

**USO DA INTELIGÊNCIA ARTIFICIAL NA IDENTIFICAÇÃO DE TUMORES
HER2 POSITIVOS E HER2-LOW, USANDO APENAS LÂMINAS CORADAS
POR HEMATOXILINA-EOSINA (H&E), EM PACIENTES COM CÂNCER DE
MAMA DO HOSPITAL AC CAMARGO CANCER CENTER**

JOSÉ MÁRCIO BARROS DE FIGUEIREDO

Trabalho aplicado apresentada ao Programa de Pós- Graduação de Mestrado Profissional na área de Cuidados oncológicos centrados no paciente da Fundação Antônio Prudente para obtenção do título de mestre

Orientador: PROF. DR. ISRAEL TOJAL DA SILVA

São Paulo

2023

FICHA CATALOGRÁFICA

Figueiredo, José Márcio Barros de.

USO DA INTELIGÊNCIA ARTIFICIAL NA IDENTIFICAÇÃO DE TUMORES HER2 POSITIVOS E HER2-LOW, USANDO APENAS LÂMINAS CORADAS POR HEMATOXILINA-EOSINA (H&E), EM PACIENTES COM CÂNCER DE MAMA DO HOSPITAL AC CAMARGO CANCER CENTER . / José Márcio Barros de Figueiredo. São Paulo, 2023.

42f.

Trabalho aplicado (Mestrado Profissional) - Fundação Antônio Prudente. Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.

Orientador: PROF. DR. ISRAEL TOJAL DA SILVA.

1. Patologia digital, 2. câncer de mama, 3. Her2 baixo

CDU 616

Folha de aprovação

Esse item é uma cópia da folha de rosto, após defesa e inserido no exemplar definitivo, **preparada pela secretaria da PG** para mestrado e doutorado com os componentes da banca, inserido no arquivo após a ficha catalográfica, **Ex:**

JOSÉ MÁRCIO BARROS DE FIGUEIREDO

USO DA INTELIGÊNCIA ARTIFICIAL NA IDENTIFICAÇÃO DE TUMORES HER2 POSITIVOS E HER2-LOW, USANDO APENAS LÂMINAS CORADAS POR HEMATOXILINA-EOSINA (H&E), EM PACIENTES COM CÂNCER DE MAMA DO HOSPITAL AC CAMARGO CANCER CENTER

Aprovado em: __22__ / __12__ / __2023__

Banca Examinadora

Dr. Luiz Juliano Neto

Instituição: AC Camargo Cancer Center

Julgamento: Aprovado

Dr. Antônio Carlos Cavalcante Godoy

Instituição: Instituto do Câncer Brasil

Julgamento: Aprovado

Dr. Israel Tojal da Silva

Instituição: AC Camargo Cancer Center

Julgamento: Aprovado

RESUMO

Barros de Figueiredo, José Márcio. **Uso da Inteligência Artificial na Identificação de Tumores HER2 Positivos e HER2-low, usando apenas Lâminas Coradas por Hematoxilina-Eosina (H&E), em Pacientes com Câncer de Mama do Hospital AC Camargo Cancer Center** [Trabalho aplicado]. São Paulo; Fundação Antônio Prudente; 2023.

INTRODUÇÃO: Até 15% dos casos de câncer de mama são classificados como HER2 positivo e 60% são classificados como HER2-low. A detecção desses biomarcadores é essencial para o prognóstico e predição de resposta ao tratamento, mas podem apresentar erros subjetivos. A inteligência artificial (IA) tem se mostrado promissora na avaliação de biomarcadores e, embora tenha apresentado boa acurácia na identificação do HER2, a avaliação de HER2-low ainda permanece a ser investigada. **OBJETIVO:** Desenvolver um modelo de IA para a detecção de HER2 e Her2-low. **MATERIAIS E MÉTODOS:** Uma abordagem de aprendizagem profunda foi utilizada em um conjunto de imagens digitalizadas a partir dos casos de câncer de mama HER2 diagnosticados no A.C. Camargo Cancer Center. **RESULTADOS:** O algoritmo alcançou uma AUC de 0,75 para HER2+ e 0,78 para Her-low. **CONCLUSÃO:** A performance do modelo é similar a outros estudos para HER2 (negativo vs positivo) e este é o segundo na literatura a descrever a identificação de HER2-low, também com uma acurácia similar. O uso de IA tem o potencial para ser uma ferramenta valiosa na identificação de biomarcadores em casos de câncer de mama diretamente a partir da imagem histológica, podendo auxiliar na escolha do tratamento adequado e no prognóstico do paciente.

Palavras-chave: Aprendizado de máquinas; Patologia digital; Her2+; Her2 baixo

ABSTRACT

Barros de Figueiredo, José Márcio. **Use of Artificial Intelligence in the Identification of HER2 Positive and HER2-low Tumors, using only Hematoxylin-Eosin (H&E) stained slides, in Breast Cancer Patients at Hospital AC Camargo Cancer Center** [Trabalho aplicado]. São Paulo; Fundação Antônio Prudente; 2023.

INTRODUCTION: Up to 15% of breast cancer cases are classified as HER2 positive and 60% are classified as HER2-low. The identification of these biomarkers is essential for prognosis and prediction of treatment response but can be subject to subjective errors. Artificial intelligence (AI) has shown good accuracy in identifying HER2 but has not been tested for HER2 low. **OBJECTIVE:** To evaluate the area under the curve (AUC) for the identification of HER2 and HER2-low using a predictive AI model. **MATERIALS AND METHODS:** Retrospective study of the AC Camargo database with breast cancer cases classified as HER2 negative and positive, defined as those with a score of 0 and 3+ by IH or FISH, respectively. **RESULTS:** The algorithm achieved an AUC of 0.75 for HER2 and 0.78 for HER2-low. **CONCLUSION:** The model's performance is like other studies for HER2 and is the second in the literature to describe the identification of HER2 low, also with good accuracy. The use of AI can be a valuable tool in identifying biomarkers in breast cancer cases, helping to choose appropriate treatment and patient prognosis.

Keywords: Machine learning; Digital pathology; Her2+; Her2 low

LISTA DE FIGURAS

- FIGURE 1** - FLUXO DE TRABALHO PARA O MÉTODO PROPOSTO. (A) AS IMAGENS DO SLIDE INTEIRO SÃO DIVIDIDAS EM FRAGMENTOS; (B) AS REGIÕES DE INTERESSE SÃO IDENTIFICADAS; (C) OS RECURSOS SÃO EXTRAÍDOS DE IMAGENS DE BLOCOS; (D) UM MODELO DE APRENDIZAGEM DE MÚLTIPLAS INSTÂNCIAS É TREINADO PARA CLASSIFICAR CONJUNTOS DE BLOCOS HER2 NEGATIVO, HER2 BAIXO (HER2-LOW) E HER2 ALTO..... 13
- FIGURE 2** - UROC DOS RESULTADOS DA VALIDAÇÃO CRUZADA DE 10 VEZES. CADA MODELO FOI TREINADO COM SEU PRÓPRIO CONJUNTO DE PONTUAÇÕES HER2 QUE COMPÕE CADA GRUPO HER2 ALVO. AS PONTUAÇÕES HER2 UTILIZADAS SÃO INDICADAS PELOS PONTOS ATIVOS, ENQUANTO A COR DO PONTO INDICA SEU GRUPO HER2 ALVO PARA O MODELO..... 14
- FIGURE 3** - LÂMINA CORADA COM H&E DE UMA AMOSTRA DE TECIDO E SEU MAPA DE CALOR DE ATENÇÃO CORRESPONDENTE GERADO PELO MODELO PROPOSTO. O MAPA DE CALOR MOSTRA AS ÁREAS ONDE O MODELO CONCENTRA MAIS ATENÇÃO, COM CORES MAIS QUENTES INDICANDO MAIOR ATENÇÃO PONTUAÇÕES. OS PONTOS MARCADOS MOSTRAM A REGIÃO SELECIONADA PELO FILTRO PRELIMINAR, DESCRITO NA FIGURA 1 (B). 16
- FIGURE 4** - ANÁLISE DE EXPLICABILIDADE DO MODELO. (A) PLOTAMOS O ESPAÇO DE INCORPORAÇÃO DOS FRAGMENTOS/TILES USANDO PCA PARA CADA MODELO E FOLDS, COLORINDO CADA REGIÃO DE ACORDO COM O GRUPO HER2 QUE PRESTA ATENÇÃO A ESSA INCORPORAÇÃO. REGIÕES VAZIAS INDICAM INCORPORAÇÕES QUE SÃO COMPARTILHADOS POR VÁRIAS CLASSES PARA CLASSIFICAÇÃO. (B) CAIXAS COLORIDAS INDICAM RECURSOS USADOS POR 40% DOS FOLDS DENTRO DE CADA MODELO. (C) ANALISAMOS A CONSISTÊNCIA DO MODELO CONTANDO QUANTOS MODELOS USAM A MESMA INCORPORAÇÃO PARA CLASSIFICAR CADA CLASSE, COLORINDO AS REGIÕES DE ACORDO. MOSTRAMOS TRÊS NÍVEIS DE CONCORDÂNCIA: I. 2 OU MAIS; II. 3 OU MAIS; III. 4 OU MAIS. (D) EXEMPLOS DE PEÇAS DO ESPAÇO DE INCORPORAÇÃO PARA CADA CLASSE, ILUSTRANDO COMO ELAS SE RELACIONAM COM AS CARACTERÍSTICAS DE INTERESSE..... 18

LISTA DE TABELAS

TABELA 1 - SUBTIPOS FENOTÍPICOS MEDIANTE A CLASSIFICAÇÃO POR IMUNO-HISTOQUÍMICA.....	2
TABELA 2 - ACHADOS HISTOLÓGICOS ASSOCIADOS A FRAGMENTOS ACIONÁVEIS NOS GRUPOS HER2.	18
TABELA 3 - MEDIANA DE AUROC EM DIFERENTES CENÁRIOS DE TREINAMENTO CONCLUÍMOS QUE A FILTRAGEM DE FRAGMENTOS COMO ETAPA DE PRÉ-PROCESSAMENTO É IMPORTANTE PARA O DESEMPENHO PREDITIVO DO MODELO. VALE RESSALTAR QUE DECIDIMOS USAR TUMO, POIS PROPORCIONOU O MELHOR EQUILÍBRIO ENTRE O DESEMPENHO PREDITIVO DO MODELO E OS REQUISITOS DE HARDWARE. CONSIDERANDO COMO O MÉTODO MIL FUNCIONA, ACREDITAMOS QUE A ETAPA DE FILTRAGEM DE FRAGMENTOS FORÇA O MODELO A APRENDER CARACTERÍSTICAS DISTINTIVAS DOS FRAGMENTOS COM INFORMAÇÕES REAIS, EM VEZ DE SE CONCENTRAR EM FRAGMENTOS IRRELEVANTES.....	20
TABELA 4 - CARACTERÍSTICAS HISTOLÓGICAS ESTRATIFICADAS POR CADA GRUPO HER2.....	20

LISTA DE ABREVIATURAS

- AI** inteligência artificial (do inglês, *artificial intelligence*)
- AUC** área sobrea a curva (do inglês, *Area Under Curve*)
- ASCO** Associação Americana de Oncologia (do inglês, *American Society of Clinical Oncology*)
- CA** câncer de mama
- CAP** Colégio Americano de Patologistas (do inglês, *College of American Pathologists*)
- CDI** carcinoma ductal invasivo
- CDIS** carcinoma ductal *in situ*
- CLI** carcinoma lobular invasivo
- CLIS** carcinoma lobular *in situ*
- EGFR** Receptor do fator de crescimento epidérmico (do inglês, *The epidermal growth factor receptor family*)
- ER** Receptor de estrogênio (do inglês, *The estrogen receptor*)
- FISH** hibridização fluorescente in situ, (do inglês, *fluorescence in situ hybridization*)
- H&E** coloração de hematoxilina e eoxina (do inglês, *hematoxylin and eosin stain*)
- HER2** Receptor do fator de crescimento epidérmico (do inglês, *The human epidermal growth factor receptor 2*)
- HER2-low** não enriquecido com HER2 (HER2+), IHQ + ou ++
- HER2 negativo** não enriquecido com HER2 (HER2+)
- HER2 positivo** enriquecido com HER2 (HER2+)
- IHQ** imuno-histoquímica
- KI67** marcador de proliferação KI67 (do inglês, *The antigen KI-67, marker of proliferation ki-67*)
- Luminal A** ER+/HER2-/PR+ e fenótipo proliferativo baixo
- Luminal B** ER+/HER2-/PR- ou fenótipo proliferativo KI67 baixo e alto (>14%) ou ER+/PR- ou baixo + e HER2 + com fenótipo proliferativo KI67 alto (>14%)
- MIL** aprendizagem de múltiplas instâncias (do inglês *multiple instance learning*)
- PARP** poli (ADP-ribose) polimerase (do inglês, *Poly ADP-ribose polymerase*)
- PDL1** ligante 1 de morte programada celular (do inglês, *programmed death-ligand 1*)
- PIK3CA** proteína fosfatidilinositol-4,5-bifosfato 3-quinase (do inglês, *The phosphatidylinositol-4,5-bisphosphate 3-kinase*).

SPHC Síndromes de predisposição hereditária ao câncer

TCGA O atlas do genoma do câncer (do inglês, The cancer genome atlas)

TNBC- ER-/HER2-/PR câncer de mama triplo negativo (do inglês, triple negative breast cancer).

TUMIN tumor invasivo

RNN rede neural recorrente (do inglês, recurrent neural network)

RP receptor de progesterona (do inglês, The progesterona receptor).

SUMÁRIO

1. INTRODUÇÃO.....	1
2. OBJETIVOS.....	8
2.1 Objetivo geral.....	8
2.2 Objetivos específicos.....	8
2.3. Hipótese.....	8
3. MÉTODOS.....	9
3.1 Desenho do estudo.....	9
3.2 Remoção do fundo e detecção do tumor.....	9
3.3 Extração de características.....	10
3.4 Definição do modelo.....	10
3.5 Modelo de treinamento.....	11
3.6 Análises dos fragmentos (tiles) acionáveis.....	11
4 RESULTADOS.....	13
4.1 Uma estrutura de aprendizado profundo de supervisão fraca para a avaliação do HER2.....	13
4.2 Desempenho dos modelos revela que a classificação do HER2-baixo é desafiadora.....	13
4.3 A validação cruzada destaca a estabilidade do modelo.....	14
4.4 Fragmentos acionáveis revelam padrões usados pelo classificador.....	15
4.5 Filtrando as regiões de interesse.....	19
5 DISCUSSÃO.....	23

6. CONSIDERAÇÕES ÉTICAS..... 27

7. REFERÊNCIAS..... 28

1. INTRODUÇÃO

As neoplasias malignas são atualmente um grave problema de saúde pública, sendo a segunda causa de morte no Brasil, com 260.000 óbitos em 2020¹. A estimativa para cada ano do triênio 2023-2025 aponta que ocorrerão 704 mil casos novos de câncer (483 mil, excluindo os casos de câncer de pele não melanoma), demonstrando o quanto o câncer impacta na mortalidade da população². Além disso, dentre os pacientes diagnosticados com câncer, aproximadamente 30% dos indivíduos apresentam história familiar com múltiplos casos de neoplasias, sendo que 5-10% estão relacionados a padrões hereditários causados por mutações germinativas em oncogenes e genes de supressão tumoral².

O câncer de mama (CA), é a neoplasia mais frequente diagnosticada em mulheres e a segunda causa mais comum de morte por câncer no mundo nessa população³. Vários fatores de risco, incluindo idade, histórico familiar, suscetibilidade genética e estilo de vida pouco saudável, contribuem para a alta incidência de câncer de mama³. Em geral, o câncer de mama tem sido considerado uma doença heterogênea e os dois subtipos morfológicos principais são divididos em carcinoma ductal: *in situ* (CDIS) e carcinoma ductal invasivo (CDI), que representam de 70% a 80% e o carcinoma lobular que também pode ser subclassificado em *in situ* (CLIS) ou invasivo (CLI)³.

A avaliação histológica do câncer de mama pelo patologista é um dos fatores prognósticos estabelecidos e ainda indispensável na prática diária⁴. A identificação dos biomarcadores receptores de estrogênio, do inglês *estrogen receptor* (ER), receptores de progesterona, do inglês *progesterone receptor* (PR), marcador de proliferação KI67, do inglês *Antigen KI-67 Marker Of Proliferation* (Ki-67%) e membros da superfamília receptor tipo 2 do fator de crescimento epidérmico humano-HER2, do inglês *human epidermal growth factor receptor-2* (HER2) são realizadas por meio de ensaios moleculares com métodos baseados em imunohistoquímica (IHQ)⁵.

É importante ressaltar que a simples avaliação adicional, por IHQ, dos receptores hormonais de PR, ER, combinado com o HER2 e o KI67 são marcadores proteicos chaves e podem inferir e categorizar o câncer de mama (Tabela 1), indiretamente, nos seguintes “subtipos moleculares like” clinicamente relevantes: (i) “enriquecido com HER2” (ER-/PR-/HER2+), (ii) “cânceres de mama triplo negativo” (ER-/PR-/HER2-, TNBC), (iii) luminal A (ER+/PR+/HER2- e fenótipo proliferativo baixo(<14%) e (iv) luminal B (ER+/PR- ou baixo/ HER- e fenótipo proliferativo KI67 alto (>14%) ou ER+/PR- ou baixo/ HER2 + com fenótipo proliferativo KI67 alto (>14%).

Tabela 1 - Subtipos fenotípicos mediante a classificação por imuno-histoquímica.

SUBTIPOS	ER	PR	HER2	ki67
HER2	negativo	negativo	positivo	variável
LUMINAL A	positivo	positivo	negativo	<14%
LUMINAL B	positivo	positivo	positivo ou negativo	>14%
TNBC	negativo	negativo	negativo	variável

Fonte: Fonte de autoria própria

Sabemos também que, dentre esses subtipos, o HER2 positivo é um dos mais agressivos, uma vez que o gene HER2, que se encontra no cromossomo 17, expressa uma proteína de transmembrana com o mesmo nome; ela pode ser encontrada em quatro isomorfos que se ligam entre elas e entre si mesmas, desencadeando um sinal intracelular que leva a proliferação celular, angiogênese e crescimento tumoral ⁶⁻⁸.

Além disso, vários trabalhos têm explorado assinaturas baseadas na expressão gênica para a estratificação do câncer de mama a fim de orientar a seleção de terapias direcionadas com precisão⁹. Por exemplo, Prat et al 2012, relataram um ensaio de expressão gênica clinicamente aplicável usando 50 genes (PAM50) que classifica o câncer de mama em quatro subtipos intrínsecos moleculares comumente conhecidos como luminal A, luminal B, enriquecido com HER2 (HER2) e basal-like (também chamado de TNBC) ¹⁰.

O (Ki-67) e, posteriormente, assinaturas genéticas surgiram como ferramentas úteis para aprimorar a estimativa de risco e tomar decisões sobre a quimioterapia (neo)adjuvante ¹⁰. A expressão do ligante 1 de morte programada celular, do inglês *programmed death-ligand 1* (PD-L1) permitiu a identificação de pacientes com câncer de mama triplo-negativo (TNBC) que provavelmente se beneficiariam com a adição de imunoterapia à quimioterapia. As mutações somáticas da proteína fosfatidilinositol-4,5-bisfosfato 3-quinase, do inglês *The phosphatidylinositol-4,5-bisphosphate 3-kinase* (PIK3CA) e as mutações germinativas dos genes câncer de mama do tipo 1, do inglês *breast cancer type 1 susceptibility protein* (BRCA) foram reconhecidas como preditivas para o benefício dos inibidores de PI3K e poli (ADP-ribose) polimerase, do inglês *Poly (ADP-ribose) polymerase* (PARP), respectivamente ¹⁰.

O último biomarcador a surgir como potencialmente preditivo para a atividade de drogas anticancerígenas no câncer de mama, na verdade, nada mais é que a expressão baixa da proteína HER2, ficando conhecida e classificada na literatura como “HER2-low”- definida como tendo expressão por IHQ de 1+ ou 2+ sem amplificação do gene HER2 detectado com um ensaio adicional de hibridização *in situ* do inglês, *in situ hybridization* (ISH) ¹¹.

Normalmente, a superexpressão da proteína HER2 na superfície das células é avaliada por imuno-histoquímica (IHQ), que fornece um sistema de pontuação bem conhecido de quatro níveis de 0 a 3+ (0: negativo; 1+: negativo; 2+: equivocal e 3+: positivo)¹². O subtipo “HER2-low” inclui aqueles que têm pontuações IHQ de 1+ ou 2+ sem amplificação por um teste adicional baseado em hibridização *in situ* (ISH) ¹³. Embora a avaliação 2+ por ISH seja mais precisa, é mais cara e demora mais para fornecer resultados, o que ainda representa um obstáculo na prática clínica atual na maioria dos hospitais ⁷. Portanto, isso apoia esforços para descobrir novas abordagens para quantificar o HER2-baixo e poderia, assim, expandir o potencial terapêutico de agentes de conjugados de anticorpos-medicamentos (ADC) no BC.

Atualmente, desde 2020, temos a informação que aproximadamente 60% dos cânceres metastáticos de mama classicamente “negativos” (classificação binária) para o HER2 expressam baixos níveis de HER2 ¹⁴. Esses tumores foram recém definidos em uma nova subclassificação conhecida com “HER2-low” e constituem uma população heterogênea, incluindo cânceres de mama positivos e negativos para receptores hormonais, que variam em

prognóstico e sensibilidade aos tratamentos sistêmicos¹⁵⁻¹⁷. Apesar dessa sub-classificação patológica, também já está claro que não se trata de um novo subtipo biológico da doença, e, portanto, a sua identificação do subtipo “HER2-low” é preditiva para o uso das novas terapias anti-HER2, mas não prognóstica por si só^{11,18,19}.

As terapias direcionadas ao HER2 até pouco tempo disponíveis não melhoraram os resultados clínicos em pacientes com HER2-*low*²⁰; portanto, o câncer de mama HER2-*low* era tratado até 2022 como HER2-negativo (HER2-baixo e HER2-zero [pontuação de IHQ 0]), com pacientes estratificados de acordo com o status de receptor hormonal^{15,17,21}. Em geral, esses pacientes tinham opções limitadas de tratamento direcionado após a progressão durante a terapia primária e, mais comumente, recebiam quimioterapia paliativa com um único agente²⁰.

Dito isto, entendemos a importância que a identificação e quantificação da proteína HER2 para o diagnóstico, prognóstico e predição da sobrevida global com o uso de inibidores de HER2 (Trastuzumabe, Pertuzumabe, Trastuzumabe Deruxtecan etc.) sendo, portanto, fundamental o seu correto diagnóstico da lâmina de parafina marcada com a técnica IHQ^{8,22}. Atualmente os padrões atuais mais aceitos para determinar amplificação do gene HER2 ocorrem por meio do ensaio de hibridização fluorescente *in situ* do inglês, *fluorescence in situ hybridization* (FISH) enquanto a superexpressão proteica é padronizada pela IHQ, respectivamente; ambos definidos pela Sociedade Americana de Oncologia Clínica, do inglês *American Society of Clinical Oncology* (ASCO) e do Colégio Americanos de Patologistas, do inglês *College of American Pathologists* (CAP), embora a discordância entre ISH e IHQ não seja tão incomum e pode levar a um excesso de diagnóstico de HER2 +²³.

Neste contexto supracitado, alguns desafios se impõem no Brasil e no mundo: (i) a carência de patologistas, geneticistas e médicos oncologistas em geral -atualmente mais concentrados na região sudeste e metropolitanas do Brasil. (ii) a heterogeneidade e discrepância dos laudos de anatomia patológica entre os diversos laboratórios do país²⁴⁻²⁶. Não menos importante, segundo dados americanos e nacionais, os erros e discordâncias entre segundas opiniões em oncologia podem chegar a quase 12% dos casos, implicando em aumento de custos, falha de

classificações e, principalmente, prejuízos sérios aos pacientes portadores de alterações como o HER2 por exemplo ^{25,26}.

Corroborando esses dados, recentemente com a nova definição de HER2-low, também ficou evidente que existe uma perda de reprodutibilidade e correta identificação da proteína HER2-low na prática clínica ^{27,28}. Rimm D et al demonstraram uma discordância de 41% na distinção entre HER2 0 e HER2 1+ ou superior, mas uma discordância de 11% entre HER2 3+ e não 3+; a porcentagem de acertos foi de apenas cerca de 60% das vezes ^{29,30}. Outro estudo feito por Viale et al, em 781 amostras, a concordância percentual geral entre os escores de HER2 reavaliados e históricos foi de 81%; a concordância foi numericamente maior para HER2-low do que para HER2 IHQ ³¹. Essa dificuldade também foi reportada por Rüschoff et al: entre 80 patologistas, a concordância geral de pontuação com uma nova categoria de HER2-low estava acima do benchmark de acordo geral de 80% para os ensaios Ventana PATHWAY 4B5 e Dako HercepTest HER2, e foi maior do que o relatado anteriormente por Fernandez et al ³².

A parametrização de práticas diagnósticas é um problema global, mais acentuado no nosso país; portanto, visando a melhora dos erros em saúde, reduções de custos, ampliação do acesso e otimização de resultados, cada vez mais e inteligência artificial (IA) vem sendo usada no setor de saúde através do desenvolvimento de algoritmos matemáticos que visam mimetizar funções cognitivas humanas e aprimorá-las ^{33,34}.

Por definição, IA é um amplo campo de pesquisa que visa projetar computadores e sistemas que simulam a inteligência humana. Dentro deste novo campo da ciência, temos o aprendizado de máquinas que é um subcampo da IA -desenvolve algoritmos que permitem ao computador adaptar-se a um novo problema sem ser reprogramado; ou seja, a máquina “aprende” a resolver um problema diretamente aplicando-se métodos estatísticos para reconhecer padrões de um conjunto de dados fornecidos sem instrução humana ^{33,34} previamente argumentado, a aplicação de IA pode utilizar uma rede neural convolucional, que é uma classe de rede neural artificial que vem sendo aplicada com sucesso no processamento e análise de imagens digitais ³³⁻³⁵.

Recentemente, Mustafa I. Jaber et al. demonstraram a possibilidade de reconhecer no microambiente tumoral padrões de imagens estruturadas usando características do tumor, estroma e padrão de infiltração de linfócitos na lâmina de parafina digitalizada, sendo marcadores prognósticos adicionais aos fatores anatomopatológicos clássicos ³⁶. Não menos importante, outras validações já reportadas na literatura com implicações diagnósticas, prognósticas e preditivas podem ser citadas: detecção do tumor primário; detecção de metástases em linfonodos; predição do grau tumoral; identificação do subtipo histológico; predição de Ki67, receptores de estrogênio (RE), progesterona (RP) e HER2 etc ^{35,37-42}. Embora a avaliação dos biomarcadores no câncer de mama seja baseada na ampliação da imunohistoquímica e no perfil de expressão gênica, os avanços na aprendizagem profunda do inglês, *deep learning* (DP) tornaram-se uma linha de frente para a tomada de decisões médicas ⁴³.

Embora a maioria das avaliações de biomarcadores no BC seja baseada em IHQ expandida e perfil de expressão gênica, os avanços em DL se tornaram um ponto focal para a tomada de decisão médica e têm o potencial de complementar os ensaios tradicionais ^{44,45}. Com base na capacidade de aprender características subjacentes diretamente de imagens histopatológicas de lâminas inteiras coradas com hematoxilina e eosina (H&E), nós e outros desenvolvemos sistemas baseados em DL para prever diversos fenótipos moleculares, incluindo mutações em câncer de pulmão não pequenas células ⁴⁵, instabilidade de microssatélites (MSI) em pacientes com câncer gastrointestinal ⁴⁶, deficiência de recombinação homóloga (HRD) ⁴⁷ e outros biomarcadores clínicos ⁴⁸. Esses estudos enfatizam ainda mais a capacidade dos métodos baseados em ML de aprender características morfológicas subjacentes que refletem a expressão de marcadores moleculares.

Embora a relevância da aprendizagem de máquina para o status do HER2 tenha sido investigada em outros estudos de BC ^{39,49} a maioria dos modelos foi treinada em imagens de microarranjo de tecido (TMA), em lâminas coradas e imagens de hibridização *in situ* por fluorescência (FISH). Além disso, os classificadores foram usados para prever o HER2 entre o status de receptor hormonal ou subtipo intrínseco, no qual o status do HER2 foi anotado apenas como positivo ou negativo. Mais importante ainda, no contexto das opções de tratamento disponíveis para pacientes com HER2-baixo, a avaliação do HER2-baixo por modelos de DL ainda precisa ser explorada.

Neste estudo, construímos uma série de modelos de aprendizado profundo para examinar diferentes conjuntos de dados de treinamento, combinando pontuações quantitativas do HER2-baixo, usando ensaios IHQ e FISH no câncer de mama. Como parte dos resultados, exploramos mecanismos de IA explicáveis (xAI) para obter *insights* sobre as causas subjacentes das decisões do modelo. Por fim, montamos um conjunto de dados online inédito, gerando mais de 10.018.656 fragmentos de imagens H&E.

Visando a produção de conhecimento, validação de dados em população brasileira e o desenvolvimento de tecnologia nacional, esse estudo de coorte retrospectivo aqui proposto tentará implementar e validar o uso da inteligência artificial na detecção de tumores HER2 positivos e HER2-low, previamente detectadas pelas técnicas IHQ ou FISH, usando apenas lâminas digitalizadas e coradas com H&E, em pacientes com tumores de mama do AC Camargo Cancer Center. Este estudo poderá servir como base e referência para um melhor refinamento dos critérios inteligência artificial aplicados a patologia diagnóstica descentralizando o acesso e reduzindo custos operacionais.

2. Objetivos

2.1 Objetivo geral

Avaliar o uso da inteligência artificial em lâminas de H&E, no reconhecimento de tumores de mama HER2 positivos e HER2-low, previamente detectados por imunohistoquímica ou FISH a partir de uma casuística disponível no AC Camargo Câncer Center.

2.2 Objetivos específicos

- 1) Realizar um levantamento dos casos com *status* HER2 detectado a partir da análise dos resultados gerados pelos ensaios de IHQ e hibridização;
- 2) Recuperar as lâminas histológicas do acervo físico institucional;
- 3) Avaliar as lâminas com qualidade aceitável de acordo com coloração e qualidade do material fixado;
- 4) Digitalizar as lâminas histológicas;
- 5) Desenvolver modelo de treino, teste e inferência com modelos de aprendizagem profunda;
- 6) Avaliar o poder discriminatório da predição de HER2+ e HER2-low por meio da área sob a curva do inglês, Area Under Curve (AUC) na avaliação lâminas coradas apenas com H&E.

2.3. Hipótese

Uso de inteligência artificial na análise de imagens digitalizadas a partir das lâminas H&E de tumores mamários como método de diagnóstico alternativo do HER2+ e HER2-low.

3. Métodos

3.1 Desenho do estudo

O conjunto de dados anonimizado provém de uma coorte de pacientes com câncer de mama do A. C. Camargo Cancer Center (ACCC, N = 504). Todos os pacientes foram tratados para doença de câncer de mama no ACCC entre os anos de 2019 e 2021. Como parte de seu diagnóstico, o status HER2 foi determinado seguindo as diretrizes da ASCO atualizadas em 2018 ⁵⁰, com avaliação visual do ensaio de imuno-histoquímica (IHQ) e, em casos com pontuação IHQ HER2 2+, foi realizado um teste de hibridização *in situ* por fluorescência (FISH) ou um teste de hibridização *in situ* por fluorescência de dupla coloração de duplo hapten em campo claro (DDISH). Todos os casos com metástases ou tratamento neoadjuvante foram excluídos e as características histológicas da coorte são apresentadas na Tabela 3. Ao total, foram digitalizadas pelos colaboradores do Laboratório de Bioinformática do AC Camargo Cancer Center, sendo no total 546 imagens de alta resolução coradas com H&E (magnificação de 40x) de biópsias e ressecções com o scanner Leica Aperio AT2. O estudo no ACCC recebeu aprovação ética do comitê de ética da Fundação Antônio Prudente.

3.2 Remoção do fundo e detecção do tumor

Começamos dividindo as lâminas em fragmentos de tamanho 256px × 256px a uma magnificação de 0,5 µm/pixel, com sobreposição de 50% em relação aos fragmentos adjacentes. Em seguida, treinamos uma rede ResNet-50 para detectar background, tumor e não tumor nas lâminas. Para fazer isso, utilizamos os conjuntos de dados denominado como NCT-CRC-HE-100K ⁵¹ e slides de tumor anotados manualmente dos conjuntos de dados TCGA ⁵². As anotações manuais de tumor no conjunto de dados TCGA contêm anotações aproximadas das regiões tumorais da maioria das lâminas desse conjunto. Neste estudo, usamos o subconjunto TCGA-BRCA para aprender um consenso aproximado de uma região tumoral para o câncer de mama.

Alternativamente, também treinamos modelos com todos os fragmentos não identificados como fundo e fragmentos identificados como tumor invasivo a partir de um conjunto de dados interno de lâminas anotadas de tumor invasivo de câncer de mama, a fim de avaliar como a seleção de região afeta o desempenho do modelo que considera um Aprendizado de Múltiplas Instâncias (MIL). Todos os fragmentos foram avaliados nessas redes, e os fragmentos identificados como tumor foram usados para treinar o modelo MIL. Apenas as lâminas com ≥ 1000 fragmentos de tumor detectados foram usados nas etapas seguintes.

3.3 Extração de características

Utilizamos o modelo pré-treinado RetCCL⁵³ para extrair um vetor de características das imagens de fragmentos. Este modelo foi treinado usando um grande conjunto de dados com mais de 22.000 imagens histológica de vários tipos de tumores, o que gera um vetor de incorporação de 2048 características que descreve as características morfológicas presentes nas imagens de fragmentos. Este vetor de 2048 características foi usado nas etapas seguintes para treinar o modelo MIL. Este método supera abordagens semelhantes que utilizam modelos pré-treinados no conjunto de dados ImageNet em tarefas de classificação de histopatologia, porque o RetCCL aprendeu características específicas da histologia⁵³.

3.4 Definição do modelo

Utilizamos um classificador de Aprendizado de Múltiplas Instâncias (MIL) de supervisão fraca, que recebe uma lâmina H&E no formato de um conjunto de fragmentos como entrada de dados e a classifica em uma das classes desejadas usando um mecanismo baseado em atenção. A arquitetura MIL foi baseada em um método de aprendizagem profundo conhecido como CLAM⁵⁴.

Resumidamente, o modelo CLAM recebe características da lâmina como um conjunto de fragmentos e a classifica em uma das classes de destino. Internamente, ele funciona modelando cada grupo HER2 de forma independente, com uma camada de atenção e uma camada de classificação. Enquanto a camada de atenção é responsável por identificar fragmentos importantes e combiná-los em um único vetor de 2048 características para a

lâmina histológica com base na importância percebida de seus fragmentos, a camada de classificação recebe este vetor de representação e produz um único valor de ativação: saídas mais altas representam uma maior probabilidade de a lâmina pertencer à classe. Finalmente, esse valor independente para cada grupo HER2 é combinado no vetor de saída usando uma ativação *softmax*, cujo valor máximo descreve a classe da lâmina. Além disso, o modelo também utiliza um classificador de instância cujo objetivo é agrupar os fragmentos de cada classe. Durante o treinamento, o classificador de instância age como um método de regularização que direciona o mecanismo de atenção para produzir características relevantes tanto para o mecanismo de atenção quanto para si próprio; no entanto, uma vez que o modelo conclui o treinamento, ele pode ser usado como um classificador de fragmentos que identifica fragmentos associados à sua classe.

3.5 Modelo de treinamento

Realizamos um procedimento de validação cruzada de 10 *folds*, onde as lâminas dos pacientes foram atribuídas aleatoriamente a um conjunto de treinamento (85%) ou conjunto de validação (15%), garantindo que a distribuição das etiquetas permanecesse inalterada em ambos os conjuntos. Todos os modelos foram treinados por um total de 200 épocas, com interrupção antecipada configurada para o escore F1 da validação. Para acelerar o treinamento e reduzir os requisitos de memória da GPU, dividimos o procedimento de treinamento em duas etapas: i) extração de características e ii) treinamento do classificador CLAM, que foi feito separadamente. Isso tem a vantagem de permitir um maior número de modelos a serem treinados em paralelo devido a requisitos mais baixos de memória e processamento. Todos os modelos foram treinados usando uma GPU NVIDIA A100 40GB.

3.6 Análises dos fragmentos (tiles) acionáveis

Os fragmentos (**tiles**) acionáveis podem ser identificados pela camada de instância do CLAM, que é um classificador binário disponível para cada grupo HER2 e retorna se o fragmento é reconhecido como pertencente a essa classe. Para cada lâmina no conjunto de validação, classificamos seus fragmentos com a camada de instância e amostramos os fragmentos

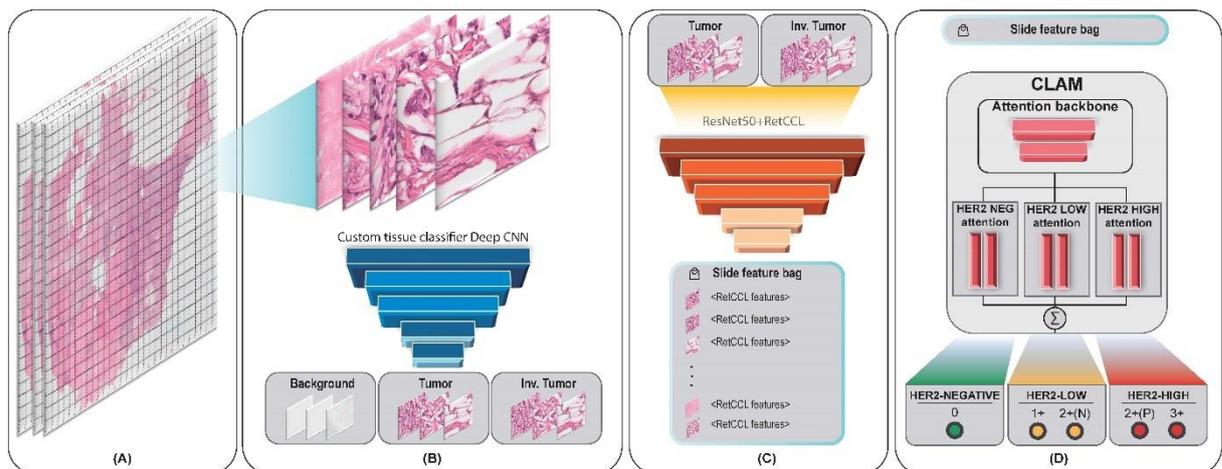
acionáveis para cada grupo HER2, independentemente da classificação da lâmina. Portanto, nosso objetivo é entender o que foi aprendido como relevante para cada grupo.

Para analisar os fragmentos acionáveis, propomos uma abordagem de visualização com base em um histograma 2D do espaço. Primeiro, projetamos os dois componentes principais das características dos fragmentos usando PCA. Em seguida, discretizamos o espaço criando um histograma 2D com 30x30 bins, onde cada bin atua como um centroide em torno das incrustações dos fragmentos amostrados, agrupando os fragmentos com características semelhantes. Essa abordagem nos permite caracterizar as diferentes características dos fragmentos, colorindo os bins usando diferentes critérios. Por exemplo, os bins podem ser coloridos com base no grupo HER2 que aparece com mais frequência; ou podemos colorir-los apenas quando dois ou mais modelos usam a característica. Essa abordagem nos permite organizar e compreender o espaço de características histológicas acionáveis.

4 RESULTADOS

4.1 Uma estrutura de aprendizado profundo de supervisão fraca para a avaliação do HER2.

O objetivo deste estudo foi desenvolver um modelo baseado em aprendizado profundo para a avaliação do HER2-baixo. Assim, propusemos uma arquitetura envolvendo 4 etapas (Figura 1). Primeiro, fragmentos foram extraídos das imagens inteiras (WSI, *whole slide imaging*) e subsequentemente classificados em grupos contendo tumor, não tumor e background (Figura 1A-B). Isso gera um conjunto de fragmentos, dos quais extraímos características (Figura 1C) para cada conjunto, que são usadas para treinar um modelo de Aprendizado de Múltiplas Instâncias (MIL) para classificar as WSI em classes HER2-negativo, HER2-baixo e HER2-alto (Figura 1D).



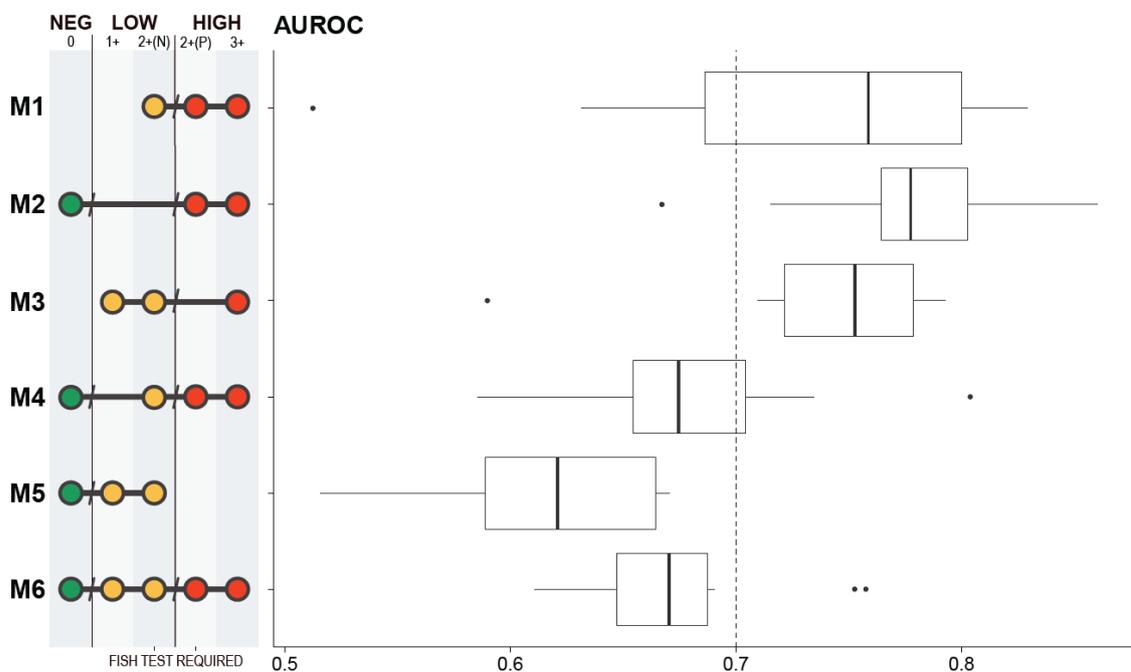
Fonte: Fonte de autoria própria

Figure 1 - Fluxo de trabalho para o método proposto. (A) As imagens do slide inteiro são divididas em fragmentos; (B) As regiões de interesse são identificadas; (C) Os recursos são extraídos de imagens de blocos; (D) Um modelo de aprendizagem de múltiplas instâncias é treinado para classificar conjuntos de blocos HER2 negativo, HER2 baixo (HER2-low) e HER2 alto.

4.2 Desempenho dos modelos revela que a classificação do HER2-baixo é desafiadora

Construímos seis modelos distintos (Figura 2) com o objetivo de explorar a capacidade dos classificadores de distinguir entre as classes HER2-negativo, HER2-baixo e HER2-alto em diferentes cenários. Enquanto o modelo M6 foi treinado com o conjunto de dados completo para classificar as 3 classes, também treinamos modelos em subconjuntos, como o M1, no qual apenas casos HER2 2+ e 3+ foram incluídos, com o objetivo de prever os resultados do teste baseado em ISH.

Observamos que os modelos treinados para separar o HER2-negativo do HER2-baixo tiveram um desempenho inferior em comparação com os modelos que previam apenas entre HER2-negativo e HER2-alto ou HER2-baixo e HER2-alto.



Fonte: Fonte de autoria própria

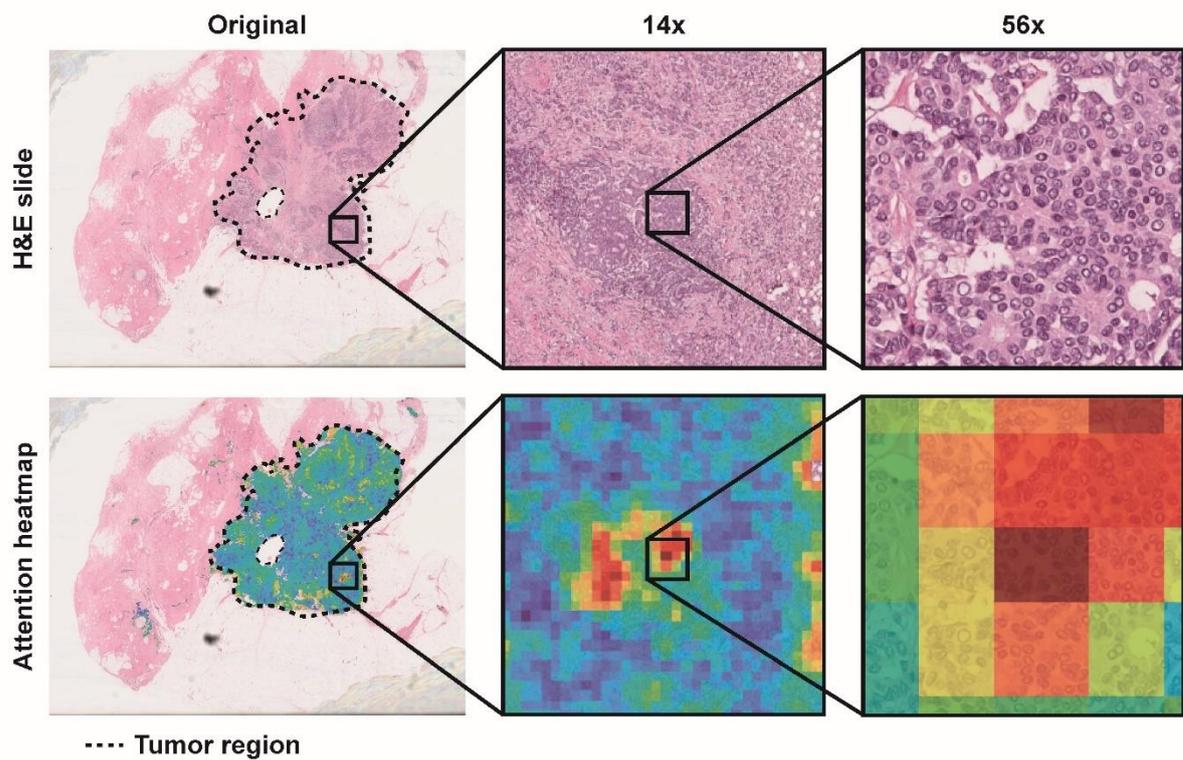
Figure 2 - UROC dos resultados da validação cruzada de 10 vezes. Cada modelo foi treinado com seu próprio conjunto de pontuações HER2 que compõe cada grupo HER2 alvo. As pontuações HER2 utilizadas são indicadas pelos pontos ativos, enquanto a cor do ponto indica seu grupo HER2 alvo para o modelo.

4.3 A validação cruzada destaca a estabilidade do modelo

Um passo importante para utilizar os resultados do aprendizado profundo em um ambiente clínico é interpretar como o modelo funciona, o que nos permite avaliar sua validade e confiabilidade. Portanto, realizamos um estudo abrangente de interpretação com foco em entender quais características morfológicas são frequentemente usadas pelo mecanismo de atenção do modelo para classificar as lâminas. Especificamente, nosso objetivo era medir a estabilidade e qualificar se as diferentes configurações do modelo dentro de cada *fold* foram capazes de identificar e aprender a prever o status do HER2 usando as mesmas características histológicas. Além disso, as características aprendidas pelo modelo dessa maneira semi-supervisionada podem fornecer *insights* sobre características histológicas clinicamente relevantes. Para avaliar a variabilidade de aprendizado desses modelos, realizamos uma validação cruzada de 10 *folds* em cada um dos subconjuntos testados. A AUC de teste para cada *fold* pode ser vista nos boxplots, conforme mostrado na Figura 2.

4.4 Fragmentos acionáveis revelam padrões usados pelo classificador.

Para entender quais padrões histológicos foram aprendidos pelo modelo para classificar as classes do HER2, precisamos examinar quais características estão sendo usadas pelo classificador. Portanto, aproveitamos as camadas de atenção fornecidas pelo aprendizado de múltiplas instâncias, que direcionam o classificador para um subconjunto de fragmentos que possuem características distintas que podem ser usadas para a classificação (Figura 3). Concentramos nossas análises nos fragmentos com alta atenção, denominados daqui em diante como "fragmentos acionáveis", devido à sua importância no resultado do classificador.

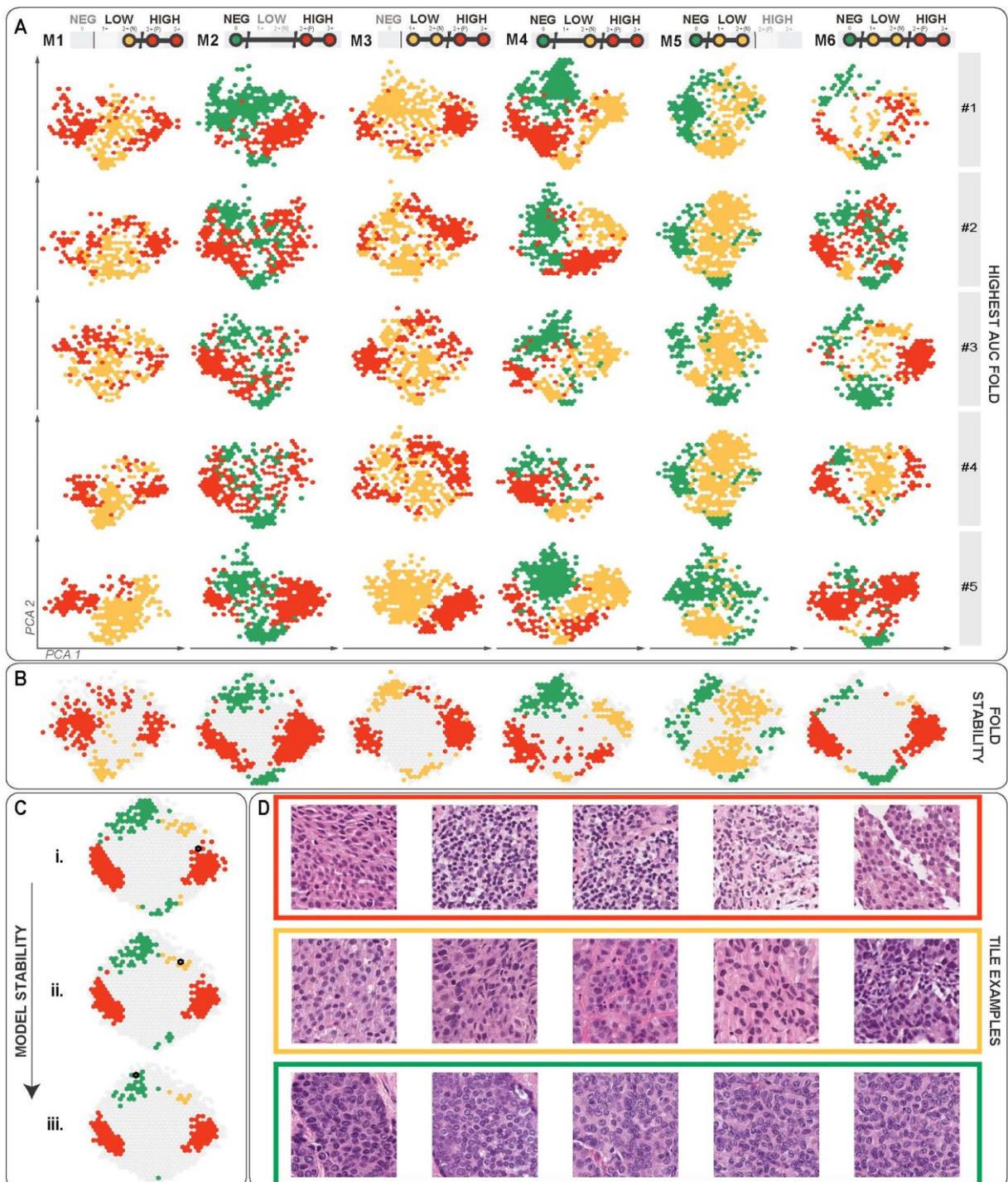


Fonte: Fonte de autoria própria

Figure 3 - Lâmina corada com H&E de uma amostra de tecido e seu mapa de calor de atenção correspondente gerado pelo modelo proposto. O mapa de calor mostra as áreas onde o modelo concentra mais atenção, com cores mais quentes indicando maior atenção pontuações. Os pontos marcados mostram a região selecionada pelo filtro preliminar, descrito na Figura 1 (B).

Para verificar se os modelos treinados com diferentes *folds* conseguiram aprender a usar características semelhantes, desenvolvemos um método para destacar quais fragmentos estão sendo usados pelos modelos nas diferentes configurações testadas (Figura 4). Primeiro, os fragmentos acionáveis foram identificados, amostrando fragmentos das lâminas do conjunto de validação e sondando as camadas de atenção do modelo. Em seguida, os fragmentos foram agrupados no espaço latente de características, e a cor indica a qual grupo HER2 alvo esses fragmentos correspondem na camada de classificação (Figura 4A). Além disso, agregamos as dobras do mesmo modelo em uma única visualização (Figura 4B), o que revelou padrões em diferentes modelos. Uma votação majoritária entre os modelos nos permitiu visualizar os padrões compartilhados (Figura 4C). Finalmente, amostramos fragmentos cujas

características estão dentro dessas regiões de padrão compartilhado para ilustrar alguns exemplos desses fragmentos acionáveis de suas respectivas classes (Figura 4D). Em conjunto, observamos que existem padrões recorrentes em dobras do mesmo modelo, o que indica que fragmentos com características semelhantes nessas áreas estão sendo usados para a classificação da mesma classe.



Fonte: Fonte de autoria própria

Figure 4 - Análise de explicabilidade do modelo. (A) Plotamos o espaço de incorporação dos fragmentos/tiles usando PCA para cada modelo e *folds*, colorindo cada região de acordo com o grupo HER2 que presta atenção a essa incorporação. Regiões vazias indicam incorporações que são compartilhados por várias classes para classificação. (B) Caixas coloridas indicam recursos usados por 40% dos *folds* dentro de cada modelo. (C) Analisamos a consistência do modelo contando quantos modelos usam a mesma incorporação para classificar cada classe, colorindo as regiões de acordo. Mostramos três níveis de concordância: i. 2 ou mais; ii. 3 ou mais; iii. 4 ou mais. (D) Exemplos de peças do espaço de incorporação para cada classe, ilustrando como elas se relacionam com as características de interesse.

Embora a avaliação histopatológica de seções de tecido coradas com hematoxilina e eosina (H&E) seja indispensável para o manejo do câncer de mama, a relação entre o padrão celular e os subgrupos do HER2 ainda é desconhecida. Portanto, examinamos o padrão celular dos fragmentos acionáveis associados aos grupos HER2, conforme resumido na Tabela 2. Especificamente, observamos que o HER2-baixo está associado a menos necrose, menor atividade mitótica e ausência de densidade de linfócitos infiltrantes de tumor (TILs) em comparação com outros grupos, o que está de acordo com observações anteriores⁵⁵⁻⁵⁷. Concluímos que os classificadores aprenderam características morfológicas na imagem associadas especificamente ao subgrupo HER2-baixo.

Tabela 2 - Achados histológicos associados a fragmentos acionáveis nos grupos HER2.

Histological features	HER2-neg	HER2-low	HER2-high
Nuclear grade	3	1 and 2	1,2 and 3
Mitotic Activity	High	Low	Moderate
Necrosis	High	Negative	High
TILs	12/473 tiles	0/100 tiles	41/650 tiles
Adipose Tissue	High	Low	Low
Desmoplasia	Low	Low	High
Morphological pattern	Solid	Solid and Cribriform	Discohesive

Fonte: Fonte de autoria própria

4.5 Filtrando as regiões de interesse

Para determinar se subconjuntos de fragmentos afetam o modelo MIL, treinamos modelos com diferentes conjuntos de dados, conforme descrito na Seção 3.2, e avaliamos a relevância de cada subconjunto. Testamos diferentes conjuntos de fragmentos de treinamento. Primeiro, tentamos filtrar apenas fragmentos sem background (NOBG) e descobrimos que o modelo é capaz de distinguir padrões de tecido únicos. No entanto, também observamos que o modelo aprende a atribuir altos valores de atenção a fragmentos do microambiente tumoral, o que nos levou a testar o uso apenas de fragmentos da região tumoral. Nos referimos a essa configuração de filtragem como apenas Tumor (TUMO) e descobrimos que, ao usar o microambiente tumoral, conseguimos alcançar o melhor equilíbrio entre o desempenho do modelo e a velocidade de treinamento. Finalmente, também testamos o uso apenas de fragmentos de Tumor Invasivo (TUMI) e descobrimos que isso afeta negativamente o desempenho do modelo. De fato, nossa seção de explicabilidade mostra que o modelo aprende a distinguir a gravidade do tumor usando o contexto tumoral, como regiões de necrose e infiltrado inflamatório, que foram filtradas nesta configuração.

Tabela 3 - Mediana de AUROC em diferentes cenários de treinamento. Concluímos que a filtragem de fragmentos como etapa de pré-processamento é importante para o desempenho preditivo do modelo. Vale ressaltar que decidimos usar TUMO, pois proporcionou o melhor equilíbrio entre o desempenho preditivo do modelo e os requisitos de hardware. Considerando como o método MIL funciona, acreditamos que a etapa de filtragem de fragmentos força o modelo a aprender características distintivas dos fragmentos com informações reais, em vez de se concentrar em fragmentos irrelevantes.

Model	NOBG	TUMO
M1	0.76 ± 0.08	0.76 ± 0.10
M2	0.77 ± 0.06	0.78 ± 0.06
M3	0.68 ± 0.06	0.75 ± 0.06
M4	0.67 ± 0.05	0.67 ± 0.06
M5	0.64 ± 0.04	0.62 ± 0.05
M6	0.64 ± 0.03	0.67 ± 0.05

Fonte: Fonte de autoria própria

Tabela 4 - Características histológicas estratificadas por cada grupo HER2

	NEG	LOW	HIGH
Nuclear grade			
1	5	7	1
1 and 2	2	4	0
2	79	97	11
2 and 3	2	4	3
3	113	109	121
Inflammatory Infiltrate			
Not found	0	2	0
Low	152	158	55
Moderate	26	29	28
High	13	2	8
Mitotic Score			
Score 1	104	116	42
Score 2	44	44	25
Score 3	39	27	20
TILs			
Greater than 10	27	22	22
Less than 10	93	76	29
Immunophenotype			
Not determined	1	4	1
TNBC	43	7	0
Luminal	30	48	1
LUMA	36	20	1
LUMB	82	110	19
Luminal B-HER2	0	0	27
HER2 overexpression	0	0	46
SBR grade			
Grade I	26	31	1
Grade II	90	102	44
Grade III	70	53	38
Tissue source			
Biopsy	106	147	95
Resection	99	86	53
Micro-calcifications			
Absent	154	153	92
Present	46	72	48
Necrosis			
Absent	153	167	83
Present	52	66	65
Histological group			
Undefined	24	27	10
In situ (DCIS)	11	36	53
No special type (NST)	134	134	71
Special type (ST)	19	18	7

Fonte: Fonte de autoria própria

Entendemos que os resultados apresentados são promissores e reforçam a importância da utilização da Inteligência Artificial como uma ferramenta custo-efetiva para ampliar a avaliação de biomarcadores com relevância preditiva e prognóstica no caso do HER2 e apenas preditiva em relação ao Her2-low; podendo, inclusive, reduzir custos em saúde e facilitar o acesso a esse teste, possibilitando um diagnóstico preciso com implicações terapêuticas que impactam positivamente na sobrevida. Não obstante, como todo desenvolvimento de um modelo de aprendizado de máquina, o volume e a qualidade dos dados são fundamentais. Portanto, iremos ainda ampliar a casuística dos casos disponíveis no A.C. Camargo conforme disponibilidade dos novos casos a partir de 2024.

Adicionalmente, buscaremos incluir na etapa de treinamento e validação externa um conjunto de imagens histológicas geradas pelo consórcio TCGA. Essa etapa, encontra-se motivada pelo fato de que o modelo pode aprender outras representações dos dados capazes de melhorar a predição da atual rede. De importância, a validação externa deve ser afetada devido potenciais efeitos, incluindo o tipo de protocolo usado na coloração além de outros ruídos no contexto da digitalização das lâminas. Finalmente, técnicas de aumento dos dados, bem como transformações de cores poderão mitigar esses efeitos.

5 DISCUSSÃO

O câncer de mama HER2-positivo constitui uma forma de câncer de mama caracterizada por níveis mais elevados do receptor do fator de crescimento epidérmico 2 (HER2), e onde medicamentos anti-HER2 têm mostrado eficácia em comparação com o câncer de mama HER2-negativo. Mais recentemente, a aprovação da terapia direcionada em tumores HER2-low amplia os tratamentos para o câncer de mama para inúmeros pacientes. Neste estudo, exploramos a capacidade de detectar o status HER2 diretamente a partir de imagens de histopatologia por meio de aprendizado profundo. Construímos seis modelos com o objetivo de explorar a capacidade dos classificadores de distinguir entre as classes HER2-negativo, HER2-low e HER2-positivo a partir de conjuntos de dados distintos com status HER2 avaliado por meio de ensaios IHQ e ISH. Primeiro, avaliamos a detecção de HER2-low em resultados HER2 2+ com base nos resultados obtidos no ensaio baseado em IHQ (M1). Neste modelo, o conjunto de treinamento abrange resultados onde há inerentemente subjetividade e alta concordância interobservador, o que reflete a garantia da capacidade discriminatória deste classificador. Vale ressaltar que este modelo tem grande potencial para ser usado de maneira custo-efetiva, uma vez que os resultados HER2 2+ por imuno-histoquímica requerem um ensaio adicional baseado em métodos de hibridização.

A seguir, procuramos avaliar a capacidade de discriminar grupos de pacientes com HER2- (negativo, pontuação 0) vs HER2+ (positivo, pontuação 2+ com teste positivo adicional realizado por FISH e pontuação 3+). Na verdade, ao considerar que as mudanças moleculares se refletem na histologia celular, o modelo M2 detecta grupos de pacientes sem atividade de expressão proteica do gene HER2 daqueles com alta expressão. Portanto, é razoável esperar um ótimo desempenho deste modelo. Em conjunto, quando consideramos os modelos (M1 e M3) destinados a detectar HER2-low em pacientes HER2 2+, observamos um desempenho um pouco inferior ao modelo M2, mas ainda alto e aceitável. Por outro lado, nos modelos destinados a detectar HER2-low, incluindo pacientes HER2 1+, observamos uma diminuição no desempenho dos classificadores (M4-M6). Coletivamente, esses resultados nos levam a concluir que existe um efeito confundidor nas classes treinadas ao considerar o grupo HER2 1+. Isso é resultado da subjetividade e da pequena diferença nos resultados (baixa concordância

interobservador) obtidos nos ensaios de imuno-histoquímica dos grupos HER2 0 e 1+ ²⁹ especialistas, o que conseqüentemente afeta o desempenho dos modelos, uma vez que os resultados dos testes moleculares são considerados na divisão dos grupos de treinamento e teste dos modelos.

Em resumo, enquanto estudos anteriores examinaram a capacidade de modelos baseados em aprendizado de máquina para prever o status HER2 apenas como positivo ou negativo, nosso trabalho tenta expandir esses subtipos incluindo o HER2-low como uma nova categoria. Além disso, nossos resultados são particularmente relevantes por duas razões principais. É o primeiro estudo a detectar o biomarcador HER2-low a partir de imagens histológicas considerando uma série retrospectiva de 504 pacientes e lança luz sobre as limitações de detectar HER2-low apenas a partir dos resultados de IHQ ²⁹. Nosso estudo revela padrões celulares locais aprendidos em fenótipos morfológicos relacionados ao subgrupo HER2-low, o que destaca a região de uma imagem importante para a previsão final.

Por fim, nossas descobertas reforçam ainda mais a importância de revisitar as diretrizes para a avaliação do HER2, mas também sugerem a inclusão de mais de um patologista na avaliação do HER2, além do uso de métricas de concordância ^{58,59} e a adoção das melhores práticas ao utilizar controles. Embora ideal, esse cenário, no entanto, está longe da realidade de muitos laboratórios devido à disponibilidade de patologistas, mas também à crescente carga de trabalho desses profissionais. Portanto, abordagens baseadas em IA, como a proposta no presente trabalho, tornam-se cada vez mais necessárias e promissoras para apoiar a tomada de decisões na avaliação de biomarcadores.

Embora nossos resultados sejam promissores, nosso modelo pode falhar na generalização devido a diferenças nos sistemas de imagem, protocolos de coloração e estamos prejudicados pelos resultados de ensaios baseados em IHQ que são subjetivos ²⁹.

Contudo, a despeito dos desafios e considerações aqui apresentados, com os nossos atuais resultados até o presente momento, apesar de uma amostragem ainda pequena, já temos um Produto Minimamente Viável, do inglês MVP- *Minimum Viable Product*. O nosso algoritmo de inteligência artificial conseguiu uma AUC compatível com a literatura para tumores HER2 e,

ao nosso conhecimento, é um dos primeiros a estabelecer a AUC para HER2-low diretamente de lâminas coradas com H&E, fato de extrema relevância dada a dificuldade de padronização e variação entre patologistas, redução de custos e recente aprovação de drogas como o Transtuzumabe Deruxtecan em câncer de mama.

Oportunidades:

De acordo com as premissas a seguir: (i) a porcentagem exata de tumores HER2++ que necessitam realizar o exame de FISH pode variar, mas em geral, é recomendado que a análise de FISH seja realizada em torno de 15% a 20% dos casos HER2++ para confirmar a amplificação do gene HER2. (ii) Os exames de FISH estão concentrados em poucos laboratórios do Brasil, necessitando de transporte de materiais biológicos em grandes distâncias. (iii) Para o ano de 2023 foram estimados 73.610 casos novos, o que representa uma taxa ajustada de incidência de 41,89 casos por 100.000 mulheres (25). (iv) Cerca de 40% dos casos são diagnosticados em fase avançada III e IV (25). (v) O preço médio unitário em reais do FISH é de R\$ 1.390,00. (vi) O Sistema Único de Saúde (SUS) paga R\$ 131,52 para cada um dos marcadores de IHQ (RE, RP, HER2, KI67); operadoras de saúde, em média, pagam R\$ 700,00 pelo mesmo exame. (vii) Na distribuição do sistema de saúde brasileira, em média, 70% das pessoas dependem exclusivamente do SUS e 30% do sistema de Saúde Complementar.

Se considerarmos os custos atualmente envolvidos com os exames de FISH e IHQ, por exemplo, apenas para avaliação do HER2, podemos estimar algo em torno de R\$ 20.463.580,00/ano e R\$ 35.788.593,00/ano respectivamente, demonstrando o tamanho do mercado.

Por fim, iremos procurar parcerias com grupos colaborativos de pesquisa como, por exemplo, o LACOG- Latin American Cooperative Oncology Group-, o grupo Oncoclínicas, o grupo Dasa e/ou grandes laboratórios nacionais que tenham interesse no contínuo desenvolvimento da ferramenta para validação externa do modelo de aprendizado, pivotando o resultado obtivo para futura comercialização final do produto.

6. Considerações éticas

O estudo foi aprovado do Comitê de Ética Médica do hospital AC Camargo Cancer Center, sendo aprovado.

7. REFERÊNCIAS

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021 May;71(3):209–49.
2. De Oliveira Santos M. Estimativa/2020 – Incidência de Câncer no Brasil. *Revista Brasileira de Cancerologia*. 2020;66(1).
3. van Seijen M, Lips EH, Thompson AM, Nik-Zainal S, Futreal A, Hwang ES, et al. Ductal carcinoma in situ: to treat or not to treat, that is the question. Vol. 121, *British Journal of Cancer*. Nature Publishing Group; 2019. p. 285–92.
4. Rakha EA, El-Sayed ME, Lee AHS, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of nottingham histologic grade in invasive breast carcinoma. *Journal of Clinical Oncology*. 2008;26(19):3153–8.
5. Jorns JM. Challenges in routine estrogen receptor, progesterone receptor, and HER2/neu evaluation. Vol. 143, *Archives of Pathology and Laboratory Medicine*. College of American Pathologists; 2019. p. 1444–9.
6. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human Breast Cancer: Correlation of Relapse and Survival with Amplification of the HER-2/neu Oncogene. *Science* (1979) [Internet]. 1987;235(4785):182–91. Available from: <https://www.science.org/doi/abs/10.1126/science.3798106>
7. Lamond DJES, Ones BLRE j, Hak SST, Uchs HFA, Aton VPI, Harm PD, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* [Internet]. 2001;344(11). Available from: www.nejm.org
8. Loibl S, Gianni L. HER2-positive breast cancer. *The Lancet* [Internet]. 2017 Nov;389(10087):2415–29. Available from: <http://www.thelancet.com/article/S0140673616324175/fulltext>
9. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N Engl J Med* [Internet]. 2009 Feb 19 [cited 2022 Feb 13];360(8):790–800. Available from: <https://pubmed.ncbi.nlm.nih.gov/19228622/>
10. Prat A, Parker JS, Fan C, Perou CM. PAM50 assay and the three-gene model for identifying the major and clinically relevant molecular subtypes of breast cancer. *Breast Cancer Res Treat*. 2012 Aug;135(1):301–6.

11. Tarantino P, Jin Q, Tayob N, Jeselsohn RM, Schnitt SJ, Vinciguilla J, et al. Prognostic and Biologic Significance of ERBB2-Low Expression in Early-Stage Breast Cancer. *JAMA Oncol*. 2022 Aug 1;8(8):1177–83.
12. Ahn S, Woo JW, Lee K, Park SY. HER2 status in breast cancer: changes in guidelines and complicating factors for interpretation. *J Pathol Transl Med* [Internet]. 2020 Jan 1 [cited 2023 Oct 6];54(1):34–44. Available from: <https://pubmed.ncbi.nlm.nih.gov/31693827/>
13. Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, et al. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. *New England Journal of Medicine*. 2022 Jul 7;387(1):9–20.
14. Schettini F, Chic N, Brasó-Maristany F, Paré L, Pascual T, Conte B, et al. Clinical, pathological, and PAM50 gene expression features of HER2-low breast cancer. *NPJ Breast Cancer*. 2021 Dec 1;7(1).
15. Tarantino P, Hamilton E, Tolaney SM, Cortes J, Morganti S, Ferraro E, et al. HER2-Low breast cancer: Pathological and clinical landscape. Vol. 38, *Journal of Clinical Oncology*. 2020.
16. Schettini F, Chic N, Brasó-Maristany F, Paré L, Pascual T, Conte B, et al. Clinical, pathological, and PAM50 gene expression features of HER2-low breast cancer. *NPJ Breast Cancer*. 2021;7(1).
17. Tarantino P, Curigliano G, Tolaney SM. Navigating the HER2-Low Paradigm in Breast Oncology: New Standards, Future Horizons. *Cancer Discov*. 2022;12(9).
18. Tarantino P, Hamilton E, Tolaney SM, Cortes J, Morganti S, Ferraro E, et al. HER2-Low Breast Cancer: Pathological and Clinical Landscape [Internet]. 2020. Available from: <https://doi.org/10.>
19. Tarantino P, Curigliano G, Tolaney SM. Navigating the HER2-Low Paradigm in Breast Oncology: New Standards, Future Horizons. *Cancer Discov*. 2022 Sep 1;12(9):2026–30.
20. Fehrenbacher L, Cecchini RS, Charles ;, Geyer E, Rastogi P, Costantino JP, et al. NSABP B-47/NRG Oncology Phase III Randomized Trial Comparing Adjuvant Chemotherapy With or Without Trastuzumab in High-Risk Invasive Breast Cancer Negative for HER2 by FISH and With IHC 1+ or 2+. *J Clin Oncol* [Internet]. 2019;38:444–53. Available from: <https://doi.>
21. Fehrenbacher L, Cecchini RS, Geyer CE, Rastogi P, Costantino JP, Atkins JN, et al. NSABP B-47/NRG oncology phase III randomized trial comparing adjuvant

- chemotherapy with or without trastuzumab in high-risk invasive breast cancer negative for HER2 by FISH and with IHC 1+ or 2+. *Journal of Clinical Oncology*. 2020;38(5).
22. Santana MF, Luiz ;, Ferreira CL. Diagnostic errors in surgical pathology Erros diagnósticos em patologia cirúrgica. *J Bras Patol Med Lab*. 2017;(2):124–9.
 23. Woo JW, Lee K, Chung YR, Jang MH, Ahn S, Park SY. The updated 2018 American Society of Clinical Oncology/College of American Pathologists guideline on human epidermal growth factor receptor 2 interpretation in breast cancer: comparison with previous guidelines and clinical significance of the proposed in situ hybridization groups. *Hum Pathol*. 2020 Apr 1;98:10–21.
 24. Woo JW, Lee K, Chung YR, Jang MH, Ahn S, Park SY. The updated 2018 American Society of Clinical Oncology/College of American Pathologists guideline on human epidermal growth factor receptor 2 interpretation in breast cancer: comparison with previous guidelines and clinical significance of the proposed in situ hybridization groups. *Hum Pathol*. 2020 Nov;98:10–21.
 25. Raab SS, Grzybicki DM, Janosky JE, Zarbo RJ, Meier FA, Jensen C, et al. Clinical impact and frequency of anatomic pathology errors in cancer diagnoses. *Cancer*. 2005 Nov;104(10):2205–13.
 26. Thomson TA, Hayes MM, Spinelli JJ, Hilland E, Sawrenko C, Phillips D, et al. HER-2/neu in breast cancer: Interobserver variability and performance of immunohistochemistry with 4 antibodies compared with fluorescent in situ hybridization. *Modern Pathology*. 2001;14(11).
 27. Fernandez AI, Liu M, Bellizzi A, Brock J, Fadare O, Hanley K, et al. Examination of Low ERBB2 Protein Expression in Breast Cancer Tissue. *JAMA Oncol*. 2022;8(4).
 28. Baez-Navarro X, van Bockstal MR, Nawawi D, Broeckx G, Colpaert C, Doebar SC, et al. Interobserver Variation in the Assessment of Immunohistochemistry Expression Levels in HER2-Negative Breast Cancer: Can We Improve the Identification of Low Levels of HER2 Expression by Adjusting the Criteria? An International Interobserver Study. *Mod Pathol*. 2023;36(1).
 29. Robbins CJ, Fernandez AI, Han G, Wong S, Harigopal M, Podoll M, et al. Multi-institutional Assessment of Pathologist Scoring HER2 Immunohistochemistry. *Mod Pathol* [Internet]. 2023 Jan 1 [cited 2023 Oct 6];36(1):100032. Available from: <https://pubmed.ncbi.nlm.nih.gov/36788069/>
 30. Han G, Schell MJ, Reisenbichler ES, Guo B, Rimm DL. Determination of the number of observers needed to evaluate a subjective test and its application in two PD-L1 studies.

- Stat Med [Internet]. 2022 Apr 15 [cited 2023 Oct 3];41(8):1361–75. Available from: <https://pubmed.ncbi.nlm.nih.gov/34897773/>
31. Viale G, Basik M, Niikura N. Retrospective study to estimate the prevalence and describe the clinicopathological characteristics, treatment patterns, and outcomes of HER2-low breast cancer. In: San Antonio Breast Cancer Symposium Annual Meeting. 2022.
 32. Rüschoff J, Penner A, Ellis IO, Hammond EH, Lebeau A, Osamura RY, et al. Proficiency assessment of HER2-low breast cancer scoring with the Ventana PATHWAY 4B5 and Dako HercepTest HER2 assays and the impact of pathologist training. In: San Antonio Breast Cancer Symposium Annual Meeting. 2022.
 33. Haykin S. Neural Networks and Learning Machines Third Edition. Vol. 936 p. New Jersey: Pearson Prentice Hall; 2008.
 34. Iqbal MJ, Javed Z, Sadia H, Qureshi IA, Irshad A, Ahmed R, et al. Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. *Cancer Cell Int* [Internet]. 2021;21:270. Available from: <https://doi.org/10.1186/s12935-021-01981-1>
 35. Farahmand S, Fernandez AI, Ahmed FS, Rimm DL, Chuang JH, Reisenbichler E, et al. Deep learning trained on hematoxylin and eosin tumor region of Interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. *Mod Pathol* [Internet]. 2022 Nov;35(1):44–51. Available from: <https://pubmed.ncbi.nlm.nih.gov/34493825/>
 36. Jaber MI, Beziaeva L, Szeto CW, Benz SC. Deep learning-based risk stratification for HER2-negative breast cancer patients. [cited 2022 Feb 13]; Available from: <https://doi.org/10.1101/2021.05.26.445720>
 37. Osareh A, Shadgar B. Microarray data analysis for cancer classification. 2010 5th International Symposium on Health Informatics and Bioinformatics, HIBIT 2010. 2010;125–32.
 38. Jaber MI, Beziaeva L, Szeto CW, Benz SC. Deep learning-based risk stratification for HER2-negative breast cancer patients. Available from: <https://doi.org/10.1101/2021.05.26.445720>
 39. Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ Breast Cancer* [Internet]. 2018 Nov;4(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/30182055/>

40. Hossain MdS, Hanna MG, Uraoka N, Nakamura T, Edelweiss M, Brogi E, et al. Automatic quantification of HER2 gene amplification in invasive breast cancer from chromogenic in situ hybridization whole slide images. *Journal of Medical Imaging* [Internet]. 2019 Nov;6(4):1. Available from: [/pmc/articles/PMC6868351/](#)
41. Naik N, Madani A, Esteva A, Keskar NS, Press MF, Ruderman D, et al. Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains. *Nat Commun*. 2020;11(1).
42. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA - Journal of the American Medical Association*. 2017;318(22).
43. Cui M, Zhang DY. Artificial intelligence and computational pathology. *Laboratory Investigation*. 2021;101(4):412–22.
44. Cui M, Zhang DY. Artificial intelligence and computational pathology. Vol. 101, *Laboratory Investigation*. Springer Nature; 2021. p. 412–22.
45. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* [Internet]. 2018 Oct 1 [cited 2023 Oct 6];24(10):1559–67. Available from: <https://pubmed.ncbi.nlm.nih.gov/30224757/>
46. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* [Internet]. 2019 Jul 1 [cited 2023 Oct 6];25(7):1054–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/31160815/>
47. Valieris R, Amaro L, de Toledo Osório CAB, Bueno AP, Mitrowsky RAR, Carraro DM, et al. Deep Learning Predicts Underlying Features on Pathology Images with Therapeutic Relevance for Breast and Gastric Cancer. *Cancers (Basel)* [Internet]. 2020 Dec 1 [cited 2023 Oct 6];12(12):1–12. Available from: <https://pubmed.ncbi.nlm.nih.gov/33316873/>
48. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer* [Internet]. 2021 Feb 16 [cited 2023 Oct 6];124(4):686–96. Available from: <https://pubmed.ncbi.nlm.nih.gov/33204028/>
49. Shamaï G, Binenbaum Y, Slossberg R, Duek I, Gil Z, Kimmel R. Artificial Intelligence Algorithms to Assess Hormonal Status From Tissue Microarrays in Patients With Breast

- Cancer. *JAMA Netw Open* [Internet]. 2019 Nov;2(7):e197700–e197700. Available from: <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2739045>
50. Wolff AC, Elizabeth Hale Hammond M, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J Clin Oncol* [Internet]. 2018 Jul 10 [cited 2023 Oct 6];36(20):2105–22. Available from: <https://pubmed.ncbi.nlm.nih.gov/29846122/>
 51. Kather JN, Halama N, Marx A. 100,000 histological images of human colorectal cancer and healthy tissue. 2018 Apr 7 [cited 2023 Oct 6]; Available from: <https://zenodo.org/record/1214456>
 52. Loeffler C, Kather JN. Manual tumor annotations in TCGA. 2021 Aug 29 [cited 2023 Oct 6]; Available from: <https://zenodo.org/record/5320076>
 53. Wang X, Du Y, Yang S, Zhang J, Wang M, Zhang J, et al. RetCCL: Clustering-guided contrastive learning for whole-slide image retrieval. *Med Image Anal* [Internet]. 2023 Jan 1 [cited 2023 Oct 6];83. Available from: <https://pubmed.ncbi.nlm.nih.gov/36270093/>
 54. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* [Internet]. 2021 Jun 1 [cited 2023 Oct 6];5(6):555–70. Available from: <https://pubmed.ncbi.nlm.nih.gov/33649564/>
 55. Li Y, Tsang JY, Tam F, Loong T, Tse GM. Comprehensive characterization of HER2-low breast cancers: implications in prognosis and treatment. *EBioMedicine* [Internet]. 2023 Nov;91. Available from: <https://pubmed.ncbi.nlm.nih.gov/37068349/>
 56. Almstedt K, Heimes AS, Kappenberg F, Battista MJ, Lehr HA, Krajnak S, et al. Long-term prognostic significance of HER2-low and HER2-zero in node-negative breast cancer. *Eur J Cancer* [Internet]. 2022 Nov;173:10–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/35839597/>
 57. van den Ende NS, Smid M, Timmermans A, van Brakel JB, Hansum T, Foekens R, et al. HER2-low breast cancer shows a lower immune response compared to HER2-negative cases. *Sci Rep* [Internet]. 2022 Nov;12(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/35902644/>
 58. Han G, Schell MJ, Reisenbichler ES, Guo B, Rimm DL. Determination of the number of observers needed to evaluate a subjective test and its application in two PD-L1 studies.

Stat Med [Internet]. 2022 Nov;41(8):1361–75. Available from: <https://pubmed.ncbi.nlm.nih.gov/34897773/>

59. Robbins CJ, Fernandez AI, Han G, Wong S, Harigopal M, Podoll M, et al. Multi-institutional Assessment of Pathologist Scoring HER2 Immunohistochemistry. Mod Pathol [Internet]. 2023 Nov;36(1):100032. Available from: <https://pubmed.ncbi.nlm.nih.gov/36788069/>