

Aplicação do *Machine Learning* no estudo da área superficial específica de nanopartículas de óxido de nióbio V para aplicação na fotocatalise heterogênea

Maurício Dalla Costa Rodrigues da Silva^{1*}, Luís Fernando Wentz Brum¹, Leandro Rodrigues Oviedo¹, Giovani Pavoski², Denise Croce Romano Espinosa², William Leonardo da Silva¹

¹Universidade Franciscana, Programa de Pós-Graduação em Nanociências, Rua Silva Jardim, 1323, CEP: 97010-491, Santa Maria-RS, Brasil

²Universidade de São Paulo, Escola Politécnica de Engenharia Química, Rua do Lago, 250, CEP: 05508-080, São Paulo – SP – Brasil

* Autor correspondente: eng.mauriciodallacosta@gmail.com

Resumo/Abstract

RESUMO – Nanopartículas metálicas vêm sendo aplicadas no tratamento de águas residuárias por fotocatalise heterogênea, destacando-se as nanopartículas de óxido de nióbio V (Nb₂O₅-NPs). Além disso, o uso de *Machine Learning* (ML) apresenta-se como uma promissora ferramenta para a predição de propriedades das Nb₂O₅-NPs. Nesse contexto, o presente estudo tem por objetivo avaliar a aplicação de ML na predição dos valores de área superficial específica das Nb₂O₅-NPs. Assim, realizou-se o levantamento dos dados disponíveis na literatura para a análise exploratória das principais variáveis de síntese e a predição dos valores área superficial específica por meio do modelo *Random Forest* (RF). Os resultados da análise exploratória evidenciaram que o aumento da temperatura ocasionou a redução da área superficial específica nas Nb₂O₅-NPs. Adicionalmente, o modelo RF conseguiu prever os valores de área superficial específica com erros relativos inferiores a 15% dentre os valores observados e preditos. Por conseguinte, o uso de ML apresenta-se como uma ferramenta a ser utilizada para avaliar a síntese de Nb₂O₅-NPs, antes da realização de etapas experimentais.

Palavras-chave: *Random Forest, Síntese, Nióbio, Nanomateriais.*

ABSTRACT - Metallic nanoparticles have been applied in the wastewater treatment by heterogeneous photocatalysis, with an emphasis on niobium oxide V nanoparticles (Nb₂O₅-NPs). Moreover, the use of Machine Learning (ML) is a promising tool for predicting important properties of Nb₂O₅-NPs. In this context, the present study aims to evaluate the application of ML in predicting the values of specific surface area of Nb₂O₅-NPs. Thus, a survey of data available in the literature was carried out for the exploratory analysis of the main synthesis variables and the prediction of specific surface area values through the Random Forest (RF) model. The results of the exploratory analysis showed that the increase in temperature led to a reduction in the specific surface area of Nb₂O₅-NPs. Additionally, the RF model was able to predict specific surface area values with relative errors of less than 15% among the observed and predicted values. Therefore, the use of ML is presented as a tool to be used to evaluate the synthesis of Nb₂O₅-NPs before carrying out experimental steps.

Keywords: *Random Forest, Synthesis, Niobium, Nanomaterials.*

Introdução

É notório a relevância que as nanopartículas de óxido de nióbio V (Nb₂O₅-NPs) ao longo dos últimos anos, em função de sua versatilidade, estabilidade térmica e química, baixa toxicidade, energia de banda proibida (ou *band gap*) na faixa de 3,1 a 4,0 eV, juntamente com o potencial de oxirredução (1,2), tornando-as com o potencial fotocatalítico para aplicação na fotocatalise heterogênea no tratamento de águas residuárias (3).

A síntese das Nb₂O₅-NPs pode ser realizada por meio de diferentes rotas, com o uso de diferentes tipos de precursores metálicos e parâmetros de processo. Assim, a escolha destas

variáveis no processamento de síntese influencia diretamente nas propriedades essenciais do sistema nanoestruturado, como o tamanho da partícula, energia de *band gap*, morfologia, estrutura cristalina e área superficial específica (1). Dentre estas características, destaca-se a área superficial específica, visto que é responsável por disponibilizar os sítios ativos catalíticos presentes sob a superfície das Nb₂O₅-NPs (4).

No entanto, mesmo com as vantagens mencionadas, o número de estudos envolvendo a síntese e aplicações das Nb₂O₅-NPs é inferior, quando comparado com outras nanopartículas metálicas (5). Nesse sentido, o uso de

estudos computacionais pode auxiliar em futuras pesquisas relacionadas à síntese das Nb₂O₅-NPs.

A utilização de estudos computacionais, como o aprendizado de máquina (ou *Machine Learning*, ML), são alternativas para a identificação de padrões com base nas informações presentes na literatura científica, utilizando algoritmos para a predição de valores de resposta e na identificação das principais variáveis que afetam um determinado processo (6). Assim, estes estudos são uma forma de análise de dados que podem ser utilizadas para reduzir o custo e tempo relacionados às etapas experimentais (7).

Dentre os algoritmos de predição do ML, destaca-se o *Random Forest* (RF), onde uma série de árvores de decisões são criadas e escolhidas aleatoriamente, a fim de realizar o treino da base de dados e, posteriormente a predição das variáveis resposta de interesse (8), indicando uma maior precisão, em comparação ao uso de uma única árvore de decisão.

Nesse sentido, o presente trabalho tem como objetivo avaliar a aplicação de ML (por meio do algoritmo RF) na predição de valores de área superficial específica das Nb₂O₅-NPs, utilizando como base os principais parâmetros de síntese reportados na literatura científica, como escolha das rotas, precursor metálico e faixa de temperatura de tratamento térmico.

Experimental

Levantamento de dados

Os dados relacionados à síntese das Nb₂O₅-NPs foram coletados, filtrados e selecionados para o estudo de ML utilizando a base de dados da plataforma Scopus (www.scopus.com), utilizando os descritores “niobium” AND “nano” AND “synthesis”, segundo a lógica Booleana, no período de 2014 – abril 2023, limitando para artigos científicos. Assim, foram coletadas informações relacionadas ao tipo de rota utilizada na síntese, tipo de precursor metálico e temperatura de tratamento térmico, como as variáveis de entrada, enquanto a área superficial específica foi escolhida com a variável resposta, totalizados em 34 artigos científicos.

Processamento de dados

Os dados coletados foram divididos em dois grupos: treino e teste. Da totalidade dos dados coletados, 70% foram utilizados para treino e 30% para teste do modelo de predição. Os dados de entrada foram normalizados conforme a Eq. (1).

$$x_{in} = \gamma \frac{x_i - x_{min}}{x_{max} - x_{min}} - \beta \gamma \quad (1)$$

Onde x_{in} é o valor normalizado, x_i é o dado bruto da variável de entrada, x_{min} é o valor mínimo para a variável de entrada e x_{max} é o valor máximo para a variável de entrada. Os valores dos coeficientes γ e β utilizados foram, respectivamente, 2 e 0,5 (9).

Machine Learning (ML)

O modelo de predição utilizado no estudo para a área superficial específica das Nb₂O₅-NPs foi o *Random Forest* (RF), sendo avaliado em Python 3.8 (Google Colaboratory), e utilizado o coeficiente de determinação (R^2) (Eq. 2) e o valor da raiz quadrada do erro quadrado médio (RMSE) (Eq. 3), como parâmetros de performance do modelo de predição empregado.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_{i,exp} - y_{i,pred})^2}{\sum_{i=1}^N (y_{i,exp} - \hat{y}_{i,pred})^2} \quad (2)$$

$$RMSE = 1 - \sqrt{\frac{\sum_{i=1}^N (y_{i,exp} - \hat{y}_{i,pred})^2}{N}} \quad (3)$$

Onde $y_{i,exp}$ e $\hat{y}_{i,pred}$ são o valor real e o valor previsto da resposta, respectivamente.

Adicionalmente, o número de árvores de decisão e a altura das árvores de decisão empregados na etapa de treino foram avaliadas em função do valor de RMSE.

Resultados e Discussão

Análise exploratória

O levantamento de dados realizado indicou a presença de três principais rotas de síntese das Nb₂O₅-NPs (decomposição térmica, sol-gel e hidrotérmica), ao passo que três precursores metálicos foram utilizados (óxido de nióbio V, oxalato amoniacal de nióbio e pentaóxido de nióbio). Sobre a faixa de temperatura de tratamento térmico, foi utilizado de 100 - 700°C e os valores de área superficial específica reportados encontram-se na faixa de 8,3 - 273,6 m² g⁻¹. Assim, foram coletados 136 dados referentes às variáveis em estudo no total.

Dentre as rotas levantadas, a rota de síntese sol-gel apresentou a maior frequência (50%), seguida da rota hidrotérmica (29%) e por decomposição térmica (21%). A maior frequência da rota sol-gel pode ser atribuída a maior aplicabilidade dessa na síntese de nanopartículas de óxidos metálicos, em virtude das suas vantagens como a obtenção de nanocristais com alta pureza em baixas temperaturas, com maior controle de tamanho e morfologia (10).

Referente aos precursores metálicos, o oxalato amoniacal de nióbio apresentou uma frequência de 59%, ao passo que o óxido de nióbio V e o pentaóxido de nióbio apresentaram

frequências de 24% e 18%, respectivamente. Esse resultado por ser explicado pelo menor custo do precursor, frente aos demais, proporcionando Nb₂O₅-NPs com propriedades catalíticas favoráveis.

No que tange a temperatura de tratamento térmico, a Figura 1 apresenta a distribuição da área superficial específica, em função da temperatura e da rota de síntese. A distribuição evidencia que as Nb₂O₅-NPs obtidas por meio das rotas hidrotérmica e de degradação possuem valores de área superficial específica maiores do que as Nb₂O₅-NPs obtidas pela rota sol gel.

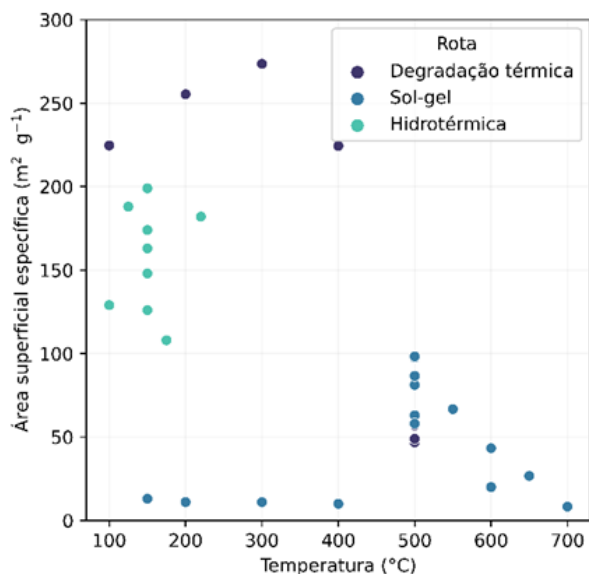


Figura 1. Distribuição da área superficial específica em função da temperatura de tratamento térmico e da rota de síntese.

O aumento da temperatura de tratamento térmico pode ocasionar a coalescência dos poros, levando ao processo de sinterização das nanopartículas, reduzindo sua área superficial específica (11). Ademais, o uso de surfactantes com alto peso molecular dificulta a posterior remoção da estrutura formada pelo método sol-gel, afetando as dimensões dos poros e, consequentemente, a área superficial específica (12).

Machine Learning (ML)

O número de árvores de decisão e a altura das árvores de decisões utilizados no treinamento do modelo de predição RF são apresentadas nas Figuras 2 e 3, respectivamente.

Conforme a Figura 2, observa-se o menor valor de RMSE (cerca de 3,7 m² g⁻¹) foi obtido para o treinamento do modelo realizado com 300 árvores de decisão, seguido de um pequeno aumento para valores superiores. Em relação à altura das árvores de decisão (Figura 3), observa-se que em 10 m foi o valor mínimo de RMSE (cerca de 4,0 m² g⁻¹) para árvores de decisão, sendo que para alturas superiores não houve alteração no RMSE. Sendo assim, a etapa de treinamento foi realizada com 300 árvores de decisão de 10

m de altura cada a fim de garantir um menor valor de RMSE e evitar sobre ajuste para o modelo RF.

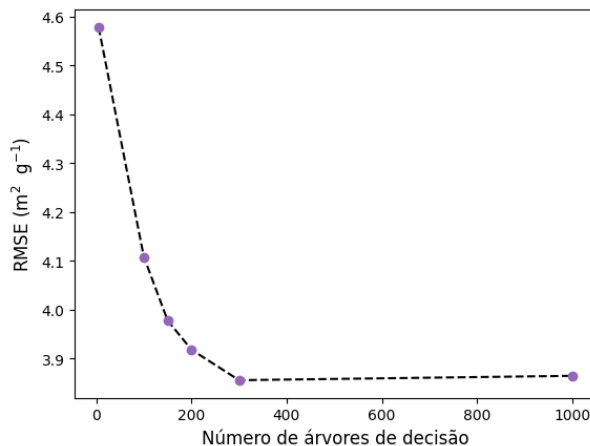


Figura 2. RMSE em função do número de árvores de decisão (altura da árvore de decisão = 10 m).

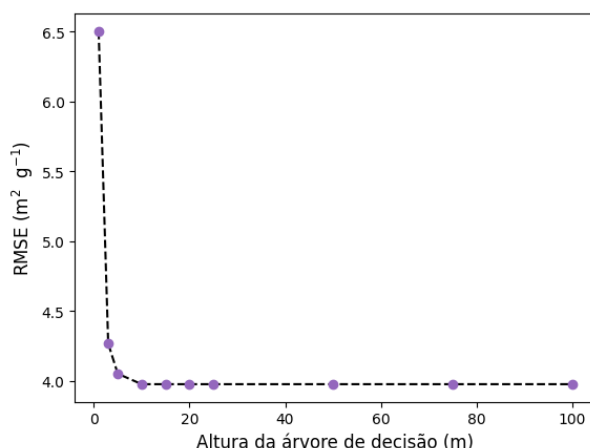


Figura 3. RMSE em função da altura da árvore de decisão (número de árvores de decisão = 300).

A Tabela 1 apresenta os parâmetros utilizados para avaliar a performance do modelo RF do ML. O coeficiente de determinação obtido para o teste ($R^2_{\text{treino}} = 0,91891$) apresentou um valor próximo a 1, indicando uma boa correlação entre os valores de área superficial específica preditos pelo modelo e os observados experimentalmente. Além disso, o valor de RMSE para o treino ($\text{RMSE}_{\text{treino}} = 3,85606 \text{ m}^2 \text{ g}^{-1}$) foi o menor do que o valor obtido para a etapa de teste ($\text{RMSE}_{\text{teste}} = 4,74493 \text{ m}^2 \text{ g}^{-1}$), confirmando a boa correlação.

Tabela 1. Parâmetros para avaliação do modelo RF pelo ML.

Treino		Teste	
R ²	RMSE (m ² g ⁻¹)	R ²	RMSE (m ² g ⁻¹)
0,91891	3,85606	0,82815	4,74493

Conforme a Tabela 1, o teste de performance do modelo RF obteve um valor de coeficiente de determinação igual a 0,82815. Esse menor valor quando comparado ao obtido no treino pode ser atribuído ao maior valor de RMSE obtido durante o teste do modelo. No entanto, o valor do coeficiente de determinação obtido no teste revela que o modelo RF consegue explicar 82,815% dos valores de área superficial específica das Nb₂O₅-NPs, em função das variáveis de entrada utilizadas (rota, precursor metálico e temperatura de tratamento térmico).

O erro relativo calculado entre o valor observado e predito é exposto na Tabela 2. A Figura 4 ilustra a correlação entre os valores preditos pelo modelo RF e os valores observados de área superficial específica (treino e teste). Grande parte dos valores observados encontram-se dentro da faixa de predição do modelo RF.

Tabela 2. Erro relativo entre valores observados e preditos pelo modelo RF.

Observado (m ² g ⁻¹)	Predito (m ² g ⁻¹)	Erro relativo (%)
58,0	66,3	14,3
174,0	163,1	6,3

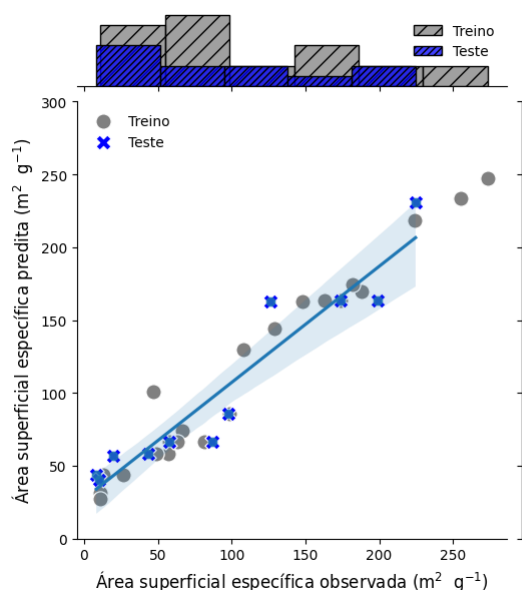


Figura 4. Correlação entre os valores de área superficial específica preditos e observados.

Conforme a Tabela 2, os erros relativos encontram-se abaixo de 15%, mostrando uma boa predição do modelo RF para valores de área superficial específica das Nb₂O₅-NPs, em função dos parâmetros de entrada adotados. Desse modo, o uso de técnicas computacionais de ML por meio do modelo RF conseguiu realizar a predição dos valores de área superficial específica das Nb₂O₅-NPs com base nas informações de rota de síntese, tipo de precursor metálico utilizado e temperatura de tratamento térmico.

Conclusões

Nesse trabalho foi possível avaliar o uso do ML como ferramenta para realizar o estudo da área superficial específica das Nb₂O₅-NPs, com base nas variáveis de síntese (rota, precursor metálico e temperatura de tratamento térmico). A análise exploratória revelou que a rota de síntese mais utilizada foi a sol-gel e o precursor majoritariamente usado foi o oxalato amoniacal de amônio. As Nb₂O₅-NPs obtidas por sol-gel apresentaram áreas superficiais específicas inferiores às obtidas por degradação térmica e hidrotérmica. Aliado a isso, o aumento da temperatura ocasionou uma redução nos valores de área superficial específica. O número de árvores de decisão e a altura das árvores de decisão para o modelo RF foram avaliadas, obtendo os valores de 300 e 10 m, respectivamente, de modo que o valor de RMSE fosse mínimo. O modelo RF utilizado para realizar a predição dos valores de área superficial específica das Nb₂O₅-NPs com base nas informações de síntese mostrou um coeficiente de determinação de 0,91891 para a etapa de treino e 0,8215 para a etapa de teste. Os erros relativos entre os valores observados e os preditos pelo modelo RF ficaram abaixo de 15%, evidenciando que o modelo RF consegue prever satisfatoriamente valores de área superficial específica de Nb₂O₅-NPs. Por conseguinte, o uso de estudos computacionais de ML mostrou-se uma ferramenta promissora para a predição de valores experimentais de área superficial específica das Nb₂O₅-NPs, podendo reduzir custos e tempo associados às etapas experimentais.

Agradecimentos

Os autores gostariam de agradecer ao Grupo de Pesquisa em Nanomateriais Aplicados (GPNAp) da Universidade Franciscana (UFN) e o Laboratório de Reciclagem, Tratamento de Resíduos e Extração (LAREX) da Universidade de São Paulo (USP) pelo suporte na realização do presente trabalho. Além disso, o apoio financeiro do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) por meio da chamada CNPq/MCTI/FNDCT Nº 23/2022 – InovaNióbio (nº 408422/2022-0).



Referências

1. O. F. Lopes *et al.*, *Quim. Nova*, **2015**, 38(1), 106-117.
2. G. H. M. Gomes *et al.*, *Ceram. Int.*, **2023**, 49, 6164-6176.
3. A. G. S. Prado *et al.*, *Appl. Catal. B*, **2008**, 82, 219-224.
4. O. F. Lopes *et al.*, *Appl. Catal. B*, **2014**, 144, 800-808.
5. A. M. Negrescu *et al.*, *J. Funct. Biomater.* **2022**, 13, 274
6. L. O. Oviedo *et al.*, *Sep. Purif. Technol.*, **2023**, 315, 123712.
7. R. Aboretti *et al.*, *Qual Reliab Engng Int.*, **2022**, 38, 1131-1156.
8. J. L. Speiser *et al.*, *Expert Syst. Appl.*, **2019**, 134, 93-101.
9. A. El Bey *et al.*, *Kem. Ind.*, **2021**, 70 (9-10), 481-488.
10. J. A. Kumar *et al.*, *J. Clean. Prod.*, **2021**, 324, 129198.
11. T. Sreethawong *et al.*, *J. Chem. Eng.*, **2013**, 215-216, 322-330.
12. R. K. Sharma *et al.*, *J. Mater. Res. Technol.*, **2023**, 24, 2629-2639.