11º CONGRESSO NACIONAL DE BIOMECÂNICA
A. P. G. Castro, C. Pina, P. R. Fernandes, C. Santos, J. Folgado & A. Amaro (Eds)
Sesimbra, Portugal, 21-22 de Fevereiro, 2025

AVALIAÇÃO DE MODELOS COMPUTACIONAIS PARA PREDIÇÃO DE SOBREVIDA NO CANCRO DE MAMA FEMININO

Daniela Carvalho [1,2], Marco Parente [2] e Priscila Goliatt [1]

[1] PPGMC - UFJF - Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, Brasil

[2] FEUP - Faculdade de Engenharia da Universidade do Porto, Portugal

daniela.schimitz@estudante.ufjf.br ; mparente@fe.up.pt ; capriles@ice.ufjf.br

PALAVRAS-CHAVE: Análise de Sobrevida, Cancro de Mama, Modelação Computacional

1 INTRODUÇÃO

O cancro de mama (CM) é a principal causa de mortalidade por cancro entre mulheres, com aproximadamente 2,3 milhões de novos casos e 666 mil mortes a nível mundial em 2022 [1]. No Brasil, foram estimados 74 mil novos casos e 19 mil óbitos em 2023 e 2022, respectivamente [2,3]. Apesar da complexidade e variabilidade dessa doença, o prognóstico tende a ser mais favorável quando o diagnóstico e o tratamento são realizados precocemente, melhorando assim as taxas de sobrevida [3]. Nos últimos anos, os modelos computacionais para predição de sobrevida, incluindo Métodos de Aprendizado de Máquina (MAM) como *Gradient Boosting Survival* (GBS), *Random Survival Forest* (RSF) e *Survival Support Vector Machine* (SSVM), têm demonstrado superioridade em relação aos métodos tradicionais, como o modelo de *Cox Proportional Hazards* (COX) [3,4,5,6]. Estes modelos estão disponíveis na biblioteca '*Scikit-Survival*', um pacote específico para análise de sobrevida [6]. O desenvolvimento e a validação destes modelos com dados clínicos usuais são essenciais para melhorar o diagnóstico, tratamento e a sobrevida das pacientes [3]. O presente estudo tem como objetivo avaliar e comparar o desempenho do modelo COX com MAM na predição da sobrevida de pacientes com CM tratadas e acompanhadas na Zona da Mata Mineira, Brasil.

2 MATERIAL E MÉTODOS

Este estudo utilizará MAM supervisionados para avaliar a predição de sobrevida em pacientes com CM feminino. Após a aprovação ética (parecer nº 5.533.296) pelo Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal de Juiz de Fora, os dados clínicos serão pré-processados, incluindo a limpeza de dados e tratamento de valores ausentes. A seleção de atributos será realizada através de validação cruzada K-fold (K=5), juntamente com a regularização Lasso. O desempenho discriminativo dos modelos será avaliado pelo *Concordance Index* (C-Index). A implementação dos modelos, a seleção de características e a divisão dos dados (75% para treino e 25% para teste) será feita em Python, seguindo as diretrizes da biblioteca *Scikit-Survival* [6].

3 RESULTADOS E DISCUSSÃO

O conjunto de dados clínicos analisados incluiu 558 pacientes diagnosticadas com CM, com 12 atributos prognósticos e análise de sobrevida de 5 anos. O tempo de seguimento foi contabilizado desde a data do laudo histopatológico até o evento adverso (óbito por CM) ou até à data de censura, totalizando 1826 dias, com 113 óbitos registrados. A Tabela 1 apresenta a

avaliação do desempenho discriminativo dos MAM com base no C-Index, comparando os resultados obtidos com os reportados na literatura. Um C-Index próximo de 1 indica uma excelente capacidade de discriminação nos dados de teste. Entre os modelos avaliados, o RSF demonstrou o melhor desempenho, seguido pelo GBS, COX e SSVM.

Tabela 1 -Comparação do C-Index entre os MAM deste estudo e os reportados na literatura.

MAM	COX	GBS	RSF	SSVM
Presente estudo	0,810	0,904	**0,924**	0,798
Carvalho et al. 2024	0,743	**0,914**	0,917	0,746
Liu et al. 2021	0,759	0,823	0,814	—
Xiao et al. 2022	**0,814**	—	0,827	**0,812**

Ao comparar com os resultados da literatura, os modelos lineares COX e SSVM apresentaram desempenhos superiores nos estudos de Xiao e colaboradores [5], que utilizaram 21 atributos e 22 mil casos de CM feminino, enquanto, no presente estudo, foram usados 12 atributos. Entre os modelos não lineares, o RSF destacou-se, superando tanto os modelos lineares quanto outros não lineares, além dos resultados reportados na literatura. O GBS obteve melhor desempenho no estudo de Carvalho e colaboradores [3], que utilizaram 70 atributos.

4 Conclusão

O modelo *ensemble* não linear RSF mostrou-se promissor para a predição de sobrevida em pacientes com CM feminino, oferecendo implicações significativas para a prática clínica e gestão da saúde.

AGRADECIMENTOS

Agradecemos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro por meio da bolsa concedida; à Faculdade de Engenharia da Universidade do Porto e ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora pelo aprendizado durante o Doutorado; ao professor Dr. Maximiliano Ribeiro Guerra e à mastologista Dra. Jane Rocha Duarte Cintra pela colaboração e fornecimento do banco de dados clínicos; e aos Hospitais 9 de Julho e Instituto Oncológico pelo apoio na pesquisa.

Referências

[1] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, et al., "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", CA Cancer J Clin, vol. 74, pp. 229–263, May-Jun 2024.

[2] M. de O. Santos, F. C. da S. de Lima, L. F. L. Martins, J. F. P. Oliveira, L. M. de Almeida, et al., "Estimativa de Incidência de Câncer no Brasil, 2023-2025", Rev. Bras. Cancerologia, vol. 69, no. 1, p. e–213700, Fev. 2023.

[3] D. S. de Carvalho, P. Capriles, and L. Goliartt, "Comparative Analysis of Machine Learning Models for Breast Cancer Patients' Survival Prediction". In: International Conference on Intelligent Systems Design and Applications, A. Abraham, A. Bajaj and H. Thomas Eds. Switzerland: Springer Nature, 2024, pp. 172-181.

[4] P. Liu, B. Fu, S. X. Yang, L. Deng, X. Zhong, and H. Zheng, "Optimizing Survival Analysis of XGBoost for Ties to Predict Disease Progression of Breast Cancer." *IEEE transactions on bio-medical engineering,* vol. 68, no.1 pp. 148-160, 2021.

[5] J. Xiao, M. Mo, Z. Wang, C. Zhou, J. Shen, J. Yuan, et al., "The Application and Comparison of Machine Learning Models for the Prediction of Breast Cancer Prognosis: Retrospective Cohort Study", JMIR Med Inform, vol. 10, no. 2, p. e33440, 2022.

[6] S. Pölsterl, "scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn", Journal of Machine Learning Research, vol. 21, no. 212, pp. 1-6, 2020.