

MACHINE LEARNING NA PREDIÇÃO DE DOENÇAS A PARTIR DO MICROBIOMA INTESTINAL

MACHINE LEARNING IN DISEASE PREDICTION BASED ON THE GUT MICROBIOME

Francileia Nogueira Albino², Amanda Ravena Soares Veras¹, Alice Maria Lima Monteles¹, Camila Nogueira Calland¹, Érica Geovanna de Carvalho Soares¹, Fabianne Ribeiro Rocha¹

¹ Discente do curso de Medicina do Centro Universitário Afya Teresina

² Docente do curso de Medicina Centro Universitário Afya Teresina

E-mail: fabianneribeiro868@gmail.com

Eixo Temático: 03

RESUMO

INTRODUÇÃO: O microbioma intestinal apresenta importante interação com o organismo humano, e alterações em sua composição têm sido associadas a diferentes condições clínicas. Nesse contexto, técnicas de *machine learning* têm sido utilizadas para identificar padrões complexos do microbioma e auxiliar na predição e classificação de doenças. **OBJETIVO:** Analisar a utilização de técnicas de *machine learning* na predição de doenças a partir do microbioma intestinal, destacando o desempenho dos modelos empregados e suas principais limitações. **MÉTODO:** Realizou-se revisão integrativa da literatura nas bases PubMed/MEDLINE, Scopus, SciELO e LILACS, considerando estudos publicados entre 2025 e 2026, nos idiomas português, inglês e espanhol. Foram incluídos estudos realizados em seres humanos ou utilizando conjuntos de dados provenientes de populações humanas que investigassem a aplicação de *machine learning* ou *deep learning* associada ao microbioma intestinal para identificação, classificação ou predição de doenças. **RESULTADOS:** Os seis estudos selecionados demonstraram aplicação de diferentes algoritmos de aprendizado de máquina em doenças neurológicas, autoimunes, inflamatórias, dermatológicas e neoplásicas. Foram observados resultados promissores, com AUC de 0,92 na classificação da doença de Parkinson, valores entre 0,75 e 0,99 para diferentes doenças autoimunes, AUC de 0,90 na validação interna e 0,82 na validação externa para câncer colorretal e AUROC de 0,9817 para dermatite atópica. Entretanto, diferenças no tamanho amostral, na composição das populações, nos métodos de análise do microbioma e nas estratégias de validação podem limitar a generalização dos modelos. **CONCLUSÃO:** O *machine learning* apresenta potencial para utilização do microbioma intestinal como ferramenta de predição e classificação de doenças. Contudo, são necessários estudos com maiores populações, padronização metodológica e validação externa para confirmar a aplicabilidade clínica desses modelos.

Palavras-chave: Microbioma intestinal; *Machine learning*; Inteligência artificial; Predição de doenças; Microbiota intestinal.

ABSTRACT

INTRODUCTION: The gut microbiome interacts extensively with the human organism, and alterations in its composition have been associated with several clinical conditions. In this context, *machine learning* techniques have been used to identify complex microbiome patterns and assist in disease prediction and classification. **OBJECTIVE:** To analyze the use of *machine learning* techniques for disease prediction based on the gut microbiome, highlighting the performance of the models used and their main limitations. **METHOD:** An integrative literature review was conducted using the PubMed/MEDLINE, Scopus, SciELO, and LILACS databases, considering studies published between 2025 and 2026 in Portuguese, English, and Spanish. Studies

involving humans or datasets derived from human populations that investigated the application of *machine learning* or *deep learning* associated with the gut microbiome for disease identification, classification, or prediction were included. **RESULTS:** The six selected studies demonstrated the application of different machine learning algorithms to neurological, autoimmune, inflammatory, dermatological, and neoplastic diseases. Promising results were observed, including an AUC of 0.92 for Parkinson's disease classification, values ranging from 0.75 to 0.99 for different autoimmune diseases, an AUC of 0.90 in internal validation and 0.82 in external validation for colorectal cancer, and an AUROC of 0.9817 for atopic dermatitis. However, differences in sample size, population characteristics, microbiome analysis methods, and validation strategies may limit model generalizability. **CONCLUSION:** Machine learning shows potential for using the gut microbiome as a tool for disease prediction and classification. Nevertheless, studies involving larger populations, methodological standardization, and external validation are needed to confirm the clinical applicability of these models.

Keywords: Gut microbiome; *Machine learning*; Artificial intelligence; Disease prediction; Gut microbiota.

INTRODUÇÃO

O microbioma intestinal é constituído por uma comunidade complexa e dinâmica de microrganismos que apresenta interação contínua com o hospedeiro e participação em diferentes processos fisiológicos e patológicos. Alterações na composição e na função dessa comunidade, conhecidas como disbiose, têm sido associadas a doenças gastrointestinais, metabólicas, inflamatórias, autoimunes e neurológicas. Entretanto, a elevada variabilidade interindividual do microbioma e a sobreposição de alterações microbianas entre diferentes doenças dificultam a utilização de marcadores isolados para fins diagnósticos e preditivos¹.

Nesse contexto, o avanço das técnicas de análise computacional possibilitou a aplicação de métodos de *machine learning* aos dados do microbioma intestinal. Esses modelos são capazes de analisar grandes conjuntos de dados e identificar padrões complexos que podem não ser facilmente reconhecidos por abordagens convencionais. Estudos recentes demonstram que algoritmos de aprendizado de máquina podem utilizar características da microbiota intestinal para diferenciar indivíduos saudáveis de pacientes com diferentes condições inflamatórias, autoimunes e neurológicas, além de identificar assinaturas microbianas potencialmente relacionadas à presença de doenças¹⁻³.

A aplicação dessa abordagem tem apresentado resultados promissores em diferentes cenários. Em doenças neurodegenerativas, modelos baseados no microbioma intestinal alcançaram área sob a curva ROC (AUC) de 0,92 para a classificação da doença de Parkinson e valores médios de AUC de 0,889 e 0,927 na predição da incidência de doença de Alzheimer em diferentes conjuntos de dados^{4,5}. Em doenças inflamatórias e autoimunes, algoritmos de *machine learning* também demonstraram capacidade de distinguir diferentes fenótipos a partir

de padrões da microbiota, enquanto modelos aplicados ao câncer colorretal apresentaram desempenho elevado na identificação de indivíduos com doença^{1-3,6}.

Apesar desses resultados, o desempenho dos modelos não depende apenas da escolha do algoritmo, mas também da qualidade e da composição dos dados utilizados para treinamento e validação. Diferenças entre populações, métodos de obtenção dos dados microbiológicos, seleção de características e estratégias de validação podem comprometer a capacidade de generalização dos modelos. Assim, torna-se relevante analisar criticamente as evidências disponíveis sobre a utilização do *machine learning* na predição de doenças a partir do microbioma intestinal, considerando tanto seu desempenho preditivo quanto suas limitações para uma futura aplicação clínica.

MÉTODO

Trata-se de uma revisão integrativa da literatura, desenvolvida a partir da seguinte pergunta norteadora: como técnicas de *machine learning* têm sido utilizadas na predição de doenças a partir do microbioma intestinal e quais são seus principais desempenhos e limitações?

A busca bibliográfica foi realizada nas bases PubMed/MEDLINE, Scopus, SciELO e LILACS, considerando estudos publicados entre 2025 e 2026, nos idiomas português, inglês e espanhol. Foram utilizados os descritores “Gut Microbiome”, “Gut Microbiota”, “Machine Learning”, “Artificial Intelligence”, “Deep Learning”, “Disease Prediction” e “Disease Diagnosis”, combinados pelos operadores booleanos AND e OR.

Foram incluídos estudos que abordassem a aplicação de técnicas de *machine learning* ou *deep learning* associadas a dados do microbioma intestinal para identificação, classificação ou predição de doenças. Foram excluídos estudos que não utilizassem dados do microbioma intestinal, publicações sem aplicação de técnicas de aprendizado de máquina e trabalhos sem relação direta com a questão norteadora.

A seleção dos estudos foi realizada por dois revisores independentes, mediante análise dos títulos, resumos e textos completos, conforme os critérios de elegibilidade. Em caso de divergência, os revisores discutiram os achados para definição dos estudos a serem incluídos.

Após a seleção, foram extraídas informações referentes ao autor, ano de publicação, condição clínica investigada, conjunto de dados, método de caracterização do microbioma, algoritmo empregado, métricas de desempenho e estratégia de validação. Posteriormente, os

dados foram organizados de forma qualitativa e comparativa, permitindo analisar os principais resultados, desempenhos e limitações do *machine learning* na predição de doenças a partir do microbioma intestinal.

RESULTADOS

Os seis estudos selecionados demonstraram a utilização de diferentes técnicas de *machine learning* e *deep learning* para identificar padrões do microbioma intestinal associados a diferentes condições clínicas. Foram investigadas doenças neurológicas, inflamatórias, autoimunes, dermatológicas e neoplásicas, utilizando diferentes estratégias de caracterização da microbiota e algoritmos de aprendizado de máquina.

Boodaghizaji et al. avaliaram padrões da microbiota intestinal associados à doença de Parkinson, doença de Crohn, retocolite ulcerativa e infecção pelo HIV. Foram utilizadas técnicas de *machine learning*, incluindo Random Forest, Support Vector Machine, redes neurais artificiais e redes neurais convolucionais. Os modelos apresentaram acurácia de até 95% na classificação das condições avaliadas e de até 90% na diferenciação entre doença de Crohn e retocolite ulcerativa¹.

An et al. investigaram dez doenças autoimunes utilizando dados de microbioma provenientes de 1.043 pacientes e 911 controles. Foram comparados diferentes algoritmos, incluindo Random Forest, Support Vector Machine, K-Nearest Neighbors, Multi-Layer Perceptron e XGBoost. O XGBoost apresentou valores de AUROC entre 0,75 e 0,99, demonstrando desempenho variável de acordo com a doença investigada².

Na doença de Parkinson, Yu et al. utilizaram uma abordagem baseada em *deep learning* para identificar diferenças na microbiota intestinal entre pacientes e controles. O modelo apresentou acurácia média de 85% e AUC de 0,92, indicando elevada capacidade de discriminação entre os grupos estudados³.

Basgaran et al. utilizaram perfis populacionais do microbioma intestinal para desenvolver modelos de *machine learning* relacionados à incidência da doença de Alzheimer. Foram analisados dois conjuntos de dados, nos quais os modelos apresentaram AUC média de 0,889 e 0,927, respectivamente⁴. Entretanto, diferentemente dos demais estudos, a análise foi realizada em nível populacional, não correspondendo à predição individual do diagnóstico de Alzheimer⁴.

No câncer colorretal, Tsai et al. desenvolveram um modelo baseado em Random Forest utilizando dados do microbioma intestinal para identificação de indivíduos com adenoma e câncer colorretal. O modelo apresentou AUC de 0,90 na validação interna e 0,82 na validação externa, demonstrando manutenção de desempenho satisfatório quando aplicado a um conjunto de dados independente⁵.

Ma et al. investigaram a associação entre o microbioma intestinal e a dermatite atópica por meio de algoritmos de *machine learning*. Entre os modelos avaliados, o Random Forest apresentou acurácia de 94,1%, sensibilidade de 91,3%, especificidade de 100% e AUROC de 0,9817. A análise de interpretabilidade identificou características específicas da microbiota associadas à classificação da doença⁶.

De forma geral, os estudos demonstraram desempenho preditivo elevado, com AUCs superiores a 0,80 em grande parte das análises. Entretanto, foram observadas diferenças importantes quanto ao tamanho das amostras, características das populações, métodos de análise do microbioma e estratégias de validação. A presença de validação externa em apenas parte dos estudos e a necessidade de amostras maiores e mais representativas constituem limitações relevantes para a generalização dos modelos e sua futura aplicação clínica.

CONCLUSÃO

A utilização de técnicas de *machine learning* associadas ao microbioma intestinal apresenta potencial promissor para a predição e classificação de diferentes doenças, com desempenho elevado observado em diferentes condições clínicas. Os estudos analisados demonstraram que a identificação de padrões complexos da microbiota pode contribuir para o desenvolvimento de modelos preditivos aplicáveis a doenças neurológicas, autoimunes, inflamatórias, dermatológicas e neoplásicas.

Entretanto, limitações relacionadas ao tamanho amostral, à heterogeneidade dos dados e à necessidade de validação externa ainda dificultam a aplicação desses modelos na prática clínica. Dessa forma, embora o *machine learning* represente uma ferramenta promissora para a utilização do microbioma como potencial biomarcador, são necessários estudos com populações maiores, padronização metodológica e validação em diferentes populações para confirmar sua aplicabilidade clínica.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Boodaghidizaji M, Jungles T, Chen T, Zhang B, Yao T, Landay A, et al. Machine learning based gut microbiota pattern and response to fiber as a diagnostic tool for chronic inflammatory diseases. *BMC Microbiol.* 2025; 25:353. doi:10.1186/s12866-025-04072-7.
2. An T, Zhang S, Li J, Wang H, Chen L, Shi Y, et al. Gut microbiota analysis reveals microbial signature for multi-autoimmune diseases based on machine learning model. *Front Microbiol.* 2025; 16:1660775. doi:10.3389/fmicb.2025.1660775.
3. Yu B, Zhang H, Zhang M. Deep learning-based differential gut flora for prediction of Parkinson's. *PLoS One.* 2025;20(1):e0310005. doi:10.1371/journal.pone.0310005.
4. Basgaran A, Lymberopoulos E, Burchill E, Reis-Dehabadi M, Sharma N. Machine learning determines the incidence of Alzheimer's disease based on population gut microbiome profile. *Brain Commun.* 2025;7(2):fcaf059. doi:10.1093/braincomms/fcaf059.
5. Tsai YJ, Lyu WN, Liao NS, Chen PC, Tsai MH, Chuang EY. Gut microbiome-based machine learning model for early colorectal cancer and adenoma screening. *Gut Pathog.* 2025; 17:80. doi:10.1186/s13099-025-00750-z.
6. Ma J, Fang Y, Li S, Zeng L, Chen S, Li Z, et al. Interpretable machine learning algorithms reveal gut microbiome features associated with atopic dermatitis. *Front Immunol.* 2025; 16:1528046. doi:10.3389/fimmu.2025.1528046.