RIO

PUC

Tese de Doutorado

Aplicação de Inteligência Artificial Explicável na Geração de Alertas Estruturais de Toxicidade e de Permeabilidade na Barreira Hematoencefálica

Cayque Monteiro de Castro Nascimento

Pontifícia Universidade Católica do Rio de Janeiro
Centro Técnico Científico
Departamento de Química

Rio de Janeiro, 03 de novembro de 2025

Pontifícia
Universidade
Católica do
Rio de Janeiro

Tese de Doutorado

Aplicação de Inteligência Artificial Explicável na Geração de Alertas Estruturais de Toxicidade e de Permeabilidade na Barreira Hematoencefálica

Cayque Monteiro de Castro Nascimento

Orientação: Professor André Silva Pimentel, D. Sc.

Tese apresentada como requisito parcial para a obtenção do grau de Doutor em Química pelo programa de Pós-Graduação em Química, no Departamento de Química.

Rio de Janeiro, 03 de novembro de 2025

Pontifícia
Universidade
Católica do
Rio de Janeiro

Aplicação de Inteligência Artificial Explicável na Geração de Alertas Estruturais de Toxicidade e de Permeabilidade na Barreira Hematoencefálica

Cayque Monteiro de Castro Nascimento

Tese apresentada como requisito parcial para a obtenção do grau de Doutor em Química, pelo Programa de Pós-Graduação em Química, aprovada pela Comissão examinadora abaixo:

Professor André Silva Pimentel, D. Sc.
Orientador
Departamento de Química – PUC-Rio

Professor Nicolás Adrián Rey, D. Sc.
Departamento de Química – PUC-Rio

Professor Camilo Henrique da Silva Lima, D. Sc.
Departamento de Química Orgânica - UFRJ

Professor Fernando Carvalho da Silva, D. Sc.
Departamento de Química Orgânica - UFF

Professor Luciano Tavares da Costa, D. Sc.
Departamento de Físico-Química - UFF

Rio de Janeiro, 03 de novembro de 2025

Pontifícia
Universidade
Católica do
Rio de Janeiro

Cayque Monteiro de Castro Nascimento
Possui Graduação em Licenciatura em Química pela Universidade Federal do Rio de Janeiro (UFRJ) e Mestrado em Química pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Atua na área de Físico-Química, com ênfase em Inteligência Artificial aplicada ao estudo de fármacos.

Ficha Catalográfica

Nascimento, Cayque Monteiro de Castro Nascimento

Aplicação de Inteligência Artificial Explicável na Geração de Alertas Estruturais de Toxicidade e de Permeabilidade na Barreira Hematoencefálica / Cayque Monteiro de Castro Nascimento; Orientador: André Silva Pimentel. – Rio de Janeiro: PUC-Rio, Departamento de Química, 2025.

126 f. : il. color. ; 30 cm

Tese (Doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Química, 2025.

Inclui bibliografia

1. Química – Teses. 2. Toxicidade de compostos orgânicos. 3. Permeabilidade na Barreira Hematoencefálica. 4. Aprendizado de Máquina. 5. LIME. I. Pimentel, André Silva. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Química. III. Aplicação de Inteligência Artificial na Geração de Alertas Estruturais de Toxicidade e de Permeabilidade na Barreira Hematoencefálica.

CDD: 540

A Deus, por estar comigo em todos os momentos.

À minha família, pelo apoio incondicional.

Ao professor André, por sua extrema dedicação, paciência e companheirismo.

Agradecimentos

Agradeço, primeiramente, a Deus, por estar comigo em todo o caminho e me fortalecer em cada desafio desta jornada.

Agradeço à minha família, por todo amor, apoio e compreensão.

Agradeço ao professor André, por sua dedicação e paciência.

Agradeço aos colegas de trabalho e de universidade, por tornarem a caminhada mais leve e significativa.

Agradeço aos colegas de laboratório pelas trocas de conhecimento, especialmente ao Caio, Lucca e Paloma, pelo empenho no projeto.

Agradeço à banca examinadora, por dedicar parte de seu precioso tempo para contribuir com este trabalho.

Agradeço à CAPES, ao CNPQ e à PUC - Rio, pelo suporte concedido a este projeto.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Código de Financiamento 001.

Resumo

Nascimento, Cayque Monteiro de Castro; Pimentel, André Silva (Orientador). **Aplicação de Inteligência Artificial Explicável na Geração de Alertas Estruturais de Toxicidade e de Permeabilidade na Barreira Hematoencefálica.** Rio de Janeiro, 2025. 126 p. Tese de Doutorado - Departamento de Química, Pontifícia Universidade Católica do Rio de Janeiro.

Uma das ferramentas mais populares de Inteligência Artificial Explicável, o método Local Interpretable Model-Agnostic Explanations (LIME), foi aplicada para interpretar modelos de Aprendizado de Máquina desenvolvidos a partir de diferentes bancos de dados: um sobre toxicidade de fármacos e outro sobre permeabilidade de compostos orgânicos na Barreira Hematoencefálica. A utilização do LIME permitiu a criação de alertas estruturais, isto é, a identificação das principais subestruturas em moléculas que causam as propriedades estudadas (toxicidade e permeabilidade). Esses alertas estruturais são uma lista de estruturas que pode ser utilizada durante a triagem de candidatos a fármacos em bibliotecas virtuais, possibilitando a exclusão precoce de compostos com alto risco de insucesso antes da síntese destes compostos e dos ensaios pré-clínicos (*in vitro* e *in vivo*) em laboratórios, agências regulatórias e indústrias. LIME identificou fragmentos tóxicos descritos na literatura e fragmentos que podem facilitar a permeabilidade na Barreira Hematoencefálica, destacando principalmente subestruturas nitrogenadas. Desta forma, a abordagem contribui para reduzir desperdícios de tempo e recursos em sínteses e ensaios pré-clínicos, fornecendo subsídios para diferentes áreas, como química medicinal, síntese orgânica, farmácia e toxicologia.

Palavras – Chave

Alertas estruturais; aprendizado de máquina; LIME.

Abstract

Nascimento, Cayque Monteiro de Castro; Pimentel, André Silva. **Application of Explainable Artificial Intelligence in the Generation of Structural Alerts for Toxicity and Blood-Brain Barrier Permeability**. Rio de Janeiro, 2025. 126 p. PhD Thesis – Department of Chemistry, Pontificial Catholic University of Rio de Janeiro.

One of the most popular tools in Explainable Artificial Intelligence, the method Local Interpretable Model-Agnostic Explanations (LIME) was applied to interpret machine learning models developed from different datasets: one addressing drug toxicity and the other focusing on the permeability of organic compounds across the Blood–Brain Barrier. The use of LIME enabled the generation of structural alerts, that is, the identification of key molecular substructures responsible for the studied properties (toxicity and permeability). These structural alerts constitute a list of chemical fragments that may be used during the virtual screening of drug candidates, allowing the early exclusion of compounds with a high risk of failure before the synthesis of these compounds and preclinical essays (*in vitro* and *in vivo*) in laboratories, regulatory agencies and industries. LIME identified toxic fragments described in the literature and fragments that can facilitate permeability in the Blood-Brain Barrier, highlighting mainly nitrogenous substructures. In this way, the approach contributes to reducing waste of time and resources in synthesis and pre-clinical assays, while providing valuable insights for different fields such as medicinal chemistry, organic synthesis, pharmacy, and toxicology.

Keywords

Structural alerts; machine learning; LIME.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de abreviações e de alvos biológicos

AM: Aprendizado de Máquina
AUC: Area Under the Curve (Área sob a curva)
BHE: Barreira Hematoencefálica
DRNs: Deep Residual Networks (Redes Residuais Profundas)
ET: Extra Trees (Árvores extremamente aleatórias)
FDA: Federal Drug Administration
IA: Inteligência Artificial
LIME: Local Interpretable Model-Agnostic Explanations
ML: Machine Learning
RF: Random Forest (Árvores Aleatórias)
RNAs: Redes Neurais Artificiais
ROC: Receiver Operating Characteristics
SNC: Sistema Nervoso Central
XAI: Inteligência Artificial Explicável
NR-AR: Receptor de Andrógeno
NR-AR-LBD: Domínio de Ligação ao Ligante do Receptor de Andrógeno
NR-AhR: Receptor de Hidrocarbonetos Aromáticos
NR-Aromatase: Enzima Aromatase
NR-ER: Receptor de Estrogênio
NR-ER-LBD: Domínio de Ligação ao Ligante do Receptor de Estrogênio
NR-PPAR-γ: Receptor Ativado por Proliferador de Peroxissoma Gama
SR-ARE: Elemento Responsivo a Antioxidantes
SR-ATAD5: Regulador de ATAD5
SR-HSE: Elemento de Choque Térmico
SR-MMP: Metaloproteinase de Matriz
SR-p53: Proteína Supressora de Tumor p53

"Existe algum conflito entre a ciência e religião? Não há conflito na mente de Deus, mas muitas vezes há conflito na mente dos homens."

Henry Eyring

"A meu ver, jamais se encontrará uma verdadeira contradição entre a fé e a ciência."

Robert Andrews Millikan

1 Introdução

O Aprendizado de Máquina é um campo da ciência da computação que emergiu da convergência entre Estatística, Ciência de Dados e Inteligência Artificial. Tradicionalmente definido como “o campo de estudo que confere aos computadores a capacidade de aprender sem serem explicitamente programados”, o Aprendizado de Máquina distingue-se de abordagens determinísticas por construir modelos a partir de dados, ajustando parâmetros internos para identificar padrões, realizar inferências e previsões (PEDREGOSA, 2011; LAVECCHIA, 2015). Essa característica permite que sistemas aprendam com erros anteriores e aprimorem seu desempenho, em contraste com programas convencionais rigidamente codificados. A frase clássica “campo de estudo que confere aos computadores a capacidade de aprender sem serem explicitamente programados” é uma definição histórica de Arthur Samuel (1959), amplamente citada em livros e artigos introdutórios de Aprendizado de Máquina. No entanto, ela pode soar imprecisa ou até contraditória para leitores técnicos contemporâneos, já que programadores de fato escrevem códigos para definir modelos, algoritmos e funções de perda usadas em Aprendizagem de Máquina. O que a definição mencionou originalmente é que o comportamento final do sistema (suas decisões, previsões e classificações) não é rigidamente programado, mas emergente do ajuste automático de parâmetros com base em dados. Ou seja, o aprendizado está nos dados, não nas regras explícitas (Draft, 2016).

A técnica permeia atualmente uma vasta gama de aplicações: sistemas de recomendação, diagnóstico médico assistido, previsão de propriedades físico-químicas de compostos, análise de risco em finanças e até o planejamento de rotas em tempo real. Em áreas como Química Medicinal, o Aprendizado de Máquina desempenha papel cada vez mais estratégico, sendo capaz de identificar padrões moleculares complexos que seriam invisíveis à análise puramente estatística (NASCIMENTO et al., 2023; ROSA et al., 2024). Apesar disso, o desempenho elevado de muitos modelos preditivos não garante interpretações transparentes. Em diversas situações, mesmo especialistas não

conseguem compreender completamente como um modelo complexo chega à determinada conclusão, dando origem à expressão “modelo caixa-preta” (ALVES et al., 2016; BARR et al., 2020; HUA et al., 2022; RODRÍGUEZ-PÉREZ; BAJORATH, 2020; VARNEK; BASKIN, 2012).

Essa falta de transparência é particularmente crítica em áreas sensíveis, como saúde, toxicologia e descoberta de fármacos, nas quais decisões equivocadas podem gerar impactos diretos na segurança humana e na viabilidade de novos fármacos (que pode levar décadas e custar até bilhões de dólares). Surge, nesse contexto, a Inteligência Artificial Explicável (XAI, do inglês eXplainable Artificial Intelligence), que reúne técnicas e ferramentas para tornar compreensíveis as previsões de modelos complexos. A Inteligência Artificial Explicável busca estabelecer uma ponte entre a acurácia de algoritmos sofisticados e a necessidade de interpretabilidade, permitindo que cientistas e tomadores de decisão compreendam, validem e, quando necessário, questionem os resultados obtidos (ALI et al., 2023; JIMÉNEZ-LUNA et al., 2020; RODRÍGUEZ-PÉREZ; BAJORATH, 2021).

Entre as técnicas mais difundidas de Inteligência Artificial Explicável, destaca-se o método Local Interpretable Model-Agnostic Explanations (LIME), proposto pelos brasileiros Marco Tulio Ribeiro e Carlos Guestrin e pelo indiano Sameer Singh, em 2016 (RIBEIRO et al., 2016). O LIME opera sob o princípio de gerar um modelo interpretável tipicamente linear que aproxime localmente o comportamento de um modelo complexo em torno de uma previsão específica. Em termos práticos, o método realiza pequenas perturbações na entrada original e observa como essas mudanças afetam a saída, permitindo identificar quais características influenciam mais a decisão. Suas principais vantagens incluem a aplicabilidade a qualquer tipo de modelo (agnosticismo), a facilidade de interpretação, a capacidade de fornecer explicações alinhadas à intuição humana e o baixo custo computacional comparado a outros métodos XAI (ZAFAR; KHAN, 2021)

Contudo, o LIME não está isento de limitações. Estudos têm apontado que o método pode apresentar instabilidade nos resultados quando diferentes amostras perturbadas são geradas, o que significa que pequenas variações nos

dados de entrada podem levar a explicações consideravelmente distintas (RIBEIRO et al., 2016) . Além disso, o LIME pode ser sensível à escolha de parâmetros, como o número de perturbações, o tipo de vizinhança definida e a métrica de proximidade adotada. Essa dependência pode comprometer a reprodutibilidade e a consistência das explicações, especialmente em cenários de alta dimensionalidade, como ocorre em representações moleculares. Ainda assim, o LIME mantém grande relevância no campo da Inteligência Artificial Explicável, principalmente devido à simplicidade conceitual, flexibilidade de aplicação em diferentes tipos de modelos e velocidade relativa em comparação a outros métodos de explicação. Por essas razões, continua sendo uma ferramenta valiosa, sobretudo quando utilizado em conjunto com outras abordagens de explicação, como Anchor (RIBEIRO et al., 2018) e SHAP (RODRÍGUEZ-PEREZ, 2020), que podem complementar suas fragilidades.

Na descoberta de fármacos, a interpretabilidade do modelo é um requisito estratégico. Um dos maiores obstáculos nessa área é a toxicidade inaceitável de moléculas candidatas, que responde por uma fração significativa das falhas em ensaios clínicos (LYSENKO et al., 2018; MAHMOUD, 2021.). A Toxicologia Computacional (*in silico*) surge como ferramenta promissora para a triagem inicial de compostos, permitindo identificar estruturas potencialmente tóxicas ainda em fases de estudo pré-laboratoriais (MARCHANT, 2012; RUSYN; DASTON, 2010). Esse processo é sustentado por bases de dados públicas e privadas, que reúnem informações sobre bioatividade, toxicidade e propriedades físico-químicas, a exemplo do PAINS (*Pan-Assay Interference Compounds*) (BAELL; HOLLOWAY, 2010), que auxilia na eliminação precoce de moléculas com grupos funcionais indesejáveis, economizando tempo e recursos experimentais.

Paralelamente, a permeabilidade na Barreira Hematoencefálica (BHE) constitui um desafio adicional no desenvolvimento de fármacos para o Sistema Nervoso Central e que não podem atravessar esta barreira (BALLABH et al., 2004; DI et al., 2013). A Barreira Hematoencefálica é uma estrutura fisiológica altamente seletiva, formada por células endoteliais interligadas por junções apertadas, que restringem a passagem de compostos neurotóxicos, microrganismos e outras substâncias potencialmente nocivas. Embora

desempenhe papel protetor essencial, essa barreira também dificulta a penetração de muitos compostos que podem ser candidatos a fármacos (BEAM; ALLEN, 1977; NAU et al., 2010; SCHREIBELT et al., 2006; WANG et al., 2015; ZIPSER et al., 2007) . A capacidade de um composto atravessar a BHE depende de fatores como lipofilicidade, área de superfície polar, massa molecular e susceptibilidade a transportadores ativos. Modelar essas interações de forma acurada é crucial para antecipar a eficácia terapêutica ou a segurança de um fármaco em potencial para não atravessar a BHE (GOODWIN; CLARK, 2005; HONG et al., 2014).

A integração entre Toxicologia Computacional e predição de permeabilidade na Barreira Hematoencefálica por meio de modelos explicáveis representa uma área de fronteira de pesquisa. Por um lado, fragmentos estruturais identificados como tóxicos podem ser evitados desde o início do processo de síntese. Por outro, padrões estruturais que favoreçam ou dificultem a penetração no Sistema Nervoso Central podem orientar a modificação racional de compostos. Ainda assim, o uso de Inteligência Artificial Explicável para gerar uma lista de fragmentos como alertas estruturais para toxicidade e permeabilidade na Barreira Hematoencefálica permanece pouco explorado (ROSA et al., 2024). Mesmo quando um composto não é classificado tipicamente como tóxico, ele pode ser ineficaz ao não atingir o Sistema Nervoso Central. Portanto, toxicidade e permeabilidade na Barreira Hematoencefálica devem ser tratadas como dimensões complementares no planejamento racional de fármacos.

Dessa forma, existe uma lacuna científica clara: a ausência de metodologias integradas que utilizem modelos explicáveis para correlacionar, de forma robusta e interpretável, as características estruturais de moléculas com sua toxicidade e capacidade de atravessar a Barreira Hematoencefálica. Essa lacuna limita a eficiência da triagem computacional e a racionalização no *design* de fármacos para o Sistema Nervoso Central. Este trabalho busca preencher essa lacuna por meio da aplicação e adaptação do LIME a conjuntos de dados de toxicidade e permeabilidade, explorando a capacidade da técnica de gerar hipóteses interpretáveis e úteis para a química medicinal. Pretende-se não apenas avançar na compreensão da relação estrutura-atividade, mas também

contribuir com ferramentas práticas para a descoberta de fármacos mais seguros e eficazes.

A motivação desta Tese é contribuir para o avanço da aplicação da Inteligência Artificial Explicável no planejamento racional de fármacos, ampliando a compreensão das relações entre estruturas químicas, propriedades toxicológicas e de permeabilidade na BHE. A toxicologia computacional desempenha papel central nesse contexto, uma vez que o conhecimento de padrões estruturais associados a efeitos adversos é fundamental para o planejamento de compostos mais seguros e eficazes (MARCHANT, 2012; RUSYN; DASTON, 2010). Assim, os resultados obtidos neste trabalho podem auxiliar o desenvolvimento de fármacos voltados para diversas doenças, incluindo distúrbios neurológicos, cardiovasculares, infecciosos e oncológicos, uma vez que a toxicidade é um fator limitante transversal em praticamente todas as etapas do processo de descoberta de medicamentos.

Além disso, a metodologia proposta apresenta caráter genérico e flexível, podendo ser adaptada para a interpretação de diferentes modelos preditivos de toxicologia, abrangendo citotoxicidade, mutagenicidade (ROSA; PIMENTEL, 2024), carcinogenicidade, disrupção endócrina (ROSA et al., 2025), neurotoxicidade, hepatotoxicidade etc. Dessa forma, esta Tese busca não apenas propor um arcabouço metodológico inovador, mas também consolidar um paradigma interpretável e generalizável para apoiar a tomada de decisão em química medicinal e segurança de fármacos.

2 Objetivos

2.1 Objetivos Gerais

O objetivo do trabalho é investigar as características estruturais que influenciam na toxicidade e na permeabilidade de compostos orgânicos na Barreira Hematoencefálica via difusão passiva, por meio de modelos explicáveis de Aprendizado de Máquina e criação de alertas estruturais, visando contribuir para a compreensão de mecanismos relacionados à farmacocinética desses compostos no Sistema Nervoso Central.

2.2 Objetivos Específicos

- Construir diferentes modelos para a previsão da permeabilidade na BHE utilizando os modelos de classificação *Extra Trees*, *Random Forests* e redes neurais residuais (DRN);
- Realizar análise estatística do desempenho dos modelos, utilizando diferentes métricas;
- Aplicar um método LIME de Inteligência Artificial Explicável para identificar os fragmentos químicos que contribuem para a toxicidade e para a permeabilidade na BHE;
- Confrontar os resultados obtidos com dados experimentais da literatura especializada, avaliando a consistência dos padrões encontrados e seu potencial de generalização;
- Discutir as implicações dos alertas estruturais para o planejamento racional de fármacos voltados para a Toxicologia *in silico* e Sistema Nervoso Central, ressaltando como a integração entre Aprendizado de Máquina e Inteligência Artificial Explicável pode acelerar a descoberta de candidatos bioativos.

3 Referencial Teórico

3.1 Inteligência Artificial e Aprendizado de Máquina

Desde a conferência de Dartmouth em 1956, a Inteligência Artificial (IA) constitui um campo multidisciplinar que emergiu com o propósito inicial em reproduzir aspectos da inteligência e cognição humana por meios computacionais. Esse é considerado o marco histórico fundador da IA (John McCarthy et al., 2015). Os pioneiros John McCarthy, Marvin Minsky, Allen Newell e Herbert Simon falavam explicitamente em “fazer o computador comportar-se como se tivesse inteligência”, abordando tarefas como raciocínio lógico e simbólico, resolução de problemas, linguagem natural, percepção e aprendizado. No decorrer das últimas décadas, a IA se consolidou como uma das áreas mais dinâmicas da Ciência e da Tecnologia, sobretudo a partir da convergência entre avanços computacionais, maior disponibilidade de dados e o desenvolvimento de algoritmos estatísticos e matemáticos sofisticados. Nesse contexto, o Aprendizado de Máquina (ML, do inglês *Machine Learning*) representa um dos pilares centrais da IA e tem como característica a capacidade de extrair conhecimento a partir de dados e melhorar seu desempenho sem a necessidade de uma programação explícita.

O ML é compreendido como um conjunto de métodos que exploram a inferência estatística para estabelecer relações entre variáveis, identificar padrões ocultos e realizar previsões. É definido como o "campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados", isto é, o ML, por meio de seus algoritmos, aprende com seus erros e faz previsões sobre dados amostrais, criando modelos a partir dos dados de entrada, o que o diferencia do aprendizado por instruções (MITCHELL, 1997).

O Aprendizado de Máquina frequentemente se apoia em fundamentos estatísticos para gerar modelos preditivos capazes de identificar padrões complexos em grandes volumes de dados. Atualmente, suas aplicações

abrangem diversas áreas, como mecanismos de busca, diagnóstico médico, reconhecimento de fala e escrita, entre muitas outras. Embora tais modelos possam alcançar desempenho superior ao humano em tarefas específicas, seus resultados não devem ser interpretados como conclusivos, sendo essencial a análise crítica do especialista para assegurar que as previsões produzidas sejam coerentes com o conhecimento e o contexto do domínio estudado.

Diferentemente de abordagens determinísticas, em que a saída é rigidamente definida por um conjunto de regras pré-estabelecidas, os modelos de ML aprendem a partir da experiência, ajustando parâmetros internos com base em exemplos previamente fornecidos. Essa característica os torna particularmente úteis em domínios marcados pela alta complexidade, interdependência entre fatores e não linearidade, características típicas de problemas em Química Medicinal e Farmacocinética (MITCHELL, 1997).

Historicamente, o campo do Aprendizado de Máquina desenvolveu-se a partir de métodos estatísticos clássicos, como a regressão que é voltada à predição de valores numéricos, e classificação que é direcionada à categorização de observações em grupos qualitativos. Com o avanço da área, foram incorporadas abordagens baseadas em árvores de decisão, redes neurais artificiais e algoritmos de otimização, as quais ampliaram de maneira expressiva a capacidade preditiva e adaptativa dos modelos. A partir dos anos 2000, o aumento da capacidade computacional, aliado à disponibilidade de grandes volumes de dados, impulsionou o surgimento e a consolidação do Aprendizado Profundo (da palavra em inglês *deep learning* e sigla DL) como uma das vertentes mais promissoras do Aprendizado de Máquina contemporâneo (CHEN et al., 2018).

A aplicação de métodos de Aprendizado de Máquina na Química Medicinal impõe desafios adicionais relacionados à interpretabilidade e à validação dos resultados. Em contextos biomédicos, não basta alcançar alto desempenho preditivo: é essencial compreender os fundamentos que sustentam as previsões, de modo a garantir a coerência científica e a relevância biológica das inferências geradas (VAMATHEVAN et al., 2019). Assim, o Aprendizado de Máquina deve ser visto não apenas como uma ferramenta preditiva, mas

também como um instrumento de apoio à descoberta científica, capaz de revelar relações estruturais e propriedades moleculares subjacentes aos dados. Nesse sentido, as abordagens baseadas em ML representam um avanço metodológico relevante para lidar com a crescente complexidade dos dados químicos e biológicos contemporâneos.

O ponto de partida de qualquer aplicação de ML é a disponibilização de bancos de dados. Estes reúnem informações experimentais ou computacionais nas mais diversas áreas de interesse, tais como propriedades físico-químicas, atividades biológicas, toxicidade ou parâmetros farmacocinéticos. A qualidade do modelo gerado é diretamente dependente da qualidade do banco de dados e da quantidade de dados, sendo essencial garantir curadoria, padronização e consistência dos registros. Erros, duplicatas e inconsistências podem comprometer seriamente a acurácia e a capacidade de generalização do modelo.

Uma etapa subsequente crucial refere-se ao tratamento desses dados, que inclui desde processos de padronização e escalonamento numérico até técnicas de implementação de valores faltantes. Em contextos aplicados à Química, também é frequente a codificação de representações moleculares (por exemplo, as representações SMILES), de modo a garantir a comparabilidade e a coerência.

A representação SMILES (do termo em inglês *Simplified Molecular Input Line Entry System*) é uma linguagem de notação linear amplamente utilizada para representar compostos químicos de forma compacta e legível por computador (WEININGER, 1988). Em vez de utilizar representações gráficas, como diagramas estruturais ou fórmulas de Lewis, o SMILES descreve a conectividade atômica e o padrão de ligações por meio de uma sequência de caracteres. Essa abordagem permite o armazenamento, a pesquisa e a manipulação de estruturas químicas em bancos de dados, além de sua aplicação em modelagem molecular e em algoritmos de aprendizado de máquina.

Na notação SMILES, cada átomo é representado pelo seu símbolo químico, e as ligações entre eles são descritas segundo regras específicas. A

sequência é construída a partir de um átomo inicial, percorrendo toda a estrutura da molécula até que todos os átomos e ramificações sejam descritos. Por exemplo, o etanol é representado como CCO, onde o primeiro “C” corresponde ao carbono do grupo metil (CH_3–), o segundo “C” ao carbono do grupo metileno (–CH_2–) e o “O” ao átomo de oxigênio do grupo hidroxila (–OH). Elementos comuns, como C, N, O e S, podem ser escritos sem colchetes quando possuem valência padrão, enquanto átomos com cargas, isótopos ou estados de oxidação específicos são indicados entre colchetes, como em [Na+] para o cátion sódio e [O-] para um oxigênio carregado negativamente.

As ligações químicas simples são normalmente implícitas, enquanto ligações duplas e triplas são indicadas pelos símbolos “=” e “#”, respectivamente. Ligações aromáticas são representadas por letras minúsculas, como em c1ccccc1 para o benzeno. Ramificações dentro da estrutura são indicadas por parênteses, como no caso do isopropanol (CC(C)O), em que o grupo “(C)” representa uma ramificação ligada ao segundo carbono. Estruturas cíclicas são descritas utilizando números que marcam o início e o fim de um anel. Assim, o ciclohexano é escrito como C1CCCCC1, e o benzeno como c1ccccc1, onde o número “1” indica o fechamento do ciclo.

O SMILES também é capaz de representar estereoquímica, incluindo quiralidade e isomeria geométrica. Centros quirais são indicados pelos símbolos “@” e “@@”, que descrevem a configuração absoluta (R ou S) do átomo central, enquanto as isomerias cis/trans em duplas ligações são representadas pelos símbolos “/” e “\”, que definem a orientação relativa dos substituintes. Essa capacidade torna o SMILES adequado para descrever moléculas tridimensionais de forma textual. Um dos principais desafios ainda em aberto na representação SMILES está relacionado à descrição completa e inequívoca de estereoisômeros.

Embora a notação SMILES permita indicar quiralidade e isomeria geométrica por meio de símbolos como “@”, “@@”, “/” e “\”, essa codificação depende da ordem de escrita dos átomos na cadeia e das convenções adotadas por diferentes implementações de software. Como consequência, uma mesma molécula pode ser representada por múltiplos SMILES distintos, e pequenas

variações na string podem levar a ambiguidades na reconstrução da estrutura tridimensional, especialmente em moléculas com múltiplos centros quirais, ligações duplas conjugadas ou conformações flexíveis. Para superar essas limitações, foram desenvolvidos sistemas alternativos capazes de descrever moléculas de forma canônica e invariável à forma de entrada, como o InChI (International Chemical Identifier), criado pela IUPAC, o SELFIES (Self-Referencing Embedded Strings) e o fragSMILES recentemente desenvolvido em 2025, projetado para garantir representações válidas e únicas mesmo em aplicações de aprendizado de máquina. Essas abordagens oferecem maior robustez e precisão estrutural, tornando-se preferíveis em contextos que exigem reprodutibilidade e interoperabilidade química.

Existem diferentes variantes de SMILES, entre elas o SMILES canônico, gerado por algoritmos que garantem uma representação única e padronizada para cada molécula; o SMILES isomérico, que inclui informações de estereoquímica e isomeria; e o SMILES arbitrário, que representa qualquer forma válida, mas não necessariamente única. Um exemplo completo pode ser dado pelo ácido lático, cuja fórmula molecular é CH_3–CH(OH)–COOH e cujo SMILES correspondente é CC(O)C(=O)O. Nessa representação, o grupo (O) indica a hidroxila ligada ao segundo carbono, enquanto C(=O)O descreve o grupo carboxílico.

O uso do SMILES é onipresente em bases de dados químicas como PubChem, ChEMBL e ZINC, além de ser essencial em tarefas de modelagem molecular, análises QSAR/QSPR e algoritmos de aprendizado de máquina, nos quais ele serve como formato de entrada textual para a construção de representações moleculares, como impressões digitais e incorporações (das palavras em inglês *fingerprints* e *embeddings*, respectivamente). Embeddings, que em inglês significa “incorporar”, é um termo utilizado em IA e Processamento de Linguagem Natural (PLN). Refere-se ao processo de “incorporar” ou “embutir” informações complexas (como palavras, frases ou documentos) em um espaço vetorial. A principal vantagem destas representações é a compactação eficiente e a compatibilidade com diversos softwares, o que facilita a interoperabilidade entre ferramentas químicas e modelos computacionais. No entanto, o SMILES

apresenta algumas limitações, como a existência de múltiplas representações possíveis para uma mesma molécula e a perda de informações tridimensionais, como geometria e conformação.

Outro desafio recorrente em bancos de dados aplicados ao ML é a presença de dados desbalanceados. Essa situação emerge quando determinadas classes (como compostos ativos ou permeáveis) são representadas em número muito inferior às demais, o que induz o modelo a favorecer a classe majoritária. Diversas estratégias têm sido desenvolvidas para mitigar esse problema, incluindo técnicas de sobreamostragem ou subamostragem, bem como a utilização de algoritmos robustos para balanceamento e métricas de avaliação que penalizam o favorecimento de uma das classes devido à falta de balanceamento do banco de dados.

No âmbito do ML aplicado à Química, um aspecto central é a forma com que os compostos são representados. Como algoritmos de ML operam sobre variáveis numéricas, estruturas químicas precisam ser codificadas por vetores de características de números binários. Nesse contexto, utilizam-se descritores moleculares e impressões digitais, que descrevem informações estruturais, topológicas e físico-químicas em formatos computacionalmente manipuláveis. Os descritores podem incluir propriedades simples, como massa molecular e lipofilicidade, até medidas mais sofisticadas, como índices topológicos ou descritores tridimensionais. Já as impressões digitais moleculares são representações que codificam a presença ou ausência de subestruturas químicas específicas.

Dentre as impressões digitais mais empregadas, destaca-se o ECFP (do termo em inglês *Extended Connectivity Fingerprints*), que são impressões digitais topológicas circulares projetadas para caracterização molecular, busca por similaridade e modelagem de estrutura-atividade (ROGERS; HAHN, 2010). Elas estão entre as ferramentas de busca por similaridade mais populares na descoberta de fármacos e são utilizadas com eficácia em uma ampla variedade de aplicações. Dentre as principais propriedades dos ECFPs, pode-se citar: as representações estruturais moleculares por meio de vizinhanças de átomos, a velocidade com que podem ser calculados, a presença de subestruturas

particulares em suas representações, o grande número de características moleculares diferentes que podem representar e o fato de poderem ser projetados para representar tanto a presença quanto a ausência de funcionalidade.

Os ECFPs são uma forma de representação molecular baseada na estrutura topológica dos compostos químicos, projetada para traduzir a conectividade atômica e o ambiente químico local de cada átomo em um vetor numérico ou binário. Diferentemente de descritores globais, que condensam propriedades físico-químicas de toda a molécula, os ECFPs capturam características locais e estruturais, permitindo que algoritmos computacionais reconheçam semelhanças moleculares de maneira mais granular. Essa representação é especialmente útil em tarefas de predição de atividade biológica, triagem virtual, QSAR (do termo em inglês *Quantitative Structure Activity Relationship*) e em modelos de aprendizado profundo voltados à descoberta de compostos bioativos (DEVILLERS, 2013; DOWEYKO, 2008).

O princípio fundamental dos ECFPs deriva do algoritmo de Morgan, proposto originalmente na década de 1960, que identifica átomos equivalentes com base em sua vizinhança química. Na versão estendida do algoritmo, utilizada para gerar ECFPs, cada átomo da molécula é inicialmente atribuído a um identificador (ou *hash*) que codifica informações sobre o tipo atômico, hibridização, carga formal, número de ligações, presença de heteroátomos vizinhos e participação em anéis. Uma função de *hash* é uma função matemática que pega dados de tamanho variável e os transforma em um valor de tamanho fixo, conhecido como *hash* ou resumo. Elas são usadas para verificar a integridade dos dados e para segurança, como o armazenamento de senhas, pois é um processo unidirecional e a informação original não pode ser recuperada a partir do *hash*. Em seguida, a representação é construída de forma iterativa e hierárquica, expandindo a vizinhança de cada átomo em sucessivos “raios” de conectividade, geralmente de raio 2, 4 ou 6 ligações químicas, dependendo da profundidade desejada. Cada iteração combina as informações do átomo central com as de seus vizinhos, gerando novos identificadores únicos que representam subestruturas químicas locais (como grupos funcionais, anéis ou fragmentos específicos).

Após a expansão, todos os identificadores gerados são convertidos em valores numéricos fixos, armazenados em um vetor binário conhecido como impressão digital. Esse vetor indica a presença (1) ou ausência (0) de determinados fragmentos estruturais na molécula. O comprimento do vetor é definido pelo usuário com valores comuns são 1.024, 2.048 ou 4.096 bits e, devido ao uso de funções de *hash*, diferentes subestruturas podem ocasionalmente ocupar a mesma posição (colisão de bits), embora esse efeito seja estatisticamente reduzido em vetores longos. O resultado é uma assinatura digital da molécula, que preserva a informação estrutural relevante e pode ser diretamente utilizada em cálculos de similaridade (como o coeficiente de Tanimoto), em buscas em bancos de dados ou como entrada em modelos de aprendizado de máquina.

Os ECFPs apresentam diversas vantagens em relação a outras representações químicas. Eles são invariantes a rotações, translações e ordenação dos átomos, garantindo que moléculas idênticas produzam a mesma impressão digital independentemente de sua forma de entrada. Além disso, são altamente interpretáveis, pois cada número binário pode ser rastreado até o fragmento químico correspondente, permitindo identificar quais subestruturas contribuem para uma determinada previsão, que é um aspecto relevante em estudos de explicabilidade de modelos. Essa característica faz dos ECFPs uma ponte eficiente entre a estrutura química e o espaço matemático de aprendizado. A figura 1 representa a lógica de funcionamento do ECFP.

Apesar de sua ampla aplicação, os ECFPs também possuem limitações. Como se baseiam unicamente na conectividade 2D, eles não capturam informações tridimensionais, como conformações espaciais ou interações intramoleculares. Além disso, a escolha do raio de expansão influencia fortemente a capacidade de generalização: raios pequenos podem perder informações estruturais importantes, enquanto raios muito grandes podem gerar redundância e aumentar o custo computacional. Ainda assim, sua eficiência e robustez tornaram os ECFPs um dos padrões de referência em representações moleculares para aprendizado de máquina, sendo implementados em bibliotecas amplamente utilizadas, como RDKit, e servindo de base para comparações com métodos mais recentes de codificação, como Graph Neural Networks (GNNs) e

incorporações moleculares (do inglês *molecular embeddings*) baseados em aprendizado profundo.

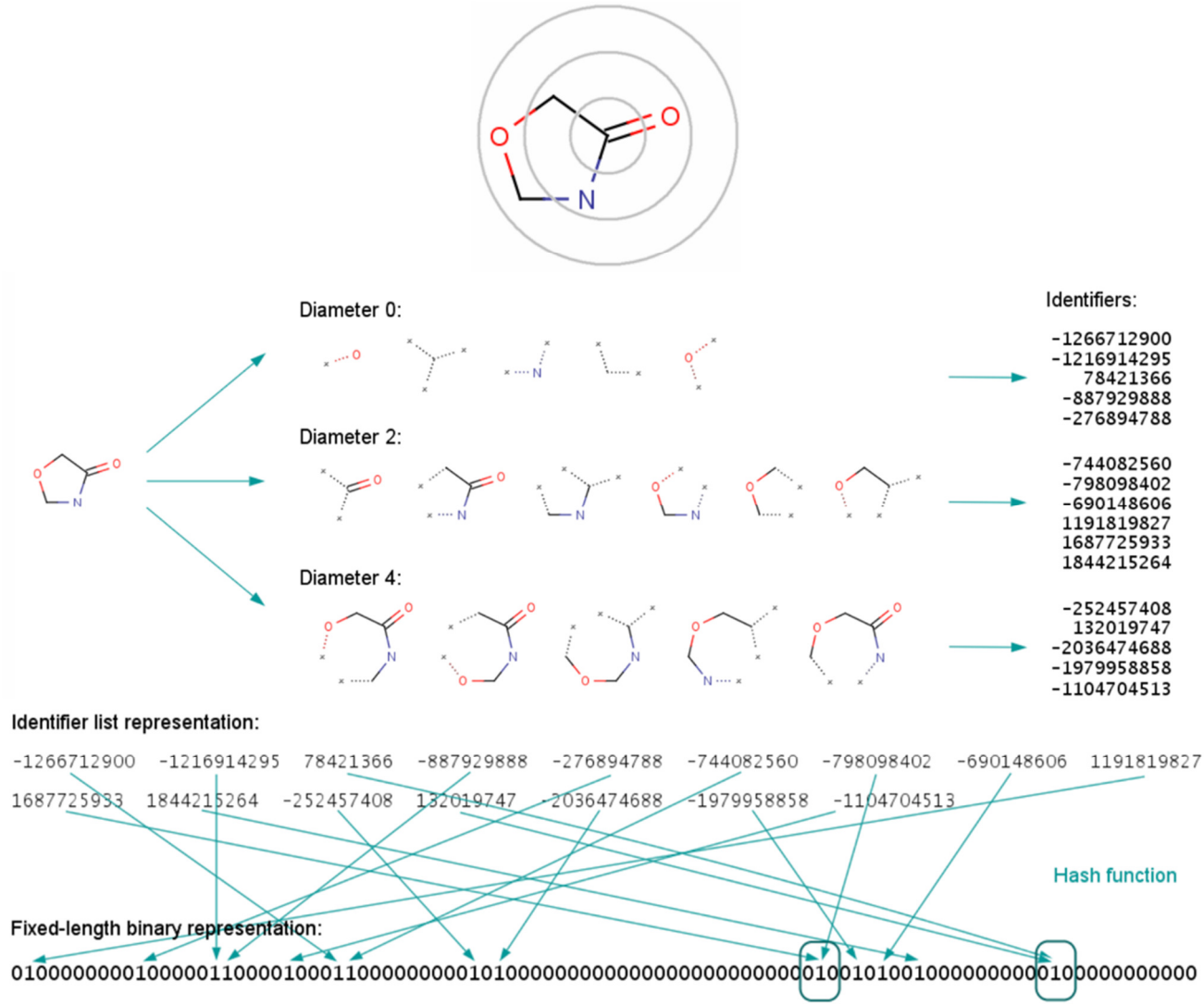


Figura 1: Ilustração da caracterização topológica de uma molécula, processo de geração do ECFP e da geração de uma sequência de números binários (Chemaxon.; Rogers and Hahn, 2010).

O arcabouço do ML aplicado a sistemas químicos e biológicos exige a integração sinérgica de diferentes etapas: construção, curadoria e tratamento de bancos de dados, além da seleção criteriosa das representações moleculares. O domínio desses conceitos constitui requisito essencial não apenas para a implementação de modelos preditivos robustos, mas também para assegurar

que as inferências extraídas possuam validade científica e possam, efetivamente, contribuir para a tomada de decisão em contextos experimentais e clínicos.

A Química é um campo fértil para a aplicação de métodos de ML em virtude da disponibilidade crescente de bases de dados experimentais e da complexidade inerente aos sistemas estudados. Em química medicinal, por exemplo, técnicas de ML são amplamente utilizadas para construção de modelos QSAR, que relacionam descritores moleculares com propriedades biológicas ou farmacológicas (CHERKASOV et al., 2014). No campo da toxicologia preditiva, algoritmos de classificação vêm sendo aplicados para identificação de compostos com potenciais efeitos adversos, permitindo antecipar riscos antes mesmo de etapas experimentais onerosas (MARCHANT, 2012; RAIES; BAJIC, 2016; RUSYN; DASTON, 2010).

Os modelos de ML frequentemente apresentam limitações no que se refere à sua interpretabilidade, sendo por isso denominados “caixas-pretas” em sua maioria (WU et al., 2023). Esse termo decorre do fato de que o processo pelo qual tais algoritmos tratam os dados para gerar previsões nem sempre é acessível ou inteligível, mesmo para os especialistas responsáveis por sua concepção. Isso ocorre porque esses modelos são ajustados diretamente a partir dos dados, explorando relações complexas e muitas vezes não lineares, cujo funcionamento interno dificilmente pode ser traduzido em regras claras. A ausência de transparência, entretanto, constitui um desafio crítico, uma vez que a explicabilidade é essencial para conferir confiabilidade, rastreabilidade e aceitação científica aos resultados. A falta desses requisitos dificulta a compreensão de quais fatores/características são, de fato, as mais importantes e que orientam as previsões (RODRÍGUEZ-PÉREZ; BAJORATH, 2021). A figura 2 representa didaticamente um modelo “caixa-preta”.

Nesse contexto, surge a Inteligência Artificial Explicável (XAI, do termo em inglês *eXplainable Artificial Intelligence*)**,** que reúne um conjunto de métodos e ferramentas voltados a tornar os modelos mais compreensíveis, contrastando diretamente com a noção de modelos opacos. Ao oferecer interpretações sobre como uma decisão é alcançada, a XAI não apenas fortalece a confiança de usuários e pesquisadores, mas também possibilita a contestação de previsões e

a identificação de oportunidades de ajuste ou aprimoramento dos algoritmos. De acordo com Ali e colaboradores (2023), a XAI tem influenciado profundamente o desenvolvimento de novos modelos de ML, nos quais a capacidade de explicação passa a ser integrada ao próprio processo de aprendizado, conferindo caráter intrínseco à interpretabilidade (ALI et al., 2023).

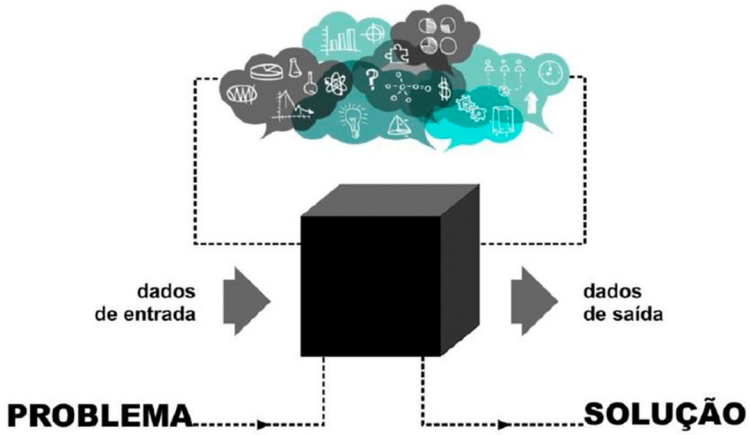


Figura 2: Representação esquemática de um modelo ‘’caixa-preta’’ (MARTINO, 2015)

No paradigma tradicional de modelagem em aprendizado de máquina, o processo de treinamento resulta em uma função preditiva complexa que mapeia dados de entrada em saídas ou recomendações, sem, contudo, oferecer clareza quanto aos critérios internos que fundamentam tais previsões. Essa característica de opacidade algorítmica, frequentemente associada a modelos de natureza não linear e de alta dimensionalidade, como redes neurais profundas e ensembles de árvores, limita a capacidade de compreender os fatores determinantes das decisões geradas. Tal limitação compromete a transparência epistemológica do modelo e dificulta a análise de aspectos essenciais, como a justificativa de uma predição específica, a identificação das causas de erros sistemáticos e a delimitação das condições sob as quais o modelo apresenta desempenho confiável. A Figura 3 ilustra de forma comparativa a distinção entre o fluxo tradicional de modelos de ML e a abordagem proposta pela XAI.

Em contraste, o paradigma da Inteligência Artificial Explicável propõe a integração explícita de mecanismos de interpretabilidade e transparência ao

longo de todo o fluxo de modelagem. Esses mecanismos podem incluir os modelos intrinsecamente explicáveis, que possuem estrutura interpretável por natureza (como árvores de decisão, regressões lineares esparsas ou modelos de regras), e os métodos pós-hoc de explicação, aplicados a modelos complexos já treinados (tais como LIME (RIBEIRO et al., 2016), SHAP (RODRÍGUEZ-PÉREZ, 2020) ou Anchor (RIBEIRO et al., 2018). O objetivo central é permitir que o usuário ou especialista compreenda a relação causal e estatística entre variáveis de entrada e saídas preditas, de modo a tornar explícitas as razões subjacentes às decisões do sistema.

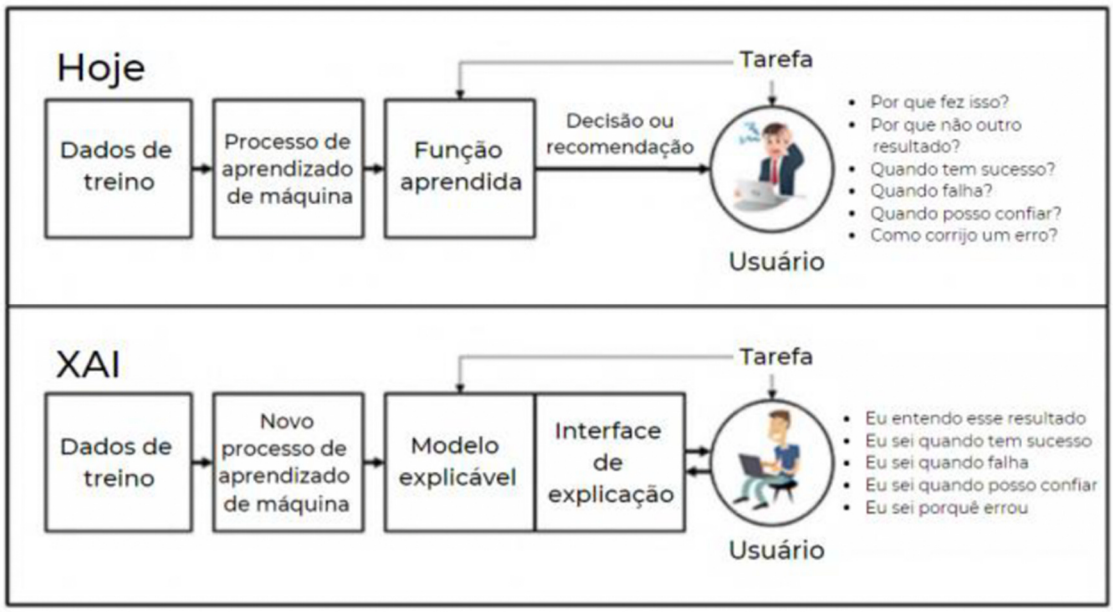


Figura 3: Implementação da etapa de explicabilidade junto aos modelos de Aprendizado de Máquina (NetApp, 2022 – ADAPTADO)

Dessa forma, o processo de interação entre o ser humano e o modelo torna-se bidirecional e interpretativo, em vez de puramente reativo. O usuário passa a compreender não apenas o resultado produzido, mas também as condições que o justificam, as fontes potenciais de incerteza, os cenários de generalização e as limitações estruturais do modelo. Esse incremento de explicabilidade e rastreabilidade promove maior confiança, auditabilidade e responsabilidade científica no uso de sistemas de aprendizado de máquina,

possibilitando a detecção de vieses, a correção de inconsistências e a validação do conhecimento extraído de forma alinhada aos princípios da metodologia científica.

A crescente complexidade dos modelos de ML aplicados às Ciências torna o desafio da explicabilidade ainda mais crítico. Em Química, não é incomum a necessidade de análises de informações em que os dados são multifacetados e altamente heterogêneos, como estruturas químicas, perfis de atividade biológica e informações de toxicidade. A natureza “caixa-preta” desses algoritmos pode comprometer a confiança e a utilidade prática dos resultados obtidos. É nesse cenário que a Inteligência Artificial Explicável se consolida como ferramenta essencial, oferecendo recursos que permitem compreender quais padrões químicos e biológicos estão sendo explorados pelo modelo para justificar determinada predição.

3.2 Modelos de Classificação e Redes Neurais

Em aprendizagem de máquina, existem três formas diferentes de aprendizado: aprendizado supervisionado, não supervisionado e semi-supervisionado. No aprendizado supervisionado, o modelo é treinado a partir de um conjunto de dados rotulados, ou seja, exemplos nos quais as variáveis de entrada (características) estão associadas a saídas conhecidas (rótulos ou valores-alvo) (Mitchell, 1997). O objetivo é aprender uma função que generalize essa relação de modo a prever corretamente o rótulo de novas amostras não vistas. Essa abordagem é amplamente utilizada em tarefas como classificação, regressão e reconhecimento de padrões, em que a disponibilidade de dados anotados garante um processo de aprendizado guiado e controlado. Por outro lado, o aprendizado não supervisionado lida com dados não rotulados, buscando identificar estruturas, agrupamentos ou regularidades implícitas nas observações. Nesse caso, o modelo não recebe instruções explícitas sobre as saídas corretas, mas procura organizar as informações de forma a revelar relações latentes entre os dados, sendo amplamente aplicado em tarefas de agrupamento, redução de dimensionalidade e detecção de anomalias. Já o

aprendizado semi-supervisionado combina aspectos dos dois paradigmas anteriores, utilizando um pequeno conjunto de dados rotulados em conjunto com um volume muito maior de dados não rotulados. Essa estratégia busca aproveitar a informação estrutural presente nos dados não rotulados para aprimorar o desempenho do modelo supervisionado, sendo especialmente útil em contextos nos quais a rotulagem é demorada ou requer conhecimento especializado, como ocorre frequentemente nas áreas biomédica, química e farmacológica.

Os modelos de classificação constituem uma das principais abordagens do aprendizado supervisionado (isto é, aprendizado cujo modelo é treinado a partir de um conjunto de dados rotulado), sendo utilizados para atribuir instâncias a categorias previamente definidas (Hastie et al., 2009; Schiff, 2011). A tarefa de classificação é especialmente relevante em contextos químicos, nos quais compostos podem ser rotulados como “ativos/inativos”, “tóxicos/não tóxicos” ou “permeáveis/não permeáveis”. O objetivo central desses modelos é aprender padrões a partir de um conjunto de treinamento para, posteriormente, realizar inferências sobre novos dados.

Dentre os algoritmos amplamente empregados, destacam-se os modelos baseados em Árvores de Decisão. A estrutura baseia-se em uma decomposição hierárquica do espaço de atributos, em que cada nó interno representa uma condição de divisão de uma variável, enquanto os ramos indicam os possíveis resultados dessa condição, culminando em folhas que correspondem às predições finais (QUINLAN, 1986). A representação esquemática de uma Árvore de Decisão é mostrada na figura 4.

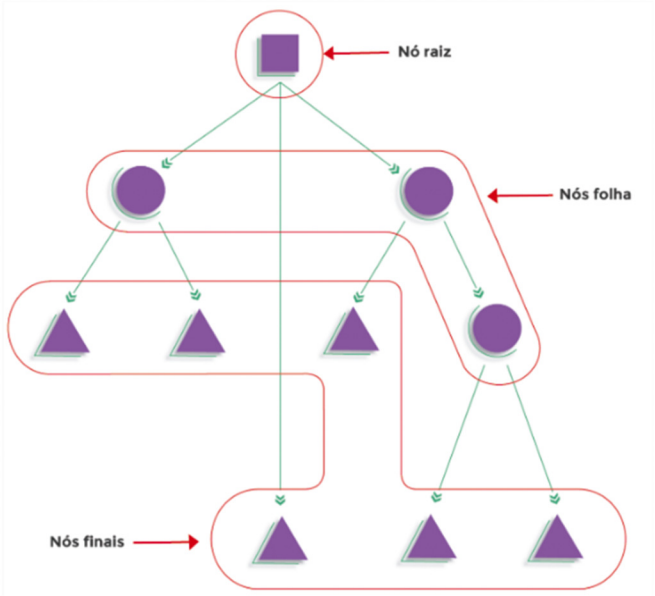


Figura 4: representação esquemática de uma Árvore de Decisão

Esse processo de particionamento sucessivo busca maximizar a pureza das classes em cada subdivisão. Árvores de Decisão isoladas são suscetíveis ao sobreajuste (do inglês *overfitting*), pois tendem a se adaptar excessivamente às particularidades do conjunto de treinamento. Por essa razão, surgiram métodos que combinam múltiplas árvores para melhorar a capacidade de generalização, explorando diversidade e estabilidade no processo de votação coletiva. Dessa forma, as Árvores de Decisão constituem tanto um modelo independente quanto a base conceitual de algoritmos mais sofisticados em aprendizado supervisionado (Quinlan, 1986).

Florestas aleatórias (do termo em inglês *Random Forest*, RF) (BREIMAN, 2001) e Árvores extras (do termo em inglês *Extra Trees*, ET) (GEURTS et al., 2006) são modelos que se caracterizam por construir múltiplas árvores de decisão e realizar a predição por meio de votação majoritária (classificação) ou média (regressão). No caso das florestas aleatórias, as árvores são treinadas com subconjuntos aleatórios de instâncias e variáveis, o que promove diversidade no conjunto de classificadores. Já o método árvores extras adota uma aleatoriedade ainda mais intensa, pois seleciona os pontos de corte dos atributos de forma randômica, reduzindo a variância e mitigando o risco de sobreajuste. Esses modelos são particularmente valorizados por sua capacidade

de lidar com bases de dados de alta dimensionalidade e pela interpretação das importâncias das variáveis, que permitem identificar quais atributos contribuem mais para o resultado.

As florestas aleatórias são um método de ensemble construído a partir de múltiplas árvores de decisão, cujo objetivo é reduzir a variância e aumentar a robustez preditiva em relação a modelos individuais. Cada árvore é treinada com uma amostra *bootstrap* dos dados, ou seja, uma amostra aleatória com reposição do conjunto de treinamento, e, a cada divisão de nó, considera-se apenas um subconjunto aleatório das variáveis explicativas. Essa combinação de bootstrap e seleção aleatória de características promove diversidade entre as árvores, reduzindo a correlação entre elas, o que, matematicamente, diminui a variância do ensemble sem aumentar significativamente o viés. A predição final das florestas aleatórias é obtida por votação majoritária em problemas de classificação ou média aritmética em problemas de regressão, garantindo uma estimativa agregada mais estável do que qualquer árvore isolada (BREIMAN, 2001a).

As árvores extras são um método com árvores extremamente randomizadas que seguem a mesma lógica de ensemble baseada em árvores, mas introduzem uma aleatoriedade adicional durante a construção das divisões. Diferentemente das florestas aleatórias, que seleciona a separação ótima para cada variável candidata com base em critérios de impureza (como entropia cruzada, por exemplo), as árvores extras escolhem o ponto de corte aleatoriamente dentro de um intervalo possível para cada característica. Essa estratégia aumenta ainda mais a diversidade entre as árvores do ensemble, reduzindo a correlação e, consequentemente, a variância do modelo final. Em termos de viés-variância, a aleatoriedade extra tende a aumentar ligeiramente o viés em comparação as florestas aleatórias, mas a redução da variância frequentemente resulta em melhor generalização, especialmente em conjuntos de dados ruidosos ou de alta dimensionalidade, além de acelerar o treinamento, pois não é necessário calcular o ponto de corte ótimo em cada divisão (Geurts et al., 2006).

Ambos os métodos são particularmente úteis em cenários com interações não lineares, alta dimensionalidade e variáveis categóricas ou contínuas

heterogêneas, sendo amplamente aplicados em química computacional, biologia computacional, análise de dados financeiros e detecção de padrões complexos. Além disso, tanto florestas aleatórias quanto árvores extras permitem calcular importâncias de variáveis, fornecendo métricas interpretáveis que auxiliam na seleção de características e na compreensão de quais fatores mais influenciam a predição, aspecto crucial em aplicações científicas e regulatórias.

No desenvolvimento da Ciência Computacional recente, as Redes Neurais Artificiais (RNAs) (LeCun et al., 2015; Schmidhuber, 2015) e as redes neurais mais complexas desenvolvidas posteriormente constituem os modelos mais expressivos do aprendizado de máquina, inspirando-se de forma abstrata na dinâmica de processamento do cérebro humano. Sua estrutura é organizada em camadas hierárquicas, sendo uma camada de entrada, que recebe os dados brutos ou pré-processados, uma ou mais camadas ocultas, responsáveis por extrair representações intermediárias, e uma camada de saída, que fornece o resultado, conforme mostrado na figura 5. Essa organização permite às RNAs capturar relações altamente não lineares e complexas, tornando-as adequadas para problemas de elevada dimensionalidade, como a predição de propriedades químicas e biológicas .

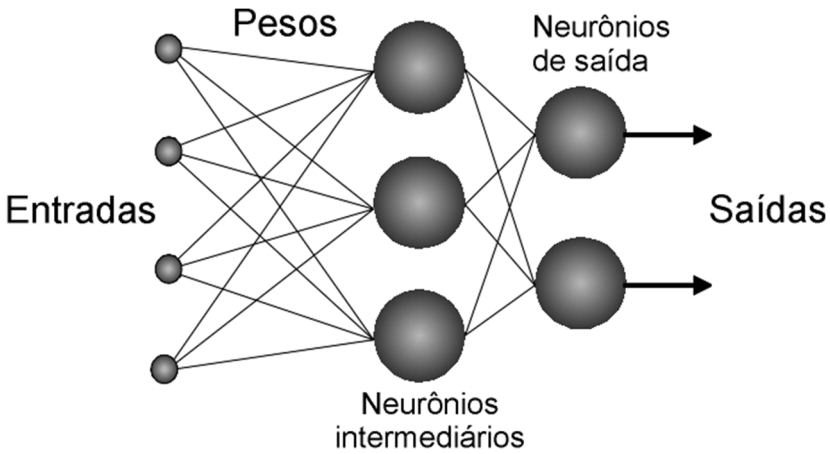


Figura 5: Ilustração das camadas das redes neurais artificiais (Revista Cérebro & Mente, 1998).

O neurônio artificial realiza uma combinação ponderada das variáveis de entrada. Formalmente, o valor de ativação z é calculado como:

$$z = \sum_{i=1}^{n} w_i x_i + b$$

em que x_i representam as entradas, w_i os pesos atribuídos a cada conexão e b, o viés, que desloca a função de ativação, garantindo maior flexibilidade. O resultado z é então processado por uma função de ativação (como ReLU, sigmoide ou tangente hiperbólica), responsável por introduzir não linearidade ao sistema, permitindo que a rede modele padrões complexos que não seriam capturados por funções lineares simples.

O processo de treinamento de uma RNA envolve a definição de uma função de custo, que quantifica o erro entre as predições e os valores esperados. Por meio do algoritmo de retropropagação do erro (do termo em inglês *backpropagation*), os pesos e vieses são ajustados iterativamente utilizando técnicas de otimização, como o gradiente descendente estocástico e suas variantes. Esse mecanismo permite que a rede refine progressivamente suas representações internas, tornando-se capaz de generalizar para dados não vistos.

O método Gradiente Descendente Estocástico (do termo em inglês *Stochastic Gradient Descent*, SGD) é um método fundamental de otimização para treinamento de modelos de aprendizado de máquina, especialmente redes neurais (BOTTOU, 2010). Apesar de sua simplicidade conceitual, a versão pura do SGD apresenta algumas limitações, como convergência lenta, oscilações em vales íngremes e sensibilidade à escolha da taxa de aprendizado. Por isso, ao longo dos anos, foram desenvolvidas diversas variantes e aprimoramentos que introduzem momentum, adaptação da taxa de aprendizado ou combinações de estratégias. Entre as principais variantes, destacam-se:

1. SGD com momento, que adiciona um termo de momento na atualização dos parâmetros, acumulando uma fração dos gradientes passados e funcionando como uma “inércia” no processo de atualização dos pesos. Com isso, o otimizador consegue avançar mais rapidamente em direções onde o

gradiente é consistente ao longo das iterações e, ao mesmo tempo, suavizar oscilações que surgem em regiões com geometria irregular no espaço de perda. A atualização dos parâmetros passa a depender tanto da taxa de aprendizado quanto da velocidade acumulada pelo coeficiente de momento.

2. Nesterov Accelerated Gradient (NAG) é uma modificação do SGD com momentum que calcula o gradiente não no ponto atual do parâmetro, mas no ponto previsto após o passo de momentum. Isso permite correções mais precisas e acelera a convergência.

3. Adagrad (do termo em inglês Adaptive Gradient Algorithm) que ajusta dinamicamente a taxa de aprendizado para cada parâmetro individual, diminuindo o passo para parâmetros que apresentam grandes gradientes e aumentando para parâmetros menos atualizados. É especialmente útil em problemas esparsos.

4. RMSProp (do termo em inglês Root Mean Square Propagation) que aprimora o Adagrad limitando o acúmulo histórico de gradientes por uma média exponencialmente ponderada, evitando que a taxa de aprendizado se torne excessivamente pequena em iterações posteriores.

5. Adam (do termo em inglês Adaptive Moment Estimation), que combina os conceitos de momento e RMSProp, mantendo médias móveis do gradiente e do quadrado do gradiente, ajustando individualmente a taxa de aprendizado para cada parâmetro de forma adaptativa. Essa variante tornou-se a mais utilizada em redes neurais profundas devido à sua eficiência e estabilidade.

6. AdaMax, que é um refinamento do Adam, com pequenas alterações na normalização ou na forma de calcular o momento, buscando melhorar a convergência ou a estabilidade em certos tipos de redes. Outra variante recente é o algoritmo AdamW, onde a degradação do peso não se acumula no momento nem na variância. Em suma, essas variantes foram criadas para superar limitações do SGD puro, como convergência lenta, oscilações, sensibilidade à escala dos gradientes e dificuldade em escapar de mínimos locais ou platôs rasos, sendo escolhidas de acordo com a arquitetura da rede, o tipo de dados e os objetivos de otimização do modelo.

Com a evolução das arquiteturas, surgiram variantes mais sofisticadas, como as Redes Residuais Profundas (DRNs, do termo em inglês *Deep Residual Networks*), que incorporam conexões de atalho entre camadas (KAIMING et al., 2016). Essas conexões residuais permitem o fluxo direto de informações e gradientes entre diferentes níveis da rede, mitigando problemas como o desvanecimento do gradiente e possibilitando o treinamento de modelos com centenas ou milhares de camadas. Tal avanço ampliou significativamente o poder preditivo das redes neurais em tarefas de classificação complexas, consolidando seu protagonismo em aplicações científicas e industriais.

O principal componente das DRNs são as conexões de atalho (do termo em inglês *skip connections*), que permitem que a entrada de uma camada seja adicionada diretamente à saída de camadas subsequentes, criando blocos residuais que aprendem a função de correção em relação à identidade, em vez de tentar modelar a transformação completa. Essa estratégia facilita o fluxo de gradientes durante a retropropagação, reduzindo a degradação do desempenho em redes muito profundas e possibilitando a construção de modelos com centenas ou milhares de camadas sem perda significativa de informação ou estabilidade no treinamento.

As DRNs também incorporam técnicas complementares, como normalização de batch e funções de ativação não lineares avançadas (ReLU, Leaky ReLU, GELU), que otimizam a convergência e melhoram a expressividade do modelo (HE et al., 2016). O resultado é uma arquitetura capaz de capturar representações hierárquicas altamente complexas, aumentando significativamente o poder preditivo em tarefas de classificação, detecção de padrões e reconhecimento de sinais, consolidando seu protagonismo em aplicações científicas, biomédicas, químicas e industriais, nas quais a modelagem de dados de alta dimensionalidade e complexidade estrutural é crítica.

Outras variantes modernas das DRNs são as ResNeXt, DenseNet e Wide ResNet, que incluem diferenças conceituais e vantagens em relação às DRNs originais, mas que não foram implementadas nesta Tese pois as florestas aleatórias foram soluções mais rápidas e satisfatórias.

Na prática, a qualidade de um modelo de ML está diretamente associada à sua capacidade de capturar os padrões subjacentes dos dados sem incorrer em desvios sistemáticos que comprometam sua capacidade preditiva (Hastie et al., 2009; YANG et al., 2019). Essa capacidade é frequentemente descrita em termos do grau de ajuste entre as predições do modelo e os dados observados. O desafio reside em alcançar um equilíbrio adequado pois o modelo não deve ser excessivamente simples a ponto de negligenciar padrões relevantes (subajuste), nem excessivamente complexo a ponto de se ajustar a ruídos e particularidades não generalizáveis dos dados de treinamento (sobreajuste). Uma representação genérica desses casos é mostrada na Figura 6.

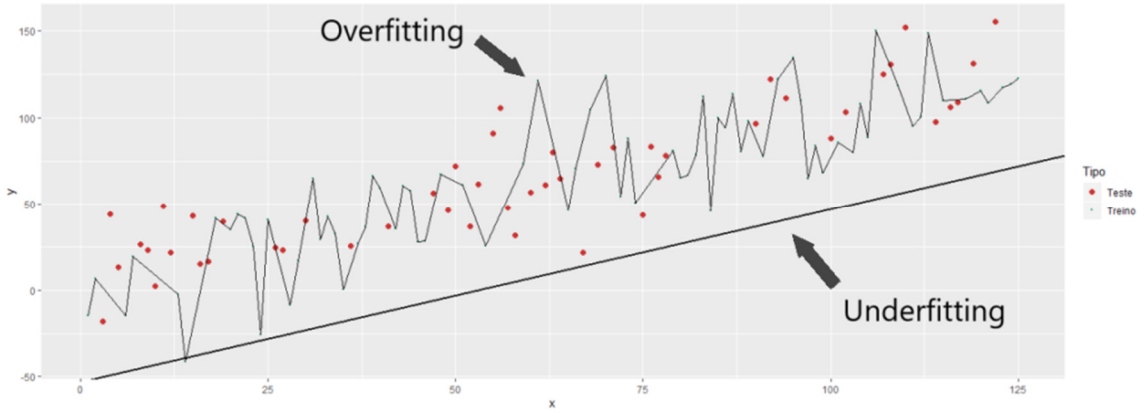


Figura 6: Representação do sobreajuste e subajuste em um modelo genérico (DIDÁTICA TECH, 2024).

O subajuste ocorre quando o modelo falha em capturar relações fundamentais presentes nos dados, resultando em baixo desempenho tanto no conjunto de treinamento quanto no de teste. Esse cenário é caracterizado pela alta polarização (viés) do modelo, que tende a produzir predições sistematicamente distantes dos valores reais. Modelos excessivamente lineares aplicados a problemas intrinsecamente não lineares são exemplos típicos de situações que conduzem ao subajuste. Além disso, conjuntos de atributos pouco representativos ou inadequados também podem agravar esse problema, limitando a expressividade do modelo (LECUN et al., 2015; MITCHELL, 1997; WELLMAN, 1995).

Em contraste, o sobreajuste emerge quando o modelo se ajusta de maneira demasiadamente precisa em relação aos dados de treinamento, incluindo variações aleatórias e ruídos experimentais. Nessa situação, observa-se um baixo erro no conjunto de treinamento, mas um aumento significativo do erro no conjunto de teste, indicando falha de generalização. O fenômeno está associado a modelos de alta complexidade (como redes neurais com número excessivo de camadas ou árvores de decisão muito profundas), que apresentam baixa polarização, mas alta variância. Em outras palavras, o modelo memoriza os exemplos disponíveis, mas perde a capacidade de inferir corretamente sobre instâncias não vistas (LeCun et al., 2015; Mitchell, 1997; Wellman, 1995).

Diversas estratégias têm sido desenvolvidas para mitigar esses problemas. O subajuste pode ser contornado com o uso de modelos mais expressivos, melhor engenharia de atributos ou aumento da qualidade e quantidade de dados. Já o sobreajuste pode ser reduzido por técnicas de validação cruzada e validação cruzada aninhada para monitorar o desempenho em diferentes partições dos dados, parada antecipada (do termo em inglês *early stopping*) em redes neurais e métodos de *ensemble*, como o Random Forest, que reduzem a variância ao combinar múltiplos modelos para se obter uma previsão final mais robusta e precisa.

Assim, compreender e mitigar os efeitos de subajuste e sobreajuste é crucial para o desenvolvimento de modelos mais robustos, capazes de equilibrar adequadamente viés e variância, garantindo não apenas bom desempenho nos dados de treinamento, mas principalmente capacidade de generalização para novos conjuntos de dados, o que representa a verdadeira finalidade do Aprendizado de Máquina.

3.3 LIME

Diferentes abordagens têm sido propostas para fornecer explicações acerca do funcionamento de modelos ML diferenciando-se, sobretudo, em relação ao objetivo e ao nível de interpretação oferecido pela técnica(FELDMANN; BAJORATH, 2022; JOHNSEN et al., 2021; LEDEL; HERBOLD, 2022; MASTROPIETRO et al., 2022; RODRÍGUEZ-PÉREZ; BAJORATH, 2020; WOJTUCH et al., 2021). De maneira geral, essas explicações podem ser classificadas em globais ou locais. As explicações globais buscam caracterizar o comportamento do modelo como um todo, identificando padrões gerais e destacando os atributos que exercem maior influência no processo de decisão. Em contraste, as explicações locais concentram-se em instâncias específicas, evidenciando quais fatores foram determinantes para que o modelo chegasse a uma determinada predição individual.

Além dessa distinção, tais métodos podem ser categorizados como específicos ao modelo, quando projetados para explicar exclusivamente a arquitetura para a qual foram desenvolvidos, ou como agnósticos ao modelo, que capazes de serem aplicados de forma independente do algoritmo de aprendizagem subjacente. Outra dimensão de classificação amplamente utilizada é a *Post-hoc*, que corresponde a técnicas aplicadas após o treinamento do modelo, tendo como finalidade a geração de explicações interpretáveis sem interferir diretamente no processo de aprendizado.

Entre os algoritmos explicativos mais difundidos e de implementação relativamente simples destaca-se o LIME (RIBEIRO et al., 2016), inicialmente aplicado em tarefas de reconhecimento de imagens e classificação de textos. O LIME constitui uma técnica agnóstica ao modelo, que atua de forma local e *Post-hoc*, permitindo identificar e interpretar as variáveis que apresentam maior correlação estatística com o resultado produzido pelo modelo. Assim, em vez de restringir-se à mera previsão de saídas, essa abordagem viabiliza a construção de representações explicativas capazes de revelar as relações mais relevantes entre as variáveis preditivas e as respostas. Tal característica confere ao LIME um papel crucial em cenários nos quais se deseja compreender os motivos

subjacentes a uma decisão específica do modelo ou avaliar como determinadas variáveis interagem entre si. Na prática, esse aspecto interpretativo assume importância equivalente ao próprio desempenho preditivo do modelo, uma vez que a ausência de explicabilidade compromete diretamente a confiabilidade e a aceitabilidade de sua aplicação.

> "LIME é a abreviação para Local Interpretable Model-Agnostic Explanations. Cada parte do nome reflete em algo que desejamos em explicações. Local se refere à fidelidade local - i.e., queremos que a explicação realmente reflita o comportamento do classificador "em volta" da instância sendo prevista. Essa explicação é inútil a não ser que seja interpretável - ou seja, a não ser que um ser humano consiga compreendê-la. LIME é capaz de explicar qualquer modelo sem precisar "espiar" dentro dele, então é agnóstica ao modelo." (RIBEIRO et al., 2016)

LIME explica a predição de qualquer classificador aprendendo um modelo interpretável e fornecendo um explicador agnóstico de modelo consistente. O método seleciona um conjunto representativo com explicações que fornecem uma compreensão intuitiva do modelo. Ele ajusta ligeiramente a entrada e testa as mudanças na previsão. Esse ajuste deve ser pequeno para que ainda esteja próximo da região local original. LIME tem as seguintes propriedades: (1) fornece uma compreensão fácil e qualitativa da resposta e das variáveis do modelo; (2) deve ser pelo menos localmente confiável, replicando o comportamento do modelo com fidelidade local; (3) deve explicar qualquer modelo sem fazer suposições sobre o modelo e sendo agnóstico em relação ao modelo; e (4) deve explicar um conjunto representativo, fornecendo uma intuição abrangente do modelo (BARR et al., 2020; FELDMANN; BAJORATH, 2022).

Sob a suposição de que um explicador seja confiável e interpretável, o LIME minimiza a seguinte equação do modelo de explicação:

$$\psi(x) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

onde *f* é o preditor inicial, *x* representa as características iniciais, *g* é o modelo explicativo e π_x é a medida de proximidade entre um exemplo de *z* e *x* que é definido localmente em torno de *x*. O primeiro termo é chamado de perda com reconhecimento de localidade, que é a medida da infidelidade de *g* aproximando *f* no local definido por π_x. O segundo termo é a medida da complexidade do modelo explicativo. A perda com reconhecimento de localidade é minimizada para garantir a fidelidade local e a interpretabilidade, mantendo a medida da complexidade baixa o suficiente para ser interpretável por humanos. Quando a perda com reconhecimento de localidade é otimizada, o LIME atinge a fidelidade local (Ribeiro et al., 2016b).

A amostragem uniforme aleatória para exploração local é realizada a partir de *x* para criar um conjunto de dados de treinamento completo, ou seja, criar múltiplos *z* a partir de uma única linha de *x*, que é ponderada por π_x para ser focada em dados *z* mais próximos de *x*. A explicação linear esparsa assume que: (1) $g(z) = w_z$; (2) a perda com reconhecimento de localidade é uma perda quadrada; e (3) a ponderação de proximidade para as amostras é:

$$\pi_x = \frac{e^{-D(x,z)^2}}{\sigma^2}$$

onde $D(x,z)$ é uma função de distância. Portanto, a perda quadrática com base na localidade é apresentada da seguinte forma:

$$\mathcal{L}(f, g, \pi_x) = \sum \pi_x(z)(f(z) - g(z))^2$$

O cerne da abordagem reside no processo de amostragem de dados sintéticos perturbados a partir de uma instância de interesse. Essas amostras são geradas com base em distribuições Gaussianas, de modo a criar variações artificiais das características originais do exemplo em análise. Diferentemente de métodos que utilizam diretamente medidas de distância Gaussianas, o LIME realiza a perturbação dos atributos e, a partir disso, produz exemplos sintéticos que permitem aproximar o comportamento do modelo na vizinhança local da instância selecionada.

A noção de “localidade” em LIME refere-se à capacidade de capturar os limites decisórios e a relevância das variáveis em torno de uma instância específica, em vez de generalizar para todo o espaço de dados. Para tanto, emprega-se uma métrica de similaridade que possibilita mensurar a proximidade entre a instância original e os exemplos perturbados. A escolha da métrica é dependente do tipo de dado e da implementação considerada.

O processo típico de aplicação do LIME envolve: (i) a seleção de uma instância-alvo cuja predição do modelo se deseja interpretar; (ii) a geração de aproximadamente 5000 amostras sintéticas por meio da perturbação de até 10 atributos de entrada, de acordo com a distribuição de treinamento; (iii) o cálculo das distâncias entre a instância original e as amostras perturbadas; (iv) a atribuição de pesos a cada amostra, de forma que exemplos mais próximos recebam maior relevância; (v) a construção de um modelo substituto (do termo em inglês *surrogate model*) simples, geralmente linear, treinado a partir das amostras perturbadas e das respectivas previsões do modelo complexo; e, por fim, (vi) a interpretação do modelo substituto, que evidencia a importância relativa de cada característica para a tomada de decisão do modelo original naquela instância específica (VISANI et al., 2020).

Portanto, o LIME não apenas fornece uma ferramenta de inspeção local do comportamento de modelos complexos, mas também constitui um elo essencial entre a opacidade algorítmica e a interpretabilidade exigida em contextos de pesquisa acadêmica e de aplicação prática. Sua relevância se amplia em virtude da crescente demanda por transparência em sistemas de

inteligência artificial, sobretudo em domínios críticos além de química medicinal, como saúde, finanças e sistemas de apoio à decisão em larga escala.

O LIME é versátil e aplicável a qualquer modelo de ML. Caracteriza-se por ser independente, fácil de compreender e didático. O LIME visa elucidar um conjunto representativo de previsões do modelo e garantir a confiabilidade local replicando o comportamento do modelo com fidelidade em uma escala menor. A principal característica do LIME é sua capacidade de construir um modelo linear na proximidade de uma instância de teste. Isso envolve inicialmente a geração de amostras artificiais permutando tais instâncias com pesos determinados por um valor de largura de núcleo (do termo em inglês *kernel width*). Consequentemente, o modelo é treinado para fazer previsões com base nos coeficientes de importância das características interpretadas do modelo (BARR et al., 2020; RIBEIRO et al., 2016; ZAFAR; KHAN, 2021).

O LIME pode aproximar qualquer modelo de ML de caixa preta a um modelo interpretável local para explicar cada previsão individual, além de ser um explicador extremamente popular. A busca por esse método mais rápido e com menor custo computacional pode ser relevante para revelar novidades à área de toxicologia, ciência de dados e quimioinformática (RIBEIRO et al., 2016).

Na utilização do LIME, a etapa de hiperparametrização assume papel central para assegurar que as explicações produzidas representem, de forma adequada, a relação entre a instância a ser interpretada e o modelo de predição subjacente. Dentre os parâmetros mais sensíveis, destaca-se a largura do núcleo, responsável por regular a distribuição de pesos atribuída às amostras perturbadas em função de sua proximidade da instância de interesse. Em termos práticos, esse parâmetro controla o escopo de localidade da explicação: valores reduzidos induzem o LIME a considerar apenas vizinhanças muito próximas, privilegiando a captura de relações altamente locais, mas potencialmente suscetíveis ao ruído. Por outro lado, valores amplos de largura de núcleo favorecem a inclusão de pontos mais distantes, o que pode aproximar a explicação de uma interpretação global (WELLAWATTE et al., 2022).

O ajuste da largura do núcleo deve equilibrar dois aspectos fundamentais: a fidelidade local, isto é, a capacidade do modelo linear substituto reproduzir o comportamento da predição na vizinhança imediata da instância, e a estabilidade das explicações, evitando que pequenas variações na geração de amostras levem a interpretações divergentes. Um núcleo excessivamente estreito pode amplificar flutuações aleatórias, resultando em explicações instáveis ou pouco generalizáveis. Por outro lado, um núcleo muito amplo compromete o objetivo central do LIME, que é justamente capturar relações locais, correndo o risco de mascarar dependências relevantes.

Além da largura do núcleo, outros parâmetros desempenham papel complementar no processo de hiperparametrização. A quantidade de amostras sintéticas geradas para a construção da vizinhança influencia diretamente a robustez estatística da explicação, uma vez que amostras insuficientes podem levar a coeficientes instáveis na regressão linear ponderada. O tipo de distância adotado na definição da proximidade (comumente a distância euclidiana) também impacta na distribuição dos pesos e pode demandar ajustes de acordo com a natureza dos atributos (contínuos, categóricos ou mistos).

Em síntese, a hiperparametrização do LIME constitui um processo não trivial que envolve decisões metodológicas com impacto direto na validade científica das explicações. Uma calibração criteriosa, baseada tanto em métricas de fidelidade quanto em análises comparativas de robustez, é indispensável para que o LIME cumpra sua função de aproximar o comportamento opaco de modelos complexos de uma representação inteligível, sem sacrificar a confiabilidade das inferências produzidas.

3.4 Desenvolvimento de Fármacos, Toxicidade de moléculas candidatas e Permeabilidade de compostos orgânicos na Barreira Hematoencefálica

O desenvolvimento de fármacos é uma das áreas mais estratégicas e desafiadoras da pesquisa contemporânea em saúde no Brasil. Esse processo envolve não apenas a busca por novas moléculas com potencial terapêutico,

mas também a integração de conhecimentos multidisciplinares que abrangem química medicinal, biologia molecular, farmacologia, bioinformática, computação científica e regulamentação sanitária (ALJOF; GAIPOV, 2024; WOUTERS et al., 2020). Estima-se que o tempo médio necessário para que uma molécula passe da etapa de descoberta até a sua disponibilização ao mercado varie entre 10 e 15 anos, com investimentos que frequentemente ultrapassam bilhões de dólares (DIMASI et al., 2016; SIMOENS; HUYS, 2021). Essa complexidade reflete não apenas o rigor exigido pelas agências regulatórias, mas também a alta taxa de insucesso em cada etapa do fluxo de trabalho.

O desenvolvimento pré-clínico de um fármaco, ou seja, as etapas que antecedem os ensaios clínicos em humanos, é um processo complexo e altamente regulamentado, que envolve descoberta, otimização, caracterização farmacológica e estudos de toxicologia. O período médio necessário para concluir essas etapas geralmente varia entre 3 e 6 anos, dependendo de fatores como a complexidade do alvo biológico, o tipo de molécula (peptídica, pequena ou biológica), a disponibilidade de modelos experimentais adequados e os requisitos regulatórios do país ou região. Durante essa fase, são realizadas atividades críticas como:

1) Síntese e otimização de compostos candidatos para garantir afinidade, seletividade e estabilidade.

2) Estudos *in vitro* e *in silico* para avaliar propriedades farmacocinéticas e farmacodinâmicas iniciais, incluindo permeabilidade, metabolismo e interação com alvos.

3) Testes de toxicidade aguda, subcrônica e genotóxica em modelos celulares e animais, necessários para identificar riscos potenciais à saúde humana.

4) Estudos de formulação e biodisponibilidade, a fim de definir a via de administração e a dosagem apropriada para futuros ensaios clínicos.

Tradicionalmente, o processo inicia-se com a identificação de um alvo biológico associado a uma condição patológica. Esse alvo, frequentemente uma

proteína, enzima ou receptor, é selecionado a partir de estudos em biologia molecular, genética e fisiopatologia. A partir daí, busca-se identificar moléculas com capacidade de modular a atividade desse alvo, um processo conhecido como *drug discovery*. A técnica de High-Throughput Screening (HTS) foi desenvolvida para permitir a triagem rápida e sistemática de grandes bibliotecas de compostos químicos em modelos *in vitro*, visando identificar candidatos promissores para desenvolvimento de fármacos (Chen et al., 2022; Hernandez et al., 2024; Roy et al., 2024). O número de compostos que pode ser testado depende da capacidade do equipamento automatizado, do formato de ensaio (por exemplo, 96, 384 ou 1.536 poços) e da complexidade do modelo celular ou bioquímico utilizado. Em termos médios, plataformas HTS modernas permitem testar de dezenas de milhares até centenas de milhares de compostos por semana. Bibliotecas de compostos comerciais ou institucionais frequentemente contêm entre 100.000 e 500.000 moléculas, enquanto iniciativas de triagem ultrarrápida podem ultrapassar 1 milhão de compostos em estudos escalonados. Entretanto, os métodos de planejamento racional de fármacos, como *structure-based drug design* (SBDD) e *ligand-based drug design* (LBDD), possibilitam a criação de moléculas otimizadas com base em informações estruturais e funcionais do alvo (Al-Karmalawy et al., 2025; El Rhabori et al., 2025; Vasilev and Atanasova, 2025).

Uma vez os candidatos iniciais são identificados, estes passam por ciclos de Design-Make-Test-Analyse (DMTA) sucessivos de otimização química, nos quais propriedades como afinidade, seletividade, estabilidade metabólica e biodisponibilidade são ajustadas (NIPPA, 2025; QIN, 2025). A otimização envolve modificações estruturais orientadas por estudos de relações estrutura-atividade (SAR) e modelagem molecular, visando maximizar a eficácia terapêutica e reduzir efeitos adversos potenciais.

Na sequência, os compostos mais promissores são submetidos a ensaios pré-clínicos, que englobam experimentação em sistemas biológicos *in vitro* e *in vivo*. Esses estudos avaliam segurança, farmacocinética e farmacodinâmica, estabelecendo parâmetros fundamentais para a transição para a fase clínica. É importante destacar que essa etapa apresenta elevada taxa de insucesso, uma

vez que moléculas eficazes em modelos experimentais muitas vezes não replicam tais resultados em organismos mais complexos.

O avanço para os ensaios clínicos em humanos representa um divisor de águas no desenvolvimento farmacêutico. Essa etapa é tradicionalmente dividida em quatro fases: (i) avaliação inicial de segurança em voluntários saudáveis (Fase I); (ii) estudo da eficácia em pequena escala em pacientes (Fase II); (iii) ampliação do número de pacientes e comparação com terapias existentes, com vistas a confirmar eficácia e monitorar segurança (Fase III); e (iv) acompanhamento pós-comercialização, no qual se monitoram efeitos adversos raros e o desempenho do fármaco em larga escala populacional (LIMA et al., 2023). O fármaco pode ser aprovado por agência regulatória, como a FDA (*Federal Drug Administration*), apenas após a conclusão bem-sucedida da fase III.

O caráter moroso e oneroso do desenvolvimento de fármacos reflete a necessidade de equilibrar inovação científica, viabilidade econômica e segurança pública. Estima-se que menos de 10% das moléculas que chegam à fase clínica obtêm aprovação final, revelando o caráter altamente seletivo do processo. Esse índice de sucesso limitado evidencia a importância de novas metodologias capazes de acelerar a descoberta, reduzir custos e mitigar a taxa de falhas.

Nos últimos anos, a incorporação de tecnologias computacionais e de IA tem promovido uma mudança de paradigma na forma como novos fármacos são concebidos com o uso de métodos alternativos (U.S FDA, 2025). Modelos de ML e XAI têm sido aplicados à análise de grandes bases de dados moleculares e clínicas, viabilizando predições mais precisas sobre propriedades farmacológicas e auxiliando no planejamento racional de moléculas (ROSA, et al., 2025; ROSA, et al., 2024). Essas inovações não substituem as etapas tradicionais, mas oferecem um complemento estratégico que pode otimizar a triagem, reduzir a dependência exclusiva de métodos empíricos e aumentar a eficiência global do pipeline de desenvolvimento.

Em síntese, o desenvolvimento de fármacos deve ser entendido não apenas como um processo técnico, mas como uma atividade de natureza interdisciplinar, que articula ciência básica, tecnologia de ponta e regulação sanitária em prol da inovação terapêutica. A crescente utilização de abordagens computacionais, em especial a inteligência artificial, representa uma das tendências mais promissoras para superar os gargalos históricos desse processo, consolidando-se como ferramenta indispensável na busca por terapias mais seguras, eficazes e acessíveis à população.

A toxicidade constitui um dos principais desafios no desenvolvimento de novos fármacos, representando uma barreira crítica tanto para o sucesso clínico quanto para a aprovação regulatória (Lysenko et al., 2018; Mahmoud, 2021.). A toxicidade pode ser entendida como a capacidade de uma molécula induzir efeitos adversos em sistemas biológicos, não sendo um fenômeno absoluto, mas sim um espectro dependente de fatores intrínsecos ao composto (como a estrutura química, propriedades físico-químicas e metabolismo) e extrínsecos ao organismo (como variações genéticas, estado fisiológico, microbiota e interações medicamentosas) (HUANG et al., 2024; WANG et al., 2025).

Do ponto de vista conceitual, a toxicidade pode ser classificada em diferentes dimensões: aguda (manifestada após exposição única em altas doses), subcrônica ou crônica (decorrente de exposições repetidas a longo prazo), reversível ou irreversível, além de sistêmica (afetando múltiplos órgãos e tecidos) ou específica (restrita a determinado órgão, como hepatotoxicidade ou cardiotoxicidade). Ainda, a toxicidade também pode estar associada a reações peculiares, nas quais fatores genéticos ou imunológicos modulam a resposta adversa de forma não previsível (DALY, 2013; JAIN et al., 2018; MORGER et al., 2020).

A mensuração da toxicidade envolve um conjunto amplo de metodologias. Os estudos *in vitro* fornecem informações iniciais sobre citotoxicidade, genotoxicidade, mutagenicidade e potenciais efeitos disruptores endócrinos, utilizando linhagens celulares, organoides e sistemas micro fisiológicos. Em paralelo, estudos *in vivo* em modelos animais permitem caracterizar parâmetros farmacocinéticos e toxicológicos, estimando, por exemplo, os valores de DL_{50}

(dose letal 50%). Recentemente, técnicas baseadas em modelagem preditiva por aprendizado de máquina vêm ganhando destaque, permitindo antecipar alertas estruturais de toxicidade e reduzir a dependência de ensaios em animais (NASCIMENTO et al., 2023).

Os mecanismos de toxicidade estão frequentemente relacionados à interação de fármacos com alvos biológicos não intencionais (*off-targets*). Moléculas que exibem alta afinidade ou baixa seletividade podem interagir com proteínas críticas, como canais iônicos, transportadores e receptores nucleares, desencadeando efeitos colaterais significativos. A toxicidade constitui uma das principais causas de falha em fases clínicas avançadas, evidenciando a necessidade de abordagens integradas de avaliação precoce. Nesse sentido, a combinação de técnicas experimentais, modelagem *in silico* e preditores baseados em IA tem se mostrado promissora para identificar subestruturas tóxicas, avaliar a relação estrutura-atividade (SAR) e apoiar a tomada de decisão durante as etapas iniciais de desenvolvimento de fármacos (CRONIN et al., 2017; VON KORFF; SANDER, 2006; LYSENKO et al., 2018; RAIES; BAJIC, 2016; RICHARD et al., 2016; ROSA et al., 2025).

O estudo da toxicidade não apenas conduz à segurança e eficácia terapêutica, mas também contribui para a racionalização do processo de descoberta de fármacos. A toxicologia moderna caminha em direção a uma abordagem preditiva e integrativa, onde a compreensão detalhada das interações moleculares, da biotransformação metabólica e das respostas adaptativas do organismo é fundamental para reduzir riscos e otimizar a inovação em química medicinal. Quanto ao futuro da química medicinal, observa-se uma tendência crescente de integrar modelagem computacional, XAI e dados biológicos de alta resolução. Essa convergência deve acelerar a identificação de novos alvos terapêuticos, permitir projetar moléculas com maior precisão e favorecer ciclos interativos mais rápidos entre predição, síntese e validação experimental, impulsionando abordagens cada vez mais personalizadas e eficientes no desenvolvimento de fármacos.

Atualmente, um dos maiores desafios terapêuticos envolvendo o sistema nervoso central (SNC) está relacionado à escassez de fármacos capazes de

atravessar a Barreira Hematoencefálica (BHE) (BALLABH et al., 2004; DI et al., 2013; GUPTA et al., 2019; LI et al., 2005; LIU et al., 2021; WU et al., 2023; ZHAO et al., 2007). A BHE é uma estrutura altamente seletiva que protege o SNC contra substâncias neurotóxicas, microrganismos e outros solutos, preservando um ambiente homeostático adequado para a função neuronal. Entretanto, essa proteção também impõe restrições ao ingresso de moléculas potencialmente terapêuticas, o que compromete a eficácia no tratamento de doenças neurológicas, como meningite (BEAM; ALLEN, 1977), esclerose múltipla (SCHREIBELT et al., 2006) e doença de Alzheimer (ZIPSER et al., 2007).

A BHE recobre os vasos sanguíneos cerebrais e é formada por células endoteliais, pericitos e astrócitos. As células endoteliais constituem os capilares cerebrais e são interligadas por junções oclusivas, que restringem a passagem celular. Já os astrócitos são células gliais de formato estrelado e os pericitos interagem diretamente com o endotélio, regulando o tônus vascular e a permeabilidade capilar. A figura 7 mostra a representação didática da BHE. Essa organização celular garante a seletividade da barreira, modulando tanto o fluxo sanguíneo cerebral quanto a passagem de solutos. A permeabilidade da BHE ocorre por diferentes mecanismos: transporte mediado por carreadores (nutrientes essenciais, como glicose e aminoácidos), transcitose mediada por vesículas, transporte ativo de efluxo por proteínas (ex.: P-glicoproteína) e difusão passiva de pequenas moléculas lipofílicas (Ballabh et al., 2004).

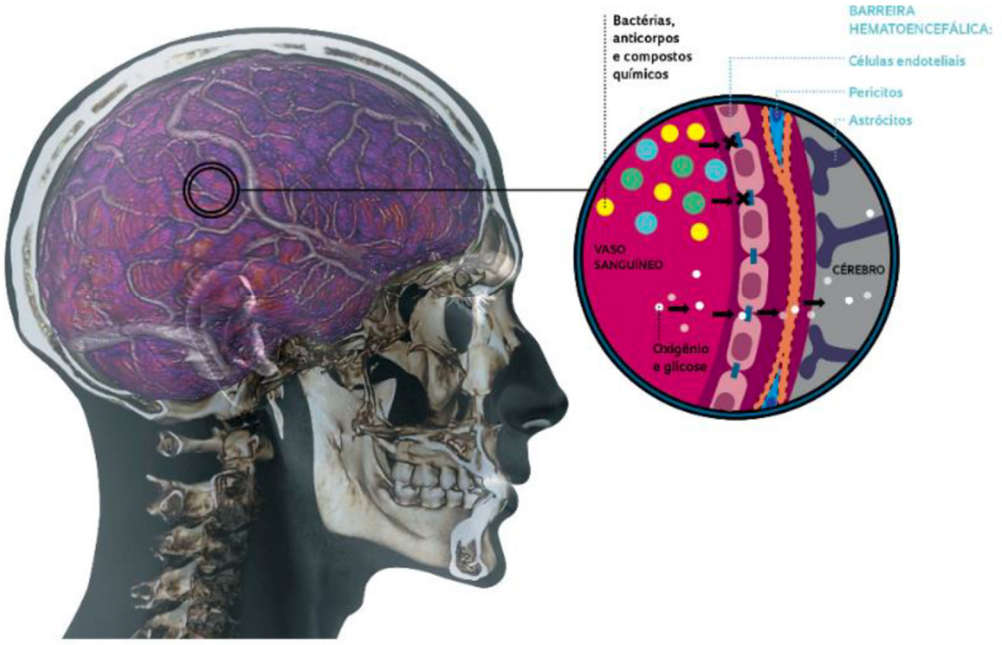


Figura 7: representação da BHE (FAPESP, 2022) - Adaptado

O transporte mediado por carreadores é responsável pela entrada de nutrientes essenciais no SNC, sendo altamente seletivo. Transportadores de glicose (GLUT1) e de aminoácidos (LAT1) são alguns dos mais relevantes, permitindo a entrada de substratos fundamentais para o metabolismo cerebral. Do ponto de vista farmacológico, esse caminho pode ser explorado por meio do princípio do mimetismo estrutural, no qual determinados fármacos ou pró-fármacos são planejados para se parecer com nutrientes endógenos e, assim, utilizar os mesmos transportadores (Patching, 2017; Puris et al., 2017). Um exemplo clássico é a L-DOPA, usada no tratamento da doença de Parkinson, que atravessa a BHE por meio de transportadores de aminoácidos neutros (HAWKINS et al., 2005).

No processo de transcitose mediada por vesículas, macromoléculas como insulina e transferrina são reconhecidas por receptores específicos presentes nas células endoteliais da BHE, sendo internalizadas em vesículas e movidas para o SNC. Do ponto de vista terapêutico, essa via tem sido explorada em estratégias inovadoras de entrega de anticorpos monoclonais e nanopartículas

funcionalizadas, que podem se ligar a receptores endoteliais e, assim, ultrapassar a barreira de forma controlada. Isso abre uma perspectiva promissora para o tratamento de doenças neurodegenerativas e tumores cerebrais, embora ainda haja limitações relacionadas à eficiência e à seletividade.

Um dos grandes obstáculos ao desenvolvimento de fármacos para o SNC é a presença de proteínas de efluxo, como a P-glicoproteína (P-gp)**.** Essas proteínas atuam como uma defesa do SNC contra xenobióticos, bombeando ativamente muitos fármacos de volta para a circulação sanguínea. Assim, mesmo moléculas lipofílicas que, em princípio, poderiam atravessar por difusão passiva podem ser rapidamente removidas. Do ponto de vista farmacológico, compreender e prever a interação de novas moléculas com esses transportadores é crucial para o sucesso no desenvolvimento de fármacos direcionados ao cérebro.

Dentre esses mecanismos, a difusão passiva é um dos principais responsáveis pela entrada de pequenas moléculas no SNC. Nesse processo, o gradiente de concentração entre sangue e cérebro impulsiona a passagem através da bicamada lipídica das células endoteliais, sem gasto de energia. Do ponto de vista físico-químico, a difusão pode ser descrita pela equação de Fick:

$$J = -D\frac{dC}{dx}$$

em que o fluxo (J) depende do coeficiente de difusão (D) e do gradiente de concentração (dC/dx). Complementarmente, a eficiência da permeação é descrita pelo coeficiente de permeabilidade:

$$J = -D\frac{P}{h}$$

que incorpora tanto a lipofilicidade (P, coeficiente de partição óleo/água) quanto a espessura da membrana (h). Assim, compostos com balanço adequado entre lipofilicidade e solubilidade aquosa, além de baixo peso molecular, apresentam maior probabilidade de atravessar a barreira por difusão passiva.

Diversas propriedades estruturais e físico-químicas modulam a difusão passiva através da BHE (GOODWIN AND CLARK, 2005; LIPINSKI, 2004). A lipofilicidade, usualmente quantificada pelo LogP, deve apresentar valores moderados: valores muito baixos indicam polaridade excessiva, enquanto valores muito altos aumentam a retenção na membrana. A polaridade também é avaliada pelo número de doadores e aceptores de ligações de hidrogênio e pela Área de Superfície Polar Topológica (TPSA). Em geral, moléculas com TPSA ≤ 90 Å² têm maior chance de atravessar a BHE, ao passo que valores superiores a 120 Å² reduzem significativamente a permeabilidade. Outro fator determinante é o grau de ionização: espécies carregadas tendem a apresentar baixa difusão passiva, sendo a forma neutra mais favorável ao processo. Nesse sentido, compostos com pKa que permitem frações neutras em pH fisiológico são candidatos mais promissores (TONG et al., 2022).

Avanços recentes em Aprendizado de Máquina têm permitido relacionar quantitativamente essas propriedades com a penetração de fármacos na BHE. Alguns padrões estruturais são associados à permeabilidade, incluindo subestruturas aromáticas e nitrogenadas. Modelos baseados em descritores físico-químicos clássicos (como TPSA e LogP) têm sido amplamente empregados, mas arquiteturas mais sofisticadas, como redes convolucionais de grafos e redes neurais artificiais, vêm ampliando a capacidade de predição da difusão passiva (MASTROPIETRO et al., 2022). Entre diferentes algoritmos, modelos de floresta aleatória (*Randon Forest)* apresentam desempenho competitivo, enquanto abordagens de Inteligência Artificial Explicável vêm sendo utilizadas para reduzir a opacidade das predições e tornar interpretáveis as relações entre estruturas químicas e permeabilidade (Martins et al., 2012).

As estratégias para superar as limitações impostas pela BHE não se restringem a métodos computacionais. Abordagens inovadoras, como o desenvolvimento de sistemas de micro e nano robôs inteligentes (MNR), vêm sendo investigadas como alternativas físicas para a entrega de fármacos ao SNC (MAIR et al., 2021). Essas tecnologias oferecem uma perspectiva translacional complementar: enquanto modelos de ML e XAI permitem prever e compreender a permeabilidade, soluções baseadas em MNR ampliam a possibilidade de

intervenção prática, estabelecendo uma ponte promissora entre predição estatística e aplicação clínica.

No desenvolvimento de fármacos, em especial nas etapas iniciais de triagem virtual e avaliação de propriedades ADMET (absorção, distribuição, metabolismo, excreção e toxicidade), a explicabilidade não é apenas desejável, mas necessária. A confiabilidade de um modelo preditivo influencia diretamente decisões estratégicas, como a priorização de compostos para síntese ou a exclusão de candidatos promissores devido a potenciais riscos toxicológicos. Dessa forma, técnicas de XAI podem atuar como uma ponte entre o raciocínio computacional e o conhecimento químico-farmacológico, permitindo que cientistas avaliem a acurácia de uma predição e a racionalidade química por trás dela (KAR et al., 2022; LEI et al., 2016; EL RHABORI et al., 2025).

Além disso, a integração de XAI no ciclo de descoberta de fármacos possibilita um ganho em rastreabilidade científica. Quando um modelo indica que determinada subestrutura química está associada ao aumento da permeabilidade na BHE, ou ainda sugere padrões relacionados a efeitos tóxicos, a interpretação transparente fornece subsídios para validar hipóteses, guiar novas sínteses e confrontar resultados experimentais. Assim, a explicabilidade reduz a distância entre predição estatística e compreensão mecanística, tornando os modelos mais alinhados com o processo de raciocínio científico empregado em química medicinal. Essa ideia refere-se ao fato de que modelos explicáveis não apenas fazem previsões numéricas, mas mostram por que as fazem, permitindo conectar padrões estatísticos identificados pelo algoritmo a mecanismos químicos plausíveis, o que aproxima a modelagem computacional da lógica de investigação na química medicinal.

Portanto, ao se aplicar métodos de ML ao desenvolvimento de novos fármacos, a incorporação de técnicas de XAI deixa de ser apenas um diferencial tecnológico para se tornar um requisito fundamental. Essa abordagem não apenas amplia a confiabilidade dos modelos, mas também favorece a adoção da inteligência artificial como parceira efetiva no processo de inovação farmacêutica, permitindo que os algoritmos não apenas “prevejam”, mas também

“expliquem” caminhos que conduzem a novas descobertas terapêuticas (CLEMENT et al., 2023).

Especialmente com o avanço da ciência de dados e do fenômeno de *big data*, a disponibilidade de informações sobre estruturas químicas aumentou consideravelmente. Filtrar estruturas que conduzem a propriedades físico-químicas indesejadas em bibliotecas virtuais pode reduzir o universo de milhões para alguns milhares de candidatos a fármacos (BRENK et al., 2008). Além disso, é altamente desejável ter filtros para grupos funcionais ou fragmentos comumente considerados inadequados para desenvolver cada vez mais triagens virtuais (BAELL; HOLLOWAY, 2010; BINIASHVILI et al., 2012; BOLOGA et al., 2019; OPREA, 2000; URSU et al., 2011). Servidores *web* e sistemas especialistas em alertas estruturais são amplamente desenvolvidos na literatura (BRUNER et al., 1996; DOWEYKO, 2008; GUHA, 2008; LEPAILLEUR et al., 2013; MARCHANT, 2012; RIDINGS et al., 1996; SUSHKO et al., 2012; YANG et al., 2018), mas modelos interpretáveis de ML e explicadores de modelos ainda não são amplamente investigados com a intenção de obter alertas estruturais e alertas tóxicos.

Embora sistemas especialistas em alertas estruturais possam destacar os perigos potenciais de subestruturas químicas, métodos automatizados com técnicas de aprendizado de máquina e redes neurais podem ser superiores em termos de desempenho preditivo (YANG et al., 2020). É importante ressaltar que, apesar de todo o progresso alcançado por essas ferramentas, a correlação entre a interpretação de dados toxicológicos e o mecanismo de ação do composto no corpo humano ainda é um desafio (DANG et al., 2017; LEPAILLEUR et al., 2013; SUSHKO et al., 2012; Yang et al., 2020, 2017).

4 Metodologia

4.1 Codificação e divisão dos dados toxicológicos

O código foi escrito utilizando Python (versão 3.7) na plataforma do *Google Colaboratory*. As seguintes bibliotecas foram importadas: DeepChem (SUGAWARA; NIKAIDO, 2014), Pandas (Reback et al., 2022), RDKit (GAVID, 2022), Matplotlib (HUNTER, 2007) e Numpy (HARRIS et al., 2020). Em seguida, o conjunto de dados Tox21 (RICHARD et al., 2021) curado foi utilizado para treinar o modelo de aprendizado de máquina de medições experimentais qualitativas com rótulo binário (0 para não tóxico e 1 para tóxico) da atividade de moléculas em doze alvos biológicos (NR-AR, NR-AR-LBD, NR-AhR, NR-Aromatase, NR-ER, NR-ER-LBD, NR-PPAR-γ, SR-ARE, SR-ATAD5, SR-HSE, SR-MMP e SR-p53 (cujos nomes são mostrados na lista de abreviações). Além disso, o conjunto de dados Tox21 foi comparado com os conjuntos de dados Sider (KUHN et al., 2016) e Clintox (TASNEEM et al., 2012) curados encontrados no MoleculeNet (WU et al., 2018). Esses conjuntos de dados incluem moléculas de fármacos com as representações SMILES (WEININGER, 1988) correspondentes.

O Tox21 é uma colaboração entre agências federais dos EUA e visa desenvolver métodos de triagem de alta capacidade para avaliar rapidamente a segurança de produtos químicos comerciais, pesticidas, aditivos alimentares, contaminantes alimentares e produtos médicos. É utilizado para prever a toxicidade de compostos químicos com base em dados de triagem *in vitro*, facilitando a identificação precoce de substâncias potencialmente perigosas.

O Sider (*Side Effect Resource*) é um banco de dados que compila informações sobre fármacos comercializados e seus efeitos adversos registrados. Os dados são extraídos de documentos públicos e bulas de fármacos, incluindo a frequência dos efeitos adversos e classificações de fármacos e efeitos colaterais. É utilizado para estudar a relação entre estruturas

químicas de fármacos e seus efeitos adversos, auxiliando no desenvolvimento de fármacos mais seguros e na compreensão dos mecanismos de toxicidade.

Já o Clintox é um banco de dados amplamente utilizado em estudos de toxicologia computacional, especialmente para a previsão de toxicidade clínica de fármacos. Ele faz parte do conjunto MoleculeNet e é frequentemente empregado para treinar e testar modelos de Aprendizado de Máquina voltados à segurança de compostos químicos. Esse banco de dados contém informações sobre compostos, incluindo os fármacos aprovados pela FDA e os fármacos que falharam em ensaios clínicos devido a problemas de toxicidade.

O conjunto de dados Tox21 foi caracterizado utilizando as impressões digitais de conectividade estendida (ECFP) (ROGERS; HAHN, 2010) sem dividir o conjunto de dados. Para obter um modelo o mais geral possível, foi necessário remover previamente do Tox21 as moléculas em comum com os outros bancos de dados (Clintox e Sider). Dessa forma, as moléculas no conjunto de dados de teste não foram encontradas nos conjuntos de dados de treinamento e validação. As moléculas em comum foram removidas utilizando o coeficiente de Tanimoto (RÁCZ et al., 2018) como uma medida de similaridade e a impressão digital de Morgan (CAPECCHI et al., 2020) como um vetor de bits. Para isso, foi necessário transformar o conjunto de dados Tox21 em um quadro de dados Pandas.

No início do método, impressões digitais de Morgan foram geradas como um vetor de bits para todas as moléculas dos quadros de dados usando a classe *rdFingerprintGenerator* para criar uma coluna. Como este comando funciona apenas quando há uma estrutura desenhada na coluna ROMol, foi necessário remover previamente as linhas de todos os SMILES que não puderam ser transformados em estruturas ROMol (Weininger, 1988).

A função droppingIdenticals foi criada para calcular a similaridade de Tanimoto de uma determinada representação de SMILES para comparar com todas as moléculas no conjunto de dados Tox21. Todas as moléculas que apresentaram similaridade igual a 1,0 (ou seja, moléculas idênticas) foram eliminadas do quadro de dados Tox21. A função *dropsIdenticals* retornou dois

quadros de dados que continham todas as moléculas em comum entre o conjunto de dados Tox21 e os conjuntos de dados Clintox e Sider.

O conjunto de dados Tox21 foi dividido aleatoriamente em conjuntos de dados de treinamento e validação para treinar e validar o modelo utilizando a técnica de *hold-out* simples, separando 80% dos dados para treinamento e 20% para validação. O modelo de treinamento foi gerado aleatoriamente para cada execução. Como os resultados dos testes são um pouco diferentes entre si, mesmo que o conjunto de dados de teste seja o mesmo para cada execução, decidiu-se não salvar o modelo de treinamento para avaliar os resultados de cada execução e verificar a consistência dos resultados. O conjunto de dados de teste foi criado com os compostos de Clintox encontrados no conjunto de dados Tox21. As únicas moléculas utilizadas no modelo foram as sobrepostas, ou seja, aquelas que os conjuntos de dados Tox21 tinha em comum com o Sider e com o Clintox. As moléculas restantes não foram utilizadas porque as informações da tarefa para cada molécula são necessárias.

Utilizando a função classificadora multitarefa da biblioteca DeepChem (SUGAWARA; NIKAIDO, 2014), 12 tarefas com 1024 características foram classificadas para criar um modelo de classificação multitarefa. O modelo MultiTaskClassifier é uma rede residual profunda totalmente conectada para classificação multitarefa, composta por blocos residuais de pré-ativação (KAIMING HE et al., 2016). A otimização dos hiperparâmetros foi feita utilizando diferentes impressões digitais e números de épocas.

Para garantir que não houvesse desvios indevidos, o método estatístico de análise da curva ROC (*Receiver Operating Characteristics*) foi utilizado para calcular a pontuação de precisão entre o modelo e o teste. A curva ROC mostra o quão bem o modelo conseguiu distinguir entre dois rótulos binários (0 para não tóxico ou 1 para tóxico), levando em consideração a taxa de verdadeiros positivos (sensibilidade) e falsos positivos (especificidade). Para simplificar, uma curva ROC pode ser avaliada pela métrica AUC (*Area Under the Curve*, ou “área sob a curva”), que calcula a área da forma bidimensional formada abaixo da curva, sendo uma forma de resumir a curva ROC em um único valor (BRADLEY, 1997).

A curva ROC é gerada alterando o limiar entre as classes; o valor da AUC não depende do limiar. Ou seja, acima desse limiar, o algoritmo classifica em uma classe e abaixo na outra. Quanto maior a AUC, melhor. Uma AUC igual a 0,5 indica um resultado de predição aleatório, e uma AUC igual a 0 indica uma predição perfeitamente reversa (os valores de predição de todas as amostras positivas estão abaixo daqueles das amostras negativas). A AUC é utilizada porque é a métrica que é invariante de escala, pois trabalha com a precisão da classificação em vez de seus valores absolutos. Além disso, ela também mede a qualidade das predições do modelo, independentemente do limiar de classificação (BRADLEY, 1997). A figura 8 representa uma curva ROC genérica e sua métrica AUC.

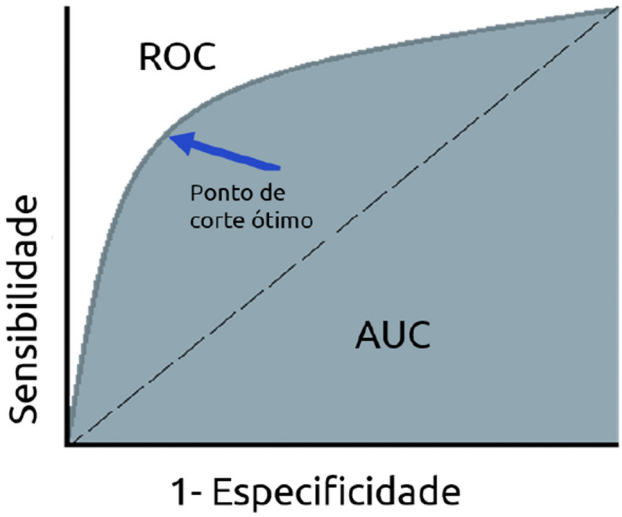


Figura 8: Representação gráfica das métricas ROC e AUC (BRADLEY, 1997).

4.2 Explicação do modelo e criação dos alertas estruturais tóxicos

O modelo treinado foi explicado utilizando a biblioteca LIME (RIBEIRO et al., 2016) (versão 0.2.0.1). O código foi escrito para descrever algumas informações sobre os fragmentos para investigar uma possível correlação entre os 12 receptores biológicos (12 tarefas) e os fragmentos que ativam uma característica específica (impressões digitais) de uma molécula.

Os dados de classificação de cada molécula ativa para cada tarefa são exportados. O número de vezes que um fragmento é encontrado em moléculas que se prevê serem tóxicas e o quanto esse fragmento contribui para a toxicidade dessas moléculas (pesos) são dois fatores importantes para entender como a toxicidade é prevista. Os dados são agrupados por tarefa. Para cada um desses grupos, são obtidas a frequência de contagem de cada fragmento e a soma total dos pesos associados a cada fragmento.

Em seguida, esses dicionários são decompostos de forma que o quadro de dados desejado seja obtido. Além disso, a nova coluna com moléculas ativas é criada neste mesmo quadro de dados e o índice é o do fragmento para um determinado grupo. Assim, uma lista é retornada com todos os IDs de moléculas pertencentes ao grupo. Finalmente, recursos do RDKit (GAVID, 2022; SUGAWARA; NIKAIDO, 2014) são usados para destacar o fragmento e visualizá-lo na molécula associada.

4.3 Codificação, divisão e caracterização dos dados de permeabilidade na Barreira Hematoencefálica

Os pacotes DeepChem (v. 2.7.1.), RDKit (v. 2023.9.1) e LIME (v. 0.2.0.1) foram instalados em uma plataforma Google Colaboratory para implementação e análise dos modelos. Adicionalmente, pacotes auxiliares do Python, como mols2grid (GITHUB, 2023.) (v. 2.0.0), Matplotlib (HUNTER, 2007) (v. 3.7.1), Scikit-learn (PEDREGOSA et al., 2011) (v. 1.2.2), Pandas (REBACK et al., 2022) (v. 1.5.3), Numpy (HARRIS et al., 2020) (v. 1.23.5) e IPython (PEREZ; GRANGER, 2007) foram utilizados para pré-processamento de dados, manipulação de tabelas e suporte computacional. Essas ferramentas constituem um conjunto fundamental de recursos para o desenvolvimento em Aprendizado de Máquina. O mols2grid permite a visualização interativa de moléculas e suas propriedades, facilitando a análise exploratória em química computacional. O matplotlib é utilizado para gerar gráficos e representações visuais dos resultados, enquanto o scikit-learn fornece algoritmos e funções para

treinamento, validação e avaliação de modelos de Aprendizado de Máquina. A biblioteca Pandas organiza e manipula bases de dados de forma eficiente e o numpy oferece operações matemáticas otimizadas para matrizes e vetores, fundamentais em cálculos de alto desempenho. Já o IPython funciona como um ambiente interativo que integra todas essas ferramentas, permitindo experimentação rápida, reprodutibilidade e maior fluidez no processo de pesquisa e modelagem.

O conjunto de dados BBBP (MARTINS et al., 2012) (*Blood-Brain Barrier Penetration*) foi carregado, incorporando rótulos binários (0 para moléculas não penetrantes e 1 para moléculas penetrantes) em propriedades de permeabilidade para mais de 2.000 moléculas, dentre elas fármacos, hormônios e neurotransmissores. O DeepChem foi utilizado para implementar divisores de dados, caracterizadores moleculares, modelos de classificadores de Aprendizado de Máquina e infraestrutura de treinamento e validação.

Cada molécula foi representada por 1024 impressões digitais de conectividade estendida (ECFP). O conjunto de dados caracterizado foi então dividido aleatoriamente em proporção 80/20 para treinar e validar os modelos usando o método de validação cruzada K-fold (K = 5) para os modelos classificadores DRN, RF e ET. Como os modelos são gerados de forma aleatória a cada execução, e o modelo não foi salvo entre rodadas, foi possível gerar diferentes subestruturas para avaliar a consistência dos resultados entre as execuções, abrindo margem para avaliação da variabilidade e robustez das subestruturas identificadas.

4.4 Explicação dos modelos, hiperparametrização e geração dos alertas estruturais de permeabilidade na Barreira Hematoencefálica

O modelo DRN (SUGAWARA; NIKAIDO, 2014) (Deep Residual Network), implementado como MultiTaskClassifier da biblioteca DeepChem, foi inicialmente treinado. Esta arquitetura consiste em uma rede residual profunda totalmente conectada para classificação multitarefa, composta por blocos residuais de pré-ativação, permitindo maior profundidade sem degradação do gradiente. Essa degradação é evitada graças às conexões residuais (*skip connections*), que criam um caminho direto para o gradiente durante o processo de retropropagação, impedindo que ele desapareça ou se amplifique de forma instável em redes muito profundas. A hiperparametrização foi realizada utilizando o *HyperparamOpt* (BERGSTRA et al., 2013) e métricas de ROC-AUC, explorando diferentes configurações, conforme abaixo:

- Tamanho das camadas: [256, 256], [256, 256, 256], [256, 256, 256, 256]
- Dropout: [0, 0], [0, 1]
- Taxa de aprendizagem: [0,005; 0,0035; 0,0025; 0,0010]
- Número de épocas: [5, 10, 20, 30, 40, 50]
- Momentum: [0, 0.9]
- Decaimento: [0, 0.1]
- Tamanho do lote: [32, 64, 128]

Posteriormente, modelos Random Forest (RF) (BREIMAN, 2001; SCHONLAU; ZOU, 2020) e Extra Trees (ET), da biblioteca Scikit-learn (PEDREGOSA, 2011), também foram avaliados. A hiperparametrização desses modelos consistiu na otimização de número de estimadores (50 e 500) e profundidade máxima das árvores (1 a 20), utilizando busca aleatória com validação cruzada (RandomizedCV) e métrica ROC-AUC. Os melhores modelos foram então aplicados aos conjuntos de treinamento e validação, avaliando desempenho com matriz de confusão, acurácia, precisão, recall e F1-score.

Para a interpretação dos modelos, foi utilizado o pacote LIME, por meio da função LimeTabularExplainer (BARR et al., 2020; RIBEIRO et al., 2016;

ZAFAR; KHAN, 2021), que cria um objeto explicador utilizando as 1024 impressões digitais circulares como características e os nomes dos descritores moleculares. O modelo previu e explicou o conjunto de validação considerando 100 recursos por explicação. O objeto explicador retorna um mapeamento entre índices de impressão digital e listas de fragmentos SMILES ativados, que correspondem a subestruturas moleculares.

A avaliação consistiu em explicar, para cada molécula, por que o modelo previu penetrabilidade ou não, permitindo identificar fragmentos críticos associados à permeação da BHE. Os pesos atribuídos pelo LIME variam de -1 a +1, indicando contribuição negativa ou positiva de cada fragmento. Cada molécula possui múltiplos fragmentos, e a classificação final considera a soma das contribuições positivas e negativas, determinando se a molécula é “penetrante” ou “não penetrante”.

Os dados resultantes incluem: ocorrência de cada subestrutura, pesos totais de contribuição e visualização molecular utilizando RDKit e mols2grid para destacar as subestruturas relevantes. Como alterações na largura do núcleo do LIME não produziram mudanças significativas, o valor padrão recomendado foi adotado. Para aumentar a interpretabilidade, subestruturas com pesos pequenos ou negativos foram descartadas, considerando que apenas os fragmentos com contribuições positivas significativas determinam o comportamento do fármaco na permeação da BHE.

5 Resultados e Discussão

5.1 Desempenho do modelo preditivo de Toxicidade

O classificador multitarefa criado com 1024 características usando impressões digitais ECFP foi desenvolvido a partir do conjunto de dados Tox21 e apresentou resultados consistentes tanto na etapa de validação quanto nos testes externos com os bancos Clintox e Sider. Os caracterizadores MACCS (DURANT et al., 2002), MAT (SUGAWARA; NIKAIDO, 2014), PubChem (SUGAWARA; NIKAIDO, 2014) e TokenizerSmiles (SCHWALLER et al., 2021) foram testados como parte da otimização de hiperparâmetros, mas todos os resultados foram semelhantes aos obtidos pelo caracterizador ECFP. A melhor função de perda ocorreu após 200 a 300 épocas. Usando o módulo explicador LIME, nenhum valor de largura do núcleo testado para explicar o modelo treinado produziu resultados instáveis. Os fragmentos produzidos na explicação são, em sua maioria, semelhantes. Portanto, foi utilizado o valor padrão sugerido pelos desenvolvedores do LIME (RIBEIRO et al., 2016), que foi de 0,75 vezes a raiz quadrada do número de colunas dos dados de treinamento.

Os valores de AUC variaram entre 0,71 e 0,77, demonstrando um equilíbrio satisfatório entre sensibilidade e especificidade (mostrados na tabela 1). Esses resultados sugerem que o modelo é capaz de generalizar de forma adequada, evitando sobreajuste e subajuste na previsão dos padrões de toxicidade. Esse desempenho é particularmente relevante considerando-se a heterogeneidade dos alvos biológicos contemplados no Tox21 e a diversidade química dos compostos presentes nos bancos de teste. A manutenção de acurácias semelhantes em bancos de dados independentes reforça a robustez do modelo e indica potencial aplicabilidade em cenários de triagem virtual de larga escala.

Tabela 1: Desempenho do modelo multitarefa, com valores de AUC para validação e teste dos conjuntos de dados Clintox e Sider em diferentes rodadas.

Dataset	Execução	AUC Validação	AUC Teste
Clintox	#1	0.719	0.75
Clintox	#2	0.746	0.765
Clintox	#3	0.735	0.761
Sider	#1	0.731	0.721
Sider	#2	0.731	0.731
Sider	#3	0.73	0.727

Outra tentativa importante de usar apenas resultados significativos na explicação é a eliminação de fragmentos sem importância. Por uma questão de concisão, foram eliminados alguns fragmentos sem importância com pesos pequenos ou negativos usando a média de todos os pesos. Ao utilizar apenas o peso máximo ou os pesos máximo e mínimo dos fragmentos, o número de alertas tóxicos gerados foi próximo de um terço daqueles produzidos usando a média dos pesos. Matematicamente, a eliminação de fragmentos com pesos negativos ou pequenos parece estar incorreta porque, desta forma, são eliminados fragmentos que podem fazer com que o comportamento tóxico em uma molécula desapareça. Quimicamente, o fragmento com peso mínimo (fragmento não tóxico) não anula o comportamento de um fragmento com peso máximo (fragmento tóxico). De todo modo, o que importa para o comportamento em uma molécula é o fragmento tóxico, justificando o uso desse corte.

Idealmente, as moléculas a serem explicadas devem ser apenas aquelas previstas como tóxicas pelo modelo. No entanto, isso não é exatamente o que é mostrado na tabela 2, que expõe um certo número de moléculas não tóxicas. Apesar de serem considerados compostos tóxicos pelo modelo para uma determinada tarefa (ou seja, para um alvo biológico), isso não é calculado corretamente quando analisado individualmente. Isso ocorre porque o saldo final da soma resultou em números negativos ao considerar todas as contribuições de peso de toxicidade de seus respectivos fragmentos. Mesmo com contribuições tóxicas positivas, a soma de todos os pesos pode ser menor que

zero, gerando a classificação não tóxica. No entanto, a maioria das moléculas que entraram na classificação foram previstas corretamente. O aparecimento de algumas moléculas não tóxicas na tabela de classificação é aceitável.

Tabela 2: Número de moléculas classificadas como tóxicas e não tóxicas para cada alvo biológico nos bancos de dados Clintox e Sider em diferentes execuções.

Execução	#1				#2				#3			
Banco de dados	**Clintox**		**Sider**		**Clintox**		**Sider**		**Clintox**		**Sider**	
Alvos	Tóxico	Não tóxico	Tóxico	Não tóxico	Tóxico	Não tóxico	Tóxico	Não tóxico	Tóxico	Não tóxico	Tóxico	Não tóxico
NR-AR	67	0	57	0	73	0	62	0	73	0	52	0
NR-AR-LBD	55	2	47	1	61	1	46	2	59	2	41	1
NR-AhR	5	0	2	1	3	3	4	0	3	1	2	1
NR-Aromatase	3	1	4	1	4	3	2	1	0	6	2	1
NR-ER	30	4	30	6	35	4	31	6	34	2	26	6
NR-ER-LBD	22	0	21	5	26	1	23	1	23	2	17	3
NR-PPAR-γ	0	0	0	0	0	0	0	0	0	0	0	0
SR-ARE	11	3	9	4	7	10	9	4	9	6	8	4
SR-ATAD5	2	0	2	0	2	1	2	1	2	0	2	0
SR-HSE	1	1	1	1	3	0	1	1	3	0	1	0
SR-MMP	0	0	0	0	12	7	0	0	0	0	0	0
SR-p53	0	0	0	0	7	3	0	0	0	0	0	0

5.2 Identificação e interpretação química dos alertas estruturais tóxicos

A utilização do LIME possibilitou interpretar localmente as previsões do modelo e gerar listas de fragmentos estruturais associados à toxicidade. Observou-se a recorrência de subestruturas simples (hidrocarbonetos saturados e insaturados) e de grupos funcionais mais complexos, incluindo nitrocompostos, derivados sulfurados e radicais halogenados. Para aumentar a relevância química das interpretações, decidiu-se prestar mais atenção ao número de átomos pesados de cada fragmento. Percebeu-se que os fragmentos com menos de 7 átomos pesados (isto é, átomos diferentes de hidrogênio) são principalmente fragmentos de hidrocarbonetos saturados e insaturados e, em menor extensão, grupos oxigenados e nitrogenados.

Mais importante ainda, esses fragmentos também fazem parte dos fragmentos maiores e complexos com mais de 6 átomos pesados que têm baixa frequência na tarefa. Assim, decidiu-se filtrar os pequenos fragmentos da lista, mantendo apenas os fragmentos com mais de 6 átomos pesados para evitar redundância. A evidência mais sólida que apoia essa escolha é que o corte está no centro do faixa de número de átomos dos compostos. Embora esse limite seja facilmente alterado pelo usuário, os limites de 4 a 10 foram testados. Parece que a extremidade inferior da faixa produz muitos fragmentos e a extremidade superior produz apenas alguns fragmentos. Portanto, o meio da faixa parece ser a melhor escolha. Essa escolha reduziu a redundância de pequenos radicais comuns e permitiu concentrar a análise em subestruturas com maior probabilidade de impactar mecanismos toxicológicos.

Alertas estruturais devem ser progressivamente desenvolvidos para permitir a compreensão do mecanismo de ação de compostos químicos (Cronin et al., 2017). Algumas evidências científicas podem ser mostradas para apoiar nossos resultados. Furanos, fenóis, nitroaromáticos e tiofenos são encontrados como alertas tóxicos na literatura (DANG et al., 2017), mas apenas alguns álcoois, compostos nitro e compostos de enxofre foram encontrados em nosso estudo. Li e colaboradores (LI et al., 2014) investigaram fragmentos tóxicos por análise de frequência e identificaram algumas subestruturas presentes em

compostos altamente tóxicos e moderadamente tóxicos. Algumas dessas subestruturas foram encontradas por LIME, como um alquilfluoreto, derivado sulfênico, cianoidrina e nitrila. Yang e colaboradores (YANG et al., 2017) compararam pequenos radicais tóxicos obtidos por diferentes técnicas (como aprendizado de máquina, grafos e sistemas especialistas).

Lei e colaboradores (LEI et al., 2016) apresentaram alguns fragmentos tóxicos gerados a partir de um banco de dados de toxicidade oral aguda em ratos, como fluoreto de alquila e aminas. Esses fragmentos também são gerados por LIME usando os conjuntos de dados Tox21/Clintox/Sider. Adicionalmente, LIME também encontrou alertas estruturais com compostos halogenados que são encontrados em estudos de disrupção endócrina com receptores de andrógeno e estrogênio (CHEN et al., 2014).

Essas associações podem estar relacionadas a diferentes mecanismos de toxicidade, como interações inespecíficas com receptores nucleares, geração de espécies reativas ou formação de metabólitos eletrofílicos. É importante destacar que a forma com que esses fragmentos atuam no aumento da toxicidade ainda é um desafio (Alves et al., 2016; Hemmerich et al., 2020; Kalgutkar, 2020). Devido ao estudo ainda ser introdutório, a necessidade da melhoria da escolha de fragmentos mais condizentes é uma realidade.

Foram identificados importantes fragmentos tóxicos, como radical alquilfluoreto, tioéster (reatividade covalente com proteínas, podendo causar hepatotoxicidade e sensibilização dérmica), enonas (que são alertas clássicos pois podem atuar na inibição de enzimas e são ligadas à toxicidade hepática e cutânea), tiol (podem ser irritantes e neurotóxicos em pequenas moléculas voláteis) e fosfina (podem ser neurotóxicas e atuar na falência respiratória).

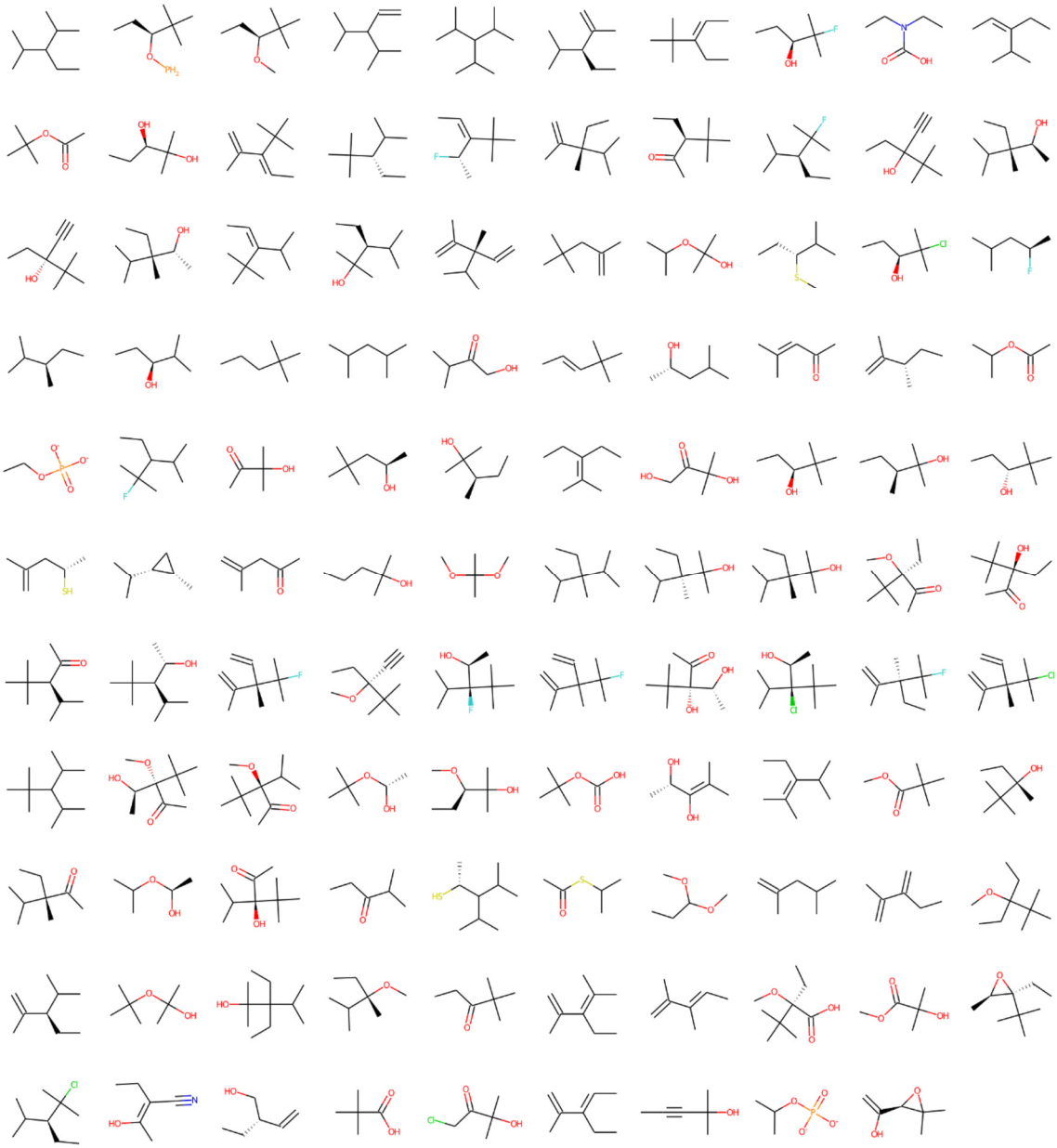


Figura 9: Alertas estruturais que influenciaram o aumento da toxicidade para os doze alvos biológicos no conjunto de dados Clintox. A ordem não representa influência tóxica de cada fragmento.

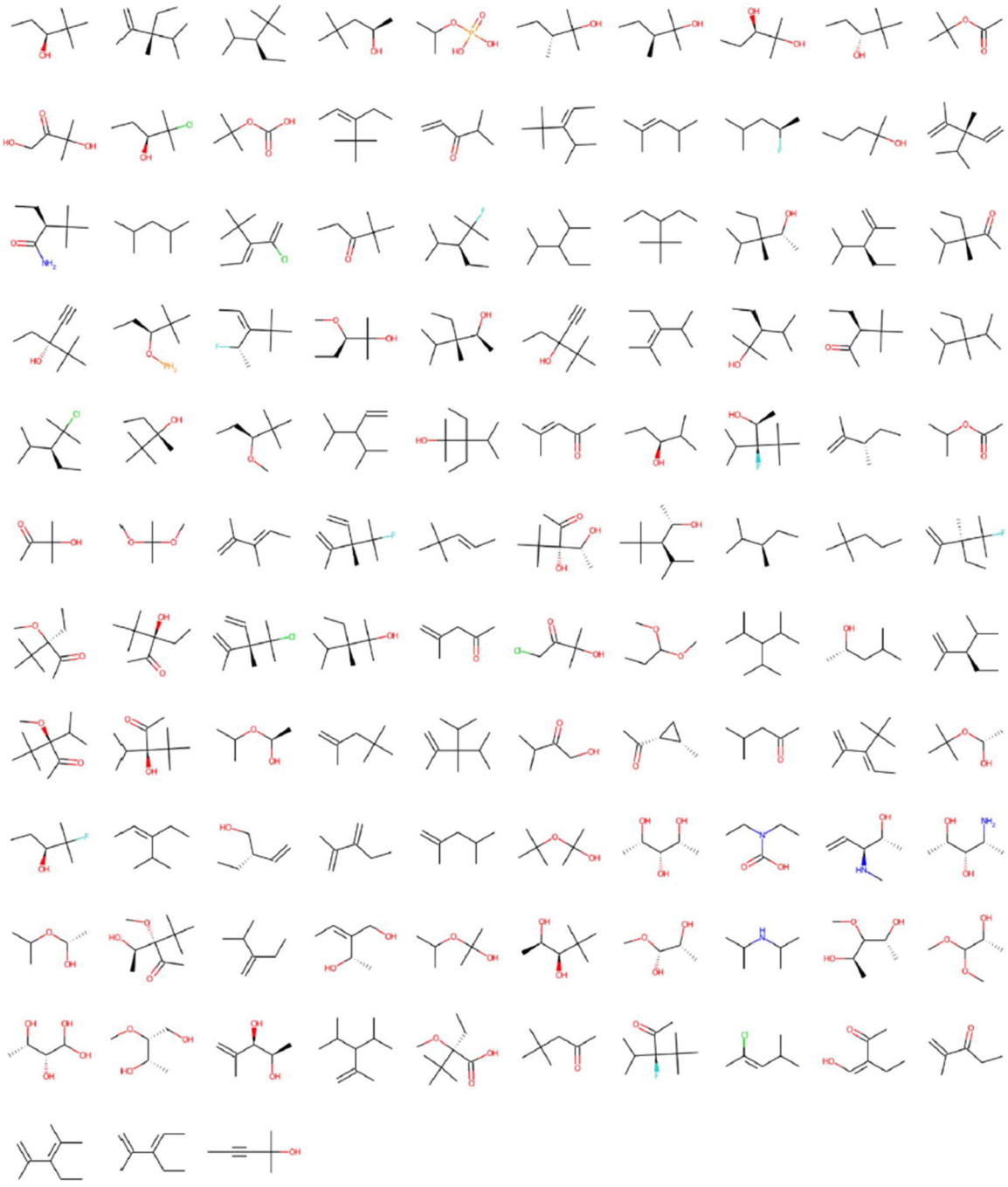


Figura 10: Alertas estruturais que influenciaram o aumento da toxicidade para os doze alvos biológicos no conjunto de dados Sider. A ordem não representa influência tóxica de cada fragmento.

5.3 Robustez e limitações metodológicas

Para analisar a consistência do modelo LIME aplicado nesses conjuntos de dados, dos 109 alertas estruturais gerados pelo conjunto de dados Clintox (Execução nº 1), 98 aparecem na Execução nº 2 e 89 são repetidos na Execução nº 3, correspondendo a 90% e 82% de similaridade, respectivamente. Além disso, as Execuções nº 2 e nº 3 têm 96 alertas estruturais semelhantes, correspondendo a 79% e 85% de suas estruturas, respectivamente.

Analisando os 113 alertas estruturais gerados pela Execução nº 1 no conjunto de dados Sider, 96 estruturas são repetidas na Execução nº 2 (85% de similaridade) e 85 na Execução nº 3 (75% de similaridade). Além disso, as execuções 2 e 3 têm 95 alertas estruturais semelhantes, correspondendo a 71% e 83% de suas estruturas, respectivamente. Finalmente, é importante analisar todas as estruturas geradas para o conjunto de dados Clintox (rodadas #1, #2 e #3), excluindo as repetidas (total de 143 alertas), e para o conjunto de dados Sider (rodadas #1, #2 e #3), excluindo as repetidas (total de 160 alertas). Há apenas 38 estruturas não repetidas, mostrando alta similaridade e consistência entre ambos os resultados. As Figuras 11, 12 e 13 apresentam esses dados.

Figura 11: Ilustração com número de alertas estruturais repetidos e não repetidos entre os bancos de dados Clintox e Sider.

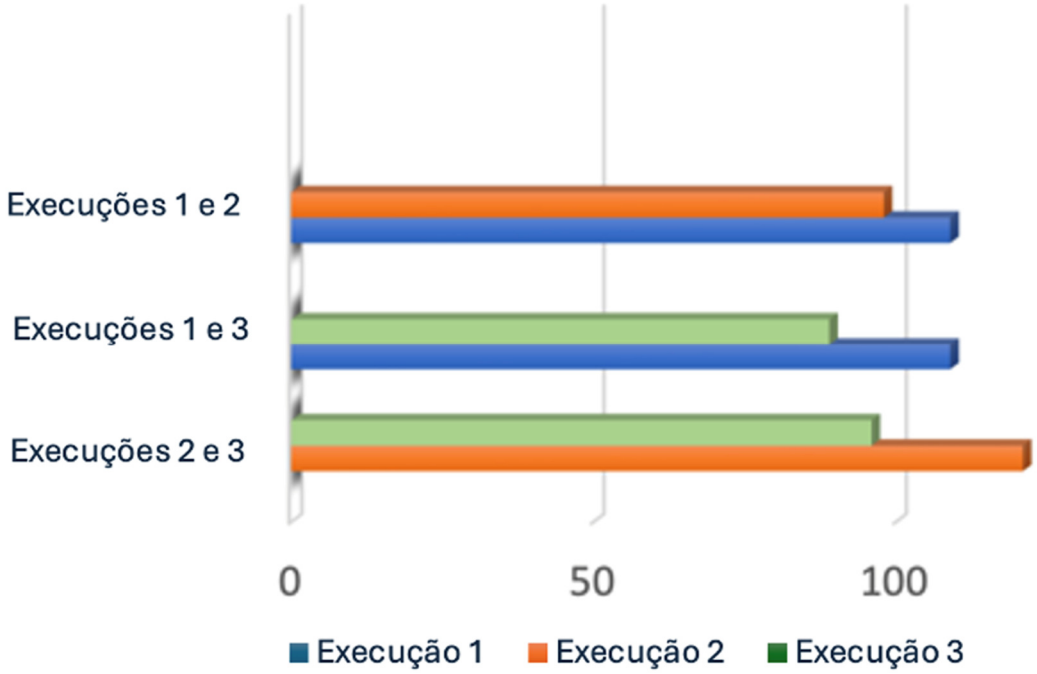


Figura 12: Similaridade entre cada par de rodadas utilizando o Clintox.

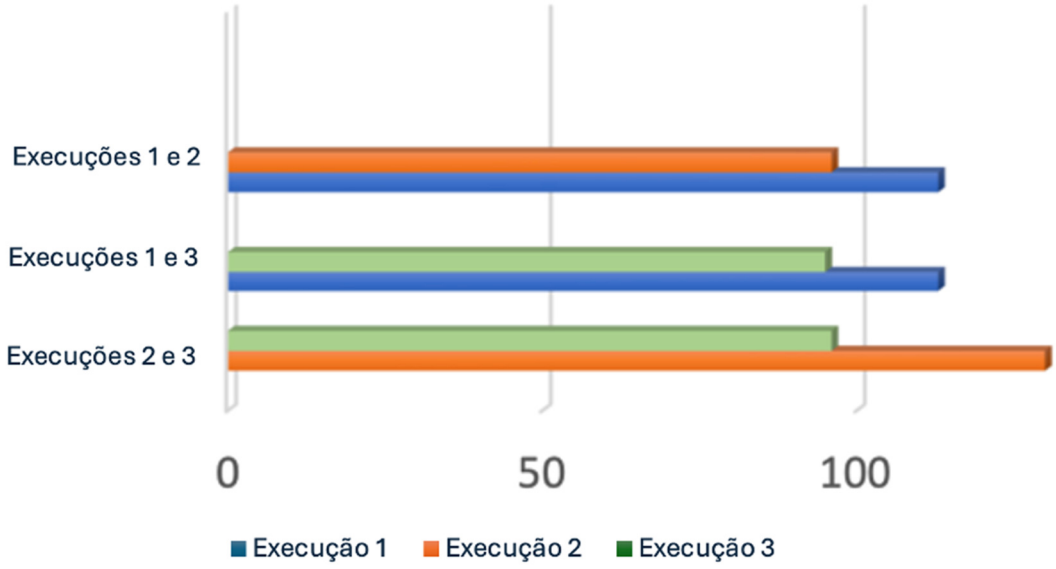


Figura 13: Similaridade entre cada par de rodadas utilizando o Sider.

O modelo foi treinado com os conjuntos de dados Tox21 e testado com os conjuntos de dados Clintox e Sider. Nenhum desses conjuntos de dados contém fármacos com problemas de mutagenicidade, portanto, o modelo não seria capaz de destacar anéis aromáticos, nitrosaminas, epóxidos e compostos nitro, por exemplo.

O código foi executado com o conjunto de dados Tox21 com divisão de dados (80:10:10) para verificar se esse conjunto de dados seria capaz de destacar esses fragmentos, mas isso não produziu nenhum destaque para eles. A explicação para esse resultado é que as tarefas (alvos biológicos) não estão relacionadas à mutagenicidade, portanto, os modelos não devem destacar anéis aromáticos. Para o modelo, o benzeno não é tóxico em nenhuma das 12 tarefas do conjunto de dados Tox21, como esperado.

Um exemplo de experimentos de prova de conceito foi prever se uma molécula que contém fragmentos mutagênicos, como um anel aromático, um grupo nitro, um grupo nitrosamina ou um grupo epóxido, seria capturada pelo LIME. Diferentemente de tarefas de toxicidade, esses experimentos devem ter uma verdade fundamental muito clara para explicações. LIME deve destacar esses grupos e funcionou como um experimento para ver se o LIME poderia escolher as estruturas corretas em tais tarefas.

Para provar esse conceito, foi realizada uma análise preliminar utilizando o conjunto de dados de mutagenicidade Bursi, um conjunto de dados contendo mais de quatro mil estruturas moleculares com os resultados relativos do ensaio de Ames (JAIN et al., 2018). Foram encontrados fragmentos que são conhecidos como mutagênicos, incluindo o grupo C–C=C (encontrado no anel aromático) e epóxido. Os cinco principais compostos mutagênicos encontrados neste experimento preliminar são apresentados na figura 14.

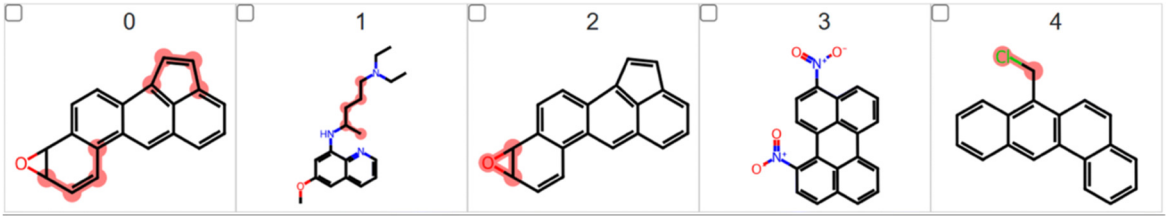


Figura 14: Os cinco principais compostos mutagênicos encontrados pelo LIME quando aplicado ao conjunto de dados de mutagenicidade de Bursi, escolhendo fragmentos maiores que um átomo pesado.

Embora os resultados demonstrem boa reprodutibilidade, algumas limitações foram observadas. Em primeiro lugar, o LIME não distingue representações alternativas da mesma molécula (SMILES não canônicos), o que pode gerar redundância na lista final de fragmentos. Além disso, a dependência do particionamento aleatório dos dados pode levar a pequenas variações na frequência e contribuição de certos fragmentos.

Outro ponto importante refere-se à natureza dos conjuntos de dados utilizados. Como o Tox21 não contempla ensaios de mutagenicidade ou carcinogenicidade clássica, fragmentos reconhecidamente mutagênicos, como anéis aromáticos ou nitrosaminas, não foram priorizados nas análises. Esse resultado evidencia que a validade das interpretações está fortemente vinculada ao tipo de ensaio biológico modelado.

Um grande desafio quando se trabalha com a criação de alertas estruturais é a existência de duas subestruturas na mesma molécula. O LIME se aplica a esse problema. Existem alguns métodos que também consideram a existência de duas ou mais subestruturas no mesmo composto (YANG et al., 2020) O padrão emergente é um método amplamente utilizado, no qual um padrão molecular é identificado como um conjunto de fragmentos moleculares. Como exemplo, o método "*jumping emerging pattern mining*" (SHERHOD et al., 2012) é um algoritmo de mineração para encontrar os padrões que atribuem pares de átomos como descritores. Além desse, o método de mineração de

padrões emergentes usa o algoritmo de árvore de padrões de contraste para encontrar características tóxicas (SHERHOD et al., 2014).

Outro problema provavelmente pode vir de um alerta estrutural gerado a partir de um conjunto de dados com falsos positivos. O modelo explicável para encontrar alertas estruturais pode não interpretar corretamente o efeito de alertas estruturais redundantes existentes em um conjunto de compostos falsos positivos. Embora existam alguns métodos, evitar falsos positivos em conjuntos de dados ainda é um desafio. Portanto, o desenvolvimento futuro de modelos explicáveis para encontrar alertas estruturais deve considerar a generalização de alertas estruturais específicos para evitar redundância.

5.4 Aplicação dos alertas tóxicos em bibliotecas de compostos

Os alertas estruturais identificados demonstram aplicabilidade prática na filtragem de bibliotecas químicas. A aplicação do conjunto de fragmentos gerado a uma base de inibidores do receptor EGFR (GAULTON et al., 2012) resultou na eliminação de centenas de compostos contendo subestruturas potencialmente tóxicas, reduzindo o espaço químico a ser investigado em etapas posteriores. Esse tipo de abordagem pode auxiliar pesquisadores em química medicinal e toxicologia regulatória, permitindo concentrar esforços em moléculas com menor risco de falha por toxicidade em fases clínicas. Mais ainda, a interpretação fornecida por modelos explicáveis oferece subsídios para modificações estruturais racionais, mitigando efeitos adversos sem comprometer a atividade farmacológica desejada.

A tabela 3 também mostra os resultados da filtragem de prováveis fragmentos indesejados obtidos usando as impressões digitais ECFP no conjunto de dados EGFR com compostos que mostram alta afinidade de ligação ao EGFR. Após a filtragem PAINS, a filtragem através da “regra dos cinco de Lipinski” e após terem sido aplicadas ao conjunto de dados EGFR original, o conjunto de dados possui 4266 compostos. O número de subestruturas indesejadas encontradas pelo LIME e o número de compostos sem

subestruturas indesejadas são apresentados para confirmar a filtragem de alertas estruturais no conjunto de dados EGFR.

Tabela 3: Resultados da filtragem de fragmentos indesejados obtidos usando as impressões digitais do ECFP no conjunto de dados do EGFR com 4.266 compostos que apresentam alta afinidade de ligação ao EGFR com conformidade com RO5 e PAINS. O número de subestruturas indesejadas encontradas e o número de compostos sem subestruturas indesejadas são apresentados para confirmar a filtragem de alertas estruturais no conjunto de dados do EGFR.

Dados	Execução	Número de subestruturas indesejadas encontradas	Número de compostos sem os fragmentos indesejados
Clintox	#1	147	4183
	#2	163	4174
	#3	110	4218
Sider	#1	237	4165
	#2	311	4123
	#3	380	4043

5.5 Curadoria dos dados de permeabilidade

As representações SMILES no conjunto de dados de referência foram retificadas manualmente, abordando um problema em que onze entradas SMILES no conjunto de dados BBBP representavam átomos de nitrogênio tetravalentes sem carga. É essencial observar que átomos de nitrogênio tetravalentes devem invariavelmente possuir carga. A presença desses erros tornou as estruturas químicas ilegíveis pelo pacote computacional RDKit. Apesar do amplo uso desse conjunto de dados em inúmeros estudos, há uma notável ausência de discussões sobre o tratamento dessas representações SMILES inválidas. Para garantir a integridade das estruturas químicas, as representações SMILES foram corrigidas e as cargas positivas necessárias foram introduzidas quando necessário nos conjuntos de dados de referência.

No conjunto de dados de penetração da BHE, 81 estruturas foram identificadas como duplicatas e 7 como triplicatas, o que é indesejável para um banco de dados. Para resolver esse problema, esses compostos duplicados ou triplicados foram removidos para evitar quaisquer problemas potenciais. Notavelmente, o conjunto de dados BBB inclui 14 compostos (13 pares e 1 trio), onde as moléculas idênticas são inconsistentemente rotuladas como, por exemplo, BHE penetrante e BHE não penetrante ao mesmo tempo. Desta forma, essas inconsistências foram removidas. Algumas dessas inconsistências foram encontradas porque as moléculas são tratadas como transportadas por P-gp em uma anotação e como transportadas passivamente na outra anotação.

5.6 Otimização de hiperparâmetros de permeabilidade

Inicialmente, a arquitetura do método classificador DRN foi explorada com tamanhos de camada de [64], [128], [256], [512] e [1024]. O melhor resultado foi alcançado com um tamanho de camada de [256]. Então, a arquitetura foi otimizada com tamanhos de camada de [[256, 256], [256, 256, 256] e [256, 256, 256, 256]]. A otimização de hiperparâmetros dos modelos classificadores DRN, RF e ET foi investigada para encontrar os melhores parâmetros. Os hiperparâmetros otimizados foram: taxa de aprendizagem, épocas, momento e número de estimadores. Os parâmetros ótimos para os modelos DRN, RF e ET variaram a cada execução.

Utilizando esses parâmetros ótimos, as métricas ROC-AUC para cada modelo de classificação foram, em sua maioria, superiores a 0,90 para os conjuntos de dados de treinamento e validação, conforme mostrado na tabela 4. Portanto, não há indicação de sobreajuste substancial. Embora as pontuações de treinamento para os modelos DRN, RF e ET tenham sido, em sua maioria, superiores a 0,98, as pontuações de validação do modelo RF são ligeiramente superiores às dos modelos DRN e ET.

Tabela 4: Métricas (ROC-AUC) dos conjuntos de dados de treinamento e validação para a permeabilidade na BHE em três diferentes execuções utilizando os modelos DRN, RF e ET.

Modelo	Execução	Treinamento	Validação
DRN	#1	0.999	0.894
	#2	0.976	0.790
	#3	0.989	0.882
RF	#1	0.999	0.931
	#2	0.999	0.943
	#3	0.999	0.936
ET	#1	0.999	0.913
	#2	0.999	0.925
	#3	0.999	0.915

Os parâmetros do explicador LIME foram investigados. O parâmetro "número de características", que define o número máximo de características dentro de uma única explicação, foi examinado. Apesar de o modelo classificador conter 1024 características, menos de uma dúzia delas se mostrou significativa para cada amostra. A modificação desse parâmetro de 20 para 500 não teve impacto perceptível nos resultados ou no tempo de execução. No entanto, uma análise de componentes principais para todas as amostras mostrou que 542 e 232 impressões digitais, do total de 1024 impressões digitais usadas na caracterização, foram significativas, representando 95% da variância explicada cumulativa nos conjuntos de dados de treinamento e validação, respectivamente, conforme apresentado na figura 15. Foi descoberto que os resultados de 542 e 1024 impressões digitais foram semelhantes. Inicialmente, uma configuração conservadora de 100 características foi escolhida para segurança.

Outro parâmetro estudado foi o número de explicações com a maior probabilidade de predição para cada subestrutura. Este parâmetro variou de zero (nenhum) a 10. Aumentar este valor resultou em resultados mais refinados devido ao aumento da complexidade nas explicações. No entanto, foi observado que valores mais altos reduziram o número de subestruturas com pesos maiores que 0,1. Consequentemente, foi tomada a decisão de definir o número de explicações para 1, maximizando a quantidade de subestruturas. Valores altos para o número de explicações correlacionaram-se com tempos de avaliação

prolongados, excedendo 1 a 2 horas. Isso ressalta a importância de evitar valores excessivamente altos para manter tempos de processamento razoáveis.

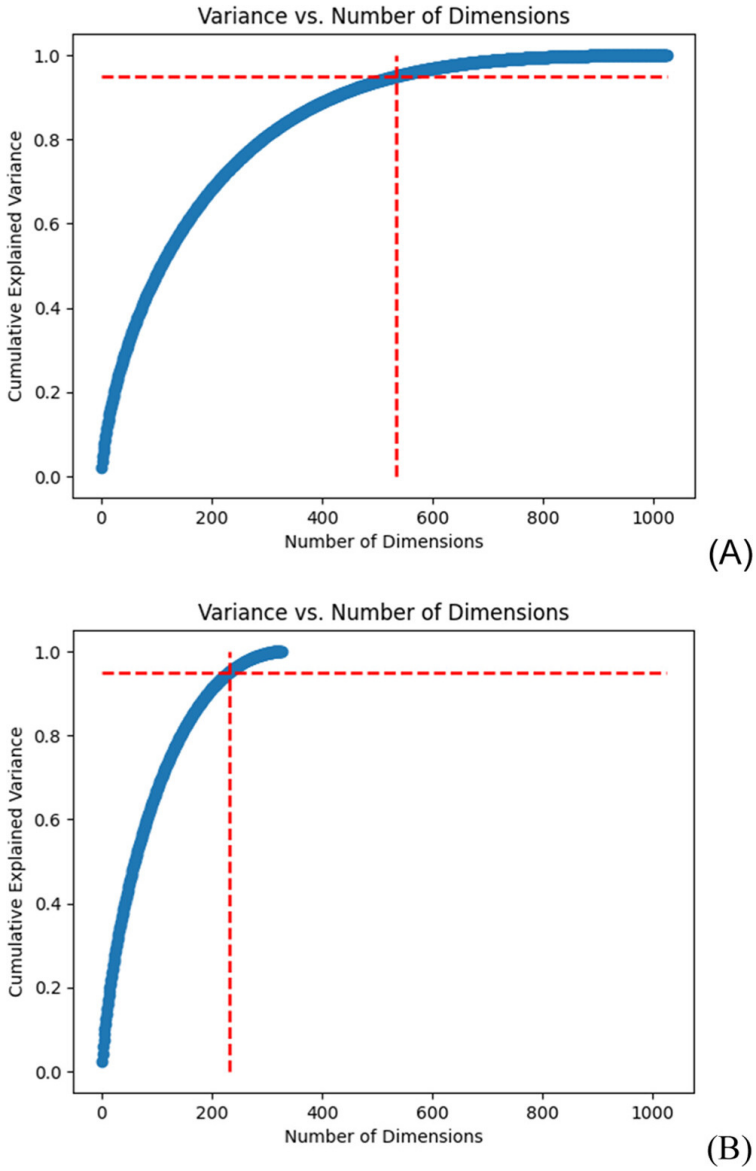


Figura 15: Análise de componentes principais do conjunto de dados de penetração da BHE representando as 542 e 232 impressões digitais (linha tracejada vertical em vermelho) que são significativas dentro de uma variância explicada cumulativa de 95% (linha tracejada horizontal em vermelho) para todas as amostras nos conjuntos de dados de treinamento e validação, respectivamente. O número de impressões digitais usadas na caracterização é 1024, representando a dimensionalidade total do conjunto de dados de treinamento.

5.7 Análise estatística

A pontuação de precisão é a razão TP/(TP + FP) e a pontuação de *recall* é a razão TP/(TP + FN), onde TP é o número de verdadeiros positivos, FP é o número de falsos positivos e FN é o número de falsos negativos. Instâncias de Falso Positivo (FP) e Verdadeiro Positivo (TP) ocorrem quando o modelo prevê um ponto de dados positivo incorretamente e quando o modelo prevê com precisão um ponto de dados positivo, respectivamente. Por outro lado, instâncias de Verdadeiro Negativo (TN) e Falso Negativo (FN) ocorrem quando o modelo prevê corretamente uma amostra de dados negativo e quando o modelo prevê incorretamente uma amostra de dados negativo, respectivamente.

A pontuação F1 é a média harmônica da precisão e da sensibilidade (da palavra em inglês *recall*). As pontuações de precisão, recall e F1 atingem seus melhores valores em 1 e piores valores em 0. Enquanto a precisão é a capacidade do classificador de não rotular uma amostra negativa como positiva, a pontuação de recall é a capacidade do classificador de encontrar todas as amostras positivas. A precisão é uma métrica usada para avaliar o desempenho de modelos de classificação. É o número de previsões corretas como uma porcentagem do número de observações no conjunto de dados. A precisão, *recall*, pontuações F1 e pontuações de acurácia para os modelos classificadores para o conjunto de dados de validação são apresentados na tabela 5.

Tabela 5: Acurácia, precisão, recall, F1 e MCC para os modelos DRN, RF e ET em três diferentes execuções dos conjuntos de dados de validação para permeabilidade na BHE.

Modelo	Execução	Acurácia	Precisão	Recall	F1	MCC
DRN	#1	0.868	0.868	0.868	0.868	0.645
	#2	0.896	0.896	0.896	0.896	0.689
	#3	0.885	0.885	0.885	0.885	0.660
RF	#1	0.893	0.908	0.956	0.893	0.696
	#2	0.883	0.900	0.954	0.883	0.653
	#3	0.891	0.899	0.963	0.891	0.688
ET	#1	0.878	0.895	0.945	0.878	0.672
	#2	0.888	0.906	0.954	0.888	0.670
	#3	0.898	0.909	0.964	0.898	0.699

A detecção de uma subestrutura falsamente identificada como irrelevante pode acontecer quando a explicação local não é precisa em relação ao comportamento subjacente do modelo. Esse comportamento pode ocorrer devido a vários motivos, como amostragem inadequada de dados, complexidade do modelo ou presença de ruído nos dados. Se os pesquisadores quiserem verificar a relevância e a importância das subestruturas identificadas por meio do LIME no contexto da permeação da BHE para ganhar mais confiança, eles podem utilizar as seguintes estratégias de validação. Primeiro, as subestruturas identificadas devem ser comparadas com a experiência de profissionais na área de permeação da BHE para garantir a relevância biológica. Segundo, experimentos podem ser conduzidos usando dados existentes para validar os efeitos dessas subestruturas. Em terceiro lugar, foram aplicados testes estatísticos para avaliar a significância das subestruturas identificadas na previsão da permeação. Em quarto lugar, foi realizada uma análise de sensibilidade para determinar se as subestruturas identificadas permanecem relevantes sob diferentes perturbações de dados. Por fim, o impacto das subestruturas identificadas nas previsões do modelo foi analisado, comparando o desempenho do modelo com e sem elas.

Como pode ser observado, os modelos classificadores realizam boa classificação para compostos penetrantes na BHE e uma classificação pior para compostos não penetrantes na BHE. Também é importante observar que os modelos RF e ET são mais precisos e exatos do que o modelo DRN. A matriz de confusão é usada para avaliar a precisão da classificação. Por definição, uma matriz de confusão C é tal que C_{ij} é igual ao número de observações conhecidas por estarem no grupo $_i$ e previstas para estarem no grupo $_j$. Assim, a contagem de TN é C_{00}, FN é C_{10}, TP é C_{11} e FP é C_{01}, conforme apresentado na figura 16 para os modelos DRN, RF e ET no conjunto de dados de validação da penetração da BHE.

Os valores de TP são grandes e os valores de TN são pequenos, o que significa que o modelo de classificação classifica com precisão os compostos penetrantes na BHE, mas não os compostos não penetrantes. Os valores de FP e FN também são baixos, o que significa que o modelo de classificação não produz muitas classificações falsas. É claro que há muito mais compostos

penetrantes no conjunto de dados de validação da penetração da BHE, portanto, o modelo de classificação deve ser usado apenas para classificar os compostos penetrantes na BHE e a classificação para os compostos não penetrantes não é suficientemente precisa para esse propósito. Vale ressaltar que um modelo de classificação diferente deve ser construído usando os compostos não penetrantes caso haja interesse na classificação dos compostos não penetrantes na BHE. Nessa situação, o elemento C_{00} na matriz de confusão seria muito maior do que o elemento C_{11}.

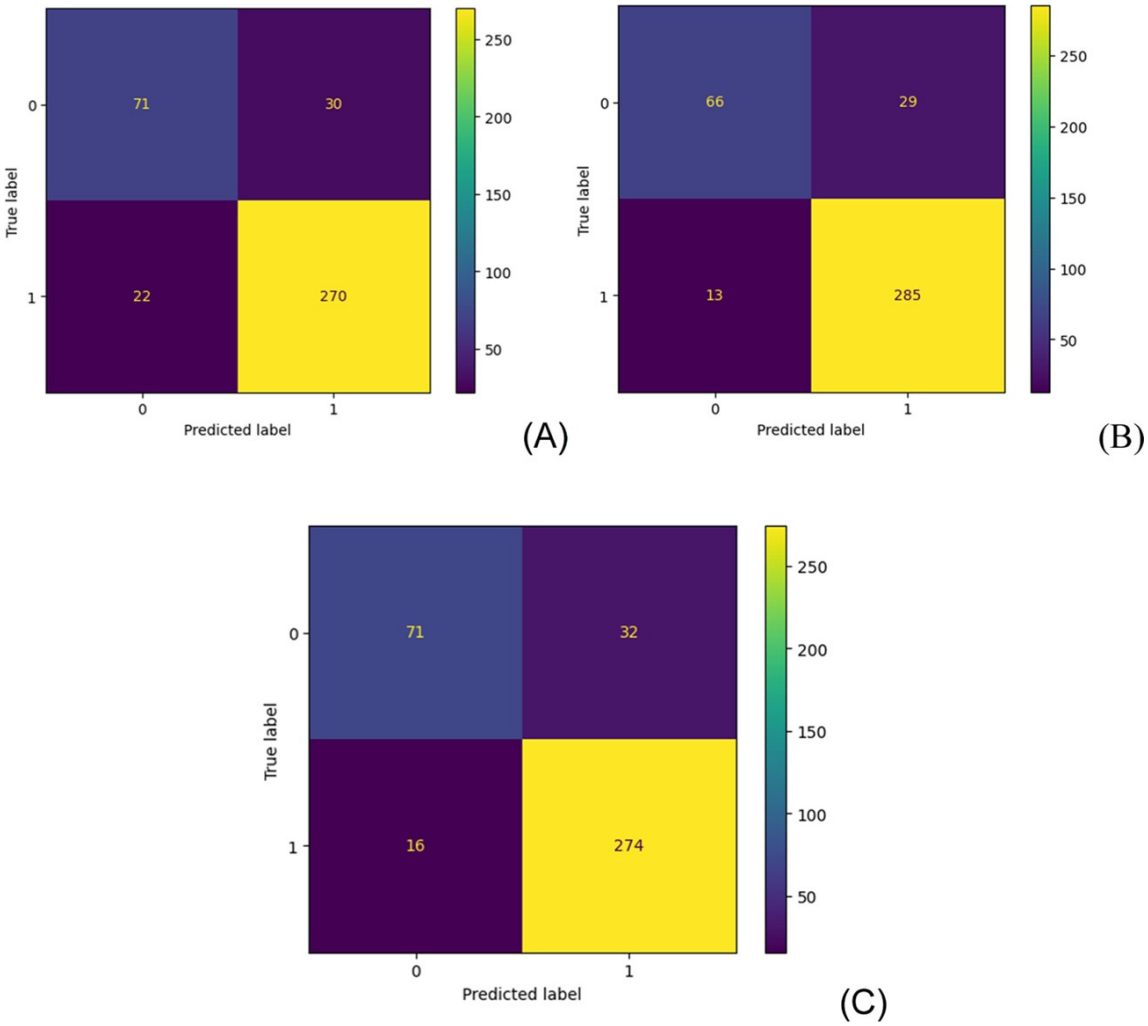


Figura 16: Matriz de confusão para os modelos de classificação DRN (A), RF (B) e ET (C) usando o conjunto de dados de validação da permeabilidade na BHE.

Não é possível comparar se algum modelo ou métrica é melhor que outro devido às armadilhas de comparar modelos de classificação e métricas de desempenho apenas com base em estatísticas. O desvio padrão é destacado como uma medida de variabilidade em vez de um teste estatístico. Ele quantifica a variação ou dispersão em um conjunto de valores. Embora forneça informações úteis sobre a dispersão dos dados, não é um teste estatístico em si. Métodos como o teste *t* de Student e a análise de variância (ANOVA) são testes estatísticos paramétricos. Esses testes têm suposições sobre a distribuição dos dados (por exemplo, normalidade e independência) que podem nem sempre ser verdadeiras em cenários do mundo real.

Para abordar as limitações dos testes paramétricos, recomenda-se o uso de alternativas não paramétricas. O teste *t* de Student pode ser substituído pelo teste de soma de postos de Wilcoxon, que é um análogo não paramétrico adequado para comparar dois métodos. Para comparar mais de dois métodos, a ANOVA é sugerida, mas com cautela sobre suas limitações. Em vez disso, recomenda-se o uso do teste de Friedman, uma alternativa não paramétrica à ANOVA, particularmente útil ao lidar com dados que não são independentes ou normalmente distribuídos. Portanto, é importante considerar a distribuição e a natureza dos dados ao comparar modelos de classificação e métricas de desempenho. Sugere-se deixar de depender exclusivamente de desvio-padrão e testes paramétricos, optando por alternativas não paramétricas, para comparações mais robustas em vários cenários.

O campo da interpretabilidade em modelos de ML pode progredir metodologicamente em várias etapas importantes. Primeiro, há a necessidade de avançar em abordagens agnósticas de modelo para oferecer explicações interpretáveis a um grupo mais diverso de modelos e conjuntos de dados. Segundo, a integração de conhecimento específico de domínio em métodos de interpretabilidade é essencial para aumentar a relevância e a precisão das explicações. Terceiro, desenvolver métodos para quantificar a incerteza associada a explicações interpretáveis é crucial para fornecer aos usuários uma compreensão diferenciada das previsões do modelo. Quarto, aprimorar a escalabilidade e a eficiência dos métodos de interpretabilidade é fundamental para lidar prontamente com grandes conjuntos de dados e modelos complexos.

Por fim, estabelecer procedimentos padronizados de validação é fundamental para avaliar o desempenho dos métodos de interpretabilidade e comparar sua eficácia em vários domínios e tarefas. Ao enfrentar esses desafios metodológicos, o campo pode progredir em direção a modelos de Aprendizado de Máquina mais confiáveis.

5.8 Explicação do modelo de permeabilidade

As explicações do LIME para a penetração na BHE são apresentadas abaixo. Os resultados foram filtrados com pesos maiores que 0,1 para remover subestruturas menos importantes. Em geral, LIME explica principalmente a penetrabilidade de compostos usando estruturas com nitrogênio. A basicidade é considerada um dos principais fatores para a penetrabilidade no SNC, especialmente para grupos com nitrogênio, porque eles podem ser protonados em pH fisiológico (como grupos piperidina e piperazina). Para a penetração na BHE, a ligação dupla C = N dos aromáticos também é considerada. LIME detectou o peso positivo dessa característica na contribuição para a penetração na BHE. Anéis aromáticos também tendem a facilitar a penetração porque a hidrofobicidade é aumentada.

A figura 17 mostra as subestruturas mais importantes encontradas pelo LIME para a penetração usando o modelo de classificação DRN para as execuções #1, #2 e #3. Da execução #1, seis compostos têm grupos contendo nitrogênio (principalmente grupos amina cíclicos e não cíclicos), mas apenas três deles são destacados no grupo nitrogênio. Nos outros três compostos, o LIME destacou a vizinhança do grupo nitrogênio. No entanto, as subestruturas destacadas nos compostos 1 e 6 são grupos contendo oxigênio e enxofre, respectivamente, embora grupos contendo flúor não destacados sejam encontrados nessas duas moléculas. Os compostos 3 e 5 têm grupos contendo cloro não destacados, com o primeiro tendo um grupo contendo fósforo destacado.

Na execução #2, sete compostos têm grupos contendo nitrogênio, e cinco deles têm grupos contendo nitrogênio destacados. Embora o composto 6 não seja destacado no grupo contendo nitrogênio, o composto 7 é destacado na vizinhança do grupo contendo nitrogênio. Os compostos 3 e 8 foram destacados como grupos contendo oxigênio e halogênio, respectivamente.

Na execução #3, oito compostos têm grupos contendo nitrogênio, mas três deles não são destacados no grupo nitrogênio. Os compostos 1 e 3 têm grupos contendo flúor, mas os destaques estão nos grupos contendo oxigênio. Os compostos 0 e 2 são repetições com destaques em diferentes grupos contendo nitrogênio. Os compostos 5 e 6 são destacados na parte hidrocarboneto dessas moléculas, mas os grupos contendo nitrogênio não são destacados nessas moléculas. O composto 9 também é destacado na parte do hidrocarboneto da molécula com dois grupos contendo enxofre não destacados.

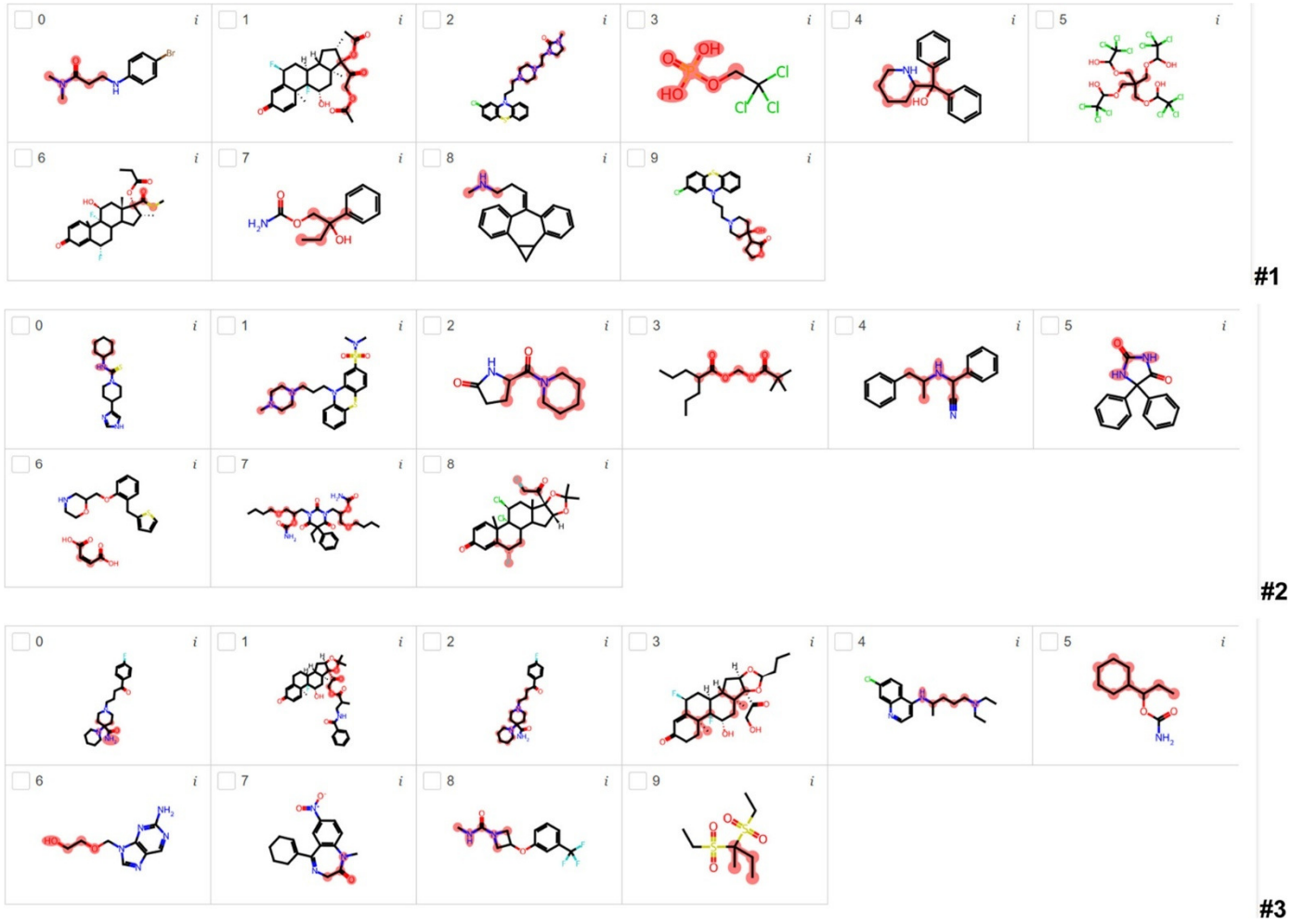


Figura 17: Subestruturas mais importantes encontradas pelo LIME para penetração na BHE usando o modelo de classificação DRN para as execuções # 1, # 2 e # 3.

A figura 18 apresenta as subestruturas mais importantes encontradas pelo LIME para a penetração na BHE usando o modelo de classificação RF para as execuções #1, #2 e #3. Na execução #1, cinco compostos são destacados com grupos contendo nitrogênio. Quatro deles (1, 2, 6 e 7) são repetições com subestruturas contendo nitrogênio destacadas muito semelhantes. Os compostos 0, 4 e 8 também são repetições, mas com diferentes grupos contendo oxigênio destacados. O composto 5 destacou subestruturas com grupos contendo oxigênio.

Na execução #2, seis subestruturas contendo nitrogênio são destacadas em compostos diferentes. Os compostos 3 e 4 são repetições com destaques bastante semelhantes, e o composto 5 é um composto muito semelhante com um grupo contendo nitrogênio destacado. Os compostos 2, 7 e 8 também são repetições com alertas estruturais bastante semelhantes. Os compostos 1 e 6 também são repetições com destaques diferentes na subestrutura cíclica com grupos contendo oxigênio e flúor. O composto 0 é uma repetição do composto 5 na execução #1, mas com destaques semelhantes (na mesma região).

Na execução #3, sete compostos contendo nitrogênio são encontrados pelo LIME, mas dois deles são destacados na vizinhança contendo nitrogênio. Os outros cinco compostos têm grupos contendo nitrogênio destacados. Os compostos 0 e 3 são compostos contendo oxigênio e flúor com destaques diferentes. Os compostos 4 e 8 contêm subestruturas contendo nitrogênio não destacadas; no entanto, o destaque está na vizinhança. Os compostos 5 e 7 são repetições com subestruturas contendo nitrogênio destacadas muito semelhantes.

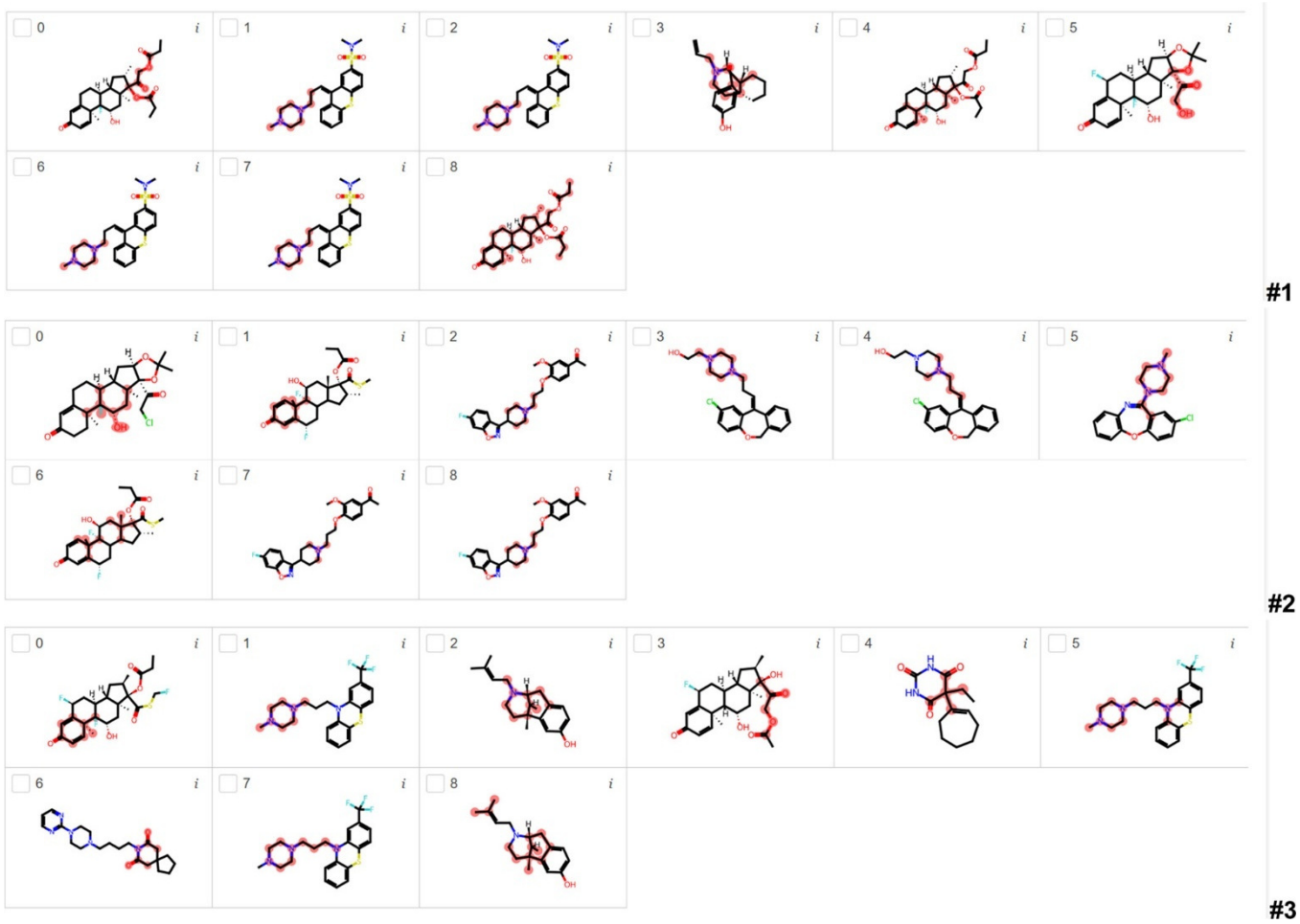


Figura 18: Subestruturas mais importantes encontradas pelo LIME para penetração na BHE usando o modelo de classificação RF para as execuções #1, #2 e #3.

A figura 19 mostra as subestruturas mais importantes encontradas pelo LIME para a penetração na BHE usando o modelo de classificação ET. Na execução #1, seis subestruturas contendo nitrogênio são destacadas em compostos diferentes. Os compostos 4, 5 e 7 são repetições com subestruturas destacadas muito semelhantes. Os compostos 3 e 6 têm grupos contendo oxigênio e halogênio com diferentes subestruturas destacadas não contendo nitrogênio. Na execução #2, seis compostos têm subestruturas destacadas contendo nitrogênio; no entanto, os compostos 1, 2 e 7 são repetições com destaques ligeiramente diferentes. Os compostos 0 e 4 são compostos contendo oxigênio e halogênio muito diferentes com subestruturas destacadas muito diferentes. Na execução #3, sete compostos contendo nitrogênio são destacados com subestruturas diferentes; no entanto, cinco deles são subestruturas contendo nitrogênio. Os compostos 4 e 7 têm subestruturas destacadas sem nitrogênio, mas o último tem um grupo contendo nitrogênio na vizinhança.

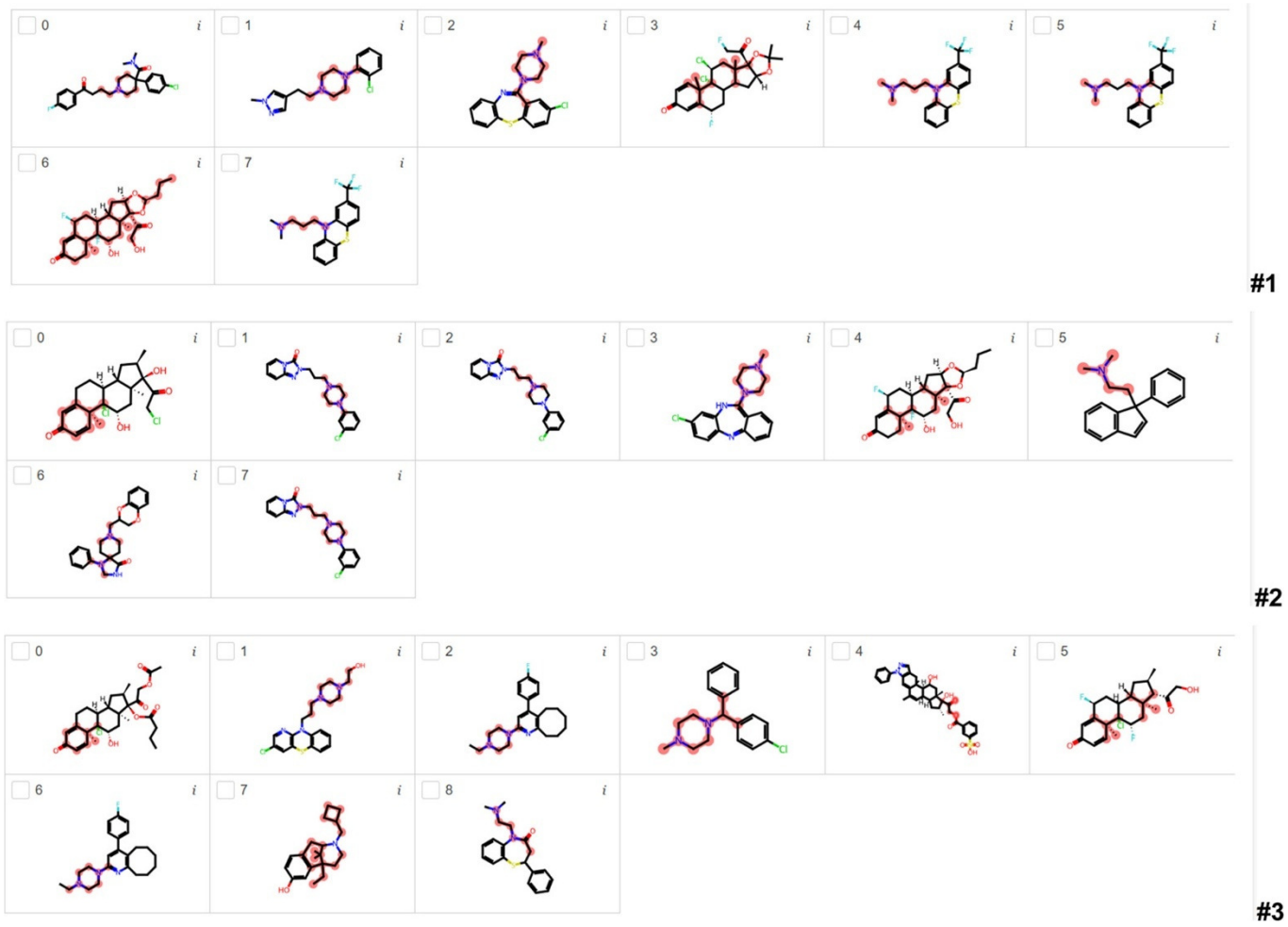


Figura 19: Subestruturas mais importantes encontradas pelo LIME para penetração da BHE usando o modelo de classificação ET para as execuções #1, #2 e #3.

A permeabilidade de compostos químicos através da BHE é uma interação complexa de vários fatores. Compreender a importância dos grupos funcionais é crucial. Embora a lipofilicidade desempenhe um papel fundamental na determinação da capacidade de penetração da BHE, diferentes grupos funcionais interagem com os componentes da BHE, influenciando na permeabilidade de um composto. No entanto, variações na lipofilicidade causadas por diferentes grupos funcionais impactam o transporte de compostos químicos, influenciando a solubilidade dos compostos químicos nas membranas ricas em lipídios da barreira.

Geralmente, compostos hidrofóbicos tendem a ter uma melhor permeação da BHE. No entanto, as faixas ótimas de lipofilicidade para uma permeação eficiente da BHE dependem das características específicas dos grupos funcionais dos compostos. Portanto, compostos muito hidrofílicos podem ter dificuldades para penetrar na BHE, enquanto compostos excessivamente lipofílicos podem enfrentar desafios relacionados à eliminação e ao transporte.

De fato, a permeabilidade de compostos químicos através da BHE é intrinsecamente influenciada por grupos funcionais específicos contendo nitrogênio. Portanto, vários grupos funcionais importantes foram identificados por seu potencial de aumentar a permeabilidade à BHE. Por exemplo, grupos amino ($-NH_2$, $-NRH$ e $-NR_1R_2$) em um composto podem participar de interações, afetando sua capacidade de penetração na BHE. A Cloroquina (composto 4 na execução do modelo DRN nº 3) é uma aminoquinolina usada para a prevenção e terapia da malária. Ela atenua a neuroinflamação. O LIME descobriu que a subestrutura da quinolina não causa penetração na BHE; no entanto, os dois grupos amina na cadeia lateral são destacados como subestruturas potenciais que aumentam a penetração na BHE. O LIME também descobriu o fármaco anfetaminil (composto 4 na execução do modelo DRN nº 2), um estimulante derivado da anfetamina que possui um grupo amina. Outro exemplo é o maleato de tioperamida (composto 0 na execução DRN nº 2), usado para prevenir convulsões ou reduzir sua gravidade. Por fim, o tiazesim (composto 8 na execução do modelo ET nº 3) é um antidepressivo heterocíclico relacionado aos antidepressivos tricíclicos. Nesta molécula, o LIME destacou o grupo amina

terciária na cadeia lateral ligada à subestrutura da amina terciária heterocíclica. Nosso estudo mostra mais detalhes com outras moléculas (ROSA et al., 2024) .

Em relação à redução da lipofilicidade do composto, é importante mencionar que outros grupos funcionais também contribuem para as variações da lipofilicidade. Grupos contendo enxofre, como o tiol (–SH), também podem influenciar a lipofilicidade, afetando potencialmente a capacidade do composto de penetrar na BHE. O átomo de enxofre nos grupos tiol pode interagir com os componentes da BHE, facilitando a penetração na BHE. Compostos contendo grupos sulfóxido ($R_2S{=}O$) e sulfona ($–SO_2R_1R_2$) podem apresentar maior permeabilidade. Certos compostos com grupos tiossulfinato, caracterizados por um átomo de enxofre ligado tanto ao oxigênio quanto ao enxofre, podem influenciar positivamente a permeabilidade à BHE. Ao explorar as interações de grupos funcionais contendo enxofre, é possível obter informações valiosas sobre o desenvolvimento de compostos com capacidades aprimoradas de penetração na BHE.

Átomos de halogênio (particularmente átomos de bromo, cloro e flúor) podem ter um impacto significativo na permeabilidade da BHE. A presença de átomos de halogênio introduz características únicas que influenciam as interações com os componentes da barreira. Átomos de flúor podem aumentar a permeabilidade da BHE porque o flúor é um átomo pequeno e eletronegativo que pode influenciar a lipofilicidade e as propriedades eletrônicas do composto. Portanto, ele potencialmente promove interações favoráveis para a permeação da BHE. Semelhante ao flúor, a presença de átomos de cloro pode alterar as propriedades físico-químicas de um composto, afetando sua capacidade de penetrar na BHE. O tamanho maior do cloro em comparação ao flúor introduz efeitos estéricos e eletrônicos específicos que contribuem para as características gerais de permeação do composto.

Os resultados obtidos encontraram muitos corticosteroides, como diacetato de diflorasona, triancinolona benetonida, dipropionato de betametasona, acetonido de fluocinolona, halcinonida, propionato de fluticasona, acetato de parametasona, rofleponida, mometasona, enbutato de icometasona e cortisuzol. De todos eles, o cortisuzol é o único corticosteroide que não possui

um átomo de halogênio em sua estrutura. LIME destacou uma subestrutura contendo oxigênio para explicar sua permeação na BHE para o cortisuzol. Observa-se que, com exceção da halcinonida, nenhum corticosteroide teve um átomo de halogênio destacado pela LIME, indicando que eles podem apenas modular a lipofilicidade dos corticosteroides. A maioria dos corticosteroides teve subestruturas contendo oxigênio destacadas pela LIME.

Na busca por desvendar os mistérios que cercam a penetração de corticosteroides na BHE devido à sua estrutura complexa com átomos de halogênio e muitas subestruturas cíclicas contendo oxigênio, esta Tese pode trazer informações importantes que convidam a uma reavaliação dos paradigmas existentes, superando o conhecimento existente obtido por meio de métodos explicáveis encontrados na literatura até o momento (GANDHI; WHITE, 2022; SINGH et al., 2020; WELLAWATTE et al., 2022) . O presente estudo investiga a intrincada relação entre corticosteroides e a BHE, esclarecendo o papel fundamental desempenhado pelos átomos de halogênio, especificamente átomos de flúor e cloro, bem como grupos contendo oxigênio.

Esses constituintes moleculares surgem como moduladores potentes da lipofilicidade dos corticosteroides, atuando como guardiões que influenciam a permeabilidade da BHE. Os resultados ressaltam a importância de considerar não apenas a composição estrutural e a lipofilicidade dos corticosteroides, mas também a interação diferenciada de átomos de halogênio e grupos contendo oxigênio na determinação de sua capacidade de transpor a BHE. Essa revelação pode ter profundas implicações para o design e desenvolvimento de fármacos.

Além disso, alguns opioides foram encontrados: levalorfano (composto 7 no modelo ET nº 3), pentazocina (compostos 2 e 8 na execução do modelo RF nº 3) e cogazocina (composto 3 na execução do modelo RF nº 1). A subestrutura contendo nitrogênio é explicada pela LIME como a causa da penetração na BHE em duas moléculas, pentazocina e cogazocina. As duas subestruturas destacadas no levalorfano estão na vizinhança da subestrutura contendo nitrogênio que foi explicada como a causa da penetração na BHE para as outras duas moléculas.

Os resultados sugerem que átomos de nitrogênio podem ter um efeito especial na facilitação da permeação de moléculas orgânicas, particularmente quando esses átomos de nitrogênio podem ser protonados em condições fisiológicas. Os estudos conduzidos por Singh e colaboradores (SINGH et al., 2020) revelam certas subestruturas comumente encontradas em compostos que permeiam a BHE. Esses resultados corroboram com o presente estudo, visto que a presença de compostos contendo nitrogênio e anéis aromáticos é mais prevalente em compostos que permeiam a BHE em comparação com compostos não permeáveis.

Embora outros grupos funcionais, como compostos contendo oxigênio, enxofre e halogênio, desempenhem um papel na regulação e no equilíbrio da lipofilicidade de um composto, grupos contendo nitrogênio parecem ser fundamentais para influenciar a permeabilidade do composto. Especificamente, subestruturas contendo nitrogênio, como carbamato, barbitúrico, piperidina, piperazina, pirazina, fenotiazina, dibenzotiazepina, benzodiazepina, butirofenona e dibenzoxepina são consideradas mais relevantes para a permeação da BHE. No entanto, tanto a lipofilicidade quanto a carga dos compostos precisam ser levadas em consideração ao determinar a permeação de um composto químico na BHE. Os corticosteroides podem servir como exemplos da importância da lipofilicidade. Além disso, os compostos opioides parecem possuir uma subestrutura contendo nitrogênio que é vital para a permeação da BHE.

5.9 Revalidação da metodologia

A validação cruzada *5-fold* serve como uma técnica valiosa para avaliar o desempenho de generalização de modelos de ML em dados não vistos. Ela facilita a otimização de hiperparâmetros e auxilia na seleção de modelos para um determinado conjunto de dados. No entanto, o uso do mesmo procedimento de validação cruzada e conjunto de dados para ajuste e seleção de modelos pode levar a avaliações tendenciosas do desempenho do modelo. Portanto, a abordagem de validação cruzada aninhada é preferível para superar esse viés. Essa abordagem inclui o procedimento de otimização de hiperparâmetros dentro

do procedimento de seleção de modelos, fornecendo uma avaliação menos tendenciosa dos modelos ajustados.

Os hiperparâmetros da maioria dos algoritmos de ML podem ser ajustados para se adequarem a um conjunto de dados específico. No entanto, existem poucas diretrizes sobre como configurá-los. Em vez disso, um procedimento de otimização é usado para encontrar os melhores hiperparâmetros para o conjunto de dados. A validação cruzada *5-fold* pode selecionar os melhores hiperparâmetros para cada modelo e a melhor configuração do modelo, sendo eficaz para estimar o desempenho do modelo, mas apresenta limitações. Quando usada várias vezes com o mesmo algoritmo, pode levar ao sobreajuste. Cada avaliação de um modelo com diferentes hiperparâmetros fornece informações sobre o conjunto de dados. Conjuntos de dados com ruído tendem a apresentar pontuações piores, e esse conhecimento pode ser usado para encontrar a melhor configuração para o conjunto de dados.

Embora a validação cruzada *5-fold* tente reduzir esse efeito, ela não consegue eliminá-lo completamente. A validação cruzada aninhada aborda o problema do sobreajuste do conjunto de dados de treinamento. Expor a busca de hiperparâmetros a apenas um subconjunto do conjunto de dados fornecido pelo procedimento de validação cruzada externa reduz o risco de sobreajuste e fornece uma estimativa menos tendenciosa do desempenho do modelo no conjunto de dados.

Por outro lado, optou-se por reamostrar o conjunto de dados de penetração da BHE desbalanceado usando a biblioteca padrão Scikit-Learn. O método de reamostragem, combinado com a validação cruzada *5-fold*, visava combater o sobreajuste, em vez de causá-lo. Esse sobreajuste ocorre quando um modelo aprende o ruído nos dados de treinamento em vez de capturar o padrão subjacente, resultando em baixo desempenho em novos dados. A técnica de reamostragem auxilia na avaliação do desempenho do modelo em dados não observados, treinando-o iterativamente em subconjuntos variados do conjunto de dados e avaliando seu desempenho nos dados restantes. Esse processo iterativo oferece uma estimativa robusta do desempenho e previne o sobreajuste, promovendo generalização eficaz para novos dados. Portanto, a

técnica de reamostragem foi fundamental para abordar potenciais problemas de sobreajuste.

A tabela 6 apresenta a média da área sob a curva ROC-AUC dos conjuntos de dados de treinamento e validação para o modelo de penetração da BHE em três execuções diferentes, utilizando os modelos classificadores DRN, RF e ET para a permeação de compostos na barreira hematoencefálica, utilizando o método de reamostragem e validação cruzada de 5 vezes.

Tabela 6: Métricas (ROC-AUC) dos conjuntos de dados de treinamento e validação para a permeabilidade na BHE em três diferentes execuções utilizando os modelos DRN, RF e ET a partir da reamostragem e validação cruzada 5-fold.

Modelo	Execução	Treinamento	Validação
DRN	#1	1.000	0.973
	#2	1.000	0.977
	#3	1.000	0.987
RF	#1	1.000	0.988
	#2	1.000	0.986
	#3	1.000	0.995
ET	#1	1.000	0.989
	#2	1.000	0.992
	#3	1.000	0.988

As pontuações de precisão, recall, F1 e exatidão, bem como o coeficiente de correlação de Matthew (MCC) para os modelos classificadores DRN, RF e ET, são apresentadas na tabela 7 para três execuções diferentes para o conjunto de dados de validação de permeação da barreira hematoencefálica de compostos usando o método de reamostragem e validação cruzada *5-fold*. A figura 20 apresenta a matriz de confusão para os modelos de classificação DRN, RF e ET usando o método de reamostragem e validação cruzada *5-fold* para o conjunto de dados de penetração da BHE.

Tabela 7: Acurácia, precisão, recall, F1 e MCC para os modelos DRN, RF e ET em três diferentes execuções dos conjuntos de dados de validação para permeabilidade na BHE utilizando a reamostragem e validação cruzada 5-fold.

Modelo	Execução	Acurácia	Precisão	Recall	F1	MCC
DRN	#1	0.952	0.952	0.952	0.952	0.906
	#2	0.924	0.924	0.924	0.924	0.850
	#3	0.935	0.935	0.935	0.935	0.873
RF	#1	0.954	0.947	0.960	0.954	0.908
	#2	0.950	0.934	0.968	0.950	0.902
	#3	0.973	0.966	0.983	0.973	0.947
ET	#1	0.959	0.948	0.972	0.959	0.919
	#2	0.968	0.966	0.972	0.968	0.936
	#3	0.954	0.950	0.957	0.954	0.908

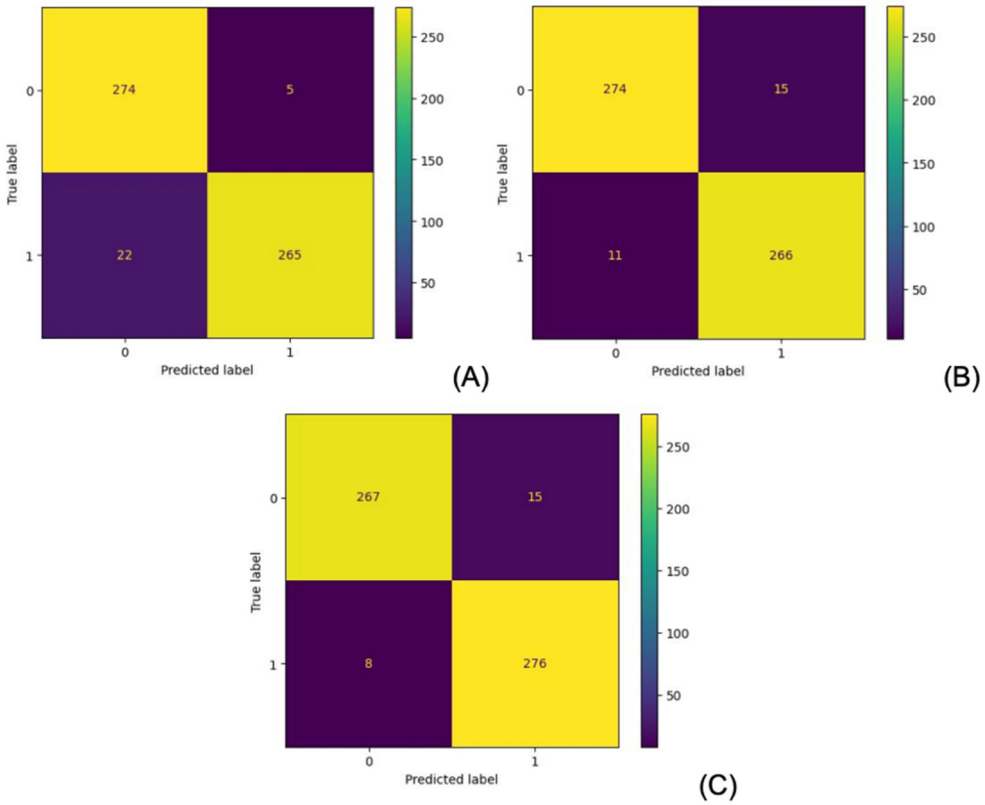


Figura 20: Matriz de confusão para os modelos DRN (A), RF (B) e ET (C) utilizando reamostragem e validação cruzada 5-fold em três execuções diferentes para a permeabilidade de compostos na BHE.

Se o modelo apresentar um desempenho significativamente melhor no conjunto de dados de treinamento em comparação com o conjunto de dados de validação, isso pode indicar sobreajuste. Considerando isso, os resultados mostram que as pontuações ROC-AUC no conjunto de dados de treinamento não são drasticamente maiores do que as do conjunto de dados de validação (tabelas 4 e 6) antes e depois da reamostragem. Portanto, isso significa que o método de reamostragem evitou o sobreajuste.

Adicionalmente, foi constatado que as pontuações de precisão, recall, F1 e exatidão do modelo no conjunto de dados de treinamento não são significativamente maiores do que as do conjunto de dados de validação (ver tabelas 5 e 7). Além disso, a precisão variou de 0,868 a 0,898. No entanto, o coeficiente de correlação de Matthew (MCC) no conjunto de dados de validação variou de 0,645 a 0,699 devido à natureza desbalanceada do conjunto de dados de penetração da BHE. O MCC fornece uma medida da qualidade das classificações binárias com base em toda a matriz de confusão. É considerada uma medida balanceada e pode ser usada mesmo quando as classes têm diferenças substanciais em tamanhos, como é o caso aqui com o conjunto de dados de penetração da BHE. O MCC varia de -1 a +1, onde +1 indica uma previsão perfeita, 0 representa uma previsão aleatória e -1 significa discordância total entre a previsão e a observação.

Embora as métricas de precisão, recall, F1 e acurácia ofereçam entendimentos valiosos sobre vários aspectos do desempenho do modelo, o MCC é frequentemente preferido, pois considera todos os elementos da matriz de confusão e é especialmente útil com conjuntos de dados desbalanceados. Vale ressaltar que, embora o modelo tenha sido construído utilizando pesos para lidar com o desequilíbrio do conjunto de dados, o MCC ainda variou apenas de 0,645 a 0,699. Consequentemente, optou-se por empregar a técnica de reamostragem, resultando em um coeficiente de correlação de Matthew aprimorado no conjunto de dados de validação, variando de 0,850 a 0,947.

Outro ponto importante em relação à metodologia é que, antes da otimização dos parâmetros, conjunto de dados foi dividido em uma proporção de 80:20. Posteriormente, utilizou-se o conjunto de dados de treinamento (que

consistia em 80% dos dados) durante o processo de otimização dos parâmetros. Essa abordagem garantiu que o modelo construído não apresentasse sobreajuste. Além disso, a validação cruzada *5-fold* e a validação cruzada aninhada foram realizadas para fins de comparação. Além disso, o conjunto de dados de penetração da BHE desbalanceado foi reamostrado, levando a resultados melhores.

As pontuações ROC-AUC nas tabelas 4 e 6 mostraram que os modelos de floresta aleatória e árvores extras não apresentam sobreajuste significativo, com pontuações em torno de 0,99 e 0,91–0,94 para treinamento e validação, respectivamente. O modelo de rede residual profunda (DRN) apresentou pontuações ROC-AUC variando de 0,976 a 0,999 para treinamento e 0,790 - 0,894 para validação, o que é uma indicação de sobreajuste. No entanto, os resultados no conjunto de dados de penetração da BHE desbalanceado foram razoáveis apenas para penetração de fármacos (rótulo 1) e ruins para penetração de não fármacos (rótulo 0), com uma pontuação MCC de apenas 0,65 - 0,70 (antes do balanceamento do conjunto de dados), conforme apresentado nas tabelas 5 e 7, embora a precisão, o F1 e a exatidão para a penetração tenham sido razoáveis.

Após a reamostragem do conjunto de dados de penetração da BHE, o modelo apresentou bom desempenho para ambos os rótulos, com MCC variando de 0,90 a 0,95 usando validação cruzada de 5 vezes e validação cruzada aninhada de 5 × 10 (5 vezes para validação interna e 10 vezes para validação externa). As pontuações ROC-AUC para treinamento e validação do modelo construído com validação cruzada aninhada 5 × 10 são em torno de 0,999 e 0,988, respectivamente (resultados semelhantes para validação cruzada de 5 vezes).

5.10 Limitações e benefícios da técnica

As explicações estruturais não podem ser vistas como regras (Alves et al., 2016), e as propriedades farmacológicas peculiares de cada molécula na permeabilidade da BHE e a biologia do sistema não devem ser subestimadas.

Para exemplificar que essas análises não são regras bem definidas, mas sim interpretações estruturais, algumas moléculas, como moléculas semelhantes a esteroides nas figuras 17, 18 e 19 possuem grupos polares, resultando em uma área de superfície polar maior em comparação com as outras moléculas. No entanto, as moléculas semelhantes a esteroides são identificadas como permeáveis, embora a maioria dessas moléculas não possua grupos contendo nitrogênio.

De fato, é importante admitir que o LIME também pode levar a interpretações equivocadas. Por exemplo, o grupo maleato destacado no maleato de teniloxazina (composto 6 na execução do modelo DRN nº 2) não é a razão para a permeação da BHE. Outro exemplo de limitação está relacionado aos pró-fármacos. Triclofos (composto 3 na execução do modelo DRN nº 1) é um medicamento sedativo raramente usado para tratar insônia. É um pró-fármaco que é metabolizado no fígado, transformando-se no fármaco ativo tricloroetanol. A figura 17 mostra o triclofos destacado como o grupo fosfato, mas não é o grupo ativo. O mesmo se aplica ao petricloral (composto 5 na execução do modelo DRN nº 1), que é um pró-fármaco sedativo e hipnótico do hidrato de cloral.

No entanto, interpretar subestruturas importantes para a permeabilidade da BHE com um certo nível de confiança pode servir de base para a análise de contrafactuais e redes neurais gráficas, que estão sendo cada vez mais exploradas por sua capacidade de sugerir mudanças estruturais em moléculas para melhorar sua capacidade de permeabilidade.

Da mesma forma, a regra dos cinco de Lipinski (LIPINSKI, 2004; LIPINSKI et al., 2001) afirma que um fármaco deve ter certas propriedades bem definidas. Ao longo dos anos, na ciência recente, a descoberta de fármacos tem aderido cada vez menos a essas regras (HARTUNG et al., 2023). Um bom fármaco para atuar no SNC deve permear a BHE. Estudos mostram que é mais fácil permear com um número menor de doadores e aceptores de ligações de hidrogênio, por exemplo. É difícil criar uma regra para prever se uma molécula pode ser um bom fármaco.

O ceticismo em torno da eficácia do LIME em auxiliar pesquisadores na síntese de novas moléculas ativas decorre de vários fatores. Primeiro, o desenvolvimento de novas moléculas com a atividade biológica desejada é um problema altamente complexo e multidimensional. Embora o LIME possa fornecer explicações para previsões de modelos individuais, ele pode não capturar toda a complexidade das interações moleculares e dos requisitos estruturais para a atividade. Segundo, as explicações do LIME são limitadas a insights locais e específicos de cada instância e podem não fornecer uma visão holística das relações estrutura-atividade subjacentes, cruciais para o projeto de moléculas. Terceiro, a interpretação dos resultados do LIME requer expertise em aprendizado de máquina e química, tornando desafiador para pesquisadores sem sólida formação em ambas as áreas a utilização eficaz do método. Quarto, quaisquer informações fornecidas pelo LIME ainda precisariam ser validadas experimentalmente, o que pode ser demorado e custoso.

Embora o LIME possa oferecer informações sobre as previsões de modelos de aprendizado de máquina na descoberta de fármacos, sua utilidade em orientar a síntese de novas moléculas ativas pode ser limitada devido à falta de descobertas experimentais sobre alertas estruturais e os pesquisadores devem considerar suas potenciais desvantagens e limitações ao incorporá-lo em seu fluxo de trabalho.

Desconsiderar o LIME e persistir no emprego de técnicas de ponta em pesquisa e descoberta de fármacos pode inadvertidamente limitar a compreensão e a interpretação de previsões geradas por modelos complexos de Aprendizado de Máquina. O LIME apresenta diversas vantagens em relação a métodos alternativos e explicáveis de IA. Essas vantagens incluem oferecer explicações localizadas e específicas para cada instância, que geralmente são mais intuitivas e fáceis de entender em comparação com explicações em nível de modelo global. Sendo agnóstico em relação ao modelo, permite sua aplicação a qualquer modelo de ML, independentemente da complexidade ou da arquitetura subjacente. O LIME gera explicações que são tipicamente mais simples, rápidas e transparentes do que outros métodos, melhorando assim a acessibilidade para não especialistas e promovendo a confiança nas previsões do modelo.

Identificar as características significativas que influenciam as previsões individuais permite que os pesquisadores entendam quais características direcionam as decisões do modelo. Além disso, fornece insights sobre modelos complexos, como redes neurais profundas, onde os métodos tradicionais de interpretabilidade podem enfrentar desafios. Ao integrar o LIME, os pesquisadores podem melhorar sua capacidade de interpretar e confiar nas previsões de modelos de aprendizado de máquina em pesquisas e descobertas de fármacos, levando a uma tomada de decisão mais informada e potencialmente acelerando a descoberta de novos fármacos.

6 Conclusão

Os modelos de classificação utilizados classificaram com sucesso a toxicidade de compostos orgânicos e permeabilidade da BHE por transporte passivo de compostos orgânicos. Os modelos das florestas aleatórias e as árvores extras foram mais eficientes do que as redes neurais residuais em termos de tempo e generalização para classificar a toxicidade e a permeabilidade da BHE. No entanto, outras redes neurais devem ser construídas para que se obtenham resultados melhores das métricas avaliadas.

O LIME foi utilizado para gerar explicações sobre a permeabilidade de compostos orgânicos na BHE. Essas explicações envolvem as subestruturas que mais contribuem para a permeabilidade da molécula, ou a falta dela, servindo como um guia útil para testes laboratoriais subsequentes. O LIME tem um tempo de execução relativamente rápido, considerando o grande número de compostos no conjunto de dados. Ele gera explicações simples e intuitivas, destacando moléculas e indicando as subestruturas importantes que facilitam a permeação na BHE. Outros métodos de explicação são importantes de serem utilizados pois a explicação de cada método é complementar e muito enriquecedora, no entanto exigem recursos computacionais não disponíveis até o momento.

Compreender essas subestruturas importantes que auxiliam na permeação na BHE é vital para a indústria farmacêutica, para os órgãos regulatórios e para grupos de pesquisa envolvidos na síntese de fármacos. De forma geral, os resultados apresentados confirmam a utilidade de modelos de aprendizagem de máquina interpretáveis na toxicologia computacional, evidenciando que o uso de explicadores como o LIME pode acelerar a identificação de padrões estruturais críticos e apoiar a tomada de decisão tanto na pesquisa acadêmica quanto na indústria farmacêutica.

7 Disponibilidade dos dados e códigos

O código, os dados de entrada e os *scripts* de processamento para o desenvolvimento desta Tese estão disponíveis no GitHub.

- SARAlerts - https://github.com/andresilvapimentel/SARAlerts
- BBBPExplainer - https://github.com/andresilvapimentel/bbbp-explainer

O conjunto de dados de penetração da BHE limpo pelos autores também está disponível no GitHub, juntamente com os demais dados. Os *scripts* de análise de dados desta Tese estão disponíveis no caderno interativo *Google Colab*.

8 Perspectivas

Apesar das contribuições relevantes, aprimoramentos ainda são necessários. Destacam-se a necessidade de desenvolver métodos que reduzam a redundância de fragmentos gerados por explicadores locais e a integração de múltiplas fontes de dados toxicológicos, incluindo informações metabólicas para ampliar a abrangência das predições.

Embora o presente trabalho tenha proporcionado avanços relevantes na interpretação de modelos de aprendizado de máquina aplicados à predição de toxicidade e permeabilidade na BHE, outras oportunidades de aprimoramento permanecem em aberto. Em etapas futuras, pretende-se aprofundar a qualidade e a robustez dos dados utilizados por meio do emprego de conjuntos curados, assegurando maior confiabilidade e consistência estatística dos resultados.

Outro eixo de desenvolvimento envolve a ampliação das representações moleculares consideradas. Além dos SMILES, pretende-se incorporar descritores baseados em SMARTS (EHMKI et al., 2019) e SELFIES (WEININGER, 1988) bem como representações que tratem de forma mais adequada a quiralidade molecular, como o fragSMILES. Essa diversificação de representações químicas visa capturar nuances estruturais que podem influenciar significativamente a acurácia preditiva e interpretabilidade dos modelos.

Do ponto de vista da modelagem, almeja-se trabalhar com bancos de dados mais equilibrados, mitigando o viés de classes e favorecendo o desempenho global dos classificadores. Estratégias adicionais incluem a adoção de modelos mais genéricos e regulares, capazes de reduzir o sobreajuste e melhorar a generalização, bem como a implementação de procedimentos rigorosos de controle de vazamento de dados, de modo a garantir a validade dos resultados obtidos.

Espera-se, com essas melhorias, alcançar métricas de desempenho superiores a 0,90 para acurácia, precisão, sensibilidade, seletividade, F1-score,

coeficiente de correlação de Matthews (MCC) e índice kappa de Cohen para consolidar a confiabilidade dos modelos e sua aplicabilidade em contextos reais.

No campo da interpretabilidade, destaca-se a intenção de empregar outros explicadores, como Anchor e SHAP, cuja implementação completa ainda é limitada pelo alto custo computacional e demanda de memória. No entanto, com o avanço de infraestruturas de processamento gráfico dotadas de maior capacidade de memória RAM, espera-se viabilizar a execução desses métodos no grupo de pesquisa, ampliando a compreensão dos mecanismos subjacentes às predições dos modelos.

Por fim, as metodologias aqui desenvolvidas e aprimoradas serão aplicadas a problemas emergentes de relevância na área biomédica, incluindo toxicologia preditiva, identificação de compostos com potencial antineoplásico e investigação de moléculas com ação moduladora em doenças neurodegenerativas, como o Alzheimer. Tais avanços poderão contribuir não apenas para o fortalecimento do campo de XAI em química medicinal, mas também para a geração de conhecimento químico interpretável de impacto.

9 Referências

Ali, S et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. Information Fusion. v. 99, 2023. https://doi.org/10.1016/j.inffus.2023.101805.

Aljofan, M; Gaipov A. Drug Discovery and Development: The Role of Artificial Intelligence in Drug Repurposing. Future Medicinal Chemistry. v. 16, n. 7, 2024. https://doi.org/10.4155/fmc-2024-0048.

Al-Karmalawy, A et al. Medicinal chemistry perspectives on anticancer drug design based on clinical applications. RSC Advances. v. 15, n. 43, 2025. https://doi.org/10.1039/D5RA05472A.

Alves, V. M et al. Alarms about structural alerts. Green Chemistry. v. 16, 2016. https://doi.org/10.1039/C6GC01492E.

Baell, J.B; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. Journal of Medicinal Chemistry. v. 53, n.7, 2010. https://doi.org/10.1021/jm901137j.

Ballabh, P et al. The blood–brain barrier: an overview. Neurobiology Disease. v. 16, n. 1, 2004, p. 1-13. https://doi.org/10.1016/j.nbd.2003.12.016.

Barr Kumarakulasinghe, N., Blomberg, T., Liu, J., Saraiva Leão, A., Papapetrou, P. Evaluating Local Interpretable Model-Agnostic Explanations on Clinical Machine Learning Classification Models. *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2020, p. 7–12. https://doi.org/10.1109/CBMS49503.2020.00009

Beam, T.R., Allen, J.C. Blood, Brain, and Cerebrospinal Fluid Concentrations of Several Antibiotics in Rabbits with Intact and Inflamed Meninges. *Antimicrob Agents Chemother*, v. 12, 1977, p. 710–716. https://doi.org/10.1128/AAC.12.6.710

Bergstra, J., Yamins, D., Cox, D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In: Dasgupta, S., McAllester, D., editors. *Proceedings of the 30th International Conference on Machine Learning*, Atlanta (USA), 2013, p. 115–123.

Biniashvili, T., Schreiber, E., Kliger, Y. Improving classical substructure-based virtual screening to handle extrapolation challenges. *J Chem Inf Model*, v. 52, 2012, p. 678–685. https://doi.org/10.1021/ci200472s

Bologa, C.G., Ursu, O., Oprea, T.I. How to prepare a compound collection prior to virtual screening. *Methods in Molecular Biology*, vol. 1939, Humana Press Inc., 2019, p. 119–138. https://doi.org/10.1007/978-1-4939-9089-4_7

Bottou, L. Large-Scale Machine Learning with Stochastic Gradient Descent. In: *Proceedings of COMPSTAT'2010*, Heidelberg, Physica-Verlag HD, 2010, p. 177–186. https://doi.org/10.1007/978-3-7908-2604-3_16

Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit*, v. 30, 1997, p. 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Breiman, L. Random Forests. *Mach Learn*, v. 45, 2001a, p. 5–32. https://doi.org/10.1023/A:1010933404324

Brenk, R et al. Lessons Learnt from Assembling Screening Libraries for Drug Discovery for Neglected Diseases. *ChemMedChem*, v. 3, 2008, p. 435–444. https://doi.org/10.1002/cmdc.200700139

Bruner, L.H et al. Validation of Alternative Methods for Toxicity Testing. 1996.

Capecchi, A., Probst, D., Reymond, J.L. One molecular fingerprint to rule them all: Drugs, biomolecules, and the metabolome. *J Cheminform*, v. 12, 2020, p. 1–15. https://doi.org/10.1186/s13321-020-00445-4

Chemaxon. Extended Connectivity Fingerprint (ECFP). https://docs.chemaxon.com/display/docs/Fingerprints_extended-Connectivity-Fingerprint-EcfpMd.

Chen, E.C., Matsson, P., Azimi, M., Zhou, X., Handin, N., Yee, S.W., et al. High Throughput Screening of a Prescription Drug Library for Inhibitors of Organic Cation Transporter 3, OCT3. *Pharm Res*, v. 39, 2022, p. 1599–1613. https://doi.org/10.1007/s11095-022-03171-8

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov Today*, v. 23, 2018, p. 1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039

Chen, Y., Cheng, F., Sun, L., Li, W., Liu, G., Tang, Y. Computational models to predict endocrine-disrupting chemical binding with androgen or oestrogen receptors. *Ecotoxicol Environ Saf*, v. 110, 2014, p. 280–287. https://doi.org/10.1016/j.ecoenv.2014.08.026

Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., et al. QSAR Modeling: Where Have You Been? Where Are You Going To? *J Med Chem*, v. 57, 2014, p. 4977–5010. https://doi.org/10.1021/jm4004285

Clement, T., Kemmerzell, N., Abdelaal, M., Amberg, M. XAIR: A Systematic Metareview of Explainable AI (XAI) Aligned to the Software Development

Process. *Mach Learn Knowl Extr*, v. 5, 2023, p. 78–108. https://doi.org/10.3390/make5010006

Cronin, M.T.D., Enoch, S.J., Mellor, C.L., Przybylak, K.R., Richarz, A.-N., Madden, J.C. In Silico Prediction of Organ Level Toxicity: Linking Chemistry to Adverse Effects. *Toxicol Res*, v. 33, 2017, p. 173–182. https://doi.org/10.5487/TR.2017.33.3.173

Daly, A.K. Pharmacogenomics of adverse drug reactions. *Genome Med*, v. 5, 2013, p. 5. https://doi.org/10.1186/gm409

DANG, N. L.; HUGHES, T. B.; MILLER, G. P.; SWAMIDASS, S. J. *Computational approach to structural alerts: furans, phenols, nitroaromatics, and thiophenes. Chemical Research in Toxicology*, v. 30, p. 1046–1059, 2017. DOI: https://doi.org/10.1021/acs.chemrestox.6b00336.

DEVILLERS, J. *Methods for building QSARs. Methods in Molecular Biology*, v. 930, p. 3–27, 2013. DOI: https://doi.org/10.1007/978-1-62703-059-5_1.

DI, L.; RONG, H.; FENG, B. *Demystifying brain penetration in central nervous system drug discovery. Journal of Medicinal Chemistry*, v. 56, p. 2–12, 2013. DOI: https://doi.org/10.1021/jm301297f.

Didática Tech: Inteligência Artificial & Data Science. Overfitting e Underfitting. Disponível em: https://didatica.tech/underfitting-e-overfitting/

DIMASI, J. A.; GRABOWSKI, H. G.; HANSEN, R. W. *Innovation in the pharmaceutical industry: new estimates of R&D costs. Journal of Health Economics*, v. 47, p. 20–33, 2016. DOI: https://doi.org/10.1016/j.jhealeco.2016.01.012.

DOWEYKO, A. M. *QSAR: dead or alive? Journal of Computer-Aided Molecular Design*, v. 22, p. 81–89, 2008. DOI: https://doi.org/10.1007/s10822-007-9162-7.

DRAFT. *Verbete Draft: o que é Machine Learning.*, 2016. Disponível em: https://www.projetodraft.com/verbete-draft-o-que-e-machine-learning/. Acesso em: 22 jun. 2023.

DURANT, J. L.; LELAND, B. A.; HENRY, D. R.; NOURSE, J. G. *Reoptimization of MDL keys for use in drug discovery. Journal of Chemical Information and Computer Sciences*, v. 42, p. 1273–1280, 2002. DOI: https://doi.org/10.1021/ci010132r.

EHMIK, E. SCHMIDT, R. RAREY, M. Comparing molecular patterns using the example of SMARTS: Applications and filter collection analysis. Journal of Chemical Information and Modeling, v. 59, n.6, p. 2572-2586, 2019. DOI: https://doi.org/10.1021/acs.jcim.9b00249

FAPESP. *Barreira Hematoencefálica.* 2022. Disponível em:

https://revistapesquisa.fapesp.br/wp-content/uploads/2017/06/054-055_barreira_256.jpg.

FELDMANN, C.; BAJORATH, J. *Calculation of exact Shapley values for support vector machines with Tanimoto kernel enables model interpretation. iScience*, v. 25, 105023, 2022. DOI: https://doi.org/10.1016/j.isci.2022.105023.

GANDHI, H. A.; WHITE, A. D. *Explaining molecular properties with natural language. Theoretical and Computational Chemistry*, 2022.

GAULTON, A. et al. *ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Research*, v. 40, 2012. DOI: https://doi.org/10.1093/nar/gkr777.

GAVID, D et al. *rdkit/rdkit: 2022_03_5 (Q1 2022) Release*, 2022. DOI: https://zenodo.org/record/6961488#.YxtZBaDMLcc.

GEURTS, P.; ERNST, D.; WEHENKEL, L. *Extremely randomized trees. Machine Learning*, v. 63, p. 3–42, 2006. DOI: https://doi.org/10.1007/s10994-006-6226-1.

GIORGIO VISANI; ENRICO BAGLI; FEDERICO CHESANI. *OptiLIME: optimized LIME explanations for diagnostic computer algorithms*, 2020.
GITHUB. *mols2grid is an interactive molecule viewer for 2D structures, based on RDKit.* Disponível em: https://zenodo.org/badge/latestdoi/348814588

GOODWIN, J. T.; CLARK, D. E. *In silico predictions of blood-brain barrier penetration: considerations to "keep in mind." Journal of Pharmacology and Experimental Therapeutics*, v. 315, p. 477–483, 2005. DOI: https://doi.org/10.1124/jpet.104.075705.

GUHA, R. *On the interpretation and interpretability of quantitative structure-activity relationship models. Journal of Computer-Aided Molecular Design*, v. 22, p. 857–871, 2008. DOI: https://doi.org/10.1007/s10822-008-9240-5.

GUPTA, M.; LEE, H. J.; BARDEN, C. J.; WEAVER, D. F. *The blood–brain barrier (BBB) score. Journal of Medicinal Chemistry*, v. 62, p. 9824–9836, 2019. DOI: https://doi.org/10.1021/acs.jmedchem.9b01220.

HARRIS, C. R. et al. *Array programming with NumPy. Nature*, v. 585, p. 357–362, 2020. DOI: https://doi.org/10.1038/s41586-020-2649-2.

HARTUNG, I. V.; HUCK, B. R.; CRESPO, A. *Rules were made to be broken. Nature Reviews Chemistry*, v. 7, p. 3–4, 2023. DOI: https://doi.org/10.1038/s41570-022-00451-0.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning.* New York, NY: Springer New York, 2009. DOI: https://doi.org/10.1007/978-0-387-84858-7.

HAWKINS, R. A.; MOKASHI, A.; SIMPSON, I. A. *An active transport system in the blood–brain barrier may reduce levodopa availability. Experimental Neurology*, v. 195, p. 267–271, 2005. DOI: https://doi.org/10.1016/j.expneurol.2005.04.008.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. p. 770-778. DOI: https://doi.org/10.1109/CVPR.2016.90.

HEMMERICH, J.; TROGER, F.; FÜZI, B.; ECKER, G. F. *Using machine learning methods and structural alerts for prediction of mitochondrial toxicity. Molecular Informatics*, v. 39, 2020. DOI: https://doi.org/10.1002/minf.202000005.

HERNANDEZ, C. C.; GIMENEZ, L. E.; STRASSMAIER, T.; ROGERS, M.; TAYMANS, J.-M. *Targeting ion channels for drug discovery: emerging challenges for high throughput screening technologies. Frontiers in Molecular Neuroscience*, v. 17, 2024. DOI: https://doi.org/10.3389/fnmol.2024.1414816.

HONG, Y.; ZHOU, Y.; WANG, J.; LIU, H. *Lead compound optimization strategy: improving blood-brain barrier permeability through structural modification. Acta Pharmaceutica Sinica*, v. 49, p. 789–799, 2014.

HUA, Y. et al. *SApredictor: an expert system for screening chemicals against structural alerts. Frontiers in Chemistry*, v. 10, 2022. DOI: https://doi.org/10.3389/fchem.2022.916614.

HUANG, Y. et al. *Role of gut microecology in the pathogenesis of drug-induced liver injury and emerging therapeutic strategies. Molecules*, v. 29, 2663, 2024. DOI: https://doi.org/10.3390/molecules29112663.

HUNTER, J. D. *MATPLOTLIB: a 2D graphics environment. Computing in Science & Engineering*, v. 9, p. 90–95, 2007.

JAIN, A. K. et al. *Models and methods for in vitro toxicity.* In: *In Vitro Toxicology*. Elsevier, 2018, p. 45–65. DOI: https://doi.org/10.1016/B978-0-12-804667-8.00003-1.

JEFF REBACK et al. *pandas-dev/pandas: pandas 1.4.4*, 2022. DOI: https://doi.org/10.5281/zenodo.3509134.

JIMÉNEZ-LUNA, J.; GRISONI, F.; SCHNEIDER, G. *Drug discovery with explainable artificial intelligence. Nature Machine Intelligence*, v. 2, p. 573–584, 2020. DOI: https://doi.org/10.1038/s42256-020-00236-4.

JOHN MC CARTHY et al. *A proposal for the Dartmouth summer research project on artificial intelligence. AI Magazine*, 2015.

JOHNSEN, P. V. et al. *A new method for exploring gene–gene and gene–environment interactions in GWAS with tree ensemble methods and SHAP values. BMC Bioinformatics*, v. 22, 230, 2021. DOI: https://doi.org/10.1186/s12859-021-04041-7

HE, K.; ZHANG, X.; REN, S.; SUN, J. *Identity mappings in deep residual networks.* 2016.

KALGUTKAR, A. S. *Designing around structural alerts in drug discovery. Journal of Medicinal Chemistry*, v. 63, p. 6276–6302, 2020. DOI: https://doi.org/10.1021/acs.jmedchem.9b00917.

KAR, S.; ROY, K.; LESZCZYNSKI, J. *In silico tools and software to predict ADMET of new drug candidates.* 2022, p. 85–115. DOI: https://doi.org/10.1007/978-1-0716-1960-5_4.

KOLESAROVA, M.; SIMKO, P.; URBANSKA, N.; KISKOVA, T. *Exploring the potential of cannabinoid nanodelivery systems for CNS disorders. Pharmaceutics*, v. 15, 204, 2023. DOI: https://doi.org/10.3390/pharmaceutics15010204.

KUHN, M.; LETUNIC, I.; JENSEN, L. J.; BORK, P. *The SIDER database of drugs and side effects. Nucleic Acids Research*, v. 44, D1075–D1079, 2016. DOI: https://doi.org/10.1093/nar/gkv1075.

LAVECCHIA, A. *Machine-learning approaches in drug discovery: methods and applications. Drug Discovery Today*, v. 20, p. 318–331, 2015. DOI: https://doi.org/10.1016/j.drudis.2014.10.012.

LECUN, Y.; BENGIO, Y.; HINTON, G. *Deep learning. Nature*, v. 521, p. 436–444, 2015. DOI: https://doi.org/10.1038/nature14539.

LEDEL, B.; HERBOLD, S. *Studying the explanations for the automated prediction of bug and non-bug issues using LIME and SHAP.* 2022. DOI: https://doi.org/10.48550/arxiv.2209.07623.

LEI, T. et al. *ADMET evaluation in drug discovery: accurate prediction of rat oral acute toxicity using relevance vector machine and consensus modeling. Journal of Cheminformatics*, v. 8, 6, 2016. DOI: https://doi.org/10.1186/s13321-016-0117-7.

LEPAILLEUR, A.; POEZEVARA, G.; BUREAU, R. *Automated detection of structural alerts (chemical fragments) in (eco)toxicology. Computational and Structural Biotechnology Journal*, v. 5, e201302013, 2013. DOI: https://doi.org/10.5936/csbj.201302013.

LI, H.; YAP, C. W.; UNG, C. Y.; XUE, Y.; CAO, Z. W.; CHEN, Y. Z. *Effect of selection of molecular descriptors on the prediction of blood–brain barrier penetrating and nonpenetrating agents by statistical learning methods. Journal of Chemical Information and Modeling*, v. 45, p. 1376–1384, 2005. DOI: https://doi.org/10.1021/ci050135u.

LI, X. et al. *In silico prediction of chemical acute oral toxicity using multi-classification methods. Journal of Chemical Information and Modeling*, v. 54, p. 1061–1069, 2014. DOI: https://doi.org/10.1021/ci5000467.

LIMA, M.; LAGO, T.; DINIZ, D.; SEIXAS, M.; MACIEL, A. C.; OLIVEIRA-FILHO, J. *Caracterização dos ensaios clínicos conduzidos por um centro de pesquisa clínica em um hospital universitário em Salvador Bahia. Jornal de Assistência Farmacêutica e Farmacoeconomia*, v. 1, 2023. DOI: https://doi.org/10.22563/2525-7323.2022.v1.s2.p.46

LIPINSKI, C. A. *Lead- and drug-like compounds: The rule-of-five revolution. Drug Discovery Today: Technologies*, v. 1, p. 337–341, 2004. DOI: https://doi.org/10.1016/j.ddtec.2004.11.007.

LIPINSKI, C. A.; LOMBARDO, F.; DOMINY, B. W.; FEENEY, P. J. *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* vol. 46, 2001.

LIU, L. et al. *Prediction of the blood–brain barrier (BBB) permeability of chemicals based on machine-learning and ensemble methods. Chemical Research in Toxicology*, v. 34, p. 1456–1467, 2021. DOI: https://doi.org/10.1021/acs.chemrestox.0c00343.

LYSENKO, A.; SHARMA, A.; BOROEVICH, K. A.; TSUNODA, T. *An integrative machine learning approach for prediction of toxicity-related drug safety. Life Science Alliance*, v. 1, 2018. DOI: https://doi.org/10.26508/lsa.201800098.

MAHMOUD, A. *A survey of computational toxicology approaches.* Journal of Information Sciences, 2021. DOI: https://doi.org/10.21608/kjis.2021.41013.1008.

MAIR, L. O. et al. *Soft capsule magnetic millirobots for region-specific drug delivery in the central nervous system. Frontiers in Robotics and AI*, v. 8, 2021. DOI: https://doi.org/10.3389/frobt.2021.702566.

MARCHANT, C. A. *Computational toxicology: A tool for all industries. Wiley Interdisciplinary Reviews: Computational Molecular Science*, v. 2, p. 424–434, 2012. DOI: https://doi.org/10.1002/wcms.100.

MARCO TULIO RIBEIRO; SAMEER SINGH; CARLOS GUESTRIN. *Local interpretable model-agnostic explanations (LIME): An introduction.* 2016. Disponível em: https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/. Acesso em: 22 jun. 2023.

MARTINO, Jarrier Andrade, Algoritmos evolutivos como método para desenvolvimento de projetos de arquitetura, 2015. DOI: 10.13140/RG.2.2.21169.38240

MARTINS, I. F.; TEIXEIRA, A. L.; PINHEIRO, L.; FALCAO, A. O. *A Bayesian approach to in silico blood-brain barrier penetration modeling. Journal of Chemical Information and Modeling*, v. 52, p. 1686–1697, 2012. DOI: https://doi.org/10.1021/ci300124c.

MASTROPIETRO, A.; PASCULLI, G.; BAJORATH, J. *Protocol to explain graph neural network predictions using an edge-centric Shapley value-based approach. STAR Protocols*, v. 3, 101887, 2022a. DOI: https://doi.org/10.1016/j.xpro.2022.101887.

MASTROPIETRO, A.; PASCULLI, G.; FELDMANN, C.; RODRÍGUEZ-PÉREZ, R.; BAJORATH, J. *EdgeSHAPer: Bond-centric Shapley value-based explanation method for graph neural networks. iScience*, v. 25, 105043, 2022b. DOI: https://doi.org/10.1016/j.isci.2022.105043.

MITCHELL, T. M. *Machine Learning.* 1997.

MORGER, A. et al. *KnowTox: Pipeline and case study for confident prediction of potential toxic effects of compounds in early phases of development. Journal of Cheminformatics*, v. 12, 2020. DOI: https://doi.org/10.1186/s13321-020-00422-x.

NetApp – Explainer AI: What is it? How does it works? And what role does data play? Disponível em: https://www.netapp.com/blog/explainable-ai/

NIPPA, D. F.; BODDY, A. J.; ATZ, K.; GRETHER, U.; BINCH, H.; MARTIN, R. E. Accelerating compound synthesis in drug discovery: the role of digitalization and automation. RSC Medicinal Chemistry, 2025. DOI: https://doi.org/10.1039/D5MD00672D

NASCIMENTO, C. M. C.; MOURA, P. G.; PIMENTEL, A. S. *Generating structural alerts from toxicology datasets using the local interpretable model-*

agnostic explanations method. Digital Discovery, v. 2, p. 1311–1325, 2023. DOI: https://doi.org/10.1039/D2DD00136E.

NAU, R.; SÖRGEL, F.; EIFFERT, H. *Penetration of drugs through the blood-cerebrospinal fluid/blood-brain barrier for treatment of central nervous system infections. Clinical Microbiology Reviews*, v. 23, p. 858–883, 2010. DOI: https://doi.org/10.1128/CMR.00007-10.

OPREA, T. I. *Property distribution of drug-related chemical databases.* vol. 14, KLUWER/ESCOM, 2000.

PATCHING, S. G. *Glucose transporters at the blood-brain barrier: Function, regulation and gateways for drug delivery. Molecular Neurobiology*, v. 54, p. 1046–1077, 2017. DOI: https://doi.org/10.1007/s12035-015-9672-6.

PEDREGOSA, F. et al. *Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

PEREZ, F.; GRANGER, B. E. *IPython: A system for interactive scientific computing. Computing in Science & Engineering*, v. 9, p. 21–29, 2007. DOI: https://doi.org/10.1109/MCSE.2007.53.

PURIS, E.; GYNTHER, M.; HUTTUNEN, J.; PETSALO, A.; HUTTUNEN, K. M. *L-type amino acid transporter 1 utilizing prodrugs: How to achieve effective brain delivery and low systemic exposure of drugs. Journal of Controlled Release*, v. 261, p. 93–104, 2017. DOI: https://doi.org/10.1016/j.jconrel.2017.06.023.

QIN, T.; CHANDRASEKARAN, A.; SHIERS, J.; CRITTALL, M.; O´REILLY, C.; SMITH, C. S. Digitalizing the design-make-test-analyze workflow in drug discovery with an electronic inventory platform, SLAS Technology, v. 32, 2025. DOI: https://doi.org/10.1016/j.slast.2025.100262

QUINLAN, J. R. *Induction of decision trees. Machine Learning*, v. 1, p. 81–106, 1986. DOI: https://doi.org/10.1007/BF00116251.

RÁCZ, A.; BAJUSZ, D.; HÉBERGER, K. *Life beyond the Tanimoto coefficient: Similarity measures for interaction fingerprints. Journal of Cheminformatics*, v. 10, p. 1–12, 2018. DOI: https://doi.org/10.1186/s13321-018-0302-y.

RAIES, A. B.; BAJIC, V. B. *In silico toxicology: Computational methods for the prediction of chemical toxicity. Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2016, p. 147–172. DOI: https://doi.org/10.1002/wcms.1240.

REVISTA CÉREBRO & MENTE. *O que são redes neurais artificiais.* 1998. Disponível em: https://cerebromente.org.br/n05/tecnologia/rna.htm.

EL RHABORI, S. et al. *Integrative computational strategy for anticancer drug discovery: QSAR-ANN modeling, molecular docking, ADMET prediction, molecular dynamics and MM-PBSA simulations, and retrosynthetic analysis.*

New Journal of Chemistry, v. 49, p. 14748–14768, 2025. DOI: https://doi.org/10.1039/D5NJ02417J.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. *Anchors: High-Precision Model-Agnostic Explanations.* Proceedings of the AAAI Conference on Artificial Intelligence, v. 32, 2018. DOI: https://doi.org/10.1609/aaai.v32i1.11491.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. *"Why should I trust you?" Explaining the predictions of any classifier.* Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, p. 1135–1144. DOI: https://doi.org/10.1145/2939672.2939778.

RICHARD, A. M. et al. *The Tox21 10K Compound Library: Collaborative Chemistry Advancing Toxicology. Chemical Research in Toxicology*, v. 34, p. 189–216, 2021. DOI: https://doi.org/10.1021/acs.chemrestox.0c00264.

RICHARD, A. M. et al. *ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. Chemical Research in Toxicology*, v. 29, p. 1225–1251, 2016. DOI: https://doi.org/10.1021/acs.chemrestox.6b00135.

RIDINGS, J. et al. *Computer prediction of possible toxic action from chemical structure: An update on the DEREK system.* vol. 106, ELSEVIER, 1996.

RODRÍGUEZ-PÉREZ, R.; BAJORATH, J. *Explainable Machine Learning for Property Predictions in Compound Optimization. Journal of Medicinal Chemistry*, v. 64, p. 17744–17752, 2021. DOI: https://doi.org/10.1021/acs.jmedchem.1c01789.

RODRÍGUEZ-PÉREZ, R.; BAJORATH, J. *Interpretation of compound activity predictions from complex machine learning models using local approximations and Shapley values. Journal of Medicinal Chemistry*, v. 63, p. 8761–8777, 2020a. DOI: https://doi.org/10.1021/acs.jmedchem.9b01101.

RODRÍGUEZ-PÉREZ, R.; BAJORATH, J. *Interpretation of machine learning models using Shapley values: Application to compound potency and multi-target activity predictions. Journal of Computer-Aided Molecular Design*, v. 34, p. 1013–1026, 2020b. DOI: https://doi.org/10.1007/s10822-020-00314-0.

ROGERS, D.; HAHN, M. *Extended-connectivity fingerprints. Journal of Chemical Information and Modeling*, v. 50, p. 742–754, 2010. DOI: https://doi.org/10.1021/ci100050t.

ROSA, L. C. S.; ARGOLO, C. O.; NASCIMENTO, C. M. C.; PIMENTEL, A. S. *Identifying substructures that facilitate compounds to penetrate the blood–brain barrier via passive transport using machine learning explainer models. ACS Chemical Neuroscience*, v. 15, p. 2144–2159, 2024. DOI: https://doi.org/10.1021/acschemneuro.3c00840.

ROSA, L. C. S.; PIMENTEL, A. S. *Applying local interpretable model-agnostic explanations to identify substructures that are responsible for mutagenicity of*

chemical compounds. Molecular Systems Design & Engineering, v. 9, p. 920–936, 2024. DOI: https://doi.org/10.1039/D4ME00038B.

ROSA, L. C. S.; SARHAN, M.; PIMENTEL, A. S. *Toxic Alerts of Endocrine Disruption Revealed by Explainable Artificial Intelligence. Environment & Health*, v. 3, p. 321–333, 2025. DOI: https://doi.org/10.1021/envhealth.4c00218.

ROY, R.; SINGH, S. K.; AHMAD, N.; MISRA, S. *Challenges and advancements in high-throughput screening strategies for cancer therapeutics. Global Translational Medicine*, v. 3, 2448, 2024. DOI: https://doi.org/10.36922/gtm.2448.

RUSYN, I.; DASTON, G. P. *Computational toxicology: Realizing the promise of the toxicity testing in the 21st century. Environmental Health Perspectives*, v. 118, p. 1047–1050, 2010. DOI: https://doi.org/10.1289/ehp.1001925.

SCHIFF, S. J. *Neural Control Engineering.* The MIT Press, 2011. DOI: https://doi.org/10.7551/mitpress/8436.001.0001.

SCHMIDHUBER, J. *Deep learning in neural networks: An overview. Neural Networks*, v. 61, p. 85–117, 2015. DOI: https://doi.org/10.1016/j.neunet.2014.09.003.

SCHONLAU, M.; ZOU, R. Y. *The random forest algorithm for statistical learning. The Stata Journal*, v. 20, p. 3–29, 2020. DOI: https://doi.org/10.1177/1536867X20909688

SCHREIBELT, G. et al. *Lipoic Acid Affects Cellular Migration into the Central Nervous System and Stabilizes Blood-Brain Barrier Integrity. The Journal of Immunology*, v. 177, p. 2630–2637, 2006. DOI: https://doi.org/10.4049/jimmunol.177.4.2630.

SCHWALLER, P. et al. *Mapping the space of chemical reactions using attention-based neural networks. Nature Machine Intelligence*, v. 3, p. 144–152, 2021. DOI: https://doi.org/10.1038/s42256-020-00284-w.

SHERHOD, R. et al. *Automating knowledge discovery for toxicity prediction using jumping emerging pattern mining. Journal of Chemical Information and Modeling*, v. 52, p. 3074–3087, 2012. DOI: https://doi.org/10.1021/ci300254w.

SHERHOD, R. et al. *Emerging pattern mining to aid toxicological knowledge discovery. Journal of Chemical Information and Modeling*, v. 54, p. 1864–1879, 2014. DOI: https://doi.org/10.1021/ci5001828.

SIMOENS, S.; HUYS, I. *Costs of New Medicines: A Landscape Analysis. Frontiers in Medicine*, v. 8, 760762, 2021. DOI: https://doi.org/10.3389/fmed.2021.760762.

SINGH, M. et al. *A classification model for blood brain barrier penetration. Journal of Molecular Graphics and Modelling*, v. 96, 107516, 2020. DOI: https://doi.org/10.1016/j.jmgm.2019.107516.

SUGAWARA, E.; NIKAIDO, H. *Deep Learning for the Life Sciences.* v. 58, 2014.

SUSHKO, I. et al. *ToxAlerts: A web server of structural alerts for toxic chemicals and compounds with potential adverse reactions. Journal of Chemical Information and Modeling*, v. 52, p. 2310–2316, 2012. DOI: https://doi.org/10.1021/ci300245q.

TASNEEM, A. et al. *The database for aggregate analysis of Clinicaltrials.gov (AACT) and subsequent regrouping by clinical specialty. PLoS ONE*, v. 7, e33677, 2012. DOI: https://doi.org/10.1371/journal.pone.0033677.

TONG, X. et al. *Blood–brain barrier penetration prediction enhanced by uncertainty estimation. Journal of Cheminformatics*, v. 14, 44, 2022. DOI: https://doi.org/10.1186/s13321-022-00619-2.

URSU, O.; RAYAN, A.; GOLDBLUM, A.; OPREA, T. I. *Understanding drug-likeness. Wiley Interdisciplinary Reviews: Computational Molecular Science*, v. 1, p. 760–781, 2011. DOI: https://doi.org/10.1002/wcms.52.

U.S. FDA. *FDA Announces Plan to Phase Out Animal Testing Requirement for Monoclonal Antibodies and Other Drugs.* 2025. Disponível em: https://www.fda.gov/news-events/press-announcements/fda-announces-plan-phase-out-animal-testing-requirement-monoclonal-antibodies-and-other-drugs.

VAMATHEVAN, J. et al. *Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery*, v. 18, p. 463–477, 2019. DOI: https://doi.org/10.1038/s41573-019-0024-5.

VARNEK, A.; BASKIN, I. *Machine learning methods for property prediction in chemoinformatics: Quo Vadis? Journal of Chemical Information and Modeling*, v. 52, p. 1413–1437, 2012. DOI: https://doi.org/10.1021/ci200409x.

VASILEV, B.; ATANASOVA, M. *A (Comprehensive) Review of the Application of Quantitative Structure–Activity Relationship (QSAR) in the Prediction of New Compounds with Anti-Breast Cancer Activity. Applied Sciences*, v. 15, 1206, 2025. DOI: https://doi.org/10.3390/app15031206.

VON KORFF, M.; SANDER, T. *Toxicity-indicating structural patterns. Journal of Chemical Information and Modeling*, v. 46, p. 536–544, 2006. DOI: https://doi.org/10.1021/ci050358k.

WANG, Q.; WU, Y.; CHEN, B.; ZHOU, J. *Drug concentrations in the serum and cerebrospinal fluid of patients treated with cefoperazone/sulbactam after craniotomy. BMC Anesthesiology*, v. 15, 33, 2015. DOI: https://doi.org/10.1186/s12871-015-0012-1.

WANG, Q.; ZHANG, Y.; ZHU, A. *Drug Metabolism and Toxicological Mechanisms. Toxics*, v. 13, 464, 2025. DOI: https://doi.org/10.3390/toxics13060464.

WEININGER, D. *SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. Journal of Chemical Information and Computer Sciences*, v. 28, p. 31–36, 1988. DOI: https://doi.org/10.1021/ci00057a005.

WELLAWATTE, G. P.; GANDHI, H. A.; SESHADRI, A.; WHITE, A. D. *A Perspective on Explanations of Molecular Prediction Models. Theoretical and Computational Chemistry*, 2022a.

WELLAWATTE, G. P.; SESHADRI, A.; WHITE, A. D. *Model agnostic generation of counterfactual explanations for molecules. Chemical Science*, v. 13, p. 3697–3705, 2022b. DOI: https://doi.org/10.1039/D1SC05259D.

WELLMAN, M. P. *The economic approach to artificial intelligence. ACM Computing Surveys*, v. 27, p. 360–362, 1995. DOI: https://doi.org/10.1145/212094.212128.

WOJTUCH, A.; JANKOWSKI, R.; PODLEWSKA, S. *How can SHAP values help to shape metabolic stability of chemical compounds? Journal of Cheminformatics*, v. 13, 74, 2021. DOI: https://doi.org/10.1186/s13321-021-00542-y.

WOUTERS, O. J.; MCKEE, M.; LUYTEN, J. *Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. JAMA*, v. 323, p. 844, 2020. DOI: https://doi.org/10.1001/jama.2020.1166.

WU, D. et al. *The blood–brain barrier: structure, regulation, and drug delivery. Signal Transduction and Targeted Therapy*, v. 8, 217, 2023. DOI: https://doi.org/10.1038/s41392-023-01481-w.

WU, Z. et al. *From Black Boxes to Actionable Insights: A Perspective on Explainable Artificial Intelligence for Scientific Discovery. Journal of Chemical Information and Modeling*, 2023. DOI: https://doi.org/10.1021/acs.jcim.3c01642.

WU, Z. et al. *MoleculeNet: A benchmark for molecular machine learning. Chemical Science*, v. 9, p. 513–530, 2018. DOI: https://doi.org/10.1039/c7sc02664a.

YANG, H.; LIU, G.; TANG, Y. *Evaluation of Different Methods for Identification of Structural Alerts Using Chemical Ames Mutagenicity Data Set as a Benchmark. Chemical Research in Toxicology*, v. 30, p. 1355–1364, 2017. DOI: https://doi.org/10.1021/acs.chemrestox.7b00083.

YANG, H.; SUN, L.; LI, W.; LIU, G.; TANG, Y. *Computational Approaches to Identify Structural Alerts and Their Applications in Environmental Toxicology*

and Drug Discovery. Chemical Research in Toxicology, v. 33, p. 1312–1322, 2020. DOI: https://doi.org/10.1021/acs.chemrestox.0c00006.

YANG, H.; SUN, L.; LI, W.; LIU, G.; TANG, Y. *In Silico Prediction of Chemical Toxicity for Drug Design Using Machine Learning Methods and Structural Alerts. Frontiers in Chemistry*, v. 6, 30, 2018. DOI: https://doi.org/10.3389/fchem.2018.00030.

YANG, X.; WANG, Y., BYRNE, R., SCHNEIDER, G., YANG, S. Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. Chemical Reviews, v. 119, n. 18, 2019. DOI: https://doi.org/10.1021/acs.chemrev.8b00728

ZAFAR, M. R.; KHAN, N. *Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability. Machine Learning Knowledge Extraction*, v. 3, p. 525–541, 2021. DOI

Zafar MR, Khan N. Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability. Mach Learn Knowl Extr 2021;3:525–41. https://doi.org/10.3390/make3030027.

Zhao YH, Abraham MH, Ibrahim A, Fish P V., Cole S, Lewis ML, et al. Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes. J Chem Inf Model 2007;47:170–5. https://doi.org/10.1021/ci600312d.

Zipser BD, Johanson CE, Gonzalez L, Berzin TM, Tavares R, Hulette CM, et al. Microvascular injury and blood–brain barrier leakage in Alzheimer's disease. Neurobiol Aging 2007;28:977–86. https://doi.org/10.1016/j.neurobiolaging.2006.05.016.

Anexo: Capas dos artigos publicados no período do Doutorado

Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Digital Discovery*, 2023, 2, 1311

Generating structural alerts from toxicology datasets using the local interpretable model-agnostic explanations method†

Cayque Monteiro Castro Nascimento, Paloma Guimarães Moura and Andre Silva Pimentel *

Received 9th December 2022
Accepted 21st July 2023

DOI: 10.1039/d2dd00136e

rsc.li/digitaldiscovery

The local interpretable model-agnostic explanations method was used to interpret a machine learning model of toxicology generated by a neural network multitask classifier method. The model was trained and validated using the Tox21 dataset and tested against the Clintox and Sider datasets, which are datasets of marketed drugs with adverse reactions and drugs approved by the Federal Drug Administration that have failed clinical trials for toxicity reasons. The stability of the explanations is proved here with a reasonable reproducibility of the sampling process, making very similar and trustful explanations. The explanation model was created to produce structural alerts with more than 6 heavy atoms that serve as toxic alerts for researchers in many fields of academics, regulatory agencies and industry such as organic synthesis, pharmaceuticals, toxicology, and so on.

ACS Chemical Neuroscience

pubs.acs.org/chemneuro

Research Article

Identifying Substructures That Facilitate Compounds to Penetrate the Blood–Brain Barrier via Passive Transport Using Machine Learning Explainer Models

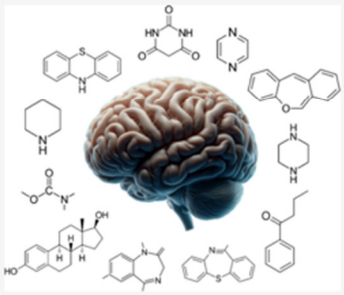
Lucca Caiaffa Santos Rosa, Caio Oliveira Argolo, Cayque Monteiro Castro Nascimento, and Andre Silva Pimentel*

Cite This: *ACS Chem. Neurosci.* 2024, 15, 2144–2159

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The local interpretable model-agnostic explanation (LIME) method was used to interpret two machine learning models of compounds penetrating the blood–brain barrier. The classification models, Random Forest, ExtraTrees, and Deep Residual Network, were trained and validated using the blood–brain barrier penetration dataset, which shows the penetrability of compounds in the blood–brain barrier. LIME was able to create explanations for such penetrability, highlighting the most important substructures of molecules that affect drug penetration in the barrier. The simple and intuitive outputs prove the applicability of this explainable model to interpreting the permeability of compounds across the blood–brain barrier in terms of molecular features. LIME explanations were filtered with a weight equal to or greater than 0.1 to obtain only the most relevant explanations. The results showed several structures that are important for blood–brain barrier penetration. In general, it was found that some compounds with nitrogenous substructures are more likely to permeate the blood–brain barrier. The application of these structural explanations may help the pharmaceutical industry and potential drug synthesis research groups to synthesize active molecules more rationally.



KEYWORDS: *explainable artificial intelligence, structural alerts, LIME, XAI, XML*

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Viewpoint

Do Large Language Models Understand Chemistry? A Conversation with ChatGPT

Cayque Monteiro Castro Nascimento and André Silva Pimentel*

Cite This: J. Chem. Inf. Model. 2023, 63, 1649−1655 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Large language models (LLMs) have promised a revolution in answering complex questions using the ChatGPT model. Its application in chemistry is still in its infancy. This viewpoint addresses the question of how well ChatGPT understands chemistry by posing five simple tasks in different subareas of chemistry.