

Explicabilidade em Modelos Preditivos de Machine Learning no Câncer de Mama

Explainability in Machine Learning Predictive Models in Breast Cancer

Explicabilidad en Modelos Predictivos de Machine Learning en Cáncer de Mama

Erika Yahata^{1,2}, Erik Paul Winnikow³, Ricardo Suyama¹, Priscyla Waleska Simões^{1,2,4}

1 Programa de Pós-Graduação em Engenharia da Informação, Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas - CECS, Universidade Federal do ABC - UFABC, Santo André (SP), Brasil.

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

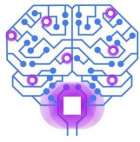
3 Curso de Medicina, Universidade do Extremo Sul Catarinense – UNESC, Criciúma (SC), Brasil.

4 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.

Autor correspondente: Profa Dra Priscyla Waleska Simões
E-mail: pritsimoes@gmail.com

Resumo

Objetivo: A Inteligência Artificial se mostra promissora como apoio à decisão no câncer de mama, porém, a interpretabilidade dos algoritmos como os de caixa preta pode contribuir na adoção na prática clínica. Esse estudo apresenta a explicabilidade em Modelos Preditivos de Aprendizado de Máquina no Câncer de Mama. **Métodos:** Avaliou-se duas abordagens distintas de Aprendizado de Máquina, Multilayer Perceptron (MLP) e Extreme Gradient Boosting (XGBoost), considerando amostra de 164 mulheres submetidas a Core Biópsia entre 2014 e 2015. Utilizou-se o Shapley Additive Explanation para a explicabilidade dos modelos. **Resultados:** Os modelos preditivos apresentaram, ambos, acurácia de 98,0% (IC95%:94,2%-100,0%) e o BI-RADS® 5 no ultrassom foi considerado como o atributo mais importante. **Conclusão:** Os modelos mostraram alta capacidade preditiva para o câncer de mama; no MLP o BI-RADS® 3 e 5 do ultrassom foram os atributos mais importantes, e no XGB, além do ultrassom, destacaram-se a idade e o nódulo palpável.



Descritores: Câncer de Mama; Aprendizado de Máquina; Inteligência Artificial

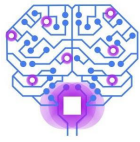
Abstract

Objective: Artificial Intelligence shows promise as decision support in breast cancer, however, the explainability of algorithms such as black box can contribute to adoption in clinical practice. This study presents the explainability in Predictive Machine Learning Models in Breast Cancer. **Methods:** Two different Machine Learning approaches were evaluated, the Multilayer Perceptron (MLP) and the Extreme Gradient Boosting (XGBoost), considering a sample of 164 women who underwent Core Biopsy between 2014 and 2015. The Shapley Additive Explanation was used to explain the models. **Results:** Both predictive models presented an accuracy of 98.0% (95%CI: 94.2%-100.0%) and the BI-RADS® 5 in the ultrasound was considered the most important attribute. **Conclusion:** The models showed high predictive capacity for breast cancer; in the MLP, the ultrasound BI-RADS® stages 3 and 5 were the most important attributes, and in the XGB model, in addition to ultrasound, age and palpable nodule were the most important.

Keywords: Breast Cancer; Machine Learning; Artificial Intelligence

Resumen

Objetivo: La Inteligencia Artificial se muestra prometedora como apoyo a la toma de decisiones en el cáncer de mama; sin embargo, la interpretabilidad de los algoritmos como la caja negra puede contribuir a su adopción en la práctica clínica. El presente estudio presenta la explicabilidad en modelos predictivos de Aprendizaje Automático en cáncer de mama. **Métodos:** Se evaluaron dos enfoques diferentes de Aprendizaje Automático, Multilayer Perceptron (MLP) y Extreme Gradient Boosting (XGBoost), considerando una muestra de 164 mujeres que se sometieron a Core Biopsy entre 2014 y 2015. Se utilizó el Shapley Additive Explanation para la explicabilidad de los modelos. **Resultados:** Ambos modelos predictivos presentaron una precisión del 98,0% (95%IC: 94,2%-100,0%) y el BI-RADS® 5 en el ultrasonido fue considerado como el atributo más importante. **Conclusión:** Los modelos mostraron alta capacidad predictiva para el cáncer de mama; en el MLP, los estadios BI-RADS® 3 y 5 del ultrasonido fueron los más



importantes, y en el modelo XGB, además del ultrasonido, se destacó la edad y el nódulo palpable.

Descritores: Câncer de Mama; Aprendizaje Automático; Inteligencia Artificial

Introdução

O câncer de mama é uma doença resultante da multiplicação de células anormais da mama que formam um tumor com potencial de invadir outros órgãos ⁽¹⁾. Em 2020, ocorreram cerca de 685.000 óbitos de mulheres por câncer de mama no mundo, correspondendo a 7% do total de mortes por câncer ⁽²⁾ sendo o câncer com maior incidência mundial, com 2,26 milhões de novos casos (11,7%), correspondendo a 47,8 casos por 100.000 habitantes ⁽³⁾.

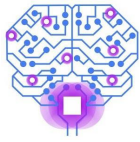
A detecção precoce, assim como a correta classificação do tipo de tumor, é fundamental para o encaminhamento do tratamento adequado de forma a aumentar a sobrevida e cura ^(4,5,6). Além disso, a análise por métodos tradicionais (estatísticos, por exemplo) de grandes quantidades de dados complexos de câncer de mama é laboriosa ⁽⁷⁾.

Nesse contexto, o Aprendizado de Máquina (AM) tem apresentado bons resultados no apoio ao diagnóstico ⁽⁸⁾, detecção ⁽⁹⁾, e prognóstico de câncer ⁽¹⁰⁾, porém, apesar dos resultados promissores, observa-se a lacuna do entendimento sobre como a predição tem sido realizada pelos modelos computacionais, em especial os “caixas pretas” (*black boxes*) ⁽⁹⁾. A explicabilidade dos modelos de AM tem sido considerada um desafio na saúde; assim evidências da interpretabilidade de tais modelos rotulados como “caixas pretas” podem contribuir na efetiva adoção na prática clínica ^(9,11,12).

Mediante o exposto, este artigo tem por objetivo apresentar a explicabilidade em Modelos Preditivos de Aprendizado de Máquina no Câncer de Mama.

Métodos

Pesquisa de natureza tecnológica e aplicada, que considerou um estudo transversal realizado no Sul do Brasil no período entre 2014 e 2015 com amostra de 164 mulheres submetidas a core biópsia e pelo menos um exame de imagem prévio,



totalizando 170 nódulos ⁽¹³⁾, aprovado pelo Comitê de Ética do local de origem da pesquisa, parecer 700.772.

Para a análise foram selecionados os atributos idade, histórico familiar, nódulo palpável, classificação Breast Imaging Reporting and Data System (BI-RADS[®]) para ultrassom (USG), classificação BI-RADS[®] para mamografia, tamanho do nódulo e tipo de nódulo (Tabela 1).

Tabela 1 - Atributos da amostra ⁽¹³⁾

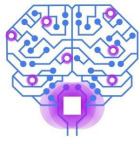
Atributo	Descrição
Idade	Maior ou igual a 50 anos, Menor que 50 anos
Histórico Familiar	Sim, Não
Nódulo Palpável	Sim, Não
BI-RADS [®] USG	BI-RADS [®] 3, BI-RADS [®] 4a, BI-RADS [®] 4b, BI-RADS [®] 4c, BI-RADS [®] 5
BI-RADS [®] Mamografia	BI-RADS [®] 0, BI-RADS [®] 1, BI-RADS [®] 2, BI-RADS [®] 3, BI-RADS [®] 4, BI-RADS [®] 5
Tamanho do nódulo	Maior que 2 cm, Menor ou igual a 2 cm
Tipo de nódulo	Maligno, Benigno

As variáveis quantitativas foram apresentadas pela média e desvio padrão, dicotomizadas e expressas por meio de frequência absoluta ou relativa.

Foram criadas variáveis indicadoras (*dummies*) para cada categoria BI-RADS[®] referente ao atributo mamografia e ultrassonografia devido as características dos algoritmos utilizados que aceitam somente dados numéricos e podem aumentar a acurácia das predições e fornecer informações mais detalhadas quando a variável é dicotomizada de acordo com as categorias originais ⁽¹⁴⁾.

Avaliou-se a correlação entre as variáveis pelo método de Pearson buscando observar a presença de alta correlação ($r > 0,9$) e que poderiam ser retiradas do modelo ⁽¹⁵⁾. A alta correlação entre variáveis resulta em estimativas instáveis devido à interdependência entre variáveis, generalização fraca do modelo e sobreajuste (*overfitting*). A correlação de 0,9 é usualmente utilizada como ponto de corte para altas correlações ^(16,17).

Dois algoritmos popularmente utilizados para dados estruturados, Multilayer Perceptron (MLP) e Extreme Gradient Boosting (XGBoost), que aceitam dados de vários tipos e são classificadores robustos ⁽¹⁸⁾, foram treinados com 70% dos dados e testados com 30% dos dados restantes, simulando novos dados desconhecidos. Todos os



modelos foram otimizados usando a busca randomizada (*randomized search*) dos valores dos hiperparâmetros e com validação cruzada (*cross validation*) de 3-fold com maximização da acurácia.

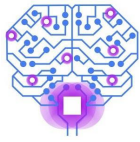
Tabela 2 - Hiperparâmetros

Modelo	Hiperparâmetro	Descrição	Valores Analisados	Valor escolhido
MLP	Hidden Layer Sizes	Número de neurônios na camada oculta da rede	{4,8,16,32,64,128}	128
	Activation	Função de ativação para as camadas ocultas	tanh,relu	tanh
	Solver	Otimizador de peso	sgd, adam	adam
	Learning Rate	Taxa de aprendizado	constant, invscaling	constant
XGB	No. Trees	Número de árvores	[10,500] espaçamento 20	410
	Subsample	Proporção de subamostras de treinamento	{0.1, 0.2, 0.3, 0.5, 0.7, 0.9, 1}	0.5
	Min. Child Weight	Soma de pesos mínimos observados	{1, 2, 3, 4}	1
	Max. Depth	Número máximo de níveis em cada árvore de decisão	{10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110}	100
	Learning Rate	Taxa de aprendizado	[0,1] espaçamento 0.01	0.19
	Colsample by Tree	Fração de colunas das amostras para cada árvore	{0.5, 0.7, 0.9, 1}	0.9

A Tabela 2 mostra os hiperparâmetros explorados para cada modelo, assim como a combinação de valores finais utilizados. A linguagem Python, pacote scikit-learn (v. 1.0.2) e xgboost (v. 0.90) foram utilizados no estudo.

Para a avaliação de desempenho dos algoritmos foram consideradas a acurácia, a sensibilidade, a especificidade, a curva característica de operação do receptor (ROC) e a área sob a curva ROC (AUC):

- Verdadeiros Positivos (VP): casos positivos corretamente classificados.
- Verdadeiros Negativos (VN): casos negativos corretamente classificados.
- Falsos Positivos (FP): casos negativos classificados incorretamente como positivos.
- Falsos Negativos (FN): casos positivos classificados incorretamente como negativos.



Na acurácia avalia-se a proporção de todas as classificações corretas, VP e VN, realizadas pelos modelos preditivos sobre todos os resultados obtidos, ou seja,

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (1)$$

Na sensibilidade avalia-se a taxa de verdadeiros positivos, isto é, a capacidade do algoritmo classificar corretamente os nódulos malignos. O cálculo da sensibilidade é dado pela divisão dos VP pela soma dos VP e FN, isto é,

$$Sensibilidade = \frac{VP}{VP + FN} \quad (2)$$

Já na especificidade avalia-se a taxa de verdadeiros negativos, isto é, a classificação correta de nódulos benignos. A especificidade é dada pela divisão dos VN pela soma dos VN e FP. Matematicamente, a especificidade é definida por:

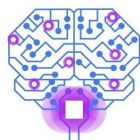
$$Especificidade = \frac{VN}{VN + FP} \quad (3)$$

A curva ROC é utilizada para avaliar graficamente o ajustamento entre especificidade e sensibilidade ⁽¹⁸⁾. A AUC é usada para determinar o desempenho de predição do modelo classificador no qual valores mais altos representam melhores preditores.

Para a compreensão dos atributos mais relevantes utilizou-se o SHapley Additive exPlanations (SHAP) devido à sua versatilidade para qualquer tipo de modelo (*model-agnostic*) e a presença de propriedades como a acurácia e a consistência que não são encontrados simultaneamente em outras técnicas ⁽¹⁰⁾.

Resultados e discussão

As características da amostra são apresentadas na Tabela 3:

**Tabela 3** - Características da amostra

Atributos	Tumor Mamário N (%)		
	Total n=170	Maligno n=45	Benigno n=125
Idade			
Maior ou igual a 50 anos	63 (37,1)	27 (42,9)	36 (57,1)
Menor que 50 anos	107 (62,9)	18 (16,8)	89 (83,2)
Histórico Familiar			
Sim	16 (9,4)	3 (18,8)	13 (81,3)
Não	154 (90,6)	42 (27,3)	112 (72,7)
Nódulo Palpável			
Sim	77 (45,3)	37 (48,1)	40 (51,9)
Não	93 (54,7)	8 (8,6)	85 (91,4)
BI-RADS USG			
BI-RADS 3	50 (29,4)	0	50 (100,0)
BI-RADS 4a	68 (40,0)	4 (5,9)	64 (94,1)
BI-RADS 4b	13 (7,7)	4 (30,8)	9 (69,2)
BI-RADS 4c	8 (4,7)	6 (75,0)	2 (25,0)
BI-RADS 5	31 (18,2)	31 (100,0)	0
BI-RADS Mamografia			
BI-RADS 0	28 (16,5)	6 (21,4)	22 (78,6)
BI-RADS 1	27 (15,9)	3 (11,1)	24 (88,9)
BI-RADS 2	6 (3,5)	0	6 (100,0)
BI-RADS 3	20 (11,8)	0	20 (100,0)
BI-RADS 4	27 (15,9)	10 (37,0)	17 (63,0)
BI-RADS 5	18 (10,5)	18 (100,0)	0
Não informado	44 (25,9)	8 (18,2)	36 (81,8)
Tamanho do nódulo			
Maior que 2 cm	56 (33,0)	24 (45,3)	29 (54,7)
Menor que 2 cm	114 (67,0)	21 (17,9)	96 (82,1)

Verificou-se que não houve correlação maior que 0,9 entre as variáveis (Figura 2), portanto, como não houve correlação forte, nenhuma variável foi excluída do modelo.

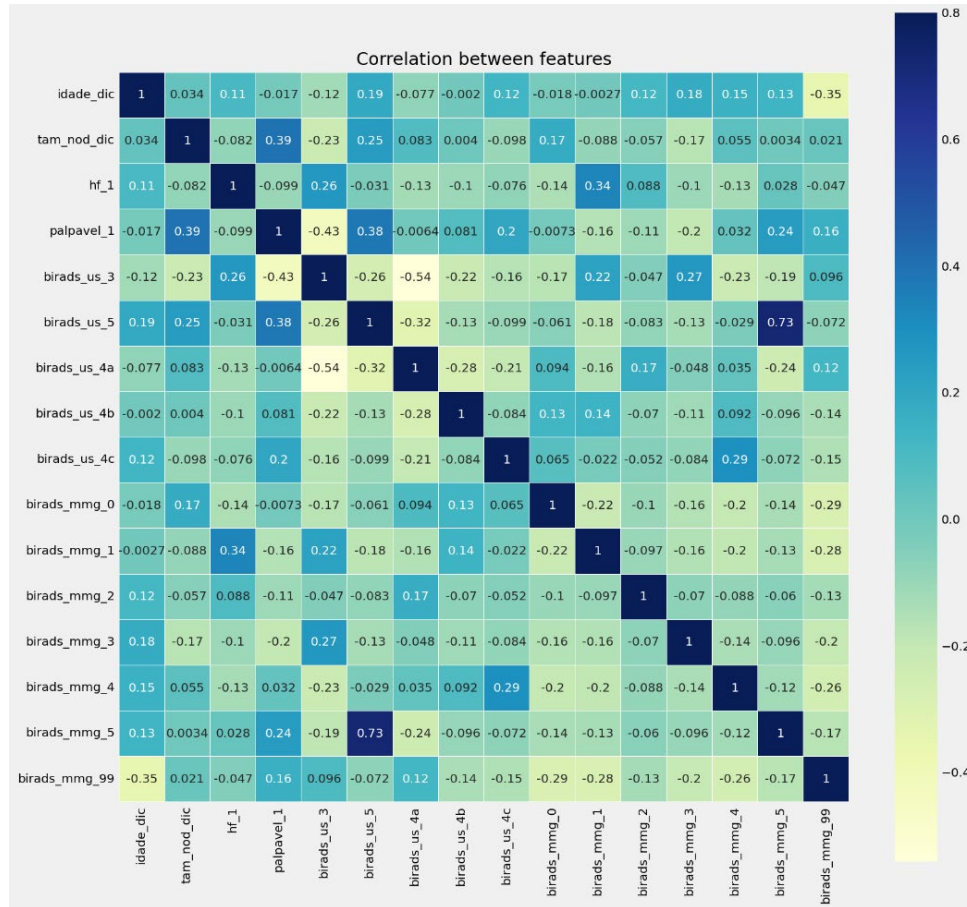
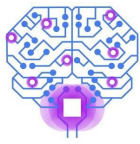


Figura 2 - Correlação entre os atributos

Após o treinamento realizou-se a avaliação dos modelos (Tabela 4). Os modelos MLP e XGBoost apresentaram acurácia de 98,0% (IC95%: 94,2%-100,0%), ou seja, proporção de 98,0% de classificações corretas considerando todos os resultados do teste. A sensibilidade de 94,1% (IC95%: 82,9%-100,0%) e especificidade de 100,0% (IC95%: 100,0%-100,0%) em ambos os modelos mostram que houve correta classificação do tumor maligno em 94,1% dos casos, e de 100,0% para os nódulos benignos. O modelo XGBoost apresentou a AUC de 0,99 (IC95%: 0,95-1,00) e o MLP, 0,98 (IC95%: 0,93-1,00).

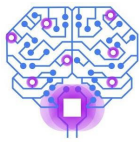


Tabela 4 - Desempenho dos modelos

Variável	Modelo	
	MLP Medida de avaliação (IC 95%)	XGBoost Medida de avaliação (IC 95%)
Acurácia	98,0% (94,2% – 100,0%)	98,0% (94,2% - 100,0%)
Sensibilidade	94,1% (82,9% - 100,0%)	94,1% (82,9% - 100,0%)
Especificidade	100,0% (100,0% – 100,0%)	100,0% (100,0% - 100,0%)
ROC (AUC)	0,98 (0,93 – 1,00)	0,99 (0,95 - 1,00)

Na Figura 3, observa-se pela curva ROC que ambos os modelos mostraram comportamento que pode ser considerado de habilidade preditiva, e os valores AUC demonstram que os modelos XGB (0,99; IC95%: 0,95-1,00) e MLP (0,98; IC95%: 0,93-1,00) têm alta proporção de realizar uma predição correta para o tumor maligno, e para o tumor benigno.

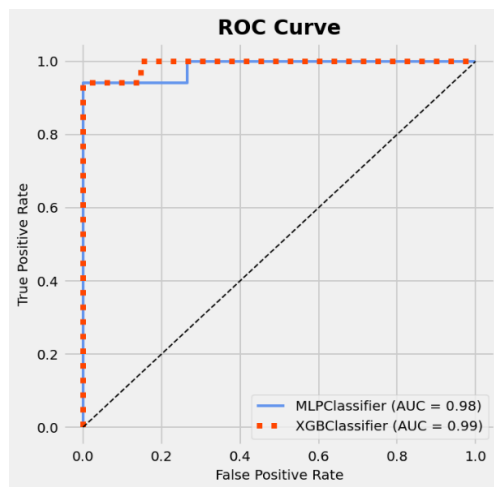
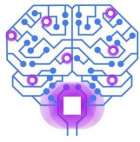


Figura 3 - Curva ROC (AUC) dos modelos MLP e XGBoost

A Figura 4 apresenta os gráficos SHAP que mostram a importância dos atributos e o efeito de seus valores nos modelos. No SHAP, o impacto de cada atributo para a predição é calculado por métodos combinatórios e avalia-se o modelo com todas as combinações de atributos utilizados ⁽¹⁹⁾. Para a interpretação gráfica, no eixo y são apresentados os atributos em ordem de importância, do mais importante para o menos importante na contribuição para o modelo; e no eixo x são apresentados os valores SHAP sendo os valores positivos sugestivos de suporte para o desfecho maligno, e os valores negativos indicam suporte ao desfecho benigno. Cada ponto do gráfico representa uma



determinada paciente, e as cores representam o valor do atributo no qual tons mais vermelhos referem-se a valores mais altos e tons azulados são valores mais baixos do atributo.

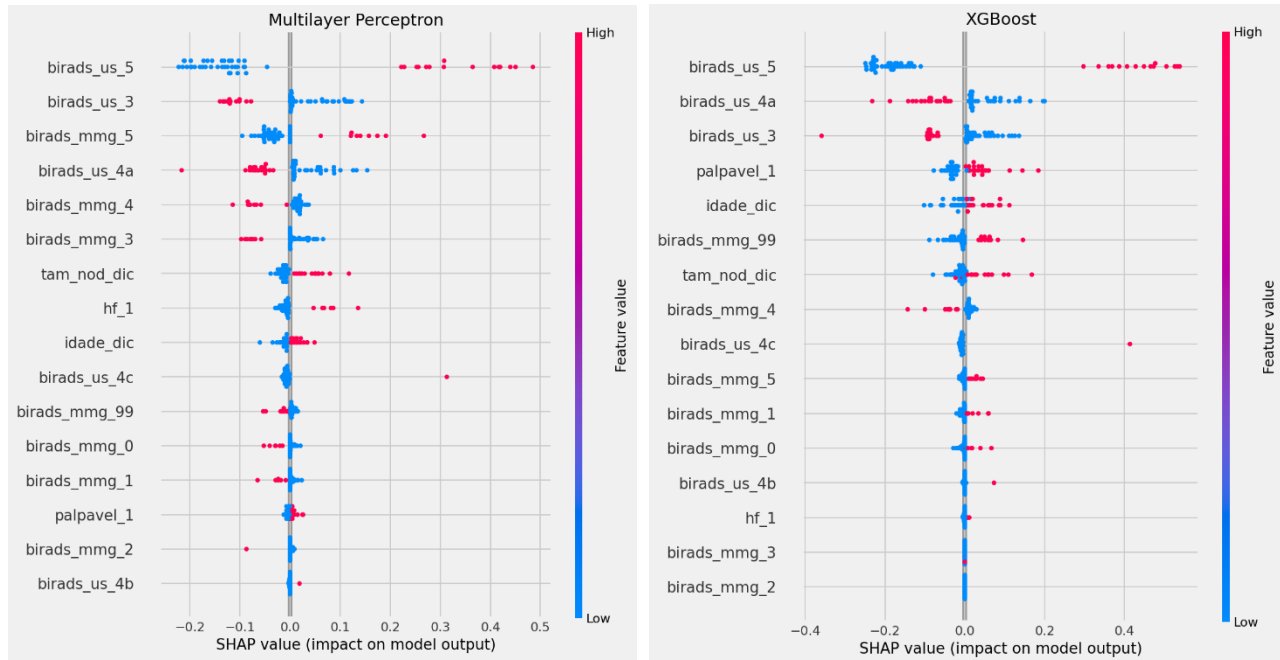
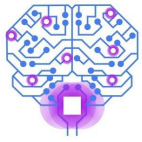


Figura 4 - Gráfico SHAP

Observa-se que o ultrassom BI-RADS® 5 foi o atributo mais importante em ambos os modelos além de apresentar divisão bem clara de cores, ou seja, pontos vermelhos e azuis em oposição. Os pontos vermelhos representam que instâncias BI-RADS® 5 foram fortes predictoras do tumor maligno. O ultrassom BI-RADS® 3 e 4a apresentaram-se entre os 5 atributos mais importantes para os modelos, sendo que instâncias nestas categorias contribuem para a predição do nódulo benigno. Porém, a mamografia BI-RADS® 5 foi considerada entre as mais importantes para o modelo MLP enquanto no modelo XGB a idade e o nódulo palpável estiveram entre os 5 atributos mais importantes para a classificação. Em ambos os modelos, o tamanho do nódulo foi classificado como o 7º atributo mais importante.

A utilização do atributo referente ao ultrassom BI-RADS® 5 como o mais importante para a predição de tumor maligno está de acordo com o observado num estudo realizado na Alemanha ⁽²⁰⁾ entre 2006 e 2009 que avaliou a acurácia da classificação BI-RADS®



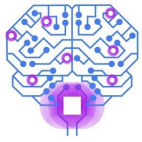
em ultrassom de mama, no qual 96,8% dos casos foram classificados como tumor maligno quando em estágio BI-RADS® 5 ⁽²⁰⁾. Além disso, um estudo retrospectivo realizado na Coreia ⁽²¹⁾ que pesquisou a associação BI-RADS® e características clínico patológicas e prognósticas do câncer de mama, considerou dados de mamografia e ultrassonografia de 44.184 pacientes diagnosticadas com câncer de mama invasivo entre abril de 1983 a dezembro de 2014, e mostrou que a proporção de pacientes da categoria BI-RADS® 5 (80,0%) foi maior se comparada ao BI-RADS® 3 e BIRADS® 4 (20,0%) ⁽²¹⁾.

O modelo MLP de nosso estudo também considerou a mamografia BI-RADS® 5 como uma das mais importantes, o que não ocorreu no modelo XGB, sendo necessário avaliar se este resultado possa estar associado ao fato de haver pacientes sem exames de mamografia (25,88%) o que possivelmente levou a classificação do atributo sem exame (birads_mmg_99) se tornar importante para o modelo XGB. Tal resultado pode sugerir também que os exames de detecção e rastreamento estão sendo pouco realizados na região pesquisada, assim, mais ações podem ser realizadas sobre a importância da mamografia no rastreamento do câncer de mama.

No modelo XGB, o atributo idade foi classificado como um dos mais importantes, característica essa corroborada por estudo que mostra a idade avançada como fator de risco para o câncer de mama ⁽²²⁾. Aproximadamente 99,3% das mortes por câncer de mama em 2016 ocorreram em mulheres com mais de 40 anos e 71,2% em mulheres com mais de 60 anos ⁽²²⁾.

No relatório do The Surveillance, Epidemiology, and End Results (SEER) ⁽²³⁾ sobre a incidência de câncer de mama nos Estados Unidos entre 2015 e 2019, mulheres acima dos 45 anos de idade representaram 89,7% dos casos de câncer, e a faixa etária entre 45 e 74 anos, maior incidência (70,9%) ⁽²³⁾. A mortalidade também foi elevada a partir dos 45 anos (94,6%), e pacientes entre 55 e 84 anos apresentaram mortalidade de 62,6% ⁽²³⁾. Corroborando essa questão, estudo do Reino Unido ⁽²⁴⁾ realizado entre 2017 e 2019, mostrou que 48% das mortes por câncer de mama foram em mulheres com idade superior a 75 anos, revelando alta incidência da doença e diminuição da sobrevivência nessa faixa etária ⁽²⁴⁾.

O nódulo papável foi considerado um atributo importante no modelo XGB corroborando sua contribuição na prática clínica. Na realização de exames clínicos, a



presença de um pequeno nódulo palpável é um dos indicativos mais importantes do câncer de mama ⁽²⁵⁾. Um estudo realizado na Coreia ⁽²¹⁾ entre 1983 e 2014 com 44.184 pacientes diagnosticadas com câncer de mama invasivo mostrou que o nódulo palpável apresentou forte associação com o BI-RADS[®] 5 tanto na mamografia (OR=1,916; $p < 0,001$) quanto na ultrassonografia (OR=2,519; $p < 0,001$), a partir do qual cerca de 90% das pacientes apresentaram BI-RADS[®] 5 em pelo menos um dos exames de imagem e possuíam nódulo palpável, reforçando que exames para a detecção do câncer de mama devem ser realizados quando houver nódulo palpável no exame físico ⁽²¹⁾.

Conclusões

Os modelos MLP e XGBoost mostraram alta capacidade preditiva para o câncer de mama. O modelo MLP apresentou os estágios BI-RADS[®] 3 e 5 do ultrassom como os mais importantes, enquanto para modelo XGB, os estágios BI-RADS[®] do ultrassom de 3 a 5, a idade e o nódulo palpável foram os mais importantes. Assim, as evidências sugerem que os resultados dos modelos são promissores para apoio a decisão e que a explicabilidade apresentada nesse estudo pode fortalecer o entendimento e incorporação de modelos de Aprendizado de Máquina na prática clínica.

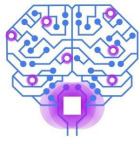
Agradecimentos

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).

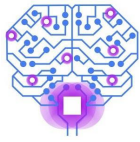
Referências

1. MINISTÉRIO DA SAÚDE. Câncer de mama: vamos falar sobre isso? [Internet]. 2021 [cited 2022 jul 01]. Available from:



<https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//cartilha-mama-6-edicao-2021.pdf>

2. IARC marks Breast Cancer Awareness Month 2021 – IARC [Internet]. 2021 [cited 2022 Aug 12]. Available from: <https://www.iarc.who.int/news-events/iarc-marks-breast-cancer-awareness-month-2021/>
3. World Health Organization International Agency for Research on Cancer (IARC). Cancer today [Internet]. 2021. [cited 2022 Aug 12]. Available from: <https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf>
4. American Cancer Society. American cancer society recommendations for the early detection of breast cancer [Internet]. 2022 [cited 2022 jul 22]. Available from: <https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/american-cancer-society-recommendations-for-the-early-detection-of-breast-cancer.html>
5. Caplan, L. Delay in breast cancer: implications for stage at diagnosis and survival. *Front Public Health*. 2014 Jul;2:87.
6. Verma B, McLeod P, Klevansky A. Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer. *Expert Systems with Applications*. 2010 Apr;37(4):3344–51
7. Chugh, G., Kumar, S., Singh, N. Survey on Machine Learning and Deep Learning Applications in Breast Cancer Diagnosis. *Cogn Comput*. 2021 Jan;13:1451–1470.
8. Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. *J Pers Med*. 2020 Mar;10(2):21.
9. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med*. 2018 Dec;284(6):603-619
10. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Sci Rep*. 2021 Mar 26;11(1):6968.
11. Amann J, Blasimme A, Vayena E, Frey D, Madai VI; Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak*. 2020 Nov 30;20(1):310.
12. Cutillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD; MI in Healthcare Workshop Working Group. Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digit Med*. 2020 Mar 26;3:47.
13. Agostin G de, Formigoni RR, Yokoi LM, Winnikow EP, Simões PW. Redes Bayesianas e Regressão Logística em pacientes submetidas a core biópsia. *Journal of Health Informatics*. 2021 Mar 15 ;12.



14. Dahouda MK, Joe I. A Deep-Learned Embedding Technique for Categorical Features Encoding. *IEEE Access*. 2021;9:114381–91
15. Khalid S, Khalil T, Nasreen S. A survey of feature selection and feature extraction techniques in machine learning. In: 2014 science and information conference; 2014 Aug. p. 372-378.
16. Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med*. 2018 Aug 7;18(3):91-93.
17. Chan JY-L, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong Z-W, et al. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics*. 2022 Apr 12;10(8):1283
18. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak*. 2019 Dec 21;19(1):281.
19. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* [Internet]. 2022. [cited 2022 jul 15]. Available from: christophm.github.io/interpretable-ml-book/
20. Hille H, Vetter M, Hackelöer BJ. The accuracy of BI-RADS classification of breast ultrasound as a first-line imaging method. *Ultraschall Med*. 2012 Apr;33(2):160-3.
21. Kim BK, Ryu JM, Oh SJ, Han J, Choi JE, Jeong J, Suh YJ, Lee J, Sun WY; Korean Breast Cancer Society. Comparison of clinicopathological characteristics and prognosis in breast cancer patients with different Breast Imaging Reporting and Data System categories. *Ann Surg Treat Res*. 2021 Sep;101(3):131-139.
22. Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP. Risk Factors and Preventions of Breast Cancer. *Int J Biol Sci*. 2017 Nov 1;13(11):1387-1397.
23. National Cancer Institute. *Female Breast Cancer - Cancer Stat Facts* [Internet]. SEER. 2022 [cited 2022 jul 01]. Available from: <https://seer.cancer.gov/statfacts/html/breast.html>
24. Cancer Research UK. *Breast cancer mortality statistics* [Internet]. 2015. [cited 2022 jul 15]. Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/mortality#heading-One>
25. IARC. *Breast Cancer Screening* [Internet]. 2016 [cited 2022 jul 15]. Available from: <https://publications.iarc.fr/Book-And-Report-Series/Iarc-Handbooks-Of-Cancer-Prevention/Breast-Cancer-Screening-2016>