



Predicción de riesgos de inundación por medio de datos sintéticos de diferencias finitas y machine learning

Flood risk prediction using synthetic finite-difference data and machine learning

Jorge Luis González Figueroa¹ , Gerardo Tinoco Guerrero² , José Alberto Guzmán Torres² , Francisco Javier Domínguez Mota^{1,2} 

¹ Facultad de Ciencias Físico Matemáticas. Universidad Michoacana de San Nicolás de Hidalgo. Edificio B. Avenida Francisco J. Múgica S/N Ciudad Universitaria C.P. 58030, Morelia, Michoacán.

² Facultad de Ingeniería Civil. Universidad Michoacana de San Nicolás de Hidalgo. Edificio B. Avenida Francisco J. Múgica S/N Ciudad Universitaria C.P. 58030, Morelia, Michoacán.

Resumen

El presente proyecto desarrolla un modelo para clasificar el riesgo de inundación en áreas con topografía irregular mediante la integración de métodos numéricos y técnicas de aprendizaje automático. Inicialmente, se implementó un modelo de diferencias finitas para simular la propagación del agua mediante una variante de la ecuación de Saint-Venant. Posteriormente, se utilizó un modelo de Random Forest Classifier para clasificar el riesgo en función del nivel del agua simulado. El modelo se evaluó mediante las métricas f1-score, exactitud, precisión y la matriz de confusión. El trabajo contribuye al desarrollo de sistemas predictivos para la gestión de riesgos hídricos mediante modelos híbridos, lo que facilita la identificación temprana de áreas vulnerables a eventos hidrometeorológicos.

Palabras clave: diferencias finitas, clasificador random forest, gestión de inundaciones, ecuación de Saint-Venant, machine learning.

Abstract

The present study proposes a classifier model to assess flood risk in areas with irregular topography by integrating numerical methods and machine learning. Initially, a finite difference model is implemented to simulate water propagation using a simplified version of the Saint-Venant equations. Subsequently, a Random Forest Classifier is used to classify the risk as a function of the simulated water level. The model was evaluated using F1-score, confusion matrix, accuracy, and precision metrics. The study aims to contribute to the development of predictive systems for hydrological risk. This is achieved through hybrid models that facilitate early identification of vulnerabilities before hydrometeorological events.

Keywords: finite differences, random forest classifier, flood management, Saint-Venant, machine learning.

Introducción

La predicción de inundaciones es un enfoque que combina modelos físicos, estadísticos y de aprendizaje automático para predecir fenómenos hidrológicos. El desarrollo de algoritmos de predicción es clave para elaborar planes de protección cuya prioridad es salvaguardar a la población y, en la medida de lo posible, los bienes materiales (Tahmasebi *et al.*, 2020).

En México existen protocolos de prevención establecidos por autoridades federales y locales, como CONAGUA, SEMARNAT, Protección Civil, entre otras, encargadas de ejecutar acciones como la limpieza de drenajes y la promoción de campañas de concientización y cultura preventiva. Otras agencias, como SACMEX y EAS-SAME, operan sistemas de monitoreo y alerta temprana, es-

pecialmente en zonas urbanas como la Ciudad de México.

La respuesta a emergencias está a cargo de organismos como Protección Civil, la SEDENA y diversas agrupaciones estatales que realizan acciones clave como desazolves, evacuaciones y asistencia comunitaria mediante la habilitación y gestión de albergues y refugios temporales

Historial

Manuscrito recibido: 31 de octubre de 2025

Manuscrito aceptado: 25 de abril de 2026

Manuscrito publicado: agosto de 2026

✉ Autor para correspondencia

Jorge Luis González Figueroa

1718717h@umich.mx

Como citar

González Figueroa JL, Tinoco Guerrero G, Guzmán Torres JA, Domínguez Mota FJ (2026). Predicción de riesgos de inundación por medio de datos sintéticos de diferencias finitas y machine learning. *Ciencia Nicolaita* 97:181-187. doi:[10.35830/cn.vi97.951](https://doi.org/10.35830/cn.vi97.951)

(CONAGUA y SEMARNAT, 2011; Secretaría de Salud/STCONAPRA, 2017).

Todas estas medidas son cruciales para la atención oportuna de fenómenos meteorológicos extremos y reflejan el resultado de décadas de experiencia y de desarrollo científico. No obstante, a pesar de contar con recursos amplios y planes de acción técnicamente eficientes, la actual crisis ambiental exige el desarrollo de nuevas metodologías y el aprovechamiento de técnicas emergentes en el análisis de datos y la predicción de riesgos, que permitan reforzar la capacidad adaptativa del país ante escenarios climáticos cada vez más complejos (Mosavi *et al.*, 2018).

En este contexto, el propósito de este trabajo fue generar un conjunto de datos sintético mediante simulaciones multiparamétricas que representan distintos escenarios de inundación en un entorno controlado. Para ello, se recurrió a la modificación sistemática de variables físicas relevantes (como la intensidad de la lluvia, la duración del evento y los terrenos irregulares) en un modelo hidrológico unidimensional, con el fin de obtener múltiples configuraciones posibles y registrar su comportamiento.

Este conjunto de datos sirvió de base para el entrenamiento de un modelo de clasificación mediante Random Forest, una técnica de aprendizaje automático no paramétrica que permite identificar patrones complejos y realizar predicciones precisas, incluso con conjuntos de datos heterogéneos. Esta metodología se planteó debido a su capacidad para manejar datos no lineales, su robustez frente al sobreajuste y su facilidad de interpretación en comparación con modelos más complejos, como redes neuronales profundas. Además, el uso de un dataset sintético permite controlar con precisión las condiciones del experimento y evaluar el desempeño del algoritmo en una variedad de escenarios representativos, sin depender exclusivamente de datos empíricos, que suelen ser escasos o incompletos. Este trabajo constituye una primera aproximación para la elaboración de modelos más complejos, contrastados con datos empíricos en trabajos futuros.

Materiales y métodos

Para la generación de datos, se implementó un modelo hidrodinámico unidimensional basado en las ecuaciones de aguas someras. Estas ecuaciones, también conocidas como ecuaciones de Saint-Venant, son simplificaciones de las ecuaciones de Navier-Stokes utilizadas cuando la profundidad de las aguas es muy pequeña en comparación con la extensión superficial del cuerpo acuífero, como en lagos, ríos y canales de agua. El modelo simula el flujo superficial en un canal abierto bajo condiciones controladas.

El modelo fue resuelto mediante un esquema de diferencias finitas hacia adelante en el tiempo y centradas en el espacio (FTCS), combinado con un modelo Lax-Friedrichs. (Celia y Gray, 1992).

Modelo matemático y discretización

Se definió un dominio de una dimensión espacial que representa un canal simplificado donde se aplica el siguiente modelo de la ecuación de Saint-Venant (Tinoco Guerrero *et al.*, 2020):

$$\frac{\partial h}{\partial t} + \frac{\partial q}{\partial x} = R(x, t) \quad (1)$$

$$\frac{\partial q}{\partial t} + gh \frac{\partial h}{\partial x} = -gh \frac{\partial z}{\partial x} \quad (2)$$

Donde:

- h es la altura en (m) del nivel del agua.
- q es el caudal en m^2/s .
- $R(x, t)$ es el término de fuente en m/s .
- g es la aceleración gravitacional m/s^2 .
- z es la altura del terreno en (m^2).

Se utilizó un esquema de diferencias finitas hacia adelante en el tiempo y centradas en el espacio, combinado con un esquema de Lax-Friedrichs para lograr estabilidad. Se muestra en las siguientes ecuaciones:

$$h_i^{k+1} = \frac{1}{2} (h_{i+1}^k + h_{i-1}^k) - \frac{\Delta t}{2\Delta x} (q_{i+1}^k - q_{i-1}^k) + R_i^k \Delta t \quad (3)$$

$$q_i^{k+1} = \frac{1}{2} (q_{i+1}^k + q_{i-1}^k) - gh_i^k \frac{\Delta t}{2\Delta x} (h_{i+1}^k - h_{i-1}^k) - gh_i^k \frac{\Delta t}{2\Delta x} (z_{i+1}^k - z_{i-1}^k) \quad (4)$$

donde:

- k es el índice para la variable temporal t .
- i es el índice para la variable espacial x .
- Δx es el intervalo espacial.
- Δt es el intervalo temporal.

Para poder generar un conjunto de datos que refleje una variación para su clasificación se realizaron las simulaciones con diversos parámetros variando las condiciones iniciales, las condiciones de frontera, condiciones de lluvia

Condiciones iniciales

- Condición de reposo:

$$h_i = 5.0(m)$$

$$q_i = A_i(m^2/s)$$

- Condición pulso en h:

$$h_i = 5.0 + A_i * \exp\left(-0.5 * \left(\frac{x-100}{10}\right)^2\right)(m) \quad (6)$$

$$q_i = A_i(m^2/s)$$

- Condición pulso en q:

$$q_i = A_i * \exp\left(-0.5 * \left(\frac{x-100}{10}\right)^2\right)(m^2/s) \quad (7)$$

Condiciones de frontera

- Condición libre:

Las condiciones no están sujetas a un valor fijo en la frontera.

- Condición pulso:

$$q = A_f * \sin\left(\frac{2\pi t}{T}\right) \quad (8)$$

- Condición crecida:

$$q = \min(A_f, 0.1 * t) \quad (9)$$

- Condición constante:

$$q = A_f \quad (10)$$

Condiciones de lluvia

- Lluvia constante:

$$R(x, t) = intensidad \quad (11)$$

- Lluvia temporal:

$$R(x, (t > 10) \& (t < 100)) = intensidad \quad (12)$$

- Lluvia espacio temporal

$$R\left(\begin{matrix} (x > L/2) \& (x < 2 * L/3), \\ (t > 10) \& (t < 100) \end{matrix}\right) = intensidad \quad (13)$$

- No lluvia

Número de condición CFL

El número de Courant-Friedrichs-Lewy (CFL) que asegura la estabilidad del algoritmo está dado por:

$$CFL = \sqrt{g} \frac{\Delta t}{\Delta x} \leq 0.5 \quad (14)$$

Generación del dataset y estadísticas

La siguiente tabla muestra los parámetros que se utilizarán tanto para variables categóricas como para variables numéricas. El dataset contiene un total de 4815 entradas correspondientes a una combinación única de parámetros (Tabla 1).

La clasificación de los datos obtenidos mediante la simulación muestra una tendencia de los datos hacia el riesgo bajo y decreciente hacia el riesgo alto. La distribución de clases presentada en la **Figura 1** revela un predominio de escenarios de riesgo ‘nulo’ y ‘bajo’. Desde la perspectiva de modelado, este comportamiento es consistente con la naturaleza de las simulaciones, donde solo combinaciones extremas de intensidad de lluvia y pendientes pronunciadas logran superar los umbrales de desbordamiento. Esta asimetría en los datos es un reflejo de la realidad hidrológica.

Tabla 1. Parámetros de simulación.

C.I	{“reposo”, “pulso en h”, “pulso en q”}
C.F	{“libre”, “pulso”, “crecida”, “flujo constante”}
C. de lluvia	{“constante”, “espacial”, “temporal”, “espacio-temporal”}
A_i	{0.1, 0.01, 0.001}
A_f	{0.01, 0.001, 0.0001}
X_c	{100, 200, 300}
Intensidad	{0.05, 0.03, 0.01}

gica, donde los eventos catastróficos son menos frecuentes pero más críticos para el entrenamiento del clasificador.

La **figura 2. a)** Muestra que la mayor distribución del valor de la altura máxima se observa para la clase de riesgo “Alto”, mientras que el riesgo “Nulo” presenta la menor variación. Se observa que la clase de riesgo “Alto” no solo presenta las alturas máximas más elevadas, sino también la mayor variabilidad. Esto sugiere que, en niveles críticos, otros factores como la topografía irregular comienzan a jugar un papel determinante en la magnitud de la inundación. Para la **figura 2. b)** Se observa que la intensidad de la lluvia está más relacionada con las simulaciones clasificadas como riesgo “Alto” y se confirma que la intensidad de valores atípicos en riesgos bajos indica que una lluvia intensa no siempre deriva en peligro si las condiciones de frontera o el terreno facilitan el drenaje rápido del flujo.

La **figura 3** muestra la distribución de la altura máxima del nivel del agua, agrupada por clases de riesgo. Se observa que el riesgo “Nulo” presenta una distribución limitada y una fuerte asimetría a la derecha. Esto refleja escenarios sin riesgo significativo de inundación. La clase “Bajo” presenta una menor densidad y una mayor dispersión. La clase “Medio” se extiende a una mayor distribución, reflejando condiciones más extremas de desborde y de elevación del nivel del agua. La **figura 4** presenta una matriz de correlación de Pearson, que muestra que la correlación lineal entre la intensidad y las alturas máximas es el único parámetro que influye en el nivel de riesgo.

Resultados

En esta sección se presentan los resultados obtenidos tras el desarrollo y la evaluación del modelo de clasificación del riesgo de inundación, basado en datos sintéticos generados mediante simulaciones multiparamétricas.

En general, los objetivos del estudio se alcanzaron con éxito. Se logró construir un conjunto de datos representativo con distintas combinaciones de variables físicas influyentes en eventos meteorológicos y se entrenó un modelo de aprendizaje automático robusto que alcanzó niveles de desempeño satisfactorios. La meteorología permitió evaluar la relación entre las características físicas del terreno y los eventos de lluvia, así como el riesgo resultante, y demostró la utilidad de los modelos de clasificación como herramientas auxiliares para la predicción temprana de inundaciones.

Además, el uso de un algoritmo Random Forest facilitó no solo una predicción precisa, sino también la interpretación de la importancia relativa de cada variable de entrada. La validación cruzada y el ajuste de hiperparámetros

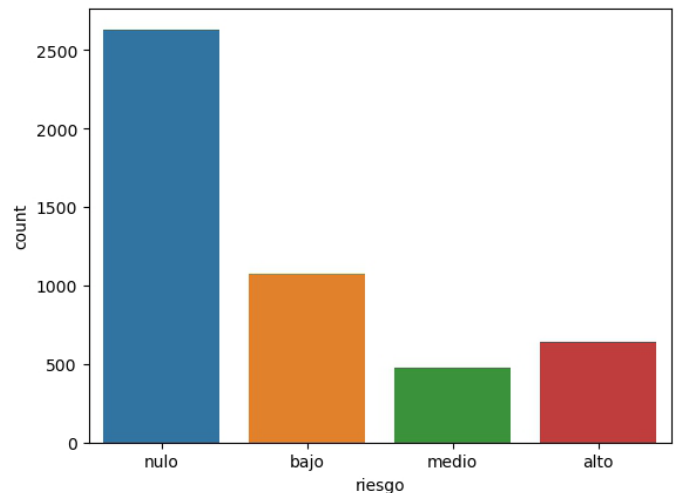


Figura 1. Distribución de las clases de riesgo en las simulaciones.

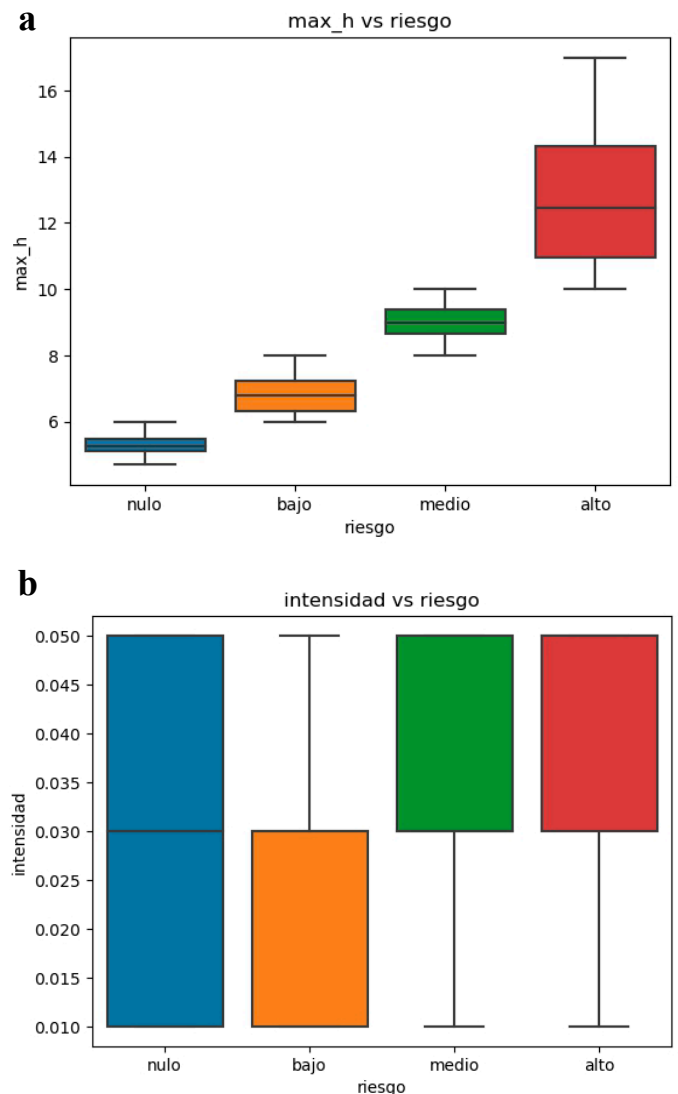


Figura 2. a) Distribución de las alturas máximas según la clasificación de riesgos. **b)** distribución de la intensidad de la lluvia según la clasificación de riesgo.

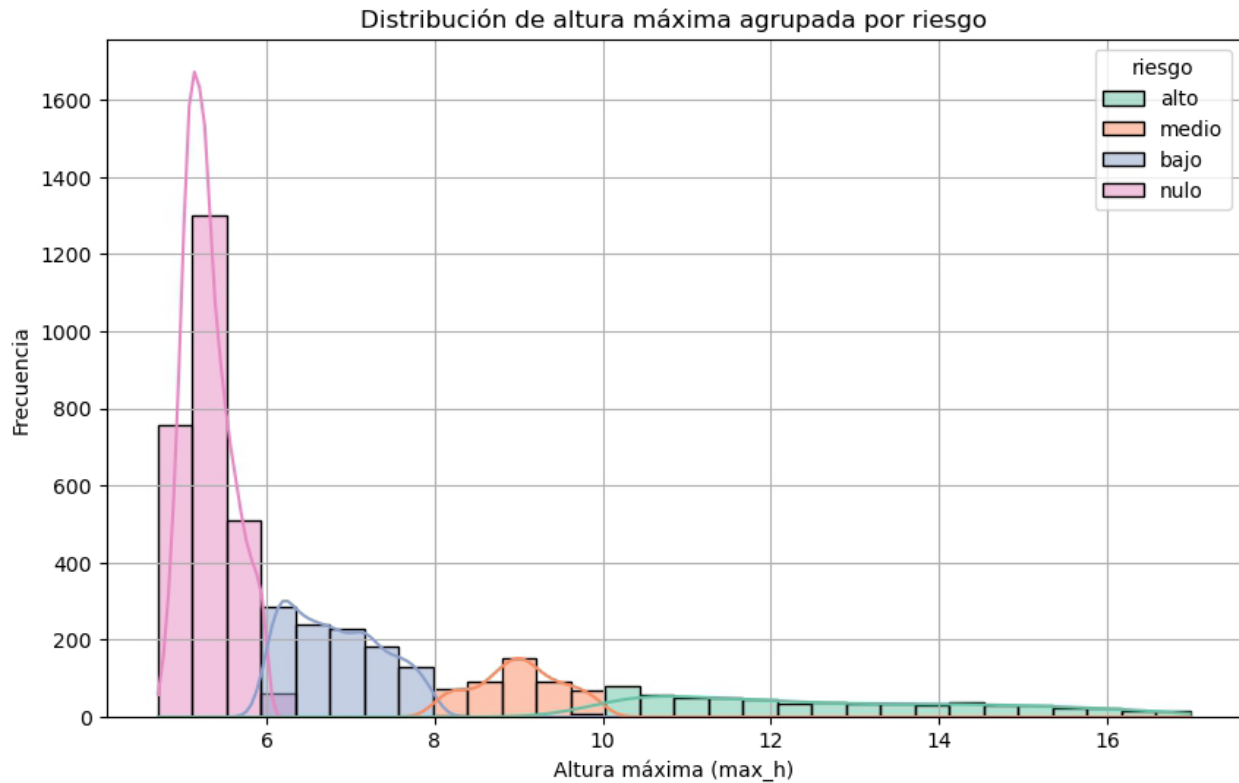


Figura 3. Histograma de la distribución de alturas máximas.

mediante búsqueda en la malla confirmaron la estabilidad y consistencia del modelo.

El modelo entrenado obtuvo un accuracy general del 99%, lo que indica un desempeño altamente satisfactorio. La matriz de clasificación muestra valores equilibrados de precisión, recall y f1-score para las cuatro clases analizadas (Alto, Medio, Bajo y Nulo), sin evidencia de

un desbalance severo. La clase “Nulo”, que representa eventos sin condiciones de inundación, fue correctamente clasificada por el modelo.

Estos resultados demuestran la capacidad del modelo para diferenciar eficazmente entre distintos niveles de riesgo, incluso en presencia de múltiples variables de entrada (**Tabla 2**). La ligera caída de rendimiento en las clases “Medio” puede atribuirse a una menor representación relativa en el conjunto de datos o a una mayor complejidad en su delimitación en comparación con las clases vecinas.

El proceso de ajuste de hiperparámetros se realizó mediante una búsqueda en malla (GridSearchCV) con validación cruzada de 5 pliegues, donde se exploraron combinaciones de parámetros como la profundidad

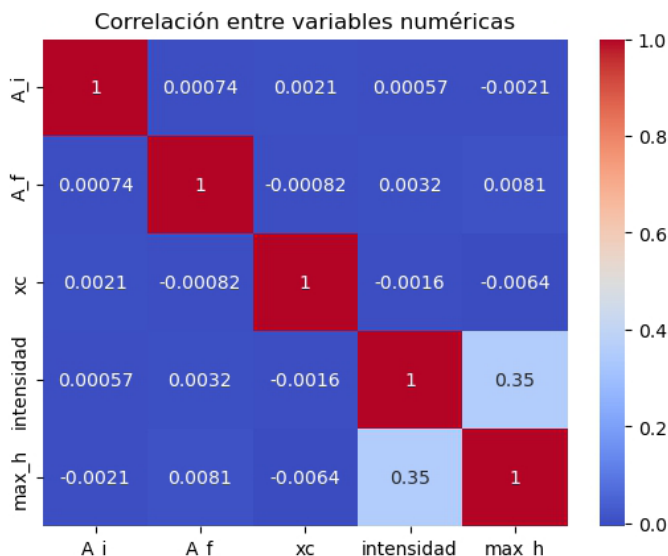


Figura 4. Matriz de correlación de Pearson.

Tabla 2. Resultados del modelo de clasificación.

	Precisión	Recall	f1-score	Support
Alto	0.99	0.99	0.99	118
Medio	0.99	0.96	0.98	106
Bajo	0.99	0.99	0.99	228
Nulo	1.00	1.00	1.00	511
Accuracy			0.99	963
Macro avg	0.99	0.99	0.99	963
Weighted avg	0.99	0.99	0.99	963

máxima del árbol (`max_depth`), el número de árboles (`n_estimators`), el número mínimo de muestras por hoja (`min_samples_leaf`) y el subconjunto de variables utilizadas en cada división (`max_features`).

El conjunto de hiperparámetros que mejor desempeño tuvo en el modelo fue el siguiente:

```
{<max_depth>: 10, <max_features>: <sqrt>, <min_samples_leaf>: 1, <min_samples_split>: 10, <n_estimators>: 200}
```

Este conjunto obtuvo la mejor puntuación en validación cruzada: 0.9929900449744806, lo que confirma la estabilidad del modelo y evita el sobreajuste.

Discusión

Los resultados obtenidos demuestran que es posible utilizar un conjunto de datos sintético, generado mediante simulaciones multiparamétricas con esquemas de diferencias finitas, para entrenar el modelo de aprendizaje automático con alta capacidad de clasificación del riesgo de inundación. La precisión alcanzada por el clasificador Random Forest, con una exactitud global del 99%, evidencia la consistencia interna del dataset y la eficacia del modelo para detectar patrones complejos entre variables hidrológicas básicas y los niveles de riesgo asociados.

Si bien el uso de un dataset sintético permitió un control riguroso de las variables y superó la escasez de registros históricos, es imperativo reconocer que estos datos son una simplificación de la física real. Al ser generados bajo un esquema de diferencias finitas unidimensional, el modelo podría no capturar fenómenos complejos como turbulencias, infiltración del suelo. Por tanto, una precisión de 99% debe interpretarse como la eficacia del Random Forest Classifier para aprender el modelo subyacente, y no necesariamente como su precisión final en un entorno hidrológico real sin un proceso previo de calibración con datos empíricos.

Este desempeño es especialmente relevante si se considera que el modelo fue entrenado con datos simulados, lo que valida el uso de datasets artificiales como una alternativa viable ante la escasez o la dificultad de recopilar datos reales de eventos extremos. La flexibilidad del enfoque permite diseñar escenarios específicos que, de otro modo, podrían no estar bien representados en los registros históricos.

En cuanto al rendimiento por clase, se observa que la categoría “Nulo” se clasificó correctamente, lo que sugiere que los casos sin condiciones de riesgo presentan características suficientemente distintivas. En cambio, la clase

“Medio” mostró una leve disminución en el valor de recall, lo que puede deberse a la cercanía de sus valores a los riesgos “Bajo” y “Alto”. Esto sugiere la posibilidad de ajustar los umbrales de clasificación o aumentar el número de muestras en esa categoría para mejorar el equilibrio del modelo.

El análisis de importancia de variables (no mostrado aquí, pero calculado durante el proceso) reveló que la variable con mayor peso predictivo fue la intensidad de la lluvia. La generación de terrenos aleatorios ofreció una mayor variabilidad en las simulaciones al representar mejor una situación real.

La metodología aquí presentada no pretende sustituir las mediciones de campo, sino actuar como un pre-entrenador robusto que reduce la necesidad de grandes volúmenes de datos reales, los cuales son costosos y difíciles de obtener en eventos extremos. En este sentido la metodología ofrece una alternativa sencilla, rápida y computacionalmente eficiente, especialmente útil en contextos de planeación preliminar, entrenamiento de sistemas de alerta temprana o como apoyo educativo y técnico.

Conclusiones

Los resultados de este trabajo muestran que la implementación de un algoritmo de predicción del riesgo de inundación basado en Machine Learning, alimentado con datos generados mediante simulaciones con métodos de Diferencias finitas, es una estrategia viable y efectiva como aproximación inicial a escenarios hidrológicos más complejos.

El modelo desarrollado, aunque idealizado, ofrece una certidumbre satisfactoria al clasificar correctamente distintos niveles de riesgo, lo que valida su utilidad como herramienta exploratoria. A pesar de su simplicidad, permite identificar las variables físicas más influyentes en la dinámica de fluidos en canales abiertos o ríos, brindando una base sólida para el desarrollo de modelos más detallados y adaptables a datos reales.

Esta metodología representa una alternativa complementaria a los enfoques tradicionales, con el potencial de integrarse en los protocolos de prevención, detección temprana y actuación ante eventos de inundación. Su aplicación práctica puede extenderse a sistemas de alerta, planes de contingencia o escenarios de evaluación rápida, especialmente en contextos donde la recolección de datos empíricos es limitada.

Finalmente, este trabajo subraya la importancia de seguir desarrollando algoritmos híbridos que combinen conocimiento físico con técnicas de inteligencia artificial,

fortaleciendo así la capacidad de respuesta ante riesgos hidrometeorológicos crecientes en un contexto de cambio climático.

Agradecimientos

Los autores del presente trabajo agradecen especialmente a la Universidad Michoacana de San Nicolás de Hidalgo, a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) y a SIIIA MATH.

Referencias

Celia MA, Gray WG (1992). *Numerical methods for differential equations: fundamental concepts for scientific and engineering applications*. Prentice-Hall International, 436 pp.

Comisión Nacional del Agua (CONAGUA) y Secretaría del Medio Ambiente y Recursos Naturales (SEMARNAT) (2011). *Manual para el control de inundaciones, Sistema Nacional de Información Estadística y Geográfica*, Ciudad de México, febrero 2011.

Mosavi A, Öztürk P, Chau K-W (2018). Flood prediction using machine learning models: literature review. *Water* 10(11):1536. doi:[10.3390/w10111536](https://doi.org/10.3390/w10111536)

Secretaría de Salud / STCONAPRA. (2017). *Manual de protección civil ante casos de emergencia, contingencias y desastres del STCONAPRA*. Primera edición 2017, Secretaría de Salud, Ciudad de México.

Tahmasebi P, Kamrava S, Bai T, Sahimi M (2020) Machine learning in geo- and environmental sciences: from small to large scale. *Advances in Water Resources* 142:103619. doi:[10.1016/j.advwatres.2020.103619](https://doi.org/10.1016/j.advwatres.2020.103619)

Tinoco Guerrero G, Domínguez Mota FJ, Tinoco Ruiz JG, Lucas Martínez JS, Tinoco Guerrero NS, Leppäranta M, Mammarella I (2020). A generalized finite-difference-volume hybrid method applied to shallow-water equations. *Revista Mexicana de Métodos Numéricos* 4(2):1-35.