

Avaliação de Modelos de Inteligência Artificial para Previsão de Variações Batimétricas com Base em Dados Meteoceanográficos de Boia

Henery Garção¹, Thiago Nascimento² & Áureo Manente³

Resumo: Este estudo avalia o desempenho de modelos de aprendizado de máquina na previsão de variações batimétricas em ambientes costeiros a partir de dados meteoceanográficos provenientes exclusivamente de boias. O conjunto de dados foi construído a partir de levantamentos batimétricos sucessivos, utilizados para estimar a evolução do leito, combinados com séries temporais de altura significativa de onda, correntes e nível do mar. As variações batimétricas foram discretizadas em cinco classes representando diferentes intensidades de assoreamento, estabilidade e erosão. Foram avaliados os algoritmos Random Forest e Gradient Boosting, com treinamento baseado em divisão temporal dos dados. Os resultados indicaram desempenho superior ao baseline, com acurácia global em torno de 0,45. No entanto, observou-se forte desbalanceamento na capacidade preditiva, com melhor desempenho na identificação de condições intermediárias e limitações na previsão de valores mais extremos. Os resultados evidenciam que dados de boia são capazes de capturar tendências gerais da dinâmica do leito, mas apresentam limitações na representação de processos mais intensos e espacialmente complexos, indicando a necessidade de integração com outras fontes de dados para aplicações operacionais.

Abstract: This study evaluates the performance of artificial intelligence models for predicting bathymetric variations in a port environment using exclusively buoy-derived metocean data. The dataset was constructed from sequential hydrographic surveys combined with time series of wave height, currents, and sea level, aggregated over survey intervals. Seabed changes were classified into five categories representing different intensities of accretion, stability, and erosion. Random Forest and Gradient Boosting models were tested using a temporal split between training and testing datasets. Results showed performance above the baseline model, with an overall accuracy of approximately 0.45. However, predictive performance was unbalanced, with better results for intermediate conditions and significant limitations in detecting extreme values. The inclusion of historical variables improved Random Forest performance but reduced Gradient Boosting accuracy. These findings indicate that buoy data can capture general sedimentary trends but are insufficient to fully represent complex spatial processes, highlighting the importance of integrating additional data sources.

Palavras-Chave: Machine Learning; Batimetria; Previsão; Random Forest; Gradient Boosting

Keywords: Machine Learning; Bathymetry; Prediction; Random Forest; Gradient Boosting

INTRODUÇÃO

A previsão de variações batimétricas em ambientes costeiros e portuários representa um desafio relevante para a engenharia costeira, especialmente em regiões sujeitas a intensa dinâmica sedimentar. Processos de assoreamento e erosão resultam da interação entre forçantes hidrodinâmicas, características sedimentares e condições morfológicas, apresentando comportamento altamente não linear e variabilidade espaço-temporal significativa (FU et al., 2025; YILDIZ; STANEV; STANEVA, 2025).

1) UMI SAN e Universidade Federal do Espírito Santo (henery.garcao@umi.com.br)

2) UMI SAN (thiago.nascimento@umi.com.br)

3) UMI SAN (aureo.manente@umi.com.br)

Tradicionalmente, a análise desses processos é realizada por meio de modelos numéricos hidrodinâmicos e morfodinâmicos ou por campanhas de monitoramento batimétrico. Embora essas abordagens sejam fundamentais, elas frequentemente demandam elevado esforço computacional, grande volume de dados de entrada e processos complexos de calibração, o que pode limitar sua aplicação em contextos operacionais (FU et al., 2025).

Nesse contexto, técnicas de aprendizado de máquina têm emergido como uma alternativa promissora para a modelagem de processos ambientais complexos, devido à sua capacidade de capturar relações não lineares entre variáveis e integrar diferentes fontes de dados (TAO et al., 2021). Estudos recentes demonstram o potencial dessas abordagens na previsão de transporte de sedimentos em rios e na reconstrução batimétrica em ambientes costeiros, especialmente quando há disponibilidade de dados espaço-temporais robustos (FU et al., 2025; YILDIZ; STANEV; STANEVA, 2025).

Entretanto, a aplicação dessas técnicas em cenários com disponibilidade limitada de dados, como aqueles baseados exclusivamente em medições pontuais (por exemplo, boias meteorológicas), ainda é pouco explorada. A maioria dos estudos utiliza conjuntos de dados extensivos, frequentemente incluindo informações espaciais detalhadas, o que reduz a aderência dessas abordagens a contextos operacionais reais (TAO et al., 2021).

Diante disso, este trabalho investiga a capacidade de modelos de aprendizado de máquina em prever variações batimétricas utilizando exclusivamente dados de boia, com o objetivo de avaliar suas limitações e potencial de aplicação em cenários com restrição de dados.

MÉTODOS

Construção do conjunto de dados

Foram utilizados levantamentos batimétricos sucessivos (LHs), convertidos para uma grade regular com resolução espacial constante por meio de interpolação linear. A partir dessas superfícies, foram calculadas as variações batimétricas entre campanhas consecutivas.

As séries temporais de dados meteorológicos (incluindo altura significativa de onda, correntes e nível do mar) foram obtidas de instrumentos de medição (boia e marégrafo) e agregadas no intervalo entre levantamentos hidrográficos. Dados faltantes foram tratados por interpolação temporal limitada.

Para cada intervalo entre levantamentos, foram extraídas variáveis estatísticas e físicas representativas, incluindo médias, desvios padrão, valores máximos e proxies de energia.

Adicionalmente, foram incorporadas variáveis morfológicas derivadas da batimetria inicial, como profundidade local e declividade do fundo (declividade), bem como variáveis históricas representando a frequência de ocorrência prévia de deposição e erosão em cada ponto.

As variações batimétricas foram classificadas em cinco classes discretas:

- *Assoreamento intenso*: $\geq 0,5$ m;
- *Assoreamento moderado*: entre 0,1 m e 0,5 m;
- *Estabilidade*: entre -0,1 m e 0,1 m;
- *Erosão moderada*: entre -0,1 m e -0,5 m;
- *Erosão intensa*: $\leq -0,5$ m.

Modelagem

Foram avaliados dois algoritmos de aprendizado de máquina voltados à classificação:

- Random Forest
- Gradient Boosting

Os modelos foram treinados com divisão temporal dos dados, garantindo que os intervalos mais recentes fossem utilizados exclusivamente para teste, evitando vazamento de informação.

Durante o treinamento, foram removidas amostras com dados incompletos, e foi aplicado balanceamento implícito no Random Forest por meio de ponderação de classes.

Avaliação

O desempenho dos modelos foi avaliado utilizando:

- Acurácia global;
- Balanced accuracy;
- Classificação por classe;
- Matrizes de confusão;
- Curvas ROC e precisão-recall.

Adicionalmente, foi implementado um modelo baseline baseado na classe dominante, permitindo avaliar o ganho real dos modelos de aprendizado de máquina. Também foi realizada uma análise comparativa com e sem variáveis históricas, a fim de avaliar o impacto da memória do sistema na capacidade preditiva.

RESULTADOS E DISCUSSÃO

Os modelos Random Forest e Gradient Boosting apresentaram desempenho superior ao modelo baseline, conforme apresentado na Figura 1. Observa-se que ambos os algoritmos atingem acurácia global próxima de 0,45, indicando capacidade de capturar padrões relevantes entre as forçantes meteoceanográficas e a resposta morfológica do sistema.

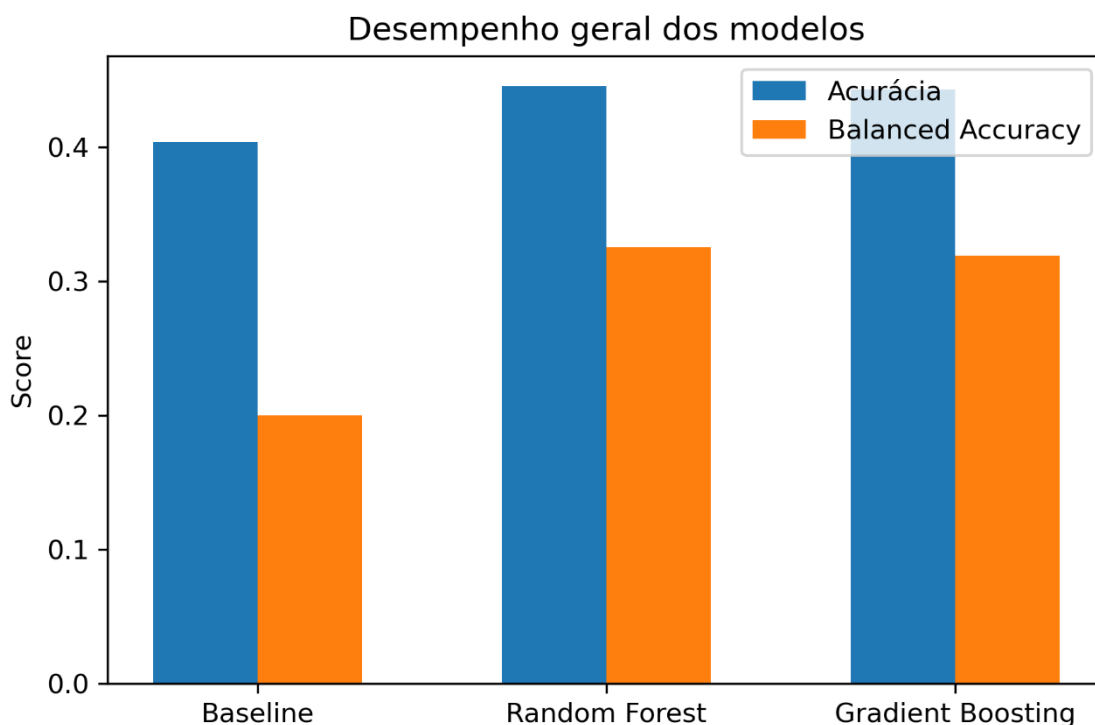


Figura 1 – Desempenho geral dos modelos de classificação. Comparação entre o modelo baseline e os algoritmos Random Forest e Gradient Boosting em termos de acurácia global e balanced accuracy.

Entretanto, a análise da balanced accuracy, inferior a acurácia global, evidencia um comportamento preditivo desbalanceado entre as classes. Esse resultado indica que, embora os modelos consigam identificar padrões gerais, há dificuldade em representar adequadamente classes menos frequentes, especialmente valores extremos de assoreamento e erosão.

A Figura 2 apresenta o desempenho por classe, permitindo uma análise mais detalhada da capacidade preditiva dos modelos. Observa-se que ambos os algoritmos apresentam melhor desempenho na identificação de condições intermediárias, especialmente assoreamento moderado, com acurácia significativamente superior às demais classes.

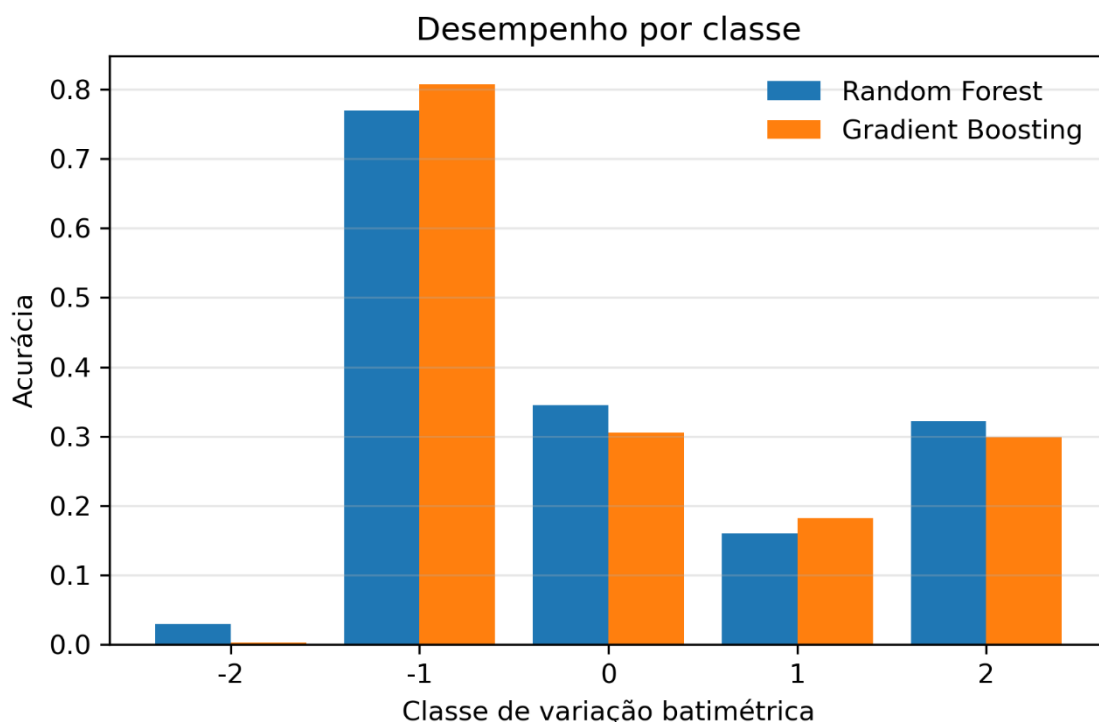


Figura 2 – Desempenho dos modelos Random Forest e Gradient Boosting por classe de variação batimétrica. Classes: “-2”: assoreamento intenso; “-1”: assoreamento moderado; “0”: estabilidade; “1”: erosão moderada; “2”: erosão intensa.

Por outro lado, valores extremos, tanto de assoreamento quanto de erosão, apresentam desempenho bastante limitado, com valores de acurácia próximos de zero em alguns casos. Esse comportamento está associado ao desbalanceamento do conjunto de dados e à menor representatividade desses eventos, além da limitação das variáveis de entrada, que representam forçantes essencialmente pontuais, em capturar processos espacialmente heterogêneos.

Os resultados indicam ainda uma tendência dos modelos em concentrar previsões nas classes mais frequentes, o que é consistente com problemas de classificação em bases desbalanceadas. Esse efeito é mais evidente no Gradient Boosting, que apresentou maior sensibilidade à distribuição das classes.

A análise comparativa com e sem variáveis históricas indicou diferenças discretas no desempenho dos modelos. O Random Forest apresentou leve melhora com a inclusão dessas variáveis, enquanto o Gradient Boosting apresentou redução de desempenho. Esses resultados sugerem sensibilidade dos algoritmos à incorporação de informação temporal, porém os efeitos observados são de pequena magnitude e não permitem concluir, de forma robusta, sobre o ganho preditivo associado ao uso de histórico.

De forma geral, os resultados demonstram que dados meteoceanográficos medidos são capazes de representar tendências globais da evolução batimétrica, porém apresentam limitações na descrição de processos extremos e altamente localizados. Essa limitação está diretamente relacionada à ausência de variáveis espaciais mais detalhadas, o que reforça a necessidade de integração com outras fontes de dados em aplicações futuras.

CONCLUSÃO

Os resultados demonstram que modelos de inteligência artificial possuem capacidade consistente de capturar relações entre forçantes meteoceanográficas e variações batimétricas, mesmo quando utilizados exclusivamente dados de boia.

O desempenho superior ao baseline confirma a existência de padrões previsíveis no sistema, especialmente para condições intermediárias de assoreamento e erosão. No entanto, a limitação na previsão de valores extremos evidencia que a dinâmica sedimentar não pode ser completamente representada apenas por variáveis hidrodinâmicas pontuais.

A análise comparativa entre modelos mostrou que o Random Forest apresenta maior robustez frente à inclusão de variáveis históricas, enquanto o Gradient Boosting demonstrou maior sensibilidade a ruídos e à estrutura temporal dos dados.

De forma geral, os resultados indicam que dados meteoceanográficos medidos são suficientes para capturar tendências globais da evolução batimétrica, mas insuficientes para descrever processos espaciais complexos de forma completa.

Esse comportamento reforça a hipótese central do projeto: a previsão robusta da dinâmica sedimentar depende da integração entre dados medidos, variáveis fisicamente consistentes e maior representatividade espacial dos processos.

AGRADECIMENTOS

Os autores agradecem à empresa UMI SAN pelo suporte técnico e pela disponibilização dos dados utilizados neste estudo. O primeiro autor agradece à Fundação de Amparo à Pesquisa e Inovação do Espírito Santo pela concessão da bolsa de doutorado.

REFERÊNCIAS

- FU, Y.; LUO, J.; ZHANG, D.; LIU, L.; LUO, G.; ZU, X. (2025). "*Machine learning for spatiotemporal prediction of river siltation in typical reach in Jiangxi, China.*" Applied Sciences, 15, 8628.
- TAO, H.; AL-KHAFI, Z. S.; QI, C.; ZOUNEMAT-KERMANI, M.; KISI, O.; TIYASHA, T.; CHAU, K.-W.; NOURANI, V.; MELESSE, A. M.; ELHAKEEM, M.; FAROOQUE, A. A.; NEJADHASHEMI, A. P.; KHEDHER, K. M.; ALAWI, O. A.; DEO, R. C.; SHAHID, S.; SINGH, V. P.; YASEEN, Z. M. (2021). "*Artificial intelligence models for suspended river sediment prediction: state-of-the-art, modeling framework appraisal, and proposed future research directions.*" Engineering Applications of Computational Fluid Mechanics, 15(1), pp. 1585–1612.
- YILDIZ, M.; STANEV, E. M.; STANEVA, J. (2025). "*Deep learning-based reconstruction and prediction of coastal bathymetry.*" Ocean Dynamics, 75, p. 36.