APLICAÇÃO DE MODELAGEM PREDITIVA PARA PREVISÃO DE SÉRIES HIDROLÓGICAS NA BACIA DO RIO PIANCÓ, PARAÍBA

Wesley Devison Melo Xavier [1]; Daniel Conceição do Nascimento [2]; Igor Fernandes Gomes [3]; Cristiano das Neves Almeida [4]; Iago Nogueira Pessoa [5]; Bruno Roberto Ferreira da Silva [6] & Eliaquim Damião de Pontes [7]

RESUMO – A gestão eficiente dos recursos hídricos no semiárido brasileiro exige projeções hidrológicas confiáveis, especialmente em bacias intermitentes e vulneráveis como a do rio Piancó, na Paraíba. Este trabalho avaliou a aplicação de uma rede neural profunda Long Short-Term Memory (LSTM) para prever os volumes mensais de precipitação em quatro estações da referida bacia. Utilizou-se uma série histórica de dez anos (2013 a 2023) para treinamento e calibração, projetando-se os cenários para o ano de 2024. A arquitetura foi otimizada com três camadas profundas, função Huber Loss e mecanismo de Early Stopping. Os resultados evidenciaram que o modelo capturou de forma consistente a variabilidade sazonal, acompanhando as transições hidroclimatológicas locais sem apresentar overfitting. O desempenho visual foi validado estatisticamente, destacando-se a estação 737071 com o melhor nível de ajuste (R^2 = 0,883). Embora tenha ocorrido a subestimação inerente de picos extremos, as magnitudes residuais mantiveram-se contidas. Conclui-se que o modelo LSTM univariado atua como uma ferramenta viável para subsidiar o planejamento estratégico do abastecimento e a alocação de outorgas na região.

ABSTRACT– Efficient water resource management in the Brazilian semi-arid region requires reliable hydrological projections, especially in vulnerable intermittent basins like the Piancó River in Paraíba. This study evaluated the application of a deep Long Short-Term Memory (LSTM) neural network to predict monthly precipitation volumes across four stations within this basin. A ten-year historical series (2013 to 2023) was used for training and calibration, forecasting scenarios for the year 2024. The architecture was optimized with three deep layers, Huber Loss, and an Early Stopping mechanism. Results showed that the model consistently captured seasonal variability, tracking local hydroclimatological transitions without exhibiting overfitting. Visual performance was statistically validated, with station 737071 achieving the best fit (R^2 = 0.883). Although the inherent underestimation of extreme peaks occurred, residual magnitudes remained well-contained. It is concluded that the univariate LSTM model acts as a viable tool to support strategic water supply planning and water rights allocation in the region.

Palavras-Chave – Modelagem Preditiva; Redes Neurais LSTM; Recursos Hídricos.

[1]) Departamento de Engenharia Civil da Universidade Federal da Paraíba – UFPB, 58051-900, João Pessoa – PB, wesley.devison@academico.ufpb.br
[2]) Departamento de Engenharia Civil da Universidade Federal da Paraíba – UFPB, 58051-900, João Pessoa – PB, daniel.conceicao@academico.ufpb.br
[3]) Departamento de Engenharia Civil da Universidade Federal da Paraíba – UFPB, 58051-900, João Pessoa – PB, igor.fernandes@ufpe.br
[4]) Departamento de Engenharia Civil da Universidade Federal da Paraíba – UFPB, 58051-900, João Pessoa – PB, almeida74br@yahoo.com.br
[5]) Departamento de Engenharia Civil da Universidade Federal da Paraíba – UFPB, 58051-900, João Pessoa – PB, iago.nogueira@academico.ufpb.br
[6]) Departamento de Engenharia Civil da Universidade Federal da Paraíba – UFPB, 58051-900, João Pessoa – PB, bruno.roberto3@academico.ufpb.br
[7]) Departamento de Engenharia Civil da Universidade Federal da Paraíba – UFPB, 58051-900, João Pessoa – PB, Eliaquim.pontes@academico.ufpb.br

1. INTRODUÇÃO

A gestão eficiente dos recursos hídricos é um dos maiores desafios no semiárido brasileiro, uma região caracterizada por sua extrema variabilidade climática espacial e temporal, com rios predominantemente intermitentes e altas taxas de evapotranspiração (Felix e Paz, 2016). Inserida nesse contexto de vulnerabilidade, a bacia do rio Piancó, no estado da Paraíba, desempenha um papel estratégico para a segurança hídrica regional, sendo a principal formadora do complexo de reservatórios Coremas-Mãe d'Água. Trata-se de uma área ambiental e socialmente sensível, onde a ocorrência de anomalias e teleconexões climáticas agrava frequentemente os episódios de secas severas e prolongadas (Reis, 2015). Essa escassez estrutural gera intensos conflitos pelo uso da água, tornando a gestão de outorgas e o planejamento do abastecimento tarefas altamente complexas e estritamente dependentes de projeções hidrológicas confiáveis (Santos et al., 2024).

A modelagem hidrológica em bacias intermitentes, como a do Piancó, esbarra na escassez de dados e nas limitações dos modelos tradicionais diante de processos físicos irregulares (Felix e Paz, 2016). Nesse cenário, a inteligência artificial destaca-se na modelagem de fenômenos extremos (Bui et al., 2019). Redes Neurais *Long Short-Term Memory* (LSTM) mostram-se altamente eficazes por reterem a memória sequencial de longo prazo, superando os modelos conceituais na predição de níveis d'água (Kim et al., 2022). A versatilidade das LSTMs é comprovada tanto na previsão de inundações em bacias brasileiras (Machado et al., 2024) quanto na modelagem de baixo fluxo (Sahoo et al., 2019), fator essencial para o manejo de rios intermitentes.

Apesar dessa consolidação, há uma lacuna metodológica na adaptação dessas arquiteturas à forte estocasticidade da precipitação do semiárido nordestino. A ausência de previsões precisas gera extrema incerteza para os órgãos gestores, comprometendo diretamente a emissão de outorgas e o controle de vazões ecológicas. Consequentemente, essa deficiência analítica inviabiliza o planejamento preventivo, perpetuando a vulnerabilidade hídrica das populações urbanas e do setor agropecuário diante de anos secos ou falhas na alocação de recursos.

Portanto, o objetivo deste artigo é avaliar a aplicação de um modelo preditivo, estruturado em uma arquitetura profunda de redes neurais, para a previsão da precipitação em 4 postos de monitoramento da Bacia do Rio Piancó. A pesquisa utiliza uma série histórica de 10 anos (2013 a 2023) para projetar os volumes mensais do ano de 2024. Adicionalmente, o estudo visa demonstrar como o ajuste da escala temporal para o nível mensal, aliado a uma janela de memória sazonal, eleva a precisão das previsões, oferecendo uma ferramenta tecnológica viável de suporte à tomada de decisão para a gestão de recursos hídricos no Nordeste brasileiro.

2. MATERIAIS E MÉTODOS

2.1. Área de estudo e banco de dados

A área de estudo compreende a bacia do rio Piancó, localizada no estado da Paraíba, conforme a Figura 1, uma região de extrema relevância estratégica e vulnerabilidade climática, atuando como a principal formadora do complexo de reservatórios Coremas-Mãe d'Água. Para a modelagem e análise espacial desta bacia, a fonte que permitiu a obtenção dos dados foi a Agência Nacional de Águas e Saneamento Básico (ANA). Por meio de suas plataformas oficiais, Metadados da ANA e HidroWeb, realizou-se o download do limite vetorial da bacia hidrográfica, bem como das séries de precipitação das estações monitoradas utilizadas no treinamento da rede.

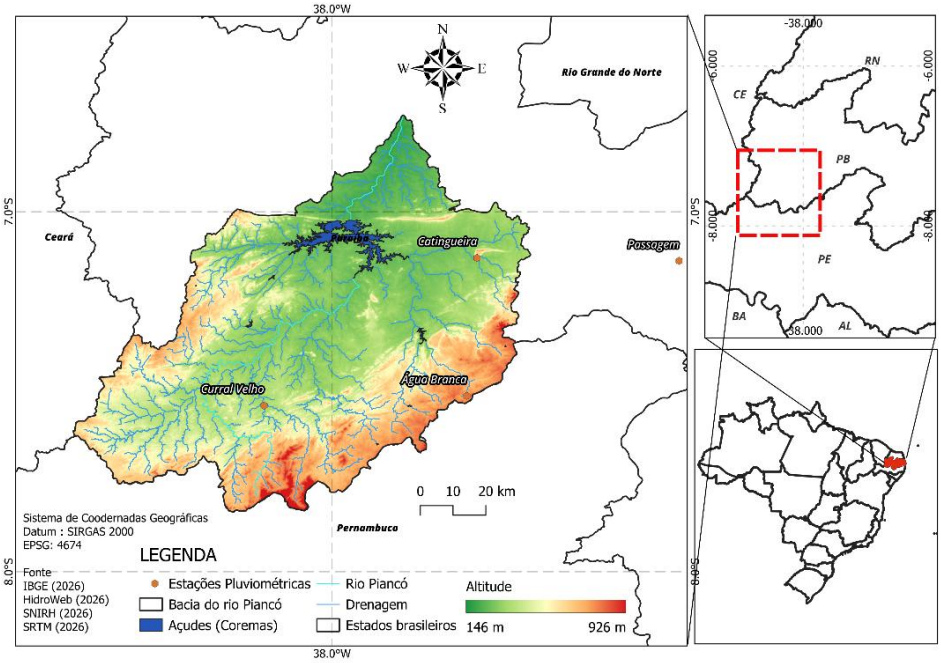


Figura 1 – Mapa de localização da bacia do rio Piancó.

FONTE: Autoria própria, (2026).

As estações escolhidas são descritas na Tabela 1.

Tabela 1 – Estações analisadas

Código	Nome da Estação	Latitude	Longitude	Período de dados
737022	Água Branca	-7,5117	-37,6367	2013-2023
737065	Catingueira	-7,1283	-37,6083	2013-2023
737071	Passagem	-7,1364	-37,0475	2013-2023
738077	Curral Velho	-7,5383	-38,1992	2013-2023

FONTE: Autoria própria, (2026).

2.2. Rede Neural Recorrente Long And Short Term Memory – LSTM

Para o processamento da série temporal, adotou-se o modelo Long Short-Term Memory (LSTM), proposto por Hochreiter e Schmidhuber (1997a). Esta arquitetura supera o problema de dissipação do gradiente típico de Redes Neurais Recorrentes (RNNs) ao utilizar de um Carrossel de Erro Constante, que mantém o sinal de erro preservado dentro da célula de cada unidade da rede (Hochreiter e Schmidhuber, 1997b).

Nesse modelo, as células atuam como redes recorrentes independentes. A versão clássica da arquitetura (vanilla LSTM) é composta pela célula de memória, responsável por reter valores sobre intervalos de tempo arbitrários, e por um conjunto de portas (gates) não-lineares que regulam o fluxo de informações. Estas portas dividem-se em: porta de entrada (input gate), porta de saída (output gate) e porta de esquecimento (forget gate), conforme a Figura 2. Esta última não fazia parte da proposição original da rede, tendo sido introduzida posteriormente por Gers et al. (2000) para permitir que a rede consiga redefinir e "esquecer" o seu próprio estado.

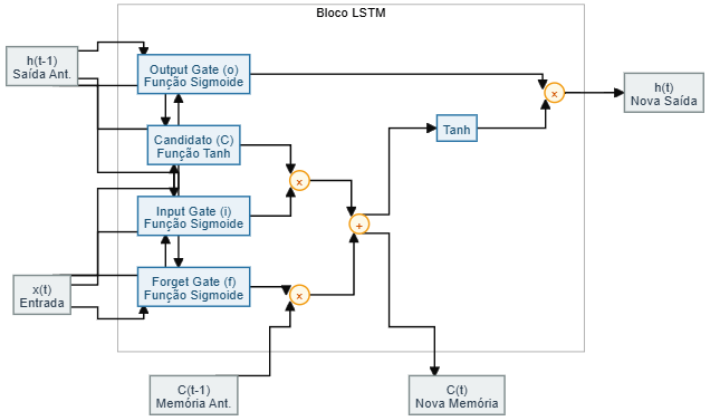


Figura 2 – Ilustração da arquitetura de uma célula de LSTM.

FONTE: Autoria própria, (2026).

O fluxo direto de processamento (forward pass) em um bloco de memória LSTM envolve uma série de operações matriciais recursivas, descritas pelas seguintes etapas:

1. Entrada do Bloco: Atualiza-se o componente de entrada combinando o dado atual x(t) e a saída da unidade LSTM no passo anterior y(t-1):

$$z(t) = g(W_z x(t) + R_z y(t-1) + b_z) \quad (1)$$

2. Porta de Entrada: Determina-se a ativação da porta de entrada e quais informações candidatas devem ser adicionadas ao estado da célula:

$$i(t) = \sigma(W_i x(t) + R_i y(t-1) + p_i \odot c(t-1) + b_i) \quad (2)$$

3. Porta de Esquecimento: O modelo decide quais informações devem ser descartadas do estado anterior da célula c(t-1):

$$f(t) = \sigma\left(W_f x(t) + R_f y(t-1) + p_f \odot c(t-1) + b_f\right) \quad (3)$$

4. Atualização da Célula: O novo estado da célula c(t) é computado através da combinação dos valores filtrados pelas portas de entrada e de esquecimento:

$$c(t) = z(t) \odot i(t) + c(t-1) \odot f(t) \quad (4)$$

5. Porta de Saída: Avalia-se o estado atual para definir qual parte da informação comporá a saída:

$$o(t) = \sigma(W_o x(t) + R_o y(t-1) + p_o \odot c(t) + b_o) \quad (5)$$

6. Saída do Bloco: Finalmente, calcula-se a saída final da unidade combinando o estado filtrado da célula com a ativação da porta de saída:

$$y(t) = g\left(c(t)\right) \odot o(t) \quad (6)$$

Nas formulações acima, W e R representam as matrizes de pesos, b é o vetor de viés (bias), p designa as conexões peephole (Gers e Schmidhuber, 2000), e o símbolo $\odot$ denota a multiplicação ponto a ponto. As ativações não-lineares são regidas pela função sigmoide logística $\sigma(x)$ para as portas de controle, e pela tangente hiperbólica $g(x) = \tanh(x)$ para as entradas e saídas do bloco.

2.3. Preparação do modelo, dos dados e definição dos hiperparâmetros

Os dados mensais (2013–2023) foram tratados em Python (Pandas) e a rede foi implementada via TensorFlow, destinando 70% da série para treinamento e 30% para calibração. O ano de 2024 foi reservado exclusivamente para a validação. A arquitetura final (Tabela 2) foi definida com três

camadas profundas (128-128-64), função Huber Loss, otimizador Adam e um mecanismo de Early Stopping (Patience = 15) para evitar overfitting. O desempenho foi avaliado estatisticamente utilizando o Erro Absoluto Médio (MAE), Erro Quadrático Médio (MSE), Raiz do Erro Quadrático Médio (RMSE) e o Coeficiente de Determinação (R^2).

Tabela 2 – Hiperparâmetros utilizados no modelo

Hiperparâmetro	Valor adotado
Arquitetura	3 Camadas LSTM
Unidades	128-128-64
Lookback	12
Epochs	100
Batch Size	16
Split Treino Val	0,7
Dropout	0,1
Função de ativação	Relu
Optimizer	Adam
Loss Function	Huber
Patience	15

FONTE: Autoria própria, (2026).

3. RESULTADOS

A análise de convergência da função de perda (Figura 3) demonstrou a estabilidade e alta capacidade de generalização da rede. A função Huber Loss garantiu o decaimento inicial do erro sem que o modelo fosse desestabilizado pelos extremos pluviométricos. O comportamento assintótico das curvas, aliado ao mecanismo de parada antecipada, confirmou o mapeamento real das não-linearidades do rio Piancó, mitigando o overfitting.



Figura 3 – Evolução do erro de treinamento e validação ao longo das épocas para as quatro estações pluviométricas analisadas.

FONTE: Autoria própria, (2026).

A Figura 4 ilustra os resultados da validação externa do modelo para o ano de 2024. Observa-se que a rede reproduziu a variabilidade sazonal da bacia do rio Piancó, acompanhando as transições entre os períodos chuvosos e de estiagem. Embora modelos LSTM tendam a atuar subestimando picos pluviométricos extremos em prol da minimização do erro global, a agregação temporal para a escala mensal atenuou essa limitação. As projeções apresentaram concordância com os dados observados, simulando de forma consistente os ciclos de alta e baixa precipitações nas estações monitoradas.

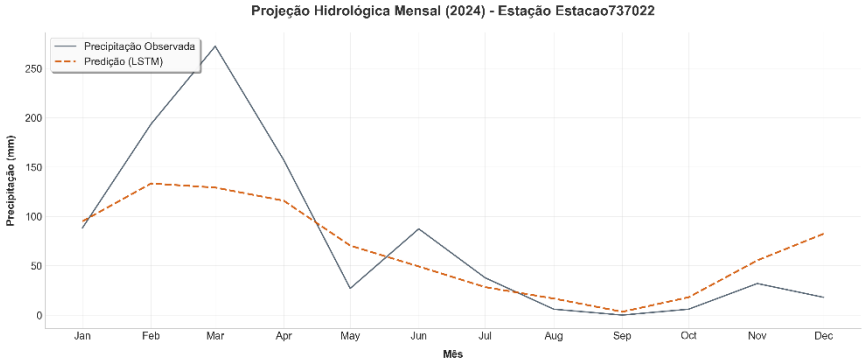




Figura 4 – Comparativo entre a precipitação mensal observada e a simulada pelo modelo LSTM para o ano de 2024.

FONTE: Autoria própria, (2026).

A avaliação quantitativa do modelo é detalhada pelos indicadores estatísticos da Tabela 3. O Coeficiente de Determinação demonstra que a rede obteve o melhor ajuste nas estações 737071 e 738077, conseguindo explicar, respectivamente, 88,3% e 83,4% da variância dos dados observados. A estação 737065 apresentou um desempenho intermediário ($R^2 = 0,796$), enquanto a estação 737022 registrou o menor índice de correlação ($R^2 = 0,592$). Esse comportamento isolado na estação 737022 reflete o impacto matemático de predições que subestimam picos pluviométricos extremos em meses singulares, reduzindo a pontuação geral da métrica de correlação linear. No dimensionamento das magnitudes residuais, a estação 737071 registrou os menores desvios, com Erro Absoluto Médio (MAE) de 13,68 mm e RMSE de 20,95 mm. Nas estações 738077, 737065 e 737022, o MAE foi de

27,71 mm, 36,67 mm e 38,01 mm, respectivamente. O distanciamento numérico observado entre os valores de MAE e RMSE (que atinge o máximo de 53,42 mm na estação 737022) evidencia a sensibilidade do RMSE em penalizar erros residuais mais altos ocorridos durante a quadra chuvosa.

Tabela 3 – Resultados das métricas

Estação	MAE	MSE	RMSE	R^2
737022	38,015	2854,673	53,4291	0,592
737065	36,677	2642,806	51,408	0,796
737071	13,687	438,914	20,950	0,883
738077	27,710	1388,619	37,264	0,834

FONTE: Autoria própria, (2026).

4. CONSIDERAÇÕES FINAIS

O presente estudo demonstrou a eficácia da arquitetura de redes neurais LSTM na previsão de precipitação mensal para a bacia do rio Piancó. A adoção da função Huber Loss e do mecanismo de parada antecipada garantiu um aprendizado estável, mitigando o risco de overfitting e permitindo que o algoritmo mapeasse satisfatoriamente a forte variabilidade sazonal e as não-linearidades inerentes ao regime pluviométrico do semiárido. Apesar da tendência matemática do modelo em atuar subestimando picos extremos, as métricas estatísticas atestaram o bom desempenho global da rede, com destaque para o ajuste alcançado na estação 737071 ($R^2 = 0,883$). Conclui-se que a agregação dos dados para a escala mensal reduziu os ruídos temporais de curto prazo, validando a aplicabilidade do algoritmo. Assim, a rede LSTM estrutura-se como uma ferramenta tecnológica promissora e viável de suporte à tomada de decisão, sendo capaz de reduzir incertezas no planejamento do abastecimento e aprimorar significativamente a gestão de outorgas no Nordeste brasileiro.

AGRADECIMENTOS

Os autores agradecem o apoio financeiro do Programa de Formação de Recursos Humanos da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (PRH-ANP), gerido pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), por meio de bolsa de graduação (Processo nº 2025/19512-9). Este trabalho foi desenvolvido no âmbito do PRH-ANP 35/UFPB – Transição Energética (Processo nº 2025/03815-2), contribuindo para a formação de recursos humanos qualificados para o setor energético no contexto da transição energética e da transformação ecológica,

e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte fornecido.

REFERÊNCIAS

BUI, D. T.; NGO, P. T. T.; PHAM, T. D. et al. (2019). “A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping”. Catena 179(October 2018), pp. 184 – 196.

FELIX, V. S.; PAZ, A. R. (2016). “Representação dos processos hidrológicos em bacia hidrográfica do semiárido paraibano com modelagem hidrológica distribuída”. Revista Brasileira de Recursos Hídricos, 21(3), pp. 556-569.

GERS, F. A.; SCHMIDHUBER, J. (2000). “Recurrent nets that time and count” in Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, vol. 3, pp. 189-194.

GERS, F. A.; SCHMIDHUBER, J.; CUMMINS, F. (2000). “Learning to forget: Continual prediction with LSTM”. Neural Computation, 12(10), pp. 2451-2471.

HOCHREITER, S.; SCHMIDHUBER, J. (1997a). “Long short-term memory”. Neural Computation, 9(8), pp. 1735-1780.

HOCHREITER, S.; SCHMIDHUBER, J. (1997b). “LSTM can solve hard long time lag problems” in Advances in Neural Information Processing Systems, 9.

KIM, D.; LEE, J.; KIM, J. et al. (2022). “Comparative analysis of long short-term memory and storage function model for flood water level forecasting of Bokha stream in NamHan River, Korea”. Journal of Hydrology 606(December 2021), p. 127415.

MACHADO, G. O.; SILVA, D. D.; MOREIRA, M. C. et al. (2024). “Aplicação de redes neurais para predição de eventos de inundação na bacia do rio Muriaé, Rio de Janeiro, Brasil” in Anais do XVII Simpósio de Recursos Hídricos do Nordeste, João Pessoa, Nov. 2024, 1, pp. 1-10.

REIS, R. L. B. S. (2015). Estudo de teleconexões climáticas para a previsão hidrológica de longo prazo da bacia do rio Piancó no semiárido paraibano. Dissertação de Mestrado, Universidade Federal de Alagoas, Maceió, AL.

SAHOO, B. B.; JHA, R.; SINGH, A.; KUMAR, D. (2019). “Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting”. Acta Geophysica 67(5), pp. 1471 – 1481.

SANTOS, K. A. O. et al. (2024). “Gestão de outorgas: um estudo de caso na bacia hidrográfica do Rio Piancó-Piranhas-Açu (PB/RN) entre os anos de 2018-2020” in Anais do 21º Simpósio Luso-Brasileiro de Engenharia Sanitária e Ambiental. Recife, PE.