

Previsão do nível da água do Rio Mundaú em Rio Largo com Redes Neurais Profundas

Lucas S. Candido¹, Christian A. de Melo Barbosa¹, Matheus B. M. Cedrim²

¹Instituto Federal de Alagoas (IFAL) – Campus Maceió

²Centro Universitário Cesmac (CESMAC)

lucas.santos.candido@gmail.com, camb1@aluno.ifal.edu.br,
matheuscedrim@hotmail.com

Abstract. *The study created an LSTM-based model to predict the daily water level of the Mundaú River in Rio Largo, Alagoas, due to recurring floods. Using historical data on rainfall and river levels (1990-2023), the model demonstrated effectiveness with an RMSE of 25.69 cm and an R^2 of 0.84 in the test phase. The results indicate applicability for flood alerts and water management, with suggestions for future improvements, such as the inclusion of environmental variables and hybrid models.*

Resumo. *O estudo criou um modelo baseado em LSTM para prever o nível diário do Rio Mundaú em Rio Largo, Alagoas, devido a cheias recorrentes. Utilizando dados históricos de chuva e nível do rio (1990-2023), o modelo demonstrou eficácia com RMSE de 25,69 cm e R^2 de 0,84 na fase de teste. Os resultados indicam aplicabilidade para alertas de cheias e gestão hídrica, com sugestões para futuras melhorias, como inclusão de variáveis ambientais e modelos híbridos.*

1. Introdução

A bacia hidrográfica do Rio Mundaú, particularmente no seu trecho que percorre o município de Rio Largo, Alagoas, é uma região de significativa relevância socioambiental. No entanto, esta área enfrenta um desafio recorrente e impactante: a ocorrência de cheias de grandes proporções, que colocam em risco uma parcela considerável da população local [Virães and Pinto 2020]. A vulnerabilidade a esses eventos hidrológicos extremos evidencia a necessidade crítica de desenvolver ferramentas capazes de prever com antecedência as variações do nível do rio.

Diante de contextos similares, modelos de aprendizado de máquina têm ganhado destaque como alternativas promissoras para previsão hidrológica. Trabalhos como o de [Adli Zakaria et al. 2023] e [Chang et al. 2014] demonstram o uso eficaz dessas técnicas na previsão de níveis de água em diferentes contextos geográficos. Viabilizando a implementação de sistemas de alerta antecipado mais eficazes, permitindo a evacuação e proteção de vidas e bens, mas também subsidia a gestão operacional dos recursos hídricos e o planejamento de ações de mitigação a médio e longo prazo.

Ainda convém mencionar que, o estudo realizado por [Zhihui Dai et al. 2023] comparou diversos algoritmos de previsão do nível de cheia de bacias hidrográficas de rios. Demonstrando como especificamente uma rede neural do tipo Long Short-Term Memory (LSTM) obteve resultados superiores a abordagens tradicionais como ARIMA,

Random Forest, Linear Regression e redes mais comuns de back-propagation. Tendo isso em vista, o presente estudo propõe e avalia a aplicação de um modelo de inteligência artificial, especificamente uma rede neural no formato LSTM, para a previsão de um dia no futuro do nível de cota diária do Rio Mundaú em Rio Largo.

2. Materiais e Métodos

2.1. Local de Estudo e Dados

O presente estudo concentra-se na análise hidrológica do Rio Mundaú, especificamente no trecho localizado as margens da fazenda Boa Fortuna no município de Rio Largo em Alagoas. A cidade localiza-se a 23 km da capital do estado, segundo os dados do IBGE [IBGE 2022], possui uma área territorial de aproximadamente 294 km² e uma população residente de 93.927 habitantes. Boa parte desta população inserida na bacia do Rio Mundaú está sujeita a cheias de grandes proporções [Virães and Pinto 2020].

A obtenção dos dados ocorreu através das estações localizadas na fazenda Boa Fortuna e publicamente disponibilizada pelo portal Hydroweb da Agência Nacional de Águas (ANA). Utilizamos os valores de chuva diária (mm) da estação pluviométrica de código 00935056, localizada nas coordenadas de 9°28'03.0"S 35°51'23.0"W. Enquanto pegamos as informações do nível diário da água (cm) do rio Mundaú através da estação fluviométrica de código 39770000, situada em 9°28'01.9"S 35°51'34.9"W. Empregamos as séries históricas entre janeiro de 1990 até dezembro de 2023 nesta pesquisa.

2.2. Processamento dos Dados

Os dados coletados apresentaram lacunas em determinados períodos, sendo preenchidas por interpolação linear para estimar os valores ausentes em cada série histórica. Para aprimorar o desempenho do algoritmos preditivo, aplicou-se a normalização dos dados por meio do escalonamento mínimo-máximo, reescalando a precipitação diária e o nível do rio para o intervalo [0,1].

O modelo preditivo adotado baseia-se na identificação de padrões temporais em séries históricas. Para isso, os dados foram organizados em janelas deslizantes de 20 dias, visando prever o nível do rio no dia seguinte. Por fim, os dados foram segmentados em três conjuntos: treinamento (1° de janeiro de 1990 a 18 de outubro de 2013), validação (19 de outubro de 2013 a 6 de agosto de 2020) e teste (7 de agosto de 2020 a 31 de dezembro de 2023).

2.3. Construção do Modelo com LSTM

As redes neurais LSTM são uma variação avançada de Redes Neurais Recorrentes (RNN), projetadas para lidar com séries temporais e dados sequenciais de forma eficiente [Adli Zakaria et al. 2023]. Diferentemente das RNN tradicionais, as LSTM possuem um mecanismo interno chamado célula de memória, que permite armazenar e reter informações ao longo do tempo, reduzindo o problema do desvanecimento do gradiente. O algoritmo foi implementado utilizando Python, Keras e Tensorflow, a Tabela 1 demonstra a construção do modelo e sua configuração de parâmetros.

Tabela 1. Arquitetura do modelo feito com LSTM

Camada	Parâmetros
Input	(20 dias, 2 variáveis)

LSTM	128 unidades, ativação ReLU, com retorno de sequências
Dropout	20% de neurônios
LSTM	32 unidades, ativação ReLU, sem retorno de sequências
Dropout	20% de neurônios
Dense	8 neurônios
Dropout	10% de neurônios
Output	1 neurônio

Para otimizar o modelo, usamos o ADAM com a função de perda de erro absoluto médio (MAE). Monitoramos o treino, salvando os pesos com o menor erro de validação, reduzindo a taxa de aprendizado em 10% a cada 5 épocas sem melhora no valor.

3. Resultados

O modelo foi treinado por 40 épocas no ambiente computacional do Google Colab, com a GPU Tesla T4 de 15gb, tendo uma duração média de 4 segundos de treino por época. Durante o treinamento, a melhor solução foi obtida na 26ª época, alcançando uma perda de validação de 1,456% (vide Figura 1). Após esse ponto, mesmo com a redução programada da taxa de aprendizado nas épocas subsequentes, não foram observados ganhos significativos na performance, indicando estabilização do treinamento e possível saturação da capacidade de generalização do algoritmo. A capacidade preditiva do modelo foi quantificada por meio das métricas Raiz do Erro Quadrático Médio (RMSE) e Coeficiente de Determinação (R^2), aplicadas aos conjuntos de dados de treinamento, validação e teste.

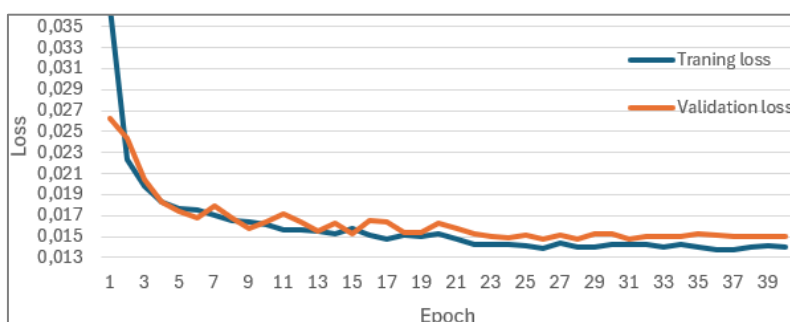


Figura 1. Histórico da perda de treinamento e perda de validação durante o processo de treinamento do modelo.

As métricas de avaliação destacaram a eficácia do modelo em capturar padrões hidrológicos. O RMSE refletiu erros médios de 19,31 cm no treino, 20,56 cm na validação e 25,69 no teste, valores que, no contexto de previsão do nível do rio, representam uma margem de erro aceitável para aplicações práticas, como alertas de cheias ou gestão operacional de recursos hídricos. O R^2 , por sua vez, demonstrou consistência entre as fases de treino (0,82), validação (0,81) e teste (0,84), indicando que o modelo explica aproximadamente 81-84% da variabilidade dos dados, mesmo em cenários não vistos durante o treinamento.

Notavelmente, o R^2 de 0.84 no conjunto de teste significa que o modelo é capaz de explicar 84% da variabilidade do nível do Rio Mundaú em dados inéditos, como demonstrado na Figura 2, confirmando sua robustez e potencial para aplicações práticas de previsão hidrológica na bacia estudada. Contudo, o aumento do RMSE na fase de teste (25,69) em relação ao treino e validação pode estar associado a eventos extremos ou variações sazonais não totalmente representadas no conjunto de treinamento.

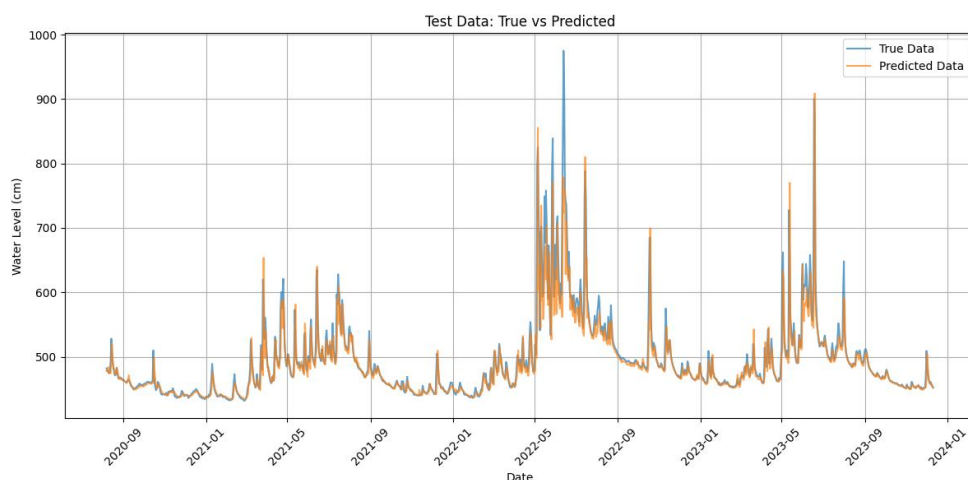


Figura 2. Comparação do nível da água (cm) previsto pelo modelo em relação aos valores verdadeiros.

4. Considerações Finais e Trabalhos Futuros

O presente estudo demonstrou a eficácia do modelo LSTM na previsão do nível do Rio Mundaú, obtendo um desempenho satisfatório conforme as métricas avaliadas. O modelo foi capaz de capturar padrões hidrológicos relevantes, alcançando um RMSE de 25,69 cm na fase de teste e um R^2 de 0,84, evidenciando sua aplicabilidade prática para alertas de cheias e gestão de recursos hídricos.

Entretanto, a ligeira elevação do RMSE na fase de teste sugere desafios na previsão de eventos extremos, o que pode ser aprimorado em estudos futuros. Dessa forma, propomos como trabalhos futuros: (i) a inclusão de novas variáveis ambientais, como temperatura e umidade do solo, para enriquecer o modelo preditivo; (ii) o emprego de arquiteturas híbridas que combinem LSTM com outros modelos estatísticos ou de machine learning; e (iii) a adaptação do modelo para previsões em múltiplos horizontes temporais, permitindo estimativas mais abrangentes e antecipadas de eventos críticos.

Referências

- Adli Zakaria, M. N., Ahmed, A. N., Abdul Malek, M., et al. (jul 2023). Exploring machine learning algorithms for accurate water level forecasting in Muda river, Malaysia. *Heliyon*, v. 9, n. 7, p. e17689.
- Chang, F.-J., Chen, P.-A., Lu, Y.-R., Huang, E. and Chang, K.-Y. (sep 2014). Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. *Journal of Hydrology*, v. 517, p. 836–846.
- IBGE (2022). IBGE | Cidades@ | Alagoas | Rio Largo | Pesquisa | Censo 2022 | Território. <https://cidades.ibge.gov.br/brasil/al/rio-largo/pesquisa/10102/122229>, [accessed on Apr 1].
- Virões, M. V. and Pinto, E. J. de A. (2020). Disponibilidade hídrica do Brasil: estudos de regionalização nas bacias hidrográficas brasileiras: análise de frequência de sistemas de alerta: Sistema de Alerta Bacia do Rio Mundaú, Rio Mundaú, Estação Fluviométrica União dos Palmares, código 39740000. CPRM. <http://rigeo.sgb.gov.br/jspui/handle/doc/21991>, [accessed on Apr 1].
- Zhihui Dai, Ming Zhang, Nadia Nedjah, Dong Xu, and FengYe (2023). A Hydrological Data Prediction Model Based on LSTM with Attention Mechanism. *Water* 2023, v. 15.