APLICAÇÃO DE APRENDIZADO DE MÁQUINA PARA DIAGNÓSTICO DO CÂNCER DE PULMÃO: UMA ABORDAGEM BASEADA EM RANDOM FOREST

https://doi.org/10.63330/cabapsamed2025-003

Daniella Rodrigues da Silva
Acadêmica do curso de Medicina pela Universidade do Estado de Minas Gerais.
E-mail: danysilvard@outlook.com

João Victor Rodrigues da Silva
Acadêmico do curso de Medicina pela Universidade do Estado de Minas Gerais.

Debora Aparecida da Silva
Mestre e Técnico Universitário da Universidade do Estado de Minas Gerais.

Resumo

Introdução: O câncer de pulmão é uma das neoplasias malignas mais letais do mundo, sendo responsável por uma elevada taxa de mortalidade devido ao diagnóstico tardio e à rápida progressão tumoral. A detecção precoce é fundamental para aumentar a sobrevida e melhorar a resposta terapêutica. Com os avanços da tecnologia, o uso de ferramentas computacionais e de aprendizado de máquina tem se tornado uma alternativa promissora no auxílio ao diagnóstico clínico. Entre essas técnicas, o modelo Random Forest, um método de aprendizado supervisionado, destaca-se pela capacidade de lidar com grandes volumes de dados e identificar padrões complexos. Objetivo: O objetivo deste estudo foi aplicar o algoritmo Random Forest para classificar nódulos pulmonares como benignos ou malignos, utilizando um conjunto de dados clínicos e radiológicos obtidos de exames de imagem. Métodos: Foi utilizado o Conjunto de Dados de Câncer de Pulmão de Low-Dose Computed Tomography (LDCT), composto por 522 amostras de nódulos com 25 variáveis radiológicas e clínicas. As variáveis independentes incluíram características como tamanho, densidade, bordas e localização dos nódulos, enquanto a variável dependente foi o diagnóstico histopatológico (benigno ou maligno). Os dados foram divididos em conjuntos de treino (80%) e teste (20%), com normalização para padronizar as escalas das variáveis. O modelo Random Forest foi treinado com 120 árvores de decisão e avaliado por métricas de acurácia, precisão, sensibilidade (recall) e F1-score. Resultados: O modelo apresentou acurácia de 95,8%, indicando excelente desempenho na diferenciação entre nódulos benignos e malignos. A precisão foi de 0,96, o recall de 0,94 e o F1-score de 0,95 para os casos malignos. As variáveis mais relevantes para a tomada de decisão foram o diâmetro médio do nódulo, irregularidade das bordas e densidade da lesão, fatores diretamente associados ao comportamento tumoral. Conclusão: O modelo Random Forest demonstrou alto potencial para aplicação no diagnóstico do câncer de pulmão, oferecendo resultados precisos e reprodutíveis. O uso do aprendizado de máquina em imagens

radiológicas pode auxiliar profissionais de saúde na detecção precoce, contribuindo para o planejamento terapêutico e redução da mortalidade. Estudos futuros com amostras maiores e o uso de diferentes algoritmos podem ampliar a robustez e generalização dos resultados obtidos.

Palavras-chave: Aprendizado de Máquina; Câncer de Pulmão; Diagnóstico Precoce; Oncologia Computacional; Random Forest.

REFERÊNCIAS

DALLAGASSA, M. R.; OLIVEIRA, A. J. C. Uso de machine learning no diagnóstico de câncer de pulmão por tomografia computadorizada. **Journal of Health Informatics**, v. 16, n. esp., p. 41–46, 2024. Disponível em: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/1292. Acesso em: 24 mar. 2025.

KHAMPARIA, A. et al. Diagnosis of lung cancer based on hybrid transfer learning. **Computers & Electrical Engineering**, v. 85, art. 106658, 2020.

RAKHLIN, A. et al. Deep convolutional neural networks for lung cancer image analysis. In: CAMPILHO, A.; KARRAY, F.; TER HAAR ROMENY, B. (ed.). **Image analysis and recognition**. Cham: Springer, 2018. p. 745–752. (Lecture Notes in Computer Science, v. 10882).

SHEN, L. et al. Deep learning to improve lung cancer detection on CT screening. **Scientific Reports**, v. 9, n. 1, art. 12501, 2019. DOI: https://doi.org/10.1038/s41598-019-48995-5.

WU, N. et al. Deep neural networks enhance radiologists' performance in lung cancer screening. **IEEE Transactions on Medical Imaging**, v. 39, n. 4, p. 1188–1192, 2020. DOI: https://doi.org/10.1109/TMI.2019.2945515.