

MONITORAMENTO INTELIGENTE DA QUALIDADE DA ÁGUA: PREDIÇÃO DE ALCALINIDADE NA BACIA DO RIO GUANDU UTILIZANDO TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

Diogo Cabral Zaror¹; Leonard Barreto Moreira²; Leila Weitzel Coelho da Silva³

Abstract: This study aimed to predict alkalinity levels in raw water samples collected at an intake point in the Guandu River Basin (Rio de Janeiro, Brazil). Physicochemical data were analyzed at 2-hour intervals between January 2020 and April 2023, totaling 14,579 samples. The methodology employed deep learning models—GRU, TFT, NHITS, and KAN—evaluated using MAE, sMAPE, and RMSE metrics. The TFT model outperformed others, achieving the lowest errors: MAE (1.129), sMAPE (2.7%), and RMSE (1.932), compared to NHITS (MAE=1.481), GRU (MAE=1.274), and KAN (MAE=1.234). Training and testing plots confirmed high prediction accuracy, validating its efficacy. Results demonstrate that TFT is the most suitable approach for alkalinity prediction in this context, with potential applications in environmental monitoring and water resource management.

Resumo: Este estudo teve como objetivo prever os níveis de alcalinidade em amostras de água bruta coletadas em um ponto de captação na bacia do Rio Guandu (RJ). Foram analisados dados físico-químicos coletados a cada 2 horas entre janeiro de 2020 e abril de 2023, totalizando 14.579 amostras. A metodologia empregou modelos de aprendizagem profunda (*deep learning*) — GRU, TFT, NHITS e KAN—avaliados pelas métricas MAE, sMAPE e RMSE. O modelo TFT destacou-se com os melhores resultados: MAE de 1,129, sMAPE de 2,7% e RMSE de 1,932, superando os demais (NHITS: MAE=1,481; GRU: MAE=1,274; KAN: MAE=1,234). Os gráficos de treinamento e teste confirmaram alta precisão nas previsões, validando sua eficácia. Concluiu-se que o TFT é a abordagem mais adequada para prever alcalinidade nesse contexto, oferecendo potencial aplicação em monitoramento ambiental e gestão hídrica.

Palavras-Chave – Previsão de alcalinidade; Aprendizagem profunda; Bacia do Rio Guandu.

INTRODUÇÃO

A água é um recurso fundamental para a manutenção da vida, desempenhando um papel vital no consumo humano, nas atividades de lazer, na irrigação agrícola e em diversos outros setores. O Brasil, reconhecido por sua abundância hídrica, enfrenta desafios significativos relacionados à degradação dos recursos aquáticos em algumas regiões (de Castro, 2022). Mudanças na disponibilidade, qualidade e quantidade de água têm impactos profundos não apenas no equilíbrio dos ecossistemas, mas também na sobrevivência de todas as formas de vida. Nesse cenário, a conservação dos recursos hídricos emerge como uma prioridade fundamental, não apenas para a sustentabilidade ambiental, mas também para o desenvolvimento econômico e social, garantindo que as gerações futuras possam usufruir desse bem essencial.

No Brasil, a gestão de recursos hídricos é compartilhada entre órgãos federais e estaduais. A Agência Nacional de Águas e Saneamento Básico (ANA) regula o uso da água e implementa a Política Nacional de Recursos Hídricos; O Conselho Nacional do Meio Ambiente (CONAMA)

1) Universidade Federal Fluminense, R. Jardim Bela Vista - Jardim Bela Vista, Rio das Ostras - RJ. +55 22 998239026, diogozaror@id.uff.br

2) Universidade Federal Fluminense, R. Jardim Bela Vista - Jardim Bela Vista, Rio das Ostras - RJ. +55 22 988077534, leonardbarreto@id.uff.br

3) Universidade Federal Fluminense, R. Jardim Bela Vista - Jardim Bela Vista, Rio das Ostras - RJ. +55 22 999058035. leila_weitzel@id.uff.br

estabelece diretrizes ambientais, incluindo os padrões de enquadramento dos corpos d'água definidos na Resolução CONAMA nº 357/2005; e o Instituto Estadual do Ambiente (INEA) fiscaliza e monitora a qualidade da água no estado do Rio de Janeiro. Esses atores, embora principais, atuam em conjunto com comitês de bacia, como o Comitê de Bacia Hidrográfica do Rio Guandu (CBH Guandu), e entidades municipais para garantir a gestão integrada e sustentável das águas.

A previsão da qualidade da água tornou-se uma área de pesquisa essencial devido à crescente demanda por recursos hídricos frente aos desafios do crescimento urbano, geração de resíduos e mudanças climáticas. O uso de tecnologias avançadas para monitoramento e predição é fundamental para a gestão sustentável desses recursos (J. Luo et al., 2024; Yao et al., 2025). Embora modelos estatísticos tradicionais, como regressão linear, Autoregressivo (AR - *AutoRegressive*) e Autoregressivo Integrado de Médias Móveis (ARIMA - *AutoRegressive Integrated Moving Average*), sejam utilizados, eles apresentam limitações diante da complexidade dos dados. Por isso, métodos baseados em aprendizado de máquina, como Redes Neurais Artificiais (ANN - *Artificial Neural Networks*), Máquina de Vetores de Suporte (SVM - *Support Vector Machine*) e Árvores de Decisão (DT - *Decision Trees*), têm se destacado por sua maior capacidade de capturar padrões não lineares, embora ainda enfrentem desafios em previsões de longo prazo (W. Luo et al., 2024).

Estudos recentes têm destacado a importância da predição de parâmetros físico-químicos da água, como a alcalinidade, em corpos hídricos como rios, córregos, reservatórios, lagos e lagoas. A alcalinidade é um parâmetro crítico, pois está diretamente relacionada à capacidade da água de neutralizar ácidos e manter a estabilidade do pH. Essa característica influencia a eficiência dos processos de tratamento, além de assegurar a qualidade da água para consumo humano e o equilíbrio dos ecossistemas aquáticos (Guerra & Silva, 2018). Nesse contexto, modelos baseados em aprendizagem profunda (*deep learning*), como Redes Neurais Recorrentes (*Recurrent Neural Networks* - RNN), Memória de Longo Curto Prazo (LSTM - *Long Short-Term Memory*) e Unidade Recorrente com Portões (GRU - *Gated Recurrent Units*), têm se destacado por sua capacidade de capturar padrões temporais complexos. Além disso, abordagens híbridas, como CNN-LSTM e BiLSTM-GRU, vêm sendo aplicadas com sucesso, aprimorando a precisão das previsões ao explorar correlações entre múltiplos parâmetros da qualidade da água (Prasad et al., 2022; Song & Yao, 2022; Zhang et al., 2023).

Diante do contexto apresentado, o principal objetivo deste estudo foi analisar os parâmetros físico-químicos de amostras de água bruta coletadas em um dado ponto de coleta ao longo da bacia hidrográfica do Rio Guandu, com o intuito de prever, por meio de técnicas de inteligência artificial, os níveis de alcalinidade no ponto de captação.

METODOLOGIA

Área de estudo

Os dados deste estudo foram coletados na Bacia do Guandu, no estado do Rio de Janeiro, que abrange uma área de 1.385 km² e é drenada pelo rio Guandu, cuja origem está no Ribeirão das Lajes, passando a ser denominado Guandu após se unir ao rio Santana. Seus principais afluentes incluem os rios dos Macacos, Santana, São Pedro, Poços, Queimados e Ipiranga. O rio se divide em dois braços com barragens operadas pela CEDAE e conecta-se à lagoa do Guandu, seguindo até a Baía de Sepetiba pelo Canal do São Francisco. Parte significativa de sua vazão provém da transposição do rio Paraíba do Sul, viabilizada pela usina hidrelétrica da *Light*³, que contribui com cerca de 60% do volume. O Guandu é a principal fonte de abastecimento de água para mais de 9 milhões de pessoas

³ Empresa responsável pela distribuição de energia elétrica na cidade do Rio de Janeiro e em municípios da região metropolitana.

na região metropolitana do Rio de Janeiro, o que torna sua preservação e monitoramento fundamentais.

Característica do conjunto de dados

No presente estudo, foram avaliados diversos tipos de dados físico-químicos, com o objetivo de identificar padrões e possíveis anomalias com relação ao parâmetro objeto da pesquisa. Para isso, 45 variáveis foram analisadas no ponto de coleta 22. Essas variáveis abrangem uma ampla gama de parâmetros, como pH, turbidez, oxigênio dissolvido, condutividade elétrica, temperatura da água, entre outros, coletados em diferentes frequências e condições. A escolha do ponto 22 deve-se à sua representatividade na bacia estudada, sendo um local estratégico para o monitoramento da qualidade da água que abastece a Estação de Tratamento de Água (ETA) do Guandu, operada pela CEDAE. Os dados foram coletados a cada 2 horas entre janeiro de 2020 e abril de 2023, totalizando 14.579 instâncias de dados.

Seleção, análise e tratamento dos dados

Neste estudo, foram adotados dois critérios para a exclusão de parâmetros: baixa frequência de coleta (menos de uma medição a cada duas horas) e alta taxa de dados faltantes ($\geq 40\%$), visando garantir a qualidade e a robustez das análises. Conforme os critérios estabelecidos, os demais pontos de monitoramento foram excluídos por apresentarem frequência de coleta insuficiente (≤ 4 observações/semana) e elevada taxa de dados faltantes ($> 40\%$).

A Análise Exploratória de Dados (EDA— *Exploratory Data Analysis*) incluiu estatísticas descritivas, como média, desvio padrão, valores extremos e percentual de dados faltantes, além da análise visual das séries por meio de gráficos, *boxplots* sazonais e decomposição em componentes de tendência, sazonalidade e ruído. Avaliaram-se a estacionariedade, por meio de testes como o *Dickey-Fuller*, e a dependência temporal, com as funções de autocorrelação (ACF – *Autocorrelation Function*) e autocorrelação parcial (PACF – *Partial Autocorrelation Function*). Também foram analisadas possíveis correlações cruzadas entre variáveis hidrológicas e eventuais mudanças estruturais nas séries. A análise de correlação foi realizada usando os coeficientes de *Spearman*, adequado para relações monotônicas, e o *Maximal Information Coefficient* (MIC), capaz de capturar associações não lineares e complexas, proporcionando uma visão abrangente das interações entre os parâmetros hidrológicos (Reshef et al. 2011). Também foi verificada a conformidade dos parâmetros com a Resolução CONAMA 357/2005. Os *outliers* da série de alcalinidade foram mantidos, pois refletem comportamentos extremos relevantes ao estudo. Picos e quedas abruptas podem estar ligados a eventos naturais ou processos no Rio Guandu, sendo essenciais para capturar a variabilidade real e melhorar a precisão das previsões.

Modelos preditivos de séries temporais

A metodologia adotada neste estudo utilizou modelos avançados de *deep learning* — GRU (*Gated Recurrent Unit*) (Cho et al., 2014), TFT (*Temporal Fusion Transformer*) (Lim et al., 2021), (Neural Hierarchical Interpolation for Time Series) e o KAN (*Kolmogorov-Arnold Network*) (Liu et al., 2025) - devido à sua capacidade de capturar não linearidades, sazonalidades e dependências de longo prazo presentes em séries hidrológicas. A implementação foi realizada com a biblioteca *Neural Forecast*, considerando a normalização dos dados pelo *RobustScaler*, adequado por ser menos sensível a *outliers* (Pedregosa et al., 2011). O modelo foi configurado para previsão da alcalinidade com horizonte de 12 horas e janelas expansivas, incorporando *lags* e covariáveis selecionados por análises estatísticas e de correlação. A otimização dos hiperparâmetros foi feita via *Grid Search*, testando combinações de parâmetros como número de unidades ocultas, taxa de aprendizado, dentre outros. Para evitar o *overfitting*, aplicou-se a técnica de *Chained Temporal Cross-Validation*

(Cerqueira et al., 2020) com 8 *folds* e horizonte de 12 horas, respeitando a sequência temporal dos dados. Complementarmente, foram adotadas estratégias como *EarlyStopping*, *dropout* e regularização L2, garantindo maior robustez dos modelos e evitando sobreajuste (Prechelt, 1998).

Métricas de avaliação

Para avaliar o desempenho dos modelos, foram adotadas as métricas MAE (*Mean Absolute Error* - Erro Absoluto Médio), sMAPE (*Symmetric Mean Absolute Percentage Error* - Erro Percentual Absoluto Médio Simétrico) e o RMSE (*Root Mean Square Error* - Raiz do Erro Quadrático Médio), amplamente utilizadas na literatura. O MAE mede o erro absoluto médio, sendo robusto a outliers e de fácil interpretação, com bom desempenho quando próximo à metade do desvio padrão dos dados Willmott e Matsuura 2005). O sMAPE, uma métrica percentual simétrica, avalia erros relativos e evita distorções causadas por valores próximos de zero, sendo considerado excelente quando abaixo de 10%. Já o RMSE penaliza mais fortemente grandes erros, sendo adequado para contextos em que *outliers* são relevantes, e indica bom ajuste quando seu valor se aproxima do desvio padrão da série (Taylor 2001). A escolha dessas métricas considera as características dos dados e proporciona uma análise equilibrada da precisão dos modelos.

RESULTADOS E DISCUSSÕES

Os parâmetros selecionados para o modelo foram Alcalinidade (mg/L), pH (grandeza adimensional), Cor Aparente (Hz) e Turbidez (NTU), conforme os critérios de inclusão e exclusão definidos na metodologia. As estatísticas descritivas desses parâmetros estão apresentadas na Tabela 1, incluindo média (\bar{x}), desvio padrão (s), valores mínimos (*min*) e máximo (*max*), permitindo uma análise detalhada da distribuição dos dados.

Tabela 1. Estatísticas básicas dos parâmetros de qualidade da água.

Parâmetro	$\bar{x} \pm s$	MIN	MAX
Alcalinidade	18.474 \pm 1.934	6	58
pH	6.650 \pm 0.158	5,4	9,4
Cor aparente	61.856 \pm 80.647	0	3600
Turbidez	19.939 \pm 26.383	1,8	1200

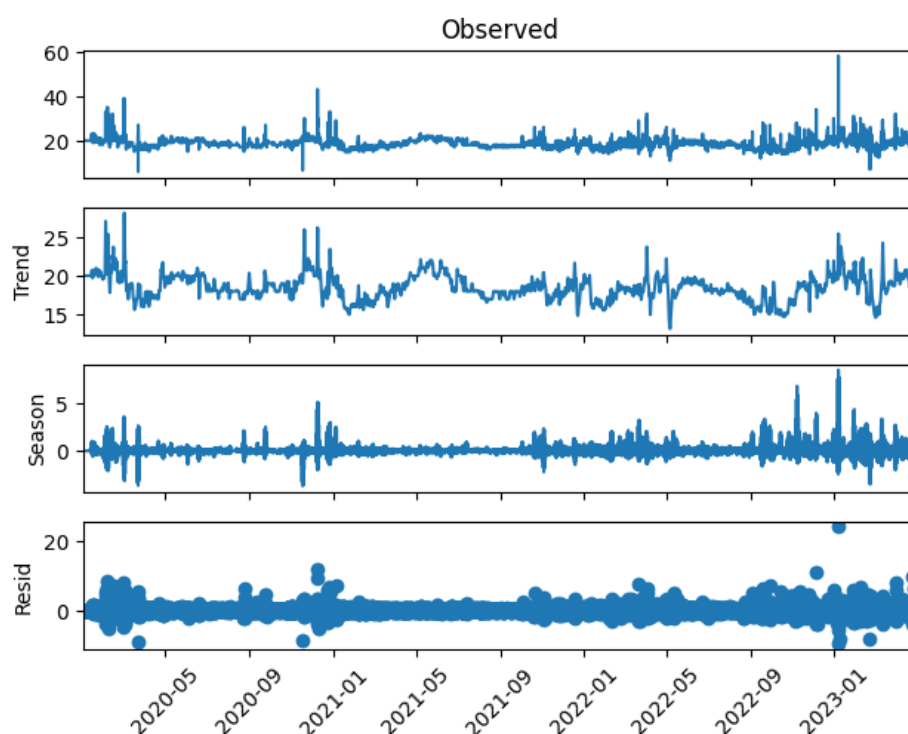
A análise dos parâmetros de qualidade da água revela que a alcalinidade apresenta média de 18,474 mg/L e baixa variabilidade ($\pm 1,934$), indicando estabilidade química e menor suscetibilidade a flutuações ambientais. Esse valor médio observado encontra-se dentro da faixa considerada normal, conforme relatórios internos que classificam os valores de referência em três níveis: Normal (15-22 ppm), Alerta (23-29 ppm) e Crítico (> 30 ppm). O pH também se mantém relativamente constante $\bar{x} \pm s = 6,650 \pm 0,158$, sugerindo um equilíbrio químico. Em contraste, a cor aparente e a turbidez exibem alta dispersão e amplas faixas de variação, com máximos de 3600 e 1200, respectivamente, sugerindo influência de fatores sazonais ou eventos pontuais de contaminação. Esses resultados indicam que, enquanto a alcalinidade e o pH tendem a permanecer estáveis, a turbidez e a cor aparente podem sofrer oscilações significativas, impactando diretamente a qualidade da água.

É importante ressaltar que, conforme estabelecido pela Resolução CONAMA 357/2005, os limites médios para águas doces (classes Especial, 1, 2 e 3), características da bacia estudada, são: pH entre 6,0 e 9,0; Cor Aparente ≤ 75 uH; e Turbidez de até 40 NTU (classes Especial e 1) e 100

NTU (classes 2 e 3) (do Meio Ambiente (CONAMA), 2005). Os parâmetros analisados indicam que: pH ($6,650 \pm 0,158$; 5,4–9,4) apresenta média dentro da faixa preconizada pela CONAMA (6,0–9,0), embora com variações pontuais; Turbidez ($19,939 \pm 26,383$; 1,8–1200 NTU) está dentro do limite para classes Especial e 1 (≤ 40 NTU), mas com picos críticos; Cor Aparente ($61,856 \pm 80,647$; 0–3600 Hz) excede o limite de 75 Hz, indicando presença significativa de substâncias dissolvidas; e Alcalinidade ($18,474 \pm 1,934$; 6–58 mg/L), embora sem limites estabelecidos, encontra-se em faixas consideradas normais.

A Figura 1 ilustra a decomposição temporal da série de alcalinidade ao longo do período estudado. O gráfico superior apresenta os dados observados, seguido pela tendência, que mostra o comportamento de longo prazo da série. O terceiro gráfico exibe a sazonalidade, destacando padrões cíclicos recorrentes. Por fim, o gráfico inferior apresenta o residual, que captura as variações não explicadas pela tendência ou sazonalidade.

Figura 1. Decomposição temporal da série de alcalinidade no ponto de coleta 22, no período de 2020 a 2023



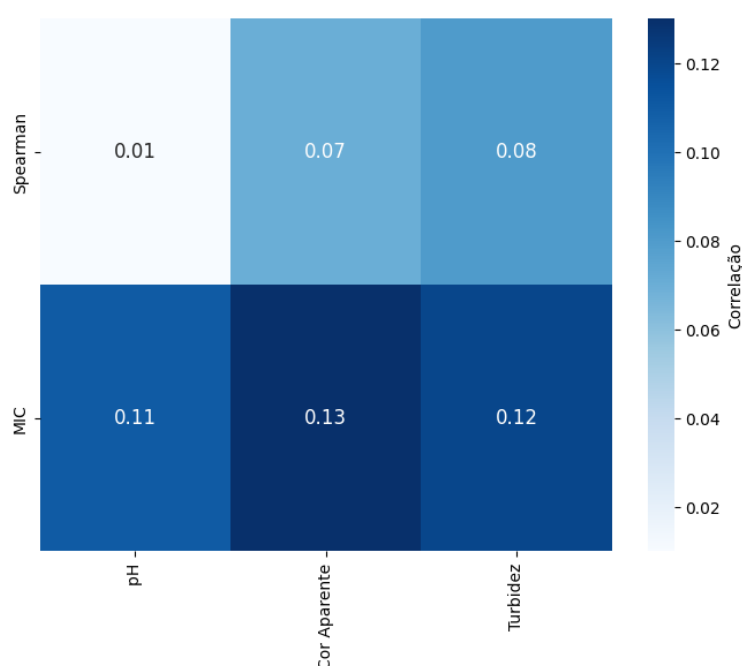
Os dados observados de alcalinidade ao longo dos anos de 2020 a 2023 (Figura 3), revela um padrão sazonal, com variações significativas associadas aos períodos de verão. Observam-se picos de alcalinidade nos meses de dezembro, janeiro e fevereiro, especialmente nos anos 2020, 2021 e 2023, coincidindo com a estação mais quente do ano. Esses picos são particularmente pronunciados nos anos analisados, sugerindo uma possível influência de fatores climáticos ou ambientais típicos do verão, como aumento da temperatura, maior taxa de evaporação ou alterações nos padrões de precipitação. A tendência geral (gráfico *Trend* da Figura 1) da alcalinidade mostra um crescimento consistente a partir do mês de setembro, atingindo seus valores máximos no mês de janeiro. Após o pico em janeiro, há uma tendência de declínio gradual a partir de março, seguida por um período de relativa estabilidade que se estende até setembro. Essa estabilidade pode indicar um equilíbrio nas

condições ambientais durante os meses mais frios, com menor variabilidade nos fatores que influenciam a alcalinidade.

Já o componente de sazonalidade (gráfico *Season* da Figura 1) apresenta um comportamento que acompanha a tendência até meados de 2021, sugerindo uma relação consistente entre os fatores que influenciam a alcalinidade e os ciclos sazonais durante esse período. No entanto, a partir de meados de 2021, observa-se um aumento na variabilidade da sazonalidade, indicando possíveis mudanças nos fatores ambientais, operacionais ou climáticos que afetam a alcalinidade. Por outro lado, o componente residual da série temporal (gráfico *Resid* da Figura 1) mostra-se uniforme ao longo de todo o período analisado, sugerindo que os modelos utilizados para decompor a série capturaram adequadamente os padrões de tendência e sazonalidade. A ausência de flutuações significativas no residual indica que não há grandes anomalias ou ruídos não explicados pelos componentes principais

A Figura 2 exibe uma matriz de correlação que compara os coeficientes obtidos pelos métodos de *Spearman* e MIC entre a Alcalinidade e os demais parâmetros analisados.

Figura 2. Análise de correlação da Alcalinidade com os parâmetros pH, Cor Aparente e Turbidez



Na análise de correlação, constatou-se que há baixa associação entre a alcalinidade e os demais parâmetros medidos no mesmo ponto de coleta, independentemente das técnicas utilizadas. Essa baixa correlação indica que a alcalinidade apresenta um comportamento relativamente independente em relação aos outros parâmetros analisados, sugerindo que sua variação não é fortemente influenciada por eles no local de estudo. Portanto, optou-se por excluir essas variáveis na elaboração dos modelos, uma vez que não contribuem significativamente para a previsão da alcalinidade, priorizando apenas os *lags* temporais da própria variável alvo.

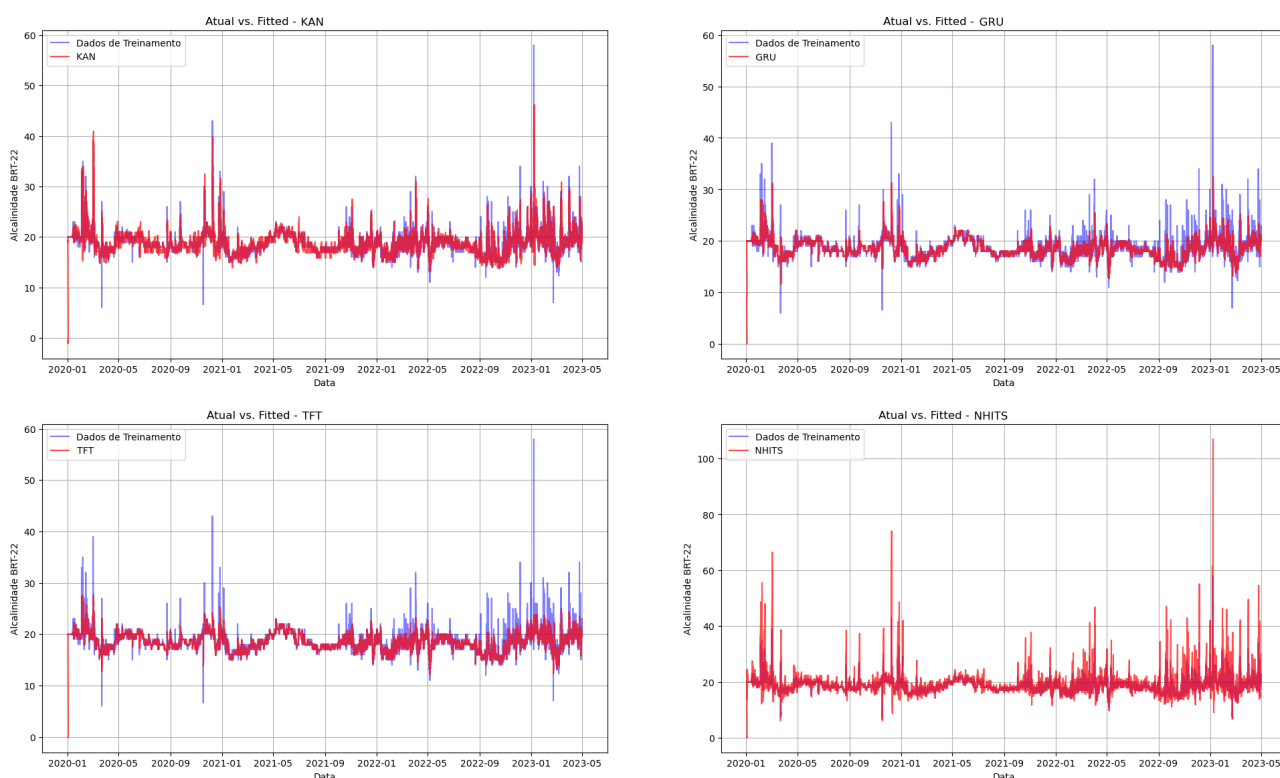
Diante desse cenário, optou-se por realizar a previsão utilizando apenas os valores históricos de alcalinidade no ponto 22 em um dos pontos de coleta de água. Essa decisão tem como objetivo

simplificar o modelo de previsão, evitando a inclusão de variáveis que não contribuem significativamente para a explicação da variabilidade da alcalinidade.

Em relação a performance qualitativa dos modelos, os gráficos entre os dados observados e os modelos KAN, GRU, TFT e NHITS treinados são ilustrados na Figura 3. Os gráficos ilustram a capacidade de cada modelo em capturar os padrões temporais da série de alcalinidade, destacando a precisão das previsões em relação aos dados reais.

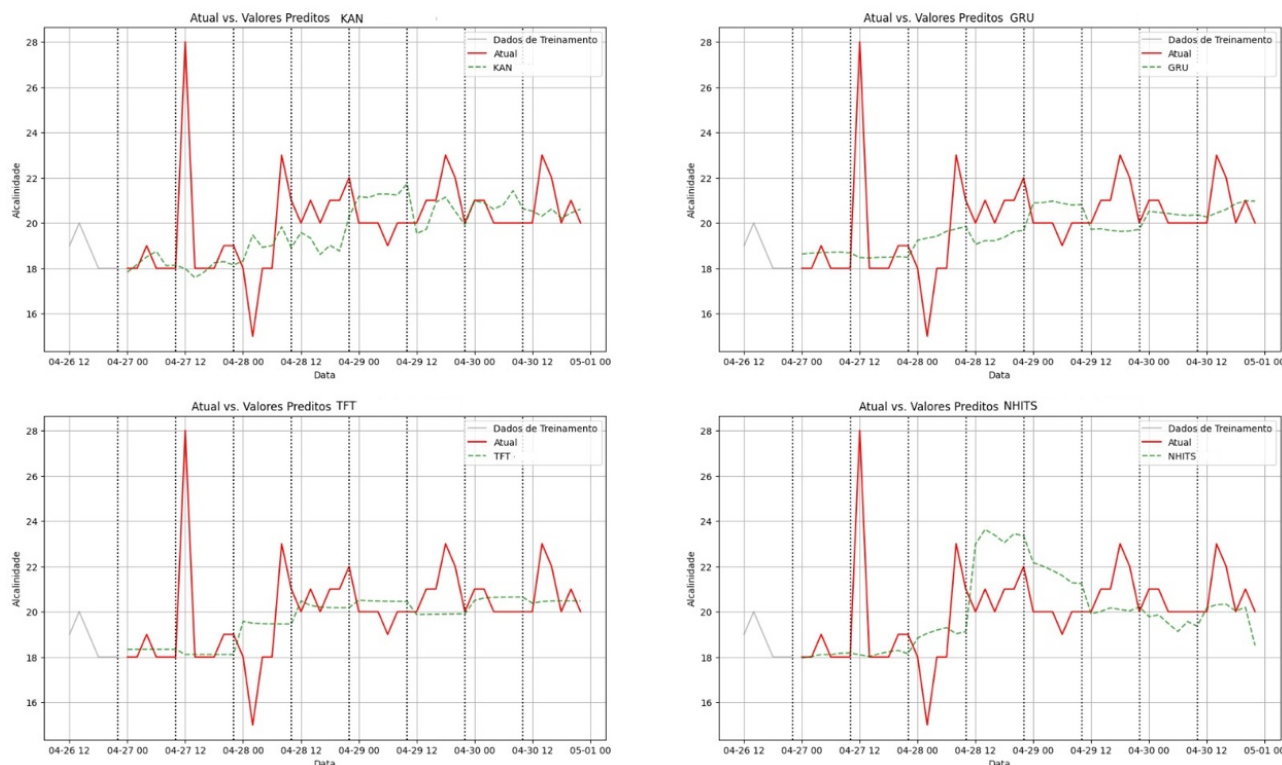
Com base na Figura 3, observa-se que o algoritmo KAN (superior esquerdo) acompanha bem a tendência dos dados, apresentando um ajuste regularizado e sendo capaz de capturar picos atípicos. O GRU (superior direito) exibe um ajuste mais suave e com menor variabilidade, porém responde menos a mudanças bruscas, como evidenciado nos picos de janeiro de 2021 e 2023. O TFT (inferior esquerdo) mantém um ajuste próximo aos valores observados, exibindo a menor variabilidade entre os modelos, mas com menor sensibilidade a variações locais, especialmente após setembro de 2022. Por outro lado, o NHITS (inferior direito) apresenta maior reatividade às oscilações e capta picos com maior sensibilidade, porém com maior variabilidade, o que pode indicar *overfitting*.

Figura 3. Gráficos dos valores observados com e os modelos treinados KAN (superior esquerdo), GRU (superior direito), TFT (inferior esquerdo) e NHITS (inferior direito)



A Figura 4 ilustra a comparação entre dados observados e valores preditos pelos modelos KAN, GRU, TFT e NHITS. Os gráficos apresentam o desempenho dos modelos na previsão dos padrões de alcalinidade ao longo de 4 dias (48 horas), dada a configuração de 8 *folds* com horizonte de previsão de 12 horas (6 passos de 2 horas cada). As linhas verticais pontilhadas demarcam os limites entre cada *fold*, que avançam sequencialmente 12 horas, permitindo avaliar a capacidade preditiva dos modelos em múltiplas janelas temporais.

Figura 4. Comparação entre dados observados e valores preditos pelos modelos KAN (superior esquerdo), GRU (superior direito), TFT (inferior esquerdo) e NHITS (inferior direito).



Com base nos gráficos da Figura 4, observa-se que os algoritmos GRU e o TFT apresentam previsões suavizadas, capturando a tendência geral da série, mas com dificuldades em responder a variações abruptas. Dentre eles, o TFT destaca-se por equilibrar melhor a estabilidade das previsões com a identificação de padrões sazonais, ainda que subestime picos extremos. O KAN exibe maior variabilidade, mas não consegue capturar adequadamente os picos e vales observados nos dados reais. O NHITS se mostra mais responsivo às oscilações da série, acompanhando de maneira mais próxima as variações dos valores reais, mas com maior sensibilidade ao ruído. No entanto, nenhum dos modelos conseguiu acompanhar com precisão as variações abruptas da série temporal.

A Tabela 2 apresenta um resumo das métricas de avaliação MAE, sMAPE e RMSE utilizadas para comparar o desempenho dos modelos TFT, GRU, NHITS e KAN na previsão da alcalinidade. Valores menores indicam melhor ajuste do modelo aos dados observados.

Tabela 2. Resumo das métricas de avaliação para os algoritmos TFT, GRU, NHITS e KAN para alcalinidade

Algoritmo	MAE	sMAPE (%)	RMSE
TFT	1,129	2,7	1,932
NHITS	1,481	3,5	2,199
GRU	1,274	3,1	1,962
KAN	1,234	3,0	1,993

Com base nos resultados da Tabela 2 os melhores resultados para a predição da alcalinidade foram obtidos pelo algoritmo TFT, com um MAE de 1,129, indicando que, em média, as previsões desviam-se dos valores observados em aproximadamente 1,13 mg/L. Esse valor representa cerca de 6% da média observada (18,474 mg/L), o que sugere uma boa precisão do modelo. O sMAPE de

0,027% reforça essa conclusão, evidenciando que o erro percentual é extremamente baixo, o que é particularmente relevante para aplicações práticas que exigem alta confiabilidade. Por fim, o RMSE de 1,932 indica que, embora haja algumas previsões com erros maiores, a dispersão geral dos erros é controlada, ficando abaixo de 2 mg/L. Considerando o desvio padrão dos dados observados (1,934), o RMSE mostra que o modelo consegue capturar a variabilidade da série temporal de forma eficiente, mantendo os erros dentro de uma faixa aceitável em relação à dispersão natural dos dados. Esses resultados destacam a capacidade do KAN em prever a alcalinidade com precisão e consistência.

CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo principal analisar os parâmetros físico-químicos de amostras de água bruta coletadas em diversos pontos ao longo da bacia hidrográfica do Rio Guandu, com o intuito de prever, por meio de técnicas de inteligência artificial, os níveis de alcalinidade no ponto de captação. Para isso, foram aplicados diversos algoritmos de aprendizado de máquina, com destaque para o KAN (Kolmogorov-Arnold Network), cuja performance foi avaliada com base em gráficos de treinamento e teste, além das métricas como o Erro Absoluto Médio (MAE), o Erro Percentual Absoluto Médio Simétrico (sMAPE) e a Raiz do Erro Quadrático Médio (RMSE).

Os resultados alcançados demonstraram que o algoritmo TFT apresentou a melhor performance para a predição de alcalinidade. Os gráficos de treinamento e teste evidenciaram uma alta aderência entre os valores observados e os preditos, com erros mínimos e consistência ao longo da série temporal. As métricas de avaliação corroboraram essa conclusão, indicando que o TFT não apenas capturou os padrões subjacentes aos dados, mas também generalizou bem para o conjunto de teste, reforçando sua aplicabilidade em cenários reais. Comparado aos demais métodos testados, o TFT destacou-se pela sua capacidade de lidar com a complexidade dos dados e pela precisão nas predições.

No entanto, o estudo apresenta algumas limitações. A análise não considerou os impactos de variáveis temporais e espaciais, como flutuações sazonais, eventos climáticos extremos ou variações geográficas ao longo da bacia hidrográfica, que podem influenciar significativamente os níveis de alcalinidade. A ausência dessas covariáveis pode ter restringido a capacidade dos modelos de capturar toda a complexidade do fenômeno estudado. Trabalhos futuros devem explorar a incorporação dessas variáveis, bem como de dados ambientais, como temperatura, precipitação, vazão, dentre outros, para potencializar o poder preditivo dos modelos e ampliar sua aplicabilidade.

Conclui-se que este estudo contribui significativamente para o avanço do uso de métodos de *deep learning* aplicados na área de hidrologia, em particular para a análise de parâmetros físico-químicos de água bruta, como a alcalinidade, em contextos de captação e tratamento em Estações de Tratamento de Água (ETA). A performance superior do KAN reforça a viabilidade de técnicas avançadas de predição para otimizar a gestão de recursos hídricos, destacando a importância de abordagens inovadoras e multidisciplinares.

AGRADECIMENTOS

Esta pesquisa foi desenvolvida no âmbito do PROJETO IAgua – Aplicação de IA em modelo computacional preditivo para a detecção de eventos de risco que impactem a qualidade da água em mananciais e ETAs, selecionado por meio da SELEÇÃO PÚBLICA MCTI/FINEP/FNDCT Subvenção Econômica à Inovação – 04/2022, e formalizado pelo Termo de Outorga de Subvenção Econômica FINEP, instrumento contratual código N° 03.23.0524.00 (Referência Finep n° 1954/22), tendo seguido as normas, acordos e diretrizes vigentes para a execução do projeto.

Os autores gostariam de expressar sua sincera gratidão à FINEP (Financiadora de Estudos e Projetos) pelo apoio financeiro e pelos recursos que tornaram esta pesquisa possível. Além disso, um agradecimento especial é direcionado à CEDAE (Companhia Estadual de Águas e Esgotos) pelo

fornecimento de dados essenciais, apoio em aspectos de ordem técnica e de infraestrutura, e à VM9 Tecnologia da Informação Ltda, pela infraestrutura computacional e suporte tecnológico especializado. Todas essas contribuições foram fundamentais para o desenvolvimento e sucesso deste trabalho.

REFERÊNCIAS

- Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83.
- Cerqueira, V., Torgo, L., Oliveira, M., & Pfahringer, B. (2020). Evaluating Time Series Forecasting Models: An Empirical Study on Performance Estimation Methods. *Machine Learning*, 109(7), 1997–2028. <https://doi.org/10.1007/s10994-020-05910-7>
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *ArXiv Preprint ArXiv:1406.1078*. <https://arxiv.org/abs/1406.1078>
- de Castro, C. N. (2022). *Água, problemas complexos e o Plano Nacional de Segurança Hídrica*. Instituto de Pesquisa Econômica Aplicada (Ipea).
- do Meio Ambiente (CONAMA), C. N. (2005). *Resolução CONAMA nº 357, de 17 de março de 2005*. <https://www legisweb.com.br/legislacao/?id=48493>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Guerra, L. V., & Silva, B. D. D. (2018). Vigilância da qualidade da água para consumo no estado do Rio de Janeiro. *Ambiente & Sociedade*, 21, 1–16. <https://doi.org/10.1590/1809-4422asoc0097r2vu18L3TD>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Lim, B., Ark, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. *International Journal of Forecasting*, 37(4), 1748–1764.
- Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M. (2025). *KAN: Kolmogorov-Arnold Networks*. <https://arxiv.org/abs/2404.19756>
- Luo, J., Zhang, W., & Li, Y. (2024). Advanced Techniques for Water Quality Prediction in Urbanized Areas. *Environmental Science and Technology*.
- Luo, W., Huang, L., Shu, J., Feng, H., Guo, W., Xia, K., Fang, K., & Wang, W. (2024). Predicting water quality in municipal water management systems using a hybrid deep learning model. *Engineering Applications of Artificial Intelligence*.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825–2830.
- Prasad, D., Venkataramana, L., Kumar, P., Gurunathan, P., Harshana, S., Srividya, S., Harrinei, K., & Indraganti, S. (2022). Analysis and prediction of water quality using deep learning and auto deep learning techniques. *Science of The Total Environment*, 821, 153311. <https://doi.org/10.1016/j.scitotenv.2022.153311>
- Prechelt, L. (1998). Early Stopping - But When? In G. B. Orr & K.-R. Müller (Eds.), *Neural Networks: Tricks of the Trade* (pp. 55–69). Springer. https://doi.org/10.1007/3-540-49430-8_3
- Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, 36(1), 75–85.
- Song, C., & Yao, L. (2022). A hybrid model for water quality parameter prediction based on CEEMDAN-IALO-LSTM ensemble learning. *Environmental Earth Sciences*, 81, 262. <https://doi.org/10.1007/s12665-022-10380-2>
- Yao, X., Xu, Z., Ren, T., & Zeng, X.-J. (2025). Feature-driven hybrid attention learning for accurate water quality prediction. *Expert Systems with Applications*, 276, 127160. <https://doi.org/https://doi.org/10.1016/j.eswa.2025.127160>
- Zhang, J., Xue, S., & Geng, P. (2023). Research on BiLSTM-GRU Water Quality PredictionModel Based on Attention Mechanism. *4th International Conference on Computer Vision, Image and Deep Learning*.