

CLASSIFICAÇÃO E PREDIÇÃO DE PROPRIEDADES DE ALVOS CONTRA CÂNCER

DANIEL FELIE PIETEZAK¹; ROBSON DA SILVA OLIBONI²

¹*Universidade Federal de Pelotas – danielfpj98@gmail.com*

²*Universidade Federal de Pelotas – rooliboni@gmail.com*

1. INTRODUÇÃO

O câncer continua sendo uma das principais causas de morte no mundo, devido ao crescimento descontrolado das células causado por alterações na regulação do ciclo celular (SUNG et. al., 2021). Entre as moléculas mais importantes nesse processo estão aquelas que controlam o DNA e as que enviam sinais para que as células cresçam e se multipliquem. Quando essas moléculas sofrem alterações ou estão em excesso, podem causar instabilidade genética e favorecer o desenvolvimento do tumor, tornando-se alvos essenciais para novos tratamentos contra o câncer. A abordagem para lidar com ambos os alvos ao mesmo tempo tem ganhado destaque na ciência nos últimos anos (GIORDANO, 2008).

Para descobrir moléculas capazes de atuar nesses alvos, os pesquisadores podem contar com a química computacional, que permite identificar e classificar rapidamente compostos com potencial terapêutico. Existem ainda diversos bancos de dados gratuitos que armazenam informações sobre milhares de moléculas promissoras para diferentes doenças. Com o uso de algoritmos, é possível acelerar o processo de seleção e análise desses compostos, passando por etapas como coleta, armazenamento, tratamento de dados e avaliação da atividade biológica de cada molécula. Essa nova abordagem se deu graças ao advento do aumento de maior capacidade computacional nos últimos anos (CARRACEDO-REBOREDO, 2021).

O objetivo deste trabalho é reunir dados de diferentes bancos de moléculas disponíveis, unificar essas informações em uma única base e, em seguida, desenvolver algoritmos capazes de identificar e classificar a atividade de moléculas, distinguindo grupos ativos e inativos em relação a alvos conhecidos no câncer. De posse desse treinamento realizado sobre as moléculas conhecidas, o intuito do trabalho também é prever a atividade ou inatividade de novos compostos antes da síntese em laboratório.

2. METODOLOGIA

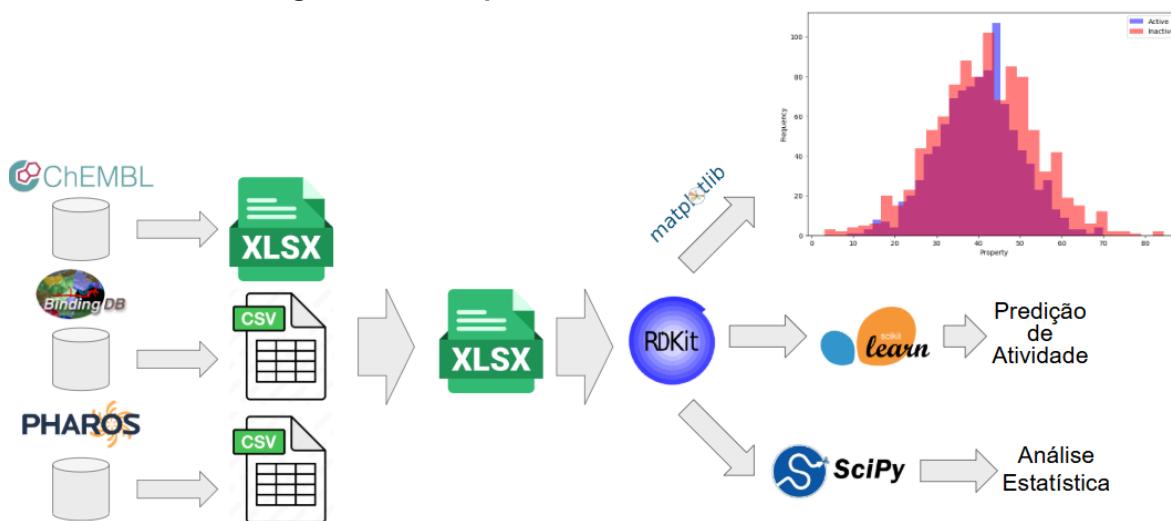
Os bancos de dados contendo as moléculas já conhecidas contra o câncer escolhidos foram ChemEMBL, BindingDB e Pharos. Um algoritmo foi escrito para coletar e unificar os dados em um arquivo padronizado, segundo Figura 1.

De posse desse arquivo padronizado, pode-se fazer uma sequência de análises estatísticas e gráficas para identificar a quantidade de moléculas ativas e inativas. Para realização disso, usam-se bibliotecas prontas como RDKit, Matplotlib, SKLearn e SciPy. O fluxo avaliado é mostrado na Figura 2.

Para realização da predição de atividade e da análise estatística, para cada molécula coletada, usou-se como atributos suas atividades moleculares: Logaritmo de partição, massa molecular, TPSA, fração de carbonos com ligações

simples, número de hidrogênios doadores, número de hidrogênio aceitadores, número de anéis aromáticos e número de ligações rotacionáveis (LAMEIRO; MONTANARI, 2023).

Figura 1: Coleta, padronização e análise dos dados

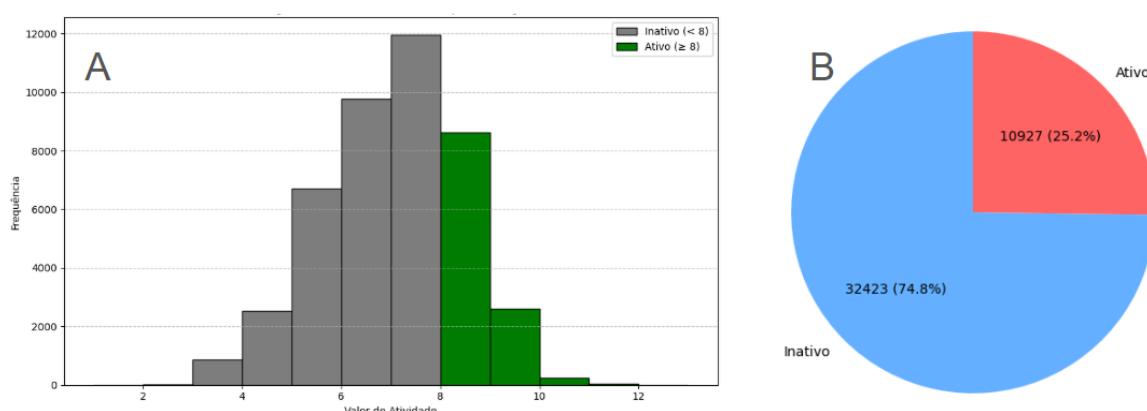


Fonte: Os autores (2025).

3. RESULTADOS E DISCUSSÃO

A análise do arquivo gerado unificado mostrou a quantidade de estruturas avaliada por nível de atividade, usando 8 como fator de corte para atividade ou inatividade, conforme mostrado na Figura 2a. Em valores absolutos, pode-se representar a quantidade relativa segundo a Figura 2b.

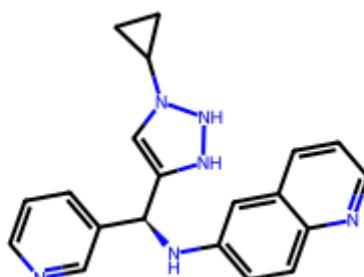
Figura 2: A) Histograma de valor de atividade por contagem. B) Gráfico de pizza para grupos de atividade



Fonte: Os autores (2025).

De posse do grupo de sistemas ativos contra o câncer, pode-se estabelecer qual o grupo molecular mais presente em moléculas ativas que não está presente no conjunto inativo, conforme Figura 3.

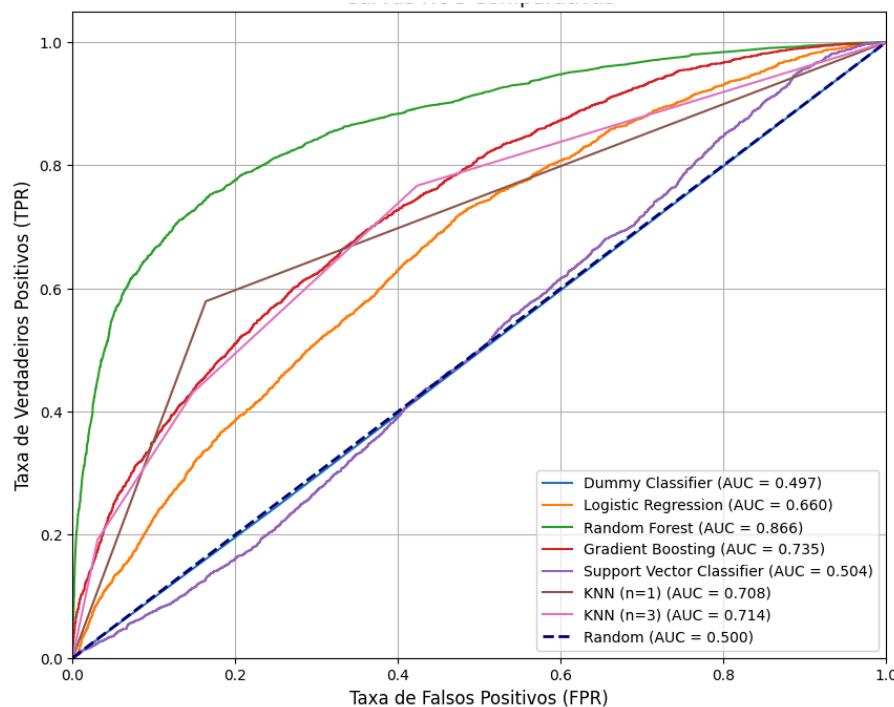
Figura 3: Grupo molecular ativo mais provável



Fonte: Os autores (2025).

Com esse treinamento realizado, é possível realizar a classificação de novos sistemas moleculares que eventualmente foram desenvolvidos posteriormente à base de dados coletada. Para esse fim, foram avaliados diferentes modelos em que Random Forest tem uma capacidade probabilística de previsão maior próxima de 90 %.

Figura 4: Curvas ROC comparativas



Fonte: Os autores (2025).

4. CONCLUSÕES

O uso de atributos moleculares mostrou-se eficaz na predição da atividade das moléculas. Esses resultados evidenciam o potencial da química computacional como ferramenta estratégica para acelerar a triagem de candidatos terapêuticos e apoiar a seleção racional de moléculas com maior probabilidade de sucesso no desenvolvimento de novas terapias anticâncer.

Os resultados obtidos neste estudo demonstram que a integração de dados de diferentes bancos de moléculas e a utilização de algoritmos computacionais permitem uma análise eficiente da atividade biológica de compostos contra alvos

conhecidos de câncer. A unificação e padronização das informações possibilitaram a identificação clara de grupos ativos e inativos, além de destacar padrões moleculares recorrentes em compostos ativos, fornecendo insights importantes para futuras investigações em química medicinal e descoberta de fármacos.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- CARRACEDO-REBOREDO, P.; LIÑARES-BLANCO, J.; RODRÍGUEZ-FERNÁNDEZ, N.; CEDRÓN, F.; NOVOA, F. J.; CARBALLAL, A.; MAJO, V.; PAZOS, A.; FERNANDEZ-LOZANO, C. A review on machine learning approaches and trends in drug discovery. *Computational and Structural Biotechnology Journal*, v. 19, p. 4538–4558, 2021.
- GIORDANO, S.; PETRELLI, A. From Single- to Multi-Target Drugs in Cancer Therapy: When Aspecificity Becomes an Advantage. *Current Medicinal Chemistry*, v. 15, n. 5, p. 422–432, 2008.
- LAMEIRO, R. F.; MONTANARI, C. A. *Investigating the Lack of Translation from Cruzain Inhibition to Trypanosoma cruzi Activity with Machine Learning and Chemical Space Analyses*. ChemMedChem, v. 18, n. 6, jan. 2023.
- SUNG, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, v. 71, n. 3, p. 209–249, 2021. Disponível em: <https://onlinelibrary.wiley.com>. Acesso em: 4 fev. 2021.