

# Modelo de aprendizagem de máquina para o diagnóstico de lesões intraepiteliais de alto grau: uma abordagem para a descoberta de marcadores metabólicos e lipídicos

Giovanna Melo Marques; Universidade Federal do Amazonas; giovannamarques1111@gmail.com;  
Luana Quadros de Souza Leão; Universidade Federal do Amazonas;  
Adriano de Souza Carolino; Universidade Federal do Amazonas;  
Anderson Ferreira Gonçalves; Fundação Centro de Controle de Oncologia do Estado do Amazonas;  
Zeliene Araújo de Souza; Fundação Centro de Controle de Oncologia do Estado do Amazonas;  
Rafael Garrett da Costa; Universidade Federal do Rio de Janeiro; Rio de Janeiro  
Marina Amaral Alves; Universidade Federal do Rio de Janeiro; Rio de Janeiro;  
Priscila Ferreira de Aquino; Instituto Leônidas e Maria Deane - Fiocruz Amazônia.

## 1. Introdução

O câncer de colo de útero (CCU) é uma das patologias de maior impacto na saúde da população feminina. Especialmente no Brasil, esta neoplasia é a terceira dentre os cânceres mais diagnosticados, os quais estimam-se 15,38 novos casos a cada 100 mil mulheres para o triênio de 2023-2025. Sobretudo, no estado do Amazonas, observa-se um cenário ainda mais alarmante, cuja taxa de incidência atinge 27,63/100 mil. Nesse cenário, estimam-se 610 novos casos por ano no estado, dos quais 420 ocorrem somente na capital, Manaus<sup>1</sup>.

A progressão desse carcinoma é antecedida por lesões denominadas neoplasias intraepiteliais cervicais (NICs) que se caracterizam pela alteração na arquitetura do tecido epitelial cervical devido a persistência da infecção pelo papilomavírus humano (HPV). E podem ser classificadas em lesão cervical de baixo grau (NIC 1) ou alto grau (NIC 2 e NIC 3)<sup>2</sup>.

Embora medidas preventivas tenham sido implementadas para conter a progressão da doença, as taxas de morbidade persistem no estado. Assim, abordagens complementares associadas a ferramentas preditivas tem se expandindo na pesquisa oncológica, especialmente com o enfoque na descoberta de biomarcadores, os quais podem contribuir no desenvolvimento de tratamentos específicos e aperfeiçoamento estratégico de outras terapias<sup>3</sup>.

Desse modo, abordagens como a metabolômica e a lipidômica são amplamente utilizadas na investigação das alterações moleculares associadas ao desenvolvimento das NICs, assim como na descoberta de indicadores metabólitos e lipídicos para a discriminação dessas neoplasias<sup>4,5</sup>. Logo, modelos de aprendizagem de máquina, especialmente o *Random Forest*, surgem como ferramentas poderosas para a classificação e análise desses dados<sup>6</sup>.

Assim, esse estudo tem como objetivo aplicar a aprendizagem de máquina para identificar painéis moleculares discriminantes de lesões pré-cancerígenas do colo de útero.

## 2. Material e Métodos

**Comitê de ética em pesquisa (CEP):** Esta pesquisa é integrada ao estudo “Avaliação das características epidemiológicas e moleculares de mulheres tratadas com lesões precursoras do câncer de colo de útero no Amazonas”, aprovado pelo Comitê de Ética em Pesquisa (CEP) da Fundação Centro de Oncologia do Estado do Amazonas (FCECON) sob n° de parecer 2.372.935.

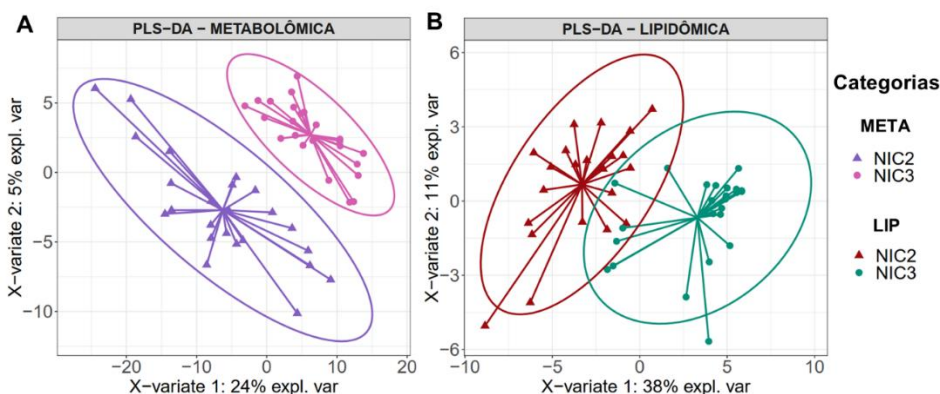
**Seleção de pacientes e coleta de amostras:** Foram selecionadas 24 mulheres com idade acima de 18 anos, que apresentaram laudo histopatológico de NIC 2 ou NIC 3 e indicação cirúrgica para o procedimento de conização. Durante o procedimento, foram coletadas 48 amostras de tecido cervical, compreendendo cerne da lesão e suas margens adjacentes.

**Análise metabolômica/lipidômica:** As amostras foram submetidas à maceração congela-descongela e em seguida à extração tipo *Matyash*<sup>7</sup>, que emprega um sistema bifásico (éter metil-terc-butílico/metanol/água) para a obtenção dos metabólitos e lipídios. As amostras foram processadas e analisadas por cromatografia líquida de ultra-alta eficiência (UHPLC) em um sistema Dionex Ultimate 3000 acoplado a um espectrômetro de massas Q Exactive Plus (Thermo Fisher Scientific), com fonte de ionização por electrospray (ESI). A obtenção dos espectros de massas foi realizada utilizando o software Xcalibur (*Thermo-Scientific*).

**Análise de dados:** Os dados brutos foram processados no software MS-DIAL v.4.9 para detecção, alinhamento e anotação de compostos, utilizando as bibliotecas MONA e MS/MS default. Foram considerados espectros com similaridade >70%. A análise estatística multivariada dos perfis metabólicos e lipídicos foi realizada por linguagem R (versão 4.3.1)

### 3. Resultados e Discussão

Inicialmente, aplicou-se a análise discriminante de mínimos quadrados parciais (PLS-DA) para explorar as diferenças metabólicas e lipídicas nos conjuntos de amostras de NIC 2 e NIC 3. Os resultados denotam uma separação clara do agrupamento dos tecidos cervicais, com excelente acurácia (AUC=1 para metabolômica e AUC= 0,95 para lipidômica). Essa distinção reflete diferenças moleculares associadas ao estágio de progressão das lesões (Figura 1).

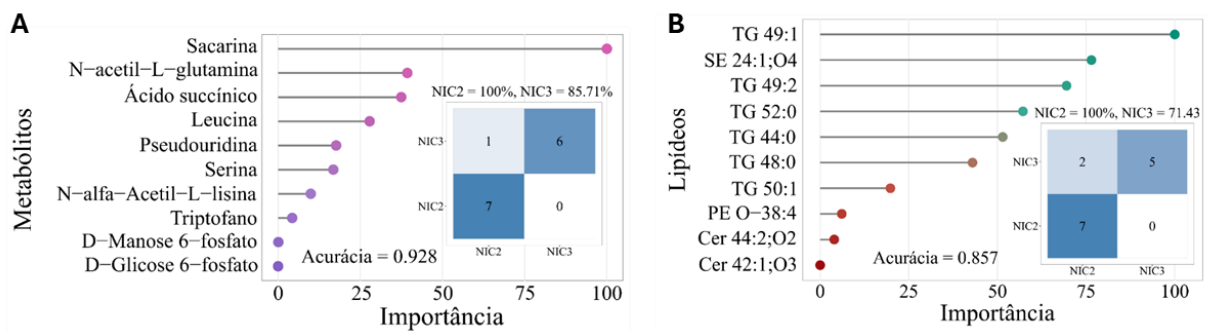


**Figura 1** – PLS-DA das amostras de NIC II (roxo) e NIC III (rosa) das amostras na abordagem metabolômica (A) e NIC II (vermelho) e NIC III (verde) das amostras na abordagem lipidômica (B).

Posteriormente, foi empregada uma abordagem complementar por aprendizagem de máquina através do método Floresta Aleatória (RF – do inglês *Random Forest*), voltado para predição de classes. A análise foi conduzida com o objetivo de avaliar a capacidade dos perfis metabólicos e lipídicos na classificação das pacientes em NIC 2 ou NIC 3, assim como, identificar os compostos com maior correlação com cada lesão.

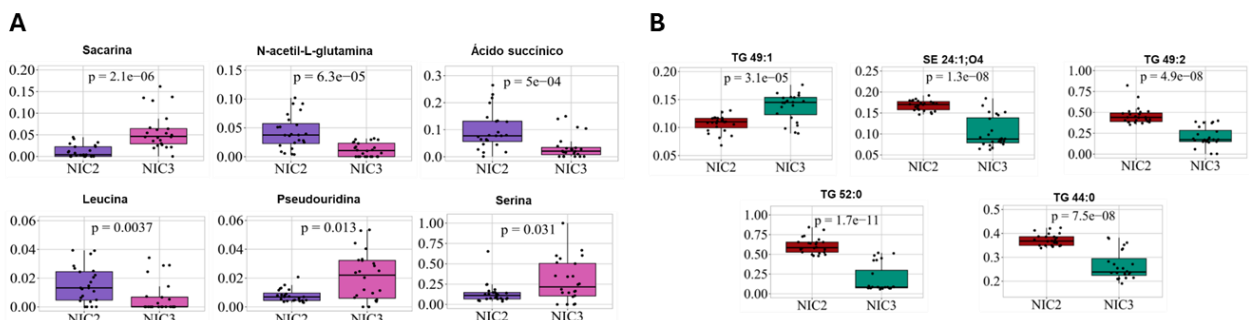
O modelo revelou alta capacidade discriminativa entre NIC 2 ou NIC 3, com acurácias de 92,8% e 85,7% para o estudo metabolômico e lipidômico (Figura 1), respectivamente. Os metabólitos sacarina, N-acetil-L-glutamina e ácido succínico e os lipídios TG 49:1, SE 24:1;O4 24:1;O4 foram os compostos moleculares de maior relevância para essa classificação.

Assim, os resultados do RF sugerem a existência de perfis metabólicos e lipídicos característicos para cada fase da displasia. Alterações nas concentrações desses três metabólitos de maior importância, associados ao metabolismo energético e de aminoácidos; e as concentrações de triglicerídeos e esteróis já foram relacionados a processos neoplásicos, visto que as células atípicas carecem de maior demanda energética e síntese de biomoléculas para a manutenção das suas atividades, como proliferação e sobrevivência<sup>8-10</sup>.



**Figura 1** - Classificação das amostras por Random Forest a partir das variáveis mais importantes para o modelo metabólico (A) e lipídico (B).

A validação gráfica dos marcadores foi realizada por *boxplots* (Figura 2), que evidenciaram diferenças estatisticamente significativas entre os grupos. Valores de  $p$  variando de  $2,1 \times 10^{-6}$  a 0,031 no estudo metabólico e  $1,7 \times 10^{-11}$  a  $3,1 \times 10^{-5}$  no estudo lipídico, denotam que esses metabólitos e lipídios estão diferencialmente abundantes entre os grupos de lesões cervicais de alto grau. Esses achados reforçam a confiabilidade do modelo de aprendizado de máquina e a consistência dos biomarcadores selecionados, que juntos compõem um painel metabólico e lipídico diferencial entre NIC 2 e NIC 3.



**Figura 2** - Box-plots mostram a eficácia de seis metabólitos (A) e cinco lipídios diferenciais (B) ( $p < 0,001$ ; VIP > 1) como potenciais biomarcadores na distinção entre NIC 2 e NIC 3.

Atualmente, não há evidências significativas que associem a presença da sacarina ao desenvolvimento e progressão do câncer, portanto, a presença desse metabólito pode refletir fatores exógenos ou variações individuais nos hábitos alimentares das pacientes<sup>11</sup>. Adicionalmente, a N-acetil-L-glutamina (NAG), participa do metabolismo derivado da glutamina, promovendo síntese de nucleotídeos, aminoácidos e lipídios. A acetilação da glutamina em NAG promove sua estabilidade e disponibilidade metabólica, aprimorando a absorção e uso pelas células, possivelmente favorecendo a progressão das NICs<sup>12</sup>. De forma semelhante, a serina é uma precursora fundamental para a biossíntese de proteínas e ácidos nucleicos, componentes-chave do DNA e RNA celular<sup>13</sup>.

Além desses, aminoácidos essenciais como a leucina também são consumidos pelas células cancerígenas para suprir demandas energéticas e ativar vias proliferativas, como mTOR<sup>14</sup>. No estudo, observou-se maior concentração de leucina em NIC 2 e redução em NIC 3, sugerindo depleção durante a progressão devido ao aumento do consumo metabólico pelas neoplasias.

De outro modo, a pseudouridina, uma molécula modificada do RNA, é associada à regulação do metabolismo e do ciclo celular, influenciando diretamente a apoptose e a proliferação das células atípicas. Seus níveis já foram encontrados em abundância em amostras de soro e urina de pacientes com câncer de ovário, mama, fígado, estômago, pulmão e outros, comparados a indivíduos saudáveis<sup>15</sup>.

Em contraste, estudos descrevem o ácido succínico como um metabólito multifuncional,

capaz de atuar tanto na resposta imunológica — ao ligar-se ao receptor SUCNR1, presente nas células imunes, promovendo a ativação de vias de sinalização intracelular e a produção de citocinas inflamatórias — quanto no avanço da progressão tumoral, por regular genes específicos envolvidos em processos biológicos como a invasão, metástase e resposta imune, podendo influenciar a carcinogênese<sup>16</sup>. Fenômeno relevante, visto que seu caráter ambíguo reflete o comportamento da lesão intermediária (na qual foi mais positivamente mais relevante), que pode regredir espontaneamente para NIC 2 ou progredir para NIC 3<sup>17</sup>.

Ademais, relatos demonstram uma associação significativa entre a dislipidemia dos níveis de triglicerídeos (TGs) à presença de HSIL e aumento do risco de câncer cervical<sup>18</sup>. Especificamente os TG 49:1 e TG 50:1 já foram detectados em níveis abundantes em linhagens de câncer de pulmão e de bexiga<sup>19,20</sup>. Assim como, os esteres de esteróis (SE) já foram relacionados à maior agressividade e progressão dos carcinomas<sup>21</sup>. Logo, sugere-se que mecanismos semelhantes podem estar ocorrendo nas NICs.

Portanto, os resultados mostram que abordagens ômicas evidenciam uma alta capacidade discriminatória entre pacientes com NIC 2 e NIC 3. A análise por RF classificou eficientemente as amostras e indicou alvos moleculares relevantes, com potencial para diagnósticos mais sensíveis e específicos, contribuindo para o rastreamento e tratamento precoce das lesões precursoras do câncer cervical.

#### 4. Conclusões

Neste estudo, abordagem metabolômica e lipidômica demonstrou alta eficácia na diferenciação de pacientes diagnosticados com NIC 2 e NIC 3. O modelo de *Random Forest* obteve alta acurácia na classificação das neoplasias, indicando painéis de biomarcadores promissores na discriminação das lesões cervicais de alto grau.

Diante do exposto, a aplicação da aprendizagem de máquina combinada com dados de espectrometria de massas mostra um grande potencial para a identificação de alvos moleculares, o que pode contribuir para o desenvolvimento de ferramentas de diagnóstico mais sensíveis e específicas para o rastreamento e tratamento precoce das lesões precursoras do câncer cervical.

**Palavras-Chave:** Neoplasias intraepiteliais cervicais; aprendizagem de máquina; metabolômica; lipidômica; espectrometria de massas.

#### Agradecimentos

Às pacientes voluntárias, ILMD-Fiocruz Amazônia, UFAM, FCECON, LabMeta (LADETEC-UFRJ), Meta2MS, LabYM, PPSUS/FAPEAM/Secretaria do Estado de Planejamento, Desenvolvimento, Ciência e Tecnologia/Governo do Estado do Amazonas, CNPq e Capes.

#### Divulgação

Os autores e revisores não relataram qualquer conflito de interesse durante a sua avaliação. Logo, o Congresso Pan-Amazônico de Oncologia detém os direitos autorais, tem a aprovação e a permissão dos autores para divulgação deste resumo, por meio eletrônico.

#### Referências

1. INCA. Estimativa 2023: incidência de câncer no Brasil. Rio de Janeiro: Instituto Nacional de Câncer José Alencar Gomes da Silva; 2023.
2. Øvestad IT, Engesæter B, Halle MK, Akbari S, Bicskei B, Lapin M, Austdal M, et al. High-

grade cervical intraepithelial neoplasia (CIN) associates with increased proliferation and attenuated immune signaling. *Int J Mol Sci.* 2022;23:373.

3. Kocheturov A, Pardalos PM, Karakitsiou A. Massive datasets and machine learning for computational biomedicine: trends and challenges. *Ann Oper Res.* 2019;276:5–34.
4. Cheng F, et al. A serum lipidomic strategy revealed potential lipid biomarkers for early-stage cervical cancer. *Life Sci.* 2021;60:118489.
5. Nam M, et al. Comparable plasma lipid changes in patients with high-grade cervical intraepithelial neoplasia and patients with cervical cancer. *J Proteome Res.* 2021;20:740–750.
6. Genuer R, Poggi JM, Tuleau-Malot C, Villa-Vialaneix N. Random Forests for Big Data. *Big Data Research.* 1º de setembro de 2017;9:28–46.
7. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J Lipid Res.* 2008;49:1137–1146.
8. Choi I, Son H, Baek JH. Tricarboxylic Acid (TCA) Cycle Intermediates: Regulators of Immune Responses. *Life (Basel).* 19 de janeiro de 2021;11(1):69. 1.
9. Ling ZN, Jiang YF, Ru JN, Lu JH, Ding B, Wu J. Amino acid metabolism in health and disease. *Signal Transduct Target Ther.* 13 de setembro de 2023;8(1):345.
10. Martin-Perez M, Urdiroz-Urricelqui U, Bigas C, Benitah SA. The role of lipids in cancer progression and metastasis. *Cell Metab.* 1º de novembro de 2022;34(11):1675–99.
11. Debras C, Chazelas E, Srouf B. Artificial sweeteners and cancer risk: results from the NutriNet-Santé population-based cohort study. *PLoS Med.* 2022;19:e1003950.
12. Zhang Y, Zhu H, Fan J, et al. A glutamine metabolism gene signature with prognostic and predictive value for colorectal cancer survival and immunotherapy response. *Front Mol Biosci.* 2025;12:1599141.
13. Handzlik MK, Metallo CM. Sources and Sinks of Serine in Nutrition, Health, and Disease. *Annu Rev Nutr.* 21 de agosto de 2023;43:123–51.
14. Aki S, Nakahara R, Maeda K, Osawa T. Cancer metabolism within tumor microenvironments. *Biochim Biophys Acta Gen Subj.* 2023;1867:130330.
15. Liu K, Zhang S, Liu Y, Hu X, Gu X. Advancements in pseudouridine modifying enzyme and cancer. *Front Cell Dev Biol [Internet].* 16 de dezembro de 2024 [citado 30 de agosto de 2025];12. Disponível em: <https://www.frontiersin.org/journals/cell-and-developmental-biology/articles/10.3389/fcell.2024.1465546/full>
16. Kuo CC, Wu JY, Wu KK. Cancer-derived extracellular succinate: a driver of cancer metastasis. *J Biomed Sci.* 2022;29:93.
17. Nourrisson A, Lepetit H, Marty M, Garrigue I, Brun JL. Regression of cervical high-grade squamous intraepithelial lesions (HSIL/CIN2) managed expectantly. *J Gynecol Obstet Hum Reprod.* 2022;51:102442.
18. Mwangi GF, Niyonzima N, Atwine R, et al. Dyslipidemia: prevalence and association with precancerous and cancerous lesions of the cervix; a pilot study. *Lipids Health Dis.* 2024;23:3.
19. Lee MY, Yeon A, Shahid M, Cho E, Sairam V, Figlin R, et al. Reprogrammed lipid metabolism in bladder cancer with cisplatin resistance. *Oncotarget.* 2 de março de 2018;9(17):13231–43.
20. Lee GB, Lee JC, Moon MH. Plasma lipid profile comparison of five different cancers by nanoflow ultrahigh performance liquid chromatography-tandem mass spectrometry. *Anal Chim Acta.* 31 de julho de 2019;1063:117–26.
21. De Gonzalo-Calvo D, et al. Intratumor cholesteryl ester accumulation is associated with human breast cancer proliferation and aggressive potential: a molecular and clinicopathological study. *BMC Cancer.* 2015;15:460.