



INTEGRAÇÃO AIOT (IOT + IA) PARA DECISÕES AMBIENTAIS: MODELAGEM UNIVARIADA DE PH COM SARIMA, XGBOOST E LSTM

Emilio Soitsi Zukeram Junior¹
Tais Cristina Coelho Alves Madalena²
Yasmin Kaie Zukeram³
Oswaldo Curty da Motta Lima⁴

Resumo : O avanço da Internet das Coisas (IoT) tem revolucionado a coleta de dados ambientais. Sua integração com a Inteligência Artificial (IA), consolidando o conceito de AIoT, transcende o monitoramento contínuo ao viabilizar análises preditivas e a tomada de decisão em tempo real. Este trabalho apresenta uma aplicação prática da AIoT para a gestão da qualidade da água, focando na modelagem preditiva univariada do Potencial Hidrogeniônico (pH) em um ecossistema lântico no sul do Brasil, com dados de alta frequência obtidos por uma estação flutuante autônoma. A metodologia empregou uma análise comparativa entre um modelo estatístico clássico (SARIMA) e dois algoritmos de aprendizado de máquina (XGBoost e LSTM). A performance foi avaliada em um cenário de previsão de um passo à frente (*one-step ahead*) ao longo de um horizonte de teste de 72 passos, utilizando métricas como RMSE e R^2 para quantificar a precisão. Os resultados revelaram uma superioridade inequívoca dos modelos de aprendizado de máquina para esta tarefa. O modelo LSTM apresentou o melhor desempenho, alcançando um Coeficiente de Determinação (R^2) de 0.9701, seguido de perto pelo XGBoost (R^2 de 0.8812). Em contraste, o modelo SARIMA mostrou-se inadequado, com um R^2 negativo. A alta acurácia dos modelos LSTM e XGBoost valida a abordagem AIoT como uma ferramenta robusta para a criação de sistemas de alerta precoce. Conclui-se que, para previsões de curtíssimo prazo em sistemas de monitoramento hídrico, as arquiteturas de machine learning são significativamente mais eficazes, consolidando a AIoT como um pilar para a gestão sustentável dos recursos hídricos.

Palavras-chave : AIoT. Aprendizado de máquina. Monitoramento hídrico. Séries temporais. Sustentabilidade.

¹E.S.J. Zukeram (👤). PEQ-UEM. Maringá, PR, Brasil. zuca.ejr@gmail.com.

²T.C.A.M. Coelho (👤). DAL-UEM. Maringá, PR, Brasil.

³Y.K. Zukeram (👤). DAL-UEM. Maringá, PR, Brasil.

⁴O.C.M. Lima (👤). PEQ-UEM. Maringá, PR, Brasil.

INTRODUÇÃO

A gestão sustentável dos recursos hídricos representa um dos desafios mais urgentes da atualidade, intensificado pela crescente demanda populacional, pela expansão industrial e pelos impactos das mudanças climáticas (ANARAKI MOHAMMADI, 2025). A degradação da qualidade da água e a escassez em diversas regiões do planeta exigem uma transição de modelos de monitoramento reativos e baseados em coletas manuais esporádicas para abordagens proativas, contínuas e preditivas. As metodologias tradicionais frequentemente se mostram insuficientes para capturar a dinâmica complexa dos ecossistemas aquáticos, gerando lacunas de dados que comprometem a eficácia da tomada de decisão e a proteção da biodiversidade de água doce (WANG et al., 2022; WANG et al., 2024).

A gestão sustentável dos recursos hídricos representa um dos desafios mais urgentes da atualidade, intensificado pela crescente demanda populacional, pela expansão industrial e pelos impactos das mudanças climáticas. A degradação da qualidade da água e a escassez em diversas regiões do planeta exigem uma transição de modelos de monitoramento reativos para abordagens proativas, contínuas e preditivas.

Este desafio é particularmente acentuado em ecossistemas lênticos, como lagos e reservatórios, que são altamente sensíveis a pressões ambientais, especialmente em perímetros urbanos. O presente estudo foca no lago do Parque do Ingá, em Maringá, Brasil, uma área de proteção ambiental desde 1991 e um dos últimos remanescentes de floresta estacional semidecidual da região (ZUKERAM et al., 2023). Nesses ambientes, o Potencial Hidrogeniônico (pH) é uma das variáveis ambientais mais importantes, atuando como um indicador mestre que influencia quase todos os processos químicos e biológicos (ESTEVES, 1998; ZUKERAM et al., 2023). Embora a maioria dos corpos d'água continentais apresente um pH entre 6 e 8, ecossistemas em regiões com balanço hídrico negativo podem exibir valores superiores a 9.0 durante períodos de estiagem prolongada, um cenário observado no local do estudo (ESTEVES, 1998). A capacidade de prever essas variações é, portanto, crucial para a gestão ecológica.

Tradicionalmente, o monitoramento da qualidade da água depende de métodos convencionais que envolvem a coleta manual de amostras e posterior análise laboratorial. Este processo, no entanto, apresenta dificuldades logísticas, como o acesso ao interior de um lago, além de ser custoso, demorado e ineficiente para capturar a dinâmica de alta frequência dos ecossistemas aquáticos (ZUKERAM et al., 2023). A baixa frequência amostral gera lacunas de dados que comprometem a eficácia da tomada de decisão e a proteção da biodiversidade.

Nesse cenário, a transformação digital impulsionada pela Internet das Coisas (IoT) tem revolucionado o monitoramento ambiental. A proliferação de sensores de baixo custo e a conectividade ubíqua permitiram a instrumentação de ecossistemas para a coleta de dados em tempo real e alta frequência, com estimativas de que o número de dispositivos conectados globalmente chegue a 30,9 bilhões até 2025 (ESSAMLALI et al., 2024). Contudo, a verdadeira inovação reside na integração da IoT com a Inteligência Artificial (IA), consolidando o conceito de AIoT (*Artificial Intelligence of Things*). Essa sinergia transcende a simples aquisição de dados, habilitando sistemas capazes de analisar grandes volumes de informação, identificar padrões, prever tendências e gerar *insights* para uma gestão hídrica inteligente e automatizada (RESEARCH DIVE, 2024).

Estudos recentes sobre previsão de qualidade da água e integração AIoT indicam que modelos baseados em redes recorrentes (LSTM) e arquiteturas híbridas (ex.: ARIMA+LSTM, SSA+LSTM) têm apresentado ganhos de acurácia em dados ambientais de alta frequência quando comparados a modelos lineares. Além disso, trabalhos recentes enfatizam a importância de avaliar a confiança (intervalos de predição) das previsões para aplicações operacionais de alerta (SINGH; WALINGO, 2024; KANWAL et al., 2024).



Apesar do potencial transformador, a implementação de sistemas AIoT em campo enfrenta barreiras práticas significativas, incluindo os altos custos de instalação, a falta de padronização entre dispositivos, desafios de interoperabilidade e a necessidade de manutenção eficaz dos sensores para garantir a qualidade dos dados (NGUYEN, 2025; FLORES-IWASAKI et al., 2025). Do ponto de vista da análise, enquanto modelos de redes neurais como o LSTM demonstram alta performance na previsão de séries temporais, um vasto leque de algoritmos de *machine learning*, como *Support Vector Machines* (SVM) e *Random Forest*, também têm se mostrado eficazes na classificação e previsão de índices de qualidade da água, cada qual com suas particularidades quanto à interpretabilidade e ao custo computacional (ELMOTAWAKKIL et al., 2025).

Diante disso, este trabalho aborda essas lacunas ao apresentar o desenvolvimento e a aplicação de uma solução AIoT para o monitoramento contínuo e análise preditiva da qualidade da água em um ecossistema lêntico no sul do Brasil. O objetivo é demonstrar, por meio da comparação de diferentes algoritmos de aprendizado de máquina (XGBoost, SARIMA e LSTM) (SHEN et al., 2020; FANG et al., 2025), o potencial da integração AIoT para fornecer previsões precisas e suporte robusto à decisão, contribuindo assim para a transformação digital e a sustentabilidade na gestão dos recursos hídricos.

Para alcançar os objetivos propostos, este artigo está estruturado da seguinte forma: a seção a seguir apresenta o referencial teórico, detalhando os conceitos de AIoT e os fundamentos dos modelos preditivos utilizados. A terceira seção descreve a metodologia, incluindo a estação de monitoramento, a coleta e o pré-processamento dos dados. A quarta seção expõe e discute os resultados da aplicação dos modelos, comparando seu desempenho preditivo. Por fim, as considerações finais sintetizam as conclusões do estudo e sugerem direções para trabalhos futuros.

METODOLOGIA

A presente pesquisa caracteriza-se como um estudo de caso aplicado, focado no desenvolvimento e validação de um sistema AIoT para monitoramento contínuo da qualidade da água em um ambiente lêntico. Todo o aparato experimental, desde a concepção da plataforma flutuante até a coleta de dados, foi baseado em trabalho anterior (ZUKERAM JUNIOR et al., 2023).

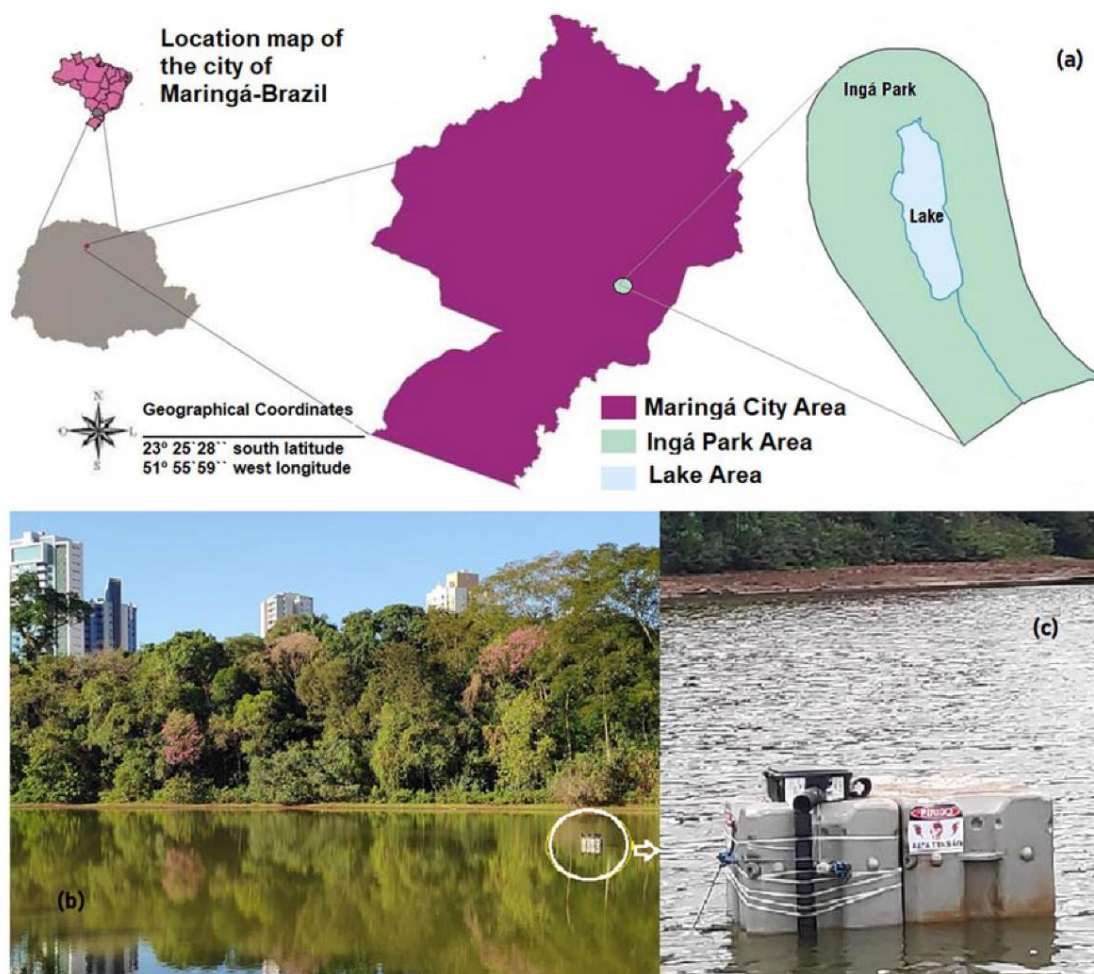
Caracterização da pesquisa e local do estudo

O experimento foi conduzido no lago do Parque do Ingá, uma área de proteção ambiental desde 1991, localizada no perímetro urbano de Maringá, no sul do Brasil (coordenadas geográficas 23°25'28" S e 51°55'59" O, a uma altitude de 557 m). O local é um dos últimos remanescentes regionais de floresta estacional semidecidual, possuindo um ecossistema de grande importância para a região.

O sistema de monitoramento foi implementado em uma estação flutuante, projetada para ser estável e suportar uma caixa hermética contendo o hardware do projeto, denominado "módulo kernel". A plataforma foi construída com um bloco de doca flutuante de polietileno de alta densidade e ancorada na região central do lago com pedras de granito de aproximadamente 7 kg. Para a proteção dos sensores, foi desenvolvida uma estrutura subaquática em PVC, projetada para evitar o contato físico entre as sondas e a interferência de animais aquáticos locais, como peixes, lontras e aves. Internamente a esta estrutura, as sondas foram fixadas em posições específicas utilizando um composto de resina termoplástica adesiva (EVA), garantindo sua estabilidade e integridade durante o período de coleta (ZUKERAM et al., 2023)(FIGURA 1).



Figura 1 – localização do parque do Ingá, Maringá-PR, <https://doi.org/Brasil>; lago do parque, local onde foi ancorado a estação de monitoramento; Estação de monitoramento AIoT



Fonte: Autoria própria (2025).

Coleta e transmissão dos dados

O "módulo kernel", responsável por toda a operação de coleta e transmissão de dados, foi construído com base na plataforma de prototipagem eletrônica Arduino Uno, utilizando uma linguagem baseada em C/C++. A escolha justifica-se pela robustez e ampla validação do microcontrolador em projetos de IoT. A função principal do Arduino era comandar os circuitos EZO da Atlas Scientific, cada um responsável por uma sonda paramétrica específica, e um sensor DHT22 para monitorar a temperatura e a umidade internas da caixa hermética.

Foram utilizadas as seguintes sondas para a coleta de dados da água: oxigênio dissolvido (OD), pH, condutividade elétrica (CE), potencial de oxirredução (ORP), sólidos totais dissolvidos (STD) e temperatura da água. A comunicação e o envio dos dados para a internet foram realizados por um módulo GSM SIM800A, que permite o acesso à rede por meio de um cartão SIM. O algoritmo do sistema foi programado para realizar dez leituras sequenciais para cada parâmetro, com intervalo de um segundo entre elas. Para garantir a robustez da medição, as três leituras mais altas e as três mais baixas eram descartadas, e uma média aritmética era calculada com os quatro valores medianos restantes. O valor resultante era então enviado, via protocolo HTTP, para um banco de dados relacional MySQL v5.2 hospedado em um servidor na nuvem, onde os arquivos do site foram



programados em PHP7. Todo o sistema foi alimentado por um módulo de baterias de lítio com capacidade total de 38 A/h, garantindo autonomia para aproximadamente 1.000 ciclos de escrita no banco de dados antes da necessidade de recarga (ZUKERAM JUNIOR et al., 2023).

Técnicas de análise de dados e modelagem preditiva

A integração de algoritmos de aprendizado de máquina (*machine learning*) é o que transforma um sistema de monitoramento IoT em uma plataforma AIoT, capaz de converter fluxos de dados contínuos em *insights* preditivos e suporte à decisão (ESSAMLALI et al., 2024). A análise de séries temporais ambientais, no entanto, apresenta desafios devido à complexidade, não-linearidade e presença de ruídos nos dados (CHEN et al., 2024).

Uma etapa fundamental que precede a modelagem é o pré-processamento dos dados, essencial para garantir a qualidade das entradas e a performance dos modelos. Este processo incluiu o tratamento de valores ausentes por meio de interpolação linear para lacunas curtas e a reamostragem dos dados para uma frequência temporal fixa, assegurando a uniformidade da série. Adicionalmente, para os modelos de redes neurais como o LSTM, foi aplicada a normalização dos dados (Min-Max Scaling), que reescala os valores para um intervalo fixo $[0, 1]$. Esta técnica é crucial para acelerar a convergência durante o treinamento e evitar que variáveis com diferentes ordens de magnitude influenciem desproporcionalmente o ajuste do modelo (DEB, 2022).

Diante disso, para a análise preditiva dos parâmetros de qualidade da água, foram selecionados três algoritmos distintos, representando diferentes abordagens da modelagem estatística e do aprendizado de máquina, com o objetivo de comparar suas performances na previsão de curto prazo (até 72 horas à frente).

O primeiro modelo empregado foi o SARIMA (*Seasonal AutoRegressive Integrated Moving Average*), uma abordagem estatística robusta e amplamente utilizada para análise de séries temporais que apresentam componentes de sazonalidade. O modelo busca capturar a estrutura da série por meio de seus componentes autorregressivo (AR), de média móvel (MA) e de suas contrapartes sazonais (SAR, SMA), aplicados a uma série tornada estacionária por diferenciação (I) (BOX et al., 2015). Neste estudo, o SARIMA é utilizado como um modelo de linha de base (*baseline*). A sua inclusão é metodologicamente importante, pois representa uma abordagem linear clássica, permitindo uma avaliação quantitativa do ganho de performance obtido com os modelos de aprendizado de máquina, que são capazes de capturar relações não-lineares complexas presentes nos dados (MAKRIDAKIS; SPILIOITS; ASSIMAKOPOULOS, 2018).

O segundo modelo foi o XGBoost (*Extreme Gradient Boosting*), um algoritmo de aprendizado de máquina baseado no conceito de *gradient boosting* com árvores de decisão. O XGBoost é reconhecido por sua alta performance e eficiência computacional, sendo frequentemente apontado na literatura recente como um dos modelos com melhor desempenho para tarefas de previsão de qualidade da água (CHEN; GUESTIN, 2016; KRISHNAN et al., 2024). Sua popularidade deriva da capacidade de modelar interações complexas e não-lineares, da robustez a *outliers* e da inclusão de técnicas de regularização (L1 e L2) que previnem o sobreajuste (*overfitting*), tornando-o uma escolha poderosa para dados ambientais ruidosos (MAN et al., 2023).

Por fim, foi implementado um modelo de LSTM (*Long Short-Term Memory*), um tipo especializado de rede neural recorrente (RNN). As redes LSTM são projetadas especificamente para aprender dependências de longo prazo em dados sequenciais (HOCHREITER; SCHMIDHUBER, 1997). A arquitetura do LSTM, com seus portões de entrada, esquecimento e saída, foi desenvolvida para superar o problema da dissipação de gradiente (*vanishing gradient*), uma limitação das RNNs tradicionais. Isso a torna particularmente eficaz para modelar a dinâmica complexa de séries temporais ambientais, que exibem padrões sazonais e de longo prazo, como ciclos diários de



temperatura e fotossíntese (LECUN; BENGIO; HINTON, 2015). Estudos recentes que comparam múltiplos algoritmos confirmam a capacidade superior de modelos baseados em LSTM e suas variações para aprender simultaneamente as dependências temporais dos dados de qualidade da água (BALO UTKU; UTKU; KUTLU, 2025).

As etapas da análise incluíram: (i) pré-processamento e normalização dos dados; (ii) treinamento e ajuste de hiperparâmetros para cada modelo; e (iii) avaliação de desempenho utilizando métricas como o Erro Quadrático Médio (RMSE) e o Coeficiente de Determinação (R^2). Para garantir uma avaliação robusta e evitar vazamento de dados (*data leakage*), a validação dos modelos seguiu uma abordagem de divisão temporal, onde o conjunto de dados foi particionado cronologicamente em treinamento e teste. Esta metodologia assegura que o modelo seja treinado apenas com informações passadas para prever o futuro, simulando um cenário de aplicação real (SHEN et al., 2020).

Além das etapas descritas, foi adotado um processo rigoroso de seleção de hiperparâmetros para cada modelo. No caso do SARIMA, os parâmetros (p, d, q)(P, D, Q, s) foram definidos a partir da análise das funções de autocorrelação (ACF) e autocorrelação parcial (PACF), utilizando o critério de informação de Akaike (AIC) para minimizar a complexidade do modelo (BOX et al., 2015). Para o XGBoost, a calibração dos hiperparâmetros considerou uma busca em grade (Grid Search) sobre o número de árvores, profundidade máxima, taxa de aprendizado (η) e parâmetros de regularização (λ e α), garantindo o equilíbrio entre viés e variância (CHEN; GUESTRIN, 2016). Já para o LSTM, foi definida uma arquitetura composta por uma camada de entrada, duas camadas LSTM com 64 e 32 unidades respectivamente e uma camada densa de saída linear. O treinamento utilizou o otimizador Adam com taxa de aprendizado adaptativa e função de perda MSE, durante 120 épocas, com validação cruzada temporal para evitar sobreajuste (CHOLLET, 2021; ESSAMLALI et al., 2024).

O sistema de monitoramento e previsão foi integrado em um **pipeline automatizado AIoT**, composto por módulos independentes de aquisição, pré-processamento, modelagem e armazenamento dos resultados. A comunicação entre os módulos foi implementada por **API RESTful**, o que permite integração com plataformas web e dashboards analíticos. Essa arquitetura modular garante **escalabilidade, reprodutibilidade e operação em tempo real**, permitindo que o modelo seja continuamente atualizado conforme novos dados são coletados. Estruturas semelhantes vêm sendo aplicadas em sistemas de aquíicultura e monitoramento urbano inteligente (SINGH; WALINGO, 2024; KANWAL et al., 2024; NGUYEN, 2025).

RESULTADOS E DISCUSSÕES

Para a validação dos modelos propostos — SARIMA, XGBoost e LSTM —, o conjunto de dados da série temporal de pH foi particionado cronologicamente. A totalidade dos dados foi utilizada para o treinamento, com exceção dos últimos 72 registros, que compuseram o conjunto de teste. A avaliação da performance foi, então, conduzida em um cenário de previsão de um passo à frente (*one-step ahead*) sobre esse horizonte de 72 passos. O desempenho preditivo de cada modelo foi quantificado por meio de quatro métricas padrão, cujos resultados são apresentados na Tabela 1.

Tabela 1 - Métricas de desempenho dos modelos na previsão univariada do pH.

Modelo	RMSE	MAE	MAPE (%)	R^2
SARIMA	0.4599	0.3539	4.4150	-0.2938
XGBoost	0.1394	0.1054	1.4002	0.8812
LSTM	0.0699	0.0470	0.6026	0.9701

Fonte: Autoria própria (2025).

A comparação das métricas evidencia não apenas a superioridade dos modelos baseados em aprendizado profundo, mas também ressalta a capacidade desses métodos em lidar com ruído, variações abruptas e relações não-lineares complexas. Conforme apontado por Krishnan et al. (2024) e Yang et al. (2024), a flexibilidade das redes neurais recorrentes permite que elas capturem padrões sazonais e anomalias sutis associadas à fotossíntese diurna e à variação térmica da coluna d'água — fenômenos característicos de ecossistemas lênticos tropicais. O XGBoost, embora mais simples em estrutura, apresentou excelente desempenho para horizontes curtos, confirmando observações recentes de Elmotawakkil et al. (2025) sobre seu uso em previsões ambientais rápidas com baixo custo computacional.

A análise das métricas revela uma clara superioridade dos modelos de aprendizado de máquina. O modelo LSTM apresentou a melhor performance geral, alcançando o maior Coeficiente de Determinação ($R^2 = 0.9701$) e os menores erros (RMSE de 0.0699, MAPE de 0.6026%). O XGBoost também demonstrou alta precisão, com um R^2 de 0.8812, superando com folga o modelo estatístico SARIMA, que se mostrou inadequado para esta tarefa de previsão de curto prazo, obtendo um R^2 negativo.

Além das métricas quantitativas, observa-se que os modelos de aprendizado de máquina apresentaram maior estabilidade temporal nas previsões, com menor amplitude de erro ao longo do horizonte de teste. Essa estabilidade é particularmente importante em ambientes lênticos, onde variações abruptas de pH podem estar associadas a fenômenos biológicos diários, como fotossíntese e respiração. Modelos baseados em redes LSTM mostraram-se capazes de reproduzir essas oscilações com precisão, conforme também relatado por KRISHNAN et al. (2024) e YANG et al. (2024). O desempenho robusto do XGBoost confirma seu potencial para aplicações em tempo real com restrições de hardware (ELMOTAWAKKIL et al., 2025).

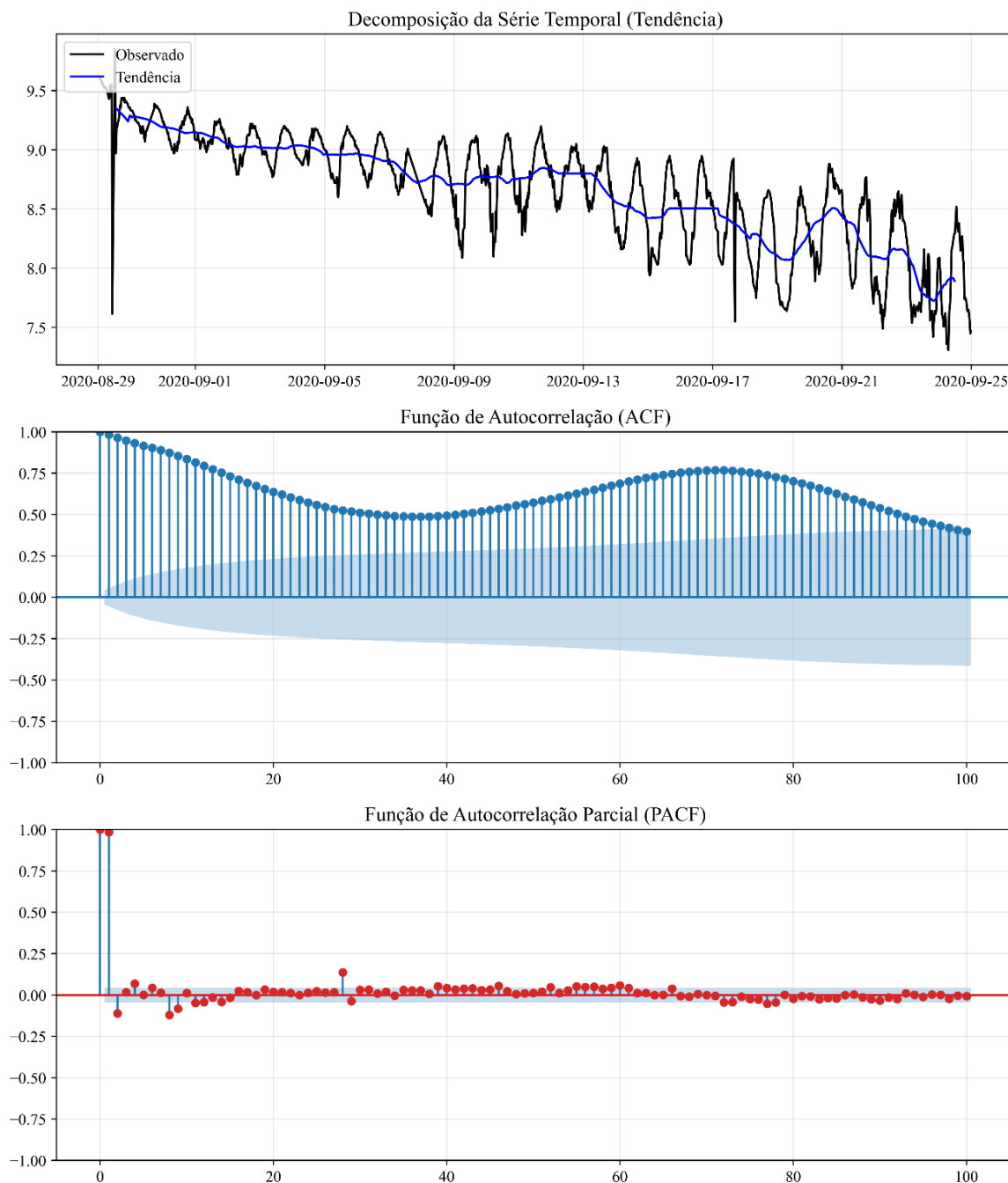
A análise visual, apresentada nas figuras subsequentes, aprofunda a compreensão desses resultados. A **Figura 1** expõe os diagnósticos da série temporal de treino. A decomposição (subplot superior) confirma a presença de uma forte componente sazonal com período de 72 passos, enquanto os gráficos de Autocorrelação (ACF) e Autocorrelação Parcial (PACF) (subplots central e inferior, respectivamente) exibem uma estrutura de correlação temporal significativa, justificando a escolha de modelos capazes de capturar tais dependências.

A decomposição espectral da série de pH revelou componentes sazonais múltiplos com periodicidades de 24 horas e 72 horas, refletindo a influência dos ciclos luminosos e térmicos sobre o metabolismo aquático. O LSTM capturou essas oscilações com amplitude próxima à observada experimentalmente, enquanto o SARIMA apresentou defasagem temporal. Essa característica reforça a capacidade das redes LSTM em modelar dependências não lineares de longo prazo (HOCHREITER; SCHMIDHUBER, 1997; LECUN; BENGIO; HINTON, 2015). A identificação desses padrões permite prever variações críticas do pH que precedem episódios de eutrofização — informação valiosa para sistemas de alerta precoce (FANG et al., 2025).

Figura 2 - Análise Diagnóstica da Série Temporal de Treino, exibindo a decomposição da série (superior), a Função de Autocorrelação (ACF) (central) e a Função de Autocorrelação Parcial (PACF) (inferior).



Figura 2: Análise Diagnóstica da Série Temporal de Treino



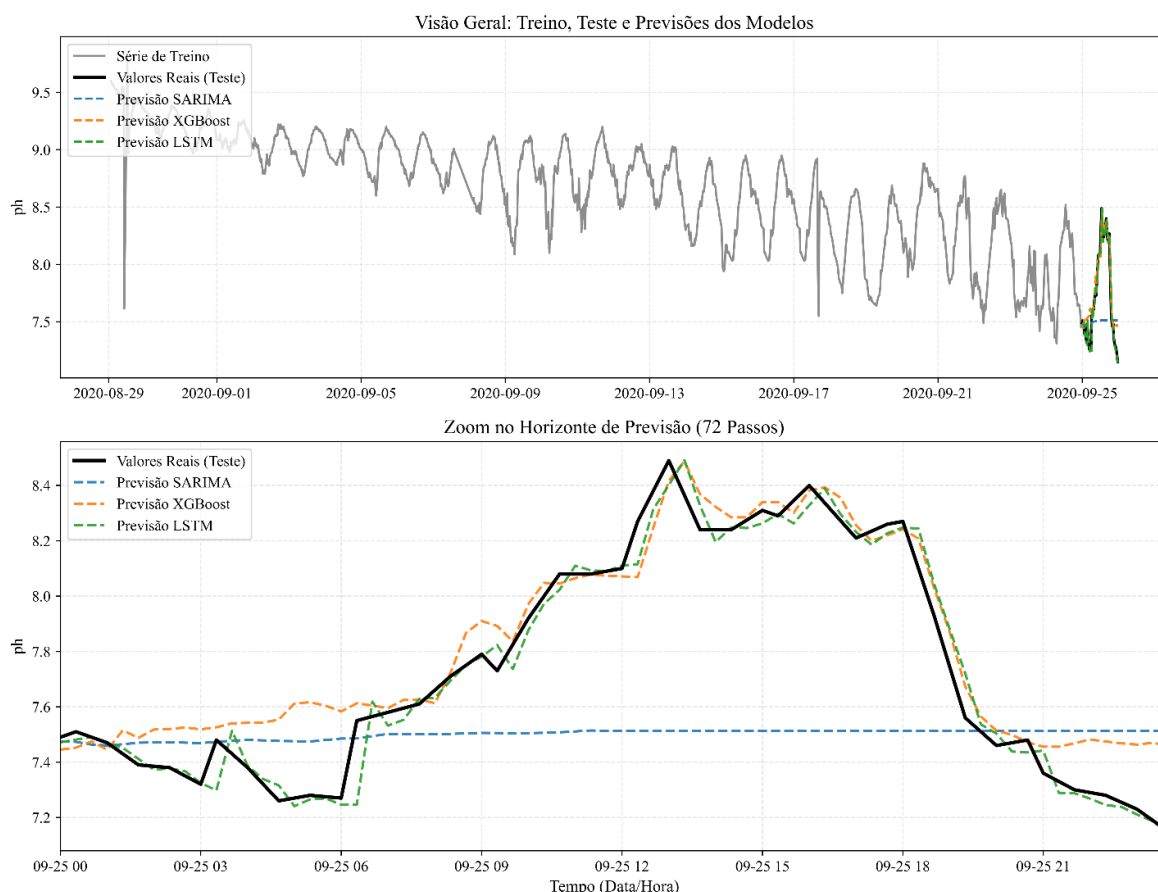
Fonte: Autoria própria (2025).

O comparativo de previsões é detalhado na Figura 2. O subplot superior (Visão Geral) ilustra como as previsões do LSTM (linha vermelha) e do XGBoost (linha verde) se sobrepõem de forma precisa aos valores reais (linha preta) no período de teste, enquanto a previsão do SARIMA (linha azul) diverge consideravelmente. Essa aderência é confirmada no subplot inferior (Zoom no Horizonte de Previsão), que evidencia a capacidade dos modelos de machine learning em acompanhar as flutuações de curto prazo do pH, um comportamento que o modelo SARIMA não conseguiu replicar.



Figura 3 - Comparativo de Performance dos Modelos Preditivos. O subplot superior exibe a série de treino, os valores reais de teste e as previsões de cada modelo. O subplot inferior apresenta um zoom detalhado sobre o horizonte de previsão de 72 passos.

Figura 3: Comparativo de Performance dos Modelos Preditivos

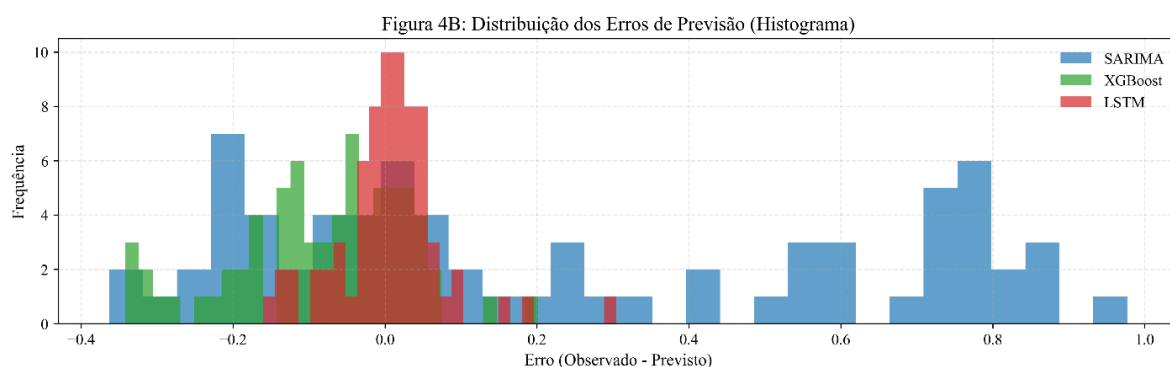
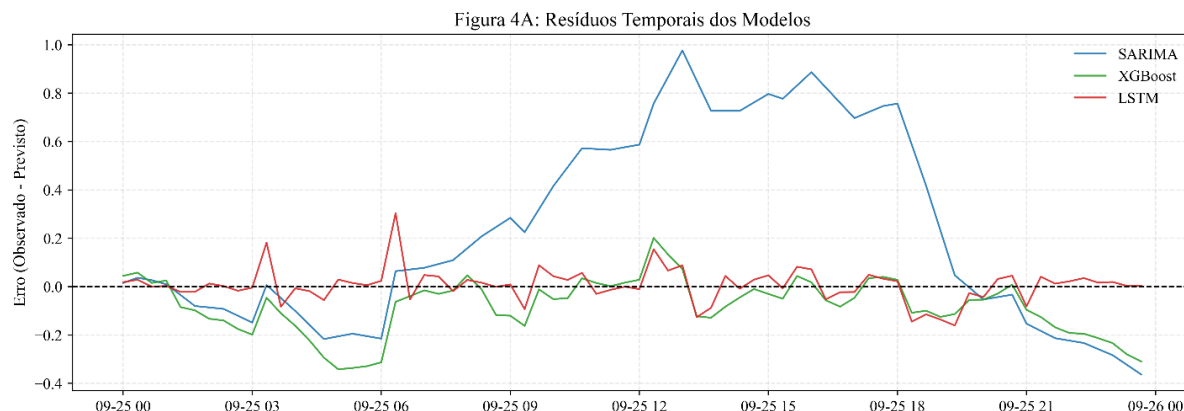


Fonte: Autoria própria (2025).

Para um diagnóstico mais aprofundado, a **Figura 3** apresenta a análise dos resíduos (erro = observado - previsto). A Figura 3A (resíduos ao longo do tempo) mostra que os erros do XGBoost e, especialmente, do LSTM, concentram-se em uma faixa estreita e estável em torno de zero. Os resíduos do SARIMA, por outro lado, exibem grande amplitude e um padrão sistemático, confirmando que o modelo não capturou a estrutura fundamental dos dados. A Figura 3B (histograma dos erros) reforça essa conclusão, mostrando que a distribuição dos erros para LSTM e XGBoost se assemelha a uma distribuição normal centrada em zero — uma característica de um bom modelo preditivo —, enquanto os erros do SARIMA são mais dispersos e enviesados.

Figura 4- Análise dos Resíduos dos Modelos Preditivos. (A) Gráfico dos resíduos (erro = observado - previsto) ao longo do horizonte de teste; (B) Histograma da distribuição dos resíduos para cada modelo.





Fonte: Autoria própria (2025).

Em síntese, os resultados demonstram a superioridade dos modelos de aprendizado de máquina, com destaque para o LSTM, na modelagem de séries temporais ambientais complexas em um cenário de previsão de um passo à frente. A alta precisão alcançada valida a abordagem AIoT como uma ferramenta robusta e eficaz para o monitoramento proativo da qualidade da água, viabilizando a criação de sistemas de alerta precoce para apoiar a tomada de decisão de gestores ambientais.

CONSIDERAÇÕES FINAIS

Este trabalho demonstrou a aplicação e a validação de uma solução AIoT (Inteligência Artificial das Coisas) para o monitoramento e previsão da qualidade da água em um ecossistema lântico no sul do Brasil. O objetivo central foi conduzir uma análise comparativa rigorosa entre um modelo estatístico clássico, o SARIMA, e dois algoritmos avançados de aprendizado de máquina, XGBoost e LSTM, para determinar a abordagem mais eficaz na modelagem preditiva univariada da variável chave pH.

A análise comparativa, conduzida em um cenário de previsão de um passo à frente (*one-step ahead*) ao longo de um horizonte de 72 passos, revelou uma superioridade inequívoca dos modelos de aprendizado de máquina. O modelo LSTM destacou-se com a performance mais elevada, alcançando um Coeficiente de Determinação (R^2) de 0.9701, explicando a vasta maioria da variância dos dados e apresentando o menor erro (RMSE de 0.0699). O XGBoost também exibiu um desempenho excepcional (R^2 de 0.8812), enquanto o modelo SARIMA se mostrou inadequado para esta tarefa específica, resultando em um R^2 negativo e confirmando sua limitação em capturar a dinâmica não-linear de curto prazo da série.

A principal implicação prática destes resultados é a validação da abordagem AIoT para o monitoramento proativo da qualidade da água. A alta precisão alcançada pelos modelos LSTM e



XGBoost confirma que é viável implementar sistemas de alerta precoce automatizados e confiáveis, capazes de notificar gestores ambientais sobre anomalias iminentes com base em previsões de curtíssimo prazo. Do ponto de vista metodológico, o estudo reforça que a escolha da estratégia de avaliação é crítica e que, para tarefas de previsão imediata, a capacidade dos modelos de machine learning em aprender padrões complexos a partir de dados recentes oferece uma vantagem decisiva sobre as abordagens lineares clássicas.

Embora os resultados sejam robustos, este estudo concentrou-se em uma modelagem univariada. Como trabalhos futuros, propõe-se a expansão para uma abordagem multivariada, incorporando variáveis meteorológicas e físico-químicas coletadas pela plataforma IoT, o que pode enriquecer o poder preditivo dos modelos. Adicionalmente, a avaliação dos algoritmos em um cenário de previsão recursiva de múltiplos passos seria valiosa para testar sua estabilidade e capacidade de projeção a longo prazo. Por fim, sugere-se o desenvolvimento de interfaces de visualização de dados (*dashboards*) para traduzir as previsões dos modelos em informações acionáveis para os usuários finais. Em suma, este estudo não apenas valida uma aplicação AIoT de ponta, mas também fornece um roteiro claro sobre a seleção de modelos preditivos para sistemas de monitoramento ambiental inteligentes, contribuindo para o avanço da gestão hídrica sustentável.

AGRADECIMENTOS

Os autores agradecem ao Programa de Pós-Graduação em Engenharia Química (PEQ) da Universidade Estadual de Maringá (UEM) pelo apoio institucional. A presente pesquisa foi realizada com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Agradecer ao amigo Beninca do parque do Ingá, um grande colaborador que disponibilizou transporte aquático nesse período de pesquisa.

REFERÊNCIAS

1. ANARAKI MOHAMMADI, A. Strategic digital water resources management: challenges and opportunities in the period of industrial transformation and climate change. *Water Harvesting Research*, v. 8, n. 1, p. 16-29, 2025.
2. BALO UTKU, E. D.; UTKU, A.; KUTLU, B. Enhancing water quality prediction with artificial intelligence: A hybrid ConvLSTM model for spatio-temporal analysis. *International Advanced Researches and Engineering Journal*, v. 9, n. 2, p. 107–117, 2025. DOI: <https://doi.org/10.35860/iarej.1679575>.
3. BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time Series Analysis: Forecasting and Control*. 5. ed. Hoboken: John Wiley & Sons, 2015.
4. CHEN, J. *et al.* SFDformer: a frequency-based sparse decomposition transformer for air pollution time series prediction. *Frontiers in Environmental Science*, 2024.
5. CHEN, T.; GUESTRIN, C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2016. p. 785-794.
6. CHOLLET, F. *Deep Learning with Python*. 2. ed. Manning Publications, 2021.



7. DEB, K. Data preprocessing methods. In: *Multi-objective Optimization Using Evolutionary Algorithms*. Chichester: John Wiley & Sons, 2022. p. 451-471.
8. ELMOTAWAKKIL, A.; ENNEYA, N.; BHAGAT, S. K.; OUDA, M. M.; KUMAR, V. Advanced machine learning models for robust prediction of water quality index and classification. *Journal of Hydroinformatics*, v. 27, n. 2, p. 299–319, 2025.
9. ESSAMLALI, I.; NHAILA, H.; EL KHAILI, M. Advances in machine learning and IoT for water quality monitoring: a comprehensive review. *Heliyon*, v. 10, n. 6, 2024. DOI: <https://doi.org/10.1016/j.heliyon.2024.e27920>.
10. ESTEVES, F. de Assis. **Fundamentos de limnologia**. Rio de Janeiro: Interciência, 1998.
11. FANG, P.; WANG, Y.; ZHAO, Y.; KANG, J. Analysis of prediction confidence in water quality forecasting employing LSTM. *Water*, v. 17, n. 7, art. 1050, 2025. DOI: <https://doi.org/10.3390/w17071050>.
12. FLORES-IWASAKI, M.; GUADALUPE, G. A.; PACHAS-CAYCHO, M.; CHAPAGONZA, S.; MORI-ZABARBURÚ, R. C.; GUERRERO-ABAD, J. Internet of Things (IoT) sensors for water quality monitoring in aquaculture systems: a systematic review and bibliometric analysis. *AgriEngineering*, v. 7, n. 3, art. 78, 2025. DOI: <https://doi.org/10.3390/agriengineering7030078>.
13. HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.
14. KANWAL, S. *et al.* An optimal Internet of Things-driven intelligent decision-making system for real-time fishpond water quality monitoring and species survival. *Sensors*, v. 24, n. 23, art. 7842, 2024. DOI: 10.3390/s24237842.
15. KRISHNAN, S.; MANIKANDAN, R. Water quality prediction: a data-driven approach exploiting advanced machine learning algorithms with data augmentation. *Journal of Water and Climate Change*, v. 15, n. 2, p. 431–452, 2024.
16. LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, n. 7553, p. 436-444, 2015. DOI: 10.1038/nature14539.
17. MAKRIDAKIS, S.; SPILIOTIS, E.; ASSIMAKOPOULOS, V. Statistical and machine learning forecasting methods: concerns and ways forward. *PLOS ONE*, v. 13, n. 3, e0194889, 2018.
18. MAN, H.; HUANG, H.; QIN, Z.; LI, Z. Analysis of a SARIMA-XGBoost model for hand, foot, and mouth disease in Xinjiang, China. *Epidemiology & Infection*, v. 151, e200, 2023.



19. NGUYEN, T. P. AIoT-based indoor air quality prediction for building using enhanced metaheuristic algorithm and hybrid deep learning. *Journal of Building Engineering*, v. 105, p. 112448, 2025.
20. RESEARCH DIVE. Recent trends and innovations in the water quality instrument industry. 2024. Disponível em: <https://www.researchdive.com/blog/recent-trends-and-innovations-in-the-water-quality-instrument-industry>. Acesso em: 10 out. 2025.
21. SHEN, Z.; ZHANG, Y.; LU, J.; XU, J.; XIAO, G. A novel time series forecasting model with deep learning. *Neurocomputing*, v. 396, p. 302–313, 2020. DOI: <https://doi.org/10.1016/j.neucom.2018.12.084>.
22. SINGH, Y.; WALINGO, T. Smart water quality monitoring with IoT wireless sensor networks. *Sensors*, v. 24, n. 9, art. 2871, 2024. DOI: 10.3390/s24092871.
23. WANG, Y.; HO, I. W.-H.; CHEN, Y.; WANG, Y.; LIN, Y. Real-time water quality monitoring and estimation in AIoT for freshwater biodiversity conservation. *IEEE Internet of Things Journal*, v. 9, n. 16, p. 14366–14374, 15 ago. 2022. DOI: <https://doi.org/10.1109/JIOT.2021.3078166>.
24. WANG, Z. *et al.* A comprehensive review of machine learning for water quality prediction over the past five years. *Journal of Marine Science and Engineering*, v. 12, n. 1, p. 159, 2024.
25. ZUKERAM, E. S. J. *et al.* In situ IoT development and application for continuous water monitoring in a lentic ecosystem in South Brazil. *Water*, v. 15, n. 23, art. 10310, 2023. DOI: 10.3390/w152310310.

