



Machine Learning e Análise Multivariada aplicados à Sobrevida do Câncer Mama

Machine Learning and Multivariate Analysis applied to Breast Cancer Survival

Machine Learning y Análisis Multivariante aplicados a la Supervivencia al Cáncer de Mama

Talita Santos Pinheiro¹, Erika Yahata², Pablo Deoclecia dos Santos¹, Felipe Soares de Oliveira¹, André Kazuo Takahata², Ricardo Suyama³, Harki Tanaka⁴, Tiago Ribeiro de Oliveira⁴, Ana Paula Romani⁵, Priscyla Waleska Simões⁶

RESUMO

Descritores: Análise de Sobrevida; Aprendizado de Máquina; Câncer mama

Objetivo: O presente estudo apresenta os resultados da Regressão Multivariada de Cox e do algoritmo Random Survival Forest na análise de sobrevida de pacientes com câncer de mama. **Métodos:** O estudo considerou uma amostra que continha dados de 4024 pacientes do sexo feminino diagnosticadas no período de 2006 a 2010 com câncer de mama (carcinoma ductal invasivo e carcinoma lobular). **Resultados:** A análise multivariada revelou a idade, tamanho do tumor e estadiamento como fatores de risco, que também foram considerados relevantes entre os atributos com maior ganho de informação pelo modelo apresentado pelo algoritmo Random Survival Forest. **Conclusão:** Temos evidências que o modelo apresentado pelo algoritmo Random Survival Forest considerado na análise de sobrevida de câncer de mama apresentou-se em concordância com os resultados da estatística multivariada, e que os resultados são promissores para a prática clínica.

ABSTRACT

Keywords: Survival Analysis; Machine Learning; Breast Cancer

Objective: This study presents the results of the application of Cox Multivariate Regression and Random Survival Forest algorithm for survival analysis of patients with breast cancer. **Methods:** The study considered a sample that contained data from 4024 female patients diagnosed in the period from 2006 to 2010 with breast cancer (invasive ductal carcinoma and lobular carcinoma). **Results:** The multivariate analysis revealed age, tumor size and stage as risk factors, which were also considered relevant among the attributes with the greatest information gain by the model presented by the Random Survival Forest algorithm. **Conclusion:** We show evidence that the model presented by the Random Survival Forest algorithm considered in breast cancer survival analysis agrees with results of multivariate statistics, and those results are promising for clinical practice.

RESUMEN

Descriptores: Análisis de Supervivencia; Aprendizaje automático; Câncer de mama

Objetivo: El presente estudio presenta los resultados de la Regresión de Cox Multivariada y el algoritmo Random Survival Forest en el análisis de supervivencia de pacientes con cáncer de mama. **Métodos:** El estudio consideró una muestra que contenía datos de 4024 pacientes mujeres diagnosticadas en el período de 2006 a 2010 con cáncer de mama (carcinoma ductal invasivo y carcinoma lobulillar). **Resultados:** El análisis multivariado reveló como factores de riesgo la edad, tamaño del tumor y estadio, los cuales también fueron considerados relevantes entre los atributos con mayor ganancia de información por el modelo presentado por el algoritmo Random Survival Forest. **Conclusión:** Tenemos evidencia de que el modelo presentado por el algoritmo Random Survival Forest considerado en el análisis de la supervivencia del cáncer de mama estuvo de acuerdo con los resultados de la estadística multivariada, y que los resultados son prometedores para la práctica clínica.

¹ Mestranda do Programa de Pós-Graduação em Engenharia Biomédica, Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas - CECS, Universidade Federal do ABC - UFABC, São Bernardo do Campo (SP), Brasil

² Professor Adjunto do Programa de Pós-Graduação em Engenharia da Informação, Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas - CECS, Universidade Federal do ABC - UFABC, São Bernardo do Campo (SP), Brasil.

³ Professor Associado do Programa de Pós-Graduação em Engenharia da Informação, Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas - CECS, Universidade Federal do ABC - UFABC, São Bernardo do Campo (SP), Brasil.

⁴ Professor Adjunto do Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas - CECS, Universidade Federal do ABC - UFABC, São Bernardo do Campo (SP), Brasil.

⁵ Professora Associada do Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas - CECS, Universidade Federal do ABC - UFABC, São Bernardo do Campo (SP), Brasil.

⁶ Professora Adjunta do Programa de Pós-Graduação em Engenharia Biomédica, Programa de Pós-Graduação em Engenharia da Informação, Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas - CECS, Universidade Federal do ABC - UFABC, São Bernardo do Campo (SP), Brasil.

Autor Correspondente: Priscyla Waleska Simões
e-mail: pritsimoes@gmail.com

Artigo recebido: 18/03/2022
Aprovado: 10/05/2022

INTRODUÇÃO

O câncer de mama constitui-se no tipo de câncer de maior incidência quando observada a população feminina brasileira e mundial, somente em 2020, 2,3 milhões de mulheres foram diagnosticadas, e 685.000 morreram, com prevalência de 7,8 milhões de casos nos últimos 5 anos⁽¹⁾. No Brasil, foram registrados 18.068 óbitos pela doença em 2019, e incidência de 66.280 novos casos em 2020, ou seja, 61,61 casos por 100 mil mulheres, equivalente a 29,7% de todos os tipos de câncer⁽²⁾.

Nos estudos de câncer de mama, é usual o registro histórico incluindo as datas correspondentes ao diagnóstico da doença, remissão na qual o paciente não apresenta mais a doença, recorrência da doença e dados de óbito. Estes dados são importantes à análise de sobrevida (ou análise de sobrevivência), relacionada ao prognóstico da doença⁽³⁾.

Se houver dados censurados (dados parciais), as técnicas estatísticas clássicas, como análise de regressão, podem ser imprecisas na progressão da doença⁽³⁾, desta forma, buscam-se métodos que possam conviver com menor nível de consistência temporal nos dados (menos censuras). Assim, uma alternativa é a utilização da análise de sobrevida, que possibilita a incorporação de séries temporais aos dados censurados⁽³⁾.

Atualmente tem-se utilizado com frequência algoritmos de machine learning (ML) para análise de sobrevida em razão da possibilidade de revelar relacionamentos ocultos entre padrões em processos prognósticos e tratamento de doenças, permitindo a leitura de grande quantidade de dados gerando, assim, modelos preditivos mais complexos com possibilidade de individualização para cada indivíduo^(4,5).

Tem-se o algoritmo Random Survival Forest (RSF) que é baseado no algoritmo Random Forest (RF) e caracteriza-se por sanar parte do problema relacionado aos dados censurados e prediz a sobrevida entre os nós filhos e não mais por impureza⁽⁶⁻⁷⁾. A saída produzida com o RSF é uma função que mostra a probabilidade, por exemplo, de o paciente sobreviver após um determinado tempo de seguimento⁽⁶⁾.

O método de ML baseado em RSF combina agregação bootstrap (amostragem com reposição) com

árvores de sobrevida aleatórias e censura à direita, quando o tempo para que se tenha o desfecho (tempo de falha) é superior ao tempo estudado⁽⁶⁻⁷⁾.

Alguns estudos têm utilizado o RSF de forma promissora para câncer de mama⁽⁸⁻⁹⁾. Considerando que a incidência do câncer de mama e a mortalidade tem aumentado nos últimos anos, tais estudos de sobrevida têm contribuído positivamente no prognóstico dos pacientes, de maneira a direcionar tratamentos mais eficazes, além de elucidar possíveis fatores de risco ou proteção⁽¹⁰⁻¹¹⁾.

Mediante o exposto, este artigo tem por objetivo apresentar um modelo preditivo baseado no RSF e regressão de Cox para análise de sobrevida de pacientes com câncer de mama.

MÉTODOS

O estudo considerou uma amostra que continha dados de 4024 pacientes do sexo feminino diagnosticadas no período de 2006 a 2010 com câncer de mama (carcinoma ductal invasivo e carcinoma lobular), disponibilizada via atualização de 2017 de forma pública, pelo National Cancer Institute (NCI).

Para análise de sobrevida foram selecionados os seguintes atributos: idade, tamanho do tumor, estadiamento⁽¹²⁻¹⁴⁾, presença do biomarcador para estrogênio, presença do biomarcador para progesterona, tempo de sobrevida (em meses) e status de sobrevida ou óbito (considerado como desfecho) (Tabela 1).

Foram realizadas transformações em todos os atributos e dicotomização no tamanho do tumor e idade pela mediana (devido a distribuição não gaussiana dos dados, previamente avaliada pelo Teste de Kolmogorov-Smirnov). O estadiamento foi dicotomizado em II e III.

Inicialmente realizou-se no software IBM SPSS®, a análise univariada que considerou para as variáveis quantitativas a mediana como medida de tendência central e o intervalo interquartilico na dispersão dos dados; e para as variáveis qualitativas, a frequência absoluta e a frequência relativa. Posteriormente foi realizada análise de sobrevida a partir da curva de Kaplan-Meier, Teste de Log-Rank, sendo sucedida pela análise multivariada que foi realizada pela regressão de Cox (precedida pela

Tabela 1 - Atributos

Atributo	Descrição
Idade	Idade em anos
Tamanho do tumor	Tamanho do tumor em centímetros
Estrogênio	Presença do biomarcador para estrogênio
Progesterona	Presença do biomarcador para progesterona
Estadiamento	Grau do estadiamento do câncer
Tempo de sobrevida	Meses de sobrevida após o diagnóstico
Desfecho	Desfecho até a data limite do estudo

Tabela 2 - Hiperparâmetros

Descrição	Hiperparâmetro
Quantidade de atributos selecionados aleatoriamente para serem candidatos a dividir um nó	mtry = 2
Tamanho mínimo de nó terminal	nodesize = 15
Quantidade de árvores solicitadas para o crescimento da floresta	ntree = 100

avaliação do pressuposto de proporcionalidade dos riscos), e tamanho do efeito mensurado por meio do Hazard Ratio (HR).

Após a análise univariada, de sobrevida e multivariada, o algoritmo RSF foi modelado no software R 4.1.0, considerando um conjunto de treinamento (70% dos dados) e outro de teste (30% dos dados). Nessa etapa foram utilizadas as bibliotecas RandomForest SRS e survival R. Os hiperparâmetros são apresentados na Tabela 2⁽¹⁵⁻¹⁸⁾.

O desempenho do algoritmo foi avaliado pelo índice de concordância C-index e pelo Integrated Brier Score (IBS). O C-index varia de 0,5 (previsão aleatória) a 1

(previsão perfeita), e reflete o poder do algoritmo em classificar duas instâncias, por exemplo, entre indivíduos de baixo a alto risco⁽¹⁹⁾. O Integrated Brier Score (IBS) é uma métrica de calibração que avalia a qualidade da predição, assim, valores mais baixos sugerem um melhor resultado, obtido pelas curvas de sobrevida que o modelo produz em relação a realidade⁽²⁰⁾.

RESULTADOS E DISCUSSÃO

As características da amostra são apresentadas na Tabela 3.

A Figura 1 apresenta as curvas de risco para: Idade, tamanho do tumor, estrogênio, progesterona e

Tabela 3 – Características da amostra

Atributos	N (% ou mediana e Intervalo Interquartil)
Idade	4024 [54 (47 -51)]
Estadiamento	
III C	472 (11,7%)
III B	67 (1,7%)
III A	1050 (26,1%)
II B	1130 (28,1%)
II A	1305 (32,4%)
Tamanho do Tumor	4024 [24 (16 -38)]
Estrogênio	
Positivo	3755 (93,3%)
Negativo	269 (6,7%)
Progesterona	
Positivo	3326 (82,7%)
Negativo	698 (17,3%)

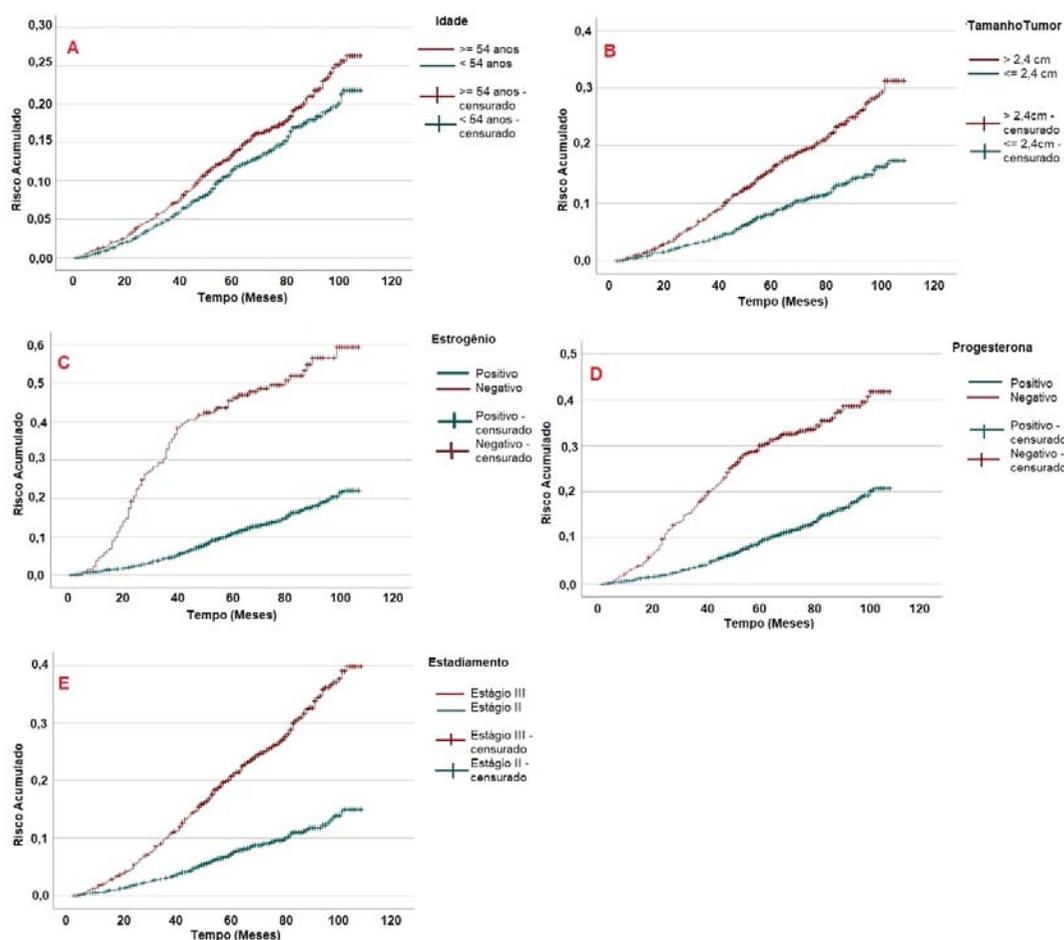


Figura 1 – Curvas de Risco

Tabela 4 – Medidas de efeito

Atributos	HR _{bruto} (IC 95%)	p	HR _{ajustado} (IC 95%)	p
Idade	1,210 (1,031 – 1,419)	0,019	1,226 (1,044 – 1,440)	0,013
Estadiamento	2,793 (2,373 – 3,288)	<0,001	2,400 (2,014 – 2,861)	<0,001
Tamanho do Tumor	1,839 (1,560 – 2,167)	<0,001	1,288 (1,080 – 1,537)	0,005
Estrogênio	0,273 (0,222 – 0,336)	<0,001	0,448 (0,347 – 0,578)	<0,001
Progesterona	0,385 (0,325 – 0,455)	<0,001	0,574 (0,467 – 0,705)	<0,001

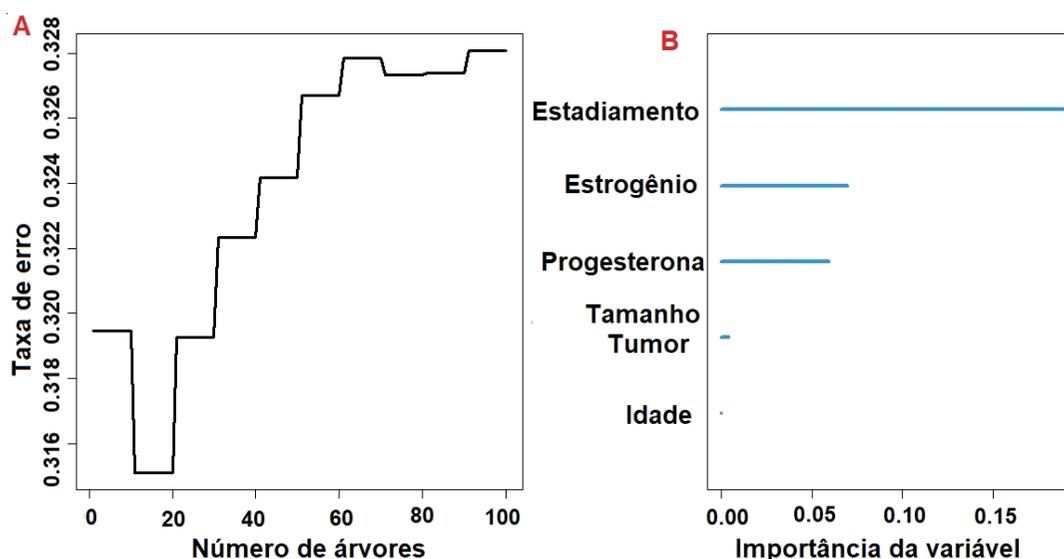


Figura 2 – Avaliação do modelo – Árvores e importância dos atributos

estadiamento. Observa-se que os perfis de risco foram significativamente distintos.

Mediante ajuste do modelo pode-se verificar também na Figura 1 e na Tabela 4 um risco mais elevado de óbito para pacientes com idade superior ou igual a 54 anos [HR=1,226 (IC95%: 1,044–1,440)] (Figura 1A); com tumores maiores que 2,4cm [HR = 1,288 (IC95%: 1,080 – 1,537)] (Figura 1B); com status negativo para estrogênio [HR = 0,448 (IC95%: 0,347 – 0,578)] (Figura 1C); com status negativo para progesterona [HR = 0,574 (IC95%: 0,467 – 0,705)] (Figura 1D); e em estágio III [IC95%: HR = 2,400 (IC95%: 2,014 – 2,861)] (Figura 1E).

A Tabela 4 apresenta em ordem decrescente os atributos com relação a sua medida de efeito (HR). Alguns fatores de risco observados (Estadiamento III, Tamanho do tumor maior 2,4cm e Idade maior ou igual a 54 anos); assim como alguns fatores de proteção (presença do biomarcador estrogênio e presença do biomarcador progesterona).

Os resultados da análise de sobrevida pelo método de Machine Learning e algoritmo RSF estão apresentados nas Figuras 2 e 3.

Na parametrização do algoritmo RSF, foi realizado o crescimento de 100 árvores, podendo-se observar de que houve estabilização na taxa de erro de previsão dos dados Out of Bag (OOB) em 32,80% (Figura 2A). O estadiamento apresentou-se como o atributo de maior contribuição ao modelo (0,1901), sucedido pelo estrogênio (0,0695), progesterona (0,0594), tamanho do tumor (0,0041), e idade (0,0002), como apresenta a Figura 2B.

O modelo RSF apresentou c-index de 0,672 nos dados de treino e 0,682 nos dados de teste, sendo ambos superiores a 0,5 que é o valor mínimo descrito na literatura para um bom modelo, sendo melhor quanto mais próximo de 1,0⁽¹²⁾. O desempenho preditivo foi avaliado pelo IBS,

resultando num valor de 0,076, enquanto para os dados de teste o valor foi de 0,090, ambos próximos de zero, característica essa que sugere boa capacidade preditiva.

A probabilidade de sobrevida de um paciente é apresentada na Figura 3A, cada linha indica um paciente do dataset. A curva vermelha representa o estimador de Kaplan-Meier que é uma métrica média da probabilidade de sobrevida, assim, pacientes com curvas abaixo (da curva vermelha) apresentam maior risco para o desfecho, ou seja, de ocorrência de óbito; enquanto pacientes com curvas acima apresentam menor risco do evento ocorrer⁽²¹⁾.

Na Figura 3A pode-se observar que à medida que aumentam os meses, menor é a probabilidade de sobrevida. As linhas de cada paciente com alto risco (as curvas mais baixas) estão bem afastadas, o que mostra a precisão do modelo, em contrapartida as linhas próximas ao estimador de Kaplan-Meier estão próximas, mostrando que para outros casos o modelo ainda pode ser ajustado melhor para o desfecho.

Na Figura 3B pode-se observar o escore OOB Brier que trata o valor de erro de predição do desfecho ao longo do tempo, que resultou em valores entre 0,00 e 0,20, próximo à 0, que é o valor pretendido para a melhor predição possível. Pode-se observar ainda tanto na linha vermelha, que representa o conjunto não estratificado com valores de 0,00 a 0,15, como nas outras linhas que simbolizam a estratificação em quatro grupos de acordo com o percentil de mortalidade (0-25,25-50,50-75,75-100), que o modelo apresentou melhor aderência à predição da sobrevida nos quartis 0-25 e 0-50, os quais também apresentaram escores OOB Brier inferiores a 0,10 por todo o seguimento de tempo.

Na Figura 3C, pode-se identificar os indivíduos que tiveram o desfecho (morte) em azul, e os indivíduos que censuram (vivos) em preto. As taxas de mortalidade mais

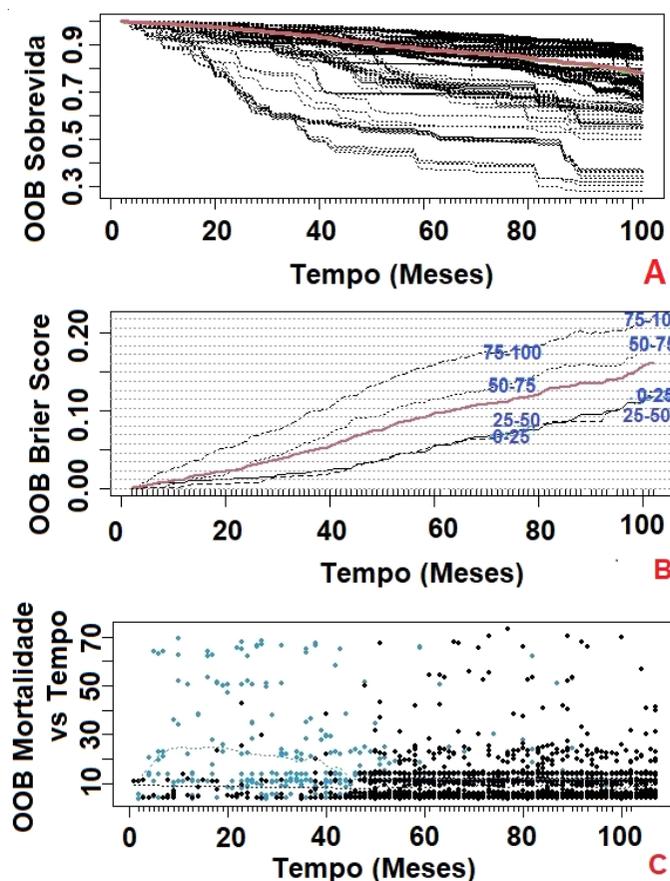


Figura 3 – Estimativas de sobrevida

altas ocorreram no início do seguimento (primeiros meses), e dessa forma, mais pontos azuis foram observados nos primeiros meses. Pode-se destacar também que a maioria da amostra foi composta por indivíduos que censuraram (sobrevida).

Numa coorte realizada no Irã⁽²²⁾ em 2009 e 2012 sobre fatores de risco associados a eventos adversos cardíacos e cerebrais através da utilização de RSF e Regressão de Cox, dentre os 13 atributos estudados, os mais importantes foram semelhantes ao RSF, com c-index de 0,648 e IBS de 0,124. Tal característica de concordância dos atributos também foi observada em nossa pesquisa.

Corroborando essa questão, um estudo de Shanghai⁽²³⁾ realizado entre os anos de 2015 e 2019 com 82 pacientes sobre previsão da progressão tumoral de glioma de alto grau (HGG) após radioterapia por feixe de partículas (PBRT), concluiu que o atributo considerado mais importante pelo RSF estava dentre os destacados com importância na regressão multivariada de Cox, resultando em c-index de 0,611 e IBS de 0,174, assemelhando-se ao que foi encontrado em nosso estudo.

Nossas análises mostraram uma melhor predição para os casos estratificados nos percentis de mortalidade 0-25 e 25-50. Assim, numa coorte chinesa⁽²⁴⁾ realizada com 121 pacientes de 1997 até 2019 sobre a sobrevida relacionada a síndrome do QT longo, o desfecho considerou a taquicardia/fibrilação ventricular espontânea. Os atributos de maior importância foram os mesmos nas duas análises (RSF e Cox) com IBS de 0,10 e c-index de 0,79, além BS estratificado de 0-25 ser o único abaixo do geral, correspondendo a melhor predição. Tal diferença pode ser atribuída a maior quantidade de pacientes no nosso estudo permitindo uma melhor

distribuição por percentil de mortalidade abrangendo assim mais uma faixa de distribuição.

CONCLUSÕES

Os resultados apresentados a partir dos dados do SEER revelaram a idade, tamanho do tumor e estadiamento como fatores de risco, que também foram considerados relevantes entre os atributos com maior ganho de informação pelo modelo apresentado pelo algoritmo Random Survival Forest. O estadiamento apresentou maior importância na sobrevida de indivíduos com câncer de mama, quando analisado pelo algoritmo Random Survival Forest. Assim, as evidências indicam que o modelo pautado no algoritmo de Random Survival Forest para análise de sobrevida de Câncer de Mama a partir dos dados do SEER apresentou-se em concordância com os resultados da estatística multivariada e com resultados promissores para a prática clínica.

AGRADECIMENTOS

Esse estudo foi financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Universidade Federal do ABC (UFABC).

O trabalho foi realizado no âmbito do Centro de Pesquisa Aplicada em Inteligência Artificial BIOS – Brazilian Institute of Data Science, apoiado pelo processo nº 2020/09838-0, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

REFERÊNCIAS

1. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: An overview. *International journal of cancer*. 2021.
2. Instituto Nacional de Câncer (INCA). Estatísticas de câncer. <https://www.inca.gov.br/numeros-de-cancer2021>.
3. Colosimo EA, Giolo SR. *Análise de sobrevivência aplicada*: Editora Blucher; 2006.
4. Johnson KW, Shameer K, Glicksberg BS, Readhead B, Sengupta PP, Björkegren JLM, et al. Enabling Precision Cardiology Through Multiscale Biology and Systems Medicine. *JACC: Basic to Translational Science*. 2017;2(3):311-27.
5. Huyut MT, Üstündağ H. Prediction of diagnosis and prognosis of COVID-19 disease by blood gas parameters using decision trees machine learning model: a retrospective observational study. *Med Gas Res*. 2022;12(2):60-6.
6. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32.
7. Bühnemann C, Li S, Yu H, Branford White H, Schäfer KL, Llombart-Bosch A, et al. Quantification of the heterogeneity of prognostic cellular biomarkers in ewing sarcoma using automated image and random survival forest analysis. *PloS one*. 2014;9(9):e107105.
8. Völkel V, Huetting TA, Draeger T, van Maaren MC, de Munck L, Strobbe LJA, et al. Improved risk estimation of locoregional recurrence, secondary contralateral tumors and distant metastases in early breast cancer: the INFLUENCE 2.0 model. *Breast cancer research and treatment*. 2021;189(3):817-26.
9. Jiang X, Xue D, Brufsky A, Khan S, Neapolitan R. A new method for predicting patient survivorship using efficient bayesian network learning. *Cancer informatics*. 2014;13:47-57.
10. Maajani K, Khodadost M, Fattahi A, Pirouzi A. Survival rates of patients with breast cancer in countries in the Eastern Mediterranean Region: a systematic review and meta-analysis. *Eastern Mediterranean health journal = La revue de sante de la Mediterranee orientale = al-Majallah al-sihhiyah li-sharq al-mutawassit*. 2020;26(2):219-32.
11. Hassanipour S, Maghsoudi A, Rezaeian S, Arab-Zozani M, Mokhtari AM, Abdzadeh E, et al. Survival Rate of Breast Cancer in Eastern Mediterranean Region Countries: A Systematic Review and Meta-Analysis. *Annals of global health*. 2019;85(1):138.
12. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. Breast schema for 1988 Based on AJCC. Washington, DC, USA: National Cancer Institute; 1988.
13. Young JL. SEER summary staging manual 2000: codes and coding instructions: National Cancer Institute, National Institutes of Health; 2001.
14. Edge SBJS. *AJCC cancer staging manual*. 2010;7:97-100.
15. Goldstein BA, Polley EC, Briggs FB. Random forests for genetic association studies. *Statistical applications in genetics and molecular biology*. 2011;10(1):32.
16. Genuer R, Poggi J-M, Tuleau C. *Random Forests: some methodological insights*. 2008.
17. Probst P, Boulesteix A-L. *JMLR*. To tune or not to tune the number of trees in random forest. 2017;18(1):6673-90.
18. Oshiro T, Perez P, Baranauskas JJP, chapter How Many Trees in a Random Forest. *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MILDM 2012, Berlin, Germany, July 13-20, 2012*. 2012:154.
19. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in medicine*. 2011;30(10):1105-17.
20. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*. 1999;18(17-18):2529-45.
21. Ishwaran H, Kogalur U. Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC). R package version 2.12.12021.
22. Farhadian M, Dehdar Karsidani S, Mozayanimonfared A, Mahjub H. Risk factors associated with major adverse cardiac and cerebrovascular events following percutaneous coronary intervention: a 10-year follow-up comparing random survival forest and Cox proportional-hazards model. *BMC cardiovascular disorders*. 2021;21(1):38.
23. Qiu X, Gao J, Yang J, Hu J, Hu W, Kong L, et al. A Comparison Study of Machine Learning (Random Survival Forest) and Classic Statistic (Cox Proportional Hazards) for Predicting Progression in High-Grade Glioma after Proton and Carbon Ion Radiotherapy. *Frontiers in oncology*. 2020;10:551420.
24. Tse G, Lee S, Zhou J, Liu T, Wong ICK, Mak C, et al. Territory-Wide Chinese Cohort of Long QT Syndrome: Random Survival Forest and Cox Analyses. *Front Cardiovasc Med*. 2021;8:608592.