

## Estimación de caudales en una cuenca tropical mediante modelación neuronal artificial basada en memoria hidrológica

### Estimating flow rates in a tropical basin using artificial neural modeling based on hydrological memory

León Darío Blandón-López<sup>1</sup>   Luis Octavio González-Salcedo<sup>2</sup> 

<sup>1</sup>Corporación Autónoma Regional del Valle del Cauca, Palmira Colombia.

<sup>2</sup>Departamento de Ingeniería, Facultad de Ingeniería y Administración, Universidad Nacional de Colombia.

## Abstract

**Introducción:** Los sistemas hidrológicos han sido severamente impactados por el aumento de eventos extremos como inundaciones y sequías, a su vez intensificados por el cambio climático. La cuenca del río Tuluá, ubicada en el Valle del Cauca – Colombia, enfrenta retos adicionales por la deforestación y limitada cobertura de monitoreo, complicando la gestión eficiente del recurso hídrico.

**Objetivos:** Desarrollar un modelo neuronal basado en redes LSTM para predecir el caudal del río Tuluá a partir de variables hidrometeorológicas, para la anticipación de escenarios extremos y el apoyo en la toma de decisiones en la gestión hídrica regional.

**Metodología:** 6875 registros diarios para cinco variables (precipitación, temperatura, evaporación, humedad relativa y caudal) de siete estaciones, fueron consolidadas. Para completar valores faltantes se aplicó preprocesamiento de datos usando KNN y análisis de correlación de Pearson. Se configuró y se entrenó el modelo LSTM, el cual se evaluó para los subconjuntos de entrenamiento, prueba y validación mediante métricas usuales para redes neuronales.

**Resultados:** Una alta capacidad predictiva con coeficientes de correlación R (0,75 – 0,98) mostró el modelo neuronal. Se evidenció que el modelo captura efectivamente las dependencias temporales y patrones no lineales, además de identificar la influencia del fenómeno ENSO en la dinámica hidrológica local.

**Conclusiones:** El modelo LSTM es una herramienta efectiva en la estimación de caudales para la cuenca en estudio, lo cual facilita tanto la gestión hídrica como la prevención de los riesgos que se asocian a la variabilidad climática. Futuros desarrollos pueden potenciar su precisión y utilidad, si consideran la inclusión de índices climático globales.

**Palabras clave:** Inteligencia artificial, cambio climático, evapotranspiración, inundación, precipitación, gestión de recursos, cuenca fluvial, modelo de simulación, desarrollo sostenible, temperatura.

## Resumen

**Introduction:** Hydrological systems have been severely impacted by the increase in extreme events such as floods and droughts, intensified by climate change. The Tuluá River basin, located in the Cauca Valley of Colombia, faces additional challenges due to deforestation and limited monitoring coverage, complicating efficient water resource management.

**Objectives:** Develop a neural model based on LSTM networks to predict the flow rate of the Tuluá River from hydrometeorological variables, for the anticipation of extreme scenarios and support in decision-making in regional water management.

**Methodology:** 6875 daily records for five variables (precipitation, temperature, evaporation, relative humidity, and streamflow) from seven stations were consolidated. To fill missing values, data preprocessing was applied using KNN and Pearson correlation analysis. The LSTM model was configured and trained, and evaluated for the training, test, and validation subsets using standard neural network metrics.

**Results:** The neural model demonstrated high predictive capacity with correlation coefficients R (0.75–0.98). It was evident that the model effectively captures temporal dependencies and nonlinear patterns, in addition to identifying the influence of the ENSO phenomenon on local hydrological dynamics.

**Conclusions:** The LSTM model is an effective tool for estimating flows for the basin under study, facilitating both water management and the prevention of risks associated with climate variability. Future developments can enhance its accuracy and usefulness if they include global climate indices.

**Keywords:** Artificial intelligence, climate change, evapotranspiration, flooding, precipitation, resource management, river basin, simulation model, sustainable development, temperature.

### ¿Cómo citar?

Blandón-López LD, González-Salcedo LO. Estimación de caudales en una cuenca tropical mediante modelación neuronal artificial basada en memoria hidrológica. Ingeniería y Competitividad, 2026, 28(1)e-20415199

<https://doi.org/10.25100/iyv.v28i1.15199>

Recibido: 31/08/25

Revisado: 05/11/25

Aceptado: 21/01/26

Online: 18/02/26

### Correspondencia

lblandon@unal.edu.co



### ¿Por qué se realizó?

Se desarrolló un modelo predictivo confiable, basado en modelado de redes neuronales artificiales, para estimar el caudal del río Tuluá, cuya cuenca es vulnerable a la variabilidad hidrológica y al cambio climático, y además sufre limitaciones de datos y presiones ambientales. Su objetivo es mejorar la anticipación de eventos extremos y fortalecer la gestión hídrica regional mediante el uso de técnicas de inteligencia artificial.

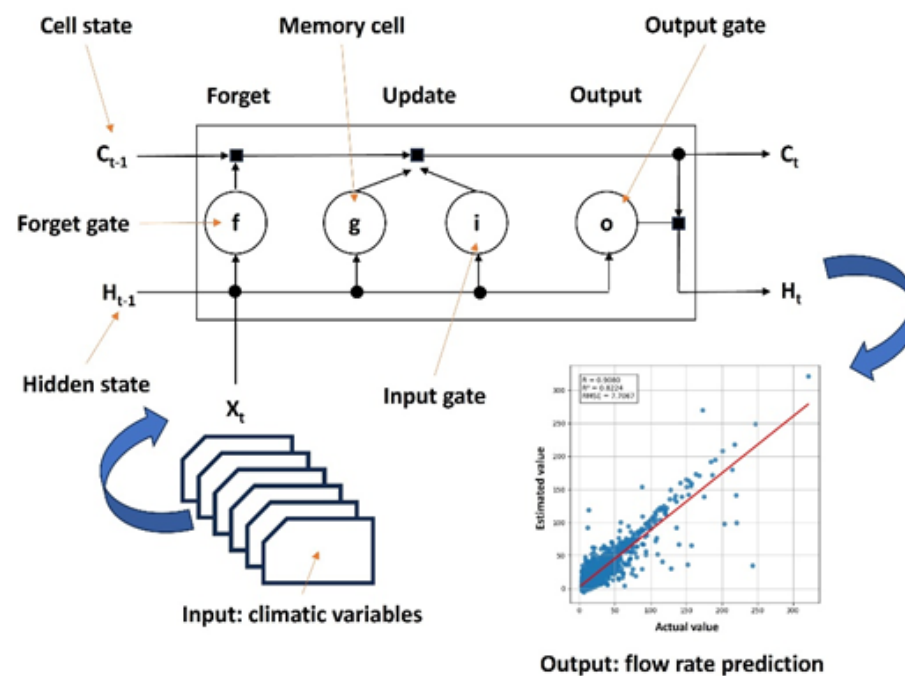
### ¿Cuáles fueron los resultados más relevantes?

El modelo neuronal artificial mostró una alta precisión en la estimación del caudal según las métricas de evaluación ( $R = 0,75 - 0,98$ ), lo que demuestra la captura de patrones no lineales y las complejas dependencias temporales de los datos hidroclimáticos, confirmando la influencia significativa del fenómeno ENSO en la dinámica del caudal.

### ¿Qué aportan?

Una herramienta robusta y adaptable para la predicción hidrológica en zonas con datos limitados y alta variabilidad climática.

## Resumen gráfico



## Introducción

Un aumento significativo, a nivel global, en la frecuencia e impacto de eventos extremos como inundaciones y sequías ha sido causado por la variabilidad hidrológica, que a su vez se intensifica por el cambio climático y las actividades humanas (1, 2). Aproximadamente el 48% de los desastres naturales registrados durante el periodo 1995–2015 estuvieron relacionados con fenómenos hidrometeorológicos (3, 4), afectando a miles de millones de personas y causando pérdidas económicas significativas (5). El crecimiento demográfico y la expansión urbana no planificada, que ejercen presión sobre los recursos hídricos globales, agravan esta situación (6, 7).

Los sistemas hidrológicos de América Latina han sufrido graves consecuencias debido al aumento de la temperatura media, el retroceso glaciar y la alteración de los patrones de precipitación (8, 9). Asimismo, fenómenos como El Niño y La Niña afectan la disponibilidad de agua, en el caso de Colombia, produciendo fluctuaciones significativas en los caudales fluviales (10). Según Ávila et al., 2019 (11), en algunos lechos de ríos pueden producirse reducciones superiores al 75% debido a estas variaciones, demostrando así que los sistemas hídricos son vulnerables a cambios climáticos y antropogénicos (12, 13).

El Valle del Cauca se caracteriza por ser una región de alta diversidad hidroclimática (14, 15); sin embargo, sus cuencas enfrentan una presión creciente, agravada aún más por la deforestación, el cambio en el uso del suelo y la limitada cobertura de las estaciones de vigilancia. Un ejemplo representativo donde la falta de datos históricos y la escasez de análisis de series temporales dificultan la gestión eficiente de los recursos hídricos es la cuenca del río Tuluá.

Esta zona, de vital importancia para el centro del departamento del Valle del Cauca, con características de bosque andino y subandino, proporciona agua para la agricultura, el consumo y la ganadería, y actualmente enfrenta desafíos relacionados con la calidad del agua debido a la contaminación y la deforestación, lo que lleva a que se le centre el interés desde una gestión ambiental que preserve su función reguladora y ecológica.

Esto da lugar a la necesidad de utilizar herramientas analíticas avanzadas que permitan identificar patrones no lineales en dinámica hidrológica. Las redes neuronales recurrentes (RNN), y en particular las redes de Memoria a Corto Plazo (LSTM), están especializadas para entrenar datos de series temporales (16). En el caso de Colombia, se han utilizado para predecir los niveles de agua en el río Atrato (17), específicamente para predecir caudales en la estación de Calamar en el río Magdalena (18). El rendimiento se evaluó utilizando una estadística de error, y los resultados mostraron que el modelo desarrollado no captura adecuadamente la dinámica hidrológica real de la cuenca estudiada. Por ello, existe un estudio limitado sobre el uso de estos modelos en Colombia, y particularmente en cuencas del Valle del Cauca.

A partir de estas premisas, se propone el desarrollo de un modelo de red neuronal artificial utilizando redes de Memoria a Corto Plazo (LSTM) para estimar el caudal del río Tuluá utilizando variables hidrometeorológicas registradas por las estaciones locales. La herramienta propuesta permitirá identificar patrones sutiles, anticipar escenarios extremos y generar insumos útiles para

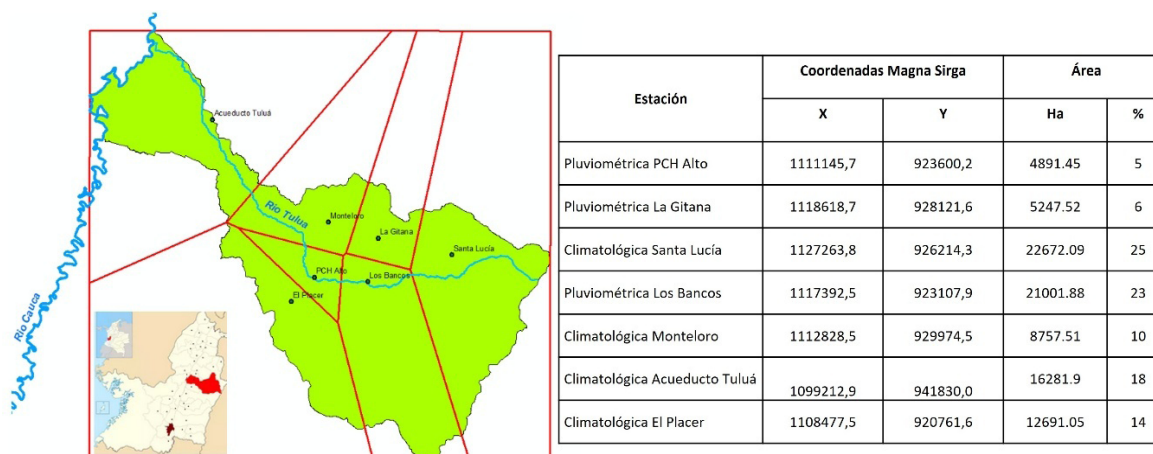
la toma de decisiones en la gestión pública del medio ambiente. De este modo, la integración de la inteligencia artificial con el análisis hidrológico (19, 20) busca cerrar brechas de información, fortalecer la resiliencia hídrica y mejorar la capacidad de respuesta a eventos extremos en esta región del suroeste de Colombia.

## Materiales y métodos

### Área de estudio y datos

El área objetivo de este estudio es la cuenca hidrográfica del río Tuluá (Figura 1). Nace en el Parque Nacional Las Hermosas, a una altitud de 4100 metros sobre el nivel del mar, y su cauce principal fluye durante 72 km a través del municipio homónimo hacia el río Cauca, a 900 metros sobre el nivel del mar, recibiendo afluentes como los ríos Cofre, Loro, San Antonio, San Marcos y Nogales, entre otros (35). La cuenca cubre una superficie de 91.485–91.495 hectáreas e incluye bosques andinos y subandinos con una biodiversidad significativa. Los ecosistemas del bosque andino contribuyen a la regulación del agua, mientras que los ecosistemas subandinos apoyan la ganadería extensiva. El uso del suelo determinado por la actividad humana consiste en la agricultura y la ganadería, así como en los asentamientos de población.

Su importancia radica en su papel como fuente de agua para uso doméstico, riego agrícola, ganadería y actividades industriales. Asimismo, para el municipio de Tuluá, actúa como un conector espacial y ambiental, integrando áreas rurales y urbanas (36). Actualmente, enfrenta problemas medioambientales debido a la contaminación en los tramos bajos causada por vertidos de aguas residuales y alteraciones en el lecho del río, así como a la deforestación por la sustitución de los bosques nativos por pastos y cultivos (35). Para esta cuenca (Figura 1), la Corporación Autónoma Regional del Valle del Cauca – CVC opera siete estaciones de monitoreo hidroclimático (Acueducto Tuluá – AT, Monteloro – ML, El Placer – EP, Santa Lucía – SL, La Gitana – LG, Los Bancos – LB y Mateguadua – MG), de donde se recogían los registros diarios, y cuya información en todas las estaciones mostraba datos faltantes para todas las estaciones y para todas las variables.



**Figura 1.** Área de estudio en la cuenca del río Tuluá y ubicación de las estaciones hidroclimáticas.

En cuanto a las variables, los datos faltantes correspondían al 5,6–11,1% para la evaporación diaria, 7,3% para el caudal, 8,6–61,4% para la humedad relativa, 0,5–3,9% para la precipitación y 7,1–58,0% para la temperatura. Monteloro fue identificada como la estación más crítica en esta situación, con los valores de datos más altos que faltaban para la mayoría de las variables consideradas.

Los datos comprenden cinco variables hidroclimáticas principales (precipitación, temperatura, evaporación, humedad relativa y caudal), Tabla 1, que suman 6785 registros diarios tras el procedimiento de preprocesamiento de datos, que implicó estructurar la base de datos, limpiarla y completarla con correlaciones lineales y no lineales. La información se organizó en un archivo de 16 columnas, en la que la variable de salida es el caudal del río Tuluá, mientras que las demás corresponden a variables de entrada, y se utilizó tanto para análisis estadístico como para entrenar el modelo predictivo.

### Análisis estadístico preliminar

#### Estadísticas descriptivas

Se calcularon medidas de tendencia central (media  $\bar{M}_e$ , mediana  $\mu$ , modo  $M_o$ ), dispersión (varianza  $\sigma^2$ , desviación estándar  $\sigma$ ) y rangos de cuartiles ( $Q_1$ ,  $Q_2$ ), según se definieron en Abubaker, 2024 (21), para cada variable y estación seleccionada (Tabla 1, Figura 2). Esto permitió identificar patrones característicos, incluyendo una mayor variabilidad en la estación Acueducto Tuluá. En cuanto al caudal, los registros se concentraron en la estación de Mateguadua, estratégicamente situada entre las zonas de producción y consumo de agua. Para la precipitación, se consideraron las estaciones Acueducto Tuluá, El Placer, La Gitana, Los Bancos, Montero y Santa Lucía. Para la evaporación, se incluyeron las estaciones Acueducto Tuluá, El Placer, Monteloro y Santa Lucía. Finalmente, para la temperatura, se utilizaron las estaciones Acueducto Tuluá, El Placer, Monteloro y Santa Lucía.

#### Prueba de normalidad

La prueba de normalidad de Shapiro-Wilk se aplicó a cada variable en cada estación (Tabla 2), con la estadística  $W$  y el valor  $P$ ; para la prueba,  $W \sim 1,0$  indica distribución normal y  $P < 0,05$  sugiere rechazo de la normalidad (22).

#### Análisis de tendencias

Se aplicó la prueba de Mann-Kendall no paramétrica (Tabla 2), para detectar tendencias significativas en la serie temporal y usando el coeficiente Tau de Kendall  $\tau$  y el valor  $P$ ; para la prueba,  $-1 < \tau < 1$  indica dirección y intensidad de la tendencia, mientras que  $P < 0,05$  indica una tendencia estadísticamente significativa (23).

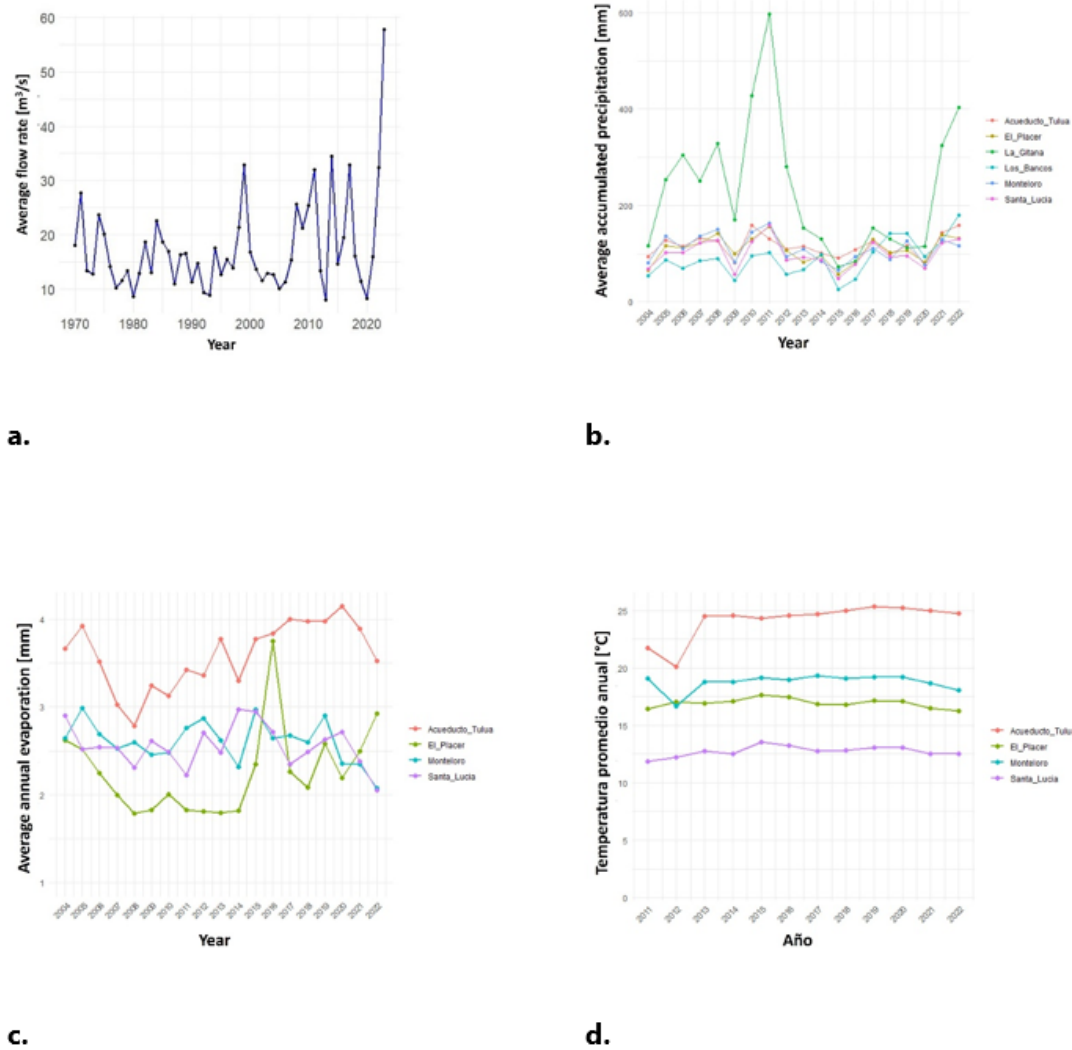
**Tabla 1.** Resumen de estadísticas descriptivas de variables hidroclimáticas por estación.

Caudal									
Estación	Mínimo	Q1	Yo	m	Mo	P3	Máximo	S	S2
	m3/s	m3/s	m3/s	m3/s	m3/s	m3/s	m3/s	m3/s	
MG	-	0,93	8,3	13,3	8,0	20,9	321,16	15,06	226,83
Precipitaciones									
Estación	Mínimo	Q1	Yo	m	Mo	P3	Máximo	S	S2
	mm	mm	mm	mm	mm	mm	mm	mm	
EN	4,0	64,0	103,0	116,2	113,0	157,0	408,0	69,51	4831,54
EP	0,0	55,0	98,0	109,5	42,0	148,0	323,0	70,81	5014,13
LG	0,0	62,0	150,0	235,7	8,0	336,0	1643,0	252,79	252,79
LB	0,0	34,0	66,0	88,88	42,0	119,0	550,0	82,25	42,0
ML	0,0	54,0	101,3	113,3	44,0	156,6	344,0	72,28	5224,15
SL	0,0	44,0	89,0	100,3	106,0	139,0	323,0	70,46	4965,2
Evaporación									
Estación	Mínimo	Q1	Yo	m	Mo	P3	Máximo	S	S2
	mm	mm	mm	mm	mm	mm	mm	mm	
EN	2,2	3,2	3,6	3,58	3,4	4,0	5,5	0,597	0,357
EP	1,0	1,7	2,1	2,22	1,7	2,5	8,1	0,790	0,624
ML	1,4	2,3	2,6	2,61	2,6	2,9	5,5	0,519	0,269
SL	1,3	2,1	2,5	2,54	2,5	2,9	4,4	0,607	0,368
Temperatura									
Estación	Mínimo	Q1	Yo	m	Mo	P3	Máximo	S	S2
	°C	°C	°C	°C	°C	°C	°C	°C	
EN	19,3	24,3	24,7	24,06	24,5	25,05	25,9	1,741	3,0
EP	15,6	16,4	16,9	16,95	17,2	17,45	19,0	0,722	0,5
ML	15,6	18,45	18,9	18,76	18,6	19,4	24,1	1,084	1,2
SL	11,0	12,2	12,8	12,75	13,3	13,3	14,5	0,736	1,084

## Desarrollo y evaluación del modelo predictivo

### Preparación de datos

Se realizó el siguiente preprocesamiento de datos: 1.- reemplazo de valores faltantes usando el método k-Nearest Neighbors (KNN), para asegurar la continuidad temporal (24), 2.- identificación de relaciones significativas entre variables mediante análisis de correlación de Pearson (25), y 3.- configuración del conjunto de datos en cuatro subconjuntos (55% entrenamiento, 20% pruebas, 25% validación y 100% simulación) para el entrenamiento y evaluación del rendimiento del modelo.



**Figura 2.** Comportamiento del promedio anual de variables climáticas: a.- Caudal, b.- Precipitación, c.- Evaporación, y d.- Temperatura.

### Arquitectura de modelos neuronales artificiales

Se implementó una variante especializada de redes neuronales recurrentes, específicamente un tipo de Memoria a Corto Plazo Largo (LSTM), descrita en Ayadi et al., 2024 (26), para el modelo. Esta variante fue seleccionada por su capacidad para capturar las complejas dependencias temporales en series hidrológicas. La estructura incluía: 1.- una capa de entrada para las variables de precipitación, temperatura, evaporación y flujo; 2.- capas ocultas de LSTM con mecanismos de compuerta (entrada, olvido y salida) para controlar el flujo de información y mantener dependencias a largo plazo; y 3.- una capa de salida densa con una función de activación lineal para estimar el caudal. El modelo fue entrenado utilizando el optimizador Adam y la función de pérdida por error cuadrático medio (MSE).

**Tabla 2.** Resultados de la prueba de normalidad de Shapiro-Wilk y de la prueba de Mann-Kendall.

Variable	Estación	Prueba de Shapiro-Wilk		Prueba de Mann-Kendall	
		W	Valor P	t	Valor P
Caudal	MG	0,62138	< 2,2E-16	-0,050570093	1,40E-04
	EN	0,48556	< 2,2E-16	0,020568889	1,71E-01
Precipitaciones	EP	0,57559	< 2,2E-16	-0,004470214	7,63E-01
	SL	0,58548	< 2,2E-16	-0,026856607	7,05E-02
	ML	0,55364	< 2,2E-16	0,018138796	2,21E-01
	LG	0,63872	< 2,2E-16	-0,059117045	4,23E-05
	LB	0,51017	< 2,2E-16	-0,037041675	1,38E-02
	EN	0,94273	< 2,2E-16	0,082426094	0,00E+00
Evaporación	EP	0,81697	< 2,2E-16	0,051181767	1,39E-04
	ML	0,96758	< 2,2E-16	-0,101046145	5,63E-14
	SL	0,93201	< 2,2E-16	-0,022796048	8,91E-02
Temperatura	EN	0,78094	< 2,2E-16	0,390442461	0,00E+00
	ML	0,74612	< 2,2E-16	0,082547799	0,00E+00
	EP	0,99503	< 2,2E-16	-0,039089784	3,43E-03
Humedad relativa	EN	0,91579	< 2,2E-16	-0,317095846	0,00E+00
	ML	0,99341	< 2,2E-16	-0,058222272	1,28E-05
	EP	0,92717	< 2,2E-16	0,139304444	0,00E+00

### Métricas de evaluación

El rendimiento del modelo se evaluó utilizando las siguientes métricas (Ecuaciones (1)-(3)): 1.- error cuadrático medio raíz (RMSE), que mide la magnitud media del error, penalizando más los errores mayores, 2.- el coeficiente de correlación (R), que indica la intensidad y dirección de la relación lineal en este caso entre el valor real y el valor predicho, y 3.- el coeficiente de determinación ( $R^2$ ), que indica la proporción de la variabilidad entre la variable dependiente que se explica por las variables independientes, corresponde al cuadrado de R, y esencialmente indica la bondad del ajuste del modelo.

$$RMSE = \sqrt{\frac{1}{N} \sum (X_{predicho} - X_{observado})^2} \quad \text{Ec. (1)}$$

$$R = \sqrt{1 - \frac{\sigma_y^2}{\sigma_r^2}} \quad \text{Ec. (2)}$$

$$R^2 = 1 - \frac{\sigma_y^2}{\sigma_r^2} \quad \text{Ec. (3)}$$

### Herramientas computacionales

Utilizando el lenguaje de programación Python (37), se desarrolla un código computacional cuya arquitectura implementa la red neuronal LSTM para fines de regresión (Figura 3), y el script se ejecuta para modelado en la plataforma Google Colab, utilizando librerías del lenguaje mencionado que permitieron la implementación y entrenamiento de la arquitectura de red LSTM (NumPy, Pandas, Scikit-learn, TensorFlow y Keras).

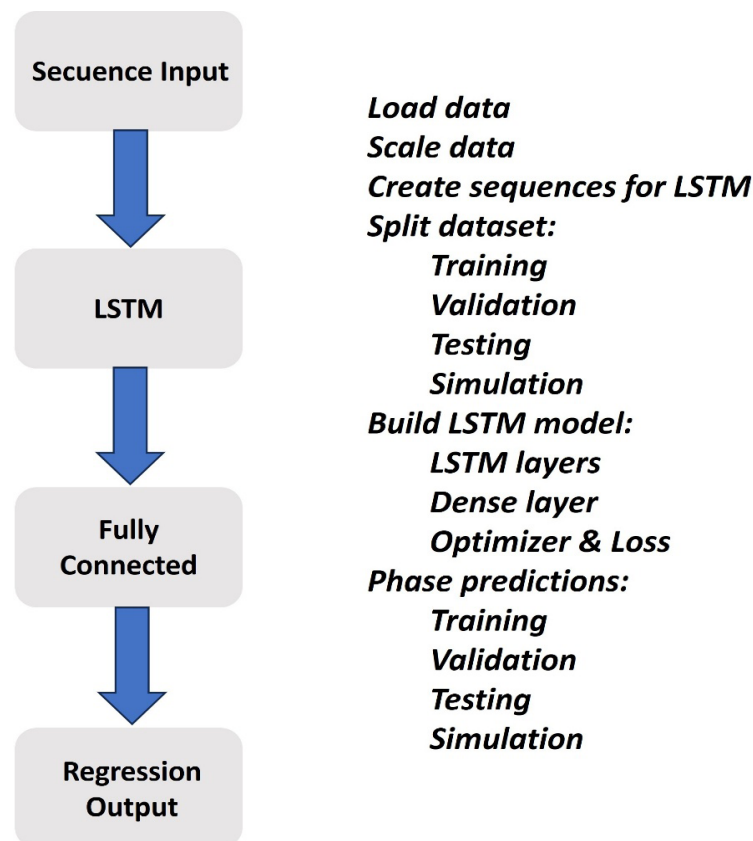
## Resultados

### Análisis estadístico de variables hidroclimáticas

El análisis estadístico de las variables hidroclimáticas reveló una distribución simétrica en las estaciones analizadas, dado que los valores medios y medianos mostraban un alto grado de consistencia. La mayor variabilidad térmica se observó en la estación Acueducto Tuluá, en el rango de 19,3-25,05 °C en el tercer cuartil, lo que la sitúa como la zona más cálida de la cuenca. Por el contrario, las temperaturas más bajas, que oscilan entre 11,0 y 13,3 °C, se registraron en la estación de Santa Lucía. La mayor varianza y desviación estándar se observaron en la estación Acueducto Tuluá, 3,0°C y 1,741°C respectivamente, lo que indica una mayor dispersión térmica de esta estación en relación con las demás.

### Prueba de normalidad

Aunque varios  $W$  coeficientes se acercaron a 1, se acercaron cerca de 1, los  $P$  valores asociados eran menores que 0,05. En particular, las variables de evaporación y humedad relativa mostraron coeficientes altos  $W$ ; sin embargo, con , confirman que no se cumple la normalidad para estas estaciones.  $P < 2,2E^{-16}$



**Figura 3.** Arquitectura para regresión (izquierda) y pseudocódigo (derecha) para entrenar la red LSTM. Fuente: elaboración propia conforme a la metodología y ejemplos de aplicación descritos en (37).

## Análisis de tendencias

- Basado en el coeficiente de Mann-Kendall, el análisis de tendencias reveló comportamientos diferenciados en las variables hidroclimáticas, destacando lo siguiente:
- El caudal en la estación de Mateguadua mostró una disminución significativa, con  $\tau = 0,051$  y  $P = 0,00014$ .
- La evaporación mostró tendencias combinadas; por un lado, en la estación de El Placer hubo un aumento significativo con  $\tau = 0,051$ , mientras que en la estación Monteloro hubo una disminución significativa con  $\tau = -0,051$  y  $P \sim 0$ .
- Las precipitaciones en las estaciones de La Gitana y Los Bancos mostraron una ligera tendencia a la baja.
- La humedad relativa disminuyó en la estación Monteloro y aumentó en la estación de El Placer, un comportamiento esperado dada la relación inversa entre temperatura y humedad.
- En cuanto a la temperatura, se observaron aumentos significativos en las estaciones Acuedcto Tuluá y Monteloro, y se observó una ligera disminución en la estación El Placer.

## Rendimiento del modelo predictivo basado en la red LSTM

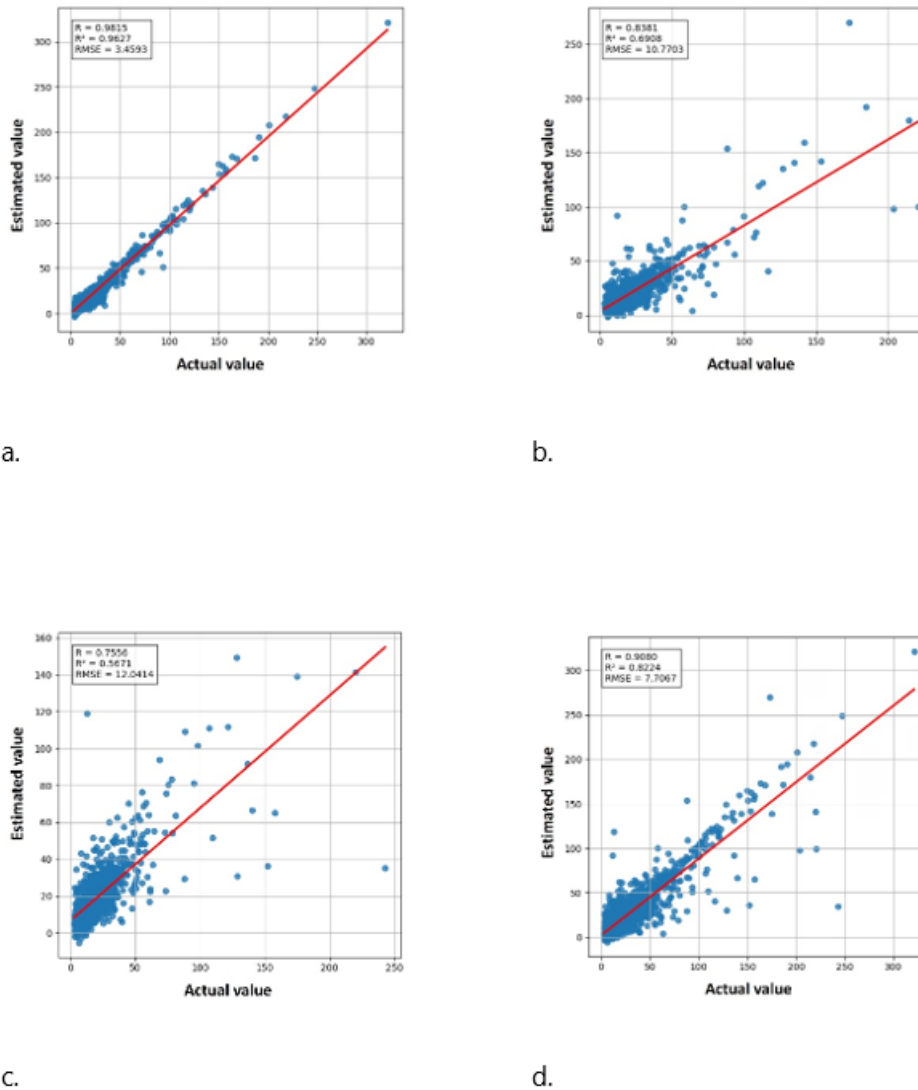
La comparación gráfica entre los valores reales y estimados en cada etapa (Figura 4) demuestra el buen ajuste del modelo y su utilidad para aplicaciones prácticas en la gestión del agua. En este sentido, las métricas de evaluación de la red neuronal LSTM confirman el alto rendimiento mostrado gráficamente al predecir el caudal del río Tuluá en las diferentes etapas del proceso de modelado.

Como se muestra en la Figura 4, el ajuste del modelo es excelente durante la fase de entrenamiento, con un 96,27% de la variabilidad del caudal explicada por las variables de entrada. Durante la fase de pruebas, el modelo mantiene una buena capacidad predictiva fuera del conjunto de entrenamiento, explicando el 69,08% de la variabilidad. Durante la fase de validación, el rendimiento sigue siendo adecuado, con un 56,71% de la variabilidad explicada por las variables de entrada. Finalmente, en la simulación global con todos los datos, se confirma la generalizabilidad y precisión del modelo, logrando el 82,24% de la variabilidad explicada.

## Discusión

En general, esta arquitectura demuestra su capacidad para manejar dependencias temporales complejas y no lineales, permitiéndole capturar con precisión las dinámicas estacionales y con retraso presentes en variables hidrológicas como precipitación, temperatura y, en particular, caudal de los ríos, que es el objetivo de la predicción. En este sentido, Waqas & Humphries, 2024 (27), destacan la eficacia de las redes LSTM para modelar sistemas hidrológicos no lineales y variables en

el tiempo, superando las limitaciones de otras redes en el modelado de dependencias a largo plazo, atribuidas al uso de celdas de memoria y mecanismos de puertas.



**Figura 4.** Comparación gráfica entre los valores reales y los valores estimados en cada una de las fases: a.- Entrenamiento, b.- Pruebas, c.- Validación y d.- Simulación.

El análisis exploratorio inicial confirmó la presencia de variaciones atípicas y registros dispersos en las variables estudiadas, sugiriendo alteraciones climáticas relevantes en la región, como se infirió en otros estudios (28, 29). Aplicar el coeficiente de correlación de Pearson en combinación con el algoritmo KNN para completar datos faltantes fortaleció la base de datos, asegurando la continuidad de la serie temporal y aumentando el número de registros diarios por variable (en el caso más crítico de datos faltantes, temperatura en ML, de 4293 a 6785), lo cual fue fundamental para un entrenamiento eficaz del modelo.

La prueba de normalidad de Shapiro-Wilk indicó que la serie temporal no cumplía con la suposición de normalidad (véase Tabla 2), justificando la elección de una técnica no paramétrica para el

análisis estadístico posterior. Esto se complementó con la identificación de tendencias utilizando el coeficiente de Mann-Kendall, que reveló patrones significativos en las variables del estudio, como disminuciones en el caudal en áreas específicas y patrones mixtos en cuanto a evaporación y humedad relativa.

Los hallazgos estadísticos anteriores destacan la necesidad de monitorizar el clima segmentado por subzonas de la cuenca debido a la presencia de heterogeneidades significativas en la respuesta hidrológica.

Finalmente, la red LSTM mostró un rendimiento robusto con coeficientes de correlación superiores a 0,75 en todas las etapas, y un valor global de 0,9080 en la simulación con todos los datos, proporcionando así un nivel fiable de precisión que confirma su potencial para la predicción del caudal hidrológico y su uso como herramienta en la gestión regional del agua. De La Fuente et al., 2024 (30), destacan la capacidad del modelado basado en LSTM para superar los métodos tradicionales de modelado basados en procesos agrupados tanto temporal como espacialmente, para la predicción de caudales; sin embargo, reconocen la complejidad estructural de la red en cuestión, que incluye operaciones de control y procesamiento secuencial de datos, así como las dificultades para interpretar los procesos internos y las ponderaciones del modelo. Por lo tanto, lo anterior reafirma el enfoque seguido para los trabajos realizados en la cuenca del río Tuluá; del mismo modo, el rendimiento obtenido es comparable al reportado en estudios previos, como se resume en la Tabla 3.

La Tabla 3 presenta los resultados de rendimiento (coeficiente de correlación R) reportados en estudios previos y en la investigación actual. Incluye diferentes modelos de redes neuronales aplicados a distintas cuencas, las fases evaluadas y las métricas obtenidas. La última fila corresponde al presente estudio, que muestra métricas de evaluación por fase para el modelo LSTM aplicado a la cuenca del río Tuluá, demostrando su capacidad para captar la dinámica hidrológica y su competitividad en comparación con otros modelos reportados en la literatura.

**Tabla 3.** Evaluación comparativa de modelos LSTM para la predicción de caudales: estudios previos y trabajo actual.

Investigación	Modelo	Cuenca / Región	Fases evaluadas	Métricas (R)	Observaciones
Li et al., 2022 (31)	CNN-LSTM	Río Hun, China	Formación / Validación	Durante todo el año: 0,894 / 0,886; Temporada de lluvias: 0,888 / 0,884; Estación seca: 0,799 / 0,721	Modelo híbrido, evaluación por períodos

Wilbran et al., 2023 (32)	Modelos Híbridos de Tiempo Multiescala-LSTM	>500 cuencas de la base de datos CAMEL-US	No específica.	Pendiente media: 0,68; Baja frecuencia de precipitación: 0,62; Conductividad hidráulica saturada: 0,63	Evaluación según las características de la cuenca
Sahoo et al., 2019 (33)	LSTM / RNN convencional	Río Mahanadi, India	Formación	LSTM: 0.971; RNN: 0.967	Cuencas de bajo caudal
Girihagama et al., 2022 (34)	LSTM / LSTM + mecanismos de atención	10 subcuencas del río Ottawa, Canadá	Formación	LSTM: 0.992; LSTM + mecanismos de atención: 0.977	Comparación entre LSTM convencional y LSTM con mecanismos de atención
<b>Esta investigación</b>	RNN convencional	Río Tuluá, Colombia	Entrenamiento / Pruebas / Validación / Simulación General	Entrenamiento: 0,9815; Pruebas: 0,8381; Validación: 0,7556; Simulación general: 0,9080	Evaluación por fases; Captura de dinámicas estacionales y retardadas

## Conclusiones

El modelo neuronal artificial basado en la red LSTM demostró ser una herramienta robusta y fiable para modelar y predecir el comportamiento del flujo en la cuenca del río Tuluá, mostrando altos niveles de ajuste y generalización.

El uso de métodos estadísticos no paramétricos y técnicas de aprendizaje profundo nos permitió abordar la complejidad y la no linealidad inherentes a las series temporales hidroclimáticas.

Los patrones de variabilidad en el caudal, evaporación y humedad relativa reflejan la influencia de fenómenos climáticos globales como el ENSO, lo que indica que su incorporación en futuros modelos podría mejorar la precisión predictiva.

El desempeño del modelo en diferentes fases y estaciones demuestra su utilidad como herramienta para la gestión sostenible de los recursos hídricos en la cuenca del río Tuluá, apoyando estrategias de planificación y adaptación frente a la variabilidad y el cambio climático.

### Declaración de contribución a la autoría de CreditT

Conceptualización – Ideas, Curación de Datos, Análisis Formal, Investigación, Metodología y Recursos: León Blandón. Gestión de Proyectos, Software y Supervisión: Luis González. Validación, Redacción – Borrador original – Elaboración y Redacción – Revisión y edición – Elaboración: León Blandón, Luis González.

## Referencias

1. Pizzorni M, Innocenti A, Tollin N. Droughts and floods in a changing climate and implications for multi hazard urban planning: A review. *City Environ Interact*. 2024;24:100169.

<https://doi.org/10.1016/j.cacint.2024.100169>

2. Furtak K, Wolińska A. The impact of extreme weather events as a consequence of climate change on the soil moisture and on the quality of the soil environment and agriculture - A review. *Catena*. 2023;231:107378

<https://doi.org/10.1016/j.catena.2023.107378>

3. Chaudhary MT, Piracha A. Natural Disasters-Origins, Impacts, Management. *Encyclopedia*. 2021;1:1101 31.

<https://doi.org/10.3390/encyclopedia1040084>

4. Skilodimou HD, Bathrellos GD. Natural and technological hazards in urban areas: assessment, planning and solutions. *Sustainability*. 2021;13(15):8301. doi: <https://doi.org/10.3390/su13158301>

5. Fischer G. El aumento de las inundaciones generado por el cambio climático afectará nuestros cultivos. *Rev Fac Nac Agron Medellín*. 2021;74(3):9619 20. [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0304-28472021000309619&lng=en&tlng=es](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0304-28472021000309619&lng=en&tlng=es)

6. Molina JE, Arboit ME, Maglione DS, Sedevich A, Mutani G. Estudio de expansión urbana, crecimiento poblacional, consumos energéticos e índices de vegetación en el Área Metropolitana de Mendoza. *AREA*. 2020;26(1):1 21. <https://dialnet.unirioja.es/servlet/articulo?codigo=8213842>

7. Zhungo Carpio JP, Baquedano Rivera PR. Impacto Ambiental de la Expansión Urbana no Planificada en los Recursos hídricos del Cantón Santo Domingo: Un Análisis crítico. *Rev Científica Salud Desarro Humano*. 2025;6(1):1600 17.

<https://doi.org/10.61368/r.s.d.h.v6i1.542>

8. Reyer C, Adams S, Albrecht T, et al. Climate change impacts in Latin America and the Caribbean and their implications for development. *Reg Environ Change*. 2017;17:1601 21.

<https://doi.org/10.1007/s10113-015-0854-6>

9. Restrepo JC, Higgins A, Escobar J, Ospino S, Hoyos N. Contribution of low frequency climatic-oceanic oscillations to streamflow variability in small, coastal rivers of the Sierra Nevada de Santa Marta (Colombia). *Hydrol Earth Syst Sci*. 2019;23:2379 400.

<https://doi.org/10.5194/hess-23-2379-2019>

10. Rodriguez Espinosa CA, Valencia CF, Ramirez CG, Valencia CD. Dynamic effect of climate change on flood damage cost in the Andean region of Colombia using an ARDL ECM model and climate change projections. *Sustainable Cities Soc*. 2025;125:106309.

<https://doi.org/10.1016/j.scs.2025.106309>

11. Ávila Á, Guerrero FC, Escobar YC, Justino F. Recent precipitation trends and floods in the Colombian Andes. *Water*. 2019;11(2):379.

<https://doi.org/10.3390/w11020379>

12. Flores P, Navarro Y. Efectos del cambio climático en los recursos hídricos: un análisis de la experiencia en Venezuela. *Novum Scientiarum*. 2016;2(4):45 55.

13. Amparo Salcedo M, Pérez Gimeno A, Navarro Pedreño J. Water security under climate change: challenges and solutions across 43 countries. *Water*. 2025;17(5):633.  
<https://doi.org/10.3390/w17050633>
14. Carvajal Y, Enciso A, Sandoval MC. Hydrological analysis of historical floods in the upper valley of Cauca river. *Inycomp* [Internet]. 2016 Jan 19 [cited 2025 Aug 19];18(1):46-57.  
<https://doi.org/10.25100/iyc.v18i1.2176>
15. González López N, Carvajal Escobar Y. Characterization of hydrological drought in the Cauca river high valley. *Tecnol Cienc Agua*. 2020;11(1):235-65.  
<https://doi.org/10.24850/j-tyca-2020-01-06>
16. Yunita A, Pratama MI, Almuzakki MZ, Ramadhan H, Akashah PAE, Mansur ABF, et al. Performance analysis of neural network architectures for time series forecasting: a comparative study of RNN, LSTM, GRU, and hybrid models. *MethodsX*. 2025;15:103462.  
<https://doi.org/10.1016/j.mex.2025.103462>
17. Renteria Mena JB, Plaza D, Giraldo E. Multivariate hydrological modeling based on long short term memory networks for water level forecasting. *Information*. 2024;15(6):358.  
<https://doi.org/10.3390/info15060358>
18. Montoya Zuluaga LM. Predicción de caudales en el sistema Magdalena Cauca utilizando modelos estadísticos y redes neuronales recurrentes [thesis]. Medellín: Universidad de Antioquia; 2023 <https://bibliotecadigital.udea.edu.co/handle/10495/37887>
19. Morante Carballo F, Arcentales Rosado M, Caicedo Potosí J, Carrión Mero P. Artificial intelligence applications in hydrological studies and ecological restoration of watersheds: a systematic review. *Watershed Ecol Environ*. 2025;7:230-48.  
<https://doi.org/10.1016/j.wsee.2025.05.004>
20. Volpi E, Kim JS, Jain S, Shrestha S. Editorial: artificial intelligence in hydrology. *Hydrol Res*. 2023;54(6):iii-iv. <https://doi.org/10.2166/nh.2023.102>
21. Abubaker Blbas HT. Descriptive statistics [Internet]. In: *Recent Advances in Biostatistics*. IntechOpen; 2024 <https://doi.org/10.5772/intechopen.1002179>
22. Das KR, Imon AHM. A brief review of tests for normality. *Am J Theor Appl Stat*. 2016;5(1):5-12.  
<https://doi.org/10.11648/j.ajtas.20160501.12>
23. Liu S, Xie Y, Fang H, Du H, Xu P. Trend test for hydrological and climatic time series considering the interaction of trend and autocorrelations. *Water*. 2022;14(19):3006.  
<https://doi.org/10.3390/w14193006>
24. Seoane Santos M, Henriques Abreu P, Wilk S, Santos J. How distance metrics influence missing data imputation with k-nearest neighbours. *Pattern Recognit Lett*. 2020;136:111-9.  
<https://doi.org/10.1016/j.patrec.2020.05.032>
25. Hernández Lalinde JD, Espinosa Castro F, Rodríguez JE, Chacón Rangel JG, Toloza Sierra CA, Arenas Torrado MK, et al. Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones. *Arch Venez Farmacol Ter*. 2018;37(5).  
[https://www.revistaavft.com/images/revistas/2018/avft\\_5\\_2018/20sobre\\_uso\\_adecuado\\_coeficiente.pdf](https://www.revistaavft.com/images/revistas/2018/avft_5_2018/20sobre_uso_adecuado_coeficiente.pdf)

26. Ayadi A, Truong TMD, Frydman C, Wemmert C. Explaining LSTM neural networks: a DEVS-based approach. In: Proceedings of the 6th European Modeling & Simulation Symposium, 21st International Multidisciplinary Modeling & Simulation Multiconference. 2024.  
<https://doi.org/10.46354/i3m.2024.emss.035>
27. Waqas M, Wannasingha Humphries U. A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*. 2024;13:102946.  
<https://doi.org/10.1016/j.mex.2024.102946>
28. Arrieta Castro M, Donado Rodríguez A, Acuña GJ, Canales FA, Teegavarapu RSV, Kaźmierczak B. Analysis of streamflow variability and trends in the Meta River, Colombia. *Water*. 2020;12(5):1451.  
<https://doi.org/10.3390/w12051451>
29. Umar DA, Ramli MF, Aris AZ, Jamil NR, Aderemi AA. Evidence of climate variability from rainfall and temperature fluctuations in semi arid region of the tropics. *Atmos Res*. 2019;224:52-64. <https://doi.org/10.1016/j.atmosres.2019.03.023>
30. De la Fuente LA, Ehsani MR, Gupta HV, Condon LE. Toward interpretable LSTM-based modeling of hydrological systems. *Hydrol Earth Syst Sci*. 2024;28:945-71. doi:10.5194/hess-28-945-2024  
<https://doi.org/10.5194/hess-28-945-2024>
31. Li X, Xu W, Ren M, Jiang Y, Fu G. Hybrid CNN-LSTM models for river flow prediction. *Water Supply*. 2022;22(5):4902-11. <https://doi.org/10.2166/ws.2022.170>
32. Wilbrand K, Taormina R, ten Veldhuis M-C, Visser M, Hrachowitz M, Nuttall J, et al. Predicting streamflow with LSTM networks using global datasets. *Front Water*. 2023;5:1166124. <https://doi.org/10.3389/frwa.2023.1166124>
33. Sahoo BB, Jha R, Singh A, et al. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophys*. 2019;67:1471-81. <https://doi.org/10.1007/s11600-019-00330-1>
34. Giriagama L, Khaliq MN, Lamontagne P, et al. Streamflow modelling and forecasting for Canadian watersheds using LSTM networks with attention mechanism. *Neural Comput Appl*. 2022;34:19995-20015. <https://doi.org/10.1007/s00521-022-07523-8>
35. Bustamante CA, Naranjo JF, Ahumada JM. Determinación de la calidad ecológica del río Tuluá Valle del Cauca. *Mutis*. 2014;4(2):35-43. <https://doi.org/10.21789/22561498.955>
36. López, JP. El río como hilo conector: caso Tuluá, Valle del Cauca [Internet]. *Sociedad-Espacio-Naturaleza*; 2022. <https://sociedadespacionaturaleza.wordpress.com/2021/09/03/el-rio-como-hilo-conector-caso-tuluá-valle-del-cauca/>
37. TechwithJules. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) - Creating an LSTM Model in Python Using TensorFlow and Keras [Internet]. 2024 Oct 12 [citado 2025 Dic 05 ]; Disponible en: <https://medium.com/@techwithjules/recurrent-neural-networks-rnns-and-long-short-term-memory-lstm-creating-an-lstm-model-in-13c88b7736e2>