



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

THAIS TONDATO ALBERTINI

**APLICAÇÃO DE REDES NEURAIIS ARTIFICIAIS PARA
A PREDIÇÃO DA ATIVIDADE BIOLÓGICA DE
COMPOSTOS VISANDO CONTRIBUIÇÃO PARA A
FARMACOLOGIA: REVISÃO BILIOGRÁFICA**

ASSIS

2022

THAIS TONDATO ALBERTINI

**APLICAÇÃO DE REDES NEURAS ARTIFICIAIS PARA
A PREDIÇÃO DA ATIVIDADE BIOLÓGICA DE
COMPOSTOS VISANDO CONTRIBUIÇÃO PARA A
FARMACOLOGIA: REVISÃO BILIOGRÁFICA**

Orientador: Dr. José Celso Rocha

Trabalho de Conclusão de Curso
apresentado à Faculdade de
Ciências e Letras da Universidade
Estadual Paulista “Júlio de
Mesquita Filho” – Campus de
Assis para obtenção do grau de
Bacharel em Engenharia
Biotecnológica.

ASSIS

2022

RESUMO

As Redes Neurais Artificiais (RNAs) possuem destaque, atualmente, ao serem utilizadas para classificação, agrupamento, reconhecimento e predição de padrões, possuindo diversas aplicações em áreas diferentes como reconhecimento de voz, reconhecimento facial, diagnóstico de doenças, previsões do tempo, entre outras. São técnicas computacionais inspiradas no funcionamento do cérebro humano, que apresentam como objetivo principal a aprendizagem de tarefas buscando solucionar problemas complexos, a partir de propriedades únicas, incluindo alta tolerância a falhas e capacidade de generalizar a partir de dados de treinamento. No âmbito da predição de padrões, quando sozinhas ou associadas a outras técnicas de Inteligência Artificial (IA), estudos mostram estque as RNAs são assertivas ao preverem a atividade biológica de compostos, facilitando nos testes laboratoriais para a descoberta de novos medicamentos no ramo da farmacologia. Dessa forma, o presente projeto tem como objetivo realizar uma revisão de trabalhos de aplicação de RNAs, associadas ou não a outras técnicas de IA, que visam a predição da atividade biológica dos mais variados compostos, com o intuito de ressaltar os benefícios trazidos por essas técnicas quanto a estes estudos.

Palavras-chave: modelos preditivos; redes neurais artificiais; atividade biológica de compostos; farmacologia.

ABSTRACT

Artificial Neural Networks (ANNs) are currently highlighted as they are used for classification, grouping, recognition and prediction of patterns, having several applications in different areas such as voice recognition, facial recognition, disease diagnosis, weather forecasts, among others. These are computational techniques inspired by the functioning of the human brain, whose main objective is the learning of tasks seeking to solve complex problems, from unique properties, including high fault tolerance and the ability to generalize from training data. In the scope of pattern prediction, when alone or associated with other Artificial Intelligence (AI) techniques, studies show that ANNs are assertive in predicting the biological activity of compounds, facilitating laboratory tests for the discovery of new drugs in the field of pharmacology. In this way, the present project aims to carry out a review of works on the application of RNAs, associated or not to other AI techniques, aimed at predicting the biological activity of the most varied compounds, in order to highlight the benefits brought by these techniques for these studies.

Keywords: predictive models; artificial neural networks; biological activity of compounds; pharmacology.

SUMÁRIO

1. Introdução.....	6
2. Objetivos.....	7
2.1. Objetivos gerais.....	7
2.2. Objetivos específicos.....	7
3. Metodologia.....	8
3.1. Critérios para pesquisa.....	8
3.2. Critérios para a seleção de artigos.....	8
3.3. Seleção dos estudos.....	8
3.4. Extração dos dados.....	9
4. Resultados.....	9
4.1. Artigos com aplicabilidade específica em RNAs.....	13
4.2. Artigos com aplicabilidade em RNAs e Algoritmos genéticos.....	18
4.3. Artigos com aplicabilidade em RNAs e técnicas de Regressão.....	19
4.4. Artigos com aplicabilidade em RNAs e outras técnicas variadas.....	22
5. Discussão.....	23
6. Conclusão.....	24
7. Referências.....	24

1 INTRODUÇÃO

A descoberta de novos medicamentos traz sempre grandes desafios, custos, além de ser um investimento arriscado (Heifetz, 2020). Em um artigo cuja pesquisa é recente, os autores estimaram, por meio de dados disponíveis relacionados a pesquisa e desenvolvimento de novos agentes terapêuticos, que o custo médio para o desenvolvimento de um novo medicamento pode variar de US\$ 314 milhões a US\$ 2,8 bilhões, levando uma década ou mais até a finalização de todos os testes para enfim a sua comercialização (Wouters *et al.*, 2020). Em outra pesquisa, os autores afirmam que as maiores empresas farmacêuticas custeiam quase US\$ 80 bilhões por ano para produzir um medicamento, cujas chances de falhas em eficiência e segurança são grandes (Freedman, 2019).

Uma das justificativas para esta problemática seria que, em tese, os medicamentos mais fáceis de serem desenvolvidos e que são considerados eficientes e seguros para o tratamento de distúrbios comuns já foram encontrados, restando a busca por medicamentos que abordam problemas com soluções complexas e que são direcionados para o tratamento de doenças específicas, que abrangem apenas pequenas porções da população, o que gera um custo muito elevado (Freedman, 2019). Principalmente na fase clínica dos testes, estágio avançado destas pesquisas, tais gastos são extremamente altos e as chances de falhas são comuns (Heifetz, 2020).

Em casos de medicamentos naturais, os quais possuem grande contribuição para o tratamento de cânceres e doenças infecciosas, inúmeros são os desafios quanto as suas descobertas devido a barreiras técnicas para triagem de compostos, isolamento, caracterização\ identificação e otimização, o que levou a uma diminuição em suas buscas, pela indústria farmacêutica, a partir da década de 1990, sendo necessário encontrar novas ferramentas que facilitem tais estudos e que tragam novas oportunidades (Atanasov *et al.*, 2021).

Atualmente, entre as alternativas visando reduzir os custos e o tempo das fases de pesquisa de novos medicamentos, está o aprendizado de máquinas (em inglês, *machine learning*), ramo da inteligência artificial (IA). A IA é uma tecnologia cuja aplicação é encontrada em diversas áreas e processos objetivando a automatização de problemas complexos (Wirtz *et al.*, 2019). Entre as técnicas se destacam as redes neurais artificiais (RNAs), utilizadas para classificação, agrupamento, reconhecimento e predição de padrões, e

que possuem propriedades únicas, incluindo alta tolerância a falhas e capacidade de generalizar resultados a partir de dados iniciais de treinamento (Abiodun *et al.*, 2018; Bertolaccini *et al.*, 2017). As RNAs são técnicas computacionais inspiradas no funcionamento do cérebro humano e que possuem como objetivo principal a aprendizagem de tarefas que não podem ser solucionadas por humanos ou estratégias computacionais tradicionais (Davenport & Kalakota, 2019; Dwivedi *et al.*, 2021).

A IA, no cenário de design de medicamentos, inclui diversas técnicas diferentes, mas que possuem em comum algoritmos com alta performance de resultados e capacidade de acúmulo de informações biológicas e químicas importantes para o desenvolvimento de novos tratamentos ou curas (Tripathi *et al.*, 2021). Este último ocorre por meio do treinamento de dados para a predição de compostos experimentais, aumentando a eficiência dos testes farmacológicos voltados a pesquisa e desenvolvimento (Heifetz, 2020) e, acelerando, assim, tanto o entendimento da atividade de compostos conhecidos, quanto a identificação e testagem de novos compostos que podem ainda serem sintetizados (Campos; Vasconcelos, 2021).

Diante do contexto apresentado, este trabalho teve por objetivo a realização de um levantamento de artigos que aplicassem técnicas de IA incluindo RNAs, para a predição de compostos, naturais ou não, com propriedades farmacológicas. Nesse sentido, buscou-se traçar um panorama geral quanto as finalidades para tais aplicações, as técnicas utilizadas em conjunto com as RNAs, além de evidenciar as vantagens da utilização do aprendizado de máquinas para o desenvolvimento de novos medicamentos.

2 OBJETIVOS

2.1 Objetivo geral

Esse trabalho teve como objetivo geral realizar uma revisão bibliográfica quanto a aplicação de RNAs, associadas ou não a outras técnicas de IA, que visam a predição da atividade biológica/farmacológica dos mais variados compostos, naturais ou sintéticos.

2.2 Objetivos específicos

Os objetivos específicos do presente trabalho foram:

- Entender a temática específica dos estudos selecionados;

- Determinar as técnicas de IA utilizadas nesses estudos na predição de compostos para a farmacologia e compará-las com as RNAs;
- Evidenciar alguns dos benefícios trazidos por essas técnicas no ramo da descoberta de novos fármacos.

3 METODOLOGIA

3.1 Critérios para a pesquisa

A pesquisa bibliográfica foi realizada através das plataformas científicas *PubMed*, *CAFe* e *ScienceDirect*, nas quais foram utilizadas as palavras-chave a seguir para a busca dos trabalhos pertinentes ao tema desta revisão. Para esta, se fez uso da busca delimitando os termos abaixo como ‘exatos’, além do ano de publicação 2017-2022.

- “*prediction model*” combinada com as palavras-chave “*artificial neural networks*”, “*biological activity*” e “*natural compounds*”;
- “*prediction model*” combinada com as palavras-chave “*artificial neural networks*”, “*biological activity*” e “*compounds*”;
- “*prediction model*” combinada com as palavras-chave “*artificial neural networks*”, “*biological activity*”, “*compounds*” e “*pharmacological*”.

3.2 Critérios para a seleção de artigos

Para a elaboração dos critérios de seleção e análise dos artigos para o presente trabalho, levou-se em consideração alguns parâmetros listados a seguir:

- Data de publicação: foram selecionados apenas artigos publicados entre os anos de 2017 a 2022 com o objetivo de entender as perspectivas atuais quanto a aplicação de técnicas de RNAs na predição de compostos com propriedades farmacológicas;
- Tipo do artigo: apenas artigos originais foram considerados. Todo e qualquer artigo voltado a revisão, publicação em congresso e afins foram descartados da análise;
- Técnica aplicada: foram selecionados artigos que apresentassem pelo menos a técnica de RNAs com o intuito de destacar a importância desta na predição da atividade de compostos;

- Objetivos: foram selecionados apenas artigos que apresentavam como objetivo a predição da atividade farmacológica de compostos, naturais ou não, podendo auxiliar em pesquisas para novos fármacos.

3.3 Seleção dos estudos

Com a finalidade de reduzir o número de artigos obtidos nas plataformas definidas e aumentar a assertividade destes quanto ao tema desta revisão, inicialmente foi realizada uma seleção prévia dos estudos encontrados. Esta seleção inicial consistiu em análises quanto ao título e resumo destes estudos. Para a seleção final dos artigos, foram realizadas análises quanto à metodologia e a clareza das pesquisas. Todos os artigos que não atenderam aos critérios definidos foram descartados.

3.4 Extração dos dados

Com o intuito de apresentar um panorama geral quanto a aplicação de RNAs em estudos farmacológicos, buscou-se selecionar nesta etapa a temática específica dos artigos, as técnicas de IA utilizadas, assim como os resultados e conclusões obtidas pelas pesquisas.

4 RESULTADOS

Após a busca nas três plataformas escolhidas, foram encontrados 164 resultados por meio de combinações de palavras-chave mencionadas, no período de 2017-2022. Após leitura do título e do resumo destes 164 resultados, foram selecionados 26 artigos para uma análise mais detalhada. Os motivos pelos quais os 138 artigos foram descartados são apresentados na figura logo abaixo (Figura 1).

Analisando a qualidade e a pertinência dos 26 artigos restantes, levou-se em consideração a descrição da técnica de IA utilizada, além da clareza diante da explicação da metodologia e dos resultados obtidos. Deste modo, 12 artigos apresentaram clareza em suas explicações e foram escolhidos para a revisão bibliográfica aqui proposta, conforme também esquema presente na Figura 1.

Vale ressaltar que além do *ScienceDirect* e *PubMed*, foram ainda selecionados, por meio do *CAFe*, artigos do banco de periódicos *MDPI*.

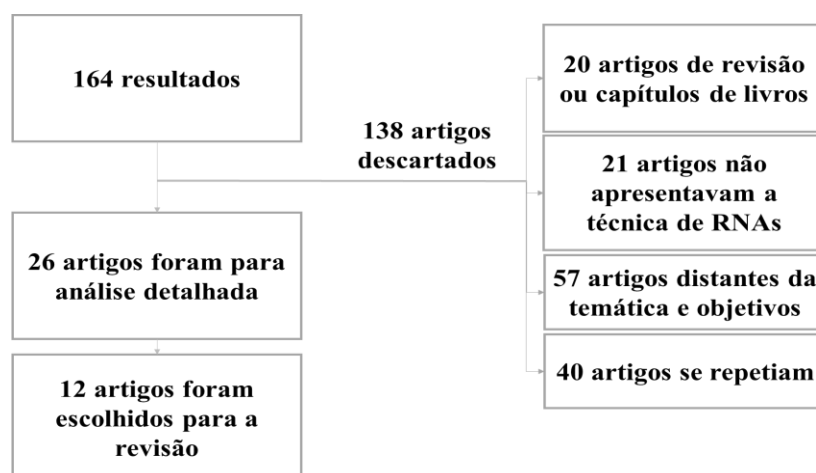


Figura 1 - Representação esquemática da seleção dos artigos a partir das plataformas de buscas e análises posteriores

Fonte: Elaborada pelo autor

A Tabela 1, a seguir, apresenta o título dos trabalhos selecionados para esta pesquisa e seus respectivos autores.

Tabela 1 – Resumo dos títulos e autores referentes aos artigos selecionados

Autores	Título dos artigos
Przybyłek, 2020	Application 2D Descriptors and Artificial Neural Networks for Beta-Glucosidase Inhibitors Screening.
Hdoufane <i>et al.</i> , 2018	In Silico SAR Studies of HIV-1 Inhibitors.
Badura <i>et al.</i> , 2021	Application of artificial neural networks to prediction of new substances with antimicrobial activity against <i>Escherichia coli</i> .
Silva <i>et al.</i> , 2021	Drug design of new 5-HT ₆ R antagonists aided by artificial neural networks.
Kothiwale <i>et al.</i> , 2017	Quantitative Structure–Activity Relationship Modeling of Kinase Selectivity Profiles.
Wu <i>et al.</i> , 2021	Artificial neural network approach for predicting blood brain barrier permeability based on a group contribution method.
Souza <i>et al.</i> , 2019	Quantitative Structure–Activity Relationships for Structurally Diverse Chemotypes Having Anti- <i>Trypanosoma cruzi</i>

Activity.

Kolawole <i>et al.</i> , 2020	In Vitro Biological Estimation of 1,2,3-Triazolo[4,5-d]pyrimidine Derivatives as Anti-breast Cancer Agent: DFT, QSAR and Docking Studies.
Havránková <i>et al.</i> , 2021	Prediction of biological activity of compounds containing a 1,3,5-triazinyl sulfonamide scaffold by artificial neural networks using simple molecular descriptors.
Badura <i>et al.</i> , 2021	Prediction of the antimicrobial activity of quaternary ammonium salts against <i>Staphylococcus aureus</i> using artificial neural networks.
Almi <i>et al.</i> , 2020	QSAR investigations and structure-based virtual screening on a series of nitrobenzoxadiazole derivatives targeting human glutathione-S-transferases.
Naboulsi <i>et al.</i> , 2018	Combining a QSAR Approach and Structural Analysis to Derive an SAR Map of Lyn Kinase Inhibition.

Fonte: Elaborada pelo autor

Para uma melhor discussão quanto à aplicação das RNAs, na predição de compostos no âmbito da farmacologia, também foram extraídas as técnicas de IA utilizadas, a descrição breve dos temas desenvolvidos pelas pesquisas, assim como os resultados e conclusões obtidos pelos artigos, sendo estas informações sintetizadas na Tabela 2.

Tabela 2 – Resumo dos temas, técnicas, resultados e conclusões referentes aos artigos selecionados

Autores	Descrição breve do tema	Técnicas	Resultados e conclusão
Przybyłek, 2020	Utilização de compostos naturais para predição da atividade inibitória de enzimas glucosidases com o intuito de aplicações farmacológicas	RNAs	Correlações positivas e AUC próximo a 1

Kothiwale <i>et al.</i> , 2017	Utilização de compostos para a predição da atividade inibitória de quinases, visando aplicação farmacológica	RNAs	Valores de AUC para 75% dos dados variaram de 0,5 a 1,0
Havránková <i>et al.</i> , 2021	Utilização de compostos para predição da atividade inibitória diante a hCA IX, visando tratamento de câncer de mama	RNAs	Valores altos para acuracidade e concordância dos resultados com o resultado teórico
Badura <i>et al.</i> , 2021	Utilização de compostos para a predição da atividade antimicrobiana contra <i>S. aureus</i> , visando novos medicamentos	RNAs	Acuracidade maior que 90% e AUC superior a 0,9
Badura <i>et al.</i> , 2021	Utilização de compostos para a predição da atividade contra <i>Escherichia coli</i> , visando novos tratamentos	RNAs	Acuracidade de 80% e AUC próximo a 1
Wu <i>et al.</i> , 2021	Utilização de compostos para a predição da atividade permeabilizante na barreira encefálica, visando aplicação farmacológica	RNAs	Correlações positivas
Silva <i>et al.</i> , 2021	Utilização de compostos para a predição de antagonistas de receptores da serotonina, visando desenvolver novas drogas contra Alzheimer	RNAs e Algoritmo genético	Correlações positivas, superiores a 0,8

Kolawole <i>et al.</i> , 2020	Utilização de compostos para predição da atividade como agente anti-câncer de mama	RNAs e Métodos de regressão linear múltipla	Erro a 0,008 e preferência pelas RNAs devido a capacidade preditiva
Almi <i>et al.</i> , 2020	Utilização de compostos para predição da atividade inibitória diante a enzimas GST, visando auxiliar no tratamento de cânceres	RNAs e Métodos de regressão linear múltipla	Valores altos para acuracidade e preferência pelas RNAs devido a capacidade preditiva
Souza <i>et al.</i> , 2019	Utilização de compostos para a predição da atividade contra Chagas, visando o desenvolvimento de novos medicamentos	RNAs e Regressão de mínimos quadrados parcial baseada em Kernel	Correlações positivas, superiores a 0,8
Naboulsi <i>et al.</i> , 2018	Utilização de compostos para a predição da atividade inibitória de quinases, visando aplicação contra cânceres	RNAs e Modelo linear generalizado (MLG)	Correlação positiva e preferência pelas RNAs devido a capacidade preditiva
Hdoufane <i>et al.</i> , 2018	Utilização de compostos para a predição da atividade contra HIV, visando novas alternativas para tratamento	Máquinas de vetores de suporte, RNAs, Florestas aleatórias e Árvores de decisão	Acuracidade de 100% e 90% nos conjuntos de aprendizado e teste. Preferência por RNAs devido a capacidade preditiva

Fonte: Elaborada pelo autor

A seguir, a descrição mais detalhada dos trabalhos é apresentada. Esta engloba os tópicos referentes as técnicas presentes nos estudos, de acordo com a ordem dos artigos mencionados na Tabela 2.

4.1 Artigos com aplicabilidade específica em RNAs

Entre as pesquisas que buscaram a aplicabilidade mais específica das RNAs, Przybyłek (2020) objetivou prever a atividade de compostos naturais como inibidores de

enzimas hidrolíticas do tipo beta-glucosidasas, as quais atuam em rotas metabólicas importantes possuindo diversas utilizações farmacológicas, voltadas para doenças tais como obesidade, diabetes, câncer, HIV e a doença de Gauche, apontadas em seu artigo. Dessa maneira, foram escolhidos 423 compostos naturais para serem classificados quanto ao seu poder inibitório diante de tais enzimas. Além disso, foram calculados os descritores moleculares destes compostos.

Os descritores moleculares são resultantes de procedimentos matemáticos e lógicos, obtidos por software, que transformam informações químicas e moleculares em informações numéricas simbólicas para que estas sejam compreendidas e processadas por computador na forma de matrizes (Alves *et al.*, 2018). Estes são classificados de acordo com suas dimensionalidades em uni, bi e tridimensional, descrevendo propriedades físico-químicas que podem ser calculadas nas conformações 1D, 2D e 3D da molécula. Os descritores moleculares são calculados em todos os outros estudos que participam desta presente revisão para servirem como dados de entrada estruturais, físicos e químicos dos compostos a serem classificados pelas RNAs.

Após o cálculo dos descritores moleculares, na predição dos compostos, Przybyłek utilizou-se da técnica de RNAs do tipo *Multilayer perceptron* (MLP), Figura 2, que funcionam de maneira similar a uma rede de neurônios biológicos, recebendo entradas e atribuindo saídas, onde define-se que as entradas se referem aos dados estruturais, físicos e químicos representativos dos compostos estudados (descritores moleculares) e as saídas aos dados de classificação/predições destes mesmos compostos (Ghritlahre; Prasad, 2018).

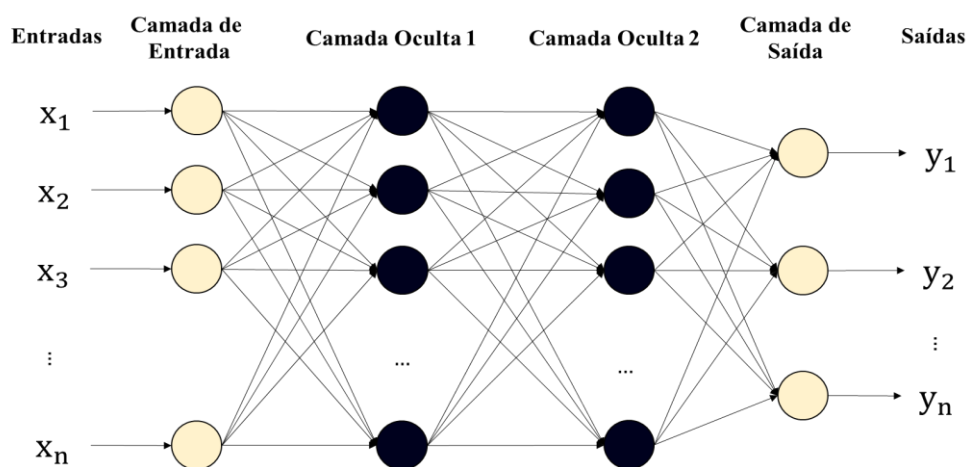


Figura 2 - Representação esquemática da arquitetura de uma RNA – MLP

Fonte: Elaborada pelo autor

Para conclusões, cálculos estatísticos foram realizados, por Przybyłek, fazendo uso do coeficiente de correlação de Matthews (MCC), curva ROC (Receiver Operator Characteristic) e AUC (Area under the ROC Curve), com o intuito de medir a qualidade das classificações. O MCC tem por finalidade retornar valores de saída das RNAs como +1, 0 e -1, apontando em +1 uma predição perfeita quanto a classificação, ou seja, idêntica ao esperado, em 0 uma predição aleatória, e em -1 uma predição inversa, facilitando as análises dos dados de saída (Kohavi; Provost, 1998).

A ROC, representada na Figura 3, é uma curva de probabilidade com dois parâmetros de avaliação: a taxa de falso positivo (dada por compostos classificados erroneamente como ativos e representada em um gráfico pelo eixo x) e a taxa de verdadeiro positivo (dada por compostos classificados verdadeiramente como ativos e representada em um gráfico pelo eixo y) (Fawcett, 2006; Mayer-Baese & Schmid, 2014). Com a ROC permite-se concluir se houve predições, pelas RNAs, ao acaso ou não. A AUC, por sua vez, resume a curva ROC em um valor único, facilitando também a avaliação (Fawcett, 2006; Mayer-Baese & Schmid, 2014).

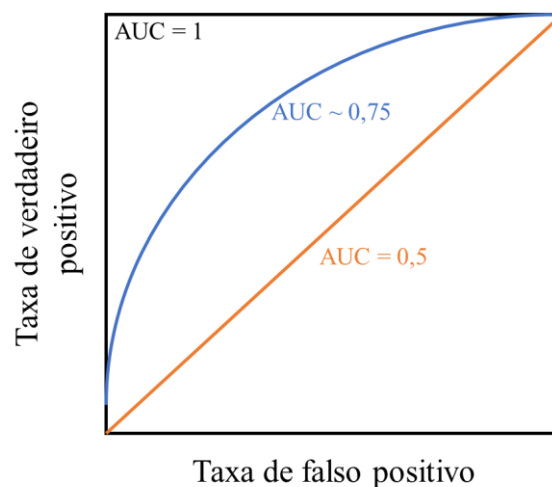


Figura 3 - Representação curva ROC e AUC. Em AUC = 1, por exemplo, observa-se 100% de verdadeiros positivos e em AUC = 0,5 observa-se 50% de verdadeiros positivos e 50% de falsos positivos

Fonte: Elaborada pelo autor

Como resultados, o autor obteve que os valores de MCC foram superiores a 0,7, indicando uma correlação positiva dos classificadores. Além disso, a AUC foi próxima a 1, apontando predominância dos dados classificados como verdadeiramente positivos e

evidenciando um bom desempenho dos modelos de RNAs desenvolvidos na predição de compostos inibidores da enzima beta-glucosidase.

Também na temática de evidenciar novos inibidores para enzimas, Kothiwale *et al.* (2017) buscou, em sua pesquisa, prever a atividade de compostos como inibidores de quinases, família multigênica envolvida na regulação de várias funções celulares e com interesse para a farmacologia. Assim, foram escolhidos 70 inibidores com possível ação para 379 quinases diferentes e as estruturas químicas dos inibidores foram calculadas por descritores moleculares.

Para a predição dos inibidores de quinases, RNAs foram desenvolvidas e analisadas estatisticamente, por Kothiwale, de acordo com curva ROC/AUC. Além disso, outros parâmetros foram analisados, incluindo sensibilidade, seletividade, precisão e MCC. As equações que definem cada um dos parâmetros mencionados anteriormente estão descritas a seguir:

- Sensibilidade (dada pela capacidade de predição correta de verdadeiros positivos) = $TP/(TP + FN)$;
- Seletividade (dada pela capacidade de predição correta de verdadeiros negativos) = $TN/(TN + FP)$;
- Precisão (dada pela proporção de predições corretas) = $(TP + TN)/(TP + TN + FP + FN)$;
- MCC (dado pela medida de qualidade para as classificações) = $((TP \times TN - FP \times FN)) / \sqrt{((TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN))}$;
- Valor preditivo positivo (dado pelo proporção de verdadeiros positivos em relação a todas as predições positivas) = $TP/(TP + FP)$;
- Valor preditivo negativo (dado pelo proporção de verdadeiros negativos em relação a todas as predições negativas) = $TN/(TN + FN)$.

Onde:

- Verdadeiros Positivos (TP) — Compostos experimentalmente ativos, classificados pelas RNAs como ativos para a atividade proposta;
- Verdadeiros Negativos (TN) — Compostos experimentalmente inativos, classificados pelas RNAs como inativos para a atividade proposta;
- Falsos Positivos (FP) — Compostos experimentalmente inativos, classificados como ativos erroneamente, pelas RNAs, para a atividade proposta;
- Falsos Negativos (FN) — Compostos experimentalmente ativos, classificados como inativos erroneamente, pelas RNAs, para a atividade proposta.

Em suas conclusões, Kothiwale obteve que os valores de AUC para 75% das quinases variavam de 0,5 a 1,0, indicando que o modelo utilizado foi um bom ponto de partida para prever a atividade de inibidores contra diferentes quinases, e evidenciando a vantagem desta técnica na triagem de compostos para testes experimentais.

Havránková *et al.* (2021), por sua vez, teve como finalidade prever a atividade inibitória de compostos derivados de 1,3,5-triazinil sulfonamidas contra a anidrase carbônica humana (hCA) IX e II, enzima apontada, em seu estudo, como associada ao câncer de mama e que desempenha um papel crucial ao promover a permanência de um microambiente adequado para tumores, aumentando o crescimento celular. Deste modo, 129 compostos foram selecionados para serem classificados e calculados seus descritores moleculares.

Quanto a técnica, as RNAs do tipo MLP foram utilizadas e a acuracidade do modelo foi comprovada a partir dos valores de atividade encontrados na literatura (dados experimentais) ao serem comparados, por Havránková, com os valores de saída para as RNAs, com acuracidade média de 98%. Neste artigo foi evidenciado a concordância entre os dados teóricos e experimentais de ambas as enzimas hCA IX e II, destacado a facilidade da utilização da metodologia, por não exigir sistema computacional avançado, diante aos ganhos com relação ao tempo na predição de compostos.

Com outra temática, Badura *et al.* (2021) buscou prever a propriedade antibacteriana de derivados do composto imidazol diante a bactéria *Staphylococcus aureus*. O imidazol é apontado, em seu artigo, por ocasionar a quebra da membrana celular e degradação de proteínas e ácidos nucleicos, levando à morte celular desta espécie de bactéria. Neste contexto, 140 compostos foram sintetizados, selecionados e calculados seus descritores moleculares.

Para as predições destes compostos, RNAs foram desenvolvidas e calculados a curva ROC e erro quadrático médio (EQM) para as análises estatísticas. O EQM é definido como a média da diferença entre as saídas verdadeiras e o que foi previsto pelas RNAs, evidenciando acuracidade dos modelos preditivos (Izbicki, 2017).

Ao final desta pesquisa, verificou-se que as RNAs se caracterizaram pela alta previsibilidade, com 93,57% de precisão de classificação. O EQM para cada conjunto de dados foi: 0,08, 0,11 e 0,15, comprovando essa proximidade entre as classificações verdadeiras e previstas. Para o conjunto de treinamento a AUC foi de 0,94, para o conjunto de

teste 0,85 e para o conjunto de validação 1,00, indicando, novamente, boa precisão dos modelos desenvolvidos.

Em outro estudo, Badura *et al.* (2021) buscou prever também a atividade bactericida de compostos derivados de imidazol, mas, neste caso, com relação a *Escherichia coli*, por promoverem a degradação da membrana celular desta bactéria relacionada principalmente a intoxicações alimentares. Nesta sua pesquisa, os mesmos 140 compostos anteriormente escolhidos no artigo referente à *Staphylococcus aureus* foram utilizados.

Fazendo uso também da mesma técnica aplicada na pesquisa anteriormente apresentada, foi possível comprovar, mais uma vez por Badura, o potencial das RNAs ao obter uma acuracidade maior que 80% nos modelos e valores de AUC satisfatórios. Para o conjunto de aprendizagem, a AUC foi de 0,973, para o conjunto de teste 0,974 e 1,00 para a validação, demonstrando, novamente, precisão das redes.

Wu *et al.* (2021), com uma abordagem diferente dos artigos anteriormente mencionados, buscou prever a permeabilidade de compostos farmacológicos quanto a barreira hematoencefálica, apontada como limitante, em seu trabalho, para que medicamentos possam atingir o cérebro em casos de doenças tais como AVC, responsável mundialmente por 5 milhões de mortes todo ano de acordo com a Organização Mundial da Saúde (OMS). Assim, no estudo em questão, 300 compostos foram selecionados para esta predição utilizando RNAs e calculados seus descritores moleculares.

Análises quanto ao coeficiente de correlação, erro relativo (ER) e EQM foram realizadas, apontando, portanto, viabilidade, robustez e precisão do modelo matemático (Justo *et al.*, 2020). Em seus resultados, Wu obteve que o coeficiente de correlação, o ER e a raiz do erro quadrático médio (REQM) foram de 0,956, 0,857 e 0,171, valores estes satisfatórios, evidenciando a importância das RNAs obtidas como modelo preditivo desenvolvido para a atividade de permeabilidade com relação a barreira hematoencefálica e, contribuindo, assim, para os avanços nas pesquisas visando o surgimento de novos medicamentos nessa área.

4.2 Artigos com aplicabilidade em RNAs e Algoritmos genéticos

Diferente dos demais trabalhos, Silva *et al* (2021) teve como finalidade, em sua pesquisa, a utilização de RNAs associadas a algoritmos genéticos (AGs). Neste estudo, buscou-se prever a atividade de compostos antagonistas do receptor serotoninérgico (5-HT₆) visando o tratamento da doença Alzheimer, a qual impacta, segundo dados fornecidos por seu

artigo, mais de 50 milhões de pessoas em todo o mundo. Desse modo, 81 compostos antagonistas foram selecionados e calculados seus descritores moleculares.

Para a predição destes compostos, como mencionado, RNAs do tipo MLP foram desenvolvidas e aplicadas, neste caso, em conjunto com os AGs. Os AGs são adicionados às RNAs visando a escolha das melhores arquiteturas de redes (Idrissi *et al.*, 2016), conforme Figura 4. Estes são inspirados nos princípios da evolução das espécies proposto por Charles Darwin, onde os indivíduos mais aptos sobrevivem. Este mesmo princípio, de evolução para a busca das melhores populações, estruturalmente é imitado nos AGs, em que as populações são as RNAs onde aquelas com as melhores estruturas são apontadas como solução para um problema (Idrissi *et al.*, 2016).

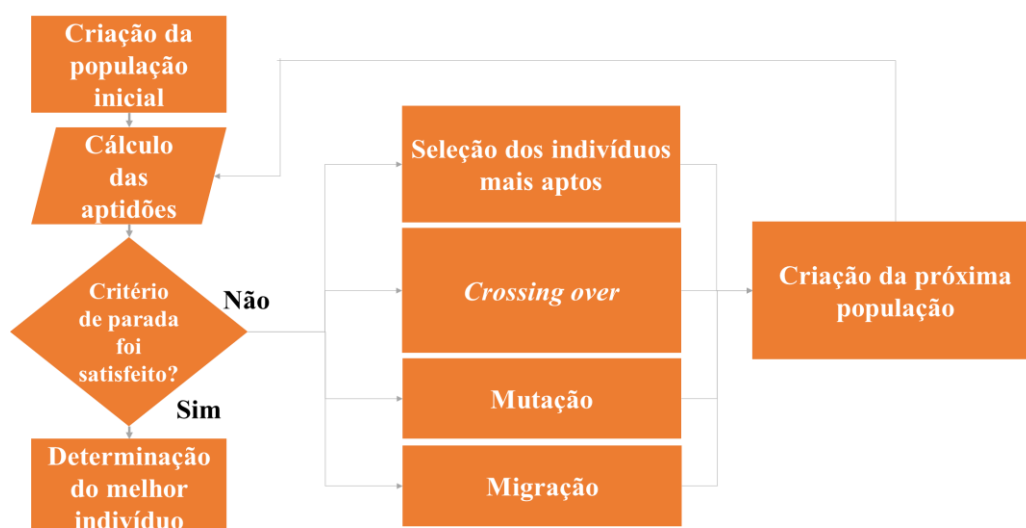


Figura 4 - Representação estrutural dos AGs

Fonte: Elaborada pelo autor

Silva *et al* (2021), em suas conclusões, mostrou que foi possível obter uma acuracidade de mais de 80% nos modelos, uma vez que o treinamento foi de 0,82, validação 0,81 e teste 0,82. Em outros testes com novos compostos antagonistas ao receptor 5-HT6, nunca antes analisados pelas RNAs, foi possível apresentar bons resultados para a atividade biológica, equivalentes aos da literatura, indicando mais de 95% de acerto nas predições.

4.3 Artigos com aplicabilidade em RNAs e técnicas de Regressão

Entre os trabalhos que visam comparar técnicas de RNAs com regressão, Kolawole *et al.* (2020), em sua pesquisa, buscou evidenciar a atividade de compostos heterocíclicos

derivados de [1,2,3] triazolo [4,5-d] pirimidina quanto sua ação anticâncer de mama, condição que, segundo seu artigo, é considerado um risco de vida comum principalmente entre as mulheres. Assim, 20 compostos foram selecionados e calculados seus descritores moleculares.

Quanto as técnicas, as RNAs do tipo MLP e a regressão linear múltipla foram utilizadas neste estudo. Na regressão linear múltipla, mais de duas variáveis são analisadas por entender que estas são capazes de explicar modelos por relação, apontando viabilidade em um estudo. Para este modelo, tem-se que várias variáveis X_1, X_2, \dots, X_n são relacionadas com uma variável Y , conforme mostrado na equação 1, a seguir, em que n é o número de variáveis utilizadas para a regressão múltipla (Pauli *et al.*, 2019).

$$Y = A_0 + A_1X_1 + A_2X_2 + A_3X_3 + \dots + A_nX_n \quad (1)$$

Por fim, para conclusões estatísticas, o coeficiente de correlação e o EQM também foram calculados por Kolawole. Os resultados mostram que os parâmetros para a regressão linear múltipla foram R-quadrado de 0,905 e EQM de 0,008. As RNAs, por sua vez, se mostraram eficientes e preditivas, com R-quadrado de 0,998 e EQM de 0,002, confirmando as hipóteses teóricas quanto aos compostos já testados em bancada diante aos dados na literatura.

Utilizando as mesmas técnicas de RNAs e regressão linear múltipla, Almi *et al.* (2020), em sua pesquisa, objetivou prever a atividade inibitória de derivados de nitrobenzoxadiazol em relação a glutathione-S-transferases (GSTP1-1), enzima de desintoxicação associada, em seu trabalho, com a resistência dos medicamentos, diminuindo a atividade de medicamentos anticâncer. Neste contexto, 38 compostos foram escolhidos, suas estruturas foram otimizadas e o cálculo dos descritores moleculares foi realizado, sendo selecionados aqueles que mais definem/representam os compostos do estudo.

Como resultados, foi possível confirmar, por Almi, o melhor desempenho das RNAs com relação a regressão linear múltipla, indicando a existência de uma relação não linear, ao obter, para as RNAs, um R-quadrado de 0,812 e REQM de 0,221 em comparação com 0,758 e 0,240, respectivamente. Houve uma boa capacidade preditiva, apontada por estudos, com compostos para teste ao determinar uma atividade inibitória com similaridade na estrutura de compostos ativos previstos.

Souza *et al.* (2019), por sua vez, em seu trabalho, buscou prever a atividade tripanocida de compostos diante ao *Trypanosoma cruzi*, causador da doença de Chagas a qual afeta, segundo a Organização Mundial da Saúde (OMS), 8 milhões de pessoas no mundo

levando a 10.000 mortes por ano, principalmente em regiões negligenciadas onde não se tem o controle do ‘barbeiro’, parasita responsável por essa condição. Desse modo, 363 compostos foram selecionados e calculados seus descritores moleculares.

Quanto as técnicas, as RNAs do tipo MLP e a regressão por mínimos quadrados parciais (Kernel PLS) foram utilizadas nesta pesquisa. A regressão do tipo Kernel PLS é utilizada em casos de não linearidade para a resolução de problemas envolvendo estimativas, classificações e reconhecimento de padrões, visando encontrar relações entre os dados (Renteria, 2004).

Ao final do trabalho, o coeficiente de correlação e o EQM também foram calculados por Souza. Conclui-se que os modelos foram capazes de apresentar uma boa capacidade preditiva para os compostos do conjunto de teste, produzindo valores para o coeficiente de correlação de 0,81 e 0,84 para RNAs e regressão do tipo Kernel, respectivamente. Foi destacado também a importância destes estudos nesse contexto em que atualmente o tratamento para Chagas é limitado e obsoleto, evidenciando a necessidade de se desenvolver novos fármacos.

Outros trabalhos também buscaram a comparação das RNAs com outras técnicas. Naboulsi *et al.* (2018), em seu artigo, buscou prever a atividade inibitória de compostos em relação a Lyn quinase, membro da família Src de proteínas tirosina quinases, expressa principalmente por várias células hematopoiéticas, tecidos neurais e adiposos, com relação a doenças tais como o câncer, representando potencial como alvo para o desenvolvimento de novas drogas antitumorais. Neste contexto, 176 compostos foram selecionados e seus descritores moleculares foram calculados.

Quanto as técnicas, as RNAs e modelo linear generalizado (MLG) foram utilizados. O MLG é uma extensão dos modelos de regressão linear simples e múltipla que aumenta as análises possíveis diante de um conjunto de dados (Pérez, 2021). Nesta pesquisa, o coeficiente de correlação e o EQM também foram calculados.

Como resultados, verificou-se por Naboulsi que o modelo de RNAs foi capaz de fornecer os melhores valores em relação a abordagem MLG, cujo poder preditivo foi considerado baixo. Para as RNAs, o coeficiente de correlação e a REQM foram satisfatórios, apresentando um valor de 0,92 e de 0,29, respectivamente.

4.4 Artigos com aplicabilidade em RNAs e técnicas variadas

Por fim, Hdoufane *et al.* (2018) fez uso de 4 técnicas de IA apresentando como objetivo prever a atividade de derivados de tetrahydroimidazo[4,5,1-jk][1,4]benzodiazepina (TIBO) como inibidores da enzima transcriptase reversa, alvo de medicamentos para impedir a infecção de novas células pelo HIV-1. Dessa maneira, foram escolhidos 84 compostos, além de terem sido calculados seus descritores moleculares.

Para a classificação das moléculas, Hdoufane fez uso dos seguintes modelos, visando compará-los: RNAs MLP, máquinas de vetores de suporte (SVMs), árvores de decisão e florestas aleatórias. Aqueles não mencionados anteriormente serão resumidos brevemente a seguir.

As SVMs são algoritmos de aprendizagem de máquinas classificadas, também, como RNAs. São capazes de prever saídas por meio de dados para treinamento e generalização com o intuito de classificação binária (Oliveira, 2010). Este modelo tem como finalidade a criação de barreiras que dividem os dados em duas classes distintas possibilitando tal classificação (Huang *et al.*, 2018), como pode ser observado na Figura 5.

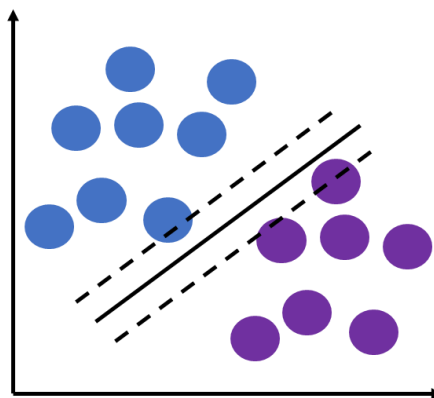


Figura 5 - Representação gráfica do algoritmo de aprendizagem de SVMs

Fonte: Elaborada pelo autor

As árvores de decisão, por sua vez, também são modelos preditivos, porém fazem uso de regras para o cálculo de um determinado valor de interesse. A partir destas regras se tem a obtenção de classificações. Estruturalmente são formadas por nós e ramos, se assemelhando a uma árvore, e representando atributos, resultados e classificações dos dados analisados (Ibanês, 2016). Por fim, as florestas aleatórias é uma técnica de aprendizado de máquina constituída por um conjunto de árvores de decisão cujos dados de entrada são selecionados de

maneira aleatória e a classificação é obtida por meio dos atributos contidos nas árvores, gerando um resultado preditivo (Ibanês, 2016). Ambos são representados na Figura 6.

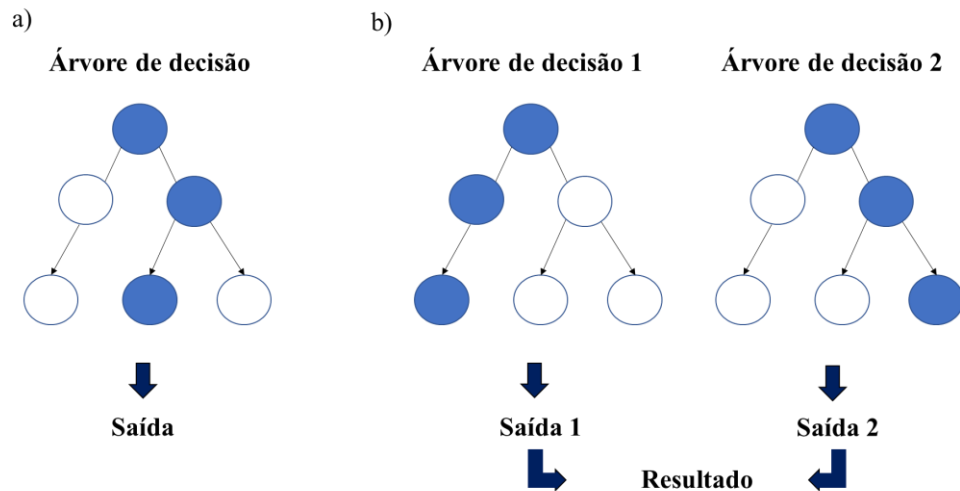


Figura 6 - Representação, em 'a', estrutural das Árvores de Decisão e, em 'b', das Florestas Aleatórias, gerando um resultado (saída) preditivo

Fonte: Elaborada pelo autor

Após a comparação dos modelos, Hdoufane obteve uma acuracidade média de 100% e 90% nos conjuntos de aprendizado e teste, respectivamente. Desta forma, todas as técnicas apresentaram bons resultados, se destacando os modelos de RNAs e SVMs por obterem maiores acertos na predição dos compostos alvo do estudo (acuracidade de 90 e 95%, respectivamente em testes, em comparação com 75% da floresta aleatória e do 70% da modelo árvore de decisão).

5 DISCUSSÃO

Nesta revisão, observou-se que a grande maioria dos artigos no ramo da IA são datados nos últimos anos e, destes, diversos foram encontrados de maneira repetida nas diferentes plataformas de pesquisa, uma vez que 40 dos estudos se encontravam nessa situação. Infere-se que os termos de busca utilizados podem ter levado a estas repetições, assim como a encontrar um número de artigos restrito, 12 ao todo.

Com relação às temáticas, os artigos aqui apresentados tiveram como objetivo, principalmente, o desenvolvimento de novos inibidores enzimáticos, dado que 6 dos 12 trabalhos apresentavam esse tema.

Entre os tratamentos/curas, observou-se que há a aplicação para as mais diversas doenças, com destaque para os cânceres, em 5 dos 12 estudos. As RNAs foram capazes de auxiliar nas fases iniciais das pesquisas para o desenvolvimento de novos medicamentos por meio da redução de recursos, como tempo e custo, de acordo com o que foi apontado pelo autor Havránková *et al.* (2021). Essas dificuldades iniciais relacionadas a recursos vão ao encontro às dificuldades evidenciadas por Wouters *et al.* (2020), onde afirma que os gastos envolvendo o desenvolvimento de novos medicamentos são de US\$ 314 milhões a US\$ 2,8 bilhões, levando uma década ou mais até sua comercialização.

A alta capacidade de poder preditivo das RNAs diante aos compostos, se sobressaindo, na maior parte das vezes, ao ser comparada com outras técnicas, como apontado, por exemplo, por Almi *et al.* (2020), Naboulsi *et al.* (2018) e Hdoufane *et al.* (2018), também foi evidenciada. Tal técnica de IA é considerada, de maneira unânime para os autores, como promissora, apresentando grandes porcentagens de acuracidade e não revelando aleatoriedade, em nenhum dos estudos, na predição dos compostos de interesse, com resultados similares, inclusive, aos resultados em bancada, como apontado por Havránková *et al.* (2021) e Silva *et al.* (2021), por exemplo.

6 CONCLUSÃO

O presente trabalho revelou as atuais aplicações da IA no âmbito do desenvolvimento de novos fármacos, abordando diferentes modelos matemáticos de regressão e de aprendizado de máquinas. Conclui-se que, apesar de recentes, as RNAs podem ser consideradas promissoras nestes estudos, principalmente nas fases iniciais do desenvolvimento de medicamentos, onde muitas são as chances de falhas em eficiência e segurança. Sendo assim, tal técnica se mostra promissora, acelerando tanto na descoberta diante a atividade de compostos conhecidos, quanto na triagem e testagem de novos compostos sintéticos, reduzindo, portanto, custos e tempo no desenvolvimento de novos tratamentos e curas.

REFERÊNCIAS

ABIODUN, O. I. et al. State-of-the-art in artificial neural network applications : A survey. **Heliyon**, v. 4, n. 11, p. 1–41, 2018. Disponível em: <<https://doi.org/10.1016/j.heliyon.2018.e00938>>.

- ALMI, I. QSAR investigations and structure-based virtual screening on a series of nitrobenzoxadiazole derivatives targeting human glutathione-S-transferases. **Journal of Molecular Structure**, 2020. Disponível em: <<https://doi.org/10.1016/j.molstruc.2020.128015>>.
- ALVES, V. M. CHEMINFORMATICS: AN INTRODUCTION. **Quim. Nova**, v. 41, n. 2, 2018. Disponível em: <<https://doi.org/10.21577/0100-4042.20170145>>.
- ATANASOV, A. G. et al. Natural products in drug discovery: advances and opportunities. **Nature Reviews Drug Discovery**, v. 20, p. 200-216, 2021. Disponível em: <<https://doi.org/10.1038/s41573-020-00114-z>>.
- BADURA, A. et al. Application of artificial neural networks to prediction of new substances with antimicrobial activity against Escherichia coli. **Journal of Applied Microbiology**, v. 130, n. 1, 2021. Disponível em : <<https://doi.org/10.1111/jam.14763>>.
- BADURA, A. et al. Prediction of the antimicrobial activity of quaternary ammonium salts against Staphylococcus aureus using artificial neural networks. **Arabian Journal of Chemistry**, v. 14, n. 7, 2021. Disponível em: <<https://doi.org/10.1016/j.arabjc.2021.103233>>.
- BERTOLACCINI, L. et al. An overview of the use of artificial neural networks in lung cancer research. **Journal of Thoracic Disease**, v. 9, n. 4, p. 924–931, 2017. Disponível em: <<https://dx.doi.org/10.21037%2Fjtd.2017.03.157>>.
- CAMPOS, T. C.; VASCONCELOS, T. C. L. Aplicação de algoritmos de machine learning na área farmacêutica: uma revisão. **Research, Society and Development**, v. 10, n. 15, 2021. Disponível em: <<https://doi.org/10.33448/rsd-v10i15.22862>>.
- DAVENPORT, T.; KALAKOTA, R. The potential for artificial intelligence in healthcare. **Future Healthcare Journal**, v. 6, n. 2, p. 94–98, 2019.
- DWIVEDI, Y. K. et al. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. **International Journal of Information Management**, v. 57, 2021.
- FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v.27, n. 8, 2006. Disponível em: <<https://doi.org/10.1016/j.patrec.2005.10.010>>.
- FREEDMAN, D. H. Hunting for New Drugs with AI. **Nature**, 2019. Disponível em: <<https://doi.org/10.1038/d41586-019-03846-0>>.

GHRITLAHRE, H. K.; PRASAD, R. K. Application of ANN technique to predict the performance of solar collector systems - A review. **Renewable and Sustainable Energy Reviews**, v. 84, p. 75–88, 2018, 2017. Disponível em: < <https://doi.org/10.1016/j.rser.2018.01.001>>.

HAVRÁNKOVÁ, E. et al. Prediction of biological activity of compounds containing a 1,3,5-triazinyl sulfonamide scaffold by artificial neural networks using simple molecular descriptors. **Bioorganic Chemistry**, v. 107, 2021. Disponível em: < <https://doi.org/10.1016/j.bioorg.2020.104565>>.

HEIFETZ, A. Quantum Mechanics in Drug Discovery, **Methods in Molecular Biology**, v. 2114, 2020. Disponível em: <[doi:10.1007/978-1-0716-0282-9](https://doi.org/10.1007/978-1-0716-0282-9)>.

HDOUFANE, I. et al. In Silico SAR Studies of HIV-1 Inhibitors. **Pharmaceuticals (Basel)**, v. 11, n. 3, 2018. Disponível em: <[10.3390/ph11030069](https://doi.org/10.3390/ph11030069)>.

HUANG, S. et al. Applications of support vector machine (SVM) learning in cancer genomics. **Cancer Genomics and Proteomics**, v. 15, n. 1, p. 41–51, 2018.

IBANÊZ, M. M. Uso de redes neurais nebulosas e florestas aleatórias na classificação de imagens em um projeto de ciência cidadã. Tese (Dissertação Mestrado) - Instituto Nacional de Pesquisas Espaciais. São José dos Campos, 2016.

IDRISSI, M. A. J. et al. Genetic algorithm for neural network architecture optimization. **3rd International Conference on Logistics Operations Management (GOL)**, p. 1-4, 2016.

IZBICKI, R. Machine Learning sob a ótica estatística. **Insper Instituto de Ensino e Pesquisa**, 2017. Disponível em: <http://est.ufmg.br/~marcosop/est171-ML/MachineLearning_Izbicki.pdf>. Acesso em: 10 de maio de 2022.

JUSTO, D. A. R. et al. Cálculo Numérico: um Livro colaborativo versão Python. **UFRGS**, 2020. Disponível em: <<https://www.ufrgs.br/reatmat/CalculoNumerico/livro-py/livro-py.pdf>>. Acesso em: 10 de maio de 2022.

KOHAVI, R.; PROVOST, F.. Glossary of terms. **Machine Learning**, v. 30, n. 2-3, p. 271-274, 1998.

KOLAWOLE, O. A. et al. In Vitro Biological Estimation of 1,2,3-Triazolo[4,5-d]pyrimidine Derivatives as Anti-breast Cancer Agent: DFT, QSAR and Docking Studies. **Curr Pharm**

Biotechnol., v. 20, n. 1, p. 70-78, 2020. Disponível em: <10.2174/1389201020666190904163003>.

KOTHIWALE, S. et al. Quantitative Structure–Activity Relationship Modeling of Kinase Selectivity Profiles. **Molecules**, v. 22, n. 9, 2017. Disponível em: <https://doi.org/10.3390/molecules22091576>.

MEYER-BAESE, A., & SCHMID, V. Statistical and Syntactic Pattern Recognition. **Pattern Recognition and Signal Analysis in Medical Imaging**, 2014. Disponível em: <https://doi.org/10.1016/B978-0-12-409545-8.00006-6>.

NABOULSI, I. Combining a QSAR Approach and Structural Analysis to Derive an SAR Map of Lyn Kinase Inhibition. **Molecules**, v. 23, n. 12, 2018. Disponível em: <10.3390/molecules23123271>.

OLIVEIRA, J. M. J. Máquina de Vetores Suporte: estudo e análise de parâmetros para otimização de resultado. Tese (Monografia) - Centro de Informática, Universidade Federal de Pernambuco. Recife, 2010.

PAULI, S. T. Z. et al. Técnicas de Redes Neurais Artificiais e Regressão Linear Múltipla aplicadas na previsão de valores de importação e exportação no setor de bebidas no Brasil. **APREPRO**, 2019. Disponível em: <http://aprepro.org.br/conbrepro/2019/anais/arquivos/08092019_090853_5d4d6ca5e4964.pdf>. Acesso em: 10 de maio de 2022.

PÉREZ, F. L. Modelos Lineares Generalizados. **UFPR**, 2021. Disponível em: <http://leg.ufpr.br/~lucambio/GLM/GLM.html>. Acesso em: 10 de maio de 2022.

PRZYBYLEK, M. Application 2D Descriptors and Artificial Neural Networks for Beta-Glucosidase Inhibitors Screening. **Molecules**, v. 25, n. 24, 2020. Disponível em: <10.3390/molecules25245942>.

RENTERIA, R. P. ALGORITMOS PARA REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS. **PUC RIO**, 2004.

SILVA, A. P. et al. Drug design of new 5-HT₆ R antagonists aided by artificial neural networks. **J Mol Graph Model**, 2021. Disponível em: <10.1016/j.jmgm.2021.107844>.

SILVA, I. N. et al. **Redes neurais artificiais: para engenharia e Ciências aplicadas**. São Paulo: Artliber, 2010.

SOUZA, A. S. et al. Quantitative Structure–Activity Relationships for Structurally Diverse Chemotypes Having Anti-Trypanosoma cruzi Activity. **Int. J. Mol. Sci.**, v. 20, n. 11, 2020. Disponível em: <<https://doi.org/10.3390/ijms20112801>>.

TRIPATHI, N. Et al. Applications of artificial intelligence to drug design and discovery in the big data era: a comprehensive review. **Mol Divers**, v. 25, n.3, p. 1643-1664. Disponível em: <<https://doi.org/10.1007/s11030-021-10237-z>>.

WIRTZ, B. W. et al. Artificial Intelligence and the Public Sector—Applications and Challenges. **International Journal of Public Administration**, v. 42, n. 7, p. 596–615, 2019.

WOUTERS, O. J. et al. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. **JAMA Network**, v. 323, n. 9, p. 844-853, 2020. Disponível em: <[10.1001/jama.2020.1166](https://doi.org/10.1001/jama.2020.1166)>.

WU, Z. et al. Artificial neural network approach for predicting blood brain barrier permeability based on a group contribution method. **Comput Methods Programs Biomed.**, 2021. Disponível em: <[10.1016/j.cmpb.2021.105943](https://doi.org/10.1016/j.cmpb.2021.105943)>.