Revista de Investigaciones y Aplicaciones Geomáticas en Geociencias – RIAG-G
UNIVERSIDAD MAYOR DE SAN ANDRÉS
Carrera de Geodesia, Topografía y Geomática
Instituto de Investigaciones y Aplicaciones Geomáticas – IIAG / ISSN: 3104-5146

EVALUACIÓN DE XGBOOST RESPECTO A HEC-RAS PARA MODELACIÓN DE INUNDACIONES EN EL RÍO JILUSAYA, LA PAZ

Lic. Cristhian Jairo Calle Huanca
Universidad Mayor de San Andrés
Correo Electrónico: cristhiancalle053@gmail.com

Resumen

La presente investigación desarrolla un modelo hidráulico para escorrentía superficial mediante herramientas de Machine Learning (ML), específicamente Extreme Gradient Boosting (XGBoost), determinando niveles de confiabilidad y coincidencia de áreas susceptibles a inundación en comparación con HEC-RAS en el río Jilusaya. El trabajo responde a la necesidad de gestión precisa del riesgo de inundación ante la intensificación de eventos hidrometeorológicos extremos asociados al cambio climático.

La metodología integró levantamiento fotogramétrico, modelación hidrológica con HEC-HMS utilizando datos del SENHAMI y análisis espacial mediante Sistemas de Información Geográfica (SIG) para periodos de retorno de 10, 30 y 50 años. Se utilizó XGBoost, un algoritmo ML reconocido por su capacidad para procesar grandes volúmenes de datos y resolver problemas no lineales complejos. Para la validación del modelo predictivo desarrollado con XGBoost, se utilizó como referencia el modelo hidráulico HEC-RAS, ya que este enfoque se fundamenta en la "Guía de delimitación de franjas de seguridad en ríos, incorporando factores de variabilidad y cambio climático" (Ministerio de Medio Ambiente y Agua, 2021), donde se establece el uso de HEC-RAS como herramienta técnica oficial para la simulación de caudales y escenarios de inundación, debido a que el río estudiado carece de registros históricos y su reducido ancho limita la disponibilidad de imágenes satelitales de alta resolución, por este motivo HEC-RAS constituye la base metodológica para contrastar y validar los resultados obtenidos con XGBoost. Donde La evaluación cuantitativa empleó métricas como el error cuadrático medio (MSE = 0,000485) y análisis de coincidencia espacial.

Se pudo observar que el modelo XGBoost demostró ventajas en velocidad de procesamiento, ya que redujo significativamente el tiempo de procesamiento a 1 hora en comparación con HEC-RAS que fueron 8 horas, utilizando el mismo entorno de procesamiento. Por otro lado, los resultados muestran niveles de coincidencia espacial del 42,83%, 52,66% y 57,39% para los periodos de retorno de 10, 30 y 50 años, respectivamente, alcanzando una precisión máxima del 57,39% en escenarios extremos. El modelo obtenido con ML presenta limitaciones que se asocian a la insensibilidad frente a variaciones de caudal y al procesamiento por bloques espaciales, lo que limita su capacidad para representar dinámicas hidráulicas complejas. Por lo que la investigación evidencia el potencial del ML como herramienta complementaria a métodos tradicionales en hidrología, para la protección de comunidades vulnerables ante eventos extremos, con aplicaciones en gestión hídrica y planificación territorial.

Abstract

This research develops a hydraulic model for surface runoff using Machine Learning (ML) tools, specifically Extreme Gradient Boosting (XGBoost), to determine levels of reliability and spatial coincidence of flood-prone areas in comparison with HEC-RAS in the Jilusaya River. The study addresses the need for precise flood risk management in the face of intensifying extreme hydrometeorological events associated with climate change.

The methodology integrated photogrammetric surveying, hydrological modeling with HEC-HMS using SENHAMI data, and spatial analysis through Geographic Information Systems (GIS) for return periods of 10, 30, and 50 years. XGBoost, a ML algorithm recognized for its ability to process large volumes of data and solve complex nonlinear problems, was employed. For the validation of the predictive model developed with XGBoost, the hydraulic model HEC-RAS was used as a reference, based on the Guide for the delimitation of river safety strips incorporating variability and climate change factors (Ministry of Environment and Water, 2021), which establishes HEC-RAS as the official technical tool for simulating flows and flood scenarios. Since the studied river lacks historical records and its narrow width limits the availability of high-resolution satellite imagery, HEC-RAS constitutes the methodological basis for contrasting and validating the results obtained with XGBoost. The quantitative evaluation employed metrics such as the mean squared error (MSE = 0.000485) and spatial coincidence analysis.

It was observed that the XGBoost model demonstrated advantages in processing speed, reducing computation time to 1 hour compared to 8 hours with HEC-RAS, under the same processing environment. On the other hand, results show spatial coincidence levels of 42.83%, 52.66%, and 57.39% for return periods of 10, 30, and 50 years, respectively, reaching a maximum accuracy of 57.39% in extreme scenarios. The ML-based model presents limitations associated with insensitivity to flow variations and block-based spatial processing, which restrict its ability to represent complex hydraulic dynamics. Therefore, the research highlights the potential of ML as a complementary tool to traditional hydrological methods, contributing to the protection of vulnerable communities against extreme events, with applications in water management and territorial planning.

Palabras clave

Modelación hidráulica, Machine Learning, XGBoost, HEC-RAS, Gestión de riesgos de inundación, SIG, Aprendizaje automático, Río Jilusaya.

1. Introducción

La ciudad de La Paz enfrenta desafíos constantes derivado de su ubicación geográfica y características topográficas e hidrológicas particulares. Situada sobre la cuenca del río La Paz, la ciudad es atravesada por 364 ríos afluentes que drenan áreas urbanas densamente pobladas, incluyendo los ríos Choqueyapu, Orckojahuira, Irpavi, Achumani y Huayñajahuira. (Ibañez Apaza, 2022). Además de que, en años recientes, la frecuencia e intensidad de eventos hidrometeorológicos extremos han incrementado debido al cambio climático, exacerbando la vulnerabilidad de la ciudad ante inundaciones y deslizamientos. El evento del 9 de marzo de 2024 ejemplifica esta problemática: intensas lluvias provocaron el desborde del río La Paz, inundando zonas críticas como Mallasa, Mallasilla y Río Abajo (ERBOL, 2024). Entonces la interacción entre precipitaciones intensas y las altas pendientes de las subcuencas no solo incrementa el caudal, sino también la carga sólida, generando erosión y acumulación de sedimentos en sectores estratégicos, evidenciando la urgente necesidad de contar con información hidrológica para una gestión eficiente del riesgo.

En este contexto, la modelación hidrológica e hidráulica se vuelve fundamental para prever el comportamiento de los ríos durante eventos extremos y mitigar los daños causados por crecidas. Tradicionalmente, se han empleado modelos convencionales como HEC-RAS para la simulación hidráulica; sin embargo, el Machine Learning (ML) ha revolucionado la capacidad de procesamiento y análisis de grandes volúmenes de datos, integrando información de sensores automáticos, imágenes satelitales y modelos topográficos de alta resolución mediante tecnología LIDAR o aeronaves remotamente pilotadas (Gonzales-Inca et al., 2022). Técnicas avanzadas de aprendizaje automático como XGBoost han demostrado mejorar significativamente la predicción de eventos extremos debido a su capacidad para procesar grandes volúmenes de datos y resolver problemas no lineales complejos (Seckin, Cobaner, & Yurtal, 2013), permitiendo la identificación de áreas vulnerables.

Los algoritmos de ML han demostrado una gran capacidad para modelar relaciones no lineales entre variables hidrológicas y topográficas, permitiendo la identificación de patrones espaciales complejos en la ocurrencia de inundaciones (Mosavi et al., 2018; Sit et al., 2024). Entre estas, los modelos que están basados en árboles de decisión como XGBoost son de gran importancia debido a la eficiencia y capacidad que tienen para manejar grandes volúmenes de datos, optimizando la predicción de eventos extremos (Seckin et al., 2013). Por otro lado, a pesar de estas ventajas, varios estudios han evidenciado que los modelos basados en ML presentan algunas limitaciones en la representación de procesos físicos del flujo, ya que estos operan como aproximadores estadísticos y no incorporan las ecuaciones hidráulicas que gobiernan la dinámica del agua. Pero, los modelos físicos como HEC-RAS tienen una mayor coherencia hidráulica, especialmente en la simulación de la propagación del flujo y la delimitación precisa de áreas inundables (Teng et al., 2020). Esta diferencia se vuelve particularmente notable en entornos complejos como cuencas urbanas andinas, donde la interacción entre topografía y variabilidad climática genera dinámicas hidráulicas altamente heterogéneas.

En Bolivia, la implementación de ML en hidrología se encuentra en fase emergente. Aunque el SENAMHI ha avanzado en políticas de acceso abierto a datos climatológicos, la infraestructura para recolección masiva de datos sigue siendo limitada, restringiendo el potencial de estas tecnologías en

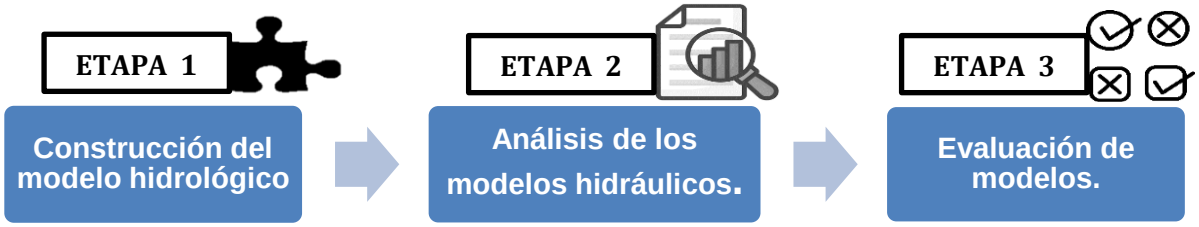
la gestión hídrica nacional, por lo que el presente estudio postula que la modelación mediante esta herramienta permitirá determinar niveles de confiabilidad y coincidencia de áreas susceptibles a inundación respecto a HEC-RAS. Este trabajo busca responder: ¿Puede XGBoost generar modelos hidráulicos con confiabilidad comparable a métodos convencionales?

Para responder esta pregunta, se desarrolla un modelo basado en XGBoost validado mediante comparación con HEC-RAS a través de análisis espacial con SIG para periodos de retorno de 10, 30 y 50 años y los resultados buscan contribuir al avance de tecnologías emergentes en Bolivia, demostrando cómo estas herramientas pueden complementar métodos tradicionales y ofrecer soluciones más ágiles para la protección de comunidades vulnerables ante eventos hidrometeorológicos extremos cada vez más frecuentes e intensos.

2. Metodología y Desarrollo

El proceso metodológico para la recolección de datos y su posterior análisis de la presente investigación se desarrollará en 3 etapas:

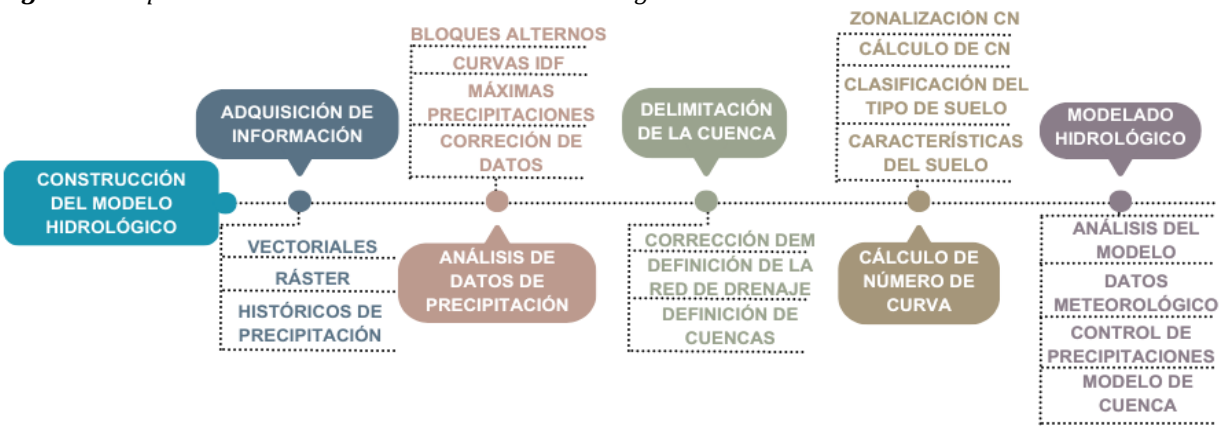
Figura 1: *Etapas para la evaluación de modelos hidráulicos*



Nota: *Cada fase incluye procedimientos relacionados con la generación de resultados parciales.*

Etapa 1: Construcción del modelo hidrológico. En esta etapa se lleva a cabo la construcción del modelo hidrológico se basa en el análisis y tratamiento de datos históricos de precipitaciones, imágenes satelitales y características de suelo. Se utiliza técnicas de SIG e hidrológicos, partiendo de la obtención de datos, hasta llegar a la obtención del modelo hidrológico.

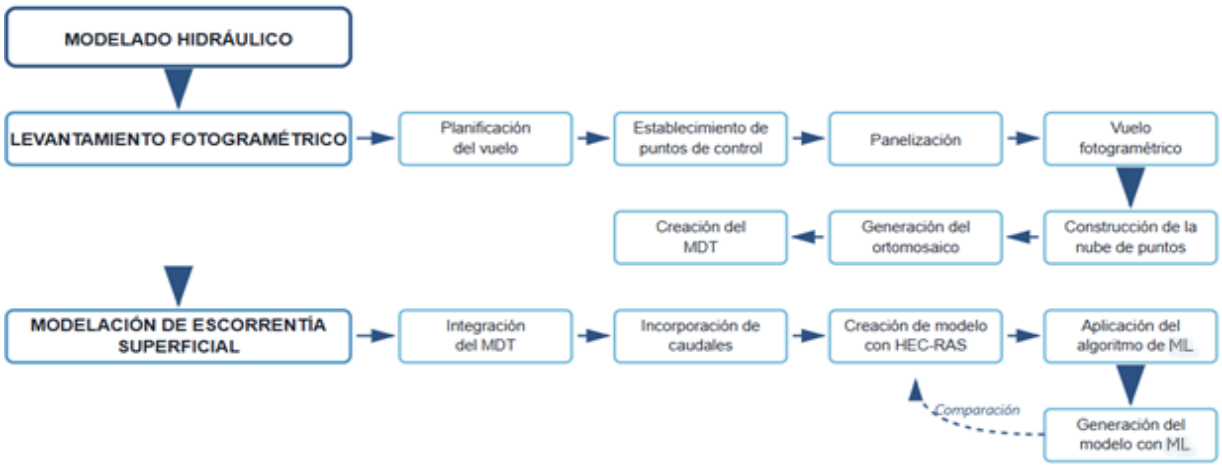
Figura 2: *Esquema de la construcción del modelo hidrológico.*



Etapa 2: Análisis de los modelos hidráulicos. Un modelo hidráulico depende de la precisión de los datos de entrada, siendo fundamental disponer de información confiable sobre geometría del terreno, topografía, características del lecho fluvial y condiciones de contorno como caudales y

precipitación. El análisis debe considerar las variaciones del terreno por el flujo de agua para determinar áreas de inundación durante eventos extremos, resultados esenciales para evaluar riesgos, identificar zonas vulnerables y diseñar medidas de mitigación. La simulación del área de inundación permite comprender el comportamiento del agua en puntos críticos, optimizando la planificación y gestión de infraestructuras hidráulicas.

Figura 3: *Esquema de la construcción del modelo hidráulico.*



El desarrollo del modelo hidráulico con HEC-RAS se fundamentó en un levantamiento fotogramétrico del tramo de estudio del río Jilusaya. La metodología incluyó la planificación detallada del vuelo, considerando cobertura, altura y traslape necesario, seguida del establecimiento de puntos de control terrestre georreferenciados mediante GNSS de alta precisión para asegurar la exactitud posicional. La panelización definió las trayectorias óptimas de la aeronave remotamente pilotada (RPA), ejecutándose posteriormente el vuelo fotogramétrico con traslape 75% entre imágenes. El procesamiento fotogramétrico generó la nube de puntos permitiendo la creación del ortomosaico georreferenciado y del Modelo Digital del Terreno (MDT) con resolución espacial adecuada para capturar las variaciones topográficas significativas. Este MDT se integró en HEC-RAS como geometría base, definiendo las secciones transversales del río, el cauce principal y las áreas de desborde potencial. Finalmente, se incorporaron los caudales de diseño para periodos de retorno de 10, 30 y 50 años, calculados mediante HEC-HMS, estableciendo las condiciones de contorno necesarias para simular el flujo en régimen permanente y delimitar las áreas susceptibles a inundación.

En cambio, la metodología basada en Machine Learning se basa en el aprendizaje de patrones espaciales de inundación a partir de las mismas variables de entrada utilizadas en el modelo hidráulico HEC-RAS. Es decir, el modelo de ML no replica resultados del modelo físico, sino que aprende relaciones implícitas entre variables topográficas e hidrológicas que condicionan la ocurrencia de inundaciones. Para ello, se utilizó el algoritmo XGBoost en lenguaje Python utilizando el entorno Spyder. En cada iteración, el modelo ajusta nuevos árboles que corrigen los errores de predicción previos, generando un ensamble que mejora progresivamente su desempeño.

Las variables de entrada que se utilizaron fueron el MDT, caudales de diseño asociados a distintos periodos de retorno y un valor representativo de rugosidad de Manning. A partir del MDT se

incorporaron variables, como la pendiente y una máscara de zonas bajas basada en percentiles de elevación, con el objetivo de resaltar áreas susceptibles a inundación. Este preproceso permitió reducir la dimensión del problema y centrar el entrenamiento en sectores relevantes. Además, el modelo fue realizado en procesamiento por bloques espaciales, debido por la alta resolución del MDT y las limitaciones de memoria del equipo que se utilizó, donde cada bloque del terreno fue procesado de manera independiente, generando subconjuntos de datos donde se construyeron matrices de características que integran elevación, pendiente, caudal, tiempo y rugosidad. En este contexto, el modelo fue entrenado como un problema de regresión, utilizando como variable objetivo la elevación del terreno en zonas previamente filtradas como susceptibles, lo que permite identificar indirectamente patrones de acumulación y desborde del flujo.

El entrenamiento se realizó mediante una partición de los datos 80% entrenamiento y 20% validación, evaluando el rendimiento del modelo a través del error cuadrático medio (MSE) y la evolución del error RMSE durante las iteraciones. En adición, se implementó un esquema de superposición entre bloques para reducir efectos de borde, seguido de un proceso de post-procesamiento mediante operaciones morfológicas, con el fin de mejorar la coherencia espacial de las áreas inundadas predichas.

Figura 4: *Entrenamiento utilizando xgb.train.*

```
# Entrenar el modelo usando xgb.train
eval_set = [(dtrain, 'train'), (dtest, 'test')]
modelo_xgb = xgb.train(params, dtrain, num_boost_round=100, evals=eval_set,
                       early_stopping_rounds=10, evals_result=evals_result, verbose_eval=False)
```

Etapa 3: Evaluación de modelos. La evaluación de modelos hidráulicos se desarrolló mediante un análisis comparativo entre el software HEC-RAS y algoritmos ML basados en XGBoost. El modelo hidráulico tradicional en HEC-RAS simuló el flujo en régimen permanente para periodos de retorno de 10, 30 y 50 años, generando perfiles de superficie del agua, distribución de velocidades y mapas de profundidad de inundación que permitieron la delimitación de zonas vulnerables en el tramo de estudio del río Jilusaya. Paralelamente, el modelo ML fue entrenado para la predicción espacial de áreas susceptibles a inundación, identificando patrones hidrológicos y relaciones no lineales entre variables topográficas, hidráulicas y morfológicas del cauce. El desempeño del modelo de ML se evaluó mediante el error cuadrático medio (MSE).

La comparación entre ambos enfoques metodológicos se ejecutó mediante tres componentes fundamentales implementados en Sistemas de Información Geográfica (SIG). Primero, se identificaron los sectores de desborde utilizando los datos de flujo de ambos modelos para los eventos de precipitación extrema simulados. Segundo, se cuantificó la coincidencia espacial mediante la superposición de mapas de inundación generados por HEC-RAS y ML, determinando el porcentaje de concordancia y las discrepancias en la predicción de zonas de riesgo. Finalmente, se evalúa la eficiencia en el procesamiento de datos y capacidad para manejar la complejidad topográfica del terreno, estableciendo así las ventajas comparativas de cada metodología en las predicciones del modelo de ML respecto al enfoque tradicional.

3. Resultados

El presente trabajo se desarrolló mediante una metodología escalonada de tres etapas, donde los resultados generados en cada fase funcionan como datos de entrada para la siguiente, estableciendo una cadena metodológica que conduce hacia los resultados finales de la investigación.

Resultados Etapa 1: Construcción del modelo hidrológico.

La construcción del modelo hidrológico arrojó como principales productos los sumarios globales para escenarios de 10, 30 y 50 años de periodo de retorno. Dada la considerable cantidad de datos generados sobre caudales de entrada y salida en todo el sistema, se realizó una selección estratégica enfocándose en el flujo hidrográfico del río R1, que representa el área de estudio prioritaria. Estos valores de caudal para los tres periodos analizados fueron extraídos y consolidados como datos de entrada para la siguiente fase metodológica.

Tabla 1: *Sumario global del modelo hidrológico.*

Para un periodo de retorno de 10 años.

Hydrologic Element	Drainage Area (KM2)	Peak Discharge (M3/S)	Time of Peak	Volume (MM)
SC3	0,9361	0,5	01ago.2024, 02:40	2,67
SC1	0,8285	0,4	01ago.2024, 02:45	2,93
J4	1,7646	0,9	01ago.2024, 02:40	2,79
R4	1,7646	0,8	01ago.2024, 02:55	2,79
SC2	4,2091	1,6	01ago.2024, 03:10	2,76
SC6	0,5215	0,4	01ago.2024, 02:30	3,65
J3	6,4952	2,7	01ago.2024, 03:00	2,84
R3	6,4952	2,5	01ago.2024, 03:30	2,84
SC4	0,9027	0,9	01ago.2024, 02:40	4,81
SC7	0,8301	1,5	01ago.2024, 02:40	8,51
J2	8,2280	3,7	01ago.2024, 03:15	3,63
R2	8,2280	3,6	01ago.2024, 03:25	3,63
SC5	0,6833	0,9	01ago.2024, 02:35	5,80
SC8	0,3396	0,5	01ago.2024, 02:20	5,36
J1	9,2509	4,1	01ago.2024, 03:20	3,85
R1	9,2509	4,1	01ago.2024, 03:40	3,85
SC9	1,2490	1,8	01ago.2024, 02:30	5,91
Sink-1	10,4999	4,6	01ago.2024, 03:30	4,10

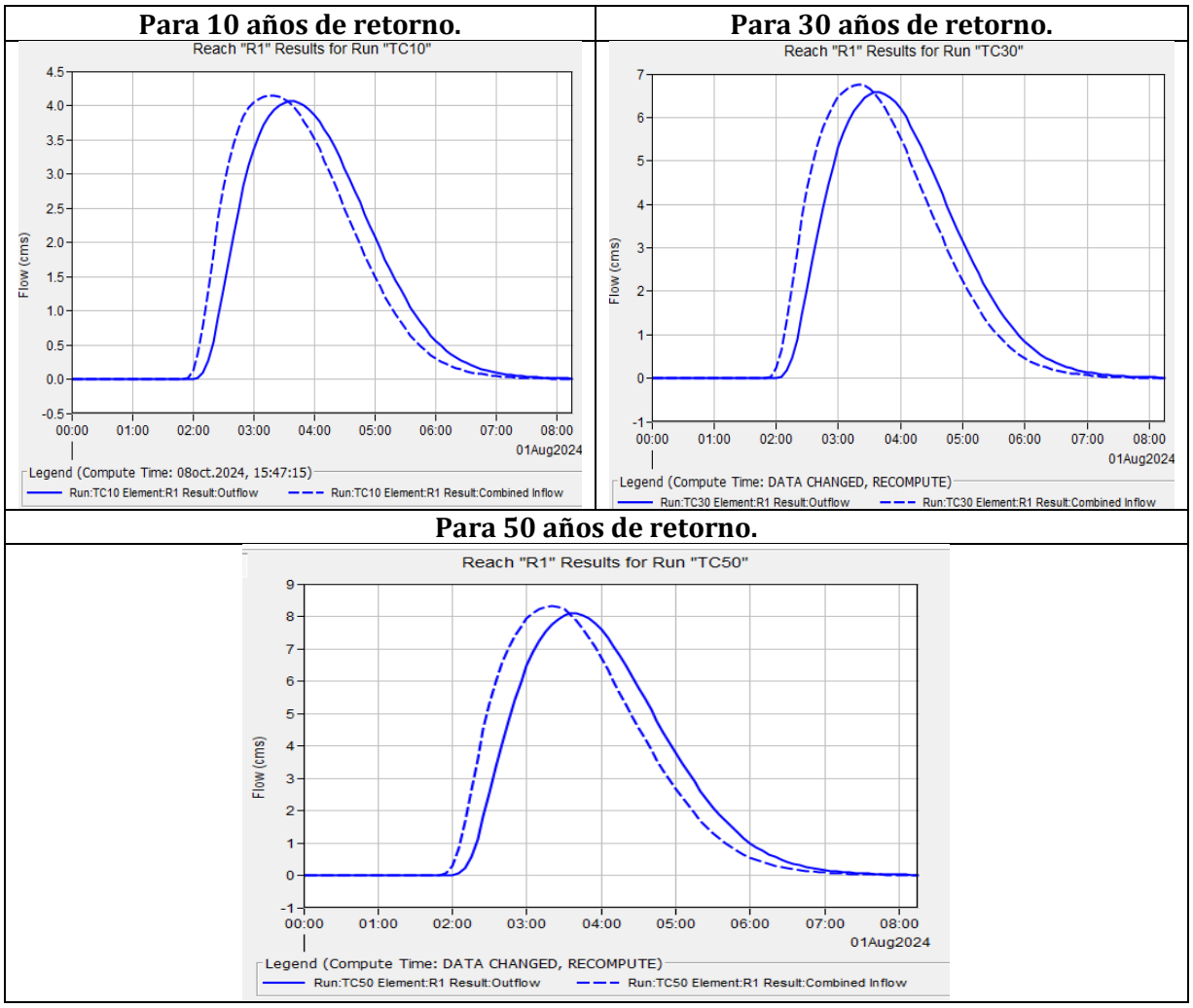
Para un periodo de retorno de 30 años.

Hydrologic Element	Drainage Area (KM2)	Peak Discharge (M3/S)	Time of Peak	Volume (MM)
SC3	0,9361	0,9	01ago.2024, 02:40	4,55
SC1	0,8285	0,8	01ago.2024, 02:40	4,90
J4	1,7646	1,7	01ago.2024, 02:40	4,71
R4	1,7646	1,5	01ago.2024, 02:55	4,71
SC2	4,2091	2,9	01ago.2024, 03:05	4,67
SC6	0,5215	0,8	01ago.2024, 02:25	5,85
J3	6,4952	4,7	01ago.2024, 03:00	4,78
R3	6,4952	4,4	01ago.2024, 03:25	4,78
SC4	0,9027	1,4	01ago.2024, 02:40	7,35
SC7	0,8301	2,2	01ago.2024, 02:35	11,91
J2	8,2280	6,1	01ago.2024, 03:15	5,78
R2	8,2280	6,0	01ago.2024, 03:25	5,78
SC5	0,6833	1,4	01ago.2024, 02:30	8,60
SC8	0,3396	0,8	01ago.2024, 02:20	8,04
J1	9,2509	6,8	01ago.2024, 03:20	6,07
R1	9,2509	6,6	01ago.2024, 03:40	6,07
SC9	1,2490	2,8	01ago.2024, 02:30	8,74
Sink-1	10,4999	7,3	01ago.2024, 03:30	6,39

Para un periodo de retorno de 50 años.

Hydrologic Element	Drainage Area (KM2)	Peak Discharge (M3/S)	Time of Peak	Volume (MM)
SC3	0,9361	1,1	01ago.2024, 02:35	5,69
SC1	0,8285	1,0	01ago.2024, 02:40	6,08
J4	1,7646	2,1	01ago.2024, 02:40	5,88
R4	1,7646	1,9	01ago.2024, 02:55	5,88
SC2	4,2091	3,7	01ago.2024, 03:05	5,82
SC6	0,5215	1,0	01ago.2024, 02:25	7,16
J3	6,4952	6,0	01ago.2024, 02:55	5,94
R3	6,4952	5,6	01ago.2024, 03:25	5,94
SC4	0,9027	1,7	01ago.2024, 02:35	8,83
SC7	0,8301	2,5	01ago.2024, 02:35	13,82
J2	8,2280	7,6	01ago.2024, 03:15	7,06
R2	8,2280	7,5	01ago.2024, 03:25	7,06
SC5	0,6833	1,6	01ago.2024, 02:30	10,21
SC8	0,3396	1,0	01ago.2024, 02:20	9,60
J1	9,2509	8,3	01ago.2024, 03:20	7,38
R1	9,2509	8,1	01ago.2024, 03:40	7,38
SC9	1,2490	3,3	01ago.2024, 02:30	10,36
Sink-1	10,4999	9,0	01ago.2024, 03:30	7,73

Figura 5: *Flujo hidrográfico del rio R1 para 10 años de retorno.*



Resultados Etapa 2: Análisis de los modelos hidráulicos. En esta etapa, el análisis hidráulico mediante HEC-RAS generó modelos diferenciados para los periodos de retorno de 10, 30 y 50 años, reflejando las variaciones en el comportamiento del flujo según la magnitud del evento hidrológico. En contraste, los modelos de ML desarrollados mediante XGBoost para los mismos periodos no exhibieron variación entre los escenarios, generando resultados equivalentes. XGBoost es un algoritmo de conjunto basado en árboles de gradiente boosting que ha demostrado alta eficiencia en la predicción de inundaciones, con tiempos de procesamiento reducidos y capacidad para manejar grandes volúmenes de datos espaciales. Por esta razón, se consideró únicamente el modelo de ML de 50 años para el análisis comparativo, correspondiente al evento más extremo evaluado. Para validar la confiabilidad de este modelo, se calcularon el Error Cuadrático Medio (MSE) y la Raíz del Error Cuadrático Medio (RMSE), permitiendo cuantificar la precisión predictiva del algoritmo respecto al modelo de referencia HEC-RAS y su capacidad para capturar los patrones espaciales de inundación en el río Jilusaya.

Figura 6: *Mapa de inundación para 10 años de retorno.*



Figura 7: *Mapa de inundación para 30 años de retorno.*



Figura 8: *Mapa de inundación para 50 años de retorno.*



Figura 9: *Áreas de inundación generadas con ML conectadas.*



Figura 10: *Gráfico MSE por bloque para 50 años de retorno.*

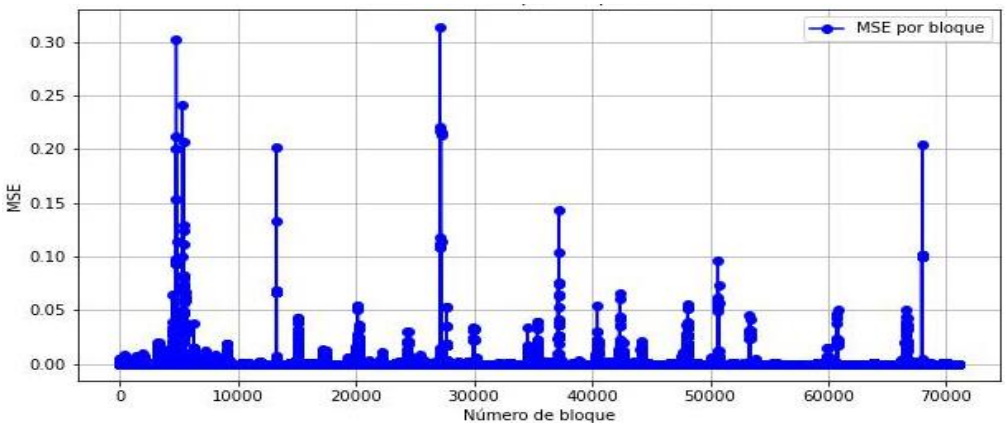


Figura 11: *Evolución del RMSE por iteraciones para 50 años de retorno.*

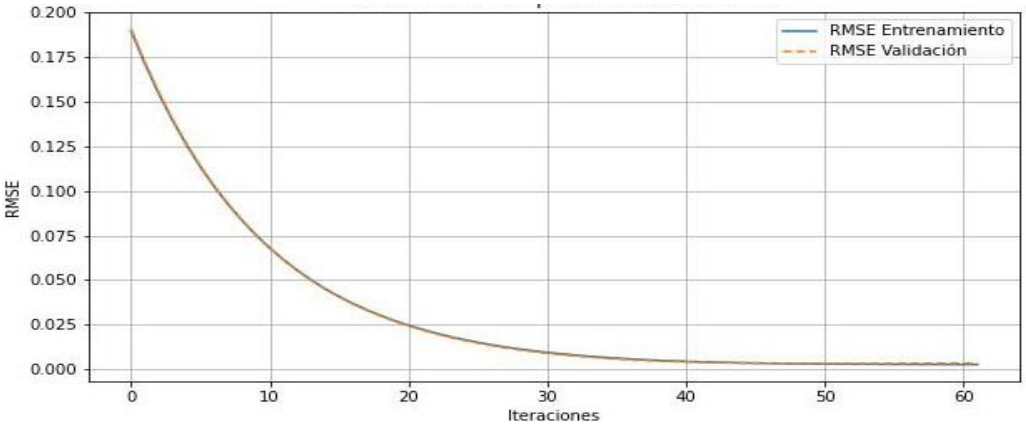


Figura 13: *Áreas de inundación Hec-Ras vs ML.*

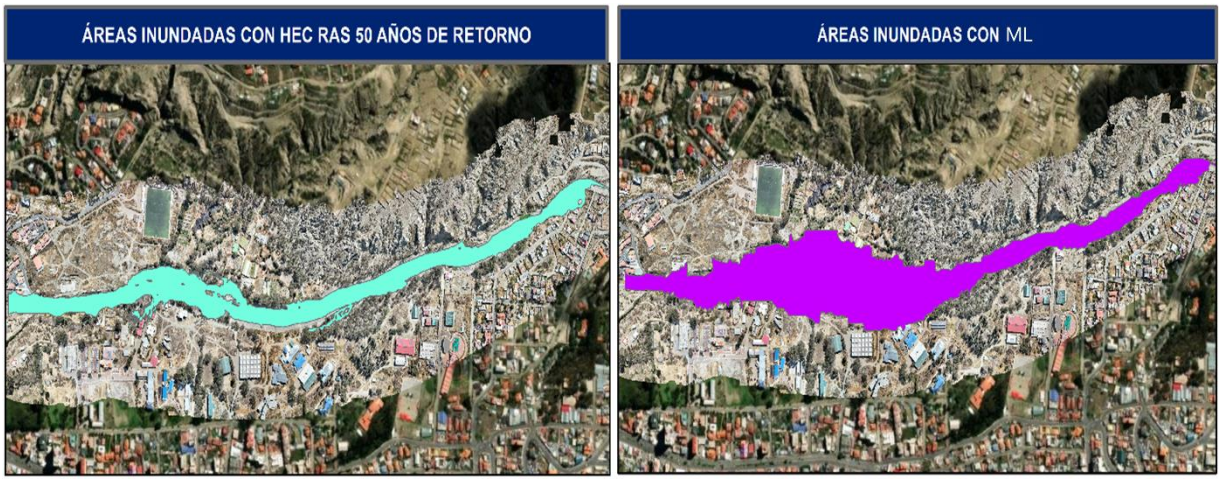


Figura 14: *Áreas de coincidencia en diferentes tramos: periodo de retorno de 50 años.*

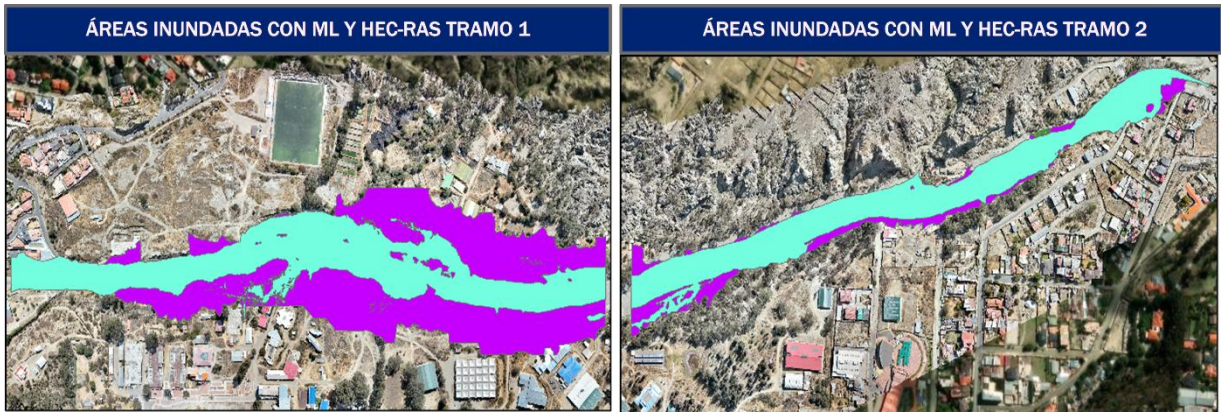


Tabla 2: *Valores de áreas de coincidencia en distintos tramos.*

MODELO	HEC-RAS TRAMO 1	ML TRAMO 1	HEC-RAS TRAMO 2	ML TRAMO 2
ÁREA m2	45874,980	99625,643	33548.317	38762.466
ÁREA Km2	0,045875	0,099626	0.033548	0.038762
COINCIDENCIA %	46,047		86,548	

Nota: *La tabla presenta la comparación de áreas inundadas estimadas mediante HEC-RAS y el modelo de Machine Learning(ML) en dos tramos del río Jilusaya.*

4. Discusión

El comportamiento del modelo hidrológico obtenido, con los caudales crecientes para mayores periodos de retorno y tiempos de respuesta rápidos, es consistente con la dinámica de cuencas andinas de altas pendientes y pequeña extensión, donde las características morfométricas controlan el tiempo de concentración, mientras la magnitud de las precipitaciones define la intensidad del caudal. No obstante, debe considerarse que el modelo asume condiciones estacionarias de uso de suelo, lo que puede introducir incertidumbre en escenarios futuros.

En cuanto al modelo de Machine Learning, la semejanza observada en las áreas inundadas para distintos escenarios de caudal nos muestra una limitada sensibilidad frente a variaciones en la magnitud del flujo, esto contrasta con los resultados obtenidos en HEC-RAS, donde el área de la inundación responde de manera diferenciada a los cambios del caudal. Esta limitación se puede asociar al procesamiento por bloques espaciales y naturaleza del modelo, que aprende relaciones estadísticas a partir de variables de entrada sin representar la dinámica hidráulica.

El enfoque por bloques permitió gestionar la alta resolución del MDT bajo restricciones computacionales, pero introdujo efectos de homogenización espacial y pérdida de continuidad hidráulica. Aunque se aplicaron estrategias de superposición y post-procesamiento, estas tienden a suavizar las diferencias entre escenarios, reduciendo la variabilidad espacial de los resultados. Este efecto explica, en parte, la similitud de las áreas inundadas para diferentes periodos de retorno.

Tabla 3: *Comparación de eficiencia computacional.*

Parámetro técnico	XGBoost	HEC-RAS
Hardware	Ryzen 5 5700U, 8 GB RAM, SSD 1 TB	Igual
Tiempo total de ejecución	1 h	8 h
Uso de memoria	Moderado	Alto

Por otro lado, el bajo valor del MSE refleja un buen ajuste del modelo, sin embargo, no necesariamente captura errores espaciales relevantes. Para esto, las métricas como el Coeficiente de Cohen Kappa y el Critical Success Index permiten una evaluación más representativa del desempeño espacial. Los resultados indican una concordancia moderada a alta en la distribución general de las áreas inundadas, con una mejora progresiva en la capacidad de detección a medida que aumenta el caudal. También, el análisis espacial muestra que el modelo reproduce de forma adecuada en el eje principal del flujo, aumentando los aciertos en el cauce, pero presenta debilidades en la representación de la expansión lateral de la inundación. La presencia de falsos negativos en zonas adyacentes y falsos positivos en áreas periféricas sugiere que el modelo simplifica procesos complejos como el desbordamiento y la interacción con la planicie de inundación, particularmente en sectores de baja pendiente.

Tabla 4: *Evaluación del desempeño espacial.*

Escenario	Kappa	CSI
PR10	0.5226	0.4069
PR30	0.6036	0.4891
PR50	0.6397	0.5279

Nota: *La tabla presenta los valores del coeficiente de Cohen's Kappa y del Critical Success Index obtenidos para los escenarios de periodos de retorno de 10, 30 y 50 años*

Además, la variación del desempeño entre distintos tramos del río indica que el modelo responde mejor en condiciones topográficas simples, mientras que en sectores más complejos su capacidad predictiva disminuye. Este comportamiento nos sugiere la idea de que los modelos basados en Machine Learning capturan patrones generales de susceptibilidad, pero presentan limitaciones cuando la dinámica hidráulica detallada es determinante.

5. Conclusiones

El análisis comparativo entre los modelos de inundación desarrollados con XGBoost y HEC-RAS reveló diferencias metodológicas significativas en la predicción de áreas susceptibles a inundación en el río Jilusaya. La coincidencia entre ambos enfoques varió espacialmente: 86,55% en el Tramo 2 y 46,05% en el Tramo 1, con una concordancia global del 57,39% para el periodo de retorno de 50 años. El modelo de ML tendió a sobreestimar áreas de inundación en sectores topográficamente complejos, mientras que HEC-RAS proporcionó delimitaciones más precisas mediante la resolución detallada de ecuaciones hidrodinámicas. Esta variabilidad evidencia que, aunque el ML identifica patrones generales de susceptibilidad, su precisión se ve limitada por el procesamiento por bloques y restricciones computacionales.

Los datos hidrológicos generados revelaron caudales máximos de 4,1 m^3/s, 6,6 m^3/s y 8,1 m^3/s para periodos de retorno de 10, 30 y 50 años, con tiempo de concentración de 03:50 horas. El MDT de alta resolución (20×20 cm) con precisión de 14 cm proporcionó la base topográfica necesaria para la modelación hidráulica. Las métricas de validación del modelo de ML (MSE de 0,000485 y convergencia progresiva del RMSE) indicaron ajuste robusto sin sobreajuste significativo, aunque la variabilidad entre bloques señala necesidad de recalibración. La principal ventaja del enfoque de ML radica en su eficiencia temporal: 45-60 minutos versus 8 horas de HEC-RAS, posicionándolo como herramienta valiosa para evaluaciones preliminares rápidas.

Los resultados indican que ambas metodologías son complementarias: HEC-RAS mantiene su rol esencial para diseño de infraestructura y gestión detallada donde se requiere máxima precisión, mientras el ML ofrece soluciones ágiles para identificación preliminar de zonas susceptibles. No obstante, los modelos de ML requieren supervisión profesional y validación continua para garantizar resultados confiables. Futuras investigaciones deberían explorar técnicas de Deep Learning, validación cruzada espacial y métodos de procesamiento integrado que superen las limitaciones del enfoque por bloques, mejorando la capacidad de capturar dinámicas hidráulicas complejas sin sacrificar precisión física.

Referencias

Angulo, M., & Saavedra, O. (2018). *Modelación hidrológica de la cuenca Maylanco utilizando HEC-HMS*. Centro de Investigación en Ingeniería Civil y Ambiental (CIICA), Universidad Privada Bolivia.

ERBOL. (2024, marzo 10). *Desborde del río La Paz anega vías de acceso y aísla a Mallasa y Río Abajo*.

GGOS Focus Area. (2023). *Artificial intelligence for geodesy (AI4G)*. https://ggos.org/about/org/fa/ai-for-geodesy/

Gonzales-Inca, C., Calle, M., Croghan, D., Torabi Haghighi, A., Marttila, H., Silander, J., & Alho, P. (2022). Geospatial artificial intelligence (GeoAI) in the integrated hydrological and fluvial systems modeling: Review of current applications and trends. *Water*, 14(22), 2211. https://doi.org/10.3390/w14142211

Ibañez Apaza, L. (2022). *Balance hídrico de la cuenca del río La Paz hasta el punto de control estación hidrométrica de Aranjuez*. Universidad Mayor de San Andrés.

Kalteh, A. M. (2013). Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Computers & Geosciences*, 54, 1–8. https://doi.org/10.1016/j.cageo.2012.12.001

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, 10(11), 1536. https://doi.org/10.3390/w10111536

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

Pilotti, M., Milanesi, L., Bacchi, V., Tomirotti, M., & Maranzoni, A. (2020). Research on flood propagation for different dam failure modes: A case study in Shenzhen. *Frontiers in Earth Science*, 8, 527363. https://doi.org/10.3389/feart.2020.527363

Rincón, J., Pérez, M., Delfín, G., Freitez, C., & Martínez, F. (2017). Análisis comparativo entre los modelos HEC-RAS e IBER en la evaluación hidráulica de puentes. *Gaceta Técnica*, 17, 9–28.

Seckin, N., Cobaner, M., & Yurtal, R. (2013). Comparison of artificial neural network methods with L-moments for estimating flood flow at ungauged sites: The case of east Mediterranean river basin. *Water Resources Management*, 27, 2103–2124. https://doi.org/10.1007/s11269-013-0261-4

Shen, C., Chen, X., & Chen, Y. (2023). Deep learning in hydrology: Recent advances and future directions. *Journal of Hydrology*.

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G., Sermet, Y., & Demir, I. (2024). A comprehensive review of deep learning applications in flood modeling. *Journal of Hydrology*.

Teng, J., Jakeman, A., Vaze, J., Croke, B., Dutta, D., & Kim, S. (2020). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental Modelling & Software*, 90, 201–216. https://doi.org/10.1016/j.envsoft.2017.01.006