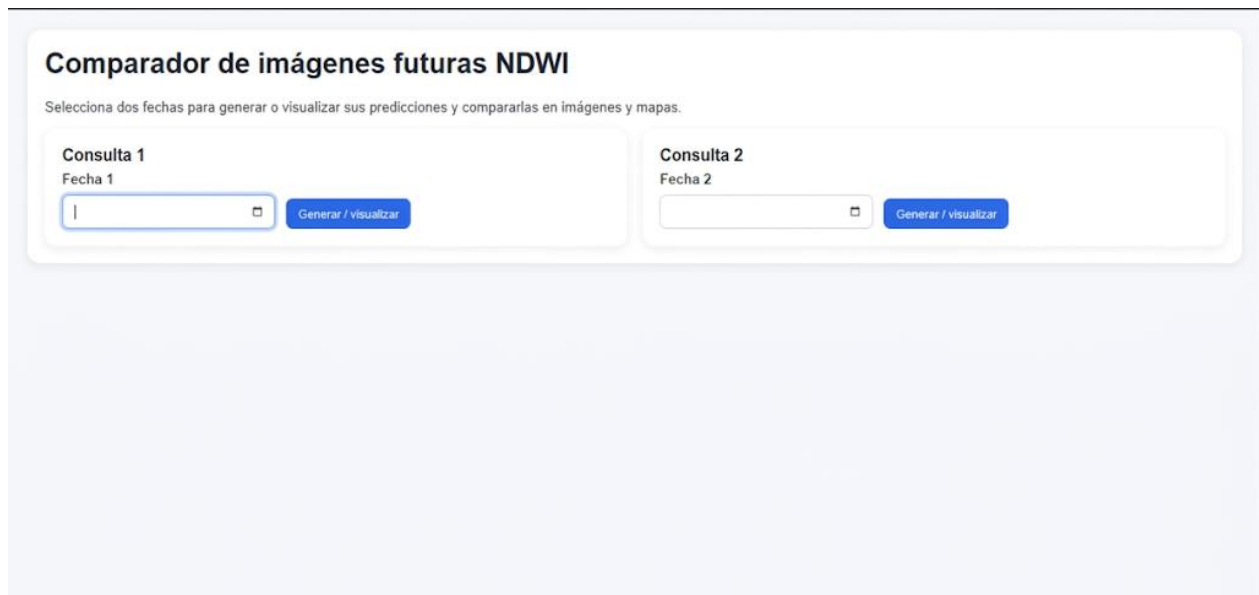


Fig. 31 Interfaz de la aplicación web para la consulta de predicciones del NDWI.

En la Fig. 31 se observa la interfaz del sistema, donde el usuario puede seleccionar la fecha de consulta y generar la predicción correspondiente.

Visualización geoespacial

Las predicciones generadas son representadas en formato imagen y superpuestas sobre un mapa base mediante el uso de herramientas de visualización geoespacial, lo que permite interpretar la distribución espacial del índice NDWI.

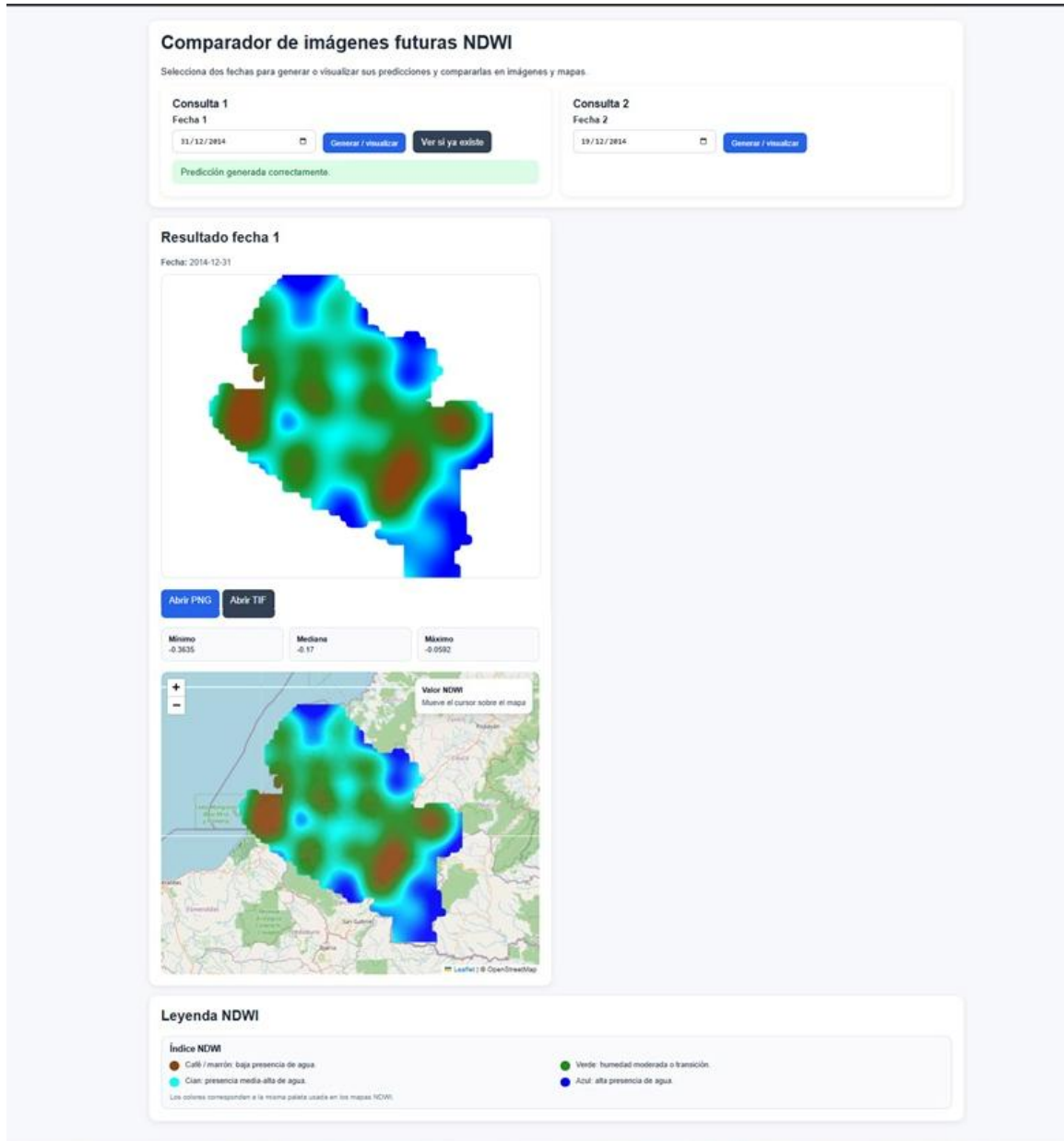


Fig. 32 Visualización de la predicción del NDWI sobre el mapa base

Cómo se muestra en la Fig. 32, la superposición de la imagen generada sobre el mapa permite identificar espacialmente las zonas con diferentes niveles de presencia de agua.

Evidencia del uso del modelo

Se evidenció que el sistema genera nuevos resultados cuando la fecha solicitada no ha sido previamente procesada, lo que confirma la ejecución del modelo predictivo. Asimismo, cuando la predicción ya existe, el sistema reutiliza los resultados almacenados, optimizando el rendimiento.

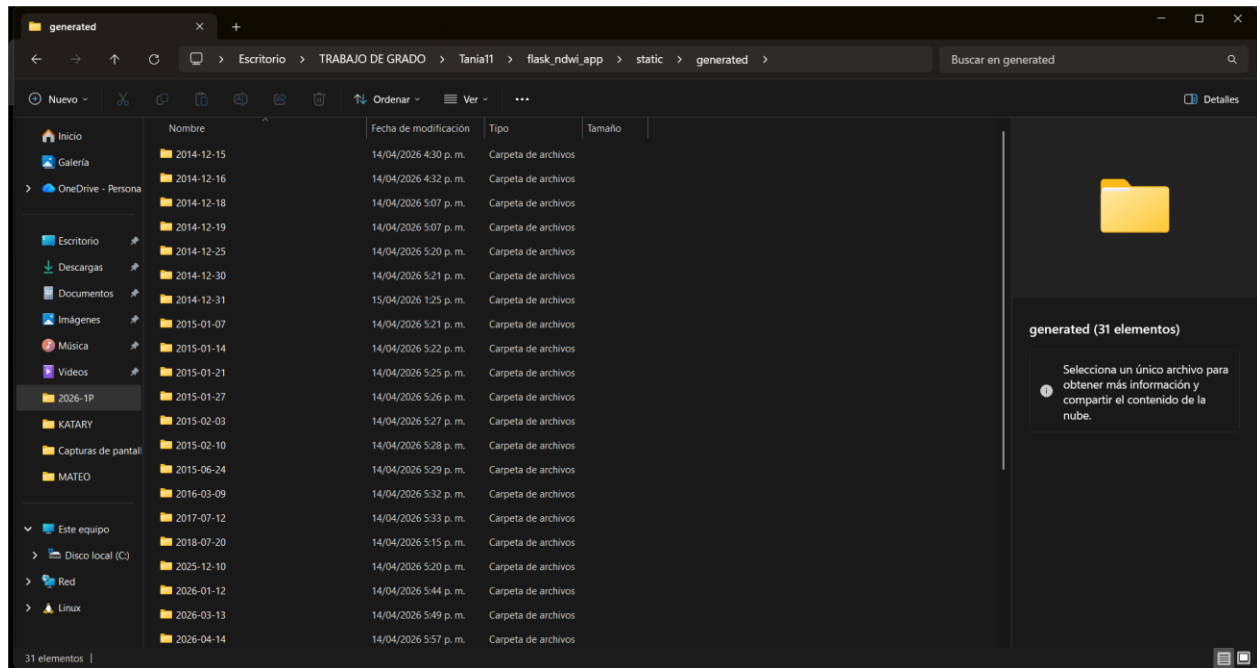


Fig. 33 Estructura de almacenamiento de resultados generados por fecha

En la Fig. 33 se presenta la organización de los resultados, donde cada predicción es almacenada en carpetas estructuradas por fecha, permitiendo su posterior consulta.

Comparación de predicciones

El sistema permite visualización simultanea de dos fechas diferentes, facilitando el análisis de cambios temporales en el NDWI.

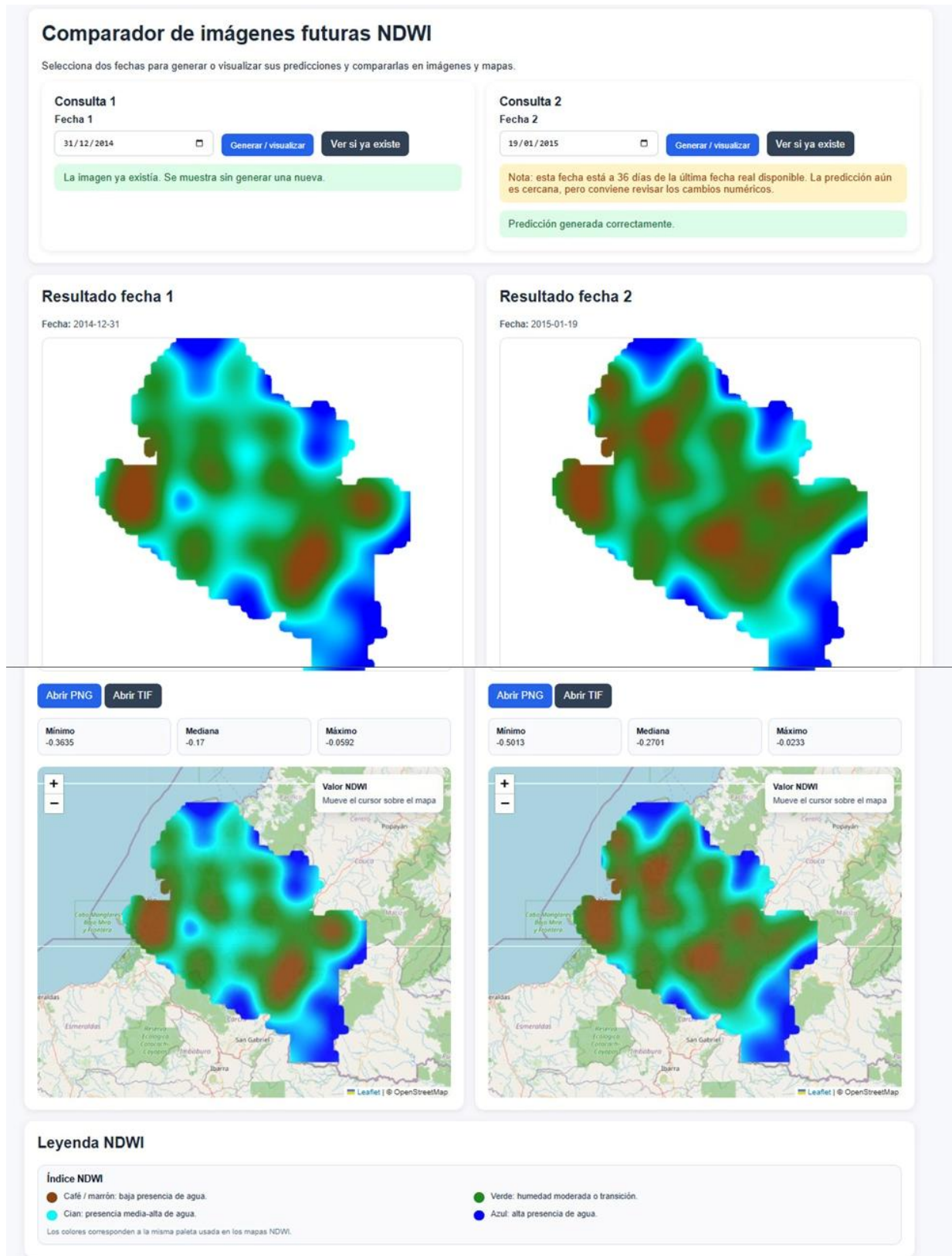


Fig. 34 Comparación de predicciones del NDWI en dos fechas diferentes

Como se observa en la Fig. 34 la comparación a lado permite identificar variaciones en el comportamiento del NDWI a lo largo del tiempo.

Estadística de los resultados

Por cada predicción generada, el sistema presenta valores estadísticos como mínimo, mediana y máximo, los cuales permiten una interpretación cuantitativa del comportamiento del índice NDWI en el área de estudio.

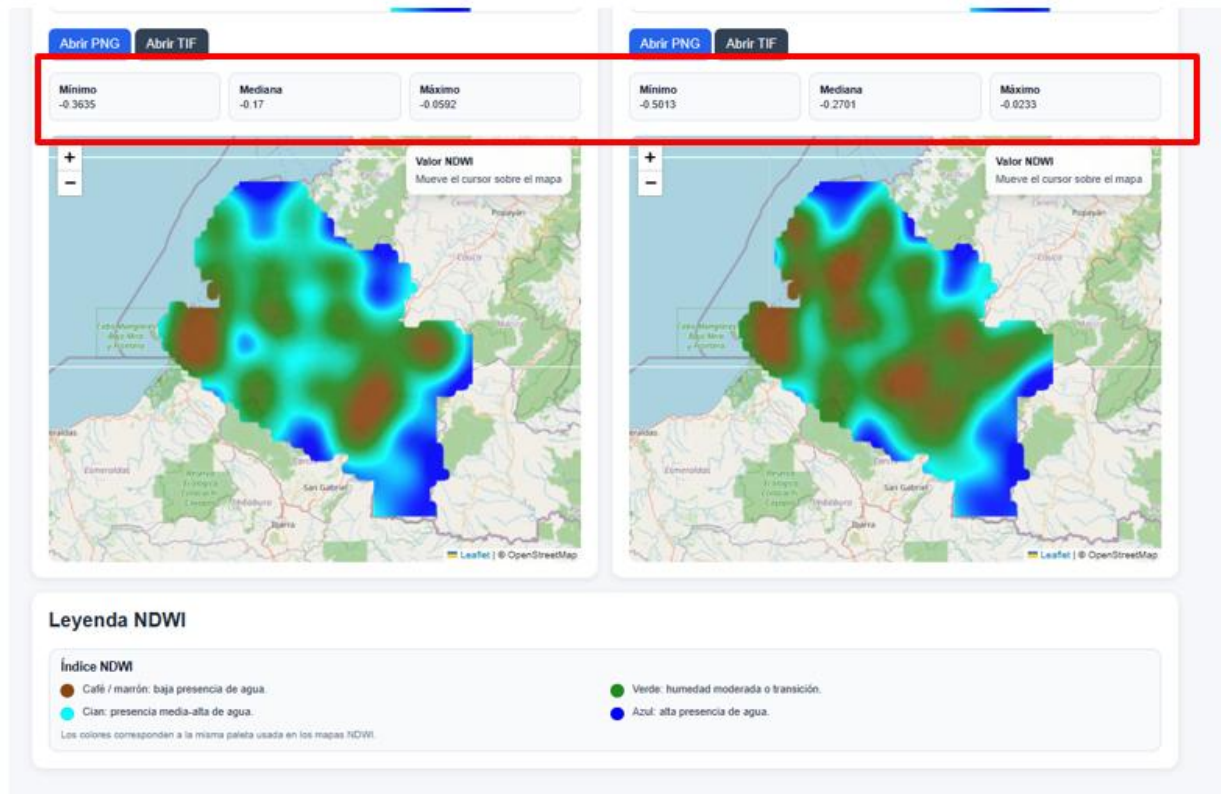


Fig. 35 Visualización de estadísticas asociadas a la predicción del NDWI

En la Fig. 35 muestra los valores facilitan la identificación de zonas con mayor o menor presencia de agua dentro del territorio analizado.

Validación temporal

Se evidenció que el sistema restringe las predicciones a fechas posteriores a la última fecha real disponible de la ventana, asegurando la coherencia con los datos utilizados en el entrenamiento del modelo y evitando resultados inconsistentes. Como se muestra en la Fig. 36.

Comparador de imágenes futuras

Selecciona dos fechas para generar o visualizar sus predicciones y comp

Consulta 1

Fecha 1

31/12/2014

Generar / visualizar

diciembre de 2014

2014

ene. feb. mar. abr.

may. jun. jul. ago.

sept. oct. nov. **dic.**

2015

2016

2017

2018

estra sin generar una nueva.

Fig. 36 Restricción de fechas

Los resultados obtenidos demuestran que el modelo predictivo fue correctamente integrado en la aplicación web, permitiendo la generación de nuevas predicciones y su visualización en un entorno geoespacial interactivo, cumpliendo así con el objetivo planteado.

V. ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

Los resultados obtenidos en el proceso de experimentación evidencian diferencias significativas en el desempeño de los modelos evaluados para la predicción del Índice de Agua de Diferencia Normalizada (NDWI), particularmente en su capacidad para modelar la dinámica espacio temporal de los datos satelitales.

En primer lugar, el modelo ConvLSTM2D con una ventana temporal de 7 días y 1000 épocas presentó el mejor desempeño general, alcanzando un coeficiente de determinación de $R^2 = 0.9925$, junto con valores bajos de error ($MSE = 0.001462$ y $MAE = 0.02480$). Adicionalmente, se evidenció un tiempo de ejecución reducido en comparación con otras configuraciones del mismo modelo, lo que demuestra un adecuado equilibrio entre precisión y eficiencia computacional. Estos resultados indican que el modelo logra capturar de manera efectiva tanto la variabilidad temporal como la estructura espacial del NDWI.

Por otra parte, la configuración de ConvLSTM2D con ventana de 3 días obtuvo un desempeño ligeramente inferior ($R^2 = 0.9885$), lo que sugiere que una ventana temporal más corta limita la capacidad del modelo para capturar dependencias más amplias en el tiempo. De manera similar, la configuración con ventana de 15 días presentó una disminución en el rendimiento ($R^2 = 0.9802$), lo cual puede atribuirse a la inclusión de información temporal excesiva que introduce ruido en el proceso de aprendizaje.

Por el contrario, el modelo MLP presentó un desempeño deficiente, evidenciado por un valor de $R^2 = -0.0133$, lo que indica que el modelo no logra representar la variabilidad del NDWI. Aunque presenta valores bajos de MSE y MAE, estos no reflejan una buena capacidad predictiva, debido a que el modelo carece de mecanismos para capturar dependencias temporales y espaciales. Por otro lado, el modelo GRU mostró un comportamiento intermedio, alcanzando un $R^2 = 0.9464$, lo que evidencia que es capaz de modelar relaciones temporales en los datos. Sin embargo, su desempeño es inferior al de ConvLSTM2D, lo cual puede explicarse por la ausencia de un componente convolucional que permita capturar patrones espaciales en la información geográfica. Estos resultados son coherentes con lo planteado en la literatura. Según Cabrera et al. (2016), el procesamiento de imágenes satelitales provenientes de sensores como Landsat y MODIS permite identificar patrones espaciales y temporales relevantes, lo cual requiere modelos capaces de integrar ambas dimensiones en el proceso de análisis [11]. En este sentido, los resultados obtenidos

confirman que los modelos que consideran la estructura espacio temporal presentan un mejor desempeño en este tipo de problemas.

Asimismo, Mora Paz (2019) señala que los modelos de aprendizaje automático permiten identificar patrones complejos en datos donde no existen relaciones lineales evidentes, destacando la importancia de seleccionar arquitecturas adecuadas para el tipo de información analizada [12]. En relación con este planteamiento, en el presente estudio se evidencia que no todas las arquitecturas neuronales presentan el mismo rendimiento, siendo el modelo ConvLSTM2D el más adecuado para la predicción del NDWI.

Del mismo modo, Pachajoa et al. (2021) resaltan que el análisis de datos satelitales mediante técnicas de aprendizaje automático permite capturar relaciones no lineales entre variables espectrales [13]. En este contexto, el desempeño superior del modelo ConvLSTM2D puede explicarse por su capacidad para integrar información espacial y temporal en un mismo esquema de aprendizaje.

Desde una perspectiva cualitativa, las predicciones generadas por el modelo ConvLSTM2D presentan coherencia espacial con respecto a los patrones observados en los datos históricos, lo cual valida el modelo no solo a nivel cuantitativo, sino también en términos de su aplicabilidad en entornos geospaciales.

Sin embargo, se identifican algunas limitaciones asociadas al proceso, como lo son la presencia de nubes en las imágenes satelitales, la pérdida de información en ciertos periodos y la dependencia de la calidad de los datos de entrada, factores que pueden influir en la precisión de las predicciones en escenarios específicos.

Para concluir, los resultados obtenidos permiten afirmar que los modelos de aprendizaje profundo con capacidad de modelado espacio temporal, como ConvLSTM2D, presentan un mejor desempeño para la predicción del NDWI en comparación con modelos tradicionales y recurrentes. Esto resalta la importancia de seleccionar arquitecturas acordes a la naturaleza de los datos en problemas de análisis geoespacial y series temporales.

CONCLUSIONES

El desarrollo de esta investigación permitió demostrar que es posible construir un sistema integral para la predicción y visualización del Índice de Agua de Diferencia Normalizada (NDWI), a partir del uso de datos satelitales y técnicas de aprendizaje automático, aportando una solución aplicable al análisis de recursos hídricos en contextos regionales como el departamento de Nariño.

En relación con el primer objetivo, el proceso de preprocesamiento de datos provenientes de imágenes Landsat y MODIS permitió la generación de series temporales consistentes del NDWI. La estructuración de la información, la segmentación temporal y la limpieza de datos garantizaron la calidad del dataset, evidenciando que un adecuado tratamiento de los datos es fundamental para obtener resultados confiables en etapas posteriores de modelado.

Respecto al segundo objetivo, la implementación y comparación de diferentes algoritmos de predicción de series temporales permitió identificar el modelo con mejor desempeño, considerando métricas como R^2 , MAE y MSE, así como el tiempo de ejecución. Se concluye que los modelos basados en aprendizaje automático tienen la capacidad de capturar patrones espacio-temporales del NDWI, logrando predicciones coherentes con el comportamiento histórico del índice.

En cuanto al tercer objetivo, la implementación de la aplicación web mediante Flask permitió integrar el modelo seleccionado en un entorno funcional, facilitando la visualización de los resultados en mapas geoespaciales interactivos. Este desarrollo evidencia la importancia de llevar los modelos teóricos a soluciones prácticas, mejorando la accesibilidad y comprensión de la información por parte de los usuarios.

De manera general, los resultados obtenidos demuestran que el sistema desarrollado no solo cumple con los objetivos planteados, sino que también ofrece una herramienta eficiente para la interpretación de dinámicas hídricas a nivel espacial y temporal, permitiendo identificar patrones asociados a la presencia de agua.

Finalmente, se concluye que la integración de técnicas de procesamiento de datos, aprendizaje automático y desarrollo web constituye una estrategia efectiva para abordar problemas complejos del entorno, destacando el potencial de este tipo de soluciones para apoyar la toma de decisiones en el ámbito ambiental y territorial.

RECOMENDACIONES

El presente trabajo puede servir como base para futuras investigaciones orientadas al análisis de recursos hídricos mediante el uso de datos satelitales y técnicas de aprendizaje automático, por lo que se recomienda continuar explorando enfoques más avanzados que permitan mejorar la precisión y robustez de las predicciones del NDWI.

Se sugiere profundizar en la implementación de modelos de Deep Learning, como redes neuronales recurrentes avanzadas y arquitecturas convolucionales espacio-temporales, con el fin de capturar de manera más eficiente las dinámicas complejas del índice en diferentes escalas espaciales y temporales.

Es recomendable integrar mayor diversidad de fuentes de datos, incluyendo variables climáticas como precipitación, temperatura y humedad, así como información geográfica adicional, lo cual permitiría enriquecer el modelo y mejorar la calidad de las predicciones.

Se aconseja optimizar la aplicación web desarrollada, incorporando funcionalidades adicionales como filtros por regiones específicas, análisis comparativos entre fechas y mejoras en la visualización geoespacial, con el propósito de ofrecer una herramienta más completa y amigable para el usuario.

Otro aspecto a considerar es la implementación del sistema en entornos reales de toma de decisiones, especialmente en entidades relacionadas con la gestión ambiental y territorial, lo que permitiría validar su utilidad práctica y generar un mayor impacto en la región.

Asimismo, se recomienda fortalecer la creación y mantenimiento de bases de datos actualizadas y de alta calidad, dado que la disponibilidad de información confiable es un factor determinante para el desempeño de los modelos predictivos.

Se recomienda que la aplicación y uso del sistema desarrollado sea interpretado con cautela, reconociendo que, aunque las predicciones generadas son coherentes con el comportamiento histórico del índice, no constituyen por sí solas evidencias suficientes para afirmar una capacidad robusta de generalización en contextos geográficos o temporales diversos. Es imperativo que futuras investigaciones incorporen una discusión más crítica sobre el uso del NDWI, considerando sus limitaciones técnicas frente a factores ambientales como la densidad de la vegetación, la turbidez de los cuerpos de agua o la variabilidad atmosférica (nubosidad persistente), factores que pueden afectar significativamente la precisión de la detección. Por tanto, se sugiere evitar generalizaciones amplias para la toma de decisiones críticas sin antes haber realizado procesos

adicionales de validación en escenarios reales e integración con otras fuentes de información, reconociendo la necesidad de fortalecer el alcance y la aplicabilidad del sistema en distintos entornos hídricos regionales.

Finalmente, se sugiere promover la interdisciplinariedad entre áreas como la ingeniería de sistemas, ciencias ambientales y geografía, con el fin de desarrollar soluciones más integrales que contribuyan al monitoreo y gestión sostenible de los recursos hídricos.

REFERENCIAS BIBLIOGRÁFICAS

- [1] L. Lozano Arias, «Un nuevo modelo para la estimación de la biomasa viva y el carbono almacenado en los bosques de manglar usando sensoramiento remoto y aprendizaje de máquina: caso de estudio Tumaco-Nariño», Trabajo de grado - Maestría, Universidad Nacional de Colombia, 2023. Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://repositorio.unal.edu.co/handle/unal/86135>
- [2] S. K. McFEETERS, «The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features», *International Journal of Remote Sensing*, may 1996, doi: 10.1080/01431169608948714.
- [3] S. S. Ahmed *et al.*, «A light-weight model to generate NDWI from Sentinel-1», 23 de enero de 2025, *arXiv*: arXiv:2501.13357. doi: 10.48550/arXiv.2501.13357.
- [4] «(PDF) Improving estimates of water resources availability over North Tropical South America: comparison of two satellite precipitation merging schemes», *ResearchGate*, oct. 2024, doi: 10.15446/esrj.v28n1.104344.
- [5] «Tecnologías de la Información y las Comunicaciones (TIC)». Accedido: 14 de febrero de 2026. [En línea]. Disponible en: <https://mintic.gov.co/portal/inicio/Glosario/T/5755:Tecnologias-de-la-Informacion-y-las-Comunicaciones-TIC>
- [6] redaccion iAgua, «¿Cómo se aplica la inteligencia artificial en la gestión del agua?», iAgua. Accedido: 14 de febrero de 2026. [En línea]. Disponible en: <https://www.iagua.es/respuestas/como-se-aplica-inteligencia-artificial-gestion-agua>
- [7] W. I. D. A. Castrillón, «Tecnofilia», Universidad CESMAG. Accedido: 14 de febrero de 2026. [En línea]. Disponible en: <https://www.unicesmag.edu.co/grupos-investigacion/tecnofilia/>
- [8] D. M. H. Posada y E. Aristizábal, «Modelo de inteligencia artificial y aprendizaje automático para la predicción espacial y temporal de eventos de sequía en el departamento del Magdalena, Colombia.», *Inge CuC*, vol. 18, n.º 2, Art. n.º 2, nov. 2022, doi: 10.17981/ingecuc.18.2.2022.20.
- [9] A. C. A. Aguilar y F. F. O.- Díaz, «APRENDIZAJE AUTOMÁTICO PARA LA PREDICCIÓN DE CALIDAD DE AGUA POTABLE», *Ingeniare*, n.º 28, Art. n.º 28, may 2020, doi: 10.18041/1909-2458/ingeniare.28.6215.

- [10] D. A. Patiño Duque y D. M. Martínez Murillo, «Estudio multitemporal de la disminución de los índices ndvi, ndwi, vigorosidad e infrarrojo para el cultivo de palma de aceite en el municipio de Puerto López, departamento del Meta, en los años 2022 y 2023», jun. 2024, Accedido: 18 de abril de 2025. [En línea]. Disponible en: <http://repository.unad.edu.co/handle/10596/62588>
- [11] O. Cabrera, B. Champutiz, A. Calderón, y A. Pantoja, «Landsat and MODIS satellite image processing for solar irradiance estimation in the department of Narino-Colombia», en *2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA)*, ago. 2016, pp. 1-6. doi: 10.1109/STSIVA.2016.7743306.
- [12] H. A. Mora-Paz, «Comparativo de kernels sobre predicción de oferta de fuentes alternativas de energía», masterThesis, 2019. Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://reunir.unir.net/handle/123456789/10020>
- [13] D.-M. Pachajoa, H. Mora-Paz, y D. Mayorca-Torres, «Comparison of Kernel Functions in the Classification of Irradiance Zones from Multispectral Satellite Images», *Revista Facultad de Ingeniería*, vol. 30, n.º 58, Art. n.º 58, dic. 2021, doi: 10.19053/01211129.v30.n58.2021.13845.
- [14] S. Montoya Correa, «Análisis multitemporal mediante teledetección espacial y Sistemas de Información Geográfica (SIG) del cambio de cobertura del suelo en el municipio de Segovia, Antioquia, en los años 2003-2018», nov. 219d. C., Accedido: 18 de abril de 2025. [En línea]. Disponible en: <https://dspace.tdea.edu.co/handle/tdea/3704>
- [15] «Política Nacional para la Gestión Integral del Recurso Hídrico | Ministerio de Ambiente y Desarrollo Sostenible». Accedido: 18 de abril de 2025. [En línea]. Disponible en: <https://archivo.minambiente.gov.co/index.php/gestion-integral-del-recurso-hidrico/direccion-integral-de-recurso-hidrico/politica-nacional-para-la-gestion-integral-del-recurso-hidrico>
- [16] «ABRHidro - ACTAS - Análisis Cuantitativo del Agua en el Embalse Jacareí-Jaguari, SP: Aplicación de Teledetección e Índice NDWI». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://anais.abrhidro.org.br/job.php?Job=16076>
- [17] G. Meléndez Salazar, «Análisis multitemporal de la dinámica de coberturas en el manglar del estero de Puntarenas mediante la teledetección para el período 1974-2024», 2024, Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://hdl.handle.net/10669/100112>

- [18] D. Cunza Robles, «Estimación del retroceso y predicción del tiempo de permanencia del nevado Ausangate mediante el análisis multitemporal años 1985-2020, Cusco», *Universidad Continental*, 2024, Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://repositorio.continental.edu.pe/handle/20.500.12394/14573>
- [19] M. M. Delgado Pascual, «Series multitemporales de imágenes de satélite para el análisis de las superficies de regadío en Castilla y León», 2024, doi: 10.14201/gredos.160347.
- [20] J. B. Renteria-Mena y E. Giraldo, «Predictive Modeling of Water Level in the San Juan River Using Hybrid Neural Networks Integrated with Kalman Smoothing Methods», *Information*, vol. 15, n.º 12, Art. n.º 12, dic. 2024, doi: 10.3390/info15120754.
- [21] A. D. A. Montaña y M. A. R. Arango, «Aplicación de índices espectrales para determinar la transformación del bosque seco en Agua de Dios (Cundinamarca, Colombia)», *Avances Investigación en Ingeniería*, vol. 18, n.º 1 (Enero-Junio), Art. n.º 1 (Enero-Junio), abr. 2021, doi: 10.18041/1794-4953/avances.1.7070.
- [22] C. L. Barajas Garzon, «Análisis de la variabilidad espacio-temporal de humedales en la ciudad de Bogotá a través de índices espectrales, como herramienta de monitoreo y gestión del recurso hídrico», Trabajo de grado - Maestría, Universidad Nacional de Colombia, 2020. Accedido: 22 de abril de 2025. [En línea]. Disponible en: <https://repositorio.unal.edu.co/handle/unal/80102>
- [23] A. Romero Contreras, «Entrevista con Luciano Boi ¿Qué es la topología? Matemáticas, ciencia, filosofía y arte», *Tópicos del Seminario*, n.º 43, pp. 213-265, jun. 2020.
- [24] «AutoCAD Map 3D 2024 Ayuda | Acerca de las topologías de aguas residuales | Autodesk». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://help.autodesk.com/view/MAP/2024/ESP/?guid=GUID-5A440ED4-21FF-4867-BF1A-B3A4839DD406>
- [25] «Acerca de las topologías de agua». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <http://help.autodesk.com/cloudhelp/2015/ESP/MAP3D-EntIndModels/files/GUID-B4F4D9F6-3C5E-4899-9048-B07FA1BB97B8.htm#!>
- [26] «¿Cuál es la calidad perfecta del agua?», Fundación Aquae. Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.fundacionaquae.org/wiki/calidad-agua/>

- [27] P. blen Siska, V. Argentina, y M. Bayumi, «Uji Kualitas Air Keran Pada Laboratorium Terpadu Universitas UIN Raden Fatah Palembang», *Jurnal Pendidikan, Sains Dan Teknologi*, vol. 2, n.º 2, Art. n.º 2, oct. 2023, doi: 10.47233/jpst.v2i4.1224.
- [28] I. Agrónomo, «NDWI: Índice Diferencial de Agua Normalizado», Brioagro. Accedido: 30 de abril de 2025. [En línea]. Disponible en: <https://brioagro.es/ndwi-indice-diferencial-de-agua-normalizado/>
- [29] «(PDF) A Complementary Remote-Sensing Method to Find Persons Missing in Water: Two Case Studies», *ResearchGate*, dic. 2024, doi: 10.3390/forensicsci3020021.
- [30] I. R. Magro, «Topological materials from a symmetry perspective», <http://purl.org/dc/dcmitype/Text>, Universidad del País Vasco - Euskal Herriko Unibertsitatea, 2021. Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://dialnet.unirioja.es/servlet/tesis?codigo=306505>
- [31] «Water Management | Tierras y Aguas | Organización de las Naciones Unidas para la Alimentación y la Agricultura | Land & Water | Food and Agriculture Organization of the United Nations». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.fao.org/land-water/water/water-management/es/>
- [32] «Documento de apoyo: Medio ambiente.», Issuu. Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://issuu.com/karollduarte/docs/5-med-ambiente>
- [33] «¿Qué es la inteligencia artificial o IA?», Google Cloud. Accedido: 2 de abril de 2025. [En línea]. Disponible en: <https://cloud.google.com/learn/what-is-artificial-intelligence>
- [34] «¿Qué es el Deep Learning y para qué sirve? | indra». Accedido: 25 de abril de 2025. [En línea]. Disponible en: <https://www.indracompany.com/es/blogneo/deep-learning-sirve>
- [35] «Manual de Forecasting PDF | PDF | Logística | Marketing», Scribd. Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://es.scribd.com/document/286468715/8-Manual-de-Forecasting-pdf>
- [36] «IBM SPSS Statistics». Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://www.ibm.com/docs/es/spss-statistics/saas?topic=forecasting-introduction-time-series>
- [37] «¿Qué es una red neuronal? | IBM». Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://www.ibm.com/es-es/think/topics/neural-networks>

- [38] «Redes neuronales convolucionales son un tipo de redes neuronales», Juan Barrios. Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://www.juanbarrios.com/redes-neurales-convolucionales/>
- [39] «¿Qué es la visualización de datos? Definición, ejemplos y recursos.» Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://www.tableau.com/es-mx/learn/articles/data-visualization>
- [40] Mariana, «Sistemas de Información Geográfica: principales usos», Cleverpy. Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://cleverpy.com/usos-sistemas-informacion-geografica/>
- [41] «Qué es Flask y ventajas que ofrece | OpenWebinars», OpenWebinars.net. Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://openwebinars.net/blog/que-es-flask/>
- [42] C. Barrera, «Imágenes satelitales: qué son, para qué sirven y sus tipos», axessnet. Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://axessnet.com/las-imagenes-satelitales-y-sus-aplicaciones-en-la-vida-cotidiana/>
- [43] «Hiperespectral Y Multiespectral: Diferencias Entre Ambas». Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://eos.com/es/blog/imagenes-multiespectrales-y-hiperespectrales/>
- [44] N. C. Baquerizo y E. J. V. Ventocilla, «Evaluación de algoritmos de machine learning en la clasificación de imágenes satelitales multiespectrales, caso: Amazonia Peruana», *Ciencia Latina Revista Científica Multidisciplinar*, vol. 6, n.º 1, Art. n.º 1, mar. 2022, doi: 10.37811/cl_rcm.v6i1.1843.
- [45] «qué son los satélites artificiales», El Economista. Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://www.eleconomista.com.mx/tecnologia/que-son-los-satelites-artificiales-20200129-0066.html>
- [46] H. He, «Qué es la reflectancia», Rayzeek. Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://www.rayzeek.com/es/glosario/que-es-la-reflectancia>
- [47] M. I. González Núñez, «Empleo de parámetros espectrales derivados de imágenes Sentinel para el seguimiento constructivo del sector sur de la ciudad de Sevilla entre los años 2015 y 2019», 2020, Accedido: 28 de abril de 2025. [En línea]. Disponible en: <https://hdl.handle.net/11441/105043>

- [48] K. Bandara, C. Bergmeir, y S. Smyl, «Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach», *Expert Systems with Applications*, vol. 140, p. 112896, feb. 2020, doi: 10.1016/j.eswa.2019.112896.
- [49] Sadhvi, «Data Preprocessing in Data Mining: A Hands On Guide», Analytics Vidhya. Accedido: 30 de abril de 2025. [En línea]. Disponible en: <https://www.analyticsvidhya.com/blog/2021/08/data-preprocessing-in-data-mining-a-hands-on-guide/>
- [50] «Data Preprocessing: A Complete Guide with Python Examples». Accedido: 30 de abril de 2025. [En línea]. Disponible en: <https://www.datacamp.com/blog/data-preprocessing>
- [51] «Detailed Explanation of Satellite Image Processing Techniques», Chang Guang Satellite Technology Co., Ltd. Accedido: 30 de abril de 2025. [En línea]. Disponible en: <https://www.jl1global.com/detailed-explanation-of-satellite-image-processing-techniques.html>
- [52] gabri, «¿Qué es el error cuadrático medio RMSE?», El blog de franz. Accedido: 30 de abril de 2025. [En línea]. Disponible en: <https://acolita.com/que-es-el-error-cuadratico-medio-rmse/>
- [53] J. Frost, «Root Mean Square Error (RMSE)», Statistics By Jim. Accedido: 30 de abril de 2025. [En línea]. Disponible en: <https://statisticsbyjim.com/regression/root-mean-square-error-rmse/>
- [54] «¿Qué es el R2 (R cuadrado) o Coeficiente de Determinación? | Estrategias de Inversión». Accedido: 30 de abril de 2025. [En línea]. Disponible en: <https://www.estrategiasdeinversion.com/herramientas/diccionario/fondos/r2-r-cuadrado-o-coeficiente-de-determinacion-t-1163#que-es-el-r2-r-cuadrado-o-coeficiente-de-determinacion>
- [55] *Pattern Recognition and Machine Learning*. Accedido: 20 de febrero de 2026. [En línea]. Disponible en: <https://link.springer.com/book/9780387310732>
- [56] A. Woodie, «Prediction Time», BigDATAwire. Accedido: 1 de mayo de 2025. [En línea]. Disponible en: <https://www.bigdatawire.com/2020/12/17/prediction-time/>
- [57] *Forecasting: Principles and Practice (2nd ed)*. Accedido: 1 de mayo de 2025. [En línea]. Disponible en: <https://otexts.com/fpp2/>
- [58] K. Al Yaqoob, M. Bououdina, M. S. Akhter, B. Al Najjar, y J. J. Vijaya, «Selectividad y eliminación eficiente de iones Pb y Cd mediante nanopartículas magnéticas MFe₂O₄ (M=Co,

- Ni, Cu y Zn)», *Materials Chemistry and Physics*, vol. 232, pp. 254-264, jun. 2019, doi: 10.1016/j.matchemphys.2019.04.077.
- [59] «(PDF) NDWI?A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space», *ResearchGate*, doi: 10.1016/S0034-4257(96)00067-3.
- [60] H. Xu, «Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery», *International Journal of Remote Sensing*, vol. 27, n.º 14, pp. 3025-3033, jul. 2006, doi: 10.1080/01431160600589179.
- [61] «Mean squared error», *Wikipedia*. 5 de abril de 2025. Accedido: 1 de mayo de 2025. [En línea]. Disponible en:
https://en.wikipedia.org/w/index.php?title=Mean_squared_error&oldid=1284160594
- [62] *Forecasting: Principles and Practice (3rd ed)*. Accedido: 20 de febrero de 2026. [En línea]. Disponible en: <https://otexts.com/fpp3/>
- [63] «Colección Digital · Metodología de la investigación. · Biblioteca Digital». Accedido: 2 de mayo de 2025. [En línea]. Disponible en: https://bibliotecadigital.uce.edu.ec/s/L-D/item/970?utm_source=chatgpt.com#?c=&m=&s=&cv=
- [64] «El portal de la tesis». Accedido: 3 de mayo de 2025. [En línea]. Disponible en: <https://recursos.ucol.mx/tesis/investigacion.php>

 UNIVERSIDAD CESMAG MTN-REG. 169.2017-7 VIGILANCIA MINISTERIO	CARTA DE ENTREGA TRABAJO DE GRADO O TRABAJO DE APLICACIÓN – ASESOR(A)	CÓDIGO: AAC-BL-FR-032
		VERSIÓN: 1
		FECHA: 09/JUN/2022

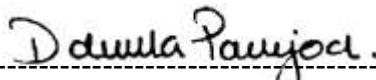
San Juan de Pasto, 10 de agosto de 2026

Biblioteca
REMIGIO FIORE FORTEZZA OFM. CAP.
 Universidad CESMAG
 Pasto

Saludo de paz y bien.

Por medio de la presente se hace entrega del Trabajo de Grado / Trabajo de Aplicación denominado **PROYECCIÓN TOPOLÓGICA DEL AGUA PARA LA TOMA DE DECISIONES INFORMADAS MEDIANTE ALGORITMOS DE FORECASTING APLICADOS AL NDWI EN IMÁGENES SATELITALES**, presentado por los autores **TANIA DEISY PINCHAO MUÑOZ** y **ANDRES JAVIER VALLEJO JARAMILLO** del Programa Académico Ingeniería de Sistemas al correo electrónico biblioteca.trabajosdegrado@unicesmag.edu.co. Manifiesto como asesora, que su contenido, resumen, anexos y formato PDF cumple con las especificaciones de calidad, guía de presentación de Trabajos de Grado o de Aplicación, establecidos por la Universidad CESMAG, por lo tanto, se solicita el paz y salvo respectivo.

Atentamente,



DALILA PACHAJOA PACHAJOA
 Número de documento: 37086647
 Ingeniería de sistemas
 Cel. 3154905682
 Correo: dmpachajoa@unicesmag.edu.co

 UNIVERSIDAD CESMAG MT: 800.109.337-7 NO INCLUYE PUBLICACIÓN	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

INFORMACIÓN DEL (LOS) AUTOR(ES)	
Nombres y apellidos del autor: Tania Deisy Pinchao Muñoz	Documento de identidad: 1080040213
Correo electrónico: tdpinchao.0213@unicesmag.edu.co	Número de contacto: 3126672260
Nombres y apellidos del autor: Andres Javier Vallejo Jaramillo	Documento de identidad: 1004215648
Correo electrónico: andresjavierv50@gmail.com	Número de contacto: 3137445747
Nombres y apellidos del asesor: Dalila Mercedes Pachajoa Pachajoa	Documento de identidad: 37086647
Correo electrónico: dmpachajoa@unicesmag.edu.co	Número de contacto: 3154905682
Título del trabajo de grado: PROYECCIÓN TOPOLÓGICA DEL AGUA PARA LA TOMA DE DECISIONES INFORMADAS MEDIANTE ALGORITMOS DE FORECASTING APLICADOS AL NDWI EN IMÁGENES SATELITALES	
Facultad y Programa Académico: Facultad de ingeniería, Ingeniería de Sistemas	

En mi (nuestra) calidad de autor(es) y/o titular (es) del derecho de autor del Trabajo de Grado o de Aplicación señalado en el encabezado, confiero (conferimos) a la Universidad CESMAG una licencia no exclusiva, limitada y gratuita, para la inclusión del trabajo de grado en el repositorio institucional. Por consiguiente, el alcance de la licencia que se otorga a través del presente documento, abarca las siguientes características:

- La autorización se otorga desde la fecha de suscripción del presente documento y durante todo el término en el que el (los) firmante(s) del presente documento conserve (mos) la titularidad de los derechos patrimoniales de autor. En el evento en el que deje (mos) de tener la titularidad de los derechos patrimoniales sobre el Trabajo de Grado o de Aplicación, me (nos) comprometo (comprometemos) a informar de manera inmediata sobre dicha situación a la Universidad CESMAG. Por consiguiente, hasta que no exista comunicación escrita de mi(nuestra) parte informando sobre dicha situación, la Universidad CESMAG se encontrará debidamente habilitada para continuar con la publicación del Trabajo de Grado o de Aplicación dentro del repositorio institucional. Conozco(conocemos) que esta autorización podrá revocarse en cualquier momento, siempre y cuando se eleve la solicitud por escrito para dicho fin ante la Universidad CESMAG. En estos eventos, la Universidad CESMAG cuenta con el plazo de un mes después de recibida la petición, para desmarcar la visualización del Trabajo de Grado o de Aplicación del repositorio institucional.
- Se autoriza a la Universidad CESMAG para publicar el Trabajo de Grado o de Aplicación en formato digital y teniendo en cuenta que uno de los medios de publicación del repositorio institucional es el internet, acepto(amos) que el Trabajo de Grado o de Aplicación circulará con un alcance mundial.
- Acepto (aceptamos) que la autorización que se otorga a través del presente documento se realiza a título gratuito, por lo tanto, renuncio(amos) a recibir emolumento alguno por la publicación, distribución, comunicación pública y/o cualquier otro uso que se haga en los términos de la presente autorización y de la licencia o programa a través del cual sea publicado el Trabajo de grado o de Aplicación.

 UNIVERSIDAD CESMAG MT: 800.109.337-7 VICERRECTORÍA	AUTORIZACIÓN PARA PUBLICACIÓN DE TRABAJOS DE GRADO O TRABAJOS DE APLICACIÓN EN REPOSITORIO INSTITUCIONAL	CÓDIGO: AAC-BL-FR-031
		VERSIÓN: 1
		FECHA: 09/JUN/2022

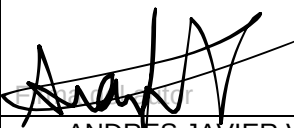
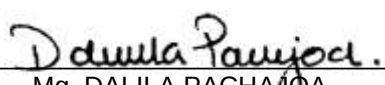
- d) Manifiesto (manifestamos) que el Trabajo de Grado o de Aplicación es original realizado sin violar o usurpar derechos de autor de terceros y que ostento(amos) los derechos patrimoniales de autor sobre la misma. Por consiguiente, asumo(asumimos) toda la responsabilidad sobre su contenido ante la Universidad CESMAG y frente a terceros, manteniéndose indemne de cualquier reclamación que surja en virtud de la misma. En todo caso, la Universidad CESMAG se compromete a indicar siempre la autoría del escrito incluyendo nombre de(los) autor(es) y la fecha de publicación.
- e) Autorizo(autorizamos) a la Universidad CESMAG para incluir el Trabajo de Grado o de Aplicación en los índices y buscadores que se estimen necesarios para promover su difusión. Así mismo autorizo (autorizamos) a la Universidad CESMAG para que pueda convertir el documento a cualquier medio o formato para propósitos de preservación digital.

NOTA: En los eventos en los que el trabajo de grado o de aplicación haya sido trabajado con el apoyo o patrocinio de una agencia, organización o cualquier otra entidad diferente a la Universidad CESMAG. Como autor(es) garantizo(amos) que he(hemos) cumplido con los derechos y obligaciones asumidos con dicha entidad y como consecuencia de ello dejo(dejamos) constancia que la autorización que se concede a través del presente escrito no interfiere ni transgrede derechos de terceros.

Como consecuencia de lo anterior, autorizo(autorizamos) la publicación, difusión, consulta y uso del Trabajo de Grado o de Aplicación por parte de la Universidad CESMAG y sus usuarios así:

- Permiso(permitimos) que mi(nuestro) Trabajo de Grado o de Aplicación haga parte del catálogo de colección del repositorio digital de la Universidad CESMAG por lo tanto, su contenido será de acceso abierto donde podrá ser consultado, descargado y compartido con otras personas, siempre que se reconozca su autoría o reconocimiento con fines no comerciales.

En señal de conformidad, se suscribe este documento en San Juan de Pasto a los 10 días del mes de agosto del año 2026

	
TANIA DEISY PINCHAO MUÑOZ	ANDRES JAVIER VALLEJO JARAMILLO
 Mg. DALILA PACHAJOA	