

INTELIGÊNCIA ARTIFICIAL E BIOINFORMÁTICA APLICADA À PESQUISA CIENTÍFICA: UMA REVISÃO SOBRE A DESCOBERTA DE NOVOS FÁRMACOS



<https://doi.org/10.22533/at.ed.1661125170313>

Data de aceite: 18/07/2025

Pâmela Gomes Santos

Universidade Federal do Maranhão-UFMA
<http://lattes.cnpq.br/7818949201806939>

Vanessa da Silva Lima

Universidade Federal do Maranhão-UFMA
<http://lattes.cnpq.br/0690064398684522>

Gabriela Cristina Baccaro

Universidade Estadual de Campinas-
UNICAMP
<http://lattes.cnpq.br/5788131997066000>

Isadora Mara Cunha Bezerra

Centro Universitário Christus-
UNICHRISTUS
<http://lattes.cnpq.br/3204829407736863>

Lailda Brito Soares

Universidade Federal do Maranhão-UFMA
<https://lattes.cnpq.br/0712430899233889>

Marcela Bacchetti Vicentini

Centro Universitário Católico de Vitoria-
UCV
<http://lattes.cnpq.br/4128271647237774>

Ayron Abraão César Xavier

Universidade Potiguar-UNP
<http://lattes.cnpq.br/4197332473894095>

Brunela Pimentel de Oliveira

Centro Universitário Vértice-UNIVÉRTIX
<http://lattes.cnpq.br/6605241150012463>

Celine Mano Andrade

Universidade Federal do Norte do
Tocantis-UFNT
<http://lattes.cnpq.br/3136555484813189>

Isadora Nascimento

Universidade do Oeste Paulista-
UNOESTE
<http://lattes.cnpq.br/2781634934901175>

Raquel Paes dos Santos

Faculdade dos Guararapes-UNIFG
<https://lattes.cnpq.br/1918846566635532>

Paloma Gomes Santos

União Brasileira de Faculdades-UNIBF
<http://lattes.cnpq.br/8850807094785171>

RESUMO: A descoberta de novos fármacos é um processo complexo, demorado e de alto custo, frequentemente ultrapassando uma década de pesquisas e bilhões em investimentos. Nesse contexto, a Inteligência Artificial (IA) e a Bioinformática são incorporadas como tecnologias na otimização das etapas iniciais da prospecção e triagem de compostos bioativos. Este capítulo apresenta uma revisão narrativa da literatura científica publicada entre 2015

e 2025, com o objetivo de identificar os principais avanços, desafios e aplicações dessas ferramentas na descoberta de medicamentos. A busca foi realizada nas bases de dados PubMed, Web of Science e Google Scholar, com base nos descritores “Artificial Intelligence”, “Drug Discovery” e “Bioinformatics”. Após os critérios de elegibilidade, foram selecionados oito artigos revisados por pares. Os resultados demonstram que abordagens baseadas em *machine learning* (ML), *deep learning* (DL), redes neurais convolucionais (CNN), redes neurais gráficas (GNN), *docking* molecular e modelos generativos (GANs) vêm sendo utilizadas com sucesso na predição de atividade biológica, identificação de alvos moleculares, redução de falsos positivos e geração de novos compostos com propriedades desejadas. As ferramentas analisadas evidenciam ganhos em precisão e agilidade nas triagens virtuais, especialmente quando associadas a grandes bancos de dados estruturais. Apesar dos avanços, há limitações quanto à necessidade de validação experimental *in vitro* e *in vivo*, além da escassez de padronização metodológica e dados de alta qualidade. Conclui-se que IA e bioinformática representam uma convergência tecnológica promissora, mas que requer amadurecimento técnico e regulatório para plena integração à pesquisa translacional.

PALAVRAS-CHAVES: *Machine learning*; *Deep learning*; Bioinformática; Inteligência artificial; Pesquisa científica.

ARTIFICIAL INTELLIGENCE AND BIOINFORMATICS APPLIED TO SCIENTIFIC RESEARCH: A REVIEW ON THE DISCOVERY OF NEW DRUGS

ABSTRACT: The discovery of new drugs is a complex, time-consuming, and high-cost process, often exceeding a decade of research and billions in investment. In this context, Artificial Intelligence (AI) and Bioinformatics are being incorporated as key technologies to optimize the early stages of bioactive compound prospecting and screening. This chapter presents a narrative review of the scientific literature published between 2015 and 2025, aiming to identify the main advances, challenges, and applications of these tools in drug discovery. The literature search was conducted in the PubMed, Web of Science, and Google Scholar databases using the descriptors “Artificial Intelligence”, “Drug Discovery”, and “Bioinformatics”. After applying eligibility criteria, eight peer-reviewed articles were selected. The results show that approaches based on *machine learning* (ML), *deep learning* (DL), convolutional neural networks (CNN), graph neural networks (GNN), molecular *docking*, and generative models (GANs) have been successfully used for predicting biological activity, identifying molecular targets, reducing false positives, and generating novel compounds with desired properties. The analyzed tools demonstrate improvements in precision and speed of virtual screening, especially when integrated with large structural databases. Despite the advances, limitations remain regarding the need for experimental validation *in vitro* and *in vivo*, as well as the lack of methodological standardization and high-quality data. It is concluded that AI and bioinformatics represent a promising technological convergence, but still require technical and regulatory maturation for full integration into translational research.

KEYWORDS: *Machine learning*; *Deep learning*; Bioinformatic; Artificial intelligence; Scientific research.

INTRODUÇÃO

A descoberta de fármacos é desafiadora, processual e demanda alto recurso financeiro e humano. O desenvolvimento completo de um novo medicamento, desde sua concepção identificando os alvos biológicos, análises fitoquímicas, ensaios pré-clínicos até a aprovação clínica, frequentemente ultrapassa 10 a 15 anos de pesquisas e investimentos superiores a 2 bilhões de dólares. Esse cenário, aliado a crescente demanda por terapias mais eficazes e personalizadas, impulsiona o uso de tecnologias disruptivas, que se propõem a otimizar e transformar a forma como esses fármacos são descobertos e desenvolvidos (Dimasi *et al.*, 2016; Neto *et al.*, 2024).

Diante disso, a Inteligência Artificial (IA) e a Bioinformática surgem como pilares de uma nova abordagem na pesquisa científica, caracterizada pela predição, análise massiva de dados e automação. A IA, principalmente em suas abordagens de *machine learning* (ML) e *deep learning* (DL), aprende padrões complexos de grandes volumes de dados e prediz a atividade biológica de moléculas, modelagem de estruturas protéicas e gera novos compostos com as propriedades almejadas. Em contrapartida, a bioinformática disponibiliza os dados genômicos, estruturais, proteômicos essenciais para alimentar esses modelos, além de possibilitar a priorização e mapeamento dos alvos terapêuticos com base em evidências biológicas (Pereira *et al.*, 2013; Shiri *et al.*, 2024; De Souza *et al.*, 2024).

Um método computacional na triagem de fármacos é o *docking* molecular, que prever a direção preferencial de uma molécula quando se liga na proteína alvo, estimando estabilidade e afinidade desse complexo. É uma técnica fundamental para o *screening* virtual, pois permite a avaliação de milhares a milhões de compostos em busca daqueles com maior potencial terapêutico. No entanto, o *docking* possui limitações na modelagem de flexibilidade molecular, o que pode comprometer a exatidão das predições, e desafios como a elevada taxa de falsos positivos (Ruyck *et al.*, 2016; Singh *et al.*, 2022).

A interação entre o *docking* molecular e a IA traz avanços significativos, pois o aprendizado de máquinas filtra resultados, reduz falsos positivos e aprimora a modelagem das interações proteína-ligante. Essa combinação potencializa o *screening* computacional, o que otimiza o processo de triagem, além de abrir campo para a geração de novos compostos por redes neurais generativas (Gupta *et al.*, 2021; Gentile *et al.*, 2022). Enquanto a interação entre a bioinformática e IA revoluciona os processos iniciais de prospecção de novos bioativos, com o *screening* virtual, a modelagem computacional e a identificação de interações proteína-ligante. Ferramentas como AlphaFold, GNNs (Graph Neural Networks) e GANs (Generative Adversarial Networks) tornaram possível o que até poucos anos atrás era inatingível: prever com alta predição o comportamento de moléculas complexas antes de produzi-las em laboratório (Ruff; Pappu, 2021; Karim *et al.*, 2023; Li *et al.*, 2025).

Frente a esse cenário de transformação, a pergunta norteadora dessa pesquisa é a seguinte: *Como a inteligência artificial e bioinformática têm sido utilizadas na descoberta de novos fármacos entre os anos de 2015 e 2025?* Por meio de uma revisão narrativa, este capítulo de livro teve como objetivo identificar os principais avanços, desafios e aplicações práticas que permeiam essa fusão de tecnologias e seus impactos na pesquisa científica.

METODOLOGIA

Este trabalho consiste em um estudo realizado por meio de uma revisão narrativa de literatura. Os artigos selecionados para compor a pesquisa foram publicados entre o período de 2015 a 2025. As buscas foram realizadas nas bases de dados Google Scholar, Web of Science e PubMed, utilizando as combinações dos seguintes descritores: “*Artificial Intelligence*” OR “*Drug Discovery*”) AND “*Bioinformatics*” AND “*Scientific Research*”. Literatura cinza, artigos de opinião e estudos sem foco claro em IA aplicada a pesquisa científica foram excluídos (Figura 1).

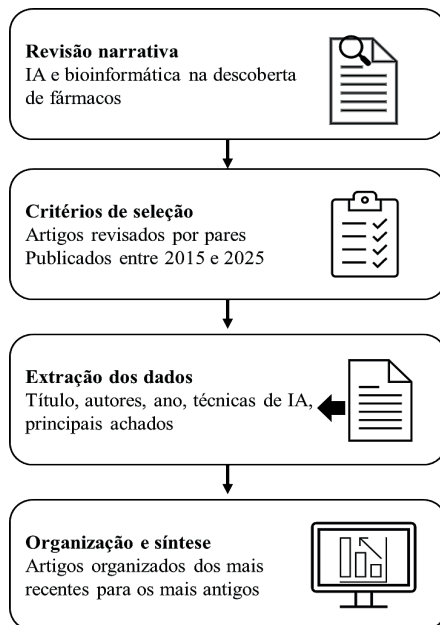


Figura 1: Fluxograma metodológico

RESULTADOS E DISCUSSÃO

A partir do levantamento bibliográfico, selecionaram-se oito artigos publicados entre os anos de 2015 a 2025 (Tabela 1), que revelam uma evolução significativa no uso da bioinformática e inteligência artificial (IA) na descoberta de novos fármacos. Análises baseadas em *machine learning* (ML), *deep learning* (DL), redes neurais convolucionais (CNN), redes neurais gráficas (GNN), *docking* molecular e redes neurais generativas (GANs) revolucionaram a triagem dos compostos bioativos, com destaque para as ferramentas que operam sobre grandes bancos de dados moleculares e estruturais.

O estudo de Noor *et al.*, (2024) emprega DL e GNN para otimizar o processo de triagem, aumentando a eficiência na previsão de interações entre ligantes e alvos moleculares. Esses dados corroboram com os resultados de Li *et al.*, (2019), que utilizaram

o *docking* molecular com o ML para identificar seis inibidores inéditos da Topoisomerase I, o que evidencia a capacidade da IA em descobrir compostos/moléculas com atividade biológica ainda não descritos na literatura.

No campo da predição da afinidade ligante-receptor, Hong *et al.*, (2024) propuseram três modelos de GNN que superaram abordagens anteriores, com o melhor desempenho na previsão de ligações e incertezas, o que reflete a tendência promissora na química computacional moderna. Já Wallach *et al.*, (2015) introduziram o AtomNet, a primeira rede neural convolucional profunda aplicada a predição de bioatividade, com 57,8% de acerto entre os alvos moleculares.

A convergência entre ferramentas de aprendizado e grandes bancos de dados também se destaca na pesquisa de Wójcikowshi *et al.*, (2017), que desenvolveram uma nova função de pontuação baseada em ML (RF-Score-Vs), treinada com mais de 900 mil compostos. Essa ferramenta alcançou uma taxa de acerto de 55,6%, superior a Vina, na identificação de moléculas ativas.

Além disso, Liu *et al.*, (2023) apresentaram o MolFilterGAN, um modelo baseado em GAN, que realiza triagens de bioativos para uma ampla gama de alvos moleculares, superando abordagens convencionais em performance. Na área da farmacocinética e toxicologia preditiva, Adeshina *et al.*, (2023) desenvolveram um conjunto de dados (D-COID) com alto desempenho na identificação de complexos ligante-alvo com afinidade significativa, o que pode ajudar na redução de falsos positivos em triagens *in silico*. Já Chen *et al.*, (2018) relataram que os modelos baseados em DL apresentam melhor desempenho na predição de bioatividade de compostos.

Embora os resultados indiquem avanços, ainda persistem limitações importantes. A qualidade dos modelos depende da diversidade e representatividade dos dados de treinamento, o que pode introduzir viés. Ademais, muitas ferramentas ainda carecem de validação robusta (ensaios *in vitro* e *in vivo*). Diante desses resultados, é evidente que a bioinformática e IA ocupam posição importante na pesquisa científica contemporânea, otimizando os processos de identificação e priorização de potenciais candidatos terapêuticos.

Título do artigo	Abordagens/ Modelos	Principais achados	Referências
<i>Deep learning pipeline for accelerating virtual screening in drug discovery.</i>	GNN + DL	DL reduziu o tempo de triagem virtual e apresentou mais combinações de ligantes e alvos	Noor <i>et al.</i> , 2024
<i>Accurate prediction of protein–ligand interactions by combining physical energy functions and graph-neural networks</i>	ML + GNN	Foram propostos 3 modelos de GNN. Dois modelos de regressão foram treinados para prever afinidade e um modelo para prever os desvios e incertezas nessa ligação. Esse modelo criado superou os já existentes e apresentou melhor desempenho na identificação dos acertos.	Hong <i>et al.</i> , 2024
<i>MolFilterGAN: a progressively augmented generative adversarial network for triaging AI-designed molecules</i>	GAN	Um novo método de filtragem molecular foi proposto, baseado em GAN (MolFilterGAN). O MolFilterGAN superou as abordagens convencionais ao possibilitar a triagem de bioativos em uma ampla gama de alvos moleculares	Liu <i>et al.</i> , 2023.
<i>Machine learning classification can reduce false positives in structure-based virtual screening</i>	ML	Foi construído um conjunto de dados de treinamento (D-COVID) capaz de identificar complexos (ligante-alvo) com alta afinidade	Adeshina <i>et al.</i> , 2023
<i>Machine Learning Models Combined with Virtual Screening and Molecular Docking to Predict Human Topoisomerase I Inhibitors</i>	ML + Docking	Foram identificadas 6 estruturas de moléculas, não relatadas na literatura antes da pesquisa, inibidores da Topoisomerase 1	Li <i>et al.</i> , 2019
<i>The rise of deep learning in drug discovery</i>	ML + DP	A DL apresenta melhor desempenho na predição de bioatividade	Chen <i>et al.</i> , 2018
<i>Performance of machine-learning scoring functions in structure-based virtual screening</i>	ML	Os autores uma nova função de pontuação (RF-Score-VS), treinada com 15.426 moléculas ativas e 893.897 moléculas inativas acopladas a um conjunto de 102 alvos moleculares. Quando comparado com o Vina, o RF-Score-VS atingiu taxa de acerto de 55,6%.	Wójcikowski <i>et al.</i> , 2017
<i>AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery</i>	CNIN + QSAR	Os autores apresentaram a primeira rede neural convolucional profunda (AtomNet). A AtomNet superou as abordagens anteriores de encaixe com taxa de 57,8% de alvos	Wallach <i>et al.</i> , 2015

Tabela 1. Principais estudos revisados sobre a aplicação da IA e inteligência artificial na descoberta de fármacos publicados de 2015 a 2025

DL: Deep Learning

ML: Machine Learning

IA: Inteligência Artificial

GNN: Redes Neurais Gráficas;

GAN: Relações Adversárias Generativas

QSAR: Relação Quantitativa Estrutura-Atividade

CNIN: Redes Neurais Convolucionais

CONCLUSÃO

Os achados nesse estudo evidenciam que a IA e bioinformática otimizam a pesquisa científica na etapa de triagem de prospecção de compostos com atividade biológica de relevância clínica. Essa convergência de tecnologias proporciona predições de ações biológicas, triagens virtuais mais rápidas e geração de novas moléculas com atividades desejadas. Ferramentas como redes neurais, algoritmos de aprendizado profundo e técnicas como QSAR e *docking* molecular se configuram como grandes aliadas nos estudos científicos. No entanto, ainda existem as seguintes lacunas: a grande maioria dos modelos carecem de validação experimental (estudos pré-clínicos e clínicos), ou seja, permanecem restritos ao ambiente *in silico*, além das bases de dados limitadas, o que impacta a reprodutibilidade.

REFERÊNCIAS

Adeshina, Y. O. *et al.* Machine learning classification can reduce false positives in structure-based virtual screening. **Proceedings of the National Academy of Sciences of the United States of America (PNAS)** v.117, n.31, p.1-12, 2020.

Chen, H. *et al.* The rise of deep learning in drug discovery. **Drug Discovery Today**. v.23, n.6, p.1241-1250, 2018.

De Souza, C. A. *et al.* The role of artificial intelligence in the discovery and development of drugs. **Brazilian Journal of Implantology and Health Sciences**. v.6, n.11, p.650-663, 2024.

Dimasi, J. A. *et al.* Innovation in the Pharmaceutical industry: New estimates of RD costs. **Journal of Health Economics**. v.47, p. 20-33, 2016.

Gentile, F. *et al.* Artificial intelligence-enabled virtual screening of ultra-large Chemical libraries with deep docking. **Nature Protocols**. v.17, n.3, p.672-697, 2022.

Gupta, R. *et al.* Artificial intelligence to deep learning: machine intelligence approach for drug discovery. **Molecular diversity**. v.25, p.1315-1360, 2021.

Hong, Y. *et al.* Accurate prediction of protein–ligand interactions by combining physical energy functions and graph-neural networks. **Journal of Cheminformatics**. v.16, n.121, p.1-11, 2024.

Karim, M. R. *et al.* Explainable AI for bioinformatics: methods, tools and applications. **Briefings in bioinformatics**. v.24, n.5, p.1-22, 2023.

Li, B. *et al.* A survey of artificial intelligence enabled channel estimation methods: recente advance, performace and outlook. **Artificial Intelligence Review**. v.58, n.6, p.1-36, 2025.

Li, B. *et al.* Machine Learning Models Combined with Virtual Screening and Molecular Docking to Predict Human Topoisomerase I Inhibitors. **Molecules**. v.4, n.24, 1-12, 2019.

Liu, X. *et al.* MolFilterGAN: a progressively augmented generative adversarial network for triaging AI-designed molecules. **Journal of Cheminformatics**. v.15, n.42, 1-15, 2023.

- Neto, C. S. *et al.* Technological advances in drug planning and development: a review. **Contemporary Journal**. v.4, n. 4, p. 1-17, 2024.
- Noor, F. *et al.* Deep learning pipeline for accelerating virtual screening in drug discovery. **Scientific Reports**. v.14, n.1, p.1-19, 2024.
- Pereira, A. C. *et al.* A bioinformática na pesquisa odontológica brasileira. **Journal of Health Informatics**. v.5, n.4, p. 1-5, 2013.
- Ruff, K. M.; Pappu, R. V. AlphaFold and implications for intrinsically disordered proteins. **Journal of molecular biology**. v.433, n. 20, p.1-11, 2021
- Ruyck, J. *et al.* Molecular docking as popular tool in drug design, na *in silico* travel. **Advances and Applications in Bioinformatics and Chemistry**. v.2, n.1-11, 2016.
- Shiri, F. M. *et al.* A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. **Journal of Artificial Intelligence**. v.1, p.1-13, 2024.
- Singh, S. *et al.* Molecular docking and molecular dynamics simulation. **Bioinformatics**. v.3, p. 291-304, 2022.
- Wallach, I. *et al.* AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. **Computer Science**. v.15, n.2, p.1-11, 2015.
- Wójcikowski, M. *et al.* Performance of machine-learning scoring functions in structure-based virtual screening. **Scientific Reports**. v.7, p.1-10, 2017.