

Técnicas de *machine learning* aplicadas a predição de dados biomoleculares

Machine learning techniques applied to biomolecular data prediction

Técnicas de aprendizaje automático aplicadas a la predicción de datos biomoleculares

Recebido: 04/02/2026 | Revisado: 05/03/2026 | Aceitado: 06/03/2026 | Publicado: 07/03/2026

Yasmin Vitória Da Silva Pedroso

ORCID: <https://orcid.org/0009-0007-3255-9158>

Universidade Anhembi Morumbi, Brasil

E-mail: yasmin.vitoria532@gmail.com

Jéssica Karoline da Cunha

ORCID: <https://orcid.org/0009-0000-8842-0584>

Universidade Anhembi Morumbi, Brasil

E-mail: jescakaroline@gmail.com

Joyce Nascimento Santos

ORCID: <https://orcid.org/0000-0002-3504-8368>

Universidade Anhembi Morumbi, Brasil

E-mail: santosjoyce@gmail.com

Paulo Bandiera Paiva

ORCID: <https://orcid.org/0000-0001-9409-3970>

Universidade Federal de São Paulo, Brasil

E-mail: paiva@unifesp.br

Resumo

O objetivo do presente estudo foi analisar a aplicação de técnicas de aprendizado de máquina *machine learning* (ML) na avaliação de dados de predição biomoleculares na área genética, comparando diferentes abordagens quanto à precisão, eficiência e aplicabilidade na análise de dados biomoleculares, bem como identificar e discutir as principais limitações e avanços dessas técnicas. Na era da Indústria 4.0, observa-se um crescimento expressivo do volume de dados digitais, incluindo informações de saúde, mídias sociais e Internet das Coisas, cuja análise demanda o uso de inteligência artificial (IA), especialmente ML e aprendizado profundo *deep learning* (DL). A IA possibilita que sistemas computacionais realizem tarefas tradicionalmente associadas à inteligência humana, sendo amplamente aplicada à análise de *big data* em saúde. Entre suas vantagens destacam-se a flexibilidade, escalabilidade e a capacidade de integrar dados heterogêneos, como informações demográficas, exames laboratoriais, imagens e textos clínicos, contribuindo para diagnóstico, prognóstico e predição de riscos. O ML apresenta ampla aplicação na biologia molecular e na descoberta de fármacos, enquanto o DL, baseado em redes neurais profundas, mostra-se particularmente eficiente na análise de grandes volumes de dados, como no diagnóstico de câncer e na identificação de alvos terapêuticos. Entretanto, desafios persistem, incluindo a necessidade de infraestrutura computacional robusta, limitada interpretabilidade dos modelos, elevados custos computacionais e questões éticas relacionadas à privacidade e segurança dos dados.

Palavras-chave: Aprendizado de máquina; Aprendizado profundo; IA.

Abstract

The objective of this study was to analyze the application of machine learning (ML) techniques in the evaluation of biomolecular prediction data in the genetic field, comparing different approaches regarding accuracy, efficiency, and applicability in the analysis of biomolecular data, as well as identifying and discussing the main limitations and advances of these techniques. In the era of Industry 4.0, there is a significant growth in the volume of digital data, including health information, social media, and the Internet of Things, whose analysis demands the use of artificial intelligence (AI), especially ML and deep learning (DL). AI enables computational systems to perform tasks traditionally associated with human intelligence, being widely applied to the analysis of big data in healthcare. Among its advantages are flexibility, scalability, and the ability to integrate heterogeneous data, such as demographic information, laboratory tests, images, and clinical texts, contributing to diagnosis, prognosis, and risk prediction. ML has wide application in molecular biology and drug discovery, while DL, based on deep neural networks, proves particularly efficient in analyzing large volumes of data, such as in cancer diagnosis and the identification of therapeutic targets. However, challenges persist, including the need for robust computational infrastructure, limited interpretability of models, high computational costs, and ethical issues related to data privacy and security.

Keywords: Machine learning; Deep learning; AI.

Resumen

El objetivo de este estudio fue analizar la aplicación de técnicas de aprendizaje automático (ML) en la evaluación de datos de predicción biomolecular en el campo de la genética, comparando diferentes enfoques en cuanto a precisión, eficiencia y aplicabilidad en el análisis de datos biomoleculares, así como identificando y discutiendo las principales limitaciones y avances de estas técnicas. En la era de la Industria 4.0, existe un crecimiento significativo en el volumen de datos digitales, incluyendo información de salud, redes sociales e Internet de las Cosas, cuyo análisis exige el uso de inteligencia artificial (IA), especialmente ML y aprendizaje profundo (DL). La IA permite a los sistemas computacionales realizar tareas tradicionalmente asociadas con la inteligencia humana, siendo ampliamente aplicada al análisis de big data en salud. Entre sus ventajas se encuentran la flexibilidad, escalabilidad y la capacidad de integrar datos heterogéneos, como información demográfica, pruebas de laboratorio, imágenes y textos clínicos, contribuyendo al diagnóstico, pronóstico y predicción de riesgos. El aprendizaje automático (ML) tiene amplias aplicaciones en biología molecular y descubrimiento de fármacos, mientras que el aprendizaje automático (AA), basado en redes neuronales profundas, resulta especialmente eficiente en el análisis de grandes volúmenes de datos, como en el diagnóstico del cáncer y la identificación de dianas terapéuticas. Sin embargo, persisten desafíos, como la necesidad de una infraestructura computacional robusta, la limitada interpretabilidad de los modelos, los altos costos computacionales y los problemas éticos relacionados con la privacidad y la seguridad de los datos.

Palabras clave: Aprendizaje automático; Aprendizaje profundo; IA.

1. Introdução

Na era atual da Quarta Revolução Industrial (4RI ou Indústria 4.0), o mundo digital possui uma riqueza de dados, como dados da Internet das Coisas (IoT), dados de cibersegurança, dados móveis, dados empresariais, dados de mídias sociais, dados de saúde. Para analisar esses dados de forma inteligente e desenvolver aplicações inteligentes e automatizadas correspondentes, o conhecimento de inteligência artificial (IA), particularmente, aprendizado de máquina ou *machine learning* (ML), é fundamental (Theodosiou & Read, 2023).

Existem muitos tipos de algoritmos de aprendizado de máquina, como aprendizado supervisionado, não supervisionado, semi-supervisionado e por reforço. Além disso, o aprendizado profundo, que faz parte de uma família mais ampla de métodos de aprendizado de máquina, pode analisar dados em larga escala de forma inteligente (Theodosiou & Read, 2023).

A inteligência artificial (IA) refere-se ao uso de algoritmos e modelos que permitem às máquinas executar tarefas que normalmente exigem inteligência humana. As primeiras aplicações de IA na área da saúde baseavam-se em regras de especialistas, nas quais o conhecimento e a experiência de especialistas humanos eram usados para formular regras, como por exemplo: requisitar hemocultura para paciente febril. O primeiro sistema desse tipo relacionado a infecções humanas foi o *MYCIN*, introduzido pela Universidade de *Stanford* na década de 1970 para sugerir antibióticos para infecções graves com base em dados específicos do paciente, apesar de esse sistema ter sido amplamente utilizado na prática clínica. Embora as regras de especialistas sejam amplamente utilizadas na medicina clínica atualmente, incluindo a maioria dos sistemas de apoio à decisão clínica e calculadoras, elas geralmente não são incluídas nas discussões contemporâneas sobre IA, que se concentram principalmente em aprendizado de máquina e, mais recentemente, em aprendizado profundo (Tran *et al.*, 2021).

A análise de *big data* por meio de aprendizado de máquina oferece vantagens consideráveis para a assimilação e avaliação de grandes quantidades de dados complexos na área da saúde. As vantagens do aprendizado de máquina incluem flexibilidade e escalabilidade em comparação com os métodos bioestatísticos tradicionais, o que o torna aplicável a muitas tarefas, como estratificação de risco, diagnóstico e classificação, e previsões de sobrevida. Outra vantagem dos algoritmos de aprendizado de máquina é a capacidade de analisar diversos tipos de dados (por exemplo, dados demográficos, resultados de exames laboratoriais, dados de imagem e anotações em texto livre de médicos) e incorporá-los em previsões de risco de doenças, diagnóstico, prognóstico e tratamentos adequados (Wessels *et al.*, 2024).

Aprendizado de máquina ou *machine learning* (ML) é uma subárea da inteligência artificial que permite que sistemas aprendam e tomem decisões baseadas em dados, sem a necessidade de programação explícita para cada tarefa. Essa abordagem é amplamente utilizada em diversos campos, como na biologia molecular e na descoberta de fármacos, onde é aplicada para identificar padrões complexos em grandes volumes de dados, otimizando processos como o acoplamento molecular ou *docking molecular* (DM) e a descoberta de novos medicamentos (Crampon *et al.*, 2022); (Dias *et al.*, 2021). Dentro do campo de ML, existe a aprendizagem profunda ou *deep learning* (DL), que é uma subcategoria que utiliza redes neurais profundas para analisar dados em camadas, permitindo a detecção de características mais abstratas e complexas. Enquanto o ML tradicional trabalha com algoritmos mais simples e necessita de maior intervenção humana na seleção de características, o DL é capaz de automatizar esse processo e tem sido especialmente eficaz em tarefas que envolvem grandes volumes de dados, como no diagnóstico de câncer e na previsão de alvos terapêuticos (Gupta *et al.*, 2021); (Ngiam & Khor, 2019). A principal diferença entre ML e DL está na profundidade e complexidade dos modelos utilizados, sendo que o DL utiliza múltiplas camadas de processamento, o que o torna mais potente em problemas complexos, como a integração de dados genômicos e a personalização de terapias (Elgart *et al.*, 2021); (Issa *et al.*, 2021).

O aprendizado de máquina também tem se destacado na descoberta de novos medicamentos, sendo utilizado para prever a eficácia e a segurança de compostos terapêuticos antes mesmo de experimentos laboratoriais (Gupta *et al.*, 2021); (Hudson, 2021). O ML representa uma abordagem poderosa e flexível para resolver problemas complexos em diversas áreas da ciência, desde o diagnóstico de doenças até a edição genética (Issa *et al.*, 2021); (Sarker, 2021).

Em termos de precisão, as abordagens de ML têm demonstrado avanços significativos em diversas aplicações clínicas. Métodos tradicionais de ML, como aqueles utilizados para a previsão de interações entre proteínas e ligantes, já apresentam alta precisão ao otimizar processos de DM, reduzindo tanto o tempo quanto o custo envolvido (Crampon *et al.*, 2022); (Dias *et al.*, 2021). No entanto, o DL, tem se destacado ainda mais quando a tarefa envolve dados mais complexos e volumosos. A utilização de redes neurais profundas permite uma análise mais detalhada, como visto em pesquisas de diagnósticos oncológicos e em tratamentos personalizados, onde o DL supera as técnicas de ML em termos de precisão e capacidade preditiva (Gupta *et al.*, 2021); (Ngiam & Khor, 2019). Além disso, no redirecionamento de fármacos para o tratamento do câncer, o uso de DL mostrou-se superior ao ML convencional ao melhorar a identificação de alvos moleculares e prever respostas ao tratamento com maior exatidão pois ele integra dados multimodais para melhorar a compreensão de processos biológicos e oferece soluções mais eficazes para o diagnóstico e tratamento de doenças (Elgart *et al.*, 2021); (Issa *et al.*, 2021); (Ngiam & Khor, 2019). Essas melhorias refletem a crescente importância do DL em áreas que demandam precisão mais apurada, como no caso de terapias gênicas utilizando *CRISPR*, onde a previsão de atividades com máxima eficiência no alvo (*on-target*) e as fora do alvo (*off-target*) foram significativamente aprimoradas (Sarker, 2021).

No entanto, a literatura global até o momento varia significativamente em escopo, qualidade e público-alvo, desde periódicos especializados em bioinformática repletos de jargões até artigos de interesse geral muito mais abrangentes. As conclusões dos autores também variam drasticamente, desde promessas quase evangelizadoras de que a IA melhorará todos os aspectos da medicina até profecias apocalípticas de que a IA substituirá completamente os médicos humanos. Diante desse tsunami, pode ser um desafio para os profissionais de saúde que atuam na linha de frente do combate a infecções (incluindo médicos infectologistas e microbiologistas clínicos) determinar a relevância da IA para sua própria prática clínica (Tran *et al.*, 2021).

Desta forma, o objetivo do presente estudo foi analisar a aplicação de técnicas de ML na avaliação de dados de predição biomoleculares na área genética para comparar as diferentes técnicas de ML em termos de precisão, eficiência e

aplicabilidade na análise de dados biomoleculares, identificar e discussão das limitações na aplicação de técnicas de ML na análise de dados biomoleculares e os principais avanços em técnicas de ML aplicadas à análise de dados biomoleculares.

2. Metodologia

Realizou-se uma pesquisa documental de fonte indireta em artigos da literatura por meio de revisão bibliográfica (Snyder, 2019; Risemberg *et al.*, 2026) num estudo de natureza quantitativa em relação aos 11 (onze) artigos selecionados para compor o corpus da pesquisa e, qualitativa em relação às discussões (Pereira *et al.*, 2018) do tipo específico de revisão integrativa (Crossetti, 2012).

A seleção do material foi conduzida por meio de pesquisas nas bibliotecas virtuais *SciELO* (*Scientific Electronic Library Online*) e *PubMed*, utilizando os descritores “aprendizagem de máquina”, “predição”, “biologia molecular” e “genética”, combinados conforme a relevância temática.

A curadoria manual envolveu a leitura dos títulos, resumos e, quando necessário, do texto completo, com o objetivo de identificar estudos diretamente relacionados à aplicação de técnicas de aprendizagem de máquina em contextos de biologia molecular e genética.

Como critérios de inclusão, foram considerados apenas artigos publicados em língua inglesa, a partir do ano de 2019, que apresentassem aplicação ou discussão metodológica relevante ao tema proposto. Foram excluídos trabalhos duplicados, estudos fora do escopo temático e publicações que não apresentassem contribuição direta para os objetivos da revisão.

3. Resultados e Discussão

Dentre a pesquisa realizada em bibliotecas virtuais utilizando os descritores e os critérios de inclusão/exclusão utilizados, o total de 134 artigos foram obtidos. Sendo 133 na *PubMed* (<https://pubmed.ncbi.nlm.nih.gov/>) e 1 na *SciELO – Scientific Electronic Library Online* (<http://www.scielo.org/index.php>). Uma curadoria manual foi realizada selecionando exclusivamente estudos que trabalhavam as técnicas de aprendizagem de máquina aplicadas a análise molecular preditiva em humanos, sendo revisão de literatura ou experimental, totalizando 11 estudos científicos.

O Quadro 1 abaixo lista artigos científicos que descrevem a aplicabilidade de *machine learning* a predição biomolecular.

Quadro 1 - Artigos científicos selecionados para aplicação neste estudo.

Autor/Ano	Título	Objetivo	Medidas Realizadas	Achados
1. CRAMPON, K., et al., 2022	Machine-learning methods for ligand–protein molecular docking.	Identificação de abordagens que possam superar limitações de métodos tradicionais, como velocidade e precisão, integrando técnicas de ML para aprimorar a previsão da afinidade de ligação e a posição do ligante no alvo proteico.	Revisão de literatura.	Melhoria da precisão de predições de <i>Docking</i> . Redução de tempo e custo. Futuras direções em <i>learning transfer</i> .

2. DIAS, M. F. R. et al., 2021.	Application of PolyPRep tools on HIV protease polyproteins using molecular docking.	Investigar e avaliar o desempenho das ferramentas <i>PolyPRep</i> para a preparação de poliproteínas virais, especificamente as proteases do HIV, e aplicar essas ferramentas no contexto de <i>docking</i> molecular.	Experimental.	Eficácia das ferramentas <i>PolyPRep</i> . Processo de <i>Docking</i> Molecular. Relevância na descoberta de fármacos. Desempenho computacional e aplicabilidade.
3. ELGART, M. et al., 2021.	Machine and deep learning in molecular and genetic aspects of sleep research.	Discussão de descobertas existentes e as metodologias relacionadas ao uso de aprendizado de máquina e aprendizado profundo na pesquisa do sono.	Revisão de literatura.	Avanços na compreensão genética do sono. Integração de dados multimodais.
4. GUPTA, R. et al., 2021.	Artificial intelligence to deep learning: machine intelligence approach for drug discovery.	Aplicabilidade de inteligência artificial (IA) e aprendizado profundo (<i>deep learning</i>) no contexto da descoberta de fármacos.	Revisão de literatura.	<i>Deep Learning</i> na descoberta de fármacos. Previsão e predição biológica. Redes Neurais na modelagem molecular.
5. HUDSON, I. L., 2021.	Data integration using advances in machine learning in drug discovery and molecular biology.	Avanços em aprendizado de máquina aplicados à descoberta de fármacos e da biologia molecular.	Experimental.	Desenvolvimento de fármacos. Integração de dados. Modelos de <i>machine learning</i> na prática biomédica.
6. ISSA, N. T., et al. 2021.	Machine and deep learning approaches for cancer drug repurposing.	Aplicação de técnicas de <i>machine learning</i> e <i>deep learning</i> ao redirecionamento de fármacos para o tratamento do câncer.	Revisão de literatura.	Eficiência no redirecionamento de medicamentos. Identificação de alvos moleculares. Modelagem preditiva de resposta ao tratamento. Redução de custo e tempo no desenvolvimento de fármacos.
7. TRAN, K. A. et al., 2021.	Deep learning in cancer diagnosis, prognosis and treatment selection.	Aplicação de técnicas de <i>deep learning</i> no contexto da oncologia, especificamente nas áreas de diagnóstico, prognóstico e seleção de tratamentos para o câncer.	Revisão de literatura.	<i>Deep learning</i> para o diagnóstico do câncer. Tratamento personalizado.
8. WESSELS, H-H. et al., 2024.	Prediction of on-target and off-target activity of CRISPR–Cas13d guide RNAs using deep learning.	Desenvolver e apresentar um modelo de <i>deep learning</i> para previsão on-target e off-target de RNAs-guia utilizados no sistema de edição genética <i>CRISPR-Cas13d</i> .	Experimental.	Desenvolvimento de modelo preditivo. Melhoria de previsão off-target. Validação experimental. Aplicabilidade de terapias gênicas.

9. SARKER, H. I., et al., 2021.	Machine Learning: Algorithms, Real-World Applications and Research Directions.	Apresentar uma visão abrangente de algoritmos de <i>Machine Learning</i> (ML) e discutir como eles são aplicados em diferentes domínios do mundo real como segurança cibernética, cidades inteligentes, saúde, agricultura, e-commerce, entre outros. O estudo também busca destacar desafios enfrentados e direções futuras de pesquisa na área.	Revisão de literatura.	Maior integração entre ML e aplicações reais. Desenvolvimento de métodos para reduzir limitações atuais.
10. THEODOSIOU, A. A., et al., 2023.	Artificial intelligence, machine learning and deep learning: Potential resources for the infection clinician.	Resumir aplicações recentes e potenciais da inteligência artificial (AI), Machine Learning (ML) e Deep Learning no contexto da pesquisa e prática clínica de infecções.	Revisão de literatura.	IA/ML úteis em diagnóstico, imagens e decisão clínica. Apoiam vigilância e controle de surtos. Falta validação no mundo real. Modelos pouco transparentes. Risco de vieses e falhas metodológicas.
11. NGIAN, K. Y., et al., 2019.	Big data and machine learning algorithms for health-care delivery.	Revisar como big data e algoritmos de Machine Learning podem ser utilizados na área de saúde para melhorar a análise de grandes volumes de dados clínicos e apoiar a prática médica.	Revisão de literatura.	Maior flexibilidade e escalabilidade que métodos estatísticos tradicionais. Capacidade de integrar múltiplos tipos de dados. Aplicações em predição de risco, diagnóstico, prognóstico e suporte ao tratamento. Principais desafios na área médica.

Fonte: Elaborado pelos Autores.

Issa *et al.* (2021) discutem que O uso de ML e DL na área da saúde vem transformando a prática médica e os serviços de saúde digital, classificando essas técnicas como essenciais para o avanço da medicina personalizada. De acordo com estudos recentes, as principais aplicações de ML e DL no setor incluem diagnóstico assistido por inteligência artificial (IA), predição e personalização de tratamentos, além do desenvolvimento de novos fármacos. Ngiam e Khor (2019) discutem que essas intervenções estão alinhadas com o conceito de "saúde digital de precisão", onde a capacidade de diagnóstico precoce e a previsão de respostas terapêuticas impactam diretamente a qualidade dos cuidados e a precisão clínica oferecida aos pacientes.

Gupta, Srivastava, Sahu, Tiwari, Ambasta e Kumar (2021) descrevem que algumas plataformas específicas, como *Google DeepVariant*, *DNAnexus*, *DeepChem* e *CRISPR-Cas13d*, são baseadas em ML para análise e integração de dados genômicos e moleculares, oferecendo resultados mais rápidos e precisos. A possibilidade de otimização do fluxo de trabalho em laboratórios e clínicas, por meio da análise automatizada de grandes volumes de dados, contribui para a eficiência no diagnóstico e nas intervenções terapêuticas.

O DL, subárea de ML que utiliza redes neurais profundas para processar dados em múltiplas camadas, é particularmente eficiente em tarefas que envolvem grandes volumes e complexidade de dados. Por exemplo, no campo da modelagem de DM, que consiste na simulação de interações entre moléculas, como ligantes e proteínas, para prever a

afinidade de ligação e identificar possíveis alvos terapêuticos, o DL proporciona maior precisão e menor tempo de análise. Crampon, Giorkallos, Deldossi, Baud e Steffene (2022) mencionam que as técnicas de DL superam métodos tradicionais de *docking* ao reduzir custos e aumentar a velocidade, benefícios que têm relevância especial na descoberta de novos fármacos. De forma similar, o trabalho de Dias *et al.* (2021) mostrou a eficiência de ferramentas específicas como o *PolyPrep*, baseada em ML, na análise das proteases do HIV, o que sugere o potencial de ML e DL na descoberta de antivirais inovadores.

No campo da oncologia, o DL se destaca por proporcionar diagnósticos e prognósticos mais detalhados e precisos. O estudo de Ngiam e Khor (2019) evidenciou que, ao processar imagens médicas e integrar dados clínicos, o DL é capaz de oferecer recomendações mais personalizadas de tratamento para pacientes com câncer, melhorando significativamente a precisão em comparação aos métodos tradicionais. Esse nível de precisão auxilia na escolha de tratamentos mais adequados para cada caso, o que é fundamental em doenças complexas como o câncer, onde cada decisão pode impactar diretamente o sucesso terapêutico.

Hudson (2021) discute que apesar das vantagens evidentes, o uso dessas técnicas enfrenta desafios. A coleta e o processamento de grandes volumes de dados de alta qualidade são requisitos fundamentais para o bom desempenho dos modelos de DL, representando um alto custo computacional e exigindo infraestrutura robusta. Além disso, a interpretabilidade dos modelos de DL, que frequentemente operam como "caixas-pretas", é uma questão em aberto, dificultando a compreensão dos mecanismos por trás das previsões feitas, o que limita sua aceitação em áreas onde a rastreabilidade dos processos de decisão é essencial.

Sarker (2021) afirma que no campo da genética, o desenvolvimento de modelos preditivos de DL aplicados a sistemas de edição gênica *CRISPR-Cas13d*, tem avançado ao prever atividades *off-target*, ou seja, efeitos fora do alvo desejado. O *off-target* representa um risco na edição genética, pois pode causar mutações não intencionais. A aplicação de DL neste contexto permite uma análise precisa das interações e dos potenciais efeitos colaterais, minimizando os riscos associados a terapias gênicas e promovendo avanços na segurança de tratamentos personalizados.

4. Conclusão

As técnicas de *machine learning* (ML) e *deep learning* (DL) vêm desempenhando um papel transformador na biologia molecular e na medicina personalizada, especialmente na análise de dados biomoleculares e na descoberta de fármacos. A inteligência artificial tem possibilitado o desenvolvimento de diversas ferramentas com potencial aplicação na pesquisa e no tratamento de infecções humanas. À medida que mais dados de ensaios clínicos, regulamentações e padrões de notificação se tornem disponíveis, a IA tende a contribuir significativamente para o aumento da precisão e da eficiência em diagnósticos laboratoriais, no suporte clínico, na saúde pública e na pesquisa em doenças infecciosas. Entretanto, não se espera que essas tecnologias substituam, em breve, os clínicos e pesquisadores que atuam diretamente na linha de frente do combate às infecções.

Apesar dos avanços, ainda persistem desafios importantes, como a necessidade de infraestrutura computacional robusta e a limitada interpretabilidade dos modelos, o que restringe a aceitação do DL em aplicações nas quais a transparência é essencial. Por outro lado, o uso de DL para prever efeitos *off-target* em sistemas de edição genética baseados em *CRISPR*, como o *CRISPR-Cas13d*, mostra-se promissor na mitigação de riscos associados às terapias gênicas.

Para que as ferramentas de aprendizado de máquina sejam efetivamente aplicadas na área da saúde, é necessário enfrentar diversas limitações e considerar questões-chave relacionadas à implementação clínica e à ética na prestação de cuidados. A aplicação dessas técnicas exige etapas rigorosas de pré-processamento de dados, treinamento adequado dos modelos e refinamento dos sistemas de acordo com problemas clínicos específicos. Além disso, são fundamentais as

considerações éticas, que incluem implicações médico-legais, a compreensão dos sistemas por parte dos profissionais de saúde, bem como a garantia da privacidade e da segurança dos dados. Nesse contexto, esta revisão discute os principais benefícios e desafios do uso de *big data* e aprendizado de máquina na área da saúde.

Referências

- Crampon, K., Giorkallos, A., Deldossi, M., Baud, S., & Steffanel, L. A. (2022). *Machine-learning methods for ligand–protein molecular docking*. *Drug Discovery Today*, 27(1), 151–164. <https://doi.org/10.1016/j.drudis.2021.09.007>
- Crossetti, M. G. O. (2012). Revisão integrativa de pesquisa na enfermagem o rigor científico que lhe é exigido. *Rev Gaúcha Enferm.* 33(2):8-9.
- Dias, M. F. R., et al. (2021). *Application of PolyPRep tools on HIV protease polyproteins using molecular docking*. *Brazilian Journal of Biology*, 84, e245592. <https://doi.org/10.1590/1519-6984.245592>
- Elgart, M., Redline, S., & Sofer, T. (2021). *Machine and deep learning in molecular and genetic aspects of sleep research*. *Neurotherapeutics*, 18(1), 228–243. <https://doi.org/10.1007/s13311-021-01014-9>
- Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). *Artificial intelligence to deep learning: Machine intelligence approach for drug discovery*. *Molecular Diversity*, 25, 1315–1360. <https://doi.org/10.1007/s11030-021-10217-3>
- Hudson, I. L. (2021). *Data integration using advances in machine learning in drug discovery and molecular biology*. In *Artificial Neural Networks* (pp. 167–184). Springer. https://doi.org/10.1007/978-1-0716-0826-5_7
- Issa, N. T., et al. (2021). *Machine and deep learning approaches for cancer drug repurposing*. In *Seminars in Cancer Biology* (pp. 132–142). Academic Press. <https://doi.org/10.1016/j.semcancer.2019.12.011>
- Ngiam, K. Y., & Khor, I. W. (2019). *Big data and machine learning algorithms for health-care delivery*. *The Lancet Oncology*, 20(5), 262–273. [https://doi.org/10.1016/S1470-2045\(19\)30149-4](https://doi.org/10.1016/S1470-2045(19)30149-4)
- Pereira, A. S. et al. (2018). *Metodologia da pesquisa científica*. Editora da UFSM. https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf?sequence=1.
- Risemberg, R. I. C. et al. (2026). A importância da metodologia científica no desenvolvimento de artigos científicos. *E-Acadêmica*. 7(1), e0171675. <https://eacademica.org/eacademica/article/view/675>.
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*. 104, 333–9. <https://doi.org/10.1016/j.jbusres.2019.07.039>.
- Sarker, I. H. (2021). *Machine learning: Algorithms, real-world applications and research directions*. *SN Computer Science*, 2(3), 1–21. <https://doi.org/10.1007/s42979-021-00592-x>
- Theodosiou, A. A., & Read, R. C. (2023). *Artificial intelligence, machine learning and deep learning: Potential resources for the infection clinician*. *Journal of Infection*, 87(4). <https://www.sciencedirect.com/science/article/pii/S0163445323003791>
- Tran, K. A., et al. (2021). *Deep learning in cancer diagnosis, prognosis and treatment selection*. *Genome Medicine*, 13, Article 1–17. <https://doi.org/10.1186/s13073-021-00968-x>
- Wessels, H. H., Stim, A., Méndez-Mancilla, A., Kim, E. J., Hart, S. K., Knowles, D. A., & Sanjana, N. E. (2024). *Prediction of on-target and off-target activity of CRISPR–Cas13d guide RNAs using deep learning*. *Nature Biotechnology*, 42(4), 628–637. <https://doi.org/10.1038/s41587-023-01830-8>