

**APLICAÇÃO DE APRENDIZAGEM
PROFUNDA NA PREDIÇÃO DE
CARACTERÍSTICAS OCULTAS EM
IMAGENS HISTOPATOLÓGICAS
COM RELEVÂNCIA TERAPÊUTICA
PARA CÂNCER DE MAMA**

**RENAN VALIERIS BUENO DE
ALMEIDA**

Tese apresentada à Fundação Antô-
nio Prudente para a obtenção do
título de Doutor em Ciências.

Área de concentração: Oncologia

Orientador: Dr. Israel Tojal da Silva

São Paulo

2024

Trata-se da versão corrigida da tese.

FICHA CATALOGRÁFICA

Almeida, Renan Valieris Bueno de.

Aplicação de aprendizagem profunda na predição de características ocultas em imagens histopatológicas com relevância terapêutica para câncer de mama. / Renan Valieris Bueno de Almeida. São Paulo, 2025.

68f.

Tese de Doutorado - Fundação Antônio Prudente. Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.

Orientador: Israel Tojal da Silva.

1. Histopatologia, 2. Deep Learning, 3. Câncer de mama

CDU 616

*Todos os direitos reservados à FAP. A violação dos direitos autorais constitui crime, previsto no art. 184 do Código Penal, sem prejuízo de indenizações cabíveis, nos termos da Lei nº 9.610/08

Renan Valieris Bueno de Almeida

Aplicação de aprendizagem profunda na predição de características ocultas em imagens histopatológicas com relevância terapêutica para câncer de mama

Aprovado em: 12/03/2025

Banca Examinadora

Orientador: Dr. Israel Tojal da Silva

Instituição: Fundação Antônio Prudente

Membro da banca: Dr. Zhao Liang

Instituição: Universidade de São Paulo

Membro da banca: Dra. Cynthia Aparecida Bueno de Toledo Osorio

Instituição: Fundação Antônio Prudente

Membro da banca: Dr. Vladimir Claudio Cordeiro de Lima

Instituição: Fundação Antônio Prudente

Membro da banca: Dr. Washington Santos

Instituição: Fundação Oswaldo Cruz - Bahia

*It is not knowledge, but the act of learning,
not possession but the act of getting there,
which grants the greatest enjoyment.*

Carl Friedrich Gauss. Letter to Bolyai. 1808.

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.*

AGRADECIMENTOS

Agradeço à minha família, principalmente à minha mãe, que sempre me incentivou a estudar mais e a ir além.

Agradeço ao meu orientador, Dr. Israel Tojal Da Silva, pelas ideias, discussões, incentivo e orientação ao longo de todo o caminho.

Agradeço ao Dr. Alexandre Defelicibus, Luan Martins e Lucas Souza, da nossa equipe de bioinformática, por toda a ajuda no processamento das lâminas, na execução das análises e pelas ideias.

Agradeço ao Dr. Rafael Rosales, do Departamento de Matemática da USP de Ribeirão Preto, que me ajudou no desenvolvimento, interpretação e escrita dos métodos matemáticos e estatísticos.

Agradeço também à Dra Adriana Passos Bueno e Dra Cynthia Aparecida Bueno Toledo Osório, do departamento de Patologia do A. C. Camargo Cancer Center, por toda a ajuda na interpretação dos exames e lâminas histopatológicas e na identificação de padrões morfológicos.

Agradeço também a Nathalia Varela Araujo e Keila Taira Macedo, do Laboratório de Histologia, por organizar e digitalizar todas as lâminas necessárias para a realização deste trabalho.

Agradeço à Dra Dirce Maria Carraro e ao Dr. Rafael Canfield Brianese, do grupo de Genômica e Biologia Molecular, pela realização e interpretação dos testes genéticos utilizados neste trabalho.

Agradeço também ao hospital A. C. Camargo, por proporcionar a infraestrutura e seu contínuo apoio à pesquisa que possibilitaram a realização deste trabalho.

Agradeço a todos os pacientes que se dispuseram a participar de estudos clínicos e possibilitaram a realização deste trabalho e de tantos outros.

RESUMO

Almeida R. **Aplicação de aprendizagem profunda na predição de características ocultas em imagens histopatológicas com relevância terapêutica para câncer de mama.** [Tese]. São Paulo; Fundação Antônio Prudente; 2024.

O tratamento do câncer está avançando em direção a uma abordagem cada vez mais personalizada e baseada em biomarcadores . No entanto, o uso de biomarcadores precisos, rápidos e de baixo custo ainda não é uma realidade na maioria dos hospitais. Nos últimos anos, abordagens baseadas em Inteligência Artificial para predição de biomarcadores a partir de imagens histopatológicas emergiram como uma alternativa promissora aos métodos tradicionais. Neste trabalho, exploramos a viabilidade de usar técnicas de redes neurais convolucionais e aprendizado com múltiplas instâncias na predição dos biomarcadores **HER2** e **HRD** diretamente de imagens histológicas coradas com hematoxilina e eosina (H&E). Os resultados demonstraram o potencial desses métodos para melhorar a detecção de padrões morfológicos associados a características moleculares, reforçando a viabilidade de integrar essas abordagens à rotina clínica, com implicações significativas para o diagnóstico, prognóstico e planejamento terapêutico de pacientes com câncer de mama.

ABSTRACT

Almeida R. [**Application of deep learning to predict hidden characteristics in pathological images with therapeutic relevance for breast cancer**]. [tese]. São Paulo; Fundação Antonio Prudente; 2024.

Cancer treatment is becoming increasingly personalized as more studies deepen our knowledge of distinct subtypes and their treatments. In order for these subtypes to be identified in clinical practice, accurate, fast and low-cost biomarkers are needed so that their treatments can be widely implemented. In recent years, deep learning methods for predicting these biomarkers from pathological images have emerged as a promising alternative to traditional methods. However, there are still significant challenges in implementing these methods. In this study we explore the application of these methods for predicting the breast biomarkers **HER2** and **HRD**, analyzing their advantages and shortcomings and demonstrating that their application in the clinic can improve the efficiency of routine examinations, allow the testing of new biomarkers and assist pathologists in identifying new patterns present in images.

LISTA DE FIGURAS

Figura 1 – Diretrizes para interpretação dos níveis de HER2	5
Figura 2 – Processos mutacionais no câncer	9
Figura 3 – Diferentes tipos de tecidos patológicos	12
Figura 4 – Segmentação computadorizada utilizando redes neurais	13
Figura 5 – Aplicações de redes neurais na prática clínica	15
Figura 6 – Exemplos de colorações H&E	26
Figura 7 – Esquematização do treinamento do modelo para predição do HRD binário.	30
Figura 8 – Distribuição de lesões nos genes conhecidos da via de recom- binação homóloga (HR) nas as coortes de treinamento (A) e independentes (B). Os números acima de cada barra indicam a quantidade de pacientes correspondentes.	34
Figura 9 – Esquematização do treinamento dos modelos para predição de HER2	35
Figura 10 – Exemplo de inferência do modelo de detecção de Tumor e fundo	36
Figura 11 – AUC da validação cruzada para os modelos de <i>HER2</i>	37
Figura 12 – AUC para os modelos de HER2 no TCGA-BRCA	39
Figura 13 – Mapa de calor da camada de atenção	40
Figura 14 – Comparação entre marcadores genéticos associados ao HRD e o escore agregado.	45
Figura 15 – Esquematização do treinamento dos modelos para predição do escore agregado de HRD	46
Figura 16 – Métricas de desempenho dos modelos do HRD agregado	48
Figura 17 – Predição do escore HRD na coorte AC-BRCA-HRD-2	50
Figura 18 – Exemplos de imagens com alta atenção ao panHRD	51
Figura 19 – Predição do escore HRD na coorte CPTAC	53

Figura 20 – Rede neural com uma camada interna.	76
Figura 21 – Cálculo de uma camada de convolução em 2D	80
Figura 22 – Arquitetura da rede neural convolucional LeNet	80
Figura 23 – Arquitetura de uma rede neural recorrente	81

LISTA DE TABELAS

Tabela 1	–	Esquema de particionamento de dados para os modelos HER2	28
Tabela 2	–	Métricas de desempenho dos modelos HER2	39
Tabela 3	–	Média e desvio padrão das AUCs comparando NOBG e TUMO	41
Tabela 4	–	Média e desvio padrão das AUCs comparando métodos diferentes	42

LISTA DE SIGLAS E ABREVIATURAS

AI	Inteligência artificial (do inglês artificial intelligence)
AUC	Área sob a curva ROC (do inglês Area Under Curve)
CNN	Rede Neural Convolucional (do inglês Convolutional Neural Network)
DSB	Quebra de fita dupla (do inglês double strand break)
ER	Receptor de Estrogênio (do inglês Estrogen Receptor)
FISH	Hibridização fluorescente in situ (do inglês Fluorescence in situ hybridization)
H&E	Hematoxilina e Eosina
HER2	Receptor Tipo 2 do Fator de Crescimento Epidérmico Humano (do inglês Human Epidermal growth factor Receptor-type 2)
HR	Recombinação Homóloga (do inglês Homologous Recombination)
HRD	Deficiência em Recombinação Homóloga (do inglês Homologous Recombination Deficiency)
IHC	Imuno-histoquímica (do inglês Immunohistochemistry)
ISH	Hibridação ‘in situ’ (do inglês In situ hybridization)
MIL	Aprendizado com Múltiplas Instâncias (do inglês Multiple Instance Learning)
MSI	Instabilidade de Microssatélite (do inglês Microsatellite instability)

NGS	Sequenciamento de nova geração (do inglês Next generation Sequencing)
PR	Receptor de Progesterona (do inglês Progesterone Receptor)
RNN	Rede Neural Recorrente (do inglês Recurrent Neural Network)
SBS3	Assinatura Mutacional 3 (do inglês Single Base Substitution 3)
TMB	Carga mutacional tumoral (do inglês Tumor Mutation Burden)
WSI	Imagem digital de Lâmina Patológica (do inglês Whole Slide Image)

SUMÁRIO

1	INTRODUÇÃO	1
1.1	O Câncer	1
1.2	Biomarcadores de Câncer	2
1.3	O Câncer de mama	4
1.3.1	HER2	4
1.3.2	Recombinação Homóloga	6
1.4	Assinaturas Mutacionais	8
1.4.1	Cicatrizes genômicas para HRD	10
1.5	Histopatologia	10
1.6	Inteligência Artificial	16
1.6.1	Oportunidades e desafios na análise de imagens H&E	17
2	OBJETIVOS	20
2.1	Objetivo Geral	20
2.2	Objetivos específicos	20
3	METODOLOGIA	21
3.1	Coortes Investigadas	21
3.1.1	TCGA	21
3.1.2	CPTAC	21
3.1.3	HEROHE	22
3.1.4	AC-BRCA	22
3.2	Técnicas de Aprendizado Profundo	23
3.3	Pré-processamento de Imagens Histopatológicas	23
3.3.1	Separação da Lâmina em Fragmentos	23
3.3.2	Remoção de Fundo	24

3.3.3	Identificação de Tumor	25
3.3.4	Normalização de Cor	25
3.3.5	Extração de Características	26
3.4	Anotação do nível de HER2	27
3.5	Estimação de Assinaturas Mutacionais para HRD	27
3.6	Cálculo do Escore Agregado para HRD	29
4	RESULTADOS E DISCUSSÃO	30
4.1	Classificação binária do HRD	30
4.1.1	Treinamento do Modelo	31
4.1.2	Performance do Modelo	31
4.2	Classificação dos subtipos de HER2	34
4.2.1	Filtragem dos fragmentos de interesse orienta o treinamento	35
4.2.2	Treinamento do Modelo	36
4.2.3	Predições do Modelo	37
4.2.4	Análise de Regiões Acionáveis	39
4.2.5	Comparação de implementações alternativas de modelo	41
4.3	Predição do escore agregado de HRD	42
4.3.1	Preditor Agregado de HRD	43
4.3.2	Validação do Escore Agregado	44
4.3.3	Treinamento do Modelo	45
4.3.4	Predição do escore de HRD em diversos tumores	47
4.3.5	Análise de coorte com triplo-negativo	49
4.3.6	Análise de Regiões Acionáveis	51
4.3.7	Validação da predição do escore agregado no CPTAC	52
4.4	Discussão	54
4.5	Trabalhos Futuros	55
4.5.1	Predição do HER2 usando o IHQ	55
4.5.2	Validação das Regiões Acionáveis	55

4.5.3	Predição da amplificação do HER2	56
4.5.4	Validar o HRD em outros tumores	56
4.6	Outros Trabalhos Realizados	56
5	CONCLUSÕES	58
6	REFERÊNCIAS	59

ANEXOS 69

Anexo A – PREMIAÇÕES	70
---------------------------------------	-----------

APÊNDICES 73

Apêndice A – TÉCNICAS DE APRENDIZADO PROFUNDO 74

A.1	Definição de Uma Rede Neural	74
A.2	Treinamento de uma Rede Neural	76
A.3	Redes Neurais Convolucionais	79
A.4	Redes Neurais Recorrentes	81
A.5	Extração de Características	82
A.6	Aprendizado com Múltiplas Instâncias	83
A.6.1	Top-K	84
A.6.2	ADMIL	85
A.6.3	CLAM	85
A.6.4	TransMIL	86
A.6.5	Regressão MIL	86
A.7	Métricas de desempenho	87
A.7.1	F1-Score	87
A.7.2	AUC	88

A.7.3	Teste Estatístico para comparação de métricas	88
A.8	Validação Cruzada	89
A.9	Redução de Dimensionalidade	89

Apêndice B – CÁLCULO DO ESCORE AGREGADO PARA HRD	90
---	-----------

Apêndice C – ARTIGO HRD – VALIERIS 2020	93
--	-----------

Apêndice D – ARTIGO HER2 – VALIERIS 2024	106
---	------------

Apêndice E – ARTIGO HRD – VALIERIS 2025	118
--	------------

1 INTRODUÇÃO

1.1 O CÂNCER

O câncer é uma das principais questões de saúde pública no mundo, sendo responsável por altas taxas de mortalidade e representando um grande obstáculo para o aumento da expectativa de vida global. Somente em 2022, foram diagnosticados quase 20 milhões de novos casos e registradas 9,7 milhões de mortes no mundo (Cancer 2022). No Brasil, as estimativas para o período de 2023 a 2025 indicam a ocorrência de 704 mil novos casos de câncer¹, sendo 483 mil quando excluídos os casos de câncer de pele não melanoma. Este último é projetado como o mais incidente, com 220 mil novos casos (31,3%), seguido pelos cânceres de mama, com 74 mil (10,5%); próstata, com 72 mil (10,2%); cólon e reto, com 46 mil (6,5%); pulmão, com 32 mil (4,6%); e estômago, com 21 mil (3,1%).

O câncer é uma doença multifacetada caracterizada por alterações genéticas, epigenéticas e microambientais, que permitem a aquisição de capacidades celulares específicas. Esses processos incluem sinais proliferativos sustentados, evasão de supressores de crescimento, resistência à morte celular, replicação ilimitada, indução de angiogênese, ativação de invasão e metástase, reprogramação metabólica e evasão imunológica (Hanahan 2022). Essas capacidades são facilitadas por características como instabilidade genômica, inflamação tumoral, reprogramação epigenética não mutacional e microbiomas polimórficos. O microambiente tumoral, composto por células cancerígenas e estromais, desempenha um papel crucial na progressão e manutenção da malignidade, destacando a complexidade do processo de tumorigênese e a interação entre células cancerígenas e seu ambiente (Sepich-Poore et al. 2021).

Existem mais de 100 tipos de câncer, que geralmente são referidos pelo

¹ <https://www.inca.gov.br/>

local, órgão ou tecido onde o câncer se forma e, mesmo dentro de um órgão específico, ainda é possível subdividi-lo de acordo com o grau histológico, histórico hereditário, alterações somáticas, entre outros (Institute 2021).

1.2 BIOMARCADORES DE CÂNCER

Os processos atuais na tomada de decisões clínicas personalizadas em oncologia estão cada vez mais ancorados no uso de biomarcadores, que desempenham um papel essencial na estratificação e manejo de pacientes. Esses biomarcadores podem ser amplamente classificados em três categorias principais: prognósticos, preditivos e diagnósticos (Ballman 2015; Zou e Wang 2020). De forma geral, biomarcadores são características claramente definidas e mensuráveis, que servem como indicadores de processos biológicos normais, processos patogênicos ou respostas biológicas a uma exposição ou intervenção específica, incluindo intervenções terapêuticas.

A integração de biomarcadores na prática clínica tem transformado a abordagem ao tratamento oncológico, permitindo uma personalização mais eficaz (Vargas e Harris 2016). Esses indicadores fornecem informações valiosas que permitem não apenas o diagnóstico precoce e mais preciso de doenças, mas também a estratificação de pacientes com base no risco de desenvolvimento de enfermidades, resposta esperada a tratamentos específicos e desfechos clínicos previstos, como progressão ou recidiva da doença (Vargas e Harris 2016). Um biomarcador de prognóstico, em particular, é aquele que fornece informações sobre a probabilidade de ocorrência de eventos clínicos futuros em uma população definida. Esses eventos incluem, entre outros, mortalidade, recorrência ou progressão da doença. Assim, um biomarcador prognóstico é essencial para identificar pacientes que apresentam maior (ou menor) risco de tais desfechos, orientando intervenções precoces e estratégias terapêuticas direcionadas (Zou e Wang 2020). Por outro lado, um biomarcador preditivo indica o provável benefício do tratamento para

o paciente, em comparação com sua condição inicial (Zou e Wang 2020). Finalmente, os biomarcadores de diagnóstico são uma classe de biomarcadores usados para identificar se um paciente tem uma condição de doença ou identificar indivíduos com um subtipo da doença.

Nos últimos anos, o conhecimento sobre biomarcadores em oncologia aumentou tremendamente, proporcionando oportunidades para melhorar a eficiência da detecção e a eficácia do tratamento. No câncer de mama, por exemplo, a avaliação dos receptores de estrógeno (**ER**), progesterona (**PR**), receptor 2 do fator de crescimento epidérmico humano (**HER2**) e de proliferação celular (**Ki-67**) é rotineira, pois esses biomarcadores têm valor prognóstico e preditivo. Na prática clínica, a investigação desses biomarcadores é realizada por meio de ensaios moleculares, como métodos baseados em imuno-histoquímica (**IHC**), seguidos pela avaliação visual do patologista (Rakha et al. 2008; Duffy et al. 2017). Adicionalmente, a avaliação da deficiência de um mecanismo de reparo do **DNA**, especificamente a via de recombinação homóloga (**HR**, *Homologous Recombination*), tem sido incluída no repertório de biomarcadores testados na rotina diagnóstica do câncer de mama. Células com deficiência nessa via resulta (**HRD**, *Homologous Recombination Deficiency*) em uma capacidade reduzida de reparar danos no DNA. Especificamente, **HRD** é um biomarcador importante para terapias que utilizam agentes que danificam o DNA, como a quimioterapia à base de platina e os inibidores de PARP (Hoppe et al. 2018).

Embora a implantação sistemática do uso de biomarcadores na rotina clínica seja cada vez mais necessária e bem-vinda, ainda há desafios a serem superados. Dentre eles, podemos destacar: i) a capacidade de análise ampla dos dados ainda é um fator impeditivo diante da diversidade e quantidade de dados gerados; ii) a disponibilidade de coortes para validação retrospectiva e prospectiva ainda é limitada a poucos centros especializados, levando a uma priorização de biomarcadores em um contexto mais restrito diante dos custos e aplicações; iii) a reprodutibilidade técnica das plataformas e a variabilidade entre laboratórios,

uma vez que a grande maioria dos estudos de biomarcadores não são realizados em laboratórios certificados por órgãos regulamentadores; iv) limitações na qualidade e quantidade do material biológico; v) problemas de reprodutibilidade e poder estatístico devido a vieses na amostra original ou coorte de validação; vi) além do alto custo que pode impedir o acesso aos testes da maioria dos pacientes.

1.3 O CÂNCER DE MAMA

O câncer de mama é o tipo mais comum de câncer invasivo em mulheres no mundo, responsável por cerca de 30% dos casos da doença entre elas, com uma proporção de mortalidade/incidência de 15%. Cerca de 10% de todos os casos de câncer de mama estão relacionados à predisposição genética ou histórico familiar (Loibl et al. 2021).

O câncer de mama é uma doença heterogênea, dividida na clínica em três subtipos principais: Luminal, HER2-positivo e Triplo Negativo. O diagnóstico desses subtipos é feito por meio de biomarcadores de receptores hormonais e **HER2**. O diagnóstico da subtipagem é essencial para o tratamento efetivo da doença, porém ainda existem controvérsias quanto à definição desses subtipos e ao melhor método de tratamento em cada caso (Loibl et al. 2021).

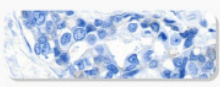
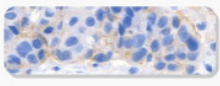
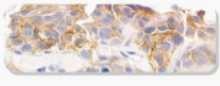
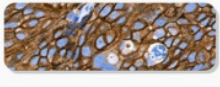
Cerca de 15-20% dos casos de câncer de mama triplo negativo estão associados a mutações germinativas em BRCA1 ou BRCA2. A avaliação de mutações germinativas em genes BRCA tem potencial para identificar pacientes que podem se beneficiar de terapias com inibidores de PARP (do inglês *poly(ADP-ribose) polymerase*) (Robson et al. 2017) ou terapias neoadjuvantes (Pohl-Rescigno et al. 2020).

1.3.1 HER2

O receptor tipo 2 do fator de crescimento epidérmico humano (**HER2**) é uma proteína de membrana codificada pelo gene *ERBB2*. A amplificação ou

superexpressão dessa proteína ocorre em aproximadamente 20-30% dos casos de câncer de mama, estando associada a tumores mais agressivos, maior taxa de recorrência e menor sobrevida (Mitri et al. 2012).

A avaliação do **HER2** é tipicamente feita por meio de **IHC**, em que um anticorpo se liga a proteínas **HER2** na superfície das células e emite uma coloração ou fluorescência, permitindo a avaliação de sua expressão por meio da análise visual de um patologista. Este processo é ilustrado na Figura 1. O patologista utiliza um sistema de pontuação para avaliar a intensidade da expressão observada: (0: negativo; 1+: negativo; 2+: equívoco e 3+: positivo) (Ahn et al. 2020). Se o patologista avaliar a pontuação 2+: equívoco, é realizado um teste adicional baseado em *hibridização in situ* (**ISH**) ou *Dual-color Dual In Situ-hybridization* (**DDISH**) para avaliar a amplificação do número de cópias do gene **HER2**.

Spectrum of HER2 positivity according to ASCO/CAP guidelines				Spectrum of HER2 positivity
	IHC score	HER2 test interpretation	HER2 status	
	0	No staining or incomplete and faint/barely perceptible membrane staining in ≤10% of tumor cells	Negative	
	1+	Incomplete and faint/barely perceptible membrane staining in >10% of tumor cells	Low	
	2+	Weak-moderate complete membrane staining in >10% of tumor cells OR intense membrane staining in ≤10% of tumor cells	ISH amplification?	
	3+	Complete and intense membrane staining in >10% of tumor cells	Positive	

Fonte: Ivanova et al. (2023).

Figura 1 – Diretrizes para interpretação dos resultados de IHC e ISH, para avaliar os níveis de HER2. A esquerda, exemplos de imagens de imuno-histoquímica que destacam as proteínas HER2 em marrom.

Em 2022, foi publicado um estudo que descreveu um ensaio clínico que comparou o desfecho de pacientes com câncer de mama metastático com baixa expressão de **HER2** (escore 1+ ou 2+ com **ISH** negativo) tratados com o medicamento *Trastuzumab deruxtecan* contra o tratamento típico de quimioterapia

(Modi et al. 2022). A droga se liga e bloqueia a via gênica do receptor **HER2**. Esse tratamento é efetivo em cânceres que dependem dessa via para seu crescimento, porém seu uso em tumores com baixa expressão de **HER2** não era recomendado. O estudo revelou que esse tratamento permitiu que pacientes tivessem sobrevida significativamente maior em comparação à quimioterapia, mesmo quando a expressão de **HER2** é baixa (Modi et al. 2022). Seguindo resultados positivos deste estudo, a agência federal do departamento de saúde dos Estados Unidos, *Food and Drug Administration*, aprovou o uso da droga na clínica para pacientes com câncer de mama metastático com **IHC** score 1+ ou 2+ e **ISH** negativo, e este subgrupo ficou conhecido como *HER2-low*.

Embora a avaliação **HER2** 2+ pelos ensaios do tipo **ISH** seja mais precisa, ela é mais cara e demorada para retornar resultados, portanto continua sendo um gargalo na prática clínica atual na maioria dos hospitais (Liwski et al. 2022). Portanto, isso endossa esforços para investigar novas abordagens para quantificar os níveis de **HER2**, o que também poderia expandir o potencial terapêutico dos agentes conjugados anticorpo-droga para pacientes com câncer de mama.

1.3.2 Recombinação Homóloga

A via de recombinação homóloga (**HR**) desempenha um papel fundamental na manutenção da estabilidade genômica, realizando a reparação eficiente de quebras de dupla fita (**DSB**, do inglês *Double-Strand Breaks*), protegendo as células de alterações mais danosas que impulsionam a tumorigênese (Moynahan e Jasin 2010). Quando um ou mais genes dessa via não desempenham seu papel corretamente, a capacidade da célula de reparar esses danos é afetada, resultando em acúmulo de mutações que podem desencadear a tumorigênese. Esse fenótipo é conhecido como Deficiência de Recombinação Homóloga (**HRD**, do inglês *Homologous Recombination Deficiency*). Por outro lado, células cancerígenas com essa deficiência são alvos eficientes para terapias antitumorais (Lord e Ashworth 2012). Na prática clínica, para testar se essa via está funcionando corretamente,

tipicamente avalia-se a presença de mutações germinativas ou somáticas nos genes supressores de tumor **BRCA 1/2** (*Breast Cancer 1/2*). Embora as mutações nesses genes estejam presentes em cerca de 10-15% dos casos associados a síndromes hereditárias, estudos envolvendo o sequenciamento em larga escala (**NGS**, *Next Generation Sequencing*) têm demonstrado a presença de mutações somáticas e eventos epigenéticos nesses genes e em outros envolvidos na via de **HR** (Krajewska et al. 2015). Além disso, estudos recentes demonstraram que outros cânceres esporádicos compartilham características moleculares e fenotípicas de tumores com mutações em *BRCA1/2*, um conceito denominado *BRCAness* (Lord e Ashworth 2016; Turner et al. 2004). A capacidade de estratificar esses tumores é importante, pois tem implicações significativas para o risco de câncer, estratégias de tratamento e sobrevida do paciente (Lord e Ashworth 2012; Pujade-Lauraine et al. 2017).

Atualmente, os principais biomarcadores para avaliar o nível de **HRD** são a detecção de mutações germinativas e somáticas com perda de função (**LoF**, *Loss of Function*) nos genes *BRCA1/2* (Hoppe et al. 2018). Outros biomarcadores para **HRD** incluem avaliar genes adicionais envolvidos no reparo de **HR**, a assinatura de expressão de genes que desempenham papéis cruciais no desenvolvimento do tumor e ensaios funcionais de expressão de proteínas. Devido à complexidade da via de **HR**, uma combinação de eventos relacionados a alterações moleculares determinantes pode melhorar a detecção de tumores com **HRD** (Polak et al. 2017; Hoppe et al. 2018). Particularmente, a avaliação do **HRD** na rotina clínica consiste em examinar as lesões que ocorrem no DNA, tanto para mutações germinativas quanto somáticas, juntamente com alterações epigenéticas nos genes ATM, PALB2, RAD51, BRCA1 e BRCA2, que estão confiavelmente associados ao fenótipo de **HRD** em cânceres como mama, ovário, pâncreas e próstata (Nguyen et al. 2020; Rempel et al. 2022). Existem também testes comerciais que avaliam HRD como *Myriad myChoiceCDx*, que são utilizados em câncer de ovário na clínica (Ray-Coquard et al. 2019), porém estes testes nem

sempre são utilizados devido ao seu custo (Fumagalli et al. 2022). Apesar das terapias serem utilizadas em tumores com mutações nesses genes, vários estudos têm demonstrado que **HRD** também pode ter outras causas, o que foi observado em vários tipos de tumores diferentes (Heeke et al. 2018).

1.4 ASSINATURAS MUTACIONAIS

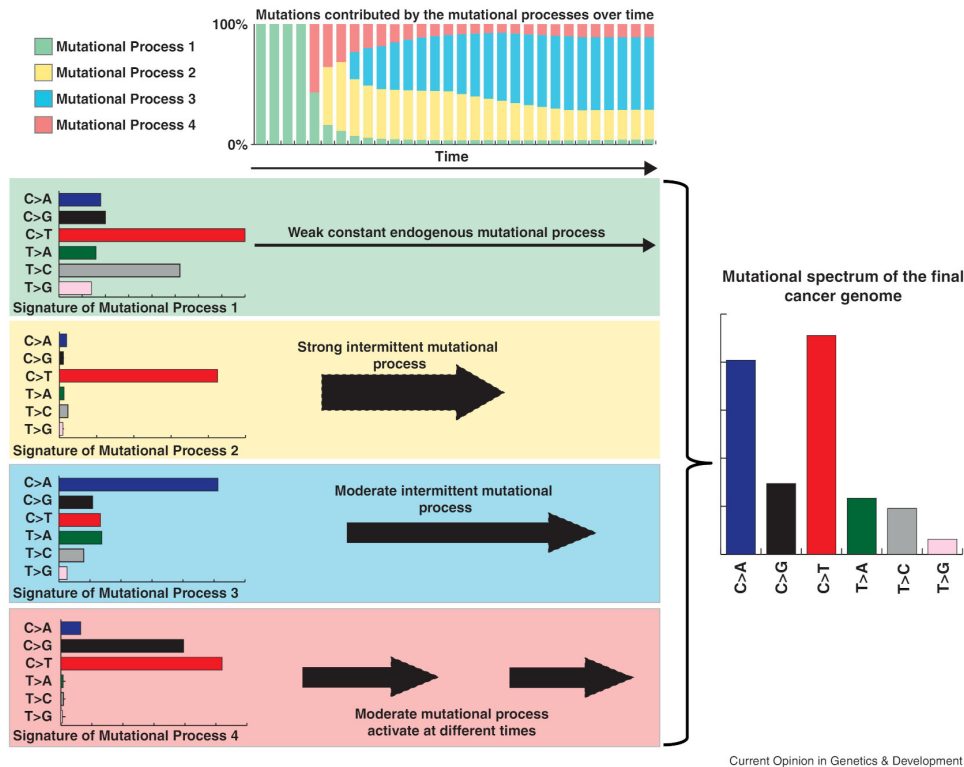
Assinaturas mutacionais são padrões distintos de mutações, consequências de um ou mais processos mutacionais endógenos ou exógenos que causam danos ao DNA. No câncer, uma proporção substancial de mutações somáticas é conhecidamente gerada por exposições a processos mutacionais. Exemplos desses processos são o fumo de tabaco em câncer de pulmão ou a exposição à radiação ultravioleta em câncer de pele (Alexandrov et al. 2013).

A figura 2 ilustra quatro processos mutacionais distintos operando em diferentes pontos na vida de uma pessoa. Cada um desses processos tem um padrão distinto caracterizado por seis tipos diferentes de mutações (Alexandrov e Stratton 2014).

Por meio do sequenciamento **NGS** de exoma ou genoma, é possível detectar a presença das mutações somáticas em uma amostra tumoral. A análise de assinaturas mutacionais permite decompor essas mutações em seus processos mutacionais de origem (Alexandrov et al. 2013; Rosales et al. 2016). Consequentemente, essa abordagem é capaz de identificar assinaturas resultantes de diversos processos mutacionais em um espectro muito mais amplo de cânceres do que anteriormente avaliado, com implicações terapêuticas significativas (Hoeck et al. 2019).

A assinatura mutacional 3 (**SBS3**, descrito no banco de dados COSMIC²) é um biomarcador promissor para **HRD** no câncer de mama (Prakash et al. 2015; Polak et al. 2017; Hoppe et al. 2018) e também reconhecido como uma

² <https://cancer.sanger.ac.uk/cosmic/signatures>



Fonte: Alexandrov e Stratton (2014).

Figura 2 – Exemplo simulado de processos mutacionais no câncer ao longo da vida de um indivíduo. No início todas as mutações em uma célula eram causadas pelo processo 1. Com o passar do tempo, outros processos mutacionais são ativados e o espectro mutacional muda. As mutações somáticas observadas por meio de sequenciamento em uma amostra de tumor (à direita) são a combinação desses processos distintos (à esquerda).

leitura confiável do nível de **HRD** em outros tumores. Essa assinatura resulta de amplas mudanças moleculares (por exemplo, alterações germinativas e somáticas em genes envolvidos em **HR** e silenciamento epigenético de *BRCA1* e *RAD51c*, entre outros) em componentes da via de **HR** que são independentes de mutações de *BRCA1/2* (Polak et al. 2017). Foi proposto que os tumores com **HRD** podem ser estratificados de acordo com a atividade mutacional da assinatura 3, um marcador considerado mais robusto do que ensaios que detectam alterações moleculares em genes específicos relacionados à via **HR** (Davies et al. 2017; Polak et al. 2017).

1.4.1 Cicatrizes genômicas para HRD

Além das assinaturas baseadas em matriz de substituições no DNA de uma única base (**SNV**, do inglês *Single Nucleotide Variation*), três tipos distintos de cicatrizes genômicas foram propostos como biomarcadores de **HRD**. Transições de estado em larga escala (**LST**, do inglês *large-scale state transitions*) definidas como uma quebra cromossômica entre regiões adjacentes de pelo menos 10 Mb (Popova et al. 2012), pontuação de desequilíbrio alélico telomérico (**TAI**, do inglês *telomeric allelic imbalance*) medida a partir de uma distribuição desigual de regiões sub-cromossômicas entre os dois alelos de um cromossomo especificamente nas regiões do telômero (Birkbak et al. 2012) e perda de heterozigosidade (**LOH**, do inglês *loss of heterozygosity*) onde o grau de **HRD** foi definida como a distribuição de comprimentos de regiões **LOH** por todo o genoma em uma amostra de tumor (Abkevich et al. 2012). Tomadas em conjunto, essas três pontuações baseadas em DNA medem características genômicas semelhantes (Sztupinski et al. 2018).

Finalmente, outras assinaturas para determinar o nível de **HRD** se estendem além dessas cicatrizes genômicas, BRCA1/2 e genes conhecidos relacionados a **HR**. A assinatura mutacional **ID6**, baseado em alterações somáticas de inserção e deleção (Alexandrov et al. 2020), e as assinatura mutacionais **CN1** e **CN17**, baseada em alterações no número de cópias do genoma **CNV** (do inglês *copy number variation*) também foram relatadas como associada ao **HRD** (Steele et al. 2022).

1.5 HISTOPATOLOGIA

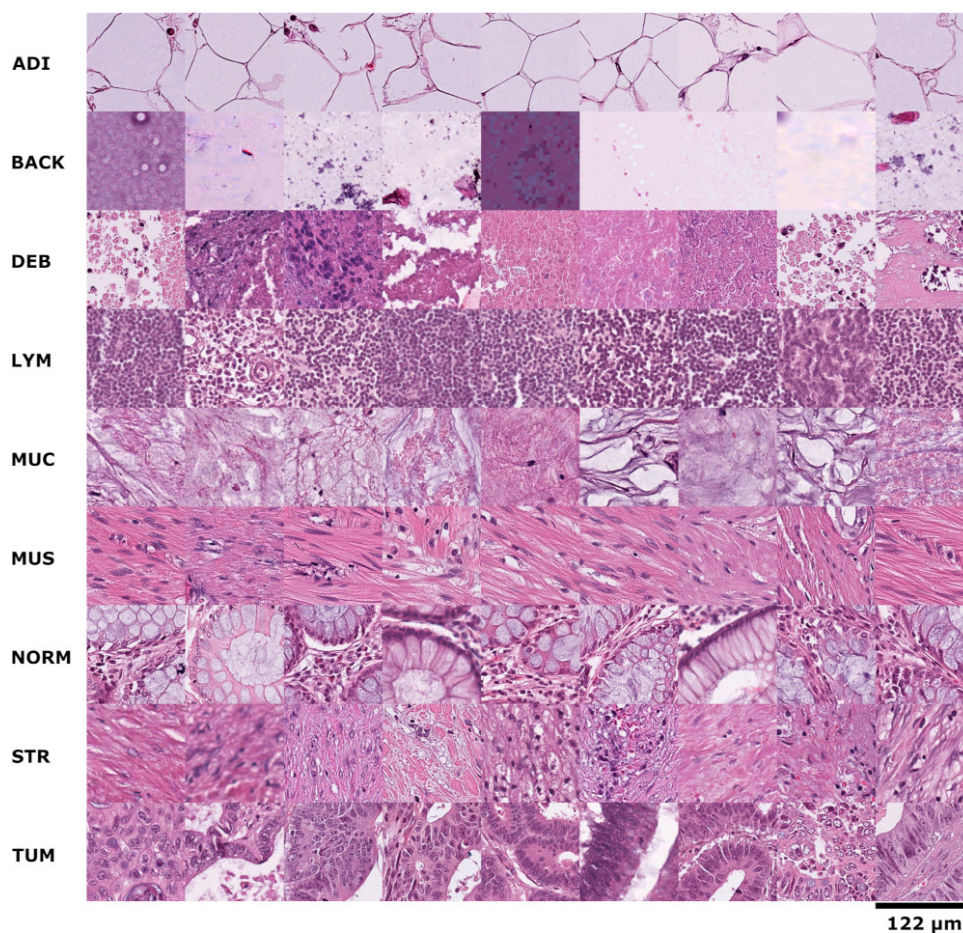
A histologia é um campo da biologia que estuda a anatomia microscópica de tecidos biológicos (Orchard e Nation 2011). A histopatologia é o estudo de como uma doença altera um tecido ou conjunto de células por meio da análise de suas características morfológicas por meio de uma análise de microscopia (Orchard

e Nation 2011).

Tipicamente, uma amostra de tecido é coletada de um paciente por meio de uma biópsia ou a partir da ressecção de tecido retirado de um paciente durante uma cirurgia. Essas amostras são fixadas em formalina e incorporadas em blocos de parafina, para preservá-las e mantê-las estáveis para armazenamento. Esses blocos podem ser cortados em fatias finas de 4 a 5 micrômetros de espessura e fixadas a lâminas de vidro. Por fim, a lâmina é tratada com corantes que realçam estruturas específicas presentes na amostra, permitindo sua análise por um patologista através de um microscópio. Uma das colorações mais comuns utilizadas na prática clínica é a Hematoxilina e Eosina (**H&E**), na qual a hematoxilina colore o núcleo das células em roxo, enquanto a eosina colore o citoplasma e o tecido conjuntivo em rosa. Essa coloração é rotineiramente utilizada na clínica e permite ao patologista identificar padrões morfológicos distintos (Figura 3), como crescimento descontrolado ou células com núcleo irregular como acontece em casos de câncer. Anticorpos também podem ser utilizados para identificar a presença de certas proteínas nas células do tecido. Esse processo é chamado de imuno-histoquímica ou **IHC** (do inglês *Immunohistochemistry*) e permite detectar a presença de proteínas importantes para o diagnóstico de certos subtipos de câncer, como por exemplo o **HER2**.

A patologia digital é um campo da histopatologia que utiliza imagens digitais para analisar, gerenciar, compartilhar e interpretar informações de patologia. Nesse processo, máquinas especializadas são capazes de digitalizar lâminas histopatológicas em alta resolução, equivalente à utilização de um microscópio, mas com a vantagem de que, uma vez digitalizada, essa imagem pode ser compartilhada e acessada remotamente por meio de um computador, permitindo a telepatologia e a análise automatizada dessas imagens por programas de computador.

Uma lâmina patológica tipicamente tem tamanho físico de 75 por 25 mm, e escâneres patológicos atuais são capazes de digitalizá-las com uma magnificação



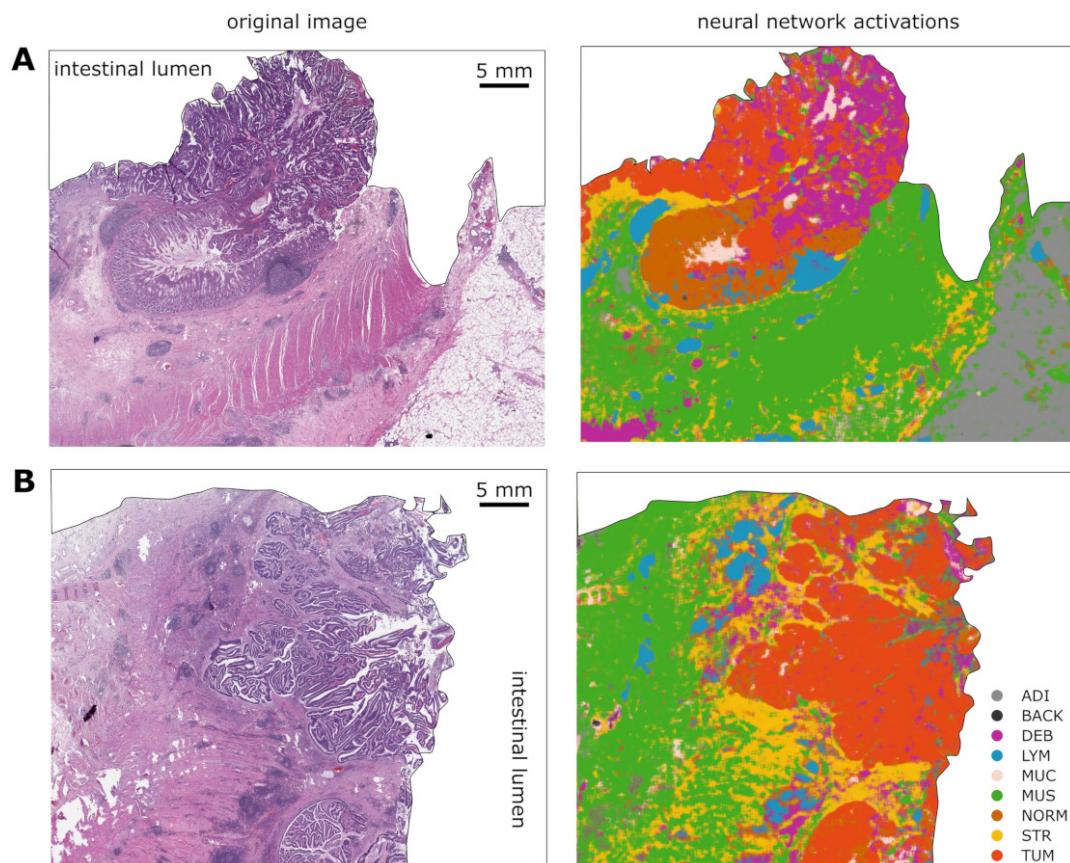
Fonte: Kather et al. (2019).

Figura 3 – Exemplos de diferentes tipos de tecidos patológicos, do conjunto de dados *NCT-CRC-HE-100K*, encontrados em imagens histopatológicas, caracterizados por seus padrões morfológicos. De cima para baixo: tecido adiposo, fundo branco, detritos, linfócitos, muco, músculo liso, mucosa normal de cólon, estroma associado ao câncer e epitélio tumoral de cólon.

que varia de 20x a 80x, o que equivale a uma resolução de 0,5 micrômetros por pixel (a 20x) a 0,125 micrômetros por pixel (a 80x). Por exemplo, com uma resolução de 0,5 micrômetros por pixel, cada lâmina resulta em uma imagem digital de 150.000 x 50.000 pixels x 3 canais de cor, totalizando 22 gigabytes de informação bruta.

Tradicionalmente, os patologistas avaliam essas imagens sob um microscópio em busca de marcadores-chave relevantes para o diagnóstico e a estratificação do tumor em seus subtipos, e suas conclusões estavam sujeitas à subjetividade da interpretação de cada profissional. Porém, avanços recentes na área de patologia

digital possibilitam a aplicação de análises computadorizadas para auxiliar o patologista nessa interpretação (Figura 4). Essas análises são capazes de destacar estruturas celulares importantes, automatizar a contagem de células, classificar tipos distintos de tecidos, detectar a presença de células mitóticas, nódulos cancerígenos, entre outros (Aeffner et al. 2019).

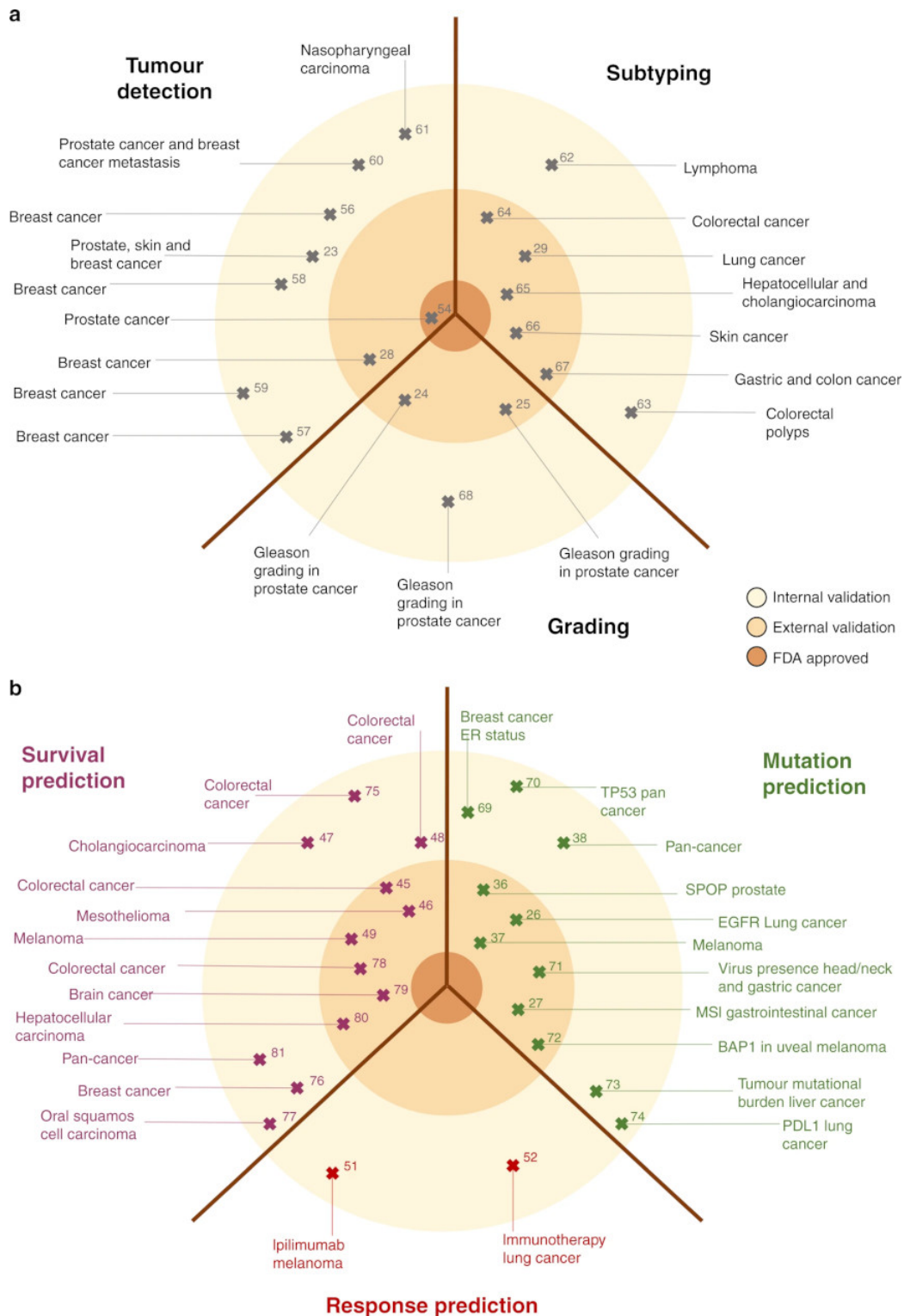


Fonte: Kather et al. (2019).

Figura 4 – Exemplo de segmentação computadorizada utilizando redes neurais. (A) e (B) ilustram exemplos de lâminas H&E de ressecção de tumor colorretal. A direita é ilustrada a classificação de uma rede neural nas respectivas lâminas. A rede neural foi treinada para identificar os tecidos distintos ilustrados na Figura 3.

Além da identificação de padrões morfológicos, estudos mostram a viabilidade de utilizar informações ocultas contidas em imagens histopatológicas para predição de características moleculares e genéticas da célula que, por sua vez, não são visíveis aos olhos do patologista, incluindo a presença de uma mutação patogênica ou expressão de uma proteína (Trivizakis et al. 2020). Como resul-

tado, vários biomarcadores têm sido investigados na detecção de tumor, subtipos, avaliação do grau de agressividade do tumor e predição de sobrevida (Figura 5) (Echle et al. 2020). Uma possível explicação para isso é a hipótese de que certas alterações moleculares e genéticas induzem alterações morfológicas no tecido que podem ser detectadas em imagens histopatológicas (Trivizakis et al. 2020).



Fonte: Echle et al. (2020).

Figura 5 – Aplicações de redes neurais na prática clínica. Cada item indica um estudo publicado, indicado pelo seu número de referência do artigo Echle et al. (2020). Cada estudo foi classificado de acordo com seu nível de evidência, indicado pela posição do ponto no círculo. (A) Análises básicas de imagem, incluindo detecção de tumor, gradação, e subtipagem. (B) Análises avançadas de imagem, onde a tarefa excede a capacidade do patologista, como predição de presença de uma mutação, prognóstico e resposta a um tratamento.

1.6 INTELIGÊNCIA ARTIFICIAL

A Inteligência Artificial (**IA**, do inglês *Artificial Intelligence*) é uma área de estudo multidisciplinar que desenvolve métodos e algoritmos que permitem que máquinas reajam ao ambiente e tomem decisões que maximizem objetivos pré-definidos (Russell e Norvig 2020). O termo Inteligência Artificial foi cunhado por John McCarthy em 1956, durante uma conferência sobre o tema. No entanto, a possibilidade das máquinas simularem o comportamento humano e realmente pensarem foi levantada anteriormente por Alan Turing, que desenvolveu o *Teste de Turing* - experimento que avalia se uma máquina é capaz de imitar o comportamento humano (Xu et al. 2021).

Dentro da área de **IA**, o subcampo Aprendizado de Máquina (**ML**, do inglês *Machine Learning*) estuda e desenvolve algoritmos estatísticos capazes de aprender padrões a partir de dados e generalizá-los para outros dados similares. Os algoritmos de aprendizado de máquina podem ser classificados em quatro tipos principais: supervisionado, não supervisionado, semi-supervisionado e por reforço. O aprendizado supervisionado, por exemplo, treina modelos com base em dados rotulados, enquanto o não supervisionado explora estruturas ocultas em dados não rotulados. Já o semi-supervisionado combina ambos, e o aprendizado por reforço utiliza interações com o ambiente para otimizar decisões. No contexto de aprendizado profundo (**DL**, *deep learning*), essas categorias são aplicadas utilizando redes neurais profundas, capazes de identificar padrões complexos em grandes volumes de dados, como reconhecimento de padrões, processamento de imagens médicas e predição de biomarcadores (Echle et al. 2020). Uma revisão dos aspectos metodológicos das técnicas atuais de **IA** e métricas para avaliação do desempenho de modelos de redes neurais pode ser encontrada no Apêndice A desta tese.

A **IA** tem um papel crescente na oncologia, especialmente na detecção precoce de câncer por meio da interpretação de exames de imagem, como ma-

mografias, tomografias e colonoscopias, oferecendo maior precisão e agilidade. Diversos dispositivos de **IA** já foram aprovados pela Administração de Alimentos e Medicamentos dos Estados Unidos (FDA) e/ou pela União Europeia (UE) para auxiliar na detecção de câncer de mama (Lotter et al. 2024).

O desenvolvimento de modelos de **IA** em oncologia segue etapas bem definidas que devem garantir sua eficácia e segurança. No contexto da análise de imagens, de maneira resumida, o processo inicia com a coleta de imagens de alta qualidade, muitas vezes provenientes de múltiplos centros para garantir a generalização do modelo. Essas imagens, por sua vez, são pré-processados, incluindo etapas de normalização, segmentação e remoção de ruídos, seguidas pela anotação, onde especialistas fornecem o *ground truth* para treinamento supervisionado. Após a construção do modelo, este é treinado em conjunto de dados, enquanto um subconjunto independente é reservado para testes. A validação cruzada e a validação independente em coortes externas são essenciais para avaliar a robustez do modelo. Métricas como acurácia, AUC-ROC, sensibilidade e especificidade são utilizadas para medir o desempenho, e técnicas de interpretabilidade, como mapas de calor, ajudam a entender as decisões do modelo e a aumentar sua aceitação clínica (Gillies et al. 2016).

1.6.1 Oportunidades e desafios na análise de imagens H&E

De fato, embora muitos biomarcadores oncológicos dependam de ensaios baseados em biologia molecular, os avanços nas tecnologias baseadas em **IA** estão ampliando significativamente a capacidade de revelar informações anteriormente inacessíveis diretamente de dados de imagem. Há uma grande quantidade de informação contida nas imagens histopatológicas, e seu potencial de uso no âmbito clínico permanece ainda a ser totalmente explorado. Processar essa grande quantidade de dados é um desafio por si só, porém, avanços recentes na inteligência artificial, em conjunto com o desenvolvimento de recursos computacionais especializados (Schmidt e Hildebrandt 2024), geraram a oportunidade de investigar

esse tipo de dado de maneira custo-efetiva.

Modelos de redes neurais, utilizando técnicas modernas de **DL**, são capazes de aprender a detectar características ocultas em imagens digitais de forma automatizada a partir de um conjunto de dados associados (Kather et al. 2019). Esses métodos possibilitam o processamento de conjuntos de dados de milhares de lâminas histopatológicas em um tempo relativamente pequeno, comparado ao trabalho de inspeção visual realizado por patologistas, além de permitir a avaliação de novos tipos de biomarcadores que não seriam possíveis sem a aplicação de métodos de aprendizado de máquina.

Nos últimos anos, estudos têm se dedicado a prever biomarcadores em tumores sólidos a partir de cortes histológicos corados com H&E (Echle et al. 2020), incluindo biomarcadores para identificar deficiências nos mecanismos de reparo do DNA. No entanto, os dados de referência (*ground truth*) para avaliação de deficiência de **HRD** em modelos de **IA** geralmente são definidos com base em cicatrizes genômicas, assinaturas mutacionais ou mutações isoladas em genes relacionados à recombinação homóloga (Valieris et al. 2020; Lazard et al. 2022; Loeffler et al. 2023; Nahhas et al. 2024; Bergstrom et al. 2024).

É fundamental reconhecer que a via **HR** é regulada por várias vias de sinalização e reguladores-chave, além dos genes BRCA1/2 e outros genes relacionados diretamente à **HR** (Krajewska et al. 2015). Essas variações podem se manifestar de forma distinta em diferentes tipos de tumores e até mesmo entre pacientes com o mesmo tipo de tumor. Além disso, embora custo e acessibilidade sejam desafios conhecidos, os testes atuais de **HRD** muitas vezes não capturam toda a complexidade ou extensão do fenótipo associado à deficiência de recombinação homóloga. Nesse contexto, há uma necessidade clínica significativa por um teste de triagem para **HRD** que seja acessível, rápido e padronizado, capaz de abranger uma ampla gama de alterações associadas aos fenótipos de **HRD**, em vez de se concentrar exclusivamente em medidas moleculares individuais.

Não obstante, embora a relevância do uso de **IA** na avaliação do *status* de

HER2 tenha sido investigada (Couture et al. 2018; Qaiser et al. 2018; Shamaï et al. 2019; Jaber et al. 2020; Rawat et al. 2020; Gamble et al. 2021; Xue et al. 2023), esses modelos foram treinados em imagens de microarranjos teciduais, lâminas coradas por imuno-histoquímica (**IHC**) e imagens de hibridização *in situ* fluorescente (**FISH**). Além disso, os classificadores existentes foram predominantemente utilizados para prever o *status* de **HER2** em conjunto com os receptores hormonais ou para classificar subtipos intrínsecos, limitando a classificação de **HER2** a positivo ou negativo.

Com as opções de tratamento recentemente disponibilizadas para pacientes com tumores **HER2** baixo (*HER2-low*), a avaliação desses tumores por modelos baseados em aprendizado profundo ainda é um campo inexplorado e apresenta um desafio promissor a ser enfrentado. Esse cenário destaca o potencial do aprendizado de máquina para superar as limitações dos métodos tradicionais, fornecendo uma base sólida para a detecção de biomarcadores complexos em câncer de mama, diretamente a partir de imagens histológicas.

2 OBJETIVOS

Motivados por esses avanços recentes tanto no âmbito da patologia digital quanto no desenvolvimento de modelos de **IA**, e partindo da hipótese de que alterações genéticas induzem mudanças celulares refletidas em características ocultas morfológicas (Trivizakis et al. 2020), decidimos investigar a viabilidade da predição de biomarcadores moleculares relevantes ao câncer de mama a partir de imagens coradas com H&E, com o objetivo de buscar novas soluções custo-efetivas em oncologia.

2.1 OBJETIVO GERAL

Investigar a capacidade de predição de biomarcadores relevantes para o câncer de mama a partir de imagens histopatológicas.

2.2 OBJETIVOS ESPECÍFICOS

- Compilar um conjunto de dados clínicos e de imagens a partir de repositórios públicos e do A.C. Camargo Cancer Center;
- Implementar um *framework* computacional capaz de integrar todas as etapas do desenvolvimento de um modelo de **IA**;
- Explorar diferentes métodos para treinamento de redes neurais com imagens histopatológicas;
- Predizer *status* de **HRD** e subgrupos de **HER2**, incluindo o *HER2-low*, a partir de imagens histopatológicas.
- Implementar técnicas analíticas capaz de fornecer interpretabilidade em nível de lâmina para classificação.

3 METODOLOGIA

3.1 COORTES INVESTIGADAS

3.1.1 TCGA

O atlas do genoma do câncer (**TCGA** , do inglês *The Cancer Genome Atlas*) , é um projeto de referência no estudo do câncer. O consórcio sequenciou o genoma de mais de 20.000 amostras de câncer de 33 tipos distintos. Este projeto, que levou mais de uma década para ser completado, gerou mais de 2,5 petabytes de dados genômicos, epigenômicos, transcriptômicos, proteômicos e de imagens (Cancer Genome Atlas Research Network et al. 2013). Os dados resultantes¹, oferecem uma grande oportunidade para desenvolver uma visão integrada das similaridades, diferenças e temas emergentes entre as linhagens tumorais.

Nós utilizamos dados de sete tipos de tumores no presente trabalho, somando um total de 1.075 lâminas de 1.060 pacientes, incluindo dados genéticos de mutações somáticas destes pacientes. Especificamente, mama (**TCGA-BRCA**) contendo 1075 lâminas, próstata (**TCGA-PRAD**) contendo 898 lâminas, pâncreas (**TCGA-PAAD**) contendo 124 lâminas, endométrio (**TCGA-UCEC**) contendo 366 lâminas, adenocarcinoma de cólon (**TCGA-COAD** contendo 342 lâminas, pulmão (**TCGA-LUAD**) contendo 416 lâminas, e estômago (**TCGA-STAD**) contendo 325 lâminas).

3.1.2 CPTAC

O Consórcio de Análise Proteômica Clínica de Tumores (**CPTAC**, do inglês *Clinical Proteomic Tumor Analysis Consortium*) é um projeto que reuniu dados genômicos, proteômicos e de imagens de mais de 1.500 pacientes com câncer de diversos tecidos com o objetivo de melhorar o diagnóstico, tratamento e

¹ <https://www.cancer.gov/ccg/research/genome-sequencing/tcga>

prevenção do câncer.

Nós utilizamos dados da coorte de câncer de mama (**CPTAC-BRCA**) neste trabalho, contendo 653 lâminas de 122 pacientes².

3.1.3 HEROHE

HEROHE³ foi uma competição organizada por uma cooperação entre o Instituto de Investigação e Inovação em Saúde e o Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, da Universidade do Porto, Portugal. Esta competição disponibilizou um conjunto de dados de imagens histopatológicas associado a anotações obtidas a partir da avaliação do **HER2**, onde equipes competiram na construção de modelos de aprendizado de máquina capazes de distinguir HER2-positivo de HER2-negativo (Conde-Sousa et al. 2022).

Notavelmente, após o encerramento da competição, o conjunto de dados foi disponibilizado publicamente e, portanto, nós utilizamos esta coorte contendo 356 lâminas de tumores de câncer de mama.

3.1.4 AC-BRCA

AC-BRCA-HER2 é um conjunto de dados anonimizado de pacientes de câncer de mama tratados no A. C. Camargo Câncer Center durante os anos de 2019 e 2021. Ao total, foram obtidas 546 lâminas H&E de 504 pacientes contendo biópsia ou ressecção do tumor. Todos os casos com metástase ou terapia neoadjuvante foram excluídos da casuística. Foram obtidos os escores de HER2 dos resultados para do teste de imuno-histoquímica para os e de hibridização in situ nos casos onde a coloração foi interpretada como 2+ equívoco. Este estudo recebeu aprovação do Comitê de Ética da Fundação Antônio Prudente (protocolo N^o 3028/20).

² <https://proteomics.cancer.gov/programs/cptac>

³ <https://ecdp2020.grand-challenge.org>

AC-BRCA-HRD-1 é um subconjunto de **AC-BRCA-HER2** de 65 pacientes com informação disponível de testes genéticos de painel **NGS** onde foi avaliada a presença de mutações relevantes ao **HRD**.

AC-BRCA-HRD-2 é um subconjunto de **AC-BRCA-HRD-1** de 22 pacientes com subtipo triplo-negativo com dados genômicos de sequenciamento de todos os exons (exoma) do genoma humano que, por sua vez, permitiram o cálculo das assinaturas mutacionais para estimação de escores de **HRD**.

3.2 TÉCNICAS DE APRENDIZADO PROFUNDO

Uma revisão dos aspectos metodológicos das técnicas atuais de **IA** e métricas para avaliação do desempenho de modelos de redes neurais pode ser encontrada no Apêndice A desta tese.

3.3 PRÉ-PROCESSAMENTO DE IMAGENS HISTOPATOLÓGICAS

3.3.1 Separação da Lâmina em Fragmentos

Uma lâmina patológica tem alta resolução e, por limitações atuais de *hardware*, não é possível construir um modelo capaz de processá-las como um todo de uma vez. Por isso, usualmente a imagem é dividida em fragmentos menores que podem ser analisados por partes e agregados em uma predição única.

Primeiramente, as imagens de um conjunto de dados são redimensionadas de modo que todas tenham a mesma resolução em termos de micrômetros por pixel, tipicamente 0,5 μm /pixel. Em seguida, as imagens são divididas em fragmentos menores, e o tamanho específico varia de acordo com a arquitetura da rede neural utilizada, mas tamanhos comuns utilizados são 224×224 pixels ou 256×256 pixels. Esse processo resulta em um conjunto de milhares de fragmentos por lâmina, que na terminologia do **MIL** é chamado de “sacola”.

Essa divisão de fragmentos pode ser realizada com ou sem sobreposição adjacente. Quando há sobreposição, é definida uma porcentagem p como a quantidade de redundância que será compartilhada entre os fragmentos adjacentes da imagem.

Seja I uma imagem $W \times H$ de uma lâmina digitalizada, um pedaço quadrado dessa imagem pode ser definido por suas coordenadas x e y contidos em I e seu tamanho quadrado t , de forma que $I_{x,y,t}$ define um pedaço da imagem I contido nas coordenadas (x, y) e $(x + t, y + t)$.

A divisão da imagem I em fragmentos sem sobreposição consiste em criar um conjunto de fragmentos S , tal que

$$S = \{I_{0,0,t}, I_{t,0,t}, I_{t,t,t}, I_{2t,t,t}, \dots, I_{W-t,H-t,t}\}$$

Enquanto que a divisão da imagem I em fragmentos com sobreposição, para uma porcentagem de sobreposição p é definida como

$$S = \{I_{0,0,t}, I_{t \cdot p, 0, t}, I_{t \cdot p, t \cdot p, t}, I_{2t \cdot p, t \cdot p, t}, \dots, I_{W-t, H-t, t}\}$$

onde p é um número real $0 < p < 1$. Neste trabalho, utilizamos 50% de sobreposição para gerar os fragmentos.

3.3.2 Remoção de Fundo

Essas “sacolas” de imagens podem ser utilizadas por algoritmos **MIL** para treinamento de modelos, porém existe uma grande diversidade de tipos de tecidos contidos em uma lâmina, inclusive regiões em branco onde nenhum tecido está presente, que também são capturadas na imagem digitalizada. A remoção desses fragmentos antes do treinamento do algoritmo de **MIL** é desejável, pois reduz consideravelmente a carga computacional necessária para o processamento desses dados, além de reduzir a possibilidade de predições equivocadas baseadas nesses fragmentos que não têm relação com o objetivo do modelo.

Existem vários algoritmos que realizam a detecção desses fragmentos em branco. Neste trabalho nós utilizamos o método de Otsu (1979). Esse método

separa os tons de cinza de uma imagem em grupos e permite uma detecção rudimentar de bordas na imagem. Fragmentos que não tiverem detecção são considerados fragmentos vazios e são descartados das etapas de treinamento subsequentes.

3.3.3 Identificação de Tumor

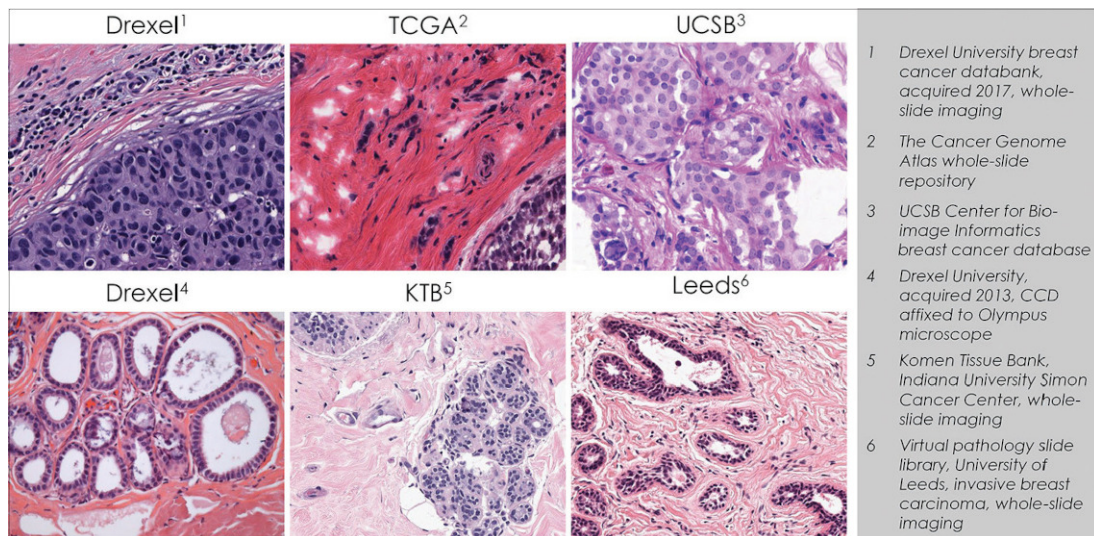
Além da remoção do fundo das imagens, também é possível detectar diferentes subtipos de tecidos e utilizar apenas uma parte de uma lâmina no treinamento de um modelo. Essa anotação pode ser feita por um patologista utilizando um programa de computador especializado (Martins et al. 2023), mas também é possível utilizar modelos de redes neurais para automatizar esta anotação.

Para identificar fragmentos contendo tecido tumoral, utilizamos o modelo de **CNN** *ConvNeXt-Tiny* (Liu et al. 2022) treinado no conjunto de dados *NCT-CRC-HE-100K* (Figura 3) (Kather et al. 2018). Esse conjunto de dados consiste em 100.000 fragmentos de imagens histopatológicas anotadas em nove tipos de padrões diferentes, incluindo fundo branco. Aplicamos esse modelo treinado no conjunto **TCGA-BRCA** para anotar as regiões com e sem fundo de cada lâmina. Combinamos essas anotações com anotações do conjunto de dados *Manual tumor annotations in TCGA* (Loeffler e Kather 2021), que contém regiões de tumor anotadas manualmente para as lâminas do **TCGA** e, ao final, utilizamos essas anotações combinadas para treinar um modelo que é capaz de classificar as regiões tumor, não-tumor e fundo.

3.3.4 Normalização de Cor

Imagens histopatológicas podem apresentar diferenças significativas em sua coloração, dependendo do processo utilizado na preparação das lâminas e do equipamento utilizado em sua digitalização (Figura 6). Isso pode ser um problema ao combinar dados de origens distintas no treinamento de modelos, pois padrões

morfológicos similares podem apresentar aparências diferentes, o que pode levar o modelo a interpretá-las como padrões distintos.



Fonte: Aeffner et al. (2019).

Figura 6 – Exemplos de colorações H&E de seis conjuntos de dados diferentes, ilustrando possíveis variações de tonalidade.

Existem duas abordagens para lidar com isso, *color augmentation* (aumento de cor), e *color normalization* (normalização de cor) (Tellez et al. 2019). Técnicas de *color augmentation* introduzem variações factíveis na coloração das imagens utilizadas no treinamento, com o objetivo de tornar o modelo mais robusto em relação a essas variações. Já as técnicas de *color normalization* buscam reduzir a variação de cor dentro de um conjunto de dados, com o objetivo de eliminar essas variações. Ambas as abordagens partem do pressuposto de que essas variações não são relevantes para a predição do modelo, e, portanto, o desempenho do modelo pode ser melhorado com a aplicação dessas técnicas (Tellez et al. 2019). Neste trabalho, aplicamos a normalização de cor pelo método de *Macenko* (Macenko et al. 2009).

3.3.5 Extração de Características

A extração de características para imagens histopatológicas é similar ao descrito na Seção A.5, porém notavelmente essa extração é feita após a lâmina ser

dividida em fragmentos. Então é feita uma inferência com o modelo extrator em cada pedaço, resultando em um conjunto de vetores unidimensionais de características para cada lâmina, a “sacola” na terminologia do **MIL**. Esses dados são armazenados e utilizados no treinamento de um modelo classificador subsequente. É possível utilizar extratores de características “genéricos”, que são treinados em conjuntos de dados expansivos, como o *ImageNet* (Deng et al. 2009), porém recentemente têm surgido vários modelos treinados especificamente para extração de características em imagens histopatológicas. Esses modelos são treinados em milhares de lâminas histopatológicas e muitos deles são disponibilizados à comunidade científica para uso em novos experimentos (Wang et al. 2022; Filiot et al. 2023; Wang et al. 2023; Chen et al. 2024; Vorontsov et al. 2024).

Neste trabalho, o modelo **UNI** (Chen et al. 2024) foi utilizado para a extração das características. Esse modelo foi treinado com um conjunto de dados contendo mais de 100 milhões de fragmentos, de 100.000 lâminas H&E, de 20 tecidos diferentes (Chen et al. 2024). A grande quantidade e diversidade de dados patológicos utilizados no treinamento desse modelo permite a detecção de diversos padrões diferentes presentes nessas imagens, que são utilizados no modelo classificador.

3.4 ANOTAÇÃO DO NÍVEL DE **HER2**

Para treinar os modelos de predição de **HER2**, os rótulos utilizados foram derivados do laudo de exames de **IHC** e **ISH** disponíveis em cada coorte utilizada. Os pacientes com dados disponíveis foram anotados nos grupos *HER2-neg*, *HER2-low* e *HER2-high*, de acordo com a Tabela 1.

3.5 ESTIMAÇÃO DE ASSINATURAS MUTACIONAIS PARA **HRD**

Para quantificar **HRD**, nós utilizamos quatro tipos de assinaturas mutacionais estimadas nas coortes do **TCGA** e **AC-BRCA**. As assinaturas de

Tabela 1 – Esquema de particionamento de dados para os modelos HER2. Cada coluna corresponde ao rótulo utilizado no treinamento, e **IHC** e **ISH** indicam o estado do teste correspondente. As linhas, **M1** ... **M6** representam os seis modelos de HER2 desenvolvidos e os respectivos dados utilizados no treinamento de cada modelo.

	HER2-neg	HER2-low		HER2-high	
IHC	0	1+	2+	3+	
ISH		Neg		Pos	
M1			●	●	●
M2	●			●	●
M3		●	●	●	●
M4	●		●	●	●
M5	●	●	●		
M6	●	●	●	●	●

mutações pontuais (**SBS**) foram estimadas a partir do conjunto *COSMIC SBS V2*, e assinaturas de indels (**ID**) foram estimas com o *COSMIC ID V3.4*. Os dados mutacionais utilizados para a estimação foram obtidos do **MC3** (Ellrott et al. 2018), e a atividade mutacional destas assinaturas foi estimado com o *sigProfiler* (Díaz-Gay et al. 2023). Assinaturas de variação no número de cópias (**CN**) foram obtidas de Steele et al. (2022), e escores do *scarHRD* de Loeffler et al. (2023). Na coorte **AC-BRCA**, a estimação das assinaturas **SBS**, **CN**, **ID** e *scarHRD* foi realizada conforme descrito em seus respectivos métodos.

Dentre esses conjuntos de assinaturas, destacamos as assinaturas relacionadas a **HRD**, especificamente: **SSBS3**, **ID6**, **CN1**, **CN17** e *scarHRD*. A assinatura **SBS3** é uma assinatura de mutações pontuais associada a defeitos na via de reparo por recombinação homóloga, caracterizado por seu padrão plano e sem características. A assinatura **ID6** também está associada ao reparo de DNA via mecanismo de recombinação homóloga. **CN1** e **CN17** são assinaturas de variação no número de cópias, **CN1** é primariamente associado a genomas com pouca ou nenhuma variação de cópias e está negativamente associado a outras assinaturas de **HRD** como *scarHRD*. **CN17** é definido por segmentos de perda de heterozigose e também foi associada ao **HRD** (Steele et al. 2022). Por fim, *scarHRD* avalia a extensão do número de cópias, perda de heterozigose, e tran-

sições em larga escala, que são características associadas a variações estruturais no genoma ligadas ao **HRD** (Sztupinski et al. 2018).

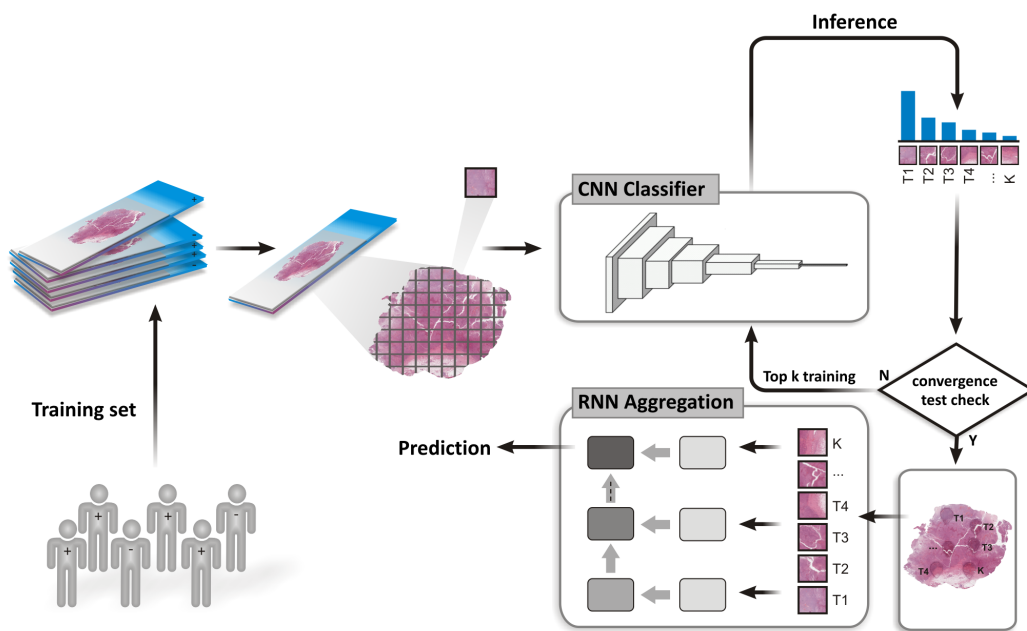
3.6 CÁLCULO DO SCORE AGREGADO PARA HRD

Apesar da relevância clínica do **HRD** como biomarcador, métodos tradicionais para sua quantificação apresentam limitações em capturar a complexidade das variações genéticas associadas a **HRD** de forma integrada. Assim, propomos o desenvolvimento de um score baseado em uma média ponderada da exposição de cada assinatura nos tumores, permitindo uma mensuração mais abrangente do **HRD**. Essa abordagem, por sua vez, tem o potencial de melhorar a estratificação dos pacientes e informar decisões terapêuticas mais personalizadas. O cálculo detalhado do score é descrito no Apêndice B.

4 RESULTADOS E DISCUSSÃO

4.1 CLASSIFICAÇÃO BINÁRIA DO HRD

Nós investigamos a predição do biomarcador de deficiência em recombinação homóloga (**HRD**) utilizando técnicas de aprendizado profundo com imagens histopatológicas. Para isso, nós obtivemos acesso ao conjunto de dados **TCGA-BRCA**, que contém lâminas digitalizadas de ressecção tumoral de câncer de mama, e treinamos um modelo baseado na técnica de **MIL** para predição binária do estado do **HRD** em cada paciente. Uma visão geral do processo de treinamento deste modelo é ilustrada na Figura 7. O estudo contendo essa análise foi publicado na revista *MDPI Cancers* (Valieris et al. 2020) em dezembro de 2020, e encontra-se anexado a este documento no Apêndice C.



Fonte: Valieris et al. (2020)

Figura 7 – Esquematisação do treinamento do modelo para predição do HRD binário.

4.1.1 Treinamento do Modelo

Para o treinamento do classificador, foi utilizado um modelo de CNN com a arquitetura *ResNet34* (He et al. 2016), pré-treinado no conjunto de dados ImageNet (Deng et al. 2009), esse modelo foi treinado utilizando a técnica **Top-K MIL**, como descrito na Seção A.6.1. O modelo agregador utilizado foi construído conforme descrito por Campanella et al. (2019).

Os rótulos utilizados no treinamento foram derivados a partir de uma análise de assinaturas mutacionais estimadas usando dados genômicos dos respectivos pacientes. Nós utilizamos o nível de atividade da assinatura 3 (**SBS3**) como preditor do nível de **HRD**. Os pacientes da coorte foram divididos em três quantis: menor que 25%, maior que 75% e “outros”, de acordo com o nível de atividade da **SBS3**. Amostras do quartil inferior (<25%) foram considerados **HRD** negativo, amostras no quartil superior (>75%) foram considerados **HRD** positivo e amostras no quartil “outros” não foram utilizadas no treinamento. O conjunto de dados foi dividido em treino (70%) e teste (30%). O código fonte utilizado para pre-processamento das imagens e treinamento do modelo encontra-se disponível em <https://github.com/rvalieris/deepHE>.

4.1.2 Performance do Modelo

Após o treinamento, utilizamos **AUC** para avaliar o desempenho do modelo no conjunto de teste, obtendo uma **AUC** de 0.80 e, em seguida, em um conjunto de dados independente (**AC-BRCA-HRD-1**) contendo amostras de 65 pacientes do A. C. Camargo Cancer Center, com seu *status* de **HRD** obtido a partir da análise de um painel de genes conhecidamente relacionados ao **HRD**. O modelo obteve **AUC** de 0.70 nessa coorte de validação independente.

Na prática clínica o teste de **HRD** é tipicamente feito através de um teste genético no qual são investigadas lesões genéticas em genes conhecidamente relacionados à via de recombinação homóloga. No entanto, esses testes tendem

a ser caros e não são capazes de detectar defeitos em outros genes reguladores que também podem estar envolvidos nesta via. Nesta análise, nós utilizamos assinaturas mutacionais para medir a presença de **HRD** nas amostras. Esse método não requer uma lista de genes a serem testados e, portanto, deve ser capaz de identificar a presença de **HRD** mesmo quando a mutação responsável não se encontra nos genes comumente testados. Particularmente, a assinatura **SBS3** reflete um mecanismo muito mais amplo de lesões em genes da via de recombinação homóloga além dos genes conhecidos, incluindo genes *BRCA1* e *BRCA2*, por exemplo (Polak et al. 2017).

Coletivamente, nós demonstramos que é possível utilizar imagens histopatológicas para predição de um biomarcador molecular. Embora já existam trabalhos publicados na literatura onde outros biomarcadores moleculares podem ser preditos usando **IA** e imagens histopatológicas, este estudo foi o primeiro a reportar predição de **HRD** especificamente. Isso é de particular importância, pois nosso estudo foi capaz de detectar alterações morfológicas ocultas associadas com características moleculares no contexto de uma via de reparo.

Não obstante, mostramos que a assinatura mutacional usada como *ground truth* durante o treinamento do modelo se mostrou robusta. Esse resultado reforça o conceito de que a assinatura reflete um espectro muito mais amplo de alterações genéticas que, por sua vez, levam a alterações histológicas que são capturadas pela abordagem proposta.

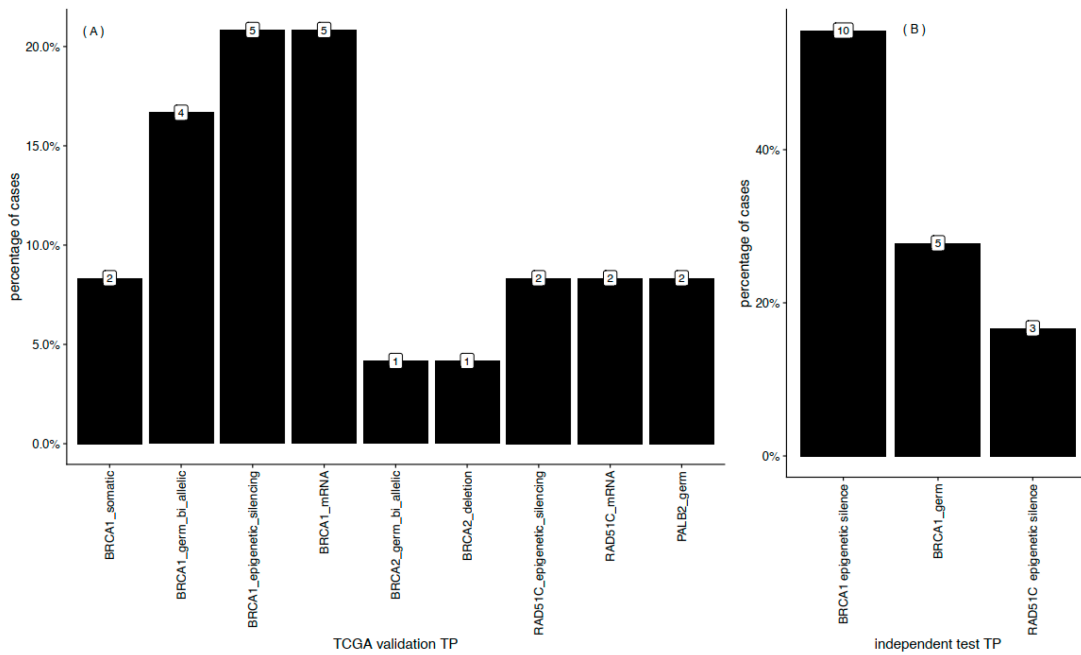
A luz de uma validação multicêntrica e extensa, reforçamos que investigações adicionais em um conjunto muito mais amplo de biomarcadores são necessárias, visando, ao final, compreender a extensão da informação contida nas imagens histopatológicas.

Classificação de imagens reflete *drivers* moleculares únicos

Após demonstrar que nossa abordagem pode identificar de forma eficaz características preditivas do *status* de **HRD** a partir de imagens H&E, exploramos os *drivers* moleculares que sustentam essas predições. Nesta etapa, utilizamos

resultados de uma análise abrangente de alterações que provavelmente contribuem **HRD** em tumores de mama (Polak et al. 2017) dos pacientes que, por sua vez, apresentam dados moleculares e imagens pareadas. Conforme ilustrado na Figura 8, é apresentada a distribuição de lesões moleculares, classificadas por tipo, nos genes específicos da via de recombinação homóloga nos casos de câncer de mama examinados no conjunto de teste (8A) e validação independente (Figura 8B).

Além das variantes em *BRCA1* (incluindo variantes germinativas bialélicas, silenciamento epigenético e mutações somáticas) e *BRCA2* (deleções e variantes germinativas bialélicas), foram identificadas alterações adicionais em outros genes da via de recombinação homóloga. Por exemplo, foram detectadas variantes germinativas em *PALB2* e silenciamento epigenético em *RAD51C*, sendo este último evidenciado por níveis baixos de expressão. Corroborando esses achados, nosso modelo identificou com precisão pacientes com deficiências funcionais causadas pela inativação de genes relacionados à via de recombinação homóloga em uma coorte independente de câncer de mama (Figura 8B). Além disso, os perfis moleculares observados estão alinhados com caracterizações anteriores do mecanismo de **HRD**, reforçando a validade de nossa abordagem.



Fonte: Valieris et al. (2020)

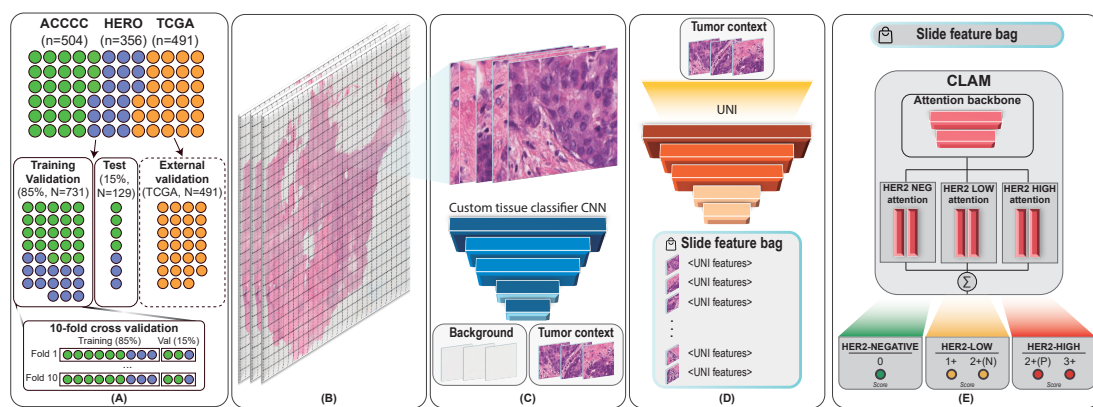
Figura 8 – Distribuição de lesões nos genes conhecidos da via de recombinação homóloga (HR) nas as coortes de treinamento (A) e independentes (B). Os números acima de cada barra indicam a quantidade de pacientes correspondentes.

4.2 CLASSIFICAÇÃO DOS SUBTIPOS DE HER2

Nós investigamos a predição do nível de **HER2** através de imagens histopatológicas. Para isso, obtivemos imagens coradas com H&E de três coortes distintas, incluindo: **TCGA-BRCA**, **HEROHE**, e **AC-BRCA-HER2**. Em conjunto, recuperamos o laudo que informa o *status* de **HER2** avaliado por um patologista através dos testes de **IHC** e **ISH**, no qual são considerados padrão ouro e como é tipicamente realizado na prática clínica.

Esses dados foram utilizados para treinar uma rede neural baseada na arquitetura **CLAM** (Lu et al. 2021), com o objetivo de prever as categorias: negativa (*HER2-neg*), baixa (*HER2-low*) e alto (*HER2-high*). Uma visão geral do processo de treinamento deste modelo é ilustrada na Figura 9: primeiro, fragmentos (*tiles*) foram extraídas de imagens inteiras (*whole-slide images*, **WSI**) e classificadas em grupos de tumor e não tumor (Figura 9B, 9C); em seguida, os

grupos gerados de fragmentos para tumor e não tumor foram utilizados para extração de características com o modelo *UNI* (Figura 9D); por fim, os conjuntos resultantes de características foram usados para treinar um modelo de aprendizado por múltiplas instâncias (*multiple instance learning*, **MIL**) para classificar as **WSI** nas categorias *HER2-neg*, *HER2-low* e *HER2-high* (Figura 9E). O estudo contendo essa análise foi publicado na revista *BMC Breast Cancer Research* (Valieris et al. 2024) em agosto de 2024, e encontra-se anexado a este documento no Apêndice D.



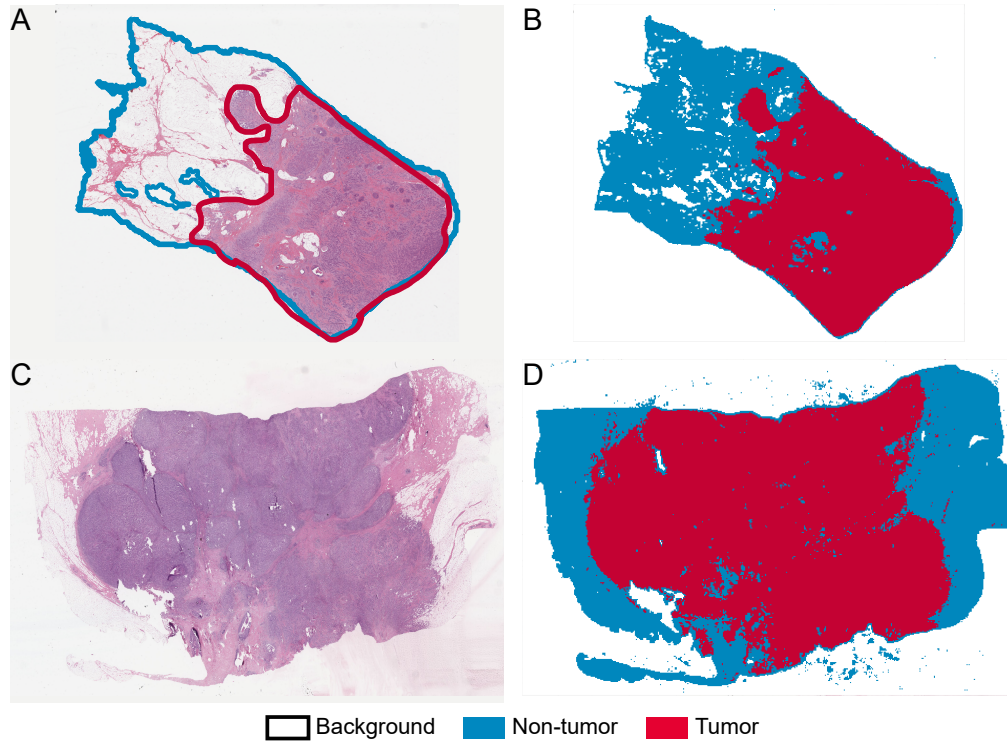
Fonte: Valieris et al. (2024)

Figura 9 – Esquematização do treinamento dos modelos para predição de **HER2**.

4.2.1 Filtragem dos fragmentos de interesse orienta o treinamento

Para processar o grande volume de imagens de maneira eficiente, nós implementamos um método para identificar somente as regiões tumorais de uma lâmina patológica (descrito na Seção 3.3.3), excluindo regiões em branco ou de tecidos não tumorais. Com essa etapa, reduzimos o espaço amostral de análise, sendo o tumor a região restrita para análise. Para isso, desenvolvemos um modelo para detectar regiões de interesse que, em sua totalidade, são exclusivas do tumor. A predição desse modelo é ilustrada na Figura 10. Notavelmente, este modelo também foi capaz de excluir regiões fora de foco ou que continham marcas de caneta marcador onde o patologista havia feito anotações sobre o vidro

da lâmina, ambos são artefatos presentes na coorte **TCGA-BRCA** e podem atrapalhar o treinamento do modelo classificador; estes fragmentos com artefatos foram identificados pelo modelo como *fundo* (*background*) e posteriormente descartados nos passos subsequentes.



Fonte: Valieris et al. (2024)

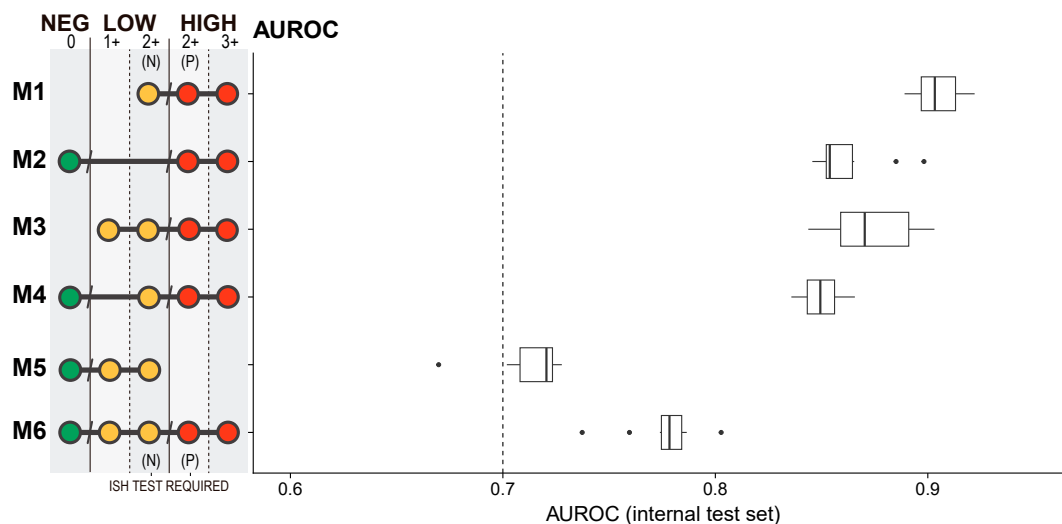
Figura 10 – Exemplo de inferência do modelo de detecção de Tumor e fundo. (A) Lâmina da coorte **TCGA-BRCA**, anotada com região tumoral (vermelho) e região não tumoral (azul). (B) Inferência do modelo na mesma lâmina. (C) Lâmina da coorte **AC-BRCA-HER2**, sem anotação disponível. (D) Exemplo de inferência do modelo na mesma lâmina.

4.2.2 Treinamento do Modelo

A partir do conjunto de dados **HEROHE** e **AC-BRCA-HER2**, nós separamos 15% dos dados para um conjunto de teste, e os dos outros 85% foram divididos em 10 conjuntos de *folds*, para utilização da técnica de validação cruzada, para avaliar a variância das predições do modelo treinado em conjuntos de dados diferentes. O conjunto de dados **TCGA-BRCA** foi utilizado somente

como um conjunto independente de teste.

Utilizando a arquitetura **CLAM** (Lu et al. 2021), e com características relevantes extraídas das regiões tumorais utilizando **UNI** (Chen et al. 2024), nós treinamos seis modelos com diferentes subconjuntos de dados, com o objetivo de avaliar o quão distinguível é cada subcategoria do **HER2**, a divisão dos dados utilizados nestes modelos está descrito na Seção 3.4 e sumarizado na Tabela 1. Após o treinamento, nós calculamos a **AUC** para cada *fold* utilizando o conjunto de dados de teste separado da validação cruzada, esses resultados são ilustrados na Figura 11. O código fonte utilizado para pré-processamento e treinamento desses modelos esta disponível em <https://github.com/tojallab/wsi-mil>.



Fonte: Valieris et al. (2024)

Figura 11 – AUC da validação cruzada para os modelos de *HER2*. Cada modelo foi treinado com um subconjunto de acordo com seu resultado do teste de *HER2*, cada ponto no *boxplot* representa um *fold* na validação cruzada.

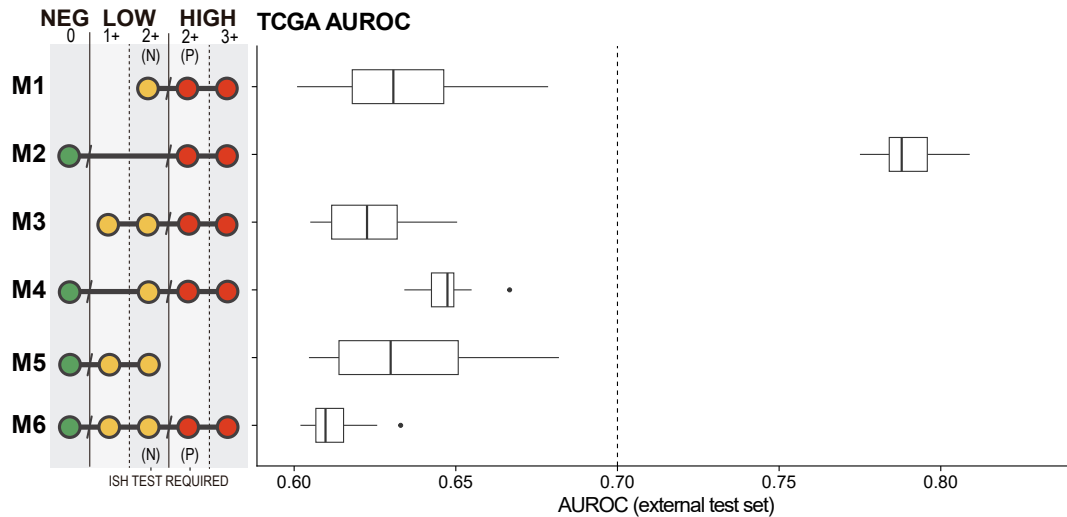
4.2.3 Predições do Modelo

Os resultados revelam que a classificação dos subgrupos de *HER2* varia significativamente de acordo com os subgrupos testados. O modelo **M1**, que busca distinguir somente entre **ISH** negativo e positivo, foi o modelo que obteve o melhor resultado dentre as configurações testadas, esse modelo é de particular

interesse pois seria um substituto ao teste de **ISH**, necessário quando o patologista interpreta o resultado do **IHC** como *2+ duvidoso*. No modelo **M2** nós avaliamos a capacidade do modelo de prever entre resultados negativos e intensos, ignorando o grupo intermediário *HER2-low*, esse modelo também obteve um resultado satisfatório, em linha com outros estudos publicados na literatura (Farahmand et al. 2022). Os modelos **M3** e **M4** estendem os modelos **M1** e **M2**, respectivamente, pela inclusão do **IHC** 1+ e **IHC** 2+ com **ISH** negativo, ambos apresentaram queda de desempenho, porém o resultado observado ainda é positivo.

No modelo **M5**, houve uma queda de desempenho mais significativa. Esse modelo busca distinguir os grupos *HER2-neg* e *HER2-low*. O modelo **M6**, que inclui os três subgrupos, também teve uma queda significativa de desempenho em comparação aos outros. Uma explicação para esses resultados seria a subjetividade na interpretação dos resultados de **IHC** que distinguem entre os grupos **IHC** 0 e **IHC** 1+. Estudos que compararam a concordância entre patologistas na interpretação desses testes reportam resultados semelhantes (Robbins et al. 2023). Esses resultados demonstram que uma limitação significativa desses modelos está na curadoria dos dados utilizados para seu treinamento (Wu et al. 2023). Nós também testamos estes modelos em um conjunto de dados independente **TCGA-BRCA**, e os resultados estão ilustrados na Figura 12.

Na coorte **TCGA-BRCA** nós observamos uma queda significativa no desempenho em comparação com os resultados do teste interno, o que indica que a generalização destes modelos para coortes diferentes ainda é um desafio. No entanto, é importante notar que esses dados utilizaram orientações distintas para a classificação do HER2. As orientações fornecidas pela *American Society of Clinical Oncology* (ASCO) estipulam regras para a interpretação dos testes patológicos e são continuamente atualizadas (Wolff et al. 2018). Notavelmente, o modelo **M2** manteve um desempenho satisfatório nesta coorte, esse modelo classifica apenas os dois grupos extremos *HER2-neg* e *HER2-high*, isso corrobora



Fonte: Valieris et al. (2024)

Figura 12 – AUC para os modelos de HER2 no **TCGA-BRCA**. Os modelos treinados foram testados no conjunto de dados independente.

Tabela 2 – Métricas de desempenho para cada modelo nos conjuntos de teste internos e externos.

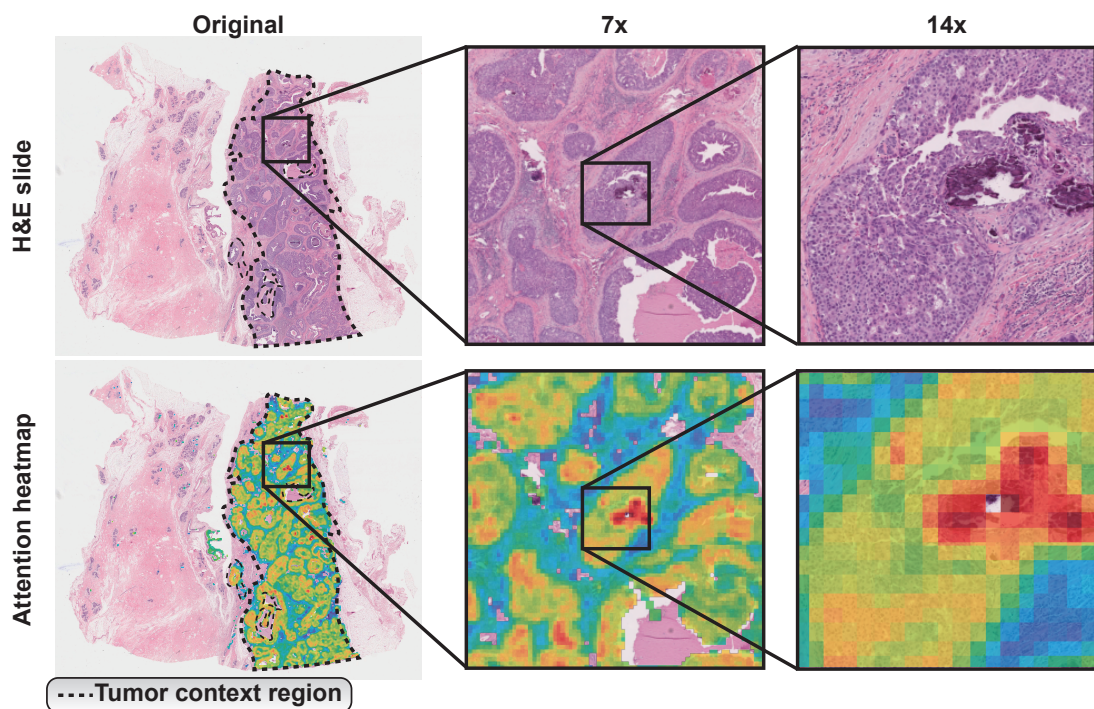
Conjunto	Modelo	AUC	Precisão	Revocação	F1-Score
Interno	M1	0.90 ± 0.01	0.82 ± 0.02	0.83 ± 0.02	0.83 ± 0.02
	M2	0.85 ± 0.02	0.77 ± 0.02	0.77 ± 0.02	0.77 ± 0.02
	M3	0.87 ± 0.02	0.81 ± 0.02	0.79 ± 0.04	0.80 ± 0.03
	M4	0.85 ± 0.01	0.70 ± 0.01	0.70 ± 0.01	0.70 ± 0.01
	M5	0.72 ± 0.02	0.65 ± 0.02	0.65 ± 0.01	0.65 ± 0.02
	M6	0.78 ± 0.02	0.57 ± 0.03	0.58 ± 0.03	0.57 ± 0.03
Externo	M1	0.63 ± 0.02	0.60 ± 0.02	0.60 ± 0.02	0.59 ± 0.03
	M2	0.79 ± 0.01	0.64 ± 0.02	0.68 ± 0.03	0.58 ± 0.04
	M3	0.62 ± 0.01	0.57 ± 0.02	0.58 ± 0.02	0.57 ± 0.02
	M4	0.65 ± 0.01	0.40 ± 0.04	0.48 ± 0.01	0.36 ± 0.02
	M5	0.63 ± 0.03	0.54 ± 0.01	0.60 ± 0.03	0.40 ± 0.02
	M6	0.61 ± 0.01	0.39 ± 0.02	0.44 ± 0.02	0.31 ± 0.02

a ideia de que a inclusão de dados no grupo intermediário *HER2-low* dificulta a classificação.

4.2.4 Análise de Regiões Acionáveis

Uma vez treinado, é possível utilizar as predições do modelo para entender quais características presentes na imagem levaram a um determinado resultado.

Para isso, nós examinamos a inferência do modelo na camada de atenção (descrita na Seção A.6.2) e ilustrada na Figura 13. Essa camada calcula escores de atenção que variam de 0 a 1 para cada fragmento da imagem. São esses escores que determinam o peso com que as características de cada fragmento são consideradas pela camada de classificação final do modelo. Portanto, entender quais padrões estão presentes em fragmentos com alto escore de atenção pode levar à descoberta de novas associações entre características morfológicas e moleculares.



Fonte: Valieris et al. (2024)

Figura 13 – Mapa de calor da camada de atenção. O mapa de calor destaca as regiões da imagem com maior escore da camada de atenção.

Com a ajuda de patologistas do hospital A. C. Camargo, nós examinamos os fragmentos com maior valor na camada de atenção em busca destes padrões morfológicos. Observamos que os fragmentos de lâminas **HER2-high** com alto valor de atenção apresentam estroma esclerótico com pouco infiltrado inflamatório, células neoplásicas isoladas, neoplasia de padrão coesivo com infiltração lobular de graus 1/2 e padrões estruturais semelhantes a características micropapilares. Em **HER2-neg**, apresentam grau nuclear 3 e estão enriquecidos por cé-

lulas adiposas. Uma possível interpretação, é que o microambiente adiposo pode estar associado a motilidade e a invasividade das células tumorais (D’Esposito et al. 2016). Em **HER2-low**, os fragmentos apresentaram células inflamatórias crônicas, tecido fibrótico e graus nucleares 1 e 2.

4.2.5 Comparação de implementações alternativas de modelo

Nós investigamos implementações alternativas à metodologia apresentada e comparamos seus resultados. Nós testamos se a remoção de fragmentos da imagem com tecidos não tumorais afeta a classificação final do modelo de forma significativa, comparando a utilização de todos os fragmentos de uma imagem exceto os em branco (**NOBG**) com a utilização de somente os com tumor (**TUMO**). As **AUCs** calculadas são apresentadas na Tabela 3.

Tabela 3 – Média e desvio padrão das **AUCs** de cada *fold* do teste interno, comparando a utilização de todos os fragmentos exceto fundo (**NOBG**) e utilização de apenas fragmentos com tumor (**TUMO**). O melhor desempenho em cada modelo foi destacado.

	NOBG	TUMO
M1	0.88 ± 0.02	0.90 ± 0.01
M2	0.87 ± 0.02	0.85 ± 0.02
M3	0.91 ± 0.02	0.87 ± 0.02
M4	0.85 ± 0.01	0.85 ± 0.01
M5	0.69 ± 0.04	0.72 ± 0.02
M6	0.78 ± 0.01	0.78 ± 0.02

Os resultados indicam que não houve alteração significativa no desempenho dos modelos após a remoção dos fragmentos sem tumor. Isso é um resultado interessante, pois significa que podemos focar o treinamento do modelo somente em áreas tumorais, reduzindo a possibilidade de incorporar padrões espúrios na predição e também diminuindo o tempo de treinamento destes modelos, considerando que fragmentos sem tumor compõem uma porcentagem significativa de uma lâmina patológica.

Nós também comparamos diferentes arquiteturas **MIL** no treinamento do modelo, além do **CLAM** (Lu et al. 2021), nós testamos o **ADMIL** (Ilse et al.

2018) e **TransMIL** (Shao et al. 2021). As **AUCs** calculadas são apresentadas na Tabela 4.

Tabela 4 – Média e desvio padrão das **AUCs** de cada *fold* do teste interno, comparando três métodos diferentes para implementação do modelo de predição. O melhor desempenho em cada modelo foi destacado.

	ADMIL	CLAM	TransMIL
M1	0.87 ± 0.02	0.90 ± 0.01	0.88 ± 0.03
M2	0.85 ± 0.02	0.85 ± 0.02	0.85 ± 0.02
M3	0.84 ± 0.02	0.87 ± 0.02	0.88 ± 0.04
M4	<i>N/A</i>	0.85 ± 0.01	<i>N/A</i>
M5	0.73 ± 0.05	0.72 ± 0.02	0.71 ± 0.02
M6	<i>N/A</i>	0.78 ± 0.02	<i>N/A</i>

Os resultados mostram que **CLAM** e **TransMIL** tiveram desempenho superior ao **ADMIL**. No entanto não foi possível calcular **AUCs** para os modelos **M4** e **M6** devido ao fato que os modelos **ADMIL** e **TransMIL** não suportam predições não binárias.

4.3 PREDIÇÃO DO ESCORE AGREGADO DE HRD

A classificação de **HRD** realizada previamente (Seção 4.1), foi baseada na medida da assinatura mutacional **SBS3**, que necessita estabelecer previamente um ponto de corte que determina se um valor é considerado **HRD** negativo ou positivo, porém não existe um consenso na literatura de qual seria o melhor ponto para esse corte. A assinatura **SBS3** também não é a única medida descrita na literatura capaz de quantificar o **HRD**, existem outras medidas relevantes que podem atuar de maneira complementar. Para que os modelos treinados a partir de dados tenham o melhor desempenho possível, é fundamental que o conjunto de dados utilizado seja anotado de forma rigorosa, coerente e o mais próximo possível da verdade e que reflita as alterações biológicas relacionadas a **HRD**. Por isso, decidimos ampliar nosso esforço para construção de um modelo para detectar **HRD** a partir de imagens H&E. Assim, consideramos utilizar um valor

contínuo para o *status* de **HRD**, derivado da combinação de quatro marcadores de **HRD** eliminando, assim, a necessidade de estabelecer um ponto de corte.

O estudo contendo essa análise foi enviado para publicação na revista *Scientific Reports* em fevereiro de 2025 e, após avaliação do editor, encontra-se em revisão por pares. Uma versão pré-publicação encontra-se anexada a este documento no Apêndice E.

4.3.1 Preditor Agregado de HRD

Nós desenvolvemos um método para combinar escores de **HRD** de várias origens em uma única medida, isso permite capturar alterações relevantes ao **HRD** de vários tipos diferentes, desde mutações pontuais a cicatrizes genômicas. Ao todo, foram combinados quatro marcadores de **HRD**, conforme descrito abaixo:

- **SBS3**: caracteriza mutações somáticas agrupadas pelo contexto e troca de um nucleotídeo (SBS, *single base substitution*);
- **ID6**: parte do conjunto de assinaturas ID (*small insertions and deletions*), que caracteriza nucleotídeos inseridos ou deletados (*indels*) e o tamanho do evento;
- **CN1** e **CN17**, parte do conjunto de assinaturas CN (*copy number variations*) que categoriza eventos de alteração no número de cópias por perda de heterozigose, número de cópias e tamanho do segmento alterado, e **CN1**, também parte do conjunto CN, é associado a amostras com baixo ou nenhum evento de alteração, porém apresenta correlação negativa com CN17 e outros escores HRD (Steele et al. 2022).
- **scarHRD**: O escore **scarHRD** é uma medida não ponderada e avalia a extensão do número de cópias, a perda de heterozigose, e as transições em

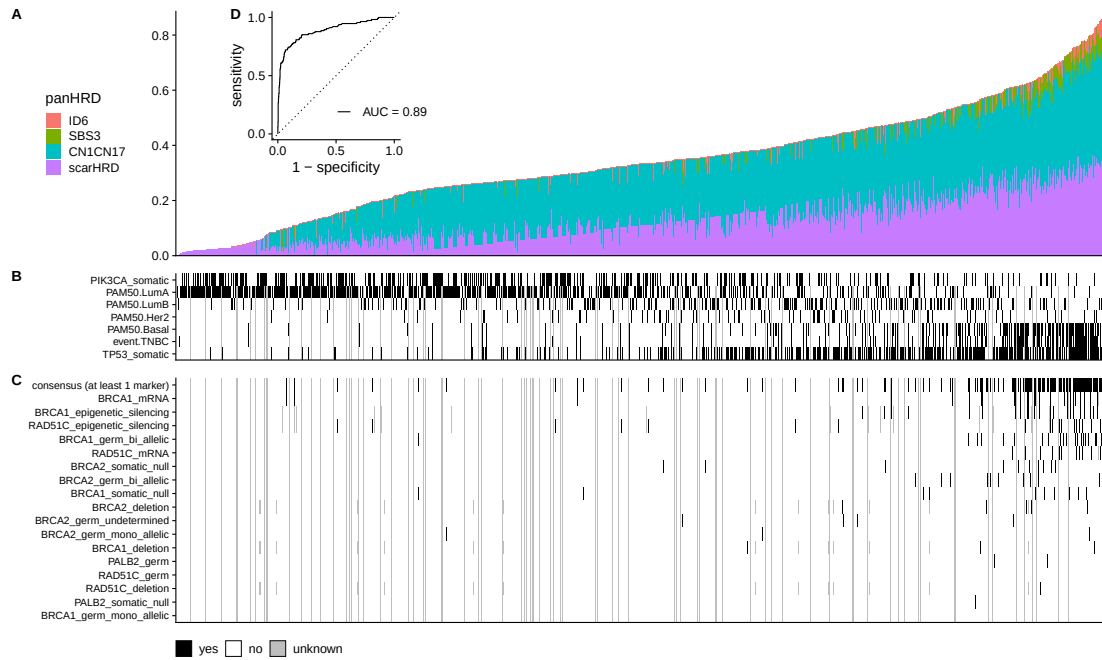
larga escala, características a nível genômico associadas a variações estruturais no genoma ligadas ao HRD. (Sztupinski et al. 2018).

De importância, esses marcadores foram combinados e ponderados de acordo com sua prevalência dentro de seus respectivos tipos tumorais. A descrição rigorosa do cálculo desse escore está descrita no Apêndice B.

4.3.2 Validação do Escore Agregado

Para avaliar a eficácia do escore agregado proposto no presente trabalho, na qual chamamos de **panHRD**, nós o testamos em amostras de tumores de mama da coorte **TCGA-BRCA**, e comparamos o escore com marcadores conhecidos associados ao **HRD**. A Figura 14A apresenta a distribuição dos valores do **panHRD** nas amostras (dispostas em colunas) da coorte do **TCGA-BRCA**. Adicionalmente, para cada amostra, incluímos a anotação dos seus respectivos subtipos (Figura 14B) de acordo com o perfil molecular PAM50 (Parker et al. 2009).

Observamos que os subtipos basal e TNBC (*triple-negative breast cancer*) são mais frequentes em pacientes com maior escore **panHRD**, além da presença de mutações somáticas no gene *TP53* nas respectivas amostras, enquanto amostras com baixo escore estão mais frequentemente associadas aos subtipos luminais. Também comparamos o escore com a presença de lesões somáticas e germinativas em genes associados a **HRD** (Figura 14C). Essa análise mostra que as amostras com alto escore de **panHRD** tendem a apresentar pelo menos uma lesão em genes conhecidos. Finalmente, nós estimamos a capacidade preditiva do **panHRD** em detectar amostras com pelo menos uma lesão e a capacidade discriminatória foi medida por meio da métrica **AUC**, resultando em uma AUC de 0,89 (Figura 14D). Podemos concluir que o escore **panHRD** é um biomarcador robusto capaz de recapitular um espectro amplo de alterações em genes conhecidos associados a **HRD**, além de estar presente em subtipos enriquecidos com esse fenótipo.



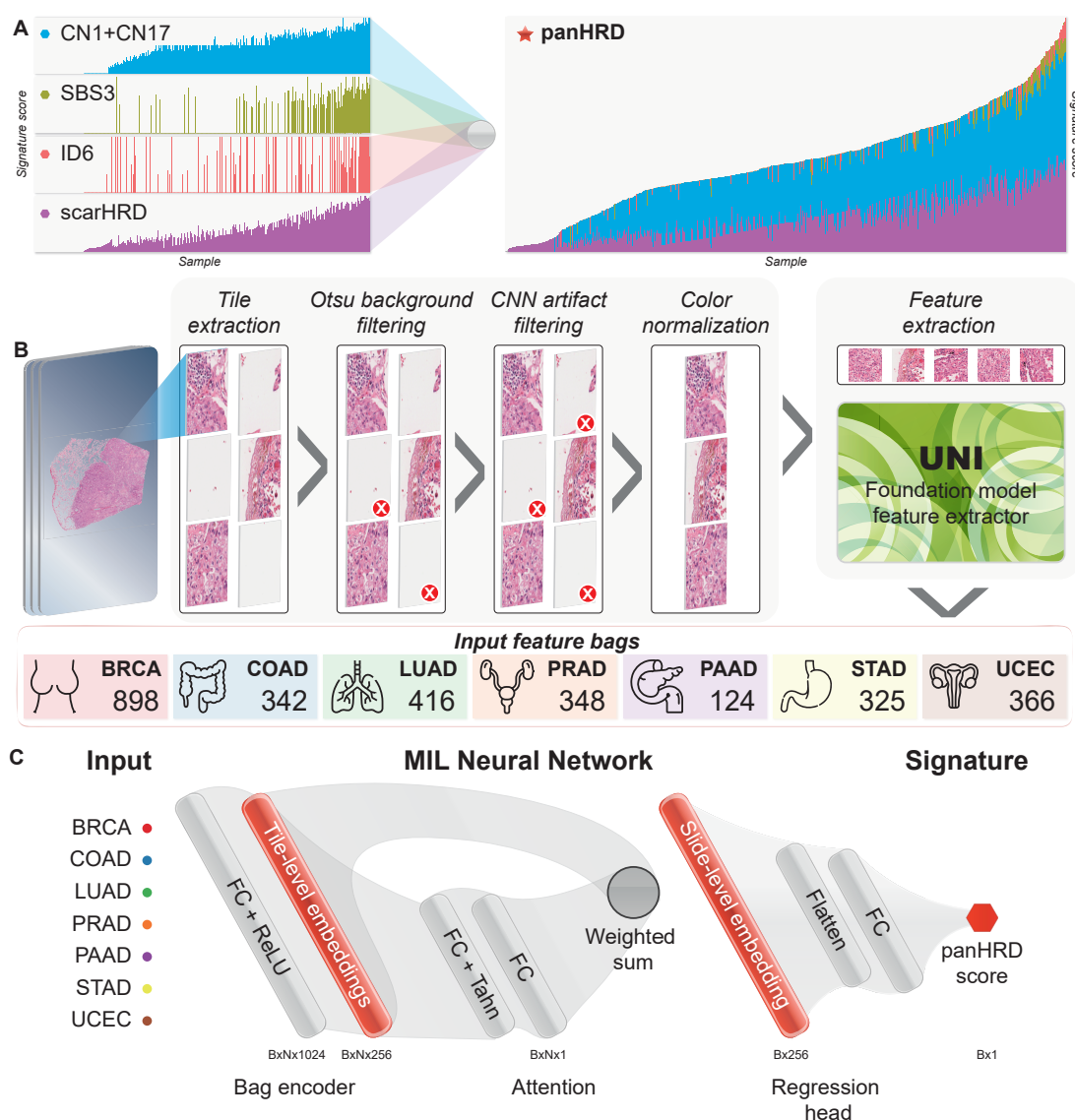
Fonte: Autoria própria.

Figura 14 – Comparação entre marcadores genéticos associados ao HRD e o escore agregado. (A) Amostras tumorais da coorte **TCGA-BRCA** ordenadas pelo escore agregado. (B) Anotações de subtipos relevantes. (C). Anotações de alterações em genes conhecidos relacionados a via **HR**. (D) Curva ROC do escore agregado como um preditor do consenso de (C).

4.3.3 Treinamento do Modelo

Após avaliar a capacidade discriminatória de recuperar eventos conhecidos associados ao fenótipo de **HRD**, nós utilizamos essa medida para desenvolver um modelo fracamente supervisionado de predição de **HRD** a partir das imagens histopatológicas de H&E, obtidas de sete tipos de tumores (Figura 15). Para preservar as nuances do escore agregado **panHRD**, desenvolvemos um modelo de regressão que prediz valores contínuos, em vez de uma classificação em categorias discretas.

O pré-processamento consiste em dividir uma lâmina H&E em fragmentos disjuntos, remover fragmentos com artefatos ou sem tecido, normalizar a cor e extrair um vetor de características (Figura 15B). Nós utilizamos o **UNI**, um modelo auto-supervisionado para patologia treinado em mais de 100.000 H&E



Fonte: Autoria própria.

Figura 15 – Esquematização do treinamento dos modelos para predição do escore agregado de HRD. (A) Combinação de quatro biomarcadores HRD em um escore agregado. (B) Pré-processamento de lâminas histopatológicas e extração de características com o UNI. (C) Treinamento do modelo de regressão MIL com sete tipos de tumores.

(Chen et al. 2024), para extrair um vetor de características de cada pedaço da imagem H&E para, então, treinar nosso modelo. A arquitetura do modelo, utiliza a técnica **MIL** com uma camada de atenção, que aprende a dar pesos distintos para fragmentos importantes para a predição, e uma camada final de agregação que combina o resultado de vários fragmentos em um escore final que varia de 0 a 1 (Figura 15C). O código fonte utilizado no pré-processamento e treinamento desses modelos está disponível em <https://github.com/TojalLab/reg-mil>.

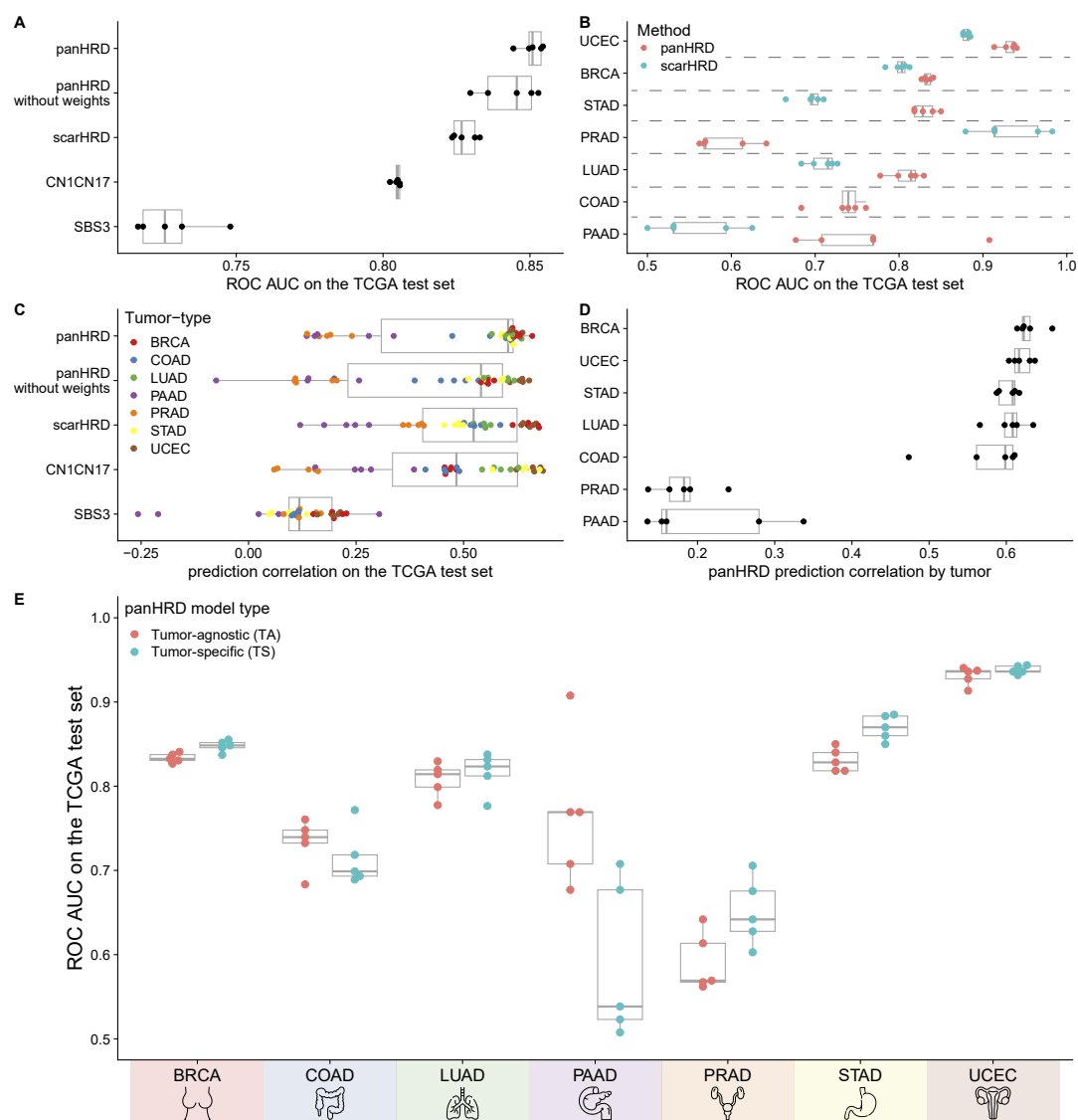
4.3.4 Predição do escore de HRD em diversos tumores

Para entender como as características distintas de cada tumor afetam a predição de **HRD**, nós treinamos um modelo agnóstico (**TA**, *Tumor-Agnostic*), utilizando dados de sete tipos de tumores diferentes combinados em um único conjunto de dados, e treinamos sete modelos diferentes (**TS**, *Tumor-specific*), cada um utilizando apenas dados do respectivo tipo de tumor.

Os dados foram divididos em conjuntos de treino (85%) e teste (15%), e o conjunto de treino foi subsequentemente separado em cinco *folds* utilizando validação cruzada, cada *fold* contendo um subconjunto de treino e validação distinto. Nós treinamos os modelos em cada *fold* para observar o impacto da seleção de amostras no treinamento e predição dos modelos. Após o treino, foi calculado **AUC** utilizando o conjunto de teste.

Como resultado, podemos observar que o modelo **TA** foi capaz de alcançar um desempenho similar comparado aos modelos **TS** (Figura 16E). A comparação estatística entre as médias das **AUCs** obtidas a partir dos modelos **TA** e **TS**, para cada tumor, revelou que não há diferença significativa. Concluimos que a predição de **HRD** não está limitada a características específicas de apenas um tipo de tumor e, portanto, o modelo **TA** é adequado para prever **HRD** em sete tumores distintos.

Nós comparamos modelos treinados com o **panHRD** com modelos treinados com seus componentes individuais, para avaliar a diferença no desempenho



Fonte: Autoria própria.

Figura 16 – Métricas de desempenho dos modelos do HRD agregado. (A) **AUC** do conjunto de dados combinado. (B) **AUC** estratificada por tipos de tumor. (C) Correlação entre previsões do modelo e o escore agregado alvo. (D) Correlação do panHRD estratificado por tipos de tumores. (E) Comparação da **AUC** entre os modelos *tumor-agnostic* (TA) e *tumor-specific* (TS).

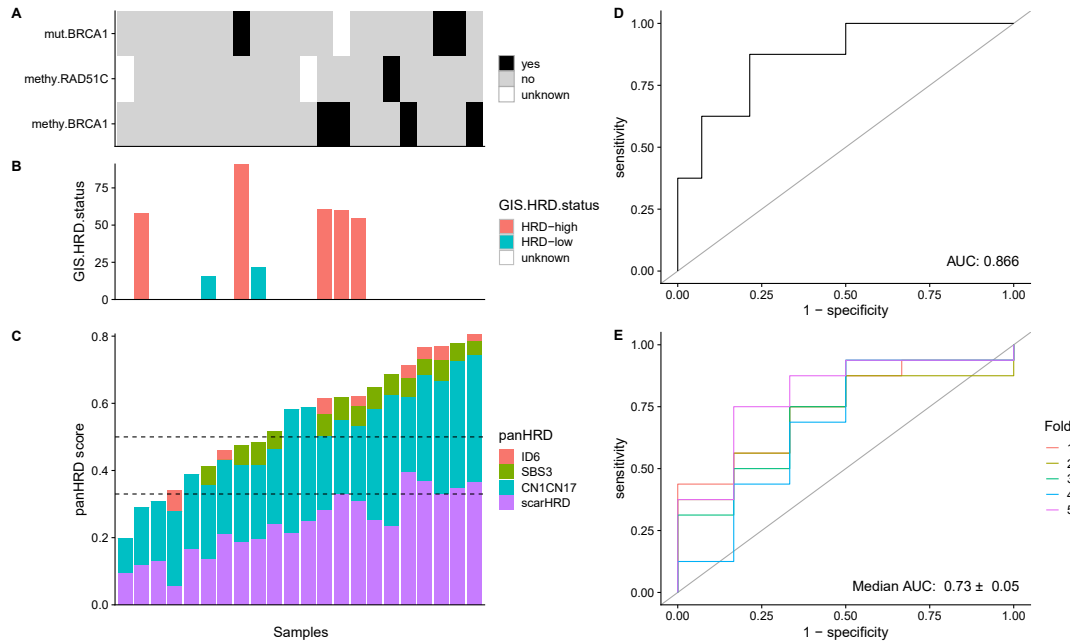
alcançado na agregação desses marcadores. As Figuras 16A e 16C apresentam as métricas de **AUC** e correlação desses modelos. Observamos que o **panHRD** teve um desempenho melhor que os seus componentes, também comparamos o **panHRD** com e sem o cálculo dos pesos por tipo de tumor e notamos que a adição desses pesos resulta em um ganho de desempenho além da redução da variância observada nas métricas calculadas.

Nós comparamos o desempenho do **panHRD** com o modelo treinado com **scarHRD** nos diferentes tipos de tumores testados (Figura 16B), e observamos que o **panHRD** superou o **scarHRD** na maioria dos tumores, exceto em **PRAD** (próstata). Na Figura 16D a correlação por tumor do **panHRD** indica que o modelo teve desempenho satisfatório na maioria dos tumores testados exceto em **PRAD** (próstata) e **PAAD** (pâncreas).

4.3.5 Análise de coorte com triplo-negativo

Nós utilizamos uma coorte independente de 22 amostras de câncer de mama com subtipo triplo-negativo (**AC-BRCA-HRD-2**), para examinar os padrões moleculares associados ao **HRD** e o escore agregado **panHRD**. Utilizando sequenciamento **NGS** de exomas e painéis em genes conhecidos para **HRD**, nós calculamos o escore agregado nessas amostras e o comparamos com testes de **HRD** praticados na clínica, os resultados estão ilustrados na Figura 17.

Nós observamos que amostras com alto escore (Figura 17C) tendem a apresentar mutações e/ou silenciamento epigenético dos genes *BRCA1* e *RAD51C* (Figura 17A). O teste genético *TrueSight Oncology 500 HRD*, que avalia alterações genéticas em 500 genes associados a **HRD**, **TMB** e **MSI**, foi realizado em 7 amostras resultando no escore ilustrado na Figura 17B. Interessantemente, nós observamos que as alterações gênicas foram mutualmente exclusivas, o que pode indicar que mecanismos alternativos para instabilidade genômica podem estar atuando neste subtipo. Nós calculamos o escore agregado **panHRD** nessas amostras e destacamos os cortes em 0.25 e 0.5, derivados da análise com o



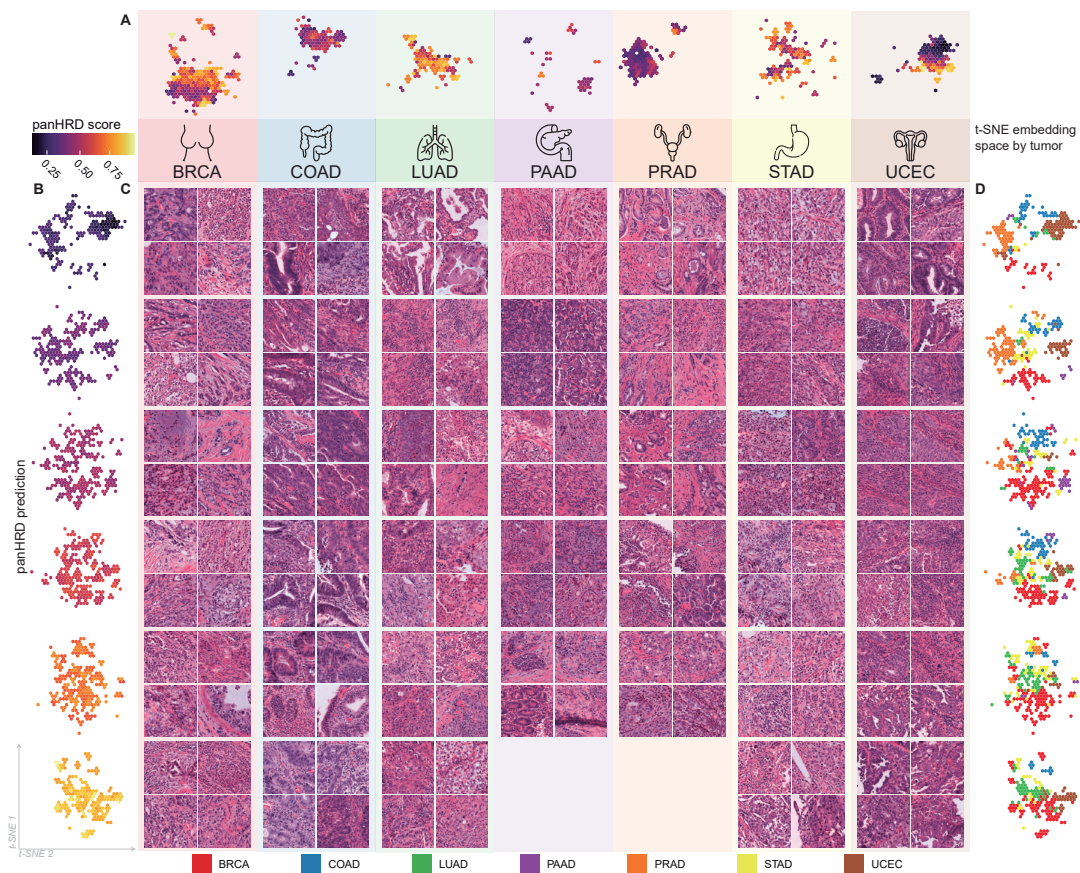
Fonte: Autoria própria.

Figura 17 – Predição do escore HRD na coorte AC-BRCA-HRD-2. (A) Alterações germinativas em genes **HR** conhecidos. (B) Nível de HRD determinado por teste genético *TSO 500 HRD*, amostras não testadas estão marcadas em branco. (C) escore **panHRD** calculado a partir dos marcadores de assinatura mutacionais. (D) Curva ROC das das predições do panHRD em dados genéticos. (E) Curva ROC das predições do panHRD a partir de imagens.

conjunto de dados combinado e somente tumores de mama respectivamente (Figura 17C), para ilustrar como uma predição poderia ser feita a partir do escore. Finalmente, nos calculamos a **AUC** para a predição dos marcadores observados utilizando o escore agregado e obtivemos uma **AUC** de 0,86 (Figura 17D), e em seguida, utilizamos o modelo agnóstico (**TA**), treinado nos dados do **TCGA**, para prever o escore a partir de suas lâminas histopatológicas, obtendo uma **AUC** mediana de 0,73 (Figura 17E). Estes resultados sugerem o potencial de uso deste método na clínica, apesar de ainda ser necessário realizar estudos com conjuntos de dados maiores para confirmar estas observações.

4.3.6 Análise de Regiões Acionáveis

Utilizando os escores aprendidos na camada de atenção do modelo, nós selecionamos os fragmentos com maior atenção de cada lâmina do conjunto de teste e projetamos o vetor de características extraído destes fragmentos em duas dimensões utilizando o método redução de dimensionalidade **t-SNE** (Figura 18).



Fonte: Autoria própria.

Figura 18 – Exemplos de imagens com alta atenção ao **panHRD**. (A) Projeção t-SNE dos pedaços com maior atenção estratificados por tipo de tumor. (B) Projeções estratificadas nível de escore **panHRD**. (C) Exemplos de imagens com alta atenção em diferentes tipos de tumores e níveis de escore. (D) Projeções coloridas por tipo de tumor e estratificadas pelo escore **panHRD**.

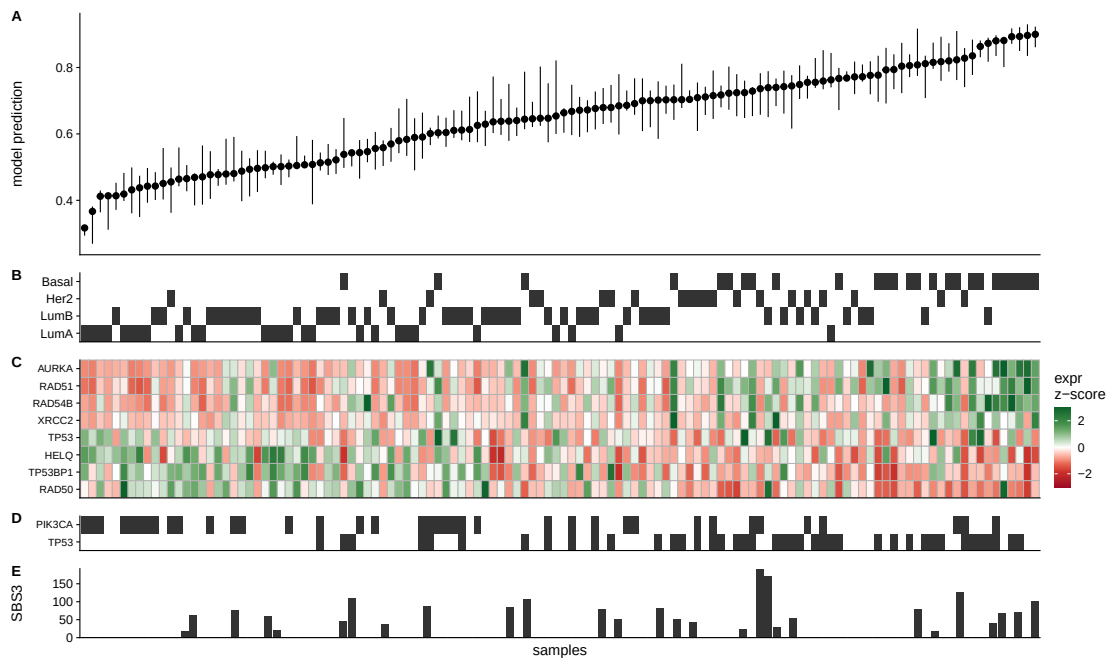
Nós estratificamos as regiões acionáveis por tipo de tumor (Figura 18A) e nível de escore (Figura 18B), e selecionamos fragmentos representativos de cada grupo para ilustrar os padrões presentes (Figura 18C). A inspeção visual desses padrões por patologistas identificou baixo grau histológico em fragmentos asso-

ciados ao baixo nível de **panHRD**, atipia nuclear leve a moderada, padrões de infiltração sólida, tubular ou lobular, desmoplasia moderada e infiltração inflamatória leve. Em contraste, casos com alto nível de **panHRD** apresentaram padrões de carcinomas de alto grau, atipia nuclear pronunciada, aumento da atividade mitótica, desmoplasia pronunciada e infiltrado inflamatório acentuado, sugerindo uma correlação entre o nível de **HRD** alto e uma resposta imune mais forte. Essas observações destacam a função do microambiente tumoral na manifestação do **HRD** e suportam o uso de características morfológicas na avaliação do nível de **HRD**.

4.3.7 Validação da predição do escore agregado no CPTAC

Nós investigamos a predição do escore agregado na coorte independente **CPTAC-BRCA**, que consiste em dados de lâminas histopatológicas, mutações somáticas e expressão gênica. Nós aplicamos o modelo **TA** (*tumor-agnostic*), treinado na coorte **TCGA**, nas lâminas do **CPTAC** para inferência do **panHRD** (Figura 19A). Este conjunto de dados não tem análise de número de cópias disponível, e por isso, não foi possível calcular os biomarcadores **CN** e **scarHRD**. Porém, nós ilustramos outros marcadores moleculares correlacionados ao escore agregado como a classificação PAM50 (Figura 19B), a presença de mutações somáticas no TP53 (Figura 19C), e níveis de expressão gênica de genes da via de recombinação homóloga (Figura 19C).

Podemos observar que amostras com alto escore tendem a ser do subtipo *basal*, enquanto amostras com escore baixo são de subtipos luminais (Figura 19B). Mutações somáticas no gene *TP53* tendem a estar presentes em amostras com alto escore, enquanto mutações somáticas no *PIK3CA* tendem a estar mais presentes em amostras com baixo escore e de subtipos luminais (Chen et al. 2018; Reinhardt et al. 2022) (Figura 19D). Nós observamos também que os padrões de expressão gênica apresentados por estas amostras (Figura 19C) corroboram os subtipos identificados, com amostras de alto escore apresentando expressão alta



Fonte: Autoria própria.

Figura 19 – Predição do escore HRD na coorte CPTAC. (A) Amostras ordenadas pelo escore predito. Os pontos indicam a mediana das predições dos 5 *folds* e as linhas indicam o valor mínimo e máximo observado para cada amostra. (B) Classificação PAM50 para cada amostra de acordo com dados de expressão gênica. (C) Níveis de expressão gênica para alguns genes de reparo. (D) Mutações somáticas de genes selecionados. (E) Exposição da assinatura mutacional **SBS3**.

nos genes *AURKA*, *RAD51*, *RAD54B* e *XRCC2*. Esses genes são componentes da via de reparo de DNA por recombinação homóloga e sua expressão elevada sugere resposta ao elevado nível de defeitos reparados por essa via (Serra et al. 2013). Por outro lado, essas amostras apresentam baixa expressão de *HELQ*, *TP53BP1*, e *RAD50*, genes que participam da regulação da resposta celular a quebras de fita dupla (Mirza-Aghazadeh-Attari et al. 2019). Nós também comparamos o escore agregado à assinatura mutacional **SBS3** (Figura 19E), porém não foi observado alto grau de concordância, o que indica que a **SBS3** por si só não seria suficiente para a predição do **HRD**. Essa análise demonstra que o escore agregado foi capaz de organizar as amostras pela intensidade do **HRD** assim como observado na coorte **TCGA-BRCA** (Figura 14).

4.4 DISCUSSÃO

A medicina de precisão requer a realização de testes genéticos capazes de caracterizar os aspectos moleculares e, assim, garantir um tratamento mais personalizado com base nessas características genéticas. Embora promissores, a avaliação dos biomarcadores genéticos não é uma realidade na maioria dos hospitais, uma vez que requer uma infraestrutura e equipe especializadas, além da disponibilidade de material biológico e nem sempre é custo-efetivo. Nesse cenário, a predição de biomarcadores a partir de imagens histopatológicas, disponíveis ao diagnóstico, representa uma oportunidade para ampliar o acesso às opções de terapias-alvo ou sistêmicas.

De fato, a vantagem principal dessa abordagem reside na disponibilidade das lâminas histológicas, uma vez que são obtidas na rotina clínica. Consequentemente, sua implementação na prática clínica não geraria um trabalho adicional por parte do patologista e, nos casos onde a patologia digital já é utilizada, seria possível re-aproveitar as lâminas já digitalizadas para a realização de predições dos biomarcadores de interesse. Como resultado, a avaliação de uma série de biomarcadores, de forma automatizada, tem o potencial de direcionar tratamentos mais personalizados de maneira rápida e em larga escala.

Embora o desenvolvimento dos modelos demande recursos computacionais e uma grande quantidade de imagens, ao final, as predições em novas lâminas são realizadas de forma rápida e custo-efetiva. Não obstante, apesar de medidas de desempenho demonstrem sua efetividade e capacidade discriminatória, também é necessária a realização de ensaios clínicos prospectivos a fim de validar os modelos no ambiente com dados do mundo real. Em paralelo, outros desafios relacionados à adesão e incorporação dessa tecnologia requerem o apoio da equipe médica, incluindo patologistas e clínicos.

De importância, apesar dos modelos permitirem predições em qualquer lâmina H&E, sua efetividade depende do domínio no qual foi treinado. Por

exemplo, se o modelo foi treinado somente em lâminas de tumores de mama, sua performance será comprometida em outros tumores. Outras variáveis que podem afetar sua performance estão relacionadas ao tipo de material utilizado, incluindo biópsias, amostras de resecção, tumores primários ou metástase. Adicionalmente, modelos treinados em lâminas obtidas de tumores ao diagnóstico, sem tratamento, não podem apresentar uma boa performance nas lâminas de tumores obtidas após tratamento, terapias neo-adjuvantes. Outros aspectos envolvem vieses durante o treinamento, ao considerar características que podem ser sub-representadas nos pacientes, incluindo demográficas, socioeconômicas e de etnia. Finalmente, outros aspectos a serem considerados são os parâmetros técnicos considerados no processo de preparo e digitalização das imagens.

4.5 TRABALHOS FUTUROS

4.5.1 Predição do HER2 usando o IHQ

Embora os modelos apresentados tenham sido desenvolvidos com imagens digitalizadas de lâminas coradas com H&E, predições a partir de imagens de lâminas obtidas de ensaios de imuno-histoquímica são promissores. Por exemplo, modelos capazes de auxiliar o patologista na interpretação dos níveis de HER2, especialmente o HER2*low*. Isso é de particular interesse pois o algoritmo permite interpretar a imagem de maneira mais eficiente auxiliando, assim, vários patologistas.

4.5.2 Validação das Regiões Acionáveis

Investigar se os padrões identificados durante a detecção dos biomarcadores são consistentes em novas lâminas com o mesmo genótipo

4.5.3 Predição da amplificação do HER2

Desenvolver modelos para avaliar a amplificação de HER2 considerando no treinamento somente lâminas onde o laudo foi gerado a partir do teste de ISH/DDISH. Esse modelo é de particular interesse uma vez que esse teste molecular é menos subjetivo comparado à imuno-histoquímica, além de ser um teste mais caro.

4.5.4 Validar o HRD em outros tumores

Ainda é necessário mais estudos em outros tumores para validar o escore agregado e avaliar seu potencial uso no apoio a decisão da escolha de terapia-alvo.

4.6 OUTROS TRABALHOS REALIZADOS

Ao longo da execução desse trabalho, também participei como autor principal e co-autor de outros estudos em temas de interesse no laboratório de Bioinformática e Biologia Computacional. Especificamente, minhas contribuições foram no planejamento e codificação dos *scripts* em *python*:

2021

- Mutational Signatures Driven by Epigenetic Determinants Enable the Stratification of Patients with Gastric Cancer for Therapeutic Intervention (Buttura et al. 2021)

2022

- Family-based whole-exome sequencing identifies rare variants potentially related to cutaneous melanoma predisposition in Brazilian melanoma-prone families (Fidalgo et al. 2022)
- A mixture model for determining SARS-Cov-2 variant composition in pooled samples (Valieris et al. 2022)

- Somatic hypermutation patterns in immunoglobulin variable regions are established independently of the local transcriptional landscape (Schoeberl et al. 2022)

2023

- WSI2ML – An Open-Source Whole Slide Image Annotation Software for Machine Learning Applications (Martins et al. 2023)
- Relating mutational signature exposures to clinical data in cancers via *sig-neR* 2.0 (Drummond et al. 2023)

2024

- Inflammatory Monocytes Increase Prior to Detectable HIV-1 Rebound Viremia (Farrell-Sherman et al. 2024)
- Tracking SARS-CoV-2 variants of concern in wastewater: an assessment of nine computational tools using simulated genomic data (Sutcliffe et al. 2024)

5 CONCLUSÕES

Neste trabalho, implementamos métodos de aprendizado profundo para identificar biomarcadores comumente investigados na rotina clínica em pacientes com câncer de mama. Exploramos a predição de **HRD** a partir de imagens histopatológicas e obtivemos resultados promissores. Nossos achados revelam que os métodos modernos baseados em **IA** são capazes de detectar características em imagens H&E que refletem alterações genéticas que, por sua vez, são ocultas aos olhos de um patologista experiente. Adicionalmente, investigamos a predição de **HER2** negativo, baixo (*HER2-low*) e positivo. Nesse estudo, os resultados revelam que a curadoria dos dados utilizados no treinamento desses modelos é essencial para a capacidade do modelo de prever um marcador. Finalmente, propomos um escore agregado para o *status* de **HRD**, representando as variáveis alvos de forma quantitativa e permitindo a predição de escores, ao invés de classificação de grupos. A derivação do escore a partir de dados genômicos evita a subjetividade de uma anotação humana e modela o problema de forma mais pragmática. Analisamos também os padrões presentes em regiões importantes apontadas pelo modelo e destacamos padrões morfológicos que poderiam ser identificados por um patologista em uma inspeção visual. Por fim, investigamos a predição do **HRD** em vários tumores e observamos que seria possível generalizar o modelo agnóstico de **HRD** para mais tumores, além daqueles realizados atualmente na prática clínica.

Apesar dos modelos treinados apresentarem desempenho satisfatório, a generalização desses modelos para outras coortes ainda é um desafio contínuo e de intensa investigação visando, ao final, predições de lâminas de forma automatizada e custo-efetiva.

6 REFERÊNCIAS

Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *British Journal of Cancer*. 2012 out.; 107:1776–1782.

Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, et al. Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology association. *Journal of Pathology Informatics*. 2019 jan.; 10:9.

Ahn S, Woo JW, Lee K, Park SY. Her2 status in breast cancer: changes in guidelines and complicating factors for interpretation. *Journal of Pathology and Translational Medicine*. 2020 jan.; 54:34–44.

Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020 fev.; 578: 94–101.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature*. 2013 ago.; 500: 415–421.

Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Current Opinion in Genetics & Development*. 2014 fev.; 24:52–60.

Babenko B. Multiple instance learning: algorithms and applications. View Article PubMed/NCBI Google Scholar. 2008; 19:.

Ballman KV. Biomarker: Predictive or prognostic? *Journal of Clinical Oncology*. 2015 nov.; 33:3968–3971.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 1995 jan.; 57:289–300.

Bergstrom EN, Abbasi A, Díaz-Gay M, Galland L, Ladoire S, Lippman SM, et al. Deep learning artificial intelligence predicts homologous recombination deficiency and platinum response from histologic slides. *J Clin Oncol*. 2024 jul.;; JCO2302641.

Berner J, Grohs P, Kutyniok G, Petersen P. The modern mathematics of deep learning. In: _____. *Mathematical Aspects of Deep Learning*. Cambridge University Press, 2022. 1–111. ISBN 9781316516782. Disponível em: <http://dx.doi.org/10.1017/9781009025096.002>.

Birkbak NJ, Wang ZC, Kim JY, Eklund AC, Li Q, Tian R, et al. Telomeric allelic imbalance indicates defective dna repair and sensitivity to dna-damaging agents. *Cancer Discovery*. 2012 abr.; 2:366–375.

Buttura JR, Provisor Santos MN, Valieris R, Drummond RD, Defelicibus A, Lima JP, et al. Mutational signatures driven by epigenetic determinants enable the stratification of patients with gastric cancer for therapeutic intervention. *Cancers*. 2021 jan.; 13:490.

Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*. 2019 jul.; 25:1301–1309.

Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The cancer genome atlas Pan-Cancer analysis project. *Nat Genet*. 2013 out.; 45:1113–1120.

for Research on Cancer IA. Global cancer observatory. [Online; acessado em 23-Nov-2024]. 2022. Disponível em: <https://gco.iarc.fr/>.

Chen L, Yang L, Yao L, Kuang XY, Zuo WJ, Li S, et al. Characterization of pik3ca and pik3r1 somatic mutations in chinese breast cancer patients. *Nature Communications*. 2018 abr.; 9:.

Chen RJ, Ding T, Lu MY, Williamson DFK, Jaume G, Song AH, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*. 2024 mar.; 30:850–862.

Conde-Sousa E, Vale J, Feng M, Xu K, Wang Y, Della Mea V, et al. Herohe challenge: Predicting her2 status in breast cancer from hematoxylin–eosin whole-slide imaging. *Journal of Imaging*. 2022 jul.; 8:213.

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*. 2018; 4:30.

Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. Hrdetect is a predictor of brca1 and brca2 deficiency based on mutational signatures. *Nature Medicine*. 2017 mar.; 23:517–525.

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009. Disponível em: <http://dx.doi.org/10.1109/CVPR.2009.5206848>.

Drummond RD, Defelicibus A, Meyenberg M, Valieris R, Dias-Neto E, Rosales RA, et al. Relating mutational signature exposures to clinical data in cancers via signer 2.0. *BMC Bioinformatics*. 2023 nov.; 24:.

Duffy MJ, Harbeck N, Nap M, Molina R, Nicolini A, Senkus E, et al. Clinical use of biomarkers in breast cancer: Updated guidelines from the european group on tumor markers (EGTM). *Eur J Cancer*. 2017 abr.; 75:284–298.

- Díaz-Gay M, Vangara R, Barnes M, Wang X, Islam SMA, Vermes I, et al. Assigning mutational signatures to individual samples and individual somatic mutations with sigproflerassignment. *Bioinformatics*. 2023 dez.; 39:.
- D'Esposito V, Liguoro D, Ambrosio MR, Collina F, Cantile M, Spinelli R, et al. Adipose microenvironment promotes triple negative breast cancer cell invasiveness and dissemination by producing ccl5. *Oncotarget*. 2016 mar.; 7:24495–24509.
- Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. *British Journal of Cancer*. 2020 nov.; 124:686–696.
- Ellrott K, Bailey MH, Saksena G, Covington KR, Kandath C, Stewart C, et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Systems*. 2018; 6:271–281.e7.
- Farahani A, Pourshojae B, Rasheed K, Arabnia HR. A concise review of transfer learning. *ArXiv*. 2021. Disponível em: <https://arxiv.org/abs/2104.02144>.
- 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. *Modern Pathology*. 2022 jan.; 35:44–51.
- Farrell-Sherman A, de la Force N, Prator C, Valieris R, Azam W, Da Silva I, et al. Inflammatory monocytes increase prior to detectable hiv-1 rebound viremia. *BioRxiv*. 2024. Disponível em: <http://dx.doi.org/10.1101/2024.12.24.630257>.
- Fidalgo F, Torrezan GT, Sá BCSd, Barros BDdF, Moredo LF, Valieris R, et al. Family-based whole-exome sequencing identifies rare variants potentially related to cutaneous melanoma predisposition in brazilian melanoma-prone families. *PLOS ONE*. 2022 jan.; 17:e0262419.
- Filiot A, Ghermi R, Olivier A, Jacob P, Fidon L, Mac Kain A, et al. Scaling self-supervised learning for histopathology with masked image modeling. *MedRxiv*. 2023. Disponível em: <http://dx.doi.org/10.1101/2023.07.21.23292757>.
- Fukushima K. Visual feature extraction by a multilayered network of analog threshold elements. *IEEE Transactions on Systems Science and Cybernetics*. 1969; 5:322–333.
- Fumagalli C, Betella I, Ranghiero A, Guerini-Rocco E, Bonaldo G, Rappa A, et al. In-house testing for homologous recombination repair deficiency (hrd) testing in ovarian carcinoma: a feasibility study comparing amoydx hrd focus panel with myriad mychoicecdx assay. *Pathologica*. 2022 set.; 114:288–294.
- Gallego V, Ríos Insua D. Current advances in neural networks. *Annual Review of Statistics and Its Application*. 2022 mar.; 9:197–222.
- Gamble P, Jaroensri R, Wang H, Tan F, Moran M, Brown T, et al. Determining breast cancer biomarker status and associated morphological features using deep learning. *Commun Med (Lond)*. 2021; 1:14.

Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology*. 2016 fev.; 278:563–577.

Godoy DV. Deep learning visuals. [Online; acessado em 21-Nov-2024]. 2021. Disponível em: <https://github.com/dvgodoy/dl-visuals>.

Graziani M, Marini N, Deutschmann N, Janakarajan N, Müller H, Martínez MR. *Attention-Based Interpretable Regression of Gene Expression in Histology*. Springer Nature Switzerland, 2022. 44–60 . ISSN 1611-3349. ISBN 9783031179761. Disponível em: http://dx.doi.org/10.1007/978-3-031-17976-1_5.

Hanahan D. Hallmarks of cancer: New dimensions. *Cancer Discov*. 2022 jan.; 12: 31–46.

He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. 770–778.

Heeke AL, Pishvaian MJ, Lynce F, Xiu J, Brody JR, Chen WJ, et al. Prevalence of homologous recombination–related gene mutations across multiple cancer types. *JCO Precision Oncology*. 2018 nov.;;1–13.

Van Hoeck A, Tjoonk NH, van Boxtel R, Cuppen E. Portrait of a cancer: mutational signature analyses for cancer diagnostics. *BMC Cancer*. 2019 maio; 19: .

Hoppe MM, Sundar R, Tan DSP, Jeyasekharan AD. Biomarkers for homologous recombination deficiency in cancer. *JNCI: Journal of the National Cancer Institute*. 2018 maio; 110:704–713.

Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Networks*. 1989 jan.; 2:359–366.

Huh M, Agrawal P, Efros AA. What makes imagenet good for transfer learning? *ArXiv*. 2016. Disponível em: <https://arxiv.org/abs/1608.08614>.

Ilse M, Tomczak JM, Welling M. Attention-based deep multiple instance learning. *ArXiv*. 2018. Disponível em: <https://arxiv.org/abs/1802.04712>.

Institute NC. What is cancer? [Online; acessado em 22-Nov-2024]. 2021. Disponível em: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>.

Ivanova M, Porta FM, D’Ercole M, Pescia C, Sajjadi E, Cursano G, et al. Standardized pathology report for her2 testing in compliance with 2023 asco/cap updates and 2023 esmo consensus statements on her2-low breast cancer. *Virchows Archiv*. 2023 set.; 484:3–14.

Jaber MI, Song B, Taylor C, Vaske CJ, Benz SC, Rabizadeh S, et al. A deep learning image-based intrinsic molecular subtype classifier of breast tumors reveals tumor heterogeneity that may affect survival. *Breast Cancer Res*. 2020 Jan; 22: 12.

Joseph R. *Contributions to Perceptron Theory*. Cornell Univ., 1960. (Report). Disponível em: <https://books.google.com.br/books?id=O9JUAAAAYAAJ>.

Kather JN, Halama N, Marx A. 100,000 histological images of human colorectal cancer and healthy tissue. Online; acessado em 01-12-2024. 2018. Disponível em: <https://zenodo.org/record/1214456>.

Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis CA, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLOS Medicine*. 2019 jan.; 16:e1002730.

Krajewska M, Fehrmann RSN, de Vries EGE, van Vugt MATM. Regulators of homologous recombination repair as novel targets for cancer treatment. *Frontiers in Genetics*. 2015 mar.; 6:.

Lazard T, Bataillon G, Naylor P, Popova T, Bidard FC, Stoppa-Lyonnet D, et al. Deep learning identifies morphological patterns of homologous recombination deficiency in luminal breast cancers from whole slide images. *Cell Reports Medicine*. 2022 dez.; 3:100872.

Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 1998; 86:2278–2324.

Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A convnet for the 2020s. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2022. Disponível em: <http://dx.doi.org/10.1109/CVPR52688.2022.01167>.

Liwski CR, Castonguay MC, Barnes PJ, Rayson D, Bethune GC. Her2 testing in metastatic breast cancer – is reflex ish testing necessary on her2 ihc-equivocal (2+) cases? *Annals of Diagnostic Pathology*. 2022 ago.; 59:151953.

Loeffler C, Kather JN. Manual tumor annotations in tcga. Online; acessado em 01-12-2024. 2021. Disponível em: <https://zenodo.org/record/5320076>.

Lavinia Loeffler CM, El Nahhas OS, Muti HS, Seibel T, Cifci D, Treeck Mv, et al. Direct prediction of homologous recombination deficiency from routine histology in ten different tumor types with attention-based multiple instance learning: a development and validation study. *MedRxiv*. 2023. Disponível em: <http://dx.doi.org/10.1101/2023.03.08.23286975>.

Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. *The Lancet*. 2021 maio; 397:1750–1769.

Lord CJ, Ashworth A. The dna damage response and cancer therapy. *Nature*. 2012 jan.; 481:287–294.

Lord CJ, Ashworth A. Brca1 revisited. *Nature Reviews Cancer*. 2016 jan.; 16: 110–120.

Lotter W, Hassett MJ, Schultz N, Kehl KL, Van Allen EM, Cerami E. Artificial intelligence in oncology: Current landscape, challenges, and future directions. *Cancer Discov.* 2024 maio; 14:711–726.

Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering.* 2021 mar.; 5:555–570.

Van der Maaten L, Hinton G. Visualizing data using t-sne. *Journal of machine learning research.* 2008; 9:2579–2605.

Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, et al. A method for normalizing histology slides for quantitative analysis. In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2009. 1107–1110. Disponível em: <http://dx.doi.org/10.1109/ISBI.2009.5193250>.

Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics.* 1947 mar.; 18:50–60.

Martins LVC, Bueno AP, Defelicibus A, Drummond RD, Valieris R, Zhu YT, et al. Wsi2ml – an open-source whole slide image annotation software for machine learning applications. In: *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web*. ACM, 2023. (WebMedia '23), 104–109. Disponível em: <http://dx.doi.org/10.1145/3617023.3617038>.

Mirza-Aghazadeh-Attari M, Mohammadzadeh A, Yousefi B, Mihanfar A, Karimian A, Majidinia M. 53bp1: A key player of dna damage response with critical functions in cancer. *DNA Repair.* 2019 jan.; 73:110–119.

Mitri Z, Constantine T, O'Regan R. The her2 receptor in breast cancer: Pathophysiology, clinical use, and new advances in therapy. *Chemotherapy Research and Practice.* 2012 dez.; 2012:1–7.

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.; 387:9–20.

Moynahan ME, Jasin M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nature Reviews Molecular Cell Biology.* 2010 mar.; 11:196–207.

El Nahhas OSM, Loeffler CML, Carrero ZI, van Treeck M, Kolbinger FR, Hewitt KJ, et al. Regression-based deep-learning predicts molecular biomarkers from pathology slides. *Nature Communications.* 2024 fev.; 15:.

Nguyen L, W M Martens J, Van Hoeck A, Cuppen E. Pan-cancer landscape of homologous recombination deficiency. *Nature Communications.* 2020 nov.; 11:.

Orchard G, Nation B. *Histopathology*. London, England: Oxford University Press, 2011. (Fundamentals of Biomedical Science).

- Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*. 1979 jan.; 9:62–66.
- Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*. 2009 mar.; 27:1160–1167.
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019. 8024–8035. Disponível em: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- Pohl-Rescigno E, Hauke J, Loibl S, Möbus V, Denkert C, Fasching PA, et al. Association of germline variant status with therapy response in high-risk early-stage breast cancer: A secondary analysis of the geparocto randomized clinical trial. *JAMA Oncology*. 2020 maio; 6:744.
- Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nature Genetics*. 2017 ago.; 49:1476–1486.
- Popova T, Manié E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with *brca1/2* inactivation. *Cancer Research*. 2012 out.; 72:5454–5462.
- Prakash R, Zhang Y, Feng W, Jasin M. Homologous recombination and human health: The roles of *brca1*, *brca2*, and associated proteins. *Cold Spring Harbor Perspectives in Biology*. 2015 abr.; 7:a016600.
- Pujade-Lauraine E, Ledermann JA, Selle F, Gebiski V, Penson RT, Oza AM, et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a *brca1/2* mutation (solo2/engot-ov21): a double-blind, randomised, placebo-controlled, phase 3 trial. *The Lancet Oncology*. 2017 set.; 18:1274–1284.
- Qaiser T, Mukherjee A, Reddy Pb C, Munugoti SD, Tallam V, aho T, et al. HER2 challenge contest: a detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology*. 2018 Jan; 72:227–238.
- Quellec G, Cazuguel G, Cochener B, Lamard M. Multiple-instance learning for medical image and video analysis. *IEEE Reviews in Biomedical Engineering*. 2017; 10:213–234.
- Rakha EA, El-Sayed ME, Lee AH, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of nottingham histologic grade in invasive breast carcinoma. *Journal of Clinical Oncology*. 2008 jul.; 26:3153–3158.
- Rawat RR, Ortega I, Roy P, Sha F, Shibata D, Ruderman D, et al. Deep learned tissue “fingerprints” classify breast cancers by er/pr/her2 status from h&e images. *Scientific Reports*. 2020 abr.; 10:.

- Ray-Coquard I, Pautier P, Pignata S, Pérol D, González-Martín A, Berger R, et al. Olaparib plus bevacizumab as first-line maintenance in ovarian cancer. *New England Journal of Medicine*. 2019 dez.; 381:2416–2428.
- Reinhardt K, Stückerath K, Hartung C, Kaufhold S, Uleer C, Hanf V, et al. Pik3ca-mutations in breast cancer. *Breast Cancer Research and Treatment*. 2022 out.; 196:483–493.
- Rempel E, Kluck K, Beck S, Ourailidis I, Kazdal D, Neumann O, et al. Pan-cancer analysis of genomic scar patterns caused by homologous repair deficiency (hrd). *npj Precision Oncology*. 2022 jun.; 6:.
- Robbins CJ, Fernandez AI, Han G, Wong S, Harigopal M, Podoll M, et al. Multi-institutional assessment of pathologist scoring her2 immunohistochemistry. *Modern Pathology*. 2023 jan.; 36:100032.
- Robbins H, Monro S. A stochastic approximation method. *The Annals of Mathematical Statistics*. 1951 set.; 22:400–407.
- Robson M, Im SA, Senkus E, Xu B, Domchek SM, Masuda N, et al. Olaparib for metastatic breast cancer in patients with a germline brca mutation. *New England Journal of Medicine*. 2017 ago.; 377:523–533.
- Rosales RA, Drummond RD, Valieris R, Dias-Neto E, da Silva IT. signer: an empirical bayesian approach to mutational signature discovery. *Bioinformatics*. 2016 set.; 33:8–16.
- Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*. 1958; 65:386–408.
- Rosenblatt F. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, 1962. (Cornell Aeronautical Laboratory. Report no. VG-1196-G-8). Disponível em: <https://books.google.com.br/books?id=7FhRAAAAMAAJ>.
- Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986 out.; 323:533–536.
- Russell S, Norvig P. *Artificial intelligence*. 4. ed. Upper Saddle River, NJ: Pearson, 2020.
- Schmidt B, Hildebrandt A. From GPUs to AI and quantum: three waves of acceleration in bioinformatics. *Drug Discov Today*. 2024 jun.; 29:103990.
- Schmidt RM. Recurrent neural networks (rnns): A gentle introduction and overview. *ArXiv*. 2019. Disponível em: <https://arxiv.org/abs/1912.05911>.
- Schoeberl UE, Fitz J, Froussios K, Valieris R, Ourailidis I, Makharova M, et al. Somatic hypermutation patterns in immunoglobulin variable regions are established independently of the local transcriptional landscape. *BioRxiv*. 2022. Disponível em: <http://dx.doi.org/10.1101/2022.05.21.492925>.

- Sepich-Poore GD, Zitvogel L, Straussman R, Hasty J, Wargo JA, Knight R. The microbiome and human cancer. *Science*. 2021 mar.; 371:eabc4552.
- Serra H, Da Ines O, Degroote F, Gallego ME, White CI. Roles of *xrcc2*, *rad51b* and *rad51d* in *rad51*-independent ssa recombination. *PLoS Genetics*. 2013 nov.; 9:e1003971.
- Shamai 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*. 2019 Jul; 2:e197700.
- Shao Z, Bian H, Chen Y, Wang Y, Zhang J, Ji X, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*. 2021; 34:2136–2147.
- Steele CD, Abbasi A, Islam SMA, Bowes AL, Khandekar A, Haase K, et al. Signatures of copy number alterations in human cancer. *Nature*. 2022 jun.; 606: 984–991.
- Sutcliffe SG, Kraemer SA, Ellmen I, Knapp JJ, Overton AK, Nash D, et al. Tracking sars-cov-2 variants of concern in wastewater: an assessment of nine computational tools using simulated genomic data. *Microbial Genomics*. 2024 maio; 10:.
- Sztupinszki Z, Diossy M, Krzystanek M, Reiniger L, Csabai I, Favero F, et al. Migrating the snp array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *npj Breast Cancer*. 2018 jul.; 4:.
- Tellez D, Litjens G, Bándi P, Bulten W, Bokhorst JM, Ciompi F, et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical Image Analysis*. 2019 dez.; 58:101544.
- Trivizakis E, Papadakis G, Souglakos I, Papanikolaou N, Koumakis L, Spandidos D, et al. Artificial intelligence radiogenomics for advancing precision and effectiveness in oncologic care (review). *International Journal of Oncology*. 2020 maio; 57:43–53.
- Turner N, Tutt A, Ashworth A. Hallmarks of “brcaness” in sporadic cancers. *Nature Reviews Cancer*. 2004 out.; 4:814–819.
- Valieris R, Amaro L, Osório CABdT, Bueno AP, Rosales Mitrowsky RA, Carraro DM, et al. Deep learning predicts underlying features on pathology images with therapeutic relevance for breast and gastric cancer. *Cancers*. 2020 dez.; 12:3687.
- Valieris R, Drummond RD, Defelicibus A, Dias-Neto E, Rosales RA, Tojal da Silva I. A mixture model for determining sars-cov-2 variant composition in pooled samples. *Bioinformatics*. 2022 fev.; 38:1809–1815.

- Valieris R, Martins L, Defelicibus A, Bueno AP, de Toledo Osorio CAB, Carraro D, et al. Weakly-supervised deep learning models enable her2-low prediction from h&e stained slides. *Breast Cancer Research*. 2024 ago.; 26:.
- Vargas AJ, Harris CC. Biomarker development in the precision medicine era: lung cancer as a case study. *Nature Reviews Cancer*. 2016 jul.; 16:525–537.
- Vorontsov E, Bozkurt A, Casson A, Shaikovski G, Zelechowski M, Severson K, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature Medicine*. 2024 jul.; 30:2924–2935.
- Wang X, Du Y, Yang S, Zhang J, Wang M, Zhang J, et al. Retccl: Clustering-guided contrastive learning for whole-slide image retrieval. *Medical Image Analysis*. 2023 jan.; 83:102645.
- Wang X, Yang S, Zhang J, Wang M, Zhang J, Yang W, et al. Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical Image Analysis*. 2022 out.; 81:102559.
- Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JM, 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. *Journal of Clinical Oncology*. 2018 jul.; 36:2105–2122.
- Wu S, Yue M, Zhang J, Li X, Li Z, Zhang H, et al. The role of artificial intelligence in accurate interpretation of her2 immunohistochemical scores 0 and 1+ in breast cancer. *Modern Pathology*. 2023 mar.; 36:100054.
- Xiong Y, Zeng Z, Chakraborty R, Tan M, Fung G, Li Y, et al. Nyströmformer: A nyström-based algorithm for approximating self-attention. *ArXiv*. 2021. Disponível em: <https://arxiv.org/abs/2102.03902>.
- Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, et al. Artificial intelligence: A powerful paradigm for scientific research. *Innovation (Camb)*. 2021 nov.; 2:100179.
- Xue T, Chang H, Ren M, Wang H, Yang Y, Wang B, et al. Deep learning to automatically evaluate HER2 gene amplification status from fluorescence in situ hybridization images. *Sci Rep*. 2023 Jun; 13:9746.
- Yao Y, Rosasco L, Caponnetto A. On early stopping in gradient descent learning. *Constructive Approximation*. 2007 abr.; 26:289–315.
- Zou J, Wang E. Cancer biomarker discovery for precision medicine: New progress. *Current Medicinal Chemistry*. 2020 jan.; 26:7655–7671.

ANEXOS

Anexo A – Premiações

O estudo anexado no apêndice C, foi um dos 3 finalistas¹ do **Prêmio Octávio Frias de Oliveira de 2021**, na modalidade *Inovação em Oncologia*, realizado pelo Instituto do Câncer do Estado de São Paulo (ICESP).

O estudo anexado no apêndice D, foi vencedor do **Prêmio Veja Saúde Oncoclínicas de Inovação Médica de 2024**, na modalidade *IA na Transformação Digital em Saúde*, realizado pela revista Veja Saúde.

Os certificados de reconhecimento seguem em anexo a este documento.

¹ Colocação final em segundo lugar



Instituto do Câncer do Estado de São Paulo

XII PRÊMIO 2021

OCTAVIO FRIAS DE OLIVEIRA



O Instituto do Câncer do Estado de São Paulo reconhece

Valieris, R.; Amaro, L.; Osório, C.A.B.T.; Bueno, A.P.; Mitrowsky, R.; Carraro, D.M.; Nunes, D.N.; Dias-Neto, E. e Silva, I.T.

como finalistas no XII Prêmio Octavio Frias de Oliveira na modalidade
Inovação Tecnológica em Oncologia, com o trabalho

*Deep Learning Predicts Underlying Features on Pathology Images
with Therapeutic Relevance for Breast and Gastric Cancer*

São Paulo, 5 de agosto de 2021

FOLHA DE S. PAULO





**Prêmio
Veja Saúde
Oncoclínicas
de Inovação
Médica**

2024

CERTIFICADO

Renan Valieris, Luan Martins, Alexandre Defelicibus, Adriana P. Bueno, Cynthia APT Osório, Dirce M. Carraro, Emmanuel Dias-Neto, Rafael Rosales, José Márcio B. de Figueiredo e Israel Tojal da Silva

foram vencedores na categoria *IA na Transformação Digital em Saúde* do Prêmio VEJA SAÚDE Oncoclínicas de Inovação Médica pelo trabalho **Desenvolvimento e validação de uma plataforma de inteligência artificial para detecção de HER2 em câncer de mama**, realizado no A.C. Camargo Cancer Center.

São Paulo, 26 de setembro de 2024

Diogo Sponchiato
Redator-Chefe
Revista VEJA SAÚDE

APÊNDICES

Apêndice A – Técnicas de Aprendizado Profundo

A.1 DEFINIÇÃO DE UMA REDE NEURAL

No contexto de aprendizado de máquina, uma rede neural é um modelo matemático utilizado para aproximar funções não lineares. Inspirado pela estrutura do cérebro, esse modelo é formado por diversos componentes chamados neurônios ou nós, interconectados que transmitem informação pela rede. Uma das primeiras implementações de redes neurais foi o Perceptron, cujo conceito foi introduzido na década de 60 (Rosenblatt 1958; Joseph 1960; Rosenblatt 1962). Desde então, este campo observou múltiplos avanços que conduziram a popularização atual das redes neurais e suas aplicações no âmbito da inteligência artificial (Gallego e Insua 2022).

Esse modelo pode ser representado por uma função Φ que, dado um vetor de entrada x e um vetor de saída y , é capaz de aproximar y a partir de x utilizando um vetor de parâmetros θ , $\Phi(x; \theta) = \hat{y} \approx y$. Os parâmetros θ podem ser otimizados de modo que uma função de distância, $\mathcal{L}(y, \hat{y}) = \varepsilon$, que calcula a diferença ε entre o resultado obtido $\hat{y}$ e o esperado y , seja minimizada (Berner et al. 2022). Funções de distância utilizadas variam de acordo com o tipo de dado e objetivo do modelo. Um exemplo de função de distância para valores contínuos é a média dos erros quadrados, $\text{MSE}(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^N (y_i - \hat{y}_i)^2$.

A função $\Phi(x; \theta)$ é composta por uma ou mais camadas internas, $\mathcal{C}_n = \sigma(w_n^\top \cdot \mathcal{C}_{n-1} + b_n)$, onde cada camada interna n possui vetores de pesos w_n e vieses b_n que parametrizam a rede neural, $\theta = \{w_0, b_0, \dots, w_n, b_n\}$. As camadas $\mathcal{C}_n$ recebem como entrada a saída da camada anterior $\mathcal{C}_{n-1}$, e utiliza seu resultado como entrada da camada sucessiva, tal como $\Phi(x; \theta) = \mathcal{C}_n \circ \mathcal{C}_{n-1} \circ \dots \circ \mathcal{C}_0 \circ x$, onde o símbolo $\circ$ é o operador de composição de funções, ou seja, $(g \circ f \circ x) = g(f(x))$.

O símbolo σ representa uma função de ativação. Essa função não linear é

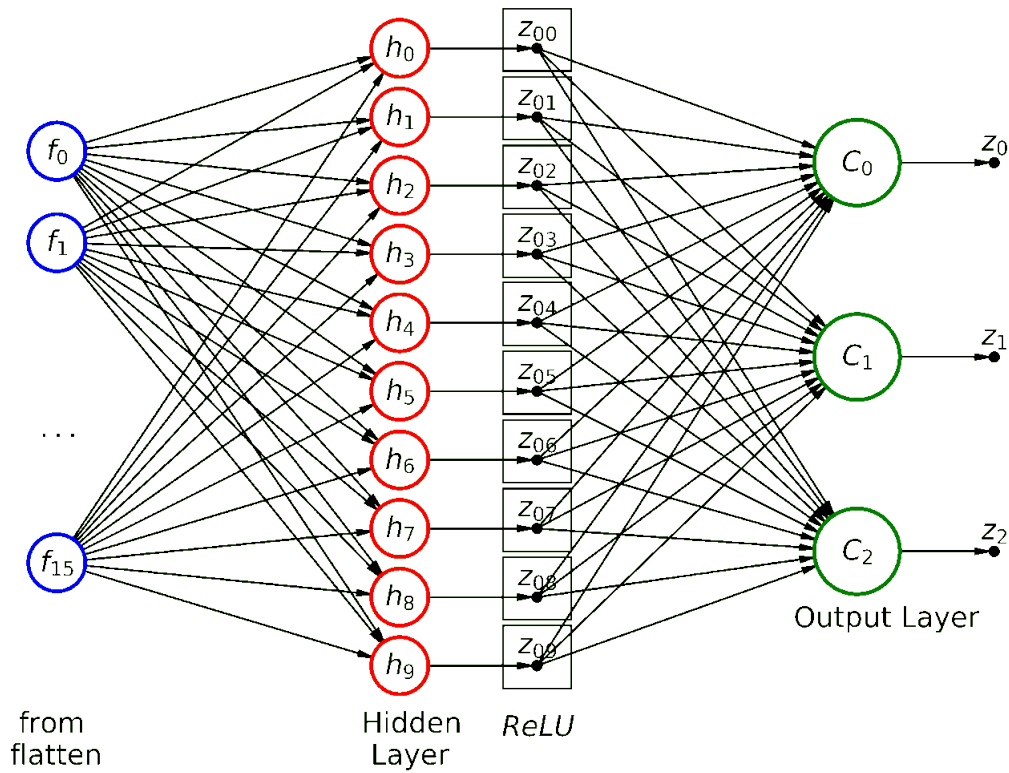
essencial para o funcionamento correto do modelo, pois a capacidade da rede neural de modelar relações complexas é resultado da combinação de diversas dessas camadas não lineares. Notavelmente, é a combinação de camadas não lineares que permite às redes neurais “aprender” qualquer função, de acordo com o teorema da aproximação universal (Hornik et al. 1989), que afirma que deve existir uma sequência de camadas não lineares capaz de modelar uma dada função, desde que o modelo tenha parâmetros suficientes para a modelar. Exemplos de funções de ativação comumente utilizadas são: $\text{Sigmoid}(x) = \frac{1}{1+e^{-x}}$, $\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$ e $\text{ReLU}(x) = \max\{0, x\}$.

A Figura 20 ilustra um exemplo de uma rede neural com 16 nós de entrada, 10 nós na camada interna e 3 nós na camada de saída. Cada nó da camada interna é uma combinação linear ponderada entre os valores de entrada e os pesos da camada, seguida pela função de ativação *ReLU*. Os nós da camada de saída, por sua vez, são uma combinação dos nós da camada interna, cada aresta tem um peso distinto associado que é otimizado durante o processo de treinamento da rede, de modo que os valores de saída preditos pelo modelo $\hat{y}$ sejam similares aos valores utilizados no treinamento y .

A última camada da rede neural calcula o resultado final do modelo. Se o objetivo for treinar um classificador, haverá um nó para cada categoria a ser predita nessa camada, no qual cada elemento do vetor corresponde a uma das categorias. Esses elementos podem assumir qualquer valor dos números reais, porém a transformação *softmax*, frequentemente utilizada na última camada, limita o resultado final de modo que a soma desse vetor é sempre 1, $\sum_{i=0}^N \text{softmax}(x)_i = 1$, tornando o resultado da última camada um vetor de probabilidades.

$$\text{softmax}(x)_i = \frac{e^{x_i}}{\sum_{j=0}^N e^{x_j}}$$

Uma rede neural treinada pode produzir previsões para novos conjuntos de dados sem a necessidade de alterar o valor dos parâmetros θ , esse processo



Fonte: Godoy (2021).

Figura 20 – Exemplo de uma rede neural com uma camada interna. Os nós azuis representam dados de entrada, cada nó interno (vermelho) e de saída (verde) é o resultado de uma combinação de nós anteriores, ilustrada pelas arestas, e seus pesos, que são otimizados durante o processo de treinamento.

é chamado de inferência. No caso de um modelo de classificação, a categoria correspondente ao valor de maior probabilidade na camada final da rede neural é considerada a predição da rede para uma dada entrada.

A.2 TREINAMENTO DE UMA REDE NEURAL

O processo de “treinar” uma rede neural consiste em encontrar valores para os parâmetros θ de modo que quando aplicada a um conjunto de dados M a rede consegue prever $\hat{y}_i$ a partir de x_i com erro ε aceitável para todo par (x_i, y_i) contido em M .

Os algoritmos mais comumente utilizados para esse processo são variantes de uma técnica conhecida como gradiente descendente estocástico. O conceito

de otimização estocástica foi introduzido por Robbins e Monro (1951). Esse é um método iterativo no qual o espaço dimensional dos parâmetros da rede neural é explorado em pequenos passos na direção oposta ao gradiente da função de distância $\mathcal{L}(y, \hat{y})$, resultando na minimização do erro ε .

A cada iteração n , o gradiente da função $\mathcal{L}(y, \Phi(x; \theta))$, $\nabla \mathcal{L}$, é calculado para cada parâmetro θ da rede neural, e seus valores são ajustados $\theta^{(n)} = \theta^{(n-1)} - \gamma \nabla \mathcal{L}_{\theta^{(n-1)}}$, onde γ é um peso fixo chamado de taxa de aprendizado (do inglês *learning rate*), que limita o tamanho do “passo” dado a cada iteração.

A derivada da função da rede neural, necessária para o cálculo do gradiente, tipicamente não tem uma definição em forma fechada. Dessa forma, utiliza-se a técnica conhecida como *backpropagation*, na qual o gradiente é calculado numericamente de trás para frente por meio da aplicação da regra da cadeia (Rumelhart et al. 1986). Existem diversas bibliotecas para linguagens de programação que disponibilizam métodos automáticos de diferenciação baseados nessa técnica. Neste trabalho, utilizamos a biblioteca *PyTorch* (Paszke et al. 2019).

Algoritmo 1 Treinamento da Rede Neural

```

1: para  $epoch = 1, 2, \dots$  faça
2:   para  $batch = 1, 2, \dots, N$  faça
3:     Calcula o resultado do modelo  $\hat{y} \leftarrow \Phi(\theta, x_{batch})$ 
4:     Calcula o erro  $\varepsilon_{batch} \leftarrow \mathcal{L}(y_{batch}, \hat{y})$ 
5:     Calcula o gradiente  $\nabla \mathcal{L}$  relativo a  $\theta$ 
6:     Ajusta os parâmetros da rede  $\theta \leftarrow \theta - \gamma \nabla \mathcal{L}_{\theta}$ 
7:   fim para
8:   Calcula média de erro obtido na época  $\varepsilon_{epoch} = \frac{1}{N} \sum_{i=0}^N \varepsilon_i$ 
9:   Compara os valores de  $\varepsilon_{epoch}$  com anteriores para decidir se deve parar o
      treinamento
10: fim para

```

Um resumo do processo de treinamento é apresentado no Algoritmo 1. O conjunto de dados utilizado no treinamento é dividido em subconjuntos, conhecidos como *batches* (do inglês, lote), o cálculo do treinamento é realizado com estes *batches* para limitar o uso de recursos de memória RAM em conjuntos de dados grandes. A cada *epoch* (do inglês, época) o conjunto de dados é utilizado

para treinar o modelo um *batch* por vez, e ao final de cada época o erro médio obtido é comparado ao erro de épocas anteriores para decidir se o treinamento deve continuar ou não.

Cada iteração idealmente reduz o erro ε , porém dependendo da arquitetura do modelo e do conjunto de dados, não é possível alcançar o valor $\varepsilon = 0$. É mais provável que o erro alcance um platô, onde novas iterações não diminuam o erro de forma significativa, neste caso o processo de treinamento é finalizado, e dizemos que a rede neural foi “treinada” para um dado conjunto de dados.

Em teoria, um modelo treinado em um conjunto de dados M é capaz de realizar novas predições em dados M' , que não foram utilizados para o seu treinamento, mas são similares a M . Porém, na prática, isso pode não funcionar por diversos motivos, sendo os mais comuns a falta de diversidade suficiente de dados em M ou a diferença significativa entre M e M' , mas a razão mais comum são os casos de *overfitting* (do inglês, ajuste excessivo), que ocorrem quando um modelo tem parâmetros suficientes para reduzir o erro ε de M de forma excessiva, fazendo com que o modelo perca a capacidade de generalizar suas predições para outros conjuntos de dados similares.

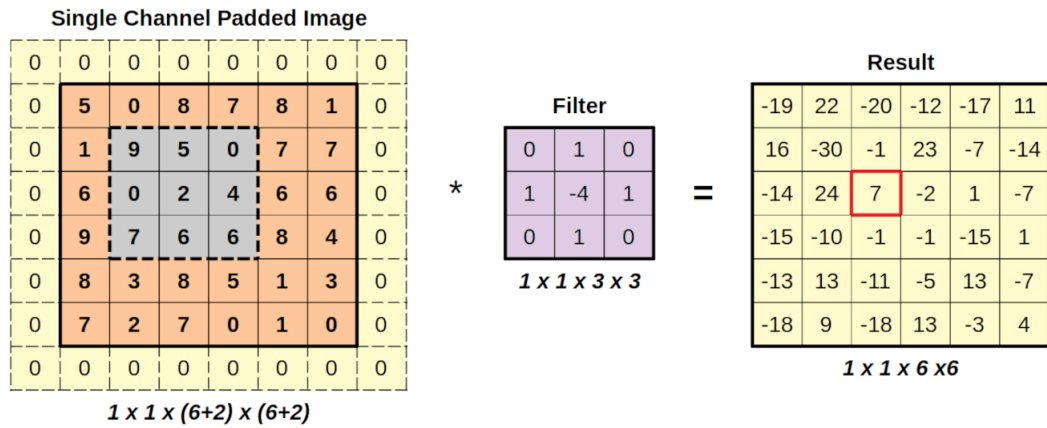
Para evitar que o *overfitting* aconteça, uma estratégia comum é dividir o conjunto de dados M em dois subconjuntos, M_{train} e M_{test} . Neste cenário o modelo é treinado utilizando apenas dados do M_{train} , mas também é calculado o erro do conjunto de teste ε_{test} . Desse modo, é possível observar empiricamente a redução do erro do modelo em um conjunto de dados que não foi utilizado no treinamento. Quando esse erro ε_{test} deixa de diminuir o processo de treinamento pode ser interrompido, esse processo é conhecido como *early stopping* (Yao et al. 2007).

A.3 REDES NEURAIAS CONVOLUCIONAIS

Redes neurais podem ser aplicadas a diversos tipos de dados, porém, em sua definição tradicional, a entrada da rede neural é sempre um vetor unidimensional, quando os dados de entrada têm mais de uma dimensão, a entrada é vetorizada em uma única dimensão, $\mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{nm}$, e portanto a estrutura espacial dos dados não é propagada pela rede neural. Esse é o caso, por exemplo, quando a entrada da rede neural é uma imagem. Tipicamente uma imagem digital é representada por uma matriz $W \times H$, onde W e H representam a largura e altura da imagem em pixels, e os valores da matriz representam tons de cinza, e no caso de uma imagem colorida, uma matriz de três dimensões $C \times W \times H$, onde C representa canais de cores diferentes, tipicamente valores de vermelho, verde e azul.

Para que essa informação espacial seja preservada e utilizada pela rede neural é necessário a definição de um tipo diferente de camada, chamado de camada convolucional. Redes neurais que incluem camadas convolucionais são chamadas de **CNN**, do inglês convolutional neural networks (Fukushima 1969). A aplicação da camada convolucional possibilita o treinamento de modelos capazes de aprender a detectar características como bordas independentemente de sua posição na imagem, e a combinação de diversas dessas camadas permite a detectar padrões complexos em duas dimensões.

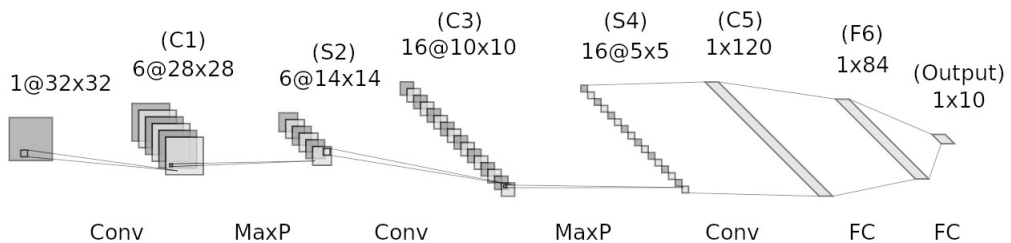
A Figura 21 ilustra o processo de cálculo da convolução em duas dimensões, $\mathcal{C}_n(x, y) = \sum_{i=-k}^k \sum_{j=-k}^k \mathcal{C}_{n-1}(x-i, y-j) \cdot w_n(i, j)$, onde $\mathcal{C}_n$ é uma camada convolucional, w_n são os pesos da camada, k é o tamanho do filtro, x e y são as coordenadas da matriz de entrada em que a convolução é calculada. Os pesos do filtro são multiplicados pelos valores na região correspondente na matriz de entrada e os valores resultantes são somados. Cada célula da matriz de resultado corresponde a uma posição distinta do filtro na matriz de entrada, que é deslocado uma coluna ou linha por vez, dependendo dos parâmetros da camada.



Fonte: Godoy (2021).

Figura 21 – Exemplo do cálculo de uma camada de convolução em 2D. Cada elemento da matriz de resultado é a soma da multiplicação dos valores da matriz de filtro na posição correspondente da matriz de entrada.

Note que a matriz de resultado da convolução também é uma matriz 2D. Apesar de ter dimensões diferentes, é possível realizar uma segunda camada de convolução no resultado. Redes neurais convolucionais tipicamente possuem diversas camadas convolucionais sucessivas (Figura 22), e uma operação de vetorização no final da rede transforma a matriz em um vetor unidimensional, este tipo de camada é conhecido em inglês como *fully-connected layer*. Existem diversas implementações de CNN descritas na literatura, uma das mais populares são as ResNets (He et al. 2016).

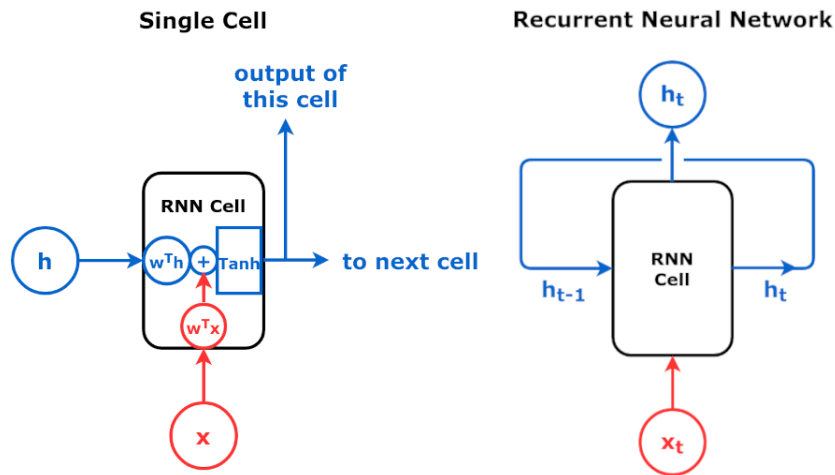


Fonte: Godoy (2021).

Figura 22 – Arquitetura da rede neural convolucional LeNet (Lecun et al. 1998). Ilustrando camadas convolucionais (C1,C3,C5), camadas de *max pooling* (S2,S4), e camadas *fully-connected* (F6,Output).

A.4 REDES NEURAIS RECORRENTES

A Rede Neural Recorrente (RNN, do inglês *Recurrent Neural Network*), é um tipo de rede neural com uma camada recorrente. Essa camada possui uma memória interna que é atualizada a cada inferência do modelo. Esse tipo de camada permite a implementação de modelos que agregam dados em série e realizam uma predição agregada (Figura 23). Um exemplo de aplicação desse tipo de modelo de predição agregada são séries temporais, como no processamento de áudio, no qual o áudio é representado digitalmente por uma série de ondas (Schmidt 2019).



Fonte: Godoy (2021).

Figura 23 – Arquitetura da rede neural recorrente. À esquerda, é ilustrado o cálculo interno da camada recorrente, onde h é a memória e x representa um item de entrada. O resultado da camada pode ser considerado a saída final ou a memória para o próximo item de entrada, ilustrado à direita.

Esse tipo de modelo possui um estado ou memória interna h , que é inicializado com um valor pré-definido. À medida que a série de entrada x_t é consumida pelo modelo, um item por vez, esse estado interno h_t é atualizado, refletindo a agregação dos itens consumidos pelo modelo até aquele ponto. Assim, o resultado do modelo representa uma predição agregada da sequência, em vez de um resultado independente para cada item.

A.5 EXTRAÇÃO DE CARACTERÍSTICAS

Quando treinamos uma rede neural para classificação de imagens, este modelo pode ser dividido em duas partes: 1- Extração de características 2- Classificação. A extração de características consiste em transformar os dados de entrada, como uma imagem digital, em um vetor de números que representam características que possibilitam a classificação desta imagem. Por exemplo, considere um conjunto de imagens de veículos com o objetivo de treinar um modelo para classificação de carros e motos. Entre as características que podem ser utilizadas nesta classificação, estão o número de rodas, o tamanho do veículo e o número de passageiros. O classificador, por sua vez, combina estas características e calcula o resultado final.

Redes neurais não necessitam definir manualmente o cálculo dessas características, pois elas são “aprendidas” pelo modelo durante o treinamento. Notavelmente, uma rede neural treinada em um conjunto de imagens M pode ser reaproveitada para extrair características de um outro conjunto de imagens M' através de uma técnica chamada de *transfer learning* (Farahani et al. 2021). Essa técnica consiste em fixar os parâmetros da rede neural previamente treinada e otimizar somente a última camada do modelo, que é responsável por calcular a classificação final. Partindo do pressuposto de que as características aprendidas pelo modelo em M também são úteis para a classificação de M' , o modelo deverá ser capaz de alcançar uma acurácia satisfatória sem a necessidade de retrainar todos os parâmetros da rede neural.

Existem diversas redes neurais pré-treinadas para uso em *transfer learning*, um dos conjuntos de dados mais utilizados para este fim é o *ImageNet* (Deng et al. 2009). Este conjunto de dados contém milhões de imagens anotadas em 1000 categorias diferentes, a grande quantidade de categorias e dados torna este conjunto um alvo popular para utilização como extrator de características (Huh et al. 2016).

A.6 APRENDIZADO COM MÚLTIPLAS INSTÂNCIAS

O treinamento de redes neurais requer um conjunto de dados no qual cada item é previamente anotado com seu resultado esperado, pois esse resultado esperado é utilizado para calcular a função de distância, o gradiente e os ajustes dos parâmetros da rede. No entanto, nem sempre é possível anotar cada item do conjunto de dados. Nesses casos é possível utilizar a técnica **MIL**, do inglês *Multiple Instance Learning*.

O **MIL** é útil quando os itens de um conjunto de dados são agrupados em “sacolas”, e essas sacolas tem uma anotação binária. Sacolas positivas contêm pelo menos um item positivo, enquanto sacolas negativas contêm apenas itens negativos. O **MIL** tem como objetivo treinar um modelo capaz de classificar itens individuais utilizando apenas informações provenientes da anotação das sacolas (Babenko 2008).

Essa abordagem é especialmente relevante e de particular interesse no contexto de análise de imagens histopatológicas, pois cada lâmina digitalizada tem alta resolução, necessária para visualizar detalhes a nível de células individuais. No entanto, devido à grande quantidade de dados, não é viável construir uma rede neural capaz de processar uma lâmina inteira de uma vez. Também é importante salientar que uma lâmina contém uma diversidade de células diferentes, as células de interesse variam de acordo com o objetivo do modelo, porém é esperado que apenas uma parcela destas células seja de fato relevante para a análise. Utilizando **MIL**, é possível dividir uma lâmina digitalizada em fragmentos (*tiles*) menores e utilizar uma anotação a nível de lâmina como a “sacola” de itens. Existe uma série de abordagens descritas na literatura que implementam soluções para o **MIL** no contexto de imagens histopatológicas (Quellec et al. 2017; Ilse et al. 2018; Campanella et al. 2019; Lu et al. 2021; Shao et al. 2021; Graziani et al. 2022).

A.6.1 Top-K

Este modelo, descrito por Campanella et al. (2019), estende a hipótese original do **MIL**, na qual sacolas positivas devem conter pelo menos um item positivo, para sacolas positivas devem ter pelo menos k itens positivos.

Algoritmo 2 Top-K MIL

```

1: para  $epoca = 1, 2, \dots$  faça
2:   para  $sacola = 1, 2, \dots, N$  faça
3:     Inferência de todos os itens  $z_{item} \leftarrow \Phi(\theta, x_{item})$ 
4:     Ordene o vetor de escores  $s \leftarrow \text{sort}(z)$ 
5:     Selecione os  $k$  maiores itens  $T_{sacola} \leftarrow \{s_{N-k+1}, \dots, s_N\}$ 
6:   fim para
7:   Efetue o treinamento de  $\Phi$  com itens selecionados  $T$ 
8: fim para
  
```

O algoritmo 2 ilustra a técnica **Top-K**, utilizando uma rede neural convolucional com um valor contínuo entre 0 e 1 como resultado. Inicialmente, é feito inferência de todos os itens pela rede neural, e seu resultado, denominado escore, é utilizado para ordenar os itens do menor para o maior. Os k itens de maior valor de escore em cada sacola são separados, e a anotação da sacola (0 para sacolas negativas; 1 para sacolas positivas) é utilizada como uma anotação individual desses itens para treinar e ajustar os pesos do modelo nessa iteração. Em seguida, uma nova inferência é feita com todos os itens, que pode selecionar itens diferentes, e esse processo é repetido até a convergência do modelo.

A vantagem desse método é a facilidade de implementação, já que podemos utilizar uma arquitetura **CNN** convencional para o treinamento do modelo. Porém, as desvantagens são a necessidade de definir previamente o valor de k , que pode não ser óbvio dependendo do conjunto de dados, e caso o objetivo final seja a classificação da sacola, também é necessário treinar um modelo agregador após o **MIL**, como por exemplo uma **RNN** como descrita na Seção A.4, que combina a informação dos k itens em uma única predição a nível de sacola.

A.6.2 ADMIL

Esse modelo, descrito por Ilse et al. (2018), define um novo tipo de camada, denominada camada de atenção, que aprende durante o treinamento, a atribuir pesos aos itens de uma sacola, de forma que a combinação destes itens ponderada por esses pesos é utilizada para uma classificação final do modelo.

Seja $H = \{h_1, \dots, h_K\}$ uma sacola de K itens contendo vetores de características previamente extraídos por uma rede pré-treinada, conforme descrito na Seção A.5. O processo de agregação via mecanismo de atenção **MIL** pode ser descrito como $z = \sum_{k=1}^K a_k h_k$, onde $a = \text{softmax}(w^\top \tanh(VH^\top))$, w e V são parâmetros a serem treinados pela rede. A operação *softmax* converte um vetor de valores reais em um vetor de probabilidades cuja soma é sempre 1. Esses pesos são utilizados em uma soma ponderada das características de todos os itens da sacola, e o resultado é um vetor de características agregado z da sacola como um todo, que é utilizado para treinar uma camada de classificação no final da rede neural.

Esse método é interessante porque ele combina a seleção de itens importantes para a predição com o passo de agregação e classificação no mesmo modelo, de modo que ambos os componentes são treinados simultaneamente, diferentemente do método **Top-K**, no qual o modelo agregador é treinado a parte. Outra vantagem deste método é que não é necessário definir previamente o valor de k .

A.6.3 CLAM

Este modelo, descrito por Lu et al. (2021), estende o algoritmo de **MIL**, que originalmente possibilita apenas predição binária (positivo/negativo), para predição de múltiplas categorias. O **CLAM** utiliza a camada de atenção, conforme definido na Seção A.6.2, não só para identificar itens relevantes da “sacola” como um todo, mas também individualmente para cada uma das categorias a ser classificadas. Isso permite que o modelo aprenda características relevan-

tes de cada categoria mesmo que uma “sacola” contenha uma mistura de itens pertencentes a múltiplas categorias.

Esse método é interessante porque ele permite a predição de múltiplas categorias. O método aprende a detectar a presença dessas categorias em nível de item em cada “sacola”, mesmo quando a sacola é composta por uma mistura, algo que não fazia parte da hipótese original do algoritmo **MIL**.

A.6.4 TransMIL

Esse modelo, descrito por Shao et al. (2021), propõem uma implementação diferente para a camada de atenção. A camada de atenção descrita em (Ilse et al. 2018) se baseia na hipótese de que todos os itens da “sacola” são independentes e identicamente distribuídos, o que não é verdade na prática pois os itens são fragmentos de uma mesma lâmina. Portanto, foi proposto que uma arquitetura que considera a correlação de informações entre itens de uma “sacola” pode melhorar o desempenho do modelo **MIL**. A camada de atenção proposta utiliza uma implementação baseada em *Transformer* que é capaz de aprender essas correlações entre itens de forma escalável (Xiong et al. 2021).

A.6.5 Regressão MIL

Esse modelo, descrito por Graziani et al. (2022), utiliza a arquitetura **MIL** com camada de atenção, como descrito na Seção A.6.2, porém a última camada é modificada para fazer uma predição de um único número real, em vez de categorias. A última camada utiliza a função *Sigmoid* para manter o resultado dentro do intervalo entre 0 e 1.

Esse método é interessante para predição de valores contínuos, como escores, pois não é necessário subdividir a faixa de valores do escore em categorias para treinar o modelo.

A.7 MÉTRICAS DE DESEMPENHO

Após o treinamento, espera-se que o modelo seja capaz de generalizar para outros conjuntos de dados, porém isso nem sempre acontece na prática, por isso é comum medir o desempenho desse modelo de forma empírica por meio do cálculo de uma ou mais métricas de desempenho em outros conjuntos de dados. Existe diversas métricas diferentes descritas na literatura para diferentes objetivos. Aqui, descrevemos as utilizadas nesse estudo.

A.7.1 F1-Score

A **F1-Score** é uma medida de desempenho para classificação binária, para calcular esta métrica as previsões do modelo são comparadas aos rótulos anotados dos itens e de acordo com o resultado cada previsão pode ser classificada em um de quatro tipos:

- **TP** – verdadeiro positivo, quando a previsão positiva é correta;
- **TN** – verdadeiro negativo, quando a previsão negativa é correta;
- **FP** – falso verdadeiro, quando a previsão positiva é incorreta;
- **FN** – falso negativo, quando a previsão negativa é incorreta.

A ocorrência de cada um destes tipos é contada, e estas contagens são utilizadas para calcular a medida:

$$\text{F1-Score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

A medida final é um número que varia de 0 a 1, e quanto mais perto de 1 melhor é o modelo.

O F1-score também pode ser calculado a partir das métricas intermediárias **Precisão** (*Precision*), que calcula a fração itens verdadeiros positivos do total de itens positivos, $\text{Precisão} = \frac{\text{TP}}{\text{TP} + \text{FP}}$, e a **Revocação** (*Recall*) que calcula a fração

do total de itens verdadeiros positivos do total de itens preditos como positivo, $\text{Revocação} = \frac{\text{TP}}{\text{TP} + \text{FN}}$, o F1-score pode então ser calculado como

$$\text{F1-Score} = 2 \cdot \frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}}$$

A.7.2 AUC

A **AUC** ou área sob a curva **ROC**, é uma medida de desempenho para classificação binária, na qual a predição do modelo é dada na forma de uma probabilidade do item ser positivo. A curva **ROC** (*receiver operating characteristic*) é um gráfico que ilustra o desempenho do modelo em diferentes limiares. Nesse gráfico o eixo x contém a taxa de falsos positivos e o eixo y contém a taxa de verdadeiros positivos, pontos são traçados considerando diferentes pontos de corte para considerar a predição do modelo uma predição positiva. Um modelo que prediz ao acaso, resulta num gráfico com uma linha diagonal e $\text{AUC} = 0.5$. Quanto melhor for o modelo mais próximo do canto superior esquerdo a curva fica. A **AUC** é a área abaixo desta curva, que sumariza o quão próxima a curva está de um classificador “perfeito”. Esta métrica também pode ser utilizada na classificação de múltiplas categorias, neste caso a **AUC** binária é calculada para cada categoria, e ao final a média das **AUCs** é calculada, esta sumarização é conhecida como *macro-averaging*.

A.7.3 Teste Estatístico para comparação de métricas

Para testar se dois conjuntos de métricas obtidas são significantemente diferentes, utilizamos o teste estatístico não-paramétrico de Mann e Whitney (1947) para cada comparação. Nos casos em que múltiplas comparações foram realizadas, foi utilizado o método de Benjamini e Hochberg (1995) para ajustar os p-valores.

A.8 VALIDAÇÃO CRUZADA

A validação cruzada é uma técnica de reamostragem utilizada em análise estatística para estimar como os resultados de uma análise podem generalizar a um conjunto de dados independente, a fim de avaliar a presença de problemas como *overfitting* ou viés de seleção. Existem várias variantes desta técnica, neste trabalho nos utilizamos a validação cruzada com *k folds*.

Na validação cruzada com *k folds*, o conjunto de dados é separado em *k* grupos de tamanho igual. Um grupo é considerado o grupo de validação e os outros $k - 1$ grupos são usados como dados de treinamento, e uma métrica de desempenho é calculada. Esse processo é repetido *k* vezes, utilizando cada um dos grupos como validação e os outros para treinamento. Ao final, teremos *k* métricas que podem ser sumarizadas em uma única estimativa.

A.9 REDUÇÃO DE DIMENSIONALIDADE

Redução de dimensionalidade se refere a um conjunto de técnicas visam transformar dados com alto número de dimensões em um espaço dimensional reduzido, de modo que esse espaço consiga reter suas propriedades do espaço dimensional original do dados. Esse tipo de técnica é tipicamente usado para simplificar análises de dados com muitas dimensões, encontrar variáveis importantes de um conjunto de dados ou criar visualizações desses dados no espaço reduzido.

Existem vários tipos diferentes de técnicas de redução de dimensionalidade, neste trabalho nos utilizamos **t-SNE** (*t-distributed stochastic neighbor embedding*) que é um método estatístico que projeta dados em um espaço dimensional $\mathbb{R}^N$ para um espaço 2-D de forma que pontos próximos em $\mathbb{R}^N$ também estarão próximos em 2-D (Maaten e Hinton 2008). Essa técnica é comumente utilizada para visualização de dados com muitas dimensões.

Apêndice B – Cálculo do Escore Agregado para HRD

Seja $\mathcal{E} = \{\mathcal{E}_k\}$, $1 \leq k \leq K$, as exposições de um superconjunto de k diferentes conjuntos de assinaturas mutacionais. O conjunto k , $\mathcal{E}_k$, contém a exposição das assinaturas para cada amostra i para m_k possíveis assinaturas do tipo k . Mais especificamente, para cada k seja

$$\mathcal{E}_k = \{E_{k,b,i}\}, \quad 1 \leq b \leq m_k, \quad 1 \leq i \leq G,$$

onde $E_{k,b,i}$ denota a exposição da amostra i na assinatura b do tipo k . Neste estudo nós consideramos quatro tipos diferentes de assinaturas para a definição do escore agregado. Especificamente, as assinaturas de mutações (**SBS**), *indels* (**ID**), assinaturas de número de cópias (**CN**) e o *scarHRD*,

$$\begin{aligned} \mathcal{E}_1 &= \{E_{1,b,i}\}, \quad b \in \{\text{SBS1}, \dots, \text{SBS30}\}, \\ \mathcal{E}_2 &= \{E_{2,b,i}\}, \quad b \in \{\text{ID1}, \dots, \text{ID23}\}, \\ \mathcal{E}_3 &= \{E_{3,b,i}\}, \quad b \in \{\text{CN1}, \dots, \text{CN21}\}, \\ \mathcal{E}_4 &= \{E_{4,b,i}\}, \quad b \in \{\text{scarHRD}\}. \end{aligned} \tag{1}$$

A exposição de cada assinatura nos conjuntos $\mathcal{E}_k$, $k = 1, 2, 3$, foi estimado a partir do método de fatoração de matrizes não-negativas (NMF) (Alexandrov et al. 2013; Rosales et al. 2016), porém estas exposições não são comparáveis entre conjuntos de assinaturas distintos. Portanto nos utilizamos a seguinte normalização,

$$\hat{E}_{k,b,i} = \frac{E_{k,b,i}}{\sum_{b=1}^{m_k} E_{k,b,i}}, \quad \text{for } k = 1, 2, 3, \tag{2}$$

ou seja, normalização da exposição das assinaturas em cada amostra, calculando a proporção da assinatura de interesse pelo total de exposição em cada amostra. De modo que $\hat{E}_{k,b,i} \in [0, 1]$ para todo k, b e i .

Para $k = 4$, a assinatura $scarHRD$, nós normalizamos com relação ao valor total observado para todas as amostras no conjunto de dados

$$\hat{E}_{4,scarHRD,i} = \frac{E_{4,scarHRD,i}}{\max_{1 \leq i \leq G} E_{4,scarHRD,i}}. \quad (3)$$

Agora focamos no subconjunto de assinaturas relacionados ao **HRD**,

$$\mathcal{E}_1 = \{E_{1,SBS3,i}\}, \quad \mathcal{E}_2 = \{E_{2,ID6,i}\}, \quad \mathcal{E}_3 = \{E_{3,CN1,i}, E_{3,CN17,i}\}, \quad \mathcal{E}_4 = \{E_{4,scarHRD,i}\}.$$

Para o subconjunto $k = 3$, nós combinamos a exposição dos marcadores $CN1$ e $CN17$,

$$\hat{E}_{3,CN1CN17,i} = \frac{(1 - \hat{E}_{3,CN1,i}) + \hat{E}_{3,CN17,i}}{2},$$

devido a ambos $CN1$ e $CN17$ estarem associados ao **HRD**, e $CN1$ apresentar correlação negativa a $CN17$. O conjunto $\mathcal{E}_3$ é, portanto, redefinido como $\mathcal{E}_3 = \{E_{3,CN1CN17,i}\}$.

Em seguida, é realizado uma regularização que considera o número diferente de amostras em cada tipo de tumor. Suponha que haja T tipos de tumores e que todas as $1 \leq i \leq G$ amostras, G_t têm o tipo t , $1 \leq t \leq T$.

Para um dado $\varepsilon \in (0, 1)$, seja

$$w_{k,b,t}(\varepsilon) = \frac{1}{G_t} \sum_{i=1}^{G_t} \mathbf{1}\{\hat{E}_{k,b,i} \geq \varepsilon\}$$

onde $\mathbf{1}\{\hat{E}_{k,b,i} \geq \varepsilon\}$ é 1 se $\hat{E}_{k,b,i} \geq \varepsilon$ ou 0 caso contrario.

$w_{k,b,t}(\varepsilon)$ denota a fração relativa de amostras no tumor do tipo t com exposição $\hat{E}_{k,b,i}$ acima de ε . A adição de pesos diferentes para cada tumor pode levar a introdução de um viés e tornar os valores entre tumores diferentes incomparáveis. Para manter tumores diferentes na mesma escala, nos rebalanceamos os pesos de cada tumor,

$$\hat{w}_{k,b,t}(\varepsilon) = \frac{w_{k,b,t}(\varepsilon)}{\sum_{k=1}^4 w_{k,b,t}(\varepsilon)}$$

O escore para a amostra i do tumor t é, portanto, calculado como

$$s_{t,i}(\varepsilon) = \sum_{k=1}^4 \sum_{b=1}^{m_k} \hat{w}_{k,b,t}(\varepsilon) \hat{E}_{k,b,i}. \quad (4)$$

A análise reportada neste trabalho foi obtida com $\varepsilon = 0.1$. Esta escolha apresentou os melhores resultados em termos de erro médio quadrado e correlação.

Apêndice C – Artigo HRD – Valieris 2020

Segue abaixo o artigo publicado na revista *MDPI Cancers* em dezembro de 2020, contendo a análise do modelo preditor de **HRD** binário, descrito na Seção 4.1.

Article

Deep Learning Predicts Underlying Features on Pathology Images with Therapeutic Relevance for Breast and Gastric Cancer

Renan Valieris ^{1,†} , Lucas Amaro ^{1,†}, Cynthia Aparecida Bueno de Toledo Osório ², Adriana Passos Bueno ^{1,2}, Rafael Andres Rosales Mitrowsky ³, Dirce Maria Carraro ⁴, Diana Noronha Nunes ⁵, Emmanuel Dias-Neto ⁵  and Israel Tojal da Silva ^{1,*} 

¹ Laboratory of Computational Biology Bioinformatics, CIPE/A.C. Camargo Cancer Center, São Paulo 01508-010, Brazil; renan.valieris@accamargo.org.br (R.V.); lucas.amaro@hey.com (L.A.); adriana.passos@accamargo.org.br (A.P.B.)

² Department of Pathology, CIPE/A.C. Camargo Cancer Center, São Paulo 01525-001, Brazil; cabtoledo@accamargo.org.br

³ Department of Computation and Mathematics, University of São Paulo, Ribeirão Preto 14040-901, Brazil; rrosales@usp.br

⁴ Laboratory of Genomics and Molecular Biology, CIPE/A.C. Camargo Cancer Center, São Paulo 01508-010, Brazil; dirce.carraro@accamargo.org.br

⁵ Medical Genomics Laboratory, CIPE/A.C. Camargo Cancer Center, São Paulo 01525-001, Brazil; dnoronha@accamargo.org.br (D.N.N.); emmanuel@accamargo.org.br (E.D.-N.)

* Correspondence: itojal@accamargo.org.br; Tel.: +55-11-2189-5000

† These authors contributed equally to this work.

Received: 13 October 2020; Accepted: 6 November 2020; Published: 9 December 2020



Simple Summary: DNA repair deficiency (DRD) is common in many cancers. This deficiency contributes to pathogenesis of the disease, but it also presents an opportunity for therapeutic targeting. However, current DRD identification assays are not available for all patients. We propose an efficient machine learning algorithm which can predict DRD from histopathological images. The utility of our method was shown by considering the detection of homologous recombination deficiency (HRD) and mismatch repair deficiency (MMRD) in breast and gastric cancer respectively. Our findings demonstrate that machine-learning approaches can be used in advanced applications to assist therapy decisions.

Abstract: DNA repair deficiency (DRD) is an important driver of carcinogenesis and an efficient target for anti-tumor therapies to improve patient survival. Thus, detection of DRD in tumors is paramount. Currently, determination of DRD in tumors is dependent on wet-lab assays. Here we describe an efficient machine learning algorithm which can predict DRD from histopathological images. The utility of this algorithm is demonstrated with data obtained from 1445 cancer patients. Our method performs rather well when trained on breast cancer specimens with homologous recombination deficiency (HRD), AUC (area under curve) = 0.80. Results for an independent breast cancer cohort achieved an AUC = 0.70. The utility of our method was further shown by considering the detection of mismatch repair deficiency (MMRD) in gastric cancer, yielding an AUC = 0.81. Our results demonstrate the capacity of our learning-base system as a low-cost tool for DRD detection.

Keywords: digital pathology; deep learning; mutational signature; biomarker; DNA repair deficiency

1. Introduction

DNA repair mechanisms have evolved to efficiently deal with various types of DNA damage from endogenous and/or exogenous sources. Independent of the source, however, DNA repair mechanisms are essential for maintaining genomic integrity in cells [1]. In certain human pathologies such as cancer, disruption of repair pathways lead to genomic instability which underlies the condition. Thus, repair pathways can also represent efficient targets for anti-tumor therapies [2]. In particular, the homologous recombination (HR) repair pathway plays a key role in maintaining genomic stability by driving high fidelity repair of double-stranded DNA breaks (DSBs) [3]. It has been documented that loss of function (LOF) mutations in genes involved in DSB repair via HR, such as *BRCA1* and *BRCA2*, confer breast and ovarian cancer susceptibility [4]. Correspondingly, tumors deficient in HR repair (HRD) are the most common forms of hereditary breast and ovarian cancer [4]. While HR-defective tumors have been found to be enriched in breast cancer cases, recent studies have demonstrated that other sporadic cancers share molecular and phenotypic characteristics of BRCA-mutant tumors, a concept termed as “BRCAness” [4,5]. An ability to stratify these tumors would have significant implications for cancer risk, treatment strategies, and patient survival [2–6].

Currently, biomarkers for HRD are used to identify tumors with events beyond germline or somatic LOF mutations in *BRCA1* or *BRCA2* [7]. Biomarkers for HRD are also used to identify tumors which carry additional genes involved in HR repair (HRR), such as somatic mutational signatures, gene expression profiles of genes which play crucial roles in tumor development, functional assays of protein expression, and “genomic scar” assays which use array-based comparative genomic hybridization. Due to the intricacy of the HR pathway, a combination of data regarding determinant molecular changes can improve the detection of tumors with HRD [7,8].

Comprehensive analyses of tumor-derived genome sequences have shown that a mutational signature known as Signature 3 (COSMIC database) is a promising biomarker for HRD in breast cancer [8,9]. Signature 3 is also recognized as a reliable readout of HR status. This signature arises from broad molecular changes (e.g., germline variants and somatic mutations in genes involved in HRR and epigenetic silencing of *BRCA1* and *RAD51c*, among others) in components of the HR pathway which are independent of *BRCA1/2* mutations [8]. It has been proposed that HRD tumors can be stratified according to their Signature 3 profile, and this profile is more robust than specific assays which detect an isolated molecular change related to the HR pathway. However, despite being a promising biomarker of HRD, obtaining a Signature 3 profile is dependent on the availability of next-generation sequencing. Consequently, this profiling approach is not widely available in all clinical settings. As a result, potential opportunities to identify patients who may benefit from poly (ADP-ribose) polymerase inhibitor treatments have been limited [10]. Indeed, not all patients are screened for HRD status at the time of their cancer diagnosis. Hence, efforts to uncover new approaches for quantifying HRD would be beneficial, and could also expand the therapeutic potential of HRD-targeting agents across many tumor types [7]. Development of a low-cost strategy to accurately and rapidly assess HRD status in any tumor sample is of particular interest.

To date, remarkable advances have been made in predicting the relationship between a cell’s genome and its phenotype, thereby facilitating personalized treatment strategies [11–15]. Most notably, the development of deep learning techniques to predict outcomes, disease recurrence, and to distinguish cancer type, tumor composition, gene mutations, and grading status are of great interest. Herein, we trained a deep-learning model on digital histopathological images to detect HRD tumors (Section 2, and further discussed in Section 3). The ground truth labels for the deep-learning process are defined according to the activity of underlying mutational signatures. These signatures represent a reliable readout of HRD. We demonstrate that histopathological images can provide an accurate and efficient predictor of HRD status. The machine learning techniques we used employ a novel end-to-end framework which is based on deep multiple instance learning (MIL), see Section 4. After demonstrating its feasibility in classifying tumors as HRD, we further explore the utility and performance of our method by applying it to a cohort of patients with gastric cancer. We detected mismatch repair deficiency

(MMRD), a repair pathway which is frequently impaired in many cancers [16]. Extensive evidence has also proved MMRD to be a reliable predictive biomarker guiding the treatment of some cancers [16,17]. Our pre-trained models are publicly available and can be successfully adapted to other specific types of DNA repair deficiencies, independent of a cancer's tissue of origin. With immunotherapy being a focus of the cancer therapeutic paradigm in recent years, we anticipate that a low-cost, screening-based tool as described here would have great potential.

2. Results

2.1. Weakly Supervised Deep Learning for DNA Repair Deficiency (DRD) Assessment

Here, we describe a deep learning model for DRD (DNA repair deficiency) assessment which is applied to hematoxylin and eosin (H&E) stained histopathological whole slide images. Briefly, breast and gastric cancer samples with matched molecular data and histopathology images were collected from The Cancer Genome Atlas (TCGA). Next, the ground truth set of positive instances in the data set, which represent DNA repair deficient tumors, was defined according to mutational activity (see Methods). The architecture of our deep learning model is depicted in Figure 1, and it comprises the following steps. First, during the inference and learning steps, a whole slide image is tessellated into tiles. Otsu's global thresholding is subsequently employed to discard all background tiles. The remaining tiles are then loaded into a convolutional neural network (CNN) in order to rank the tiles according to their probability of being positive (Figure 2). With learning gained from the top-ranking tiles per slide, a recurrent neural network (RNN) is used to establish a final slide classification according to slide-level aggregation. The latter process also takes into account the most promising tiles previously detected from the CNN.

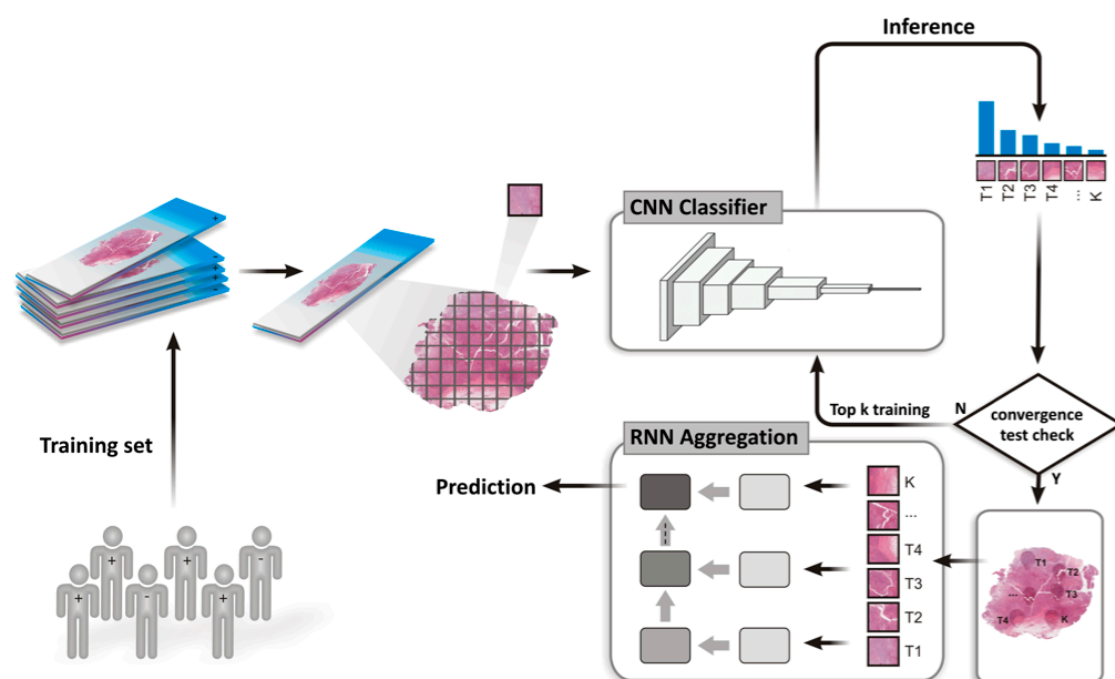


Figure 1. Overview of the end-to-end framework of training and testing for DNA repair assessment on whole slide images.

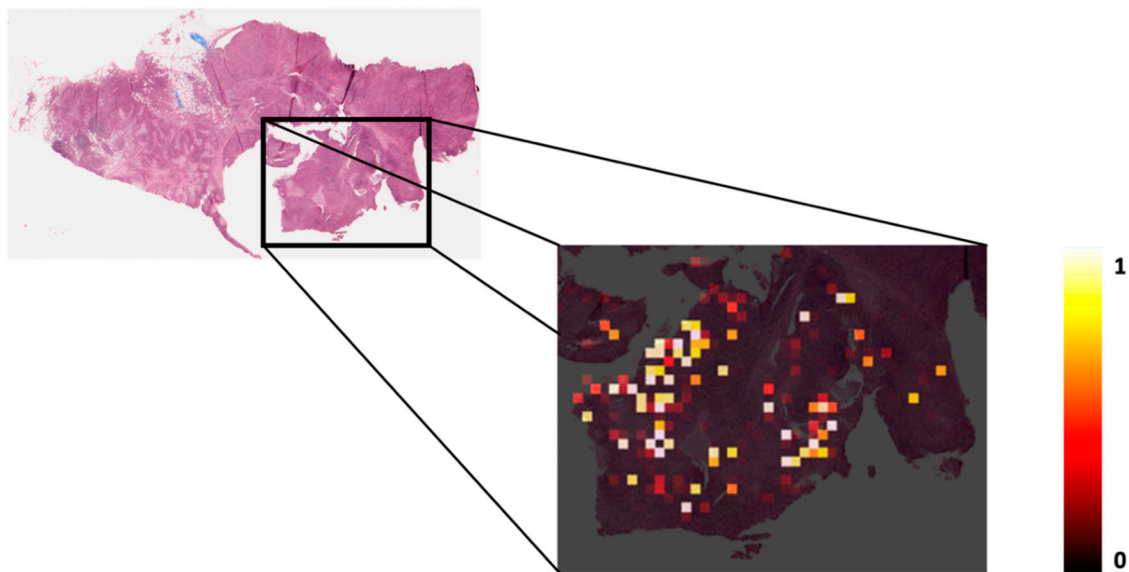


Figure 2. A whole slide image (32x) and corresponding heatmap for a single patient in the independent test set. The intensity of coloring in the heatmap represents the probability of being class-positive in the CNN (convolutional neural network) step.

Notably, when tracking down the spatial patterns learned by the CNN while using an independent validation dataset, we noticed that our model selects regions which are composed of tiles that include several cell types (Figure 3 and Figure S1). This result suggests that tile-embedded cell diversity harbors underlying features which can predict DRD.

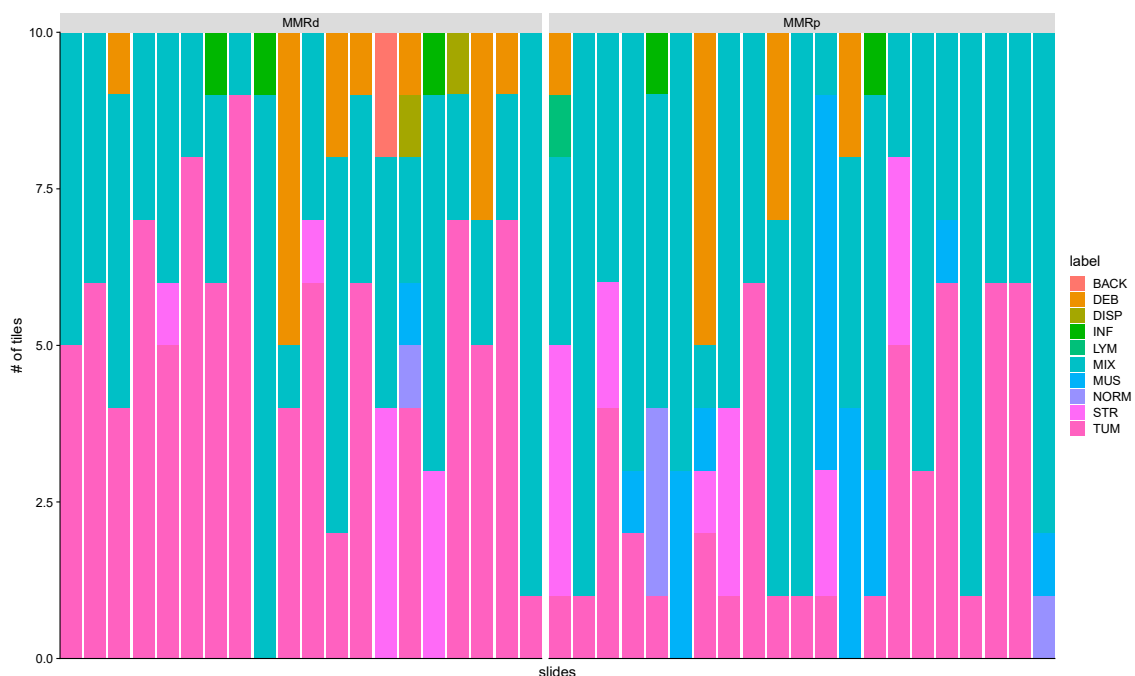


Figure 3. Tissue classes learned from convolutional neural network (CNN) in stomach cancer. Examples from mismatch repair deficiency (MMRD) and mismatch repair proficiency (MMRP) groups. The bar represents the spectrum of histologic diversity in the test set and it contains the top-ranking tiles per slide learned from CNN. These tiles were manually annotated by two pathologists among the following ten tissue labels: ADI, adipose tissue; BACK, background; DEB, debris; DIS, dysplasia; INF, inflammation; LYM, lymphocytes; MUS, smooth muscle; NORM, normal gastric mucosa; STR, cancer-associated stroma; TUM, gastric adenocarcinoma epithelium.

Our model detected HR deficiency in breast tumors with an area under curve (AUC) of 0.80 (95% confidence interval (CI): 0.709–0.889, Figure 4A). When the model was challenged with an independent cohort of triple-negative breast cancer (TNBC) whole slide images taken from formalin-fixed paraffin embedded (FFPE) sections (65 slides from 65 patients), the AUC was 0.70 (95% CI: 0.542–0.849, Figure 4B). These breast cancers were previously characterized as proficient or deficient in HRR as described in the Methods section [18]. Taken together, these results demonstrate that our model is consistent.

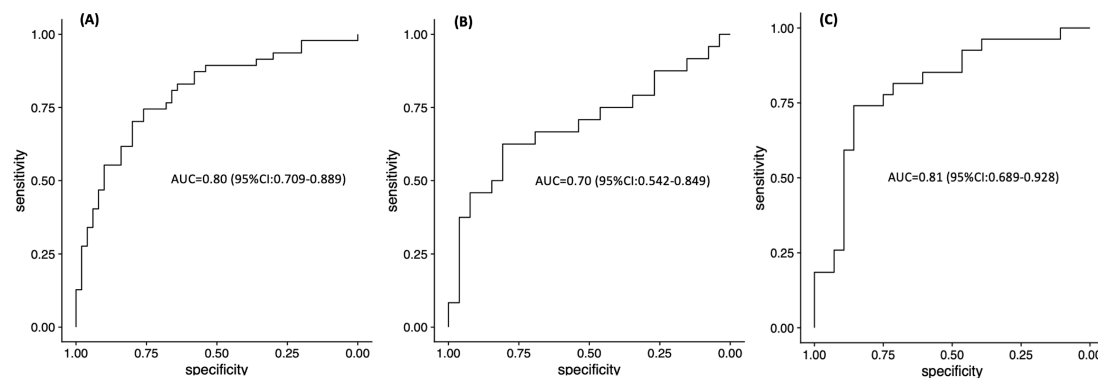


Figure 4. ROC analysis of classifier predicting DRD (DNA repair deficiency) status on (A) TCGA-BRCA training cohort, (B) ACCC-BRCA independent validation cohort and (C) TCGA-STAD training cohort.

Next, we tested the performance of our approach to predict MMRD in gastric cancer samples. For the cases analyzed, the AUC was 0.81 (95% CI: 0.689–0.928, Figure 4C). In addition, the MMRD prediction was consistent with a mutator phenotype known as microsatellite instability (MSI). The latter is characterized by an accumulation of insertions or deletions of small tandem DNA repeats, known as microsatellites [19], and high overall mutation rates. Our model was trained to predict mismatch repair (MMR) status, yet we observed that most of the patients (83%) who were predicted to have MMRD also harbored a MSI phenotype (see Section 2.3). This important result highlights the value of our approach in considering mutational signatures as a surrogate for MMRD during our training tests. Consequently, our model achieved a similar potential in predicting MSI status.

Taken together, our results demonstrate and substantiate the capacity for machine-learning methods to be used in advanced applications to support therapeutic decisions.

2.2. Mutational Signature as a Quantitative Label for DNA Damage Repair Status

While it is acknowledged that targeted genetic assays facilitate the classification of DRD tumors, comprehensive studies which have assessed exome, whole-genome, and sequencing panels have also revealed that mutational signatures are physiological readouts of DNA damage and DNA repair processes based on broader molecular changes in repair pathway components [8,9,20–22]. Therefore, we initially considered groups of samples with high versus low mutational signature activities for a given DNA repair pathway as a ground truth set for specific repair pathways. Mutational signatures were used since histomorphologic footprints are often preceded by harbinger DNA mutations which can be collectively identified in mutational signatures. Our results suggest that these mutational signatures are reflected in the morphological alterations detected by our approach. Moreover, the identification of signatures facilitates the development of a rapid and low-cost method which can be applied to a range of clinically relevant information involving tumor etiology.

2.3. Image Classification Recapitulates Unique Molecular Drivers

After demonstrating that our approach can accurately distinguish underlying features which are predictive of DNA repair status, we examined the molecular drivers involved. In this second stage, we took advantage of comprehensive alterations which are likely to contribute to HRD in breast

tumors [9]. In Figure 5A, the proportion of molecular lesions are shown by type in selected genes of the HR pathway in the breast cancer cases examined [8]. Independent of variants in *BRCA1* (germline biallelic variants, epigenetic silencing, and somatic variants) and *BRCA2* (deletion and germline biallelic variants), other events were detected among HR pathway genes. For example, germline variants in *PALB2* and epigenetic silencing in *RAD51C* were detected. The latter is reflected by low expression levels. Consistent with these findings, our model also correctly identified patients with functional deficiencies incurred from inactivation of genes related to the HR pathway in an independent breast cancer cohort (Figure 5B). Moreover, the molecular profiles obtained are consistent with previous characterizations of HRD mechanisms [10].

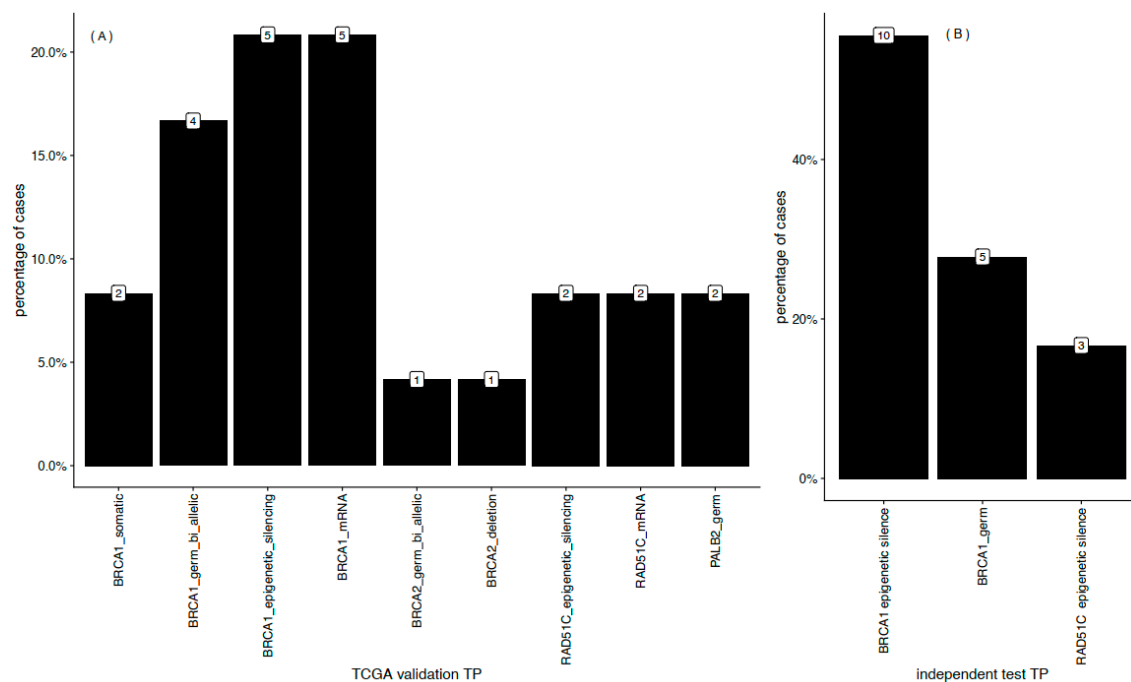


Figure 5. HR (homologous recombination)-associated molecular drivers. The distribution of lesions in known HR-pathway genes of the MMRD predicted patients among the training (A) and independent cohorts (B). Numbers above each bar count the patients for the corresponding lesion.

To evaluate the molecular landscape of gastric cancer patients predicted as MMRD and mismatch repair proficiency (MMRP) by our TCGA-based model, we first examined the absence of *MLH1* with assays of promoter hypermethylation (Figure 6A) and transcriptional downregulation (Figure 6B). We observed a negative correlation between decreased expression of *MLH1* and hypermethylation of its promoter (Figure 6D). Taken together, these results support the hypothesis that epigenetic silencing of *MLH1* functions as a repressive transcriptional signal. The commonly accepted association of *MLH1* and MMR-deficient gastric cancers is also confirmed by these data [23–25]. We further evaluated how mutational activity is modulated in our MMRD and MMRP groups (Figure 6C,E,F). Our analysis revealed high mutational activity in MMRD patients, which is expected when activity of *MLH1*, a protein which plays a critical role in correcting replication errors, is low (Figure 6F). Consistent with previous reports, we also observed higher tumor mutation burden (TMB) in the MMRD group compared to the MMRP group (Figure 6G). Cells with a high degree of TMB correlate with a greater probability of abnormal protein expression which can give rise to mutation-derived antigens (neoantigens) and recruitment of tumor-infiltrating leukocytes. Significant differences were observed between MMRD and MMRP tumors in regard to immune-related gene expression, including genes related to pro-inflammatory markers (Figure 6H). Furthermore, our data support the findings of other studies where MMRD tumors have exhibited increased mutation burden in combination with upregulated immune-related signaling [19].

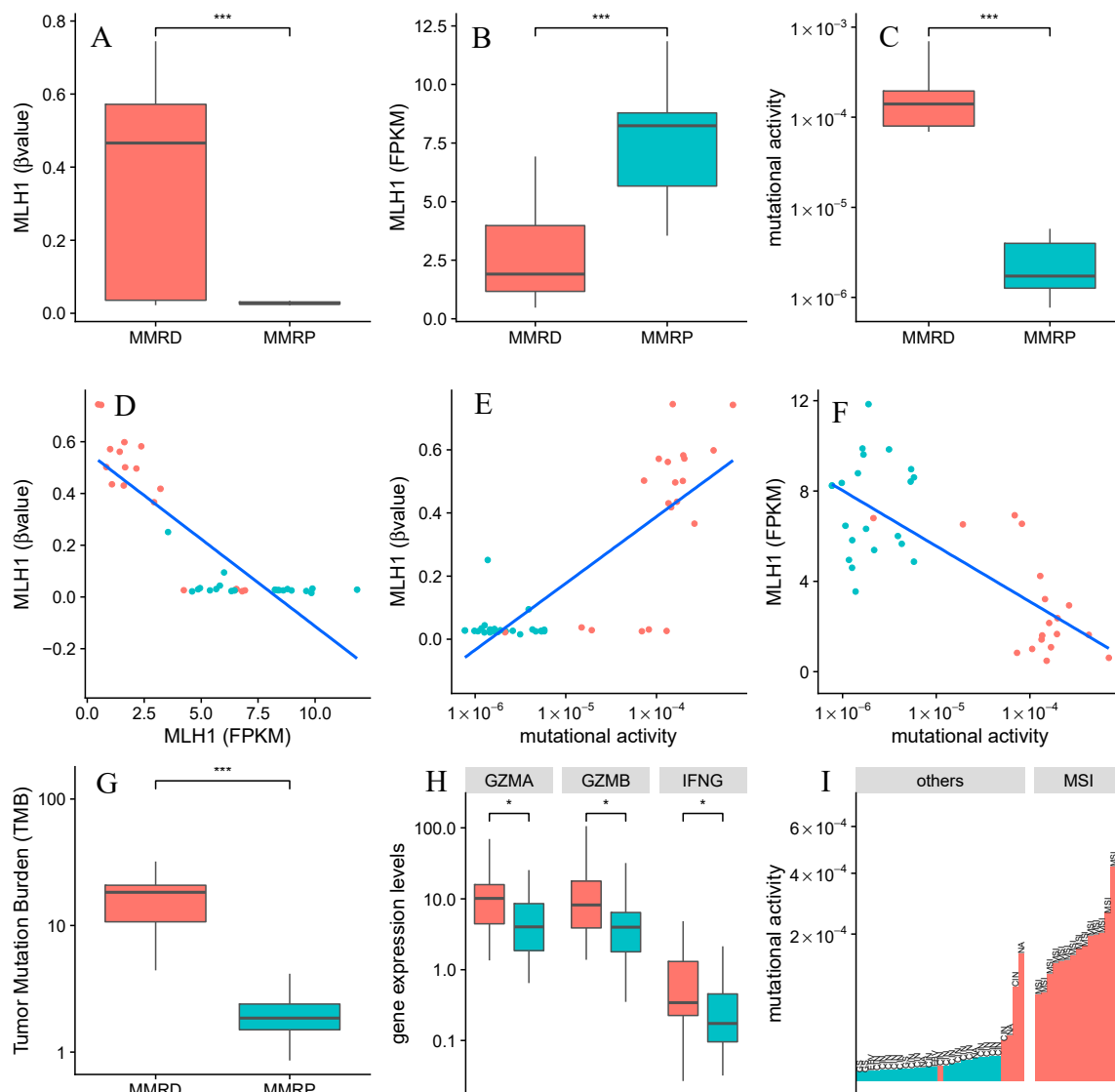


Figure 6. MMR (mismatch repair)-associated molecular drivers. This figure shows the molecular interplay among the MMRD and MMRP predicted groups, which are colored in red and blue, respectively. (A) Changes in methylation of *MLH1* promoter, (B) *MLH1* expression levels and (C) mutational activity of Signature 20 were detected by comparing MMRD and MMRP groups; Scatterplots showing (D) methylation and gene expression levels of *MLH1*, (E) methylation levels of *MLH1* and mutational activity of Signature 20, (F) gene expression levels of *MLH1* and mutational activity of Signature 20; (G) Changes of Tumor Mutation Burden and (H) genes related to pro-inflammatory markers across MMRD and MMRP groups. (I) Mutational activity of Signature 20 across MSI subtype and others molecular subtypes. Statistical significance was assessed by the two-tailed Wilcoxon Rank Sum and it is indicated by * $p < 0.05$, *** $p < 0.001$. The MSI (Microsatellite Instability) and other subtypes (GS: Genomically Stable; EBV: EBV-positive and CIN: Chromosomal INstability) were collected from The Cancer Genome Atlas (TCGA) study.

3. Discussion

We evaluated the feasibility of a weakly supervised deep learning system for assessing DRD from whole slide images of FFPE tissues. We show that our TCGA-derived models are able to detect HR deficiency in breast cancer (AUC = 0.80), and subsequently, we successfully applied the inferred model to a dataset of FFPE tissues from an independent cohort (AUC = 0.71). Our classifier was further applied to predictions of DNA MMR deficiency in gastric cancer patients (AUC = 0.81). These results

show that our model works reasonably well while dealing with DRD. Further, our methods can be used to detect deficiencies more generally in a broader context of other repair deficient pathways.

Classification of patients according to mutational signatures is strongly consistent with well known molecular profiles of repair deficiency components [8,9,20,22]. To further confirm this, we characterized our findings with known genotype-phenotype relationships by integrating data regarding gene expression levels, methylation status, and tumor-infiltrating lymphocyte composition across our test sets. Our findings (Figures 5 and 6) show that patients that are classified as having both HRD and MMRD, harbor the molecular determinants that drive these deficiencies.

Previous studies have demonstrated the potential of digital pathology analysis by using deep learning [13–15]. However, manual segmentation remains a bottleneck step in the analysis of whole slide images. As a result, the development and implementation of this approach in the clinic has been severely hampered. Here, we overcome this shortcoming by using MIL, a machine learning paradigm which includes a full pass through a dataset to detect relevant local patterns in images [26].

Similar to other studies which have used FFPE-based samples, improvements are needed to increase the performance of the methods described [12–15,27]. First, weakly supervised deep learning on whole slide images requires appropriate exclusion of artifacts such as tissue preparation, tissue folds, torn tissue, and pen marks. We addressed this issue by considering Otsu's thresholding methods by extensive experimentation. However, it is possible these methods could be further optimized. Larger training cohorts could also improve classification performance if the network learned from a broader range of morphological variants. Furthermore, improvements in classification accuracy of independent patient cohorts with different staining protocols and whole slide scanners could prevent overfitting.

In conclusion, the deep learning system for DRD assessment described here is a low-cost, screening-based tool which holds extraordinary potential for identifying patients who may be eligible for either genetic counseling or immunotherapy. In addition, this screening tool could identify which patients would derive the most benefit from receiving cancer treatment as early as possible. Our results demonstrate and add strength to the belief that machine-learning methods can be used for advanced applications to assist therapy decisions.

4. Materials and Methods

Classification of whole slides was modeled as previously described [27]. Briefly, whole slides are split into tiles and each tile is classified with a multiple instance learning model. Next, tile-level classifications of each slide are aggregated by using a recurrent network resulting in a single classification for the whole-slide (Figure 1). The following two sections further describe these steps in some detail.

4.1. Slide Pre-Processing

Initially, the available set of histopathological slide images was randomly divided into a training set (70%) and a validation set (30%). Samples with multiple slides were kept together in their respective sets (details in Tables S1 and S2). Next, each slide was resized to 0.5 μm /pixel magnification and split into 224×224 pixel non-overlapping tiles with *libvips* [28] and *openslide* [29]. An *Otsu* thresholding algorithm [30] was calculated for each tile. Tiles with a threshold value < 0.1 were discarded. After image preprocessing, a total of 37,140 tiles (TCGA-BRCA, $n = 32250$; TCGA-STAD, $n = 3940$; ACCC-BRCA, $n = 1950$) were generated.

4.2. Model Definition and Training

Classification of a slide was modeled as a multiple instance binary classification problem [26] (Figure 1), where each slide is a bag of tiles. Positive bags must have at least K positive tiles, while negative bags must have all negative tiles. Given a convolutional model, $M1$, all tiles are classified and ranked according to a $M1$ prediction. The top K tiles of each bag are used to train $M1$,

and this process is repeated until **M1** predictions converge for a given dataset. The result is a model which can distinguish tiles that are relevant to the whole-slide label from the rest.

Training tiles were fed into a resnet34 model [31] pre-trained with the ImageNet dataset [32]. For every epoch, the inference of each tile is calculated and the top **K** tiles are used to train this model according to whole slide labels. Test accuracy is calculated by comparing the prediction of the top tile of each slide with whole slide labels. Classification of a whole slide is modeled with a RNN. RNNs have an internal state which can be used to aggregate a sequence of inputs into a single prediction.

To train the RNN model, **M2**, the best tile-classifier model of the first step (**M1**) is fixed and the top **K** tiles of each slide are fed into the **M1** model. After the intermediate learned features are extracted, these features are fed into **M2** to aggregate all of the features of the top **K** tiles and output a single prediction variable. **M2** is trained with the same dataset that was used in the first step.

The M2 RNN architecture takes a 512-vector as input, which matches the size of the intermediate features output of resnet34, and a state (“memory”) vector. This vector aggregates the features of previous tiles fed into the RNN. The input and the state vector is combined with ReLU activation step resulting in a new state output. The latter is finally fed into another fully connected layer that outputs a single prediction variable.

All of our models were built and trained with PyTorch (available at <https://pytorch.org>). The computational framework for DRD status, including all parameters used in this paper, are freely available and described at <https://github.com/rvalieris/deepHE>. All models were trained by using NVIDIA TESLA V100 GPUs provided by AWS (Amazon Web Services).

4.3. Datasets

Clinical and molecular information for cases of breast cancer (TCGA-BRCA, 1011 patients, 1075 slides) and stomach cancer (TCGA-STAD, 369 patients, 394 slides) were gathered from the Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov/>). H&E-stained histologic images derived from the TCGA-based cohort were retrieved from The Cancer Imaging Archive (TCIA) repository. As an external validation cohort, we retrospectively analyzed archived H&E-stained pathology slides obtained from breast cancer patients treated at the A.C. Camargo Cancer Center (ACCC-BRCA, 65 patients, 65 slides). The HRD status for this independent cohort validation set was assessed by considering pathogenic germline mutations and epigenetic silencing in HR-related genes that were previously detected and reported [18]. Briefly, TNBC samples were characterized as HR proficient or HR deficient according to the presence of *BRCA1/2* germline or somatic pathogenic mutations in DNA from leukocytes and tumors, or *BRCA1* epigenetic silencing analyses of tumor DNA, respectively [18]. Additionally, a *RAD51c* epigenetic silencing analysis of tumor DNA was performed (data not published). Both clinical and molecular data from the patient cohorts analyzed during this study are presented in Supplementary Table S1.

4.4. Biomarker for HRD and MMRD Status

A biomarker score for HR and MMR status was used. Briefly, a pattern of genome-wide mutations (mutational process) known as a mutational signature (https://cancer.sanger.ac.uk/cosmic/signatures_v2) was used. This mutational signature reflected the proficiency of multiple upstream components of repair pathways [8,22]. Mutational signature activities for breast cancer (TCGA-BRCA) and their lesions in HR genes were previously detected and described [8]. For stomach cancer (TCGA-STAD), all somatic single nucleotide variants (SNVs) (e.g., C > A, C > G, C > T, T > A, T > C, and T > G) were mapped onto trinucleotide sequences by including the 5′ and 3′ neighboring base-contexts. Next, the SNV spectrum with 96 types of trinucleotide mutations for all samples was loaded into *signer* [33]. A projection analysis was performed and signature activity across the samples was inferred by using a MMR-associated COSMIC signature. Finally, we performed assignments of per-patient signature activity. Thus, samples with higher mutational signature activity for a given DNA repair

pathway were labeled as a ground truth set for that pathway. We used Signature 3 and Signature 20 as readouts of HRD and MMRD for breast and stomach cancers, respectively.

4.5. Data Availability

The dataset analyzed in this study is described in our Methods and Supplementary Data. Further information regarding this dataset is available from the authors upon request.

5. Conclusions

Our study demonstrates the use of an AI-based workflow capable of detecting DRD-positive tumors directly from H&E-stained histologic images. The method is illustrated by considering the detection of homologous recombination deficiency and mismatch repair deficiency across 1445 cancer patients, including breast and gastric cancer samples. The reliability of the results is shown by evaluating the model in an independent dataset. The success of detecting DRD-status, suggests that deep-learning models can be used to assess other clinically relevant information, that can potentially enhance traditional pathology reports.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-6694/12/12/3687/s1>, Figure S1: Tissue classes learned from Convolutional Neural Network (CNN) in breast cancer.

Author Contributions: R.V. and L.A. performed slide pre-processing and developed and trained the deep learning algorithms used. Acquisition and analysis of data: R.V., L.A., and I.T.d.S. D.M.C. provided FFPE sections and sample-level annotations with regard to genetic lesions in the HR-pathway gene of breast cancer for independent validation. C.A.B.d.T.O. and A.P.B. contributed to the evaluation of tissue slides. E.D.-N., D.N.N., and R.A.R.M. contributed to data interpretation and critical revision of the manuscript. I.T.d.S. conceived and designed the study and also wrote the manuscript with input from the other authors. All of the authors have reviewed this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: I.T.S. was supported by the Amazon Web Services Research Cloud Credits program and a FAPESP grant (#15/19324-6). ED-N was supported by FAPESP grant (#14/26897-0).

Acknowledgments: ED-N is a research fellow from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq–Brazil), and also acknowledges the support given by Associação Beneficente Alzira Denise Hertzog Silva (ABADHS–Brazil). The authors thank Bernardo Peixoto, Eloisa Olivieri, and Nathalia Araújo at the pathologic facility of the A.C. Camargo Cancer Center for whole slide imaging scanning service and guidance. The authors also thank Rafael Canfield Brianese for the *RAD51c* epigenetic silencing analysis.

Conflicts of Interest: The authors declare no conflict of interest

Abbreviations

DNA repair deficiency (DRD), microsatellite instability (MSI), homologous recombination (HR)

References

1. Hakem, R. DNA-damage repair; the good, the bad, and the ugly. *EMBO J.* **2008**, *27*, 589–605. [[CrossRef](#)] [[PubMed](#)]
2. Lord, C.J.; Ashworth, A. The DNA damage response and cancer therapy. *Nature* **2012**, *481*, 287–294. [[CrossRef](#)] [[PubMed](#)]
3. Moynahan, M.E.; Jasin, M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 196–207. [[CrossRef](#)] [[PubMed](#)]
4. Turner, N.; Tutt, A.; Ashworth, A. Hallmarks of ‘BRCAness’ in sporadic cancers. *Nat. Rev. Cancer* **2004**, *4*, 814–819. [[CrossRef](#)] [[PubMed](#)]
5. Lord, C.J.; Ashworth, A. BRCAness revisited. *Nat. Rev. Cancer* **2016**, *16*, 110–120. [[CrossRef](#)]
6. Pujade-Lauraine, E.; Ledermann, J.A.; Selle, F.; Gebbski, V.; Penson, R.T.; Oza, A.M.; Korach, J.; Huzarski, T.; Poveda, A.; Pignata, S.; et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): A double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Oncol.* **2017**, *18*, 1274–1284. [[CrossRef](#)]
7. Hoppe, M.M.; Sundar, R.; Tan, D.S.P.; Jeyasekharan, A.D. Biomarkers for Homologous Recombination Deficiency in Cancer. *J. Natl. Cancer Inst.* **2018**, *110*, 704–713. [[CrossRef](#)]

8. Polak, P.; Kim, J.; Braunstein, L.Z.; Karlic, R.; Haradhavala, N.J.; Tiao, G.; Rosebrock, D.; Livitz, D.; Kübler, K.; Mouw, K.W.; et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat. Genet.* **2017**, *49*, 1476–1486. [[CrossRef](#)]
9. Davies, H.; Glodzik, D.; Morganella, S.; Yates, L.R.; Staaf, J.; Zou, X.; Ramakrishna, M.; Martin, S.; Boyault, S.; Sieuwerts, A.M.; et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat. Med.* **2017**, *23*, 517–525. [[CrossRef](#)]
10. Prakash, R.; Zhang, Y.; Feng, W.; Jasin, M. Homologous recombination and human health: The roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harb. Perspect Biol.* **2015**, *7*, a016600. [[CrossRef](#)]
11. Skrede, O.-J.; De Raedt, S.; Kleppe, A.; Hveem, T.S.; Liestøl, K.; Maddison, J.; Askautrud, H.A.; Pradhan, M.; Nesheim, J.A.; Albrechtsen, F.; et al. Deep learning for prediction of colorectal cancer outcome: A discovery and validation study. *Lancet* **2020**, *395*, 350–360. [[CrossRef](#)]
12. Kather, J.N.; Pearson, A.T.; Halama, N.; Jäger, D.; Krause, J.; Loosen, S.H.; Marx, A.; Boor, P.; Tacke, F.; Neumann, U.P.; et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **2019**, *25*, 1054–1056. [[CrossRef](#)] [[PubMed](#)]
13. Coudray, N.; Ocampo, P.S.; Sakellaropoulos, T.; Narula, N.; Snuderl, M.; Fenyö, D.; Moreira, A.L.; Razavian, N.; Tsirigos, A. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **2018**, *24*, 1559–1567. [[CrossRef](#)] [[PubMed](#)]
14. Couture, H.D.; Williams, L.A.; Geradts, J.; Nyante, S.J.; Butler, E.N.; Marron, J.S.; Perou, C.M.; Troester, M.A.; Niethammer, M. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ Breast Cancer* **2018**, *4*, 30. [[CrossRef](#)]
15. Acs, B.; Rantalainen, M.; Hartman, J. Artificial intelligence as the next step towards precision pathology. *J. Intern. Med.* **2020**, *288*, 62–81. [[CrossRef](#)]
16. Gavande, N.S.; VanderVere-Carozza, P.S.; Hinshaw, H.D.; Jalal, S.I.; Sears, C.R.; Pawelczak, K.S.; Turchi, J.J. DNA repair targeted therapy: The past or future of cancer treatment? *Pharmacol. Ther.* **2016**, *160*, 65–83. [[CrossRef](#)]
17. Eso, Y.; Shimizu, T.; Takeda, H.; Takai, A.; Marusawa, H. Microsatellite instability and immune checkpoint inhibitors: Toward precision medicine against gastrointestinal and hepatobiliary cancers. *J. Gastroenterol.* **2020**, *55*, 15–26. [[CrossRef](#)]
18. Brianese, R.C.; Nakamura, K.D.M.; Almeida, F.G.S.R.; Ramalho, R.F.; Ferreira, E.N.; Barros, B.D.F.; Formiga, M.N.C.; Andrade, V.P.; Lima, V.C.C.; Carraro, D.M. BRCA1 deficiency is a recurrent event in early-onset triple-negative breast cancer: A comprehensive analysis of germline mutations and somatic promoter methylation. *Breast Cancer Res. Treat.* **2018**, *167*, 803–814. [[CrossRef](#)]
19. Zhao, P.; Li, L.; Jiang, X.; Li, Q. Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy. *J. Hematol. Oncol.* **2019**, *12*, 54. [[CrossRef](#)]
20. Gulhan, D.C.; Lee, J.J.-K.; Melloni, G.E.M.; Cortés-Ciriano, I.; Park, P.J. Detecting the mutational signature of homologous recombination deficiency in clinical samples. *Nat. Genet.* **2019**, *51*, 912–919. [[CrossRef](#)]
21. Secrier, M.; Xiaodun, L.; de Silva, N.; Eldridge, M.D.; Contino, D.; Bornschein, J.; MacRae, S.; Grehan, N.; O'Donovan, M.; Miremedi, A.; et al. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat. Genet.* **2016**, *48*, 1131–1141. [[CrossRef](#)] [[PubMed](#)]
22. Van Hoeck, A.; Tjoonk, N.H.; van Boxtel, R.; Cuppen, E. Portrait of a cancer: Mutational signature analyses for cancer diagnostics. *BMC Cancer* **2019**, *19*, 457. [[CrossRef](#)] [[PubMed](#)]
23. Fleisher, A.S.; Esteller, M.; Wang, S.; Tamura, G.; Suzuki, H.; Yin, J.; Zou, T.T.; Abraham, J.M.; Kong, D.; Smolinski, K.N.; et al. Hypermethylation of the hMLH1 gene promoter in human gastric cancers with microsatellite instability. *Cancer Res.* **1999**, *59*, 1090–1095. [[PubMed](#)]
24. Leung, S.Y.; Yuen, S.T.; Chung, L.P.; Chu, K.M.; Chan, A.S.; Ho, J.C. hMLH1 promoter methylation and lack of hMLH1 expression in sporadic gastric carcinomas with high-frequency microsatellite instability. *Cancer Res.* **1999**, *59*, 159–164. [[PubMed](#)]
25. Kwak, Y.; Seo, A.N.; Lee, H.E.; Lee, H.S. Tumor immune response and immunotherapy in gastric cancer. *J. Pathol. Transl. Med.* **2020**, *54*, 20–33. [[CrossRef](#)]
26. Dietterich, T.G.; Lathrop, R.H.; Lozano-Perez, T.; Pharmaceutical, A. Solving the multiple-instance problem with axis-parallel rectangles. *Artif. Intell.* **1997**, *89*, 31–71. [[CrossRef](#)]

27. Campanella, G.; Hanna, M.G.; Geneslaw, L.; Miraflor, A.; Silva, V.W.K.; Busam, K.J.; Brogi, E.; Reuter, V.E.; Klimstra, D.S.; Fuchs, T.J. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **2019**, *25*, 1301–1309. [[CrossRef](#)]
28. Martinez, K.; Cupitt, J. VIPS—A highly tuned image processing software architecture. In Proceedings of the IEEE International Conference on Image Processing 2, Genoa, Italy, 11–14 September 2005; pp. 574–577.
29. Goode, A.; Gilbert, B.; Harkes, J.; Jukic, D.; Satyanarayanan, M. OpenSlide: A vendor-neutral software foundation for digital pathology. *J. Pathol Inf.* **2013**, *4*, 27.
30. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [[CrossRef](#)]
31. Deep Residual Learning for Image Recognition. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016.
32. ImageNet: A large-scale hierarchical image database. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009.
33. Rosales, R.A.; Drummond, R.D.; Valieris, R.; Dias-Neto, E.; da Silva, I.T. Signer: An empirical Bayesian approach to mutational signature discovery. *Bioinformatics* **2017**, *33*, 8–16. [[CrossRef](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Apêndice D – Artigo HER2 – Valieris 2024

Segue abaixo o artigo publicado na revista *BMC Breast Cancer Research* em agosto de 2024, contendo as análises dos modelos preditores de **HER2**, descrito na Seção 4.2.

RESEARCH

Open Access



Weakly-supervised deep learning models enable HER2-low prediction from H &E stained slides

Renan Valieris^{1†}, Luan Martins^{1,7†}, Alexandre Defelicibus¹, Adriana Passos Bueno^{1,2}, Cynthia Aparecida Bueno de Toledo Osorio², Dirce Carraro³, Emmanuel Dias-Neto^{4,5}, Rafael A. Rosales⁶, Jose Marcio Barros de Figueiredo^{1†} and Israel Tojal da Silva^{1*†}

Abstract

Background Human epidermal growth factor receptor 2 (HER2)-low breast cancer has emerged as a new subtype of tumor, for which novel antibody–drug conjugates have shown beneficial effects. Assessment of HER2 requires several immunohistochemistry tests with an additional in situ hybridization test if a case is classified as HER2 2+. Therefore, novel cost-effective methods to speed up the HER2 assessment are highly desirable.

Methods We used a self-supervised attention-based weakly supervised method to predict HER2-low directly from 1437 histopathological images from 1351 breast cancer patients. We built six distinct models to explore the ability of classifiers to distinguish between the HER2-negative, HER2-low, and HER2-high classes in different scenarios. The attention-based model was used to comprehend the decision-making process aimed at relevant tissue regions.

Results Our results indicate that the effectiveness of classification models hinges on the consistency and dependability of assay-based tests for HER2, as the outcomes from these tests are utilized as the baseline truth for training our models. Through the use of explainable AI, we reveal histologic patterns associated with the HER2 subtypes.

Conclusion Our findings offer a demonstration of how deep learning technologies can be applied to identify HER2 subgroup statuses, potentially enriching the toolkit available for clinical decision-making in oncology.

Keywords Breast cancer, HER2, Artificial intelligence, Digital pathology

[†]Renan Valieris, Luan Martins, Jose Marcio Barros de Figueiredo and Israel Tojal da Silva contributed equally to this work.

⁷ Institute of Mathematics and Computer Sciences, Universidade de São Paulo, São Carlos, São Paulo 13566-590, Brazil

*Correspondence:

Israel Tojal da Silva

itojal@accamargo.org.br

¹ Laboratory of Computational Biology and Bioinformatics, CIPE/A.C.Camargo Cancer Center, São Paulo, São Paulo 01508-010, Brazil

² Department of Pathology, CIPE/A.C.Camargo Cancer Center, São Paulo, São Paulo 01508-010, Brazil

³ Laboratory of Genomics and Molecular Biology, CIPE/A.C.Camargo Cancer Center, São Paulo, São Paulo 01508-010, Brazil

⁴ Laboratory Medical Genomics, CIPE/A.C.Camargo Cancer Center, São Paulo, São Paulo 01508-010, Brazil

⁵ Division of Cancer Biology, Department of Radiation Oncology, Rutgers New Jersey Medical School, Newark, NJ 07103, USA

⁶ Departamento de Computação e Matemática, Universidade de São Paulo, Ribeirão Preto, São Paulo 14040-901, Brazil



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Breast cancer (BC) is the most frequently diagnosed malignancy in women and the second most common cause of death from cancer among women worldwide [1]. BC is a heterogeneous disease with two main morphological subtypes: ductal carcinoma *in situ* (DCIS) and invasive ductal carcinoma (IDC). The latter accounts for 70–80% of cases [2]. The morphological assessment of BC is perhaps the main prognostic factor and is still indispensable in routine practice [3]. The additional evaluation of the estrogen (ER) and progesterone (PR) hormone receptors coupled with human epidermal growth factor receptor 2 (HER2) categorizes BC into the following clinically relevant immunophenotypes: (i) luminal A (LUMA), (ii) luminal B (LUMB), (iii) triple-negative breast cancers (TNBC), and (iv) HER2-enriched (HER2+).

The assessment of the proliferative index (Ki-67) together with ER, PR, and HER2 status is currently determined by immunohistochemistry (IHC)-based methods followed by visual evaluation by a pathologist [3]. Several research efforts have explored gene expression-based signatures for specific BC stratification [4], informing the selection of precisely targeted intervention therapies. For example, a clinically applicable gene expression assay using 50 genes (PAM50) that classifies breast cancer into four molecular intrinsic subtypes shown above has been reported [5].

In particular, the HER2 assessment has been reshaped by considering the first Food and Drug Administration approved targeted therapy for a new classification of the HER2 subtype called HER2-low [6]. Usually, the assessment of HER2 protein overexpression on the surface of cells is evaluated by IHC using a well known four-tier scoring system of 0 to 3+ (0: negative; 1+: negative; 2+: equivocal and 3+: positive) [7]. The HER2-low subtype includes tumors with IHC scores of 1+ or 2+ without amplification by an additional *in situ* hybridization (ISH)-based test [6]. Although the 2+ assessment by ISH is the most accurate, it is more expensive, and obtaining the results takes more time, thus remaining a bottleneck step in current clinical practice in most hospitals [8]. This justifies efforts for the development of novel approaches that aim to quantify HER2-low and should benefit patients with a faster and more accurate classification method that could better guide therapeutics.

Although the assessment of most biomarkers in BC is based on expanded IHC and gene expression profiling analyses, recent advances in deep learning (DL) have the potential to complement traditional assays [9]. Motivated by the ability to learn from underlying features directly from hematoxylin and eosin (H & E) stained histopathological whole slide images (WSIs), we and others have

developed DL-based systems to predict diverse molecular phenotypes including mutations in non-small cell lung [10], microsatellite instability (MSI) in patients with gastrointestinal cancer [11], homologous recombination deficiency (HRD) [12] and other clinical biomarkers [13]. These studies further emphasize the ability of machine learning methods to learn underlying morphologic features that mirror the expression of molecular markers.

Although the relevance of machine learning for BC applications has been investigated [14–21], most models have been trained on images of tissue microarrays, IHC-stained slides and fluorescence *in situ* hybridization (FISH)-images. Furthermore, the classifiers were used to predict HER2 either within the hormone receptor status or intrinsic subtype, in which HER2 status was classified solely as positive or negative. Most importantly, in light of treatment options available for HER2-low patients, the assessment of HER2-low tumors by DL models remains to be explored.

In this study, we built a series of deep learning models to examine different training datasets by combining quantitative scores from either IHC tests or ISH assays for HER2-low assessment in BC patients. Moreover, we explored the potential of explainable AI (xAI) mechanisms for providing insights into the underlying causes of the model's decisions. Lastly, we assembled a novel online dataset, containing more than 2 million H & E image patches.

Methods

In total, data were gathered from 1437 H & E-stained WSI of 1351 patients across three multicentric cohorts. These included an in-house cohort (ACCCC, N=546) and two public repositories (TCGA, N=535 and HEROHE, N=356).

ACCCC cohort

We used an anonymized dataset from patients treated for BC at A.C. Camargo Cancer Center (ACCCC, N = 504) between 2019 and 2021. The HER2 tumor status within the ACCCC cohort was determined in accordance with standard procedures outlined by the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) guidelines. In total, we obtained and digitized 546 H & E-stained WSI at 0.25 $\mu\text{m}/\text{px}$ magnification from both biopsy and resection tissue samples using a Leica Aperio AT2 scanner. This study received ethical approval (#3028/20) from the ethics committee of the Fundação Antônio Prudente, with all cases with metastasis or neoadjuvant treatment were excluded. Supplementary Table 2 presents the detailed histological characteristics of our in-house cohort.

HERO cohort

The HERO dataset was downloaded from a publicly available database, published by The HEROHE Challenge [22, 23], which aimed to create new tools for predicting HER2 status in H & E-stained WSI from biopsy samples of patients with invasive breast cancer. The WSI are available at <https://ecdp2020.grand-challenge.org>. These slides were scanned at Ipatimup Diagnostics using a 3D Histech Panoramic 1000 digital scanner at 0.25 $\mu\text{m}/\text{px}$ magnification, and saved in the MIRAX file format [23]. The HERO dataset is composed of 509 slides from distinct patients and after an initial background tile filtering were reduced to 356 slides. Details about the distribution of IHC scores and HER2 status are summarized in Table 1, which were classified by experienced pathologists according to the latest ASCO/CAP guidelines [23].

TCGA-BRCA cohort

A total of 1072 H & E stained WSI at 0.25 $\mu\text{m}/\text{px}$ magnification from 1063 patients was retrieved from the GDC portal TCGA-BRCA project, which contains clinical information and is available at <https://portal.gdc.cancer.gov/>. Only high-resolution H & E stained WSI were selected for analysis, excluding those with artifacts such as significant pen marks, tissue folds, or blurriness, we also only selected slides that had HER2 test results available in the clinical data, after filtering a total of 535 slides remained for the analysis. The HER2 status in TCGA was determined using both the IHC score and FISH status.

Ground truth for HER2 status

All models were trained with ground truth data derived from a combination of IHC tests and DDISH/FISH tests. Specifically, the HER2 protein expression is initially assessed using an immunohistochemistry assay, which involves staining breast cancer tissue samples with antibodies specific to HER2. The staining intensity and pattern are then evaluated by a pathologist resulting in four possible scores (0: negative; 1+: negative; 2+: equivocal and 3+: positive). In case of an equivocal result, an additional in-situ hybridization test is performed either by Fluorescence in situ hybridization (FISH) or dual-color dual-hapten brightfield in situ hybridization (DDISH).

Table 1 Distribution of slides across the three cohorts and HER2 groups

Cohort	HER2-neg	HER2-low	HER2-high	Total
ACCCC	196	217	133	546
HERO	23	187	146	356
TCGA-BRCA	40	377	118	535
Total	259	781	397	1437

For the purposes of our study, all cases were subdivided into three groups according to their IHC score and ISH status: HER2-neg, HER2-low, and HER2-high, as described in Table 2.

Data preprocessing

In the pre-processing phase, we developed a model to remove background and detect tumor regions, aiming to segment informational tissue regions from the WSI. We employed the following procedures to achieve this goal. First, we trained a ConvNeXt-Tiny model [24] with the *NCT-CRC-HE-100K* dataset [25], which contains examples of background tiles. Applying this model on the TCGA-BRCA, we constructed a set of background region annotations. Next, we used annotations from the Manual Tumor annotations dataset [26] to create a set of non-tumor annotations from the regions in the slide that are neither tumor nor background. Finally, we trained a combined model (tumor-background) on the entirety of TCGA-BRCA slides to identify tumor, non-tumor, and background regions based on these annotation sets. The prediction of the tumor-background model is illustrated in Supplementary Figure 1.

To process the slides prior to the MIL model training, each slide was divided into 256 px $\times$ 256 px tiles at 0.5 $\mu\text{m}/\text{pixel}$ magnification, if the original slide had higher magnification a corresponding down-scaling transformation was applied, with adjacent tiles having a 50% overlap. Next, the tumor-background model was used to detect background regions, tumors, and non-tumors in all cohorts. Only tiles identified as tumor were used to train the MIL model, slides with less than 1000 tumor tiles detected were excluded from further analysis.

Feature extraction

Following the color normalization procedure by [27], we normalized slides from all cohorts to a reference image from the ACCCC dataset. Next, a 1024-feature embedding vector was extracted from each color-normalized tile by using the *UNI* model. This is a general-purpose self-supervised model for pathology, trained with more than 100 million images from over 100,000 diagnostic H & E-stained WSIs across 20 major tissue types [28]. The 1024-feature vector was used in the following steps to train the MIL model.

MIL model definition

We use a weakly supervised multiple instance learner (MIL) classifier, which takes an H & E slide in the format of a bag-of-tiles as input data and classifies it into one of the desired classes using an attention-based mechanism. The MIL architecture considered throughout was based

on clustering-constrained-attention multiple-instance learning (CLAM) [29].

Briefly, the CLAM model considers the features of a slide as a bag of tiles and classifies the slide into one of the target classes. Internally, it works by modeling each HER2 group independently with an attention layer and a classification layer. The attention layer is responsible for identifying important tiles and combining them into a single 1024-feature vector for the slide based on the perceived importance of its tiles, while the classification layer takes this representation vector and outputs a single activation value: higher outputs represent a higher likelihood of the slide belonging to the class. Finally, this independent value for each HER2 group is then combined in the output vector using a *softmax* activation function, whose maximum value describes the slide's HER2 group. Additionally, the model also employs an *instance classifier* whose goal is to cluster the tiles of each class. During training, the *instance classifier* acts as a regularization method that biases the attention backbone into producing relevant features for both the attention mechanism and itself; however, once the model completes the training stage, it can be used as a tile classifier that identifies tiles that are associated with its target HER2 group.

Model training

The ACCCC and HERO cohorts were combined into a single dataset and 15% of the cases were separated for a hold-out test set. The remaining 85% of the cases were further split into 10 cross-validation folds where cases were randomly assigned to a training (85%) or validation set (15%) while ensuring that the label distribution and slide dataset proportion of both sets remained unchanged. All models were trained independently for each fold using the same set of fixed hyper-parameters with the goal of evaluating the model variance on different data. The TCGA-BRCA cohort was used as an external test dataset.

We tested six models, each considering a different subset of cases with assay results of HER2 status analyzed at the protein level by IHC and ISH techniques (Table 2). Each model was trained multiple times with a 10-fold cross-validation procedure. All models were trained for a total of 200 epochs with the early stopping criterion set to the loss of the validation set. To speed up training and diminish GPU memory requirements, we divided the training procedure into two stages: i) feature extraction and ii) CLAM classifier training. This approach has the advantage of allowing more models to be trained in parallel due to the lower memory and processing requirements of training each model. All models were trained using an NVIDIA A100 40 GB GPU.

We employed the CLAM's originally proposed loss function as the optimization objective, which is a two-part linear combination consisting of a 'bag' loss and an 'instance' loss. The 'bag' loss, a Cross-Entropy loss, is weighted by 0.7, while the 'instance' loss is a Smooth SVM Loss. The training process utilized the ADAM optimizer with a fixed learning rate of 0.0001 and weight decay of 0.0001. The resulting model has 2.1M parameters.

Analysis of actionable tiles

Actionable tiles can be identified by the CLAM's instance layer, which is a set of binary classifiers available for each class (in our case, HER2 groups). Its goal is to identify whether a given tile belongs to its specific class. For each slide in the internal test set, we classify the tiles with the instance layer, selecting those with a predicted probability above 0.5 for their respective class. The top 15 tiles with the highest attention value of each slide identified by the instance classifier are selected as actionable tiles. In total, we sampled 100,280 actionable tiles across all HER2 groups and models, regardless of the final whole-slide classification. Hence, our goal is to understand what features were learned as being relevant for each group.

To analyze the actionable tiles, we propose a visualization approach based on a 2D histogram of the latent feature space. First, we reduce the dimensionality of the tile feature vector with PCA, projecting the first two principal components of the features onto two dimensions. Then, we discretize this space into 30×30 bins, with each of the bins acting as a centroid around the sampled tile's embeddings, clustering the tiles with similar features. This approach allows us to characterize the continuous space of tile features by coloring the bins based on different criteria. For instance, bins can be colored using the HER2 group that appears more frequently, or we can color them only when two or more models use their corresponding feature. This approach allows us to organize and understand the space of actionable histological features.

Results

The purpose of this study was to develop a deep learning model for the assessment of HER2-low tumors. To this end, we develop a computational pipeline involving the following three steps (Fig. 1):

1. Tiles were extracted from whole-slide images (WSI) and subsequently classified into tumors and nontumor groups (Fig. 1B, C).
2. The generated bag of tiles for the tumor and nontumor groups were then used for feature extraction with the UNI model [28] (Fig. 1D).

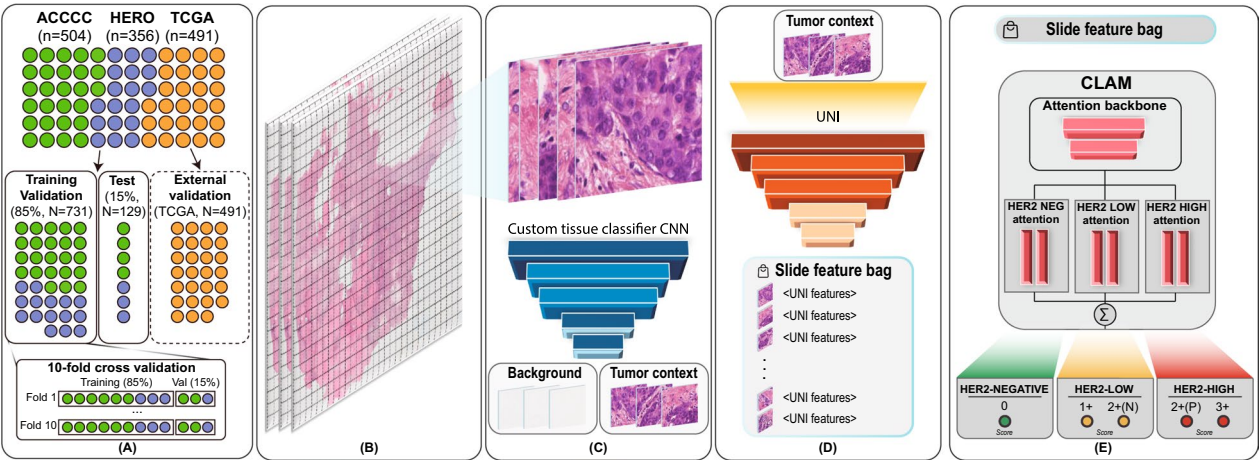


Fig. 1 Workflow of the proposed method. **A** Datasets (ACCC, HERO, TCGA) are split by patients (N) into four categories: training and validation (85%, ACCC+HERO); test (15%, ACCC+HERO); external validation (TCGA). **B** Whole-slide Images are split into tiles; **C** regions of interest are identified; **D** features are extracted from tile images; **E** the CLAM model is trained to classify tile sets into HER2-negative, HER2-low, and HER2-high groups

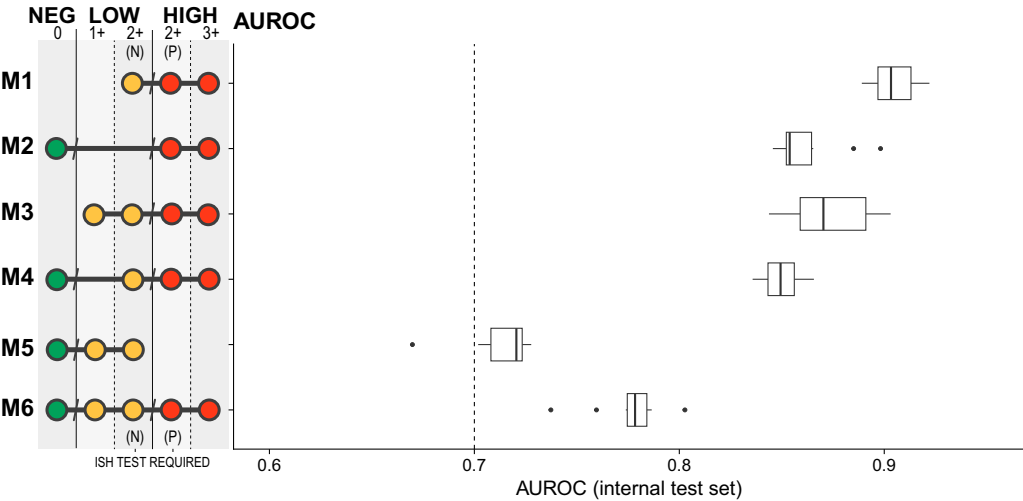


Fig. 2 AUROC of the 10-fold cross-validation results for the test set. Each model was trained with its own set of HER2 scores, i.e., each target HER2 group. The HER2 scores used are indicated by the dots, while the dot's color indicates its target HER2 group for the model

Table 2 Data partitioning scheme of each model

	HER2-neg		HER2-low		HER2-high	
	0		1+	2+	3+	
IHC						
ISH				Neg	Pos	
M1				•	•	•
M2	•			•	•	•
M3			•	•	•	•
M4	•		•	•	•	•
M5	•		•	•	•	•
M6	•		•	•	•	•

Data partitioning scheme of each model. We divided the cases into the following 3 groups according to the results of the IHC and ISH status: HER2-neg, HER2-low, and HER2-high

3. The resulting bags of features were used to train a multiple instance learning (MIL) model for classifying the WSI into HER2-negative, HER2-low, and HER2-high classes (Fig. 1E).

Model performance reveals that classifying HER2-low is challenging

We built six distinct models to explore the ability of classifiers to distinguish between the HER2-negative, HER2-low, and HER2-high classes in different scenarios. The detailed trained set used in these models is summarized in Table 2 and the performance is shown in Fig. 2.

Specifically, while the M6 model was trained to classify the 3 classes with the whole dataset, we also trained models with subsets; for example, M1 was trained with only HER2 2+ and 3+ cases, to detect HER2-low on the HER2 2+ subgroup without the need for a confirmatory additional ISH-based test.

To assess the learning variability of these models we performed a 10-fold cross-validation on each of the subsets tested, the hold-out internal test AUC for each fold can be seen on boxplot variance as shown in Fig. 2. ROC curves for each model and fold are included in the Supplementary Figure 2. Precision-Recall curves are shown in Supplementary Figure 3.

We assessed the detection of HER2-low in HER2 2+ results based on the IHC-based assay results (M1). In this model, the training dataset includes results where there is an inherently subjective and high interobserver agreement, which guarantees a good discriminatory capacity of this classifier. Notably, this model has great potential for use in a cost-effective manner, since HER2 2+ IHC results require confirmation by an additional assay based on hybridization methods. Next, we sought to evaluate the ability to discriminate groups of patients with HER2 0 (negative, scoring 0) vs. HER2 2+ (with scoring 2+ with additional positive ISH results and scoring 3+ being equivalent). In fact, when considering that molecular changes precede and results altered cellular histology, the M2 model was able to detect groups of patients with a basal HER2 expression (without HER2 amplification) as distinct from those with high expression. Therefore, it is reasonable to expect great performance from this model as compared with the former model. Notably, by considering either groups labeled as HER2-low (HER2 1+ and HER2 2+ with additional ISH negative), we noted a performance gain to discriminate HER2-low. However, the M4 and M5 models show a performance drop when we either include the HER2 0 and subtract the HER2-high (HER2 2+ with additional ISH positive and HER2 3+) groups. Although we are using labels ground truth results from well-standardized clinical practice, these results suggest a possible confounding effect on the trained classes when considering the HER2 1+ group. This is a result of the subjectivity and slight difference in the results (low interobserver agreement) obtained from the IHC assays by the experts of the HER2 0 and 1+ groups [30], which consequently affects the performance of the models since the results of the molecular tests are considered in the division of the training and testing groups of the models. Although building a model to classify samples into 3 groups is challenging, we observed that the M6 (ideal) model shows a discriminatory power to predict HER2 group directly from H & E histology. The worst performance is obtained from the fM5 model, which

must detect HER2-low type tumors among the HER2 0, 1+, and 2+ groups with negative ISH results; ultimately, this demonstrates a limitation of the model when considering results from assays where there is discordance in distinguishing HER2 0 from HER2 1+ [30, 31].

Importantly, HER2 assessment protocols are continuously updated according to the recommendations of the American Society of Clinical Oncology/College of American Pathologists (CAP) [32] and thus gathering high-quality datasets with consistent labeling of the HER2 groups (neg, low and high) is challenging, this hampers the evaluation of the models capability to generalize to independent datasets. Indeed, we used the TCGA breast cancer data for HER2 prediction as an external cohort. In particular, we observed the same trend for all models built considering the distinct datasets (Supplementary Table 1, Supplementary Figure 4). However, since the TCGA-BRCA dataset used a different set of guidelines for HER2 assessment, we observed a drop in performance for all models, as expected. Furthermore, the TCGA database aggregates image datasets from numerous institutions, each employing varied protocols for sample preparation and image capture. Such technical diversity also affects classification performance [19].

Altogether, our findings suggest that the performance of classifiers is dependent on the reproducibility and reliability of assay-based tests for HER2 since the results of these assays serve as the ground truth for model training.

Actionable tiles reveal patterns used by the classifier

A key step in deep learning in a clinical setting is interpreting how the generated models work, thus allowing us to evaluate their validity and trustworthiness. To understand which histological patterns are used by the model to classify the HER2 classes, we need to examine which features are being used by the MIL classifier. To this end, we considered the attention layers provided by the MIL algorithm [33]. These layers direct the classifier to a subset of tiles with features that allow for the classification (Fig. 3). We focus our analyses on tiles that are given high levels of attention, denoted as actionable tiles, due to their importance on the final result of the classifier.

The resulting representation revealed patterns across different models and inside the folds for the same model (Fig. 4). The analysis of different folds by the same model reveals the occurrence of recurring patterns, indicating that tiles of similar features in these areas are being used for the classification of the same class.

Tiles classified into a given HER2 class tend to form clusters on the two dimensional principal component space. Tiles from these clusters share morphological characteristics that define each class. A sample of

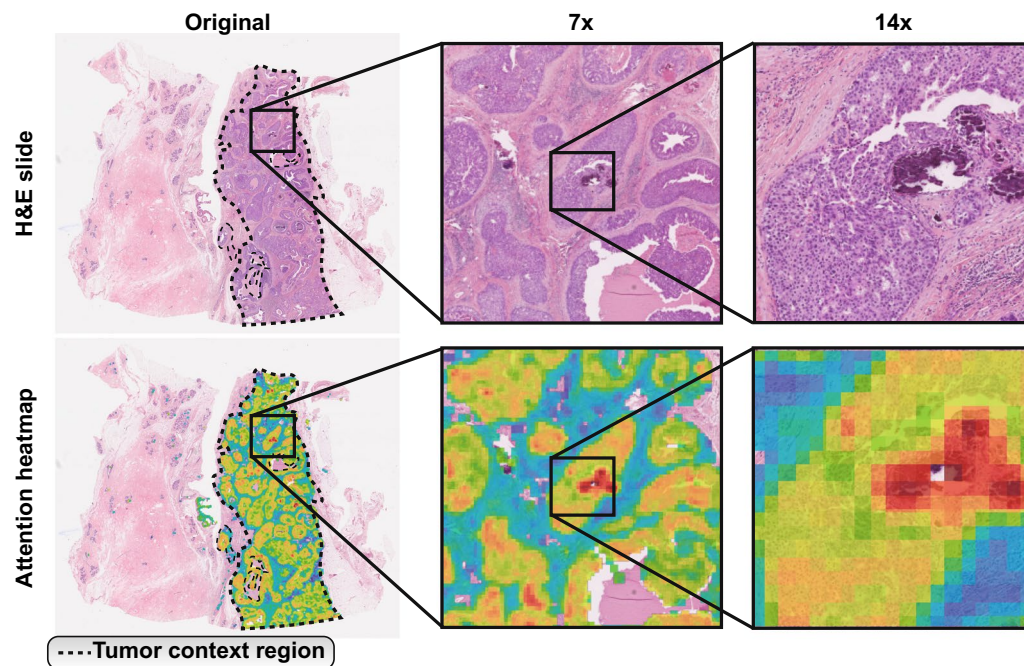


Fig. 3 H & E-stained slide of a tissue sample and its corresponding attention heatmap generated by the proposed model. Attention maps highlight meaningful patches at different magnifications. The heatmap shows the areas where the model focuses more of its attention, with warmer colors indicating higher attention scores. The dotted line shows the region identified as tumor

the image tiles of each such cluster in Fig. 4C–D supports this (with more tiles examples in Supplementary Figure 5).

While histopathological evaluation of tissue sections stained with hematoxylin and eosin (H & E) is indispensable for the management of breast cancer, the relationship between cell pattern and HER2 subgroups is still unclear. We thus examined the cellular patterns of the actionable tiles associated with the HER2 groups. We observed actionable tiles in HER2-high with the presence of hyalinized stroma with minimal inflammatory infiltrate, possible isolated neoplastic cells, and various distinct histologic patterns including cohesively patterned neoplasia with lobular infiltration patterns of nuclear grades 1/2, and structural patterns that resemble micro-papillary features. This subtype also exhibited tiles with nuclear grades 2/3 but predominantly grade 2, and rare instances of mitotic activity. Conversely, we noted actionable tiles in HER2-neg presented nuclear grade 3 and were enriched for adipose cells. Importantly, this adipose microenvironment could be associated with motility and invasiveness of breast cancer cells [34]. Reinforcing this observation, we noticed some tiles with muscular tissue infiltrated by isolated neoplastic cells, a feature that can be associated with aggressiveness since muscular infiltration of thoracic wall muscles characterizes advanced-stage tumors. Putting it all together, the presence of

apocrine features, higher nuclear grade and a more infiltrative pattern indicate a more aggressive tumor, often associated with triple-negative breast cancers (TNBC). Finally, actionable tiles in HER2-low were enriched for chronic inflammatory cells, fibrotic tissue, and predominantly nuclear grades 1 and 2. These histology findings provide a nuanced understanding of the heterogeneity within breast cancer, underlining the importance of explainable visualizations for elucidating how certain spatial pathology features correlate with HER2 groups and specific diagnoses.

Filtering tiles of interest guides training

Given a tumor identified as HER2-negative through an IHC stain, it is reasonable to infer that its corresponding H & E WSI lacks discernible features indicative of HER2-positivity. Conversely, in cases where the patient is identified as HER2-positive, it is reasonable to assume that at least certain regions within their H & E WSI exhibit features relevant to HER2-positivity. Consequently, models can be trained using WSI stains as input data and IHC annotations as data labels. Hence, this problem formulation aligns effectively with the principles of multiple instance learning (MIL) [35]. However, as digital slides may incorporate distinctive artifact patterns resulting from the staining protocol or digitalization process,

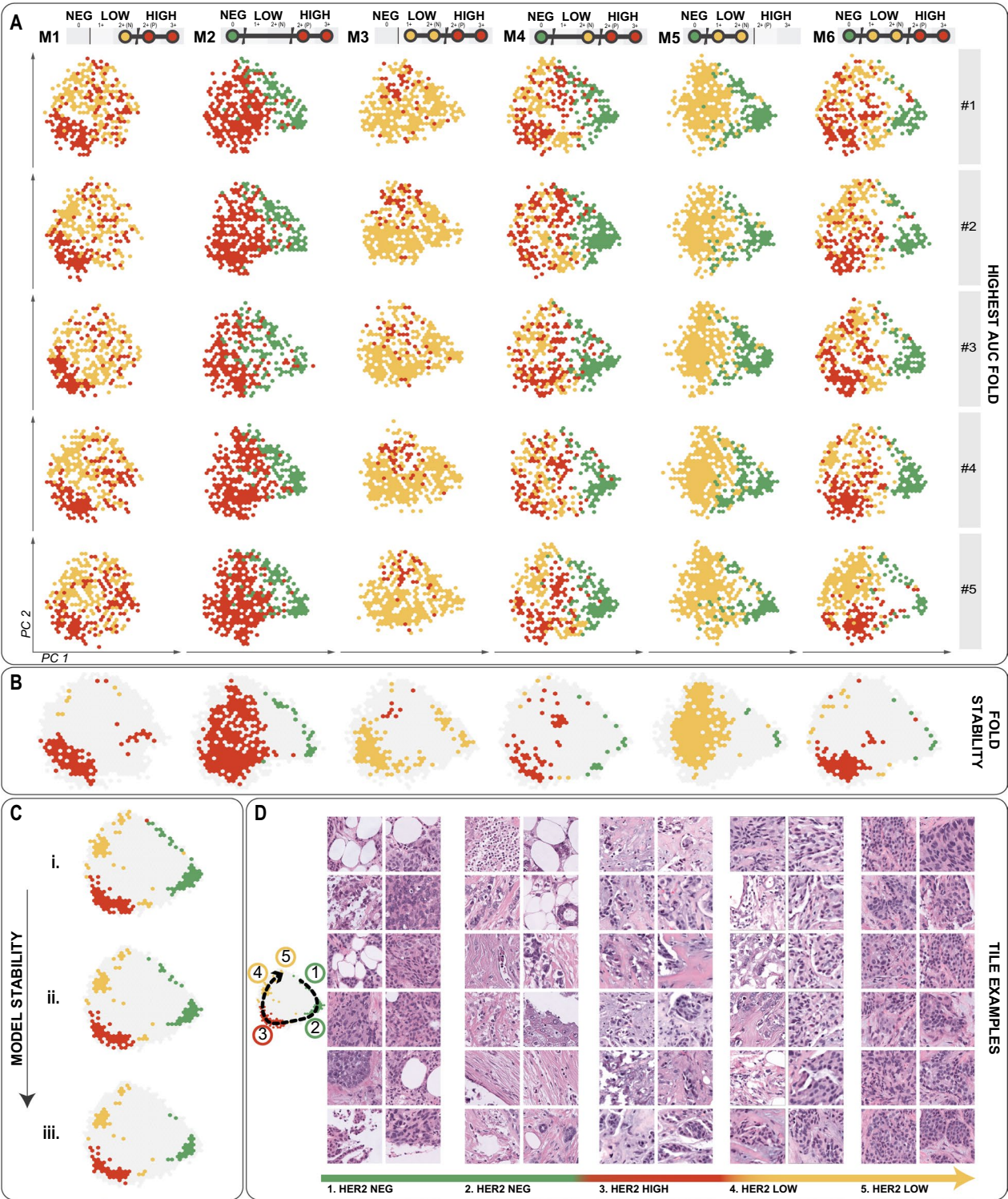


Fig. 4 Model explicability analysis. **A** Tile embedding is plotted using the PCA results for each model and fold, only coloring regions according to the HER2 group classification that relies on attention to that embedding. **B** Colored bins indicate features used by 50% of the folds within each model. **C** The consistency of the models highlights how many models use the same embedding for classifying each HER2 group. There are three levels of agreement: i. 2 or more; ii. 3 or more; iii. 4 or more. **D** Examples of tiles from the embedding space for each class

which might be captured by the MIL algorithm, we conduct background detection and removal of tiles from the dataset prior to MIL model training. This is undertaken to mitigate the potential impact of such tiles and minimize any spurious effects they may introduce.

Importantly, one of the most common methods for background removal is a thresholding-based algorithm such as the Otsu method [36]. However, such methods are not able to detect other types of artifacts such as out-of-focus regions and colored pen markings, which are present in some datasets, including TCGA dataset.

To investigate whether the inclusion of specific subsets of tiles impact the results of the MIL model, we developed a neural-network method to identify and filter background, tumor, and non-tumor regions, as described in the Data preprocessing section. We illustrate the effect of this filter by training models with different subsets and evaluating the performance of each model (Table 3, Supplementary Figure 6). First, we attempted to filter only background tiles (NOBG), and we found that the model is able to distinguish tissue-specific patterns; however, we also found that the model learns to give high levels of attention to tiles showing the tumor environment, which prompted us to test the models using only tiles from the tumor region. We refer to this filtering configuration as *tumor only* (TUMO), and we found that by using the tumor environment, we were able to achieve the best balance between model performance and training speed, as the models only need a fraction of tiles (Supplementary Figures 7 and 8) to achieve similar performance to models using the full set (Supplementary Figure 6).

Our analysis indicates that models built only with tumor tiles (TUMO) performed similarly to models that used all tiles except background (NOBG). This suggests that by filtering the tiles, the model can focus on the most important areas, lowering the risk of attending spurious unique patterns such as artifacts or background noise. We chose to use TUMO because it strikes the optimal

balance between model predictive performance and hardware requirements.

Comparison of alternative implementations

We compared the results from the CLAM model with two other implementations of the MIL method: ADMIL [33] and TransMIL [37], the results are summarized in Table 4, Supplementary Image 9. The methods presented comparable performance in most models, however CLAM showed better overall performance. Importantly, only the CLAM method permits multi-class predictions, while the ADMIL and TransMIL method implements only binary prediction, thus we choose to employ the CLAM method.

We also compared the use of the *UNI* feature extractor with a *ResNet-50* trained with the ImageNet dataset [38]. Notably, all models trained with *UNI* outperformed models trained with *ImageNet* features (Supplementary Figure 10). This highlights the importance of foundation models trained for a specific domain.

Discussion

HER2-positive BC is characterized by higher levels of epidermal growth factor receptor 2 (HER2), where anti-HER2 targeted therapies have shown to be effective. More recently, the approval of targeted therapies in HER2-low tumors has expanded the efficacy of BC treatments to benefit countless patients. Herein, we explore the ability to detect HER2 status directly from H &E images using deep learning. We built six models aiming to explore the ability of classifiers to distinguish between HER2-negative, HER2-low, and HER2-high classes from distinct datasets of IHC and ISH-based assays for HER2 status assessment. We show that models trained to distinguish HER2-negative from HER2-low had lower performance compared to those that only classified between HER2-negative and HER2-high or HER2-low and HER2-high. Thus, we argue that the known challenges [7]

Table 3 Median and standard deviation of the AUROC of the internal test

	NOBG	TUMO
M1	0.88 +0.02	0.90 +0.01
M2	0.87 +0.02	0.85 +0.02
M3	0.91 +0.02	0.87 +0.02
M4	0.85 +0.01	0.85 +0.01
M5	0.69 +0.04	0.72 +0.02
M6	0.78 +0.01	0.78 +0.02

Median and standard deviation of the AUROC of the internal test set on the 10-fold cross-validation with different tile filtering strategies. The best performance for each model is highlighted in bold

Table 4 Median and standard deviation of AUROC performance

	ADMIL	CLAM	TransMIL
M1	0.87 +0.02	0.90 +0.01	0.88 +0.03
M2	0.85 +0.02	0.85 +0.02	0.85 +0.02
M3	0.84 +0.02	0.87 +0.02	0.88 +0.04
M4	N/A	0.85 +0.01	N/A
M5	0.73 +0.05	0.72 +0.02	0.71 +0.02
M6	N/A	0.78 +0.02	N/A

Median and standard deviation of AUROC performance on the internal test set for ADMIL, CLAM, and TransMIL. ADMIL and TransMIL can only work on binary models. The best performance for each model is highlighted in bold

in distinguishing HER2 score of 0 from 1+ may affect the quality of data needed for training the algorithms. Improvements in the pathologic workflow, including stringent quality controls [30, 39] and novel quantitative testing that allows measuring a range of antibody concentrations may help to maximize the sensitivity of HER2 expression.

On the external test set, we observed an overall drop in performance compared to the validation set, which is expected, considering that the TCGA cohort is a multicentric heterogeneous dataset. This effect can also be explained as a result of overfitting. Notably, the M2 model performed significantly better on the external test set compared to the other models, suggesting that the distinction between HER2-neg and HER2-high is better defined, and the classification accuracy might more easily be translated to new cohorts.

When selecting which slide tiles to utilize, many studies eliminate all background tiles and make predictions on what is left [16, 19–21]. We tested whether using solely tumor tiles was sufficient for accurate predictions, and our findings were encouraging. Importantly, while it is anticipated that the tumor environment would hold the essential information for prediction, it remains unclear if surrounding tissues also carry significant information.

In the clinical setting, IHC and ISH assays remain the gold standard, however, it is important to note that a wide range of potential candidates for anti-HER2 treatment incurs screening costs and is hindered in the routine due to various reasons, such as the constrained accessibility of molecular assays, turnaround time, tissue requirements, and health plan coverage. Given that the majority of pathology slides are already routinely stained with H & E, the use of these models enables a fast and cost-effective testing in routine daily practice. For example, in cases when IHC result is 2+ “equivocal” and the HER2 status needs to be tested with additional ISH assay, the model M1 can be used to discriminate among HER2-low or HER2-high. Additionally, the M2 model might be useful for pre-screening patients aiming to detect HER2-neg vs HER2-high before the IHC assay. Finally, the M6 model could be used to discriminate HER2-neg, HER2-low or HER2-high without the need for additional tests.

While past studies have examined the ability of machine learning-based models to predict HER2 status as either positive or negative, our work attempts to expand upon the classification of these subtypes by including HER2-low as a new category. Our results are particularly relevant for two main reasons. This is the first study to detect the HER2-low biomarker from histological images considering a retrospective series of 804 patients from two distinct datasets (ACCC and HEROHE), and it elucidates the limitations of detecting HER2-low solely

from IHC results [30]. Our study reveals the local cellular patterns learned for the morphological phenotypes related to the HER2-low and other subgroups, highlighting which regions of an image are important to the final prediction.

Finally, our findings further reinforce the importance of revisiting the guidelines for HER2 assessment while also supporting the inclusion of more than one pathologist in the HER2 assessment alongside the use of agreement metrics and, following the best practices for utilizing controls [7, 30, 39]. Although this practice would be ideal, many laboratories are not able to make it a reality due to the availability of pathologists and the increasing workload of these professionals.

In addition, by using H & E slides we are also investigating the potential discovery of new H & E tissue patterns that pathologists could learn from and adopt in their evaluation. Therefore, AI-based approaches, such as the one proposed in the present work, have become increasingly in demand and promise to support decision-making in the evaluation of biomarkers.

Although our results are promising, our model may fail to generalize due to overfitting, differences in imaging systems and staining protocols. Our findings may also be hampered by subjective results from IHC-based assays [30].

Conclusion

In our retrospective study using multicentric cohort, we evaluated HER2 subgroups status using the start-of-art of deep learning models in histological images. We show that a weakly-supervised deep learning model can predict HER2 subtypes directly from H & E slides. However, we found that the performance of a model is dependent on the reliability of ground truth, which is derived from the experts interpretation of molecular assays. Future works should focus on curating datasets with high inter-expert agreement in order to mitigate subjective variability of the HER2 assessment.

Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s13058-024-01863-0>.

Supplementary information

Acknowledgements

The authors thank the pathology facility of the A.C. Camargo Cancer Center for whole slide imaging scanning service and guidance.

Author Contributions

RV designed and implemented the AI-based algorithms. LM performed the explainability analysis and generated the figures of the article. AD contributed with the data acquisition and preparation. CT and AB contributed with pathology expertise and histologic interpretations. RR, EDN and DMC contributed to

data analysis and interpretation. JM provided clinical expertise. ITS conceptualized and supervised the study. ITS, RV and LM wrote the final version of the manuscript, on which all the authors provided comments. All authors agreed to the submission of this paper. The work reported in the paper has been performed by the authors, unless specified in the text.

Funding

This work was funded by FAPESP under grant #22/12991-0.

Data Availability

The TCGA and HERO data used in this study are deposited in a public repository described in the Methods. We generated a dataset containing over 2 million tiles of tumor only regions, and it is publicly available at <https://doi.org/10.5281/zenodo.8383580>, under the Creative Commons Attribution 4.0 International license. Should there be a need for further details to reassess the data presented in this manuscript, such information can be obtained by reaching out to the primary contact upon request.

Code Availability

All code used in training the models is available at <https://github.com/tojallab/wsi-mil>, under the GPLv3 license.

Declarations

Conflict of interest

All authors declare that they have no conflict of interest.

Ethics approval and consent to participate

This study received ethical approval (#3028/20) from the ethics committee of the Fundação Antônio Prudente.

Received: 7 March 2024 Accepted: 21 June 2024

Published online: 19 August 2024

References

- Afifi AM, et al. Causes of death after breast cancer diagnosis: a US population-based analysis. *Cancer*. 2020;126:1559–67.
- van Seijen M, et al. Ductal carcinoma in situ: to treat or not to treat, that is the question. *Br J Cancer*. 2019;121:285–92.
- Rakha EA, et al. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *J Clin Oncol*. 2008;26:3153–8.
- Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N Engl J Med*. 2009;360:790–800.
- Parker JS, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27:1160–7.
- Modi S, et al. Trastuzumab Deruxtecan in previously treated HER2-low advanced breast cancer. *N Engl J Med*. 2022;387:9–20.
- 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*. 2020;54:34–44.
- Liwski CR, Castonguay MC, Barnes PJ, Rayson D, Bethune GC. HER2 testing in metastatic breast cancer - is reflex ISH testing necessary on HER2 IHC-equivocal (2+) cases? *Ann Diagn Pathol*. 2022;59: 151953.
- Cui M, Zhang DY. Artificial intelligence and computational pathology. *Lab Invest*. 2021;101:412–22.
- Coudray N, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med*. 2018;24:1559–67.
- Kather JN, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med*. 2019;25:1054–6.
- Valieris R, et al. Deep learning predicts underlying features on pathology images with therapeutic relevance for breast and gastric cancer. *Cancers (Basel)* 2020;12.
- Echle A, et al. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer*. 2021;124:686–96.
- Shamai G, et al. Artificial intelligence algorithms to assess hormonal status from tissue microarrays in patients with breast cancer. *JAMA Netw Open*. 2019;2: e197700.
- Kaiser T, et al. HER2 challenge contest: a detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology*. 2018;72:227–38.
- Gamble P, et al. Determining breast cancer biomarker status and associated morphological features using deep learning. *Commun Med (Lond)*. 2021;1:14.
- Rawat RR, et al. E images. *Sci Rep*. 2020;10:7275.
- Xue T, et al. Deep learning to automatically evaluate HER2 gene amplification status from fluorescence in situ hybridization images. *Sci Rep*. 2023;13:9746.
- Lazard T, et al. Deep learning identifies morphological patterns of homologous recombination deficiency in luminal breast cancers from whole slide images. *Cell Rep Med*. 2022;3: 100872.
- Jaber MI, et al. A deep learning image-based intrinsic molecular subtype classifier of breast tumors reveals tumor heterogeneity that may affect survival. *Breast Cancer Res*. 2020;22:12.
- Couture HD, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ Breast Cancer*. 2018;4:30.
- La Barbera D, Polonia A, Roitero K, Conde-Sousa E, Della Mea V. Detection of HER2 from haematoxylin-eosin slides through a cascade of deep learning classifiers via multi-instance learning. *J Imaging* 2020;6.
- Conde-Sousa E, et al. HEROHE challenge: predicting HER2 status in breast cancer from hematoxylin-eosin whole-slide imaging. *J Imaging* 2022;8.
- Liu Z, et al. A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) 2022.
- Kather JN, Halama N, Marx A. 100,000 histological images of human colorectal cancer and healthy tissue 2018.
- Loeffler C, Kather JN. Manual tumor annotations in tcga 2021.
- Macenko M, et al. A method for normalizing histology slides for quantitative analysis 2009.
- Chen RJ, et al. Towards a general-purpose foundation model for computational pathology. *Nat Med*. 2024;30:850–62.
- Lu MY, et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng*. 2021;5:555–70.
- Robbins CJ, et al. Multi-institutional assessment of pathologist scoring HER2 immunohistochemistry. *Mod Pathol*. 2023;36: 100032.
- Wu S, et al. The role of artificial intelligence in accurate interpretation of HER2 immunohistochemical scores 0 and 1+ in breast cancer. *Mod Pathol*. 2023;36: 100054.
- Wolff AC, 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*. 2018;36:2105–22.
- Ilse M, Tomczak JM, Welling M. Attention-based deep multiple instance learning. *arXiv preprint arXiv:1802.04712* 2018.
- D'Esposito V, et al. Adipose microenvironment promotes triple negative breast cancer cell invasiveness and dissemination by producing CCL5. *Oncotarget*. 2016;7:24495–509.
- Babenko B. Multiple instance learning: algorithms and applications. Tech. Rep., Dept. of Computer Science and Engineering University of California, San Diego (2008).
- Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*. 1979;9:62–6.
- Shao Z, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Adv Neural Inf Process Syst*. 2021;34:2136–47.
- Deng J, et al. Imagenet: a large-scale hierarchical image database (2009).
- 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*. 2022;41:1361–75.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Apêndice E – Artigo HRD – Valieris 2025

Segue abaixo o artigo enviado para publicação na revista *Scientific Reports* em fevereiro de 2025, contendo as análises dos modelos preditores de escore de **HRD** com valores contínuos, descrito na Seção 4.3.

Integrating homologous recombination deficiency signatures enhances deep learning prediction from whole-slide images

Renan Valieris¹, Luan Martins¹, Alexandre Defelicibus¹, Lucas Rosa¹, Iara Santana², Mariangela Brunaldi³, Gabriel Oliveira Santos⁴, Louise de Brot⁴, Diana Noronha Nunes^{5,6}, Dirce Maria Carraro^{7,9}, Emmanuel Dias-Neto^{5,6,9}, Rafael A. Rosales^{8,+}, and Israel Tojal da Silva^{1,9,+,*}

¹Laboratory of Computational Biology and Bioinformatics, A.C.Camargo Cancer Center, São Paulo, Brazil.

²Departamento de Patologia, Barretos Cancer Hospital, Barretos, Brazil.

³Departamento de Patologia e Medicina Legal, Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, Brazil.

⁴Departamento de Anatomia Patológica, A.C.Camargo Cancer Center, São Paulo, Brazil.

⁵Laboratory Medical Genomics, A.C.Camargo Cancer Center, São Paulo, Brazil.

⁶Division of Cancer Biology, Department of Radiation Oncology, Rutgers New Jersey Medical School, Newark, United States.

⁷Clinical and Functional Genomics Group, A.C.Camargo Cancer Center, São Paulo, Brazil.

⁸Departamento de Computação e Matemática, Universidade de São Paulo, São Paulo, Brazil.

⁹National Institute of Science and Technology in Oncogenomics and Therapeutic Innovation (INCITO), São Paulo, Brazil.

*itojal@accamargo.org.br

+these authors contributed equally to this work

ABSTRACT

Homologous recombination deficiency (HRD) has emerged as a central biomarker to guide targeted therapies for various types of solid tumors. While deep-learning approaches for HRD prediction from histopathological images show great promise as alternatives to conventional HRD testing methods, establishing a reliable and consistent ground truth remains a significant challenge. To address this challenge, we examined 3,060 whole-slide images from 2,963 patients in three cohorts across seven tumor types. We demonstrate that genome-wide signatures associated with HRD vary across and within distinct tumor types. These signatures are leveraged to define an aggregated score, which allows the construction of a classifier to identify patients with the HRD phenotype. The score encompasses the molecular diversity of a host of processes related to HRD. We then used this aggregated score as the ground truth for a deep-learning model, built to predict HRD directly from histology images. We show that both the classifier and the derived deep-learning analysis are able to predict HRD and are generalizable across diverse datasets. The analysis of histology images proposed here reduces the time required to determine HRD status, decreases the need for expensive mutational assays, and enables HRD assessment of a broader range of patients who may benefit from PARP inhibitors.

Introduction

Homologous recombination (HR) is a fundamental cellular process that is central for maintaining genomic stability while generating genetic diversity and repairing DNA damage. In this key evolutionary process, new combinations of DNA sequences are generated through the shuffling of DNA segments, thereby contributing to genetic variation in the offspring. The HR process also occurs in somatic cells¹ where it repairs double-stranded DNA breaks (DSBs)², and interstrand crosslinks and DNA gaps are caused by, for example, radiation and chemical exposures³. While the inability of a cell to efficiently mend DSBs contributes to tumorigenesis, the homologous recombination deficiency (HRD) phenotype reveals specific vulnerabilities that can be exploited as a valuable diagnostic tool, with the potential to improve patient selection for Poly (ADP-ribose) polymerase (PARP) inhibitor therapies⁴⁻⁷.

The assessment of HRD consists of scrutinizing DNA lesions for both germline and somatic mutations, along with epigenetic changes in the genes *ATM*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, and *RAD51*. These genes are consistently involved in the HRD phenotype in several cancers such as breast, ovarian, pancreatic, and prostate cancers^{8,9}. In addition, using SNP array-based

signatures, three distinct types of genomic scars were proposed as biomarkers of HRD. Briefly, these include large-scale state transitions (LSTs), which are defined as chromosomal breaks between adjacent regions of at least 10 Mb¹⁰, telomeric allelic imbalance (TAI) score, which are measured from an unequal distribution of subchromosomal regions between the two alleles of a chromosome specifically at telomeric regions¹¹ and loss of heterozygosity (LOH), where the HRD score is defined as the distribution of lengths of LOH regions throughout the genome in a tumor sample¹². Taken together, these three DNA-based scores measure similar genomic features. Other efforts to determine HRD status extend beyond genomic scars, *BRCA1/2*, and known HR-related genes. For example, recent studies have shown a distinctive pattern of mutations known as mutational signatures that are associated with deficiencies in DNA repair mechanisms within a genome^{13–15}, including HRD. In particular, the mutational signature SBS3 serves as a robust HR-status indicator, enabling the classification of genetic variants whose functional implications have not yet been fully characterized¹³.

Although many oncology biomarkers rely on bench experimental molecular biology analysis, recent advances in artificial intelligence (AI)-based technologies have increased the ability to uncover previously inaccessible information directly from imaging data. For example, studies in the recent have sought to predict biomarkers in solid tumors directly from hematoxylin and eosin (H&E)-stained histopathological slides¹⁶, including biomarkers to detect DNA repair deficiency. However, the ground truth of studies for HRD assessments via AI models is usually defined according to either genomic scars, mutational signatures, or mutations in HR-related genes alone^{17–21}. Importantly, HR is regulated by numerous signaling pathways, additional key regulators beyond *BRCA1/2* and other HR-related genes²². These variations can manifest differently across various tumor types and also vary among individual patients with the same tumors. Additionally, beyond cost and accessibility, individual HRD assays may not fully capture the complexity or the extent of HRD. Hence, there is a significant clinical demand for affordable, rapid, and standardized screening HRD tests that encompass a broad range of alterations associated with HRD phenotypes, rather than focusing solely on individual molecular measures.

In this study, we evaluated whether the integration of genome-wide mutational footprints that harbor various HRD-related genetic determinants, or a mixture thereof would reveal significant underlying features on histological whole slide images, capable of enhancing HRD detection using a deep learning (DL) model. To achieve this goal, we developed a new score, called the panHRD, by integrating genome-wide mutational signatures of biomarkers associated with HRD. This score served as a foundation for constructing a classifier, which we first used to identify patients with known HRD-related genetic lesions. We subsequently employed the panHRD score as the ground truth for training a regression based DL model to detect HRD directly from histological images. The validation of both the classifier and the derived DL model was conducted by using breast cancer samples from two independent cohorts, which revealed that our method is robust and generalizable across diverse datasets.

Results

Aggregated signature score as a reliable proxy of HRD status

As HRD manifests through a complex interplay of genomic alterations, with minor and major consequences for proteins encoded by different genes, relying on a single biomarker to detect this deficiency is misleading²³. In this study, we devised a method to aggregate HRD scores from different sources by considering broader mechanisms encompassing specific patterns such as single nucleotide changes and larger-scale genomic alterations. Specifically, we used 5 known HRD-related biomarkers: SBS3, ID6, CN17, CN1, and the scarHRD. SBS3 is part of the single-base substitution (SBS) signature set, which characterizes somatic mutations by grouping them by nucleotide alteration and context. ID6 is part of the small insertion and deletion (ID) signature set, which characterizes somatic indels by the number of nucleotides inserted or deleted and the length of the event. Both signatures SBS3 and ID6 have been previously associated with HRD²⁴. The signature CN17 is part of the Copy Number variations (CN) signature set, which categorizes copy number events by loss-of-heterozygosity status, number of copies, and segment length, and it is also associated with HRD²⁵. In contrast, signature CN1, another member of the CN signature set, is associated with samples with little or no copy number alteration but is significantly negatively correlated with CN17 and other HRD scores. Finally, genomic scars (scarHRD) encompass measures of loss of heterozygosity (LOH), large-scale state transitions (LSTs), and telomeric allelic imbalance (TAI), all of which reflect accumulated DNA damage over time²⁶.

We quantified the activities of the five HRD-related biomarkers described previously in individual samples of seven tumor types (Figure 1 [A-E]) by using their mutational signatures as described in the COSMIC database. Figure 1 shows that HRD-related signatures vary considerably across tumor types, indicating the relevance of further aspects related to a tumor-specific context²⁷. An integration of HRD scores should therefore account for this inter-tumor variability. This motivated us to develop an aggregation method that estimates weights for each tumor cohort and each signature component (Supplementary Figure 1), leading to the panHRD score. Integrating these diverse genomic features into a combined score allows the capture of a more comprehensive mutational spectrum of HRD (Figure 1 F). Importantly, similar to our findings on HRD, the aggregated score we developed is highly versatile and can, in principle, be applied to other genetic signatures.

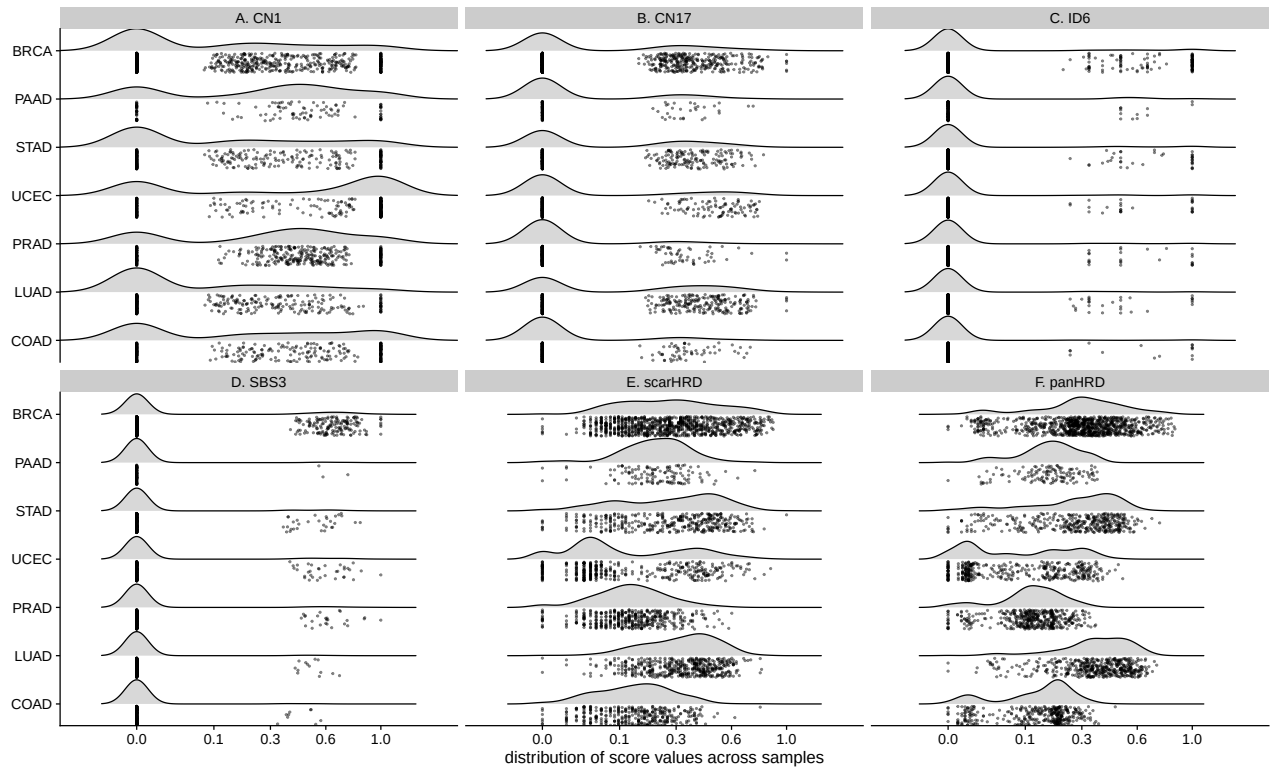


Figure 1. Exposure landscape of several HRD-biomarkers across tumor types. The x -axis in the panels ranging from A to E corresponds to the normalized exposure to each biomarker. From A to D normalization was made by using equation (2) and for E by equation (3). The x -axis for panel F represents the normalized exposure defined by the panHRD score defined in (4).

panHRD is associated with somatic, germline, and epigenetic alterations in HR components

We examined whether our panHRD score is associated with known genomic lesions from HRD (Figure 2). For this purpose, we used the TCGA breast cancer cohort, the largest available cohort with genomic alterations annotated in our dataset. In agreement with published results²⁸, we observed the prevalence of HRD in samples with basal and triple-negative (TNBC) phenotypes (Figure 2B). Together with known genetic lesions in key HR genes (Figure 2C), these findings recapitulate HRD tumor phenotypes in intrinsic subtypes of breast cancer^{13,28}. In addition, the analysis displayed in Figure 2B shows that somatic mutations in the *TP53* gene are more frequent across samples with higher panHRD as expected. Indeed, when *TP53* is mutated, its ability to regulate DNA repair processes is compromised, leading to increased genomic instability and defects in HR^{29,30}. In contrast, we observed a trend toward *PIK3CA* mutations in HR-proficient samples, which, in turn, are associated with the luminal A (LumA) subtype, supporting findings of previous studies that reported a similar relationship³¹.

The predictive capability of the panHRD score can be evaluated as a classifier score allows the identification of HRD-positive samples, using the set of at least one variant known to impair homologous recombination (HR) (Figure 2C) as the ground truth positive set. The panHRD score achieved an area under the ROC curve (AUC) of 0.89 (Figure 2E). We next sought to assess whether the panHRD is predictive of the detection of subsets of HRD-related alterations (Figure 2D,E and Supplementary Table 1). These results indicate that the panHRD score retains its utility even when focusing on a reduced subset of alterations.

In summary, the panHRD score serves as a predictor of somatic, germline, and epigenetic alterations affecting homologous recombination (HR) components. By capturing a broader spectrum of genomic instability beyond individual gene mutations, it provides a framework for identifying HR-deficient tumors.

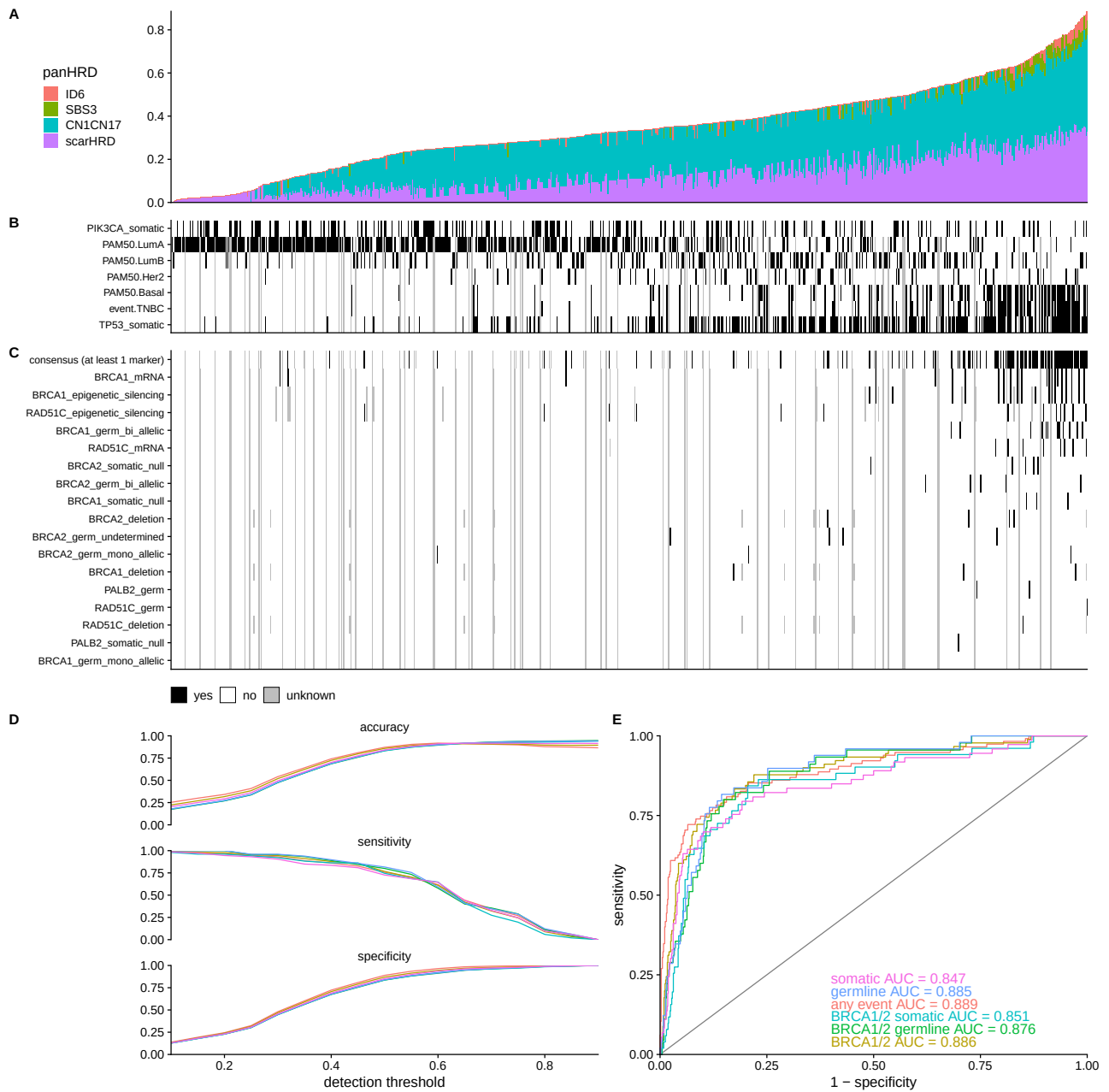


Figure 2. Overview of the mutational landscape, PAM50 intrinsic subtype data, and panHRD prediction based on genomic data of breast cancer samples. (A) Tumor samples from the TCGA-BRCA cohort were sorted by ascending panHRD score, with the individual components of the aggregated score annotated. (B) Sample-level annotations of the PAM50 subtype, triple-negative status, *TP53* and *PIK3CA* somatic status. (C) Sample-level annotations of lesions in known homologous recombination (HR) genes. (D) Performance metrics evaluated at varying threshold levels. (E) ROC curves of each variant subset and their AUC.

Predictability of panHRD from histology

Having demonstrated that our panHRD score effectively predicts genetic lesions in HR-related genes using genomic data, we next investigated whether our regression model could infer the genomic panHRD score directly from histology. To this end, we used the panHRD score as a target to train a multiple instance learning (MIL) regression model from a large cohort of 3,060 H&E whole-slide histological images, encompassing 2,819 subjects and seven tumor types. MIL models are capable of learning

from the data which portions of the image contain the information relevant to the prediction. Furthermore, the regression approach allows for more nuanced predictions of HRD, avoiding the need to predefine binary negative/positive groups.

First, we assessed how specific tumor characteristics affect score prediction by building a tumor-agnostic (*TA*) model using data from seven tumor types simultaneously. Additionally, we trained seven tumor-specific models (*TS*), each one using data from a specific tumor type. Then, we calculated the correlation coefficient for each model using a held-out subset of the original data (Figure 3E). We found that the *TA* model matched or outperformed the *TS* models in most tumor types, demonstrating that HRD prediction was not restricted to a single tumor type. Training on multiple cohorts provides regularization, reducing over-fitting to specific cohort characteristics and encouraging the learning of more generalized and robust features.

Next, we trained models to predict the individual components of the panHRD score to evaluate how well these scores can be predicted using histopathology data. Given that histology reflects the cumulative effects of molecular alterations, we investigated whether the specific lesions used as ground truth influence the predictive performance of the models (Figure 3 [A-D]). The overall performance of the aggregated cohort reveals that SBS3 has lower performance than the other signatures do, whereas CNICN17 and scarHRD perform significantly better, and the aggregation of signatures (panHRD) further increases the correlation measured. We observed an improvement in the prediction correlation when tumor-specific weights were used for the panHRD calculation (Figure 3A). We conclude that the combined measure provides a more comprehensive representation of HRD in histopathology data, leading to the enhanced correlation observed across tumors. These results suggest that the panHRD may serve as a novel readout of HRD, specifically in direct histology.

We compared our model with a previous regression-based approach²⁰, referred to as CAMIL, which was employed for the prediction of HRD across multiple cancer types. Importantly, the authors highlighted the relevance of CAMIL for biomarker prediction considering HRD as a continuous numerical value. Indeed, as demonstrated in their benchmark, CAMIL enhances the accuracy of biomarker prediction²⁰. Herein, we used the correlation as the evaluation metric to compare our approach with the CAMIL model as reported in Supplementary Table 4 from Nahhas et al.²⁰. The panHRD was used as the ground truth for the HRD score in our model, while CAMIL considers HRD by adding up the LOH, TAI, and LST scores. Our evaluation highlights the strengths of panHRD in predicting HRD when compared to CAMIL. The results for the aggregated dataset with seven cancer types are shown in Figure 3D, and those for the tumor-specific models are shown in Figure 3E. The CAMIL approach in these figures is referred to as scarHRD.

Finally, unlike other approaches¹⁷⁻¹⁹, we built a model considering an aggregated dataset, which was able to predict HRD across seven tumor types. Accordingly, the tumor-agnostic model is able to detect underlying HRD features (Figure 3). Beyond the optimized changes in the regression-based model, we believe that our performance benefits arise from the use of UNI as our feature extractor, which is a general-purpose self-supervised model for pathology³², the inclusion of images from multiple tumor types, which benefits the generalization of the model, and the aggregated panHRD score, which better represents the HR deficiency across tumor types.

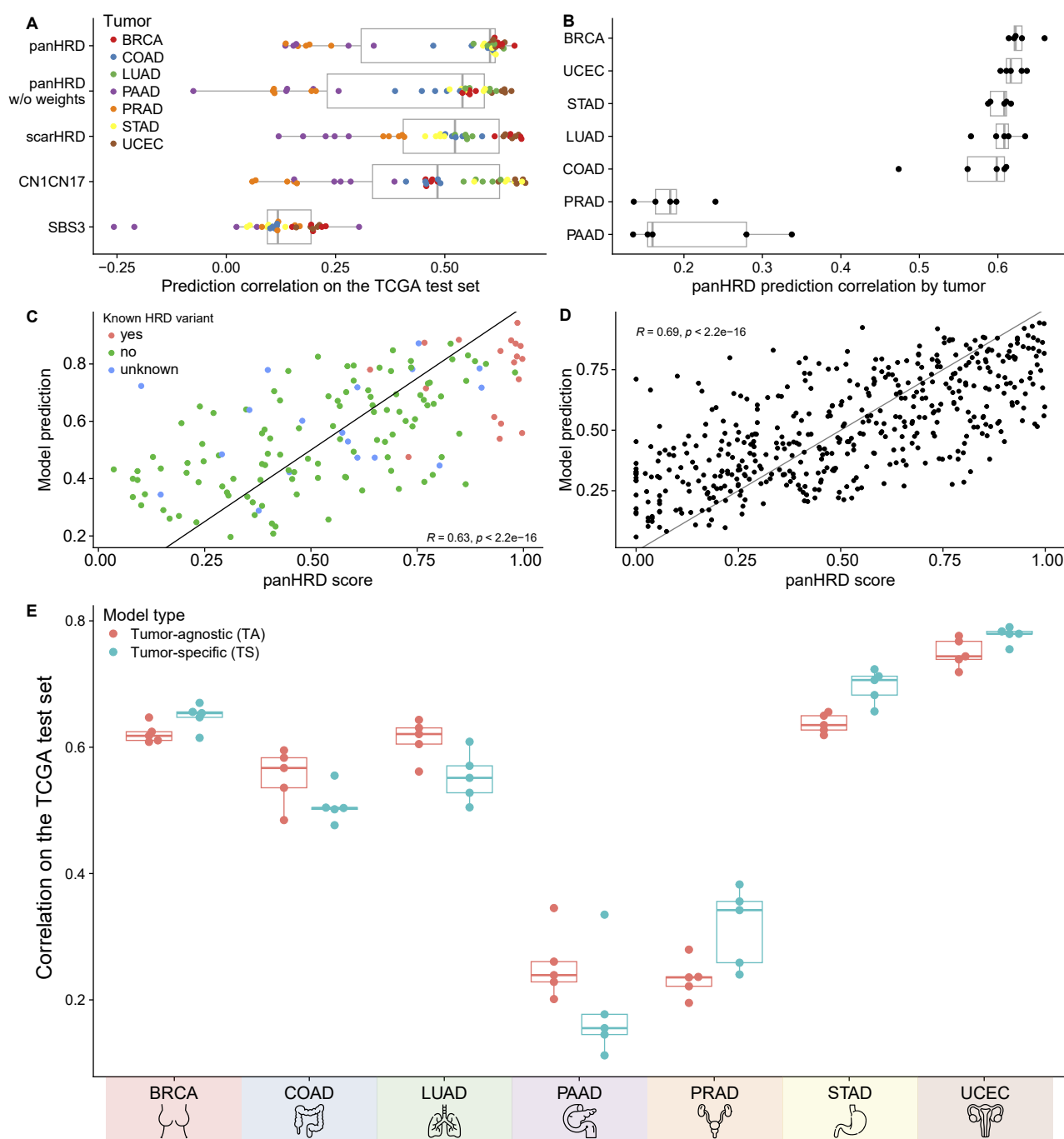


Figure 3. Overview of HRD prediction performance on H&E slides using panHRD-based deep learning models. (A) Correlation of the model predictions with the targets for each HRD score; (B) panHRD prediction correlation stratified by tumor type; (C) Scatter plot of the model predictions against the genomic score in the TCGA-BRCA test subset, the black line indicates the identity $y = x$; (D) Scatter plot of the model prediction against genomic score for the TCGA test subset, the black line indicates the identity $y = x$ (E) Correlation coefficients of the Tumor-specific (TS) and Tumor-agnostic (TA) models for each of the tumors.

Optimal panHRD thresholds for HRD detection on histological images

We defined the panHRD as a continuous score that reflects the aggregated effect of deficiencies in HR repair. We then employed a regression-based modeling approach to investigate the relationship between the panHRD and morphological features from a

Event type	Metric	Image-predicted panHRD thresholds for HRD detection on TCGA-BRCA																
		0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9
BRCA1/2	accuracy	0.12	0.12	0.13	0.16	0.20	0.26	0.35	0.45	0.54	0.58	0.63	0.68	0.78	0.82	0.87	0.89	0.87
	sensitivity	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93	0.93	0.80	0.73	0.73	0.60	0.53	0.27	0.00
	specificity	0.00	0.00	0.01	0.05	0.09	0.16	0.27	0.38	0.48	0.54	0.61	0.67	0.79	0.85	0.91	0.97	0.99
BRCA1/2 germline	accuracy	0.05	0.05	0.06	0.09	0.13	0.19	0.28	0.38	0.47	0.51	0.58	0.62	0.72	0.80	0.84	0.91	0.95
	sensitivity	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.83	0.83	0.67	0.50	0.50	0.50	0.33	0.17	0.00
	specificity	0.00	0.00	0.01	0.04	0.08	0.15	0.25	0.35	0.45	0.50	0.57	0.63	0.74	0.81	0.87	0.95	0.99
BRCA1/2 somatic	accuracy	0.07	0.07	0.08	0.11	0.15	0.21	0.31	0.40	0.50	0.55	0.61	0.68	0.78	0.82	0.88	0.92	0.92
	sensitivity	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89	0.89	0.89	0.67	0.67	0.33	0.00
	specificity	0.00	0.00	0.01	0.04	0.09	0.15	0.25	0.36	0.47	0.52	0.59	0.66	0.77	0.83	0.90	0.97	0.99
any event	accuracy	0.14	0.14	0.15	0.18	0.22	0.28	0.38	0.47	0.56	0.59	0.64	0.69	0.79	0.83	0.87	0.90	0.87
	sensitivity	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.89	0.78	0.72	0.72	0.61	0.56	0.33	0.06
	specificity	0.00	0.00	0.01	0.05	0.09	0.17	0.28	0.39	0.50	0.54	0.62	0.68	0.80	0.86	0.93	0.99	1.00

Table 1. Variation in performance metrics for the image-predicted panHRD score across different detection thresholds. The ground-truth HRD alterations were separated into groups and performance was tested on the test subset of the TCGA breast cancer cohort.

histological image. Since panHRD is associated with the HRD burden, we sought to establish a clinically applicable threshold for binary phenotypic classification. Indeed, this task is challenging because of the lack of a standardized method to unequivocally determine the presence of HRD. Thus, to standardize a measure for estimating the predictive power of panHRD in identifying samples with alterations in known HR-related genes, we established a cutoff to distinguish HR-deficient from HR-proficient cancers. To achieve this, we utilized breast cancer samples¹³ with deleterious alterations in known HR genes (Figure 2) as the ground truth positive set. Specifically, given that HRDs induced by BRCA1 and BRCA2 mutations, both germline and somatic, are well-established predictive markers for the response to PARP inhibitors and platinum-based therapies in breast cancer^{33,34}, we initially used BRCA1/2 alterations as the endpoint for determining an optimal cutoff in this context. Additionally, our choice is further supported by results from the OlympiA trial, which highlights the need for an efficient method to screen for BRCA1 and BRCA2 mutations across all breast cancer phenotypes³⁵.

To determine the optimal thresholds for HRD detection, we calculated performance metrics, including accuracy, sensitivity, and specificity across a range of cutoffs (Table 1). The choice of a detection threshold depends on how the results will be interpreted and applied in a clinical context. For example, we considered a scenario in which an image-based HRD prediction could serve as an initial screening tool before performing an expensive gene panel test. In this context, prioritizing high sensitivity is essential. As a result, the optimal cutoff was set at 0.55, yielding 93% sensitivity in detecting both germline and somatic BRCA1/2 alterations in the test set.

Independent validation confirms the predictability of genomic and image-derived panHRD

To further validate the ability of the genomic panHRD to detect samples with HR pathway gene alterations and its association with morphological patterns extracted from histological images, we conducted an independent validation analysis using data from two separate cohorts. First, we used a cohort of 65 TNBC patients, for whom histological images and panel sequencing data of HR-related genes were available. Additionally, a subset of 22 patients had whole-exome sequencing (WES) data, for which we calculated the panHRD score (Figure 4C). As a first step, we evaluated the association between panHRD and HR-related alterations in this subset of 22 patients. The analysis revealed elevated panHRD in the samples with validated genetic alterations in *BRCA1* through germline pathogenic mutations and epigenetic silencing in the tumor, and in *RAD51C* through epigenetic silencing in the tumor (Figure 4A). The *TrueSight Oncology 500 HRD* assay was performed on 7 samples, and the resulting GIS score can be seen in Figure 4B. Interestingly, we observed lesions mutually exclusive in specific HR genes, which may represent alternative mechanisms for achieving genomic instability and the development of disease in this specific cancer subtype. Next, we calculated the area under the ROC curve for the prediction of known biomarkers using the panHRD score and obtained an AUC of 0.86 (Figure 4D). As our second step, we applied our tumor-agnostic model trained through cross-validation on TCGA data to investigate whether the panHRD score derived from morphological features is associated with HR gene alterations detected in these samples. Using a threshold of 0.55, the tumor-agnostic model predicts BRCA1/BRCA2 deficiency with a sensitivity of 88% in a cohort of 65 patients who had a panel of genes tested for HRD status (Supplementary Table 2).

Notably, unlike previous approaches^{18,20} that primarily only measured the capacity of the model to predict the given score or group, our method establishes a detection threshold directly from the mutations in HR pathway genes, ensuring a more biologically relevant and precise stratification of patients.

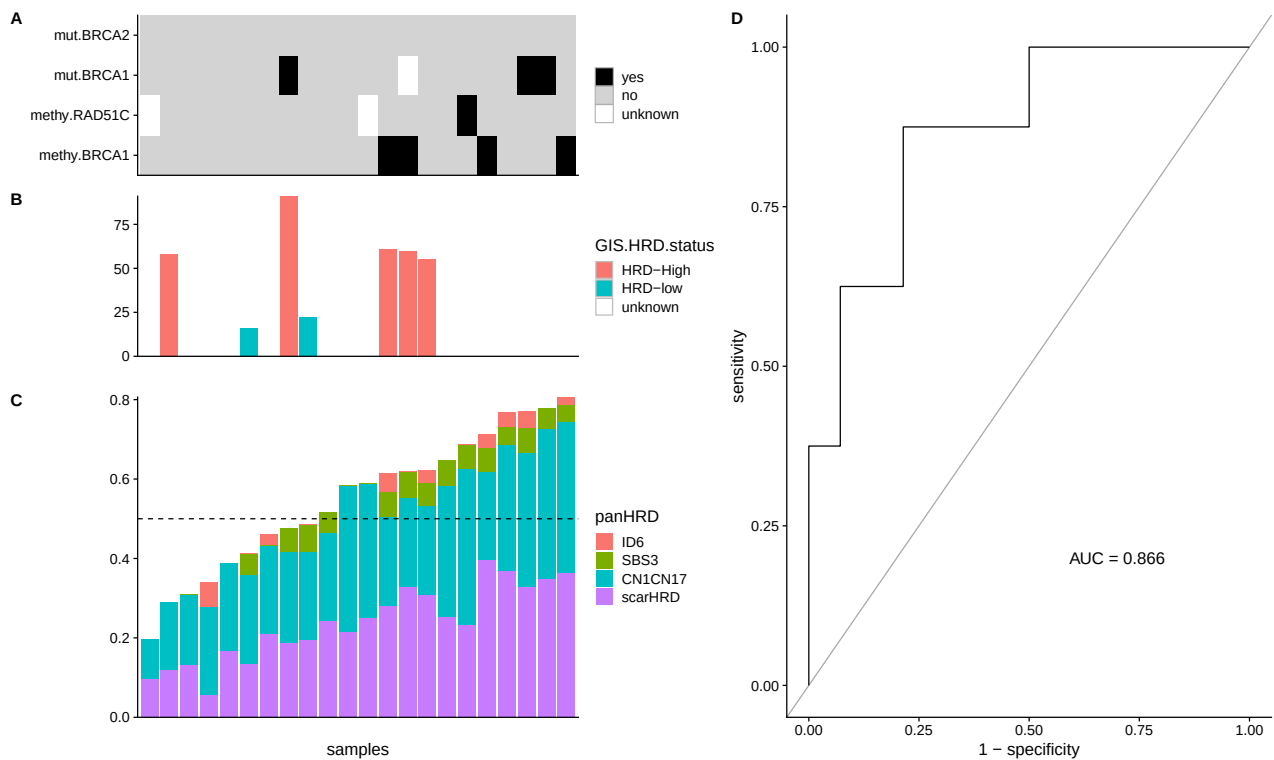


Figure 4. Overview of the external AC-BRCA cohort using trained models for HRD prediction from H&E slides. (A) Germline alterations in known HR genes. (B) HRD status determined by the TSO 500 HRD assay, with untested samples marked as unknown. (C) panHRD score and contributions of mutational signatures in analyzed subjects. (D) ROC curve of the panHRD predictions of known biomarkers from the genomic data. The cutoff value at 0.55 was derived from the TCGA dataset with known HRD biomarkers

Next, the CPTAC breast cancer cohort was utilized for additional external validation. This cohort consists of 402 slides from 122 subjects, with available H&E-stained images, somatic mutations, and gene expression data. We applied our tumor-agnostic model, trained on the TCGA cohort, to the CPTAC slides³⁶ for inference. Owing to the unavailability of copy number information, the ground truth HRD scores could not be calculated. However, we leveraged other HRD-related markers, including the associations between HRD and breast cancer subtypes inferred from the PAM50 classification (Figure 5B), the gene expression levels of homologous recombination repair genes (Figure 5C), *TP53* and *PIK3CA* mutations (Figure 5D), and SBS3 signature activity. The analysis revealed that patients with elevated panHRD scores presented increased expression levels of genes such as *AURKA*, *RAD51*, *RAD54B*, and *XRCC2* (Figure 5C). These genes are relevant components of DNA repair pathways, and their increased expression in panHRD-high tumors suggests a compensatory upregulation of repair mechanisms in response to defective homologous recombination, potentially reflecting an adaptive response to genomic instability³⁷. Conversely, patients with lower panHRD scores presented higher expression of *HELQ*, *TP53BP1*, and *RAD50*. These genes are regulators of the cellular response to DSBs³⁸. This differential gene expression highlights the distinct molecular adaptations that tumors undergo depending on their HRD status. Furthermore, we observed a greater prevalence of somatic *TP53* mutations in samples with higher predicted scores, whereas lower scores were associated with somatic *PIK3CA* mutations which are enriched in Luminal subtypes^{39,40}. These findings are consistent with the patterns observed in the TCGA cohort (Figure 2B).

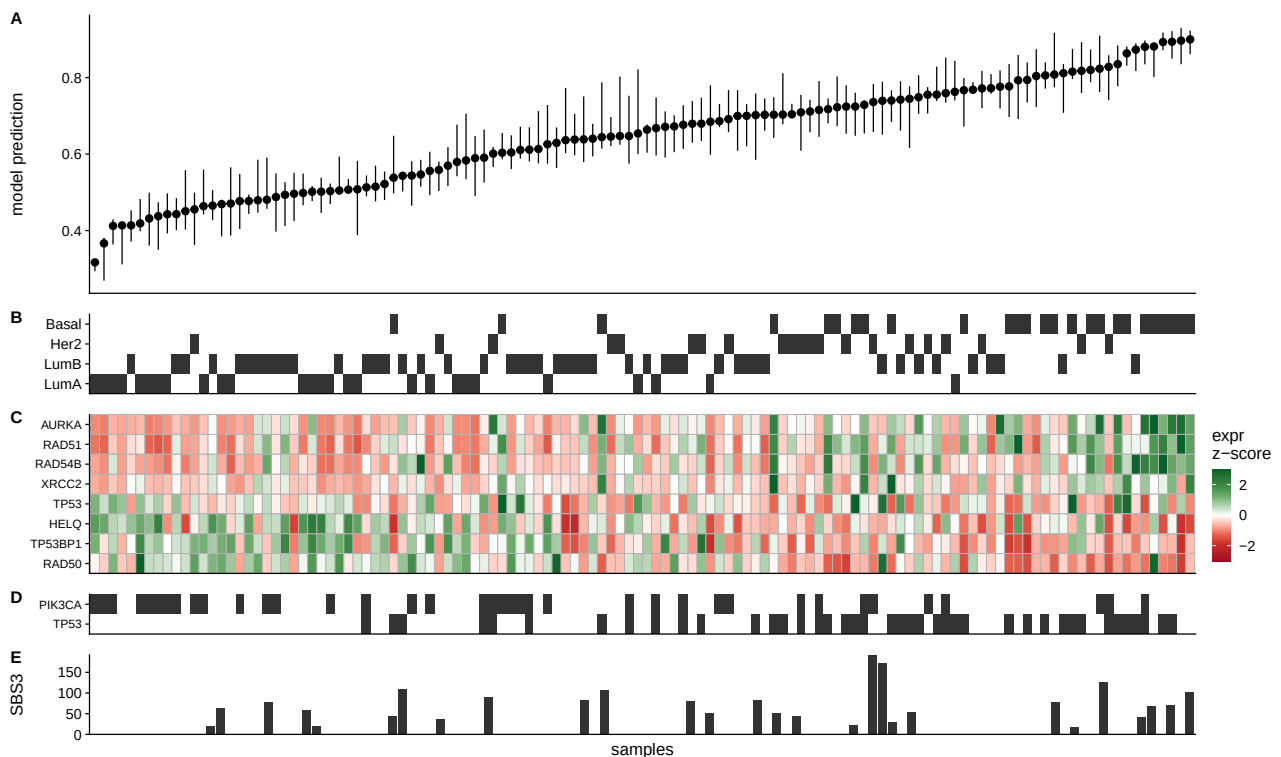


Figure 5. Tumor agnostic model prediction on the H&E slides of the CPTAC breast cancer cohort. (A) The point denotes the median prediction of the five folds, and the whiskers are the minimum and maximum prediction values obtained for a given subject. (B) PAM50 classification of the samples according to their gene expression. (C) Gene expression levels of selected DNA repair genes. (D) Somatic mutations of selected genes. (E) Mutational signature SBS3 exposure.

Collectively, the independent validation across multiple cohorts reinforces the generalizability of the panHRD score as a predictive marker for HRD-associated gene alterations, demonstrating its ability to infer these alterations from both genomic data and histological images, thereby enabling the detection of HRD status through multiple modalities.

Attention maps reveal HRD-related histological patterns

The MIL architecture contains an attention module that permits the model to learn from the data in which portions of the image are relevant for the target prediction. Thus, by analyzing the image regions with high model attention we can learn which morphological patterns are relevant for the prediction of the HRD score.

We selected the top 10 highest attention tiles from each slide of the test set and projected the UNI's 1024-feature vector toward the plane by using the t-SNE visualization tool⁴¹ (Figure 6A). We observed that there were distinct patterns across slides from different tumor types, indicating that the model was able to predict HRD. The patterns learned are consistent across several folds (Supplementary Figure 2), indicating that the model learned to rely on the same morphological patterns when trained with different subsets of the data.

We stratified the tiles by the panHRD score attributed to their respective slide and tumor-type (Figure 6). We observed tumor-specific patterns across the strata (Figure 6A and D). This stratification allows the analyses of tiles associated with these regions to identify potential morphological characteristics associated with HRD (Figure 6C). For example, visual inspection of the slides made by pathologists revealed distinct histopathological patterns associated with varying levels of HRD in patients with breast cancer. Cases with lower HRD exhibited breast carcinomas of low to intermediate histological grade, with mild to moderate nuclear atypia, solid, tubular, or lobular infiltration patterns, moderate desmoplasia, and mild inflammatory infiltration. In contrast, cases with higher HRD were characterized by high-grade carcinomas with pronounced nuclear atypia, increased mitotic activity, pronounced desmoplasia, and marked inflammatory infiltration, suggesting a correlation between higher HRD status and stronger immune response. Moreover, necrosis was observed in colon adenocarcinoma with increased HRD, highlighting the potential link between HRD and tumor cell death mechanisms⁴². These findings highlight the role of the tumor microenvironment in the manifestation of HRD and support the use of histological features in assessing HRD status and guiding treatment strategies, particularly with DNA-damaging therapies such as PARP inhibitors⁴³.

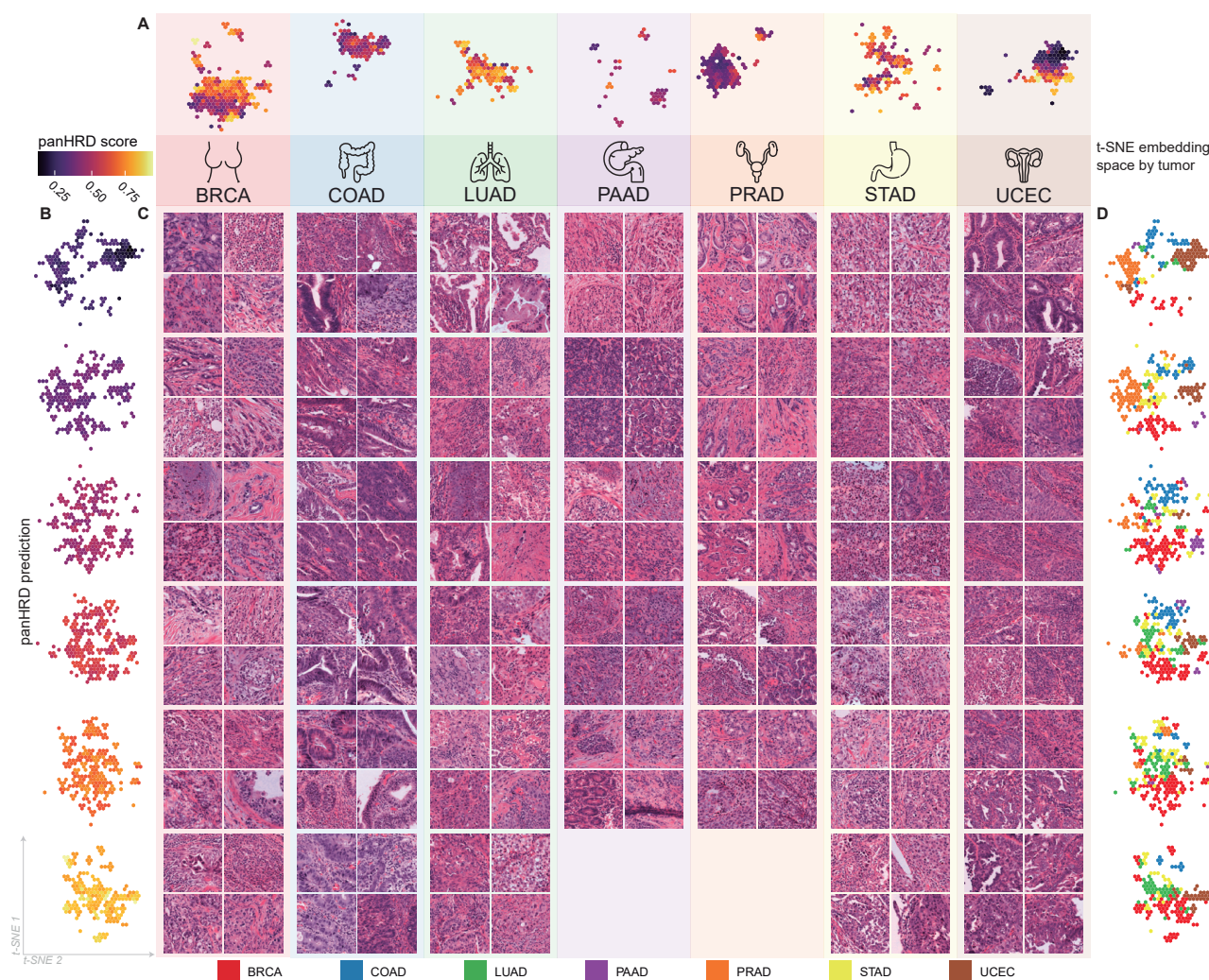


Figure 6. Examples of actionable tiles across tumors and panHRD scores using the t-SNE. (A) Tumor-specific embeddings across panHRD scores; (B) Embeddings stratified by the panHRD score; (C) Examples of actionable tiles across tumor types and panHRD scores; (D) Embeddings colored by tumor type stratified by the panHRD score.

Discussion

Homologous Recombination Deficiency (HRD) is a key predictive biomarker for identifying patients who are likely to benefit from therapies involving PARP inhibitors (PARPi). However, large-scale screening for HRD using genomic assays for *BRCA1/2* mutations and HRD status is not only expensive and time-consuming but also requires high-quality tumor DNA/RNA samples, making widespread implementation difficult.

While deep-learning (DL) methods for predicting HRD directly from tissue slides have emerged as promising alternatives to traditional DNA-based and functional assays, a major challenge in these approaches is establishing a reliable and universally accepted ground truth. Herein, we developed the panHRD, a novel aggregated measure that integrates multiple genome-wide mutational signatures associated with homologous recombination deficiency (HRD). We then employed a regression-based deep learning model to investigate the relationships between panHRD and morphological features extracted from histologic whole-slide images. Our findings demonstrate that the panHRD score serves as a robust proxy for HRD burden, allowing us to define a reliable framework for AI-driven HRD prediction directly from histopathology slides.

We assessed the predictive power of the panHRD score by calculating the Area Under the ROC Curve (AUC), which was 0.89 for detecting deleterious HR gene alterations when TCGA-BRCA patients were used as the ground truth. We then validated the panHRD in independent breast cancer cohorts (AC-BRCA and CPTAC-BRCA), confirming its efficacy in detecting HR gene deficiencies. Notably, a detailed examination of the AC-BRCA cohort revealed that four subjects presented high panHRD

scores despite having no detectable events in the known HR genes or those identified by the TSO 500 HRD assay. These findings suggest that HRD in these subjects may be driven by alterations in genes beyond the well-characterized HR genes. A key challenge we faced was the lack of ground truth HRD scores in the CPTAC breast cancer cohort, which does not provide copy number variation (CNV) data required for calculating HRD scores. To address this, we used alternative markers associated with HRD, such as the gene expression levels of DNA repair pathway genes and mutational data from the *TP53* and *PIK3CA* genes. Our findings demonstrated that somatic *TP53* mutations, which are known to be linked with HRD, were more frequent in samples with higher predicted HRD scores. In addition, *PIK3CA* mutations, typically associated with Luminal breast cancer subtypes, were more common in samples with lower predicted scores, further reinforcing the biological relevance of our model's predictions.

Importantly, the set of alterations used as the ground truth does not encompass all possible alterations that may lead to HRD, and further studies are needed to fully elucidate the complete spectrum of such alterations⁴⁴. Additionally, we must distinguish between mutations that directly cause HRD—by disrupting the function of genes within the HR pathway—and HRD scores, which measure the downstream effects of an impaired HR pathway. In this context, the panHRD score defined here may be able to identify patients with HR deficiencies caused by yet-undiscovered alterations.

Our regression-based deep learning model, trained on data from multiple tumor types, takes into account shared patterns across tumors, enhancing its performance. This approach increased the sample size, reduced overfitting, and allowed for greater generalization. The consideration of continuous HRD scores instead of binary groups avoids the loss of information due to discretization, accommodates the varying degrees of HRD across samples, and addresses the lack of consensus on a threshold for HRD detection.

Although our panHRD is a continuous score, we defined a clinically validated cutoff for breast cancer, derived from HR pathway gene mutations. This threshold ensures a biologically meaningful stratification of patients, enhancing the applicability of histology-based HRD detection. While validated in breast cancer, this approach lays the groundwork for expanding HRD assessment across multiple tumor types, supporting broader clinical adoption and integration into biomarker-driven treatment strategies. Future efforts should focus on integrating comprehensive HR gene annotations in diverse malignancies to further enhance the applicability of our method in a broader oncological context.

Finally, reliance on surrogate markers introduces the possibility of overlooking some of the complex genomic alterations involved in HRD. This limitation highlights the need for future studies to validate our model with more comprehensive datasets that include both histological and full genomic data to ensure more accurate predictions and broader applicability across cancer types. Subsequent studies should focus on refining the model with expanded datasets and further external validation to fully establish its clinical utility.

To conclude, our study demonstrates that a regression-based deep learning model can effectively predict homologous recombination deficiency (HRD) directly from histological images by utilizing a novel panHRD score. By training on combined data from multiple tumor types, we enhanced the model's robustness and generalizability, leveraging shared patterns across different cancers. Our approach of predicting continuous HRD scores, rather than relying on predefined thresholds, addresses the challenges of existing classification methods and offers a more nuanced assessment of HRD. Importantly, our findings highlight that HRD may be driven by genetic alterations beyond the well-characterized and more well-known HR genes, suggesting the potential for broader applicability of the panHRD score in identifying HRD across diverse patient populations. This work paves the way for further research to refine and validate these predictive models, ultimately contributing to more personalized and effective cancer treatments.

Methods

Workflow design for HRD prediction from histologic slides

Our proposed end-to-end pipeline proceeds in several steps illustrated in Figure 7. We start by computing the panHRD score for each slide image i of tumour type t , $s_{t,i}(\epsilon)$, as defined by (4). A second stage of our workflow begins with the preprocessing of whole slide images (WSIs) across various tumor types, as described in Section "Slide preprocessing and feature extraction". Briefly, slides are divided into 256x256 patches, and background regions are systematically removed to focus on the relevant tissue areas. Color normalization is then applied to standardize the appearance of the histology images, mitigating variations due to differences in staining. Features are extracted from these patches using the UNI model³², as described in section "MIL Model Definition and Training". The UNI model is a deep learning architecture designed for effective feature representation. These features are fed to the attention model that reduces the tile-level features to a slide-level vector representation of 256 features. As a result, a vector $X_{t,i,j}$ with $1 \leq j \leq 256$ features, defined for each (t, i) pair, was generated by this procedure. The extracted features are subsequently fed into the regression model $s_{t,i}(\epsilon) = \beta_0 + \sum_{j=1}^{256} \beta_j X_{t,i,j} + e_{t,i}$, specifically designed to predict the panHRD score. This model is adjusted via multiple layers, including an attention module that scores the patches, an aggregation module that computes a representative slide-level vector, and a decision module that generates the final prediction.

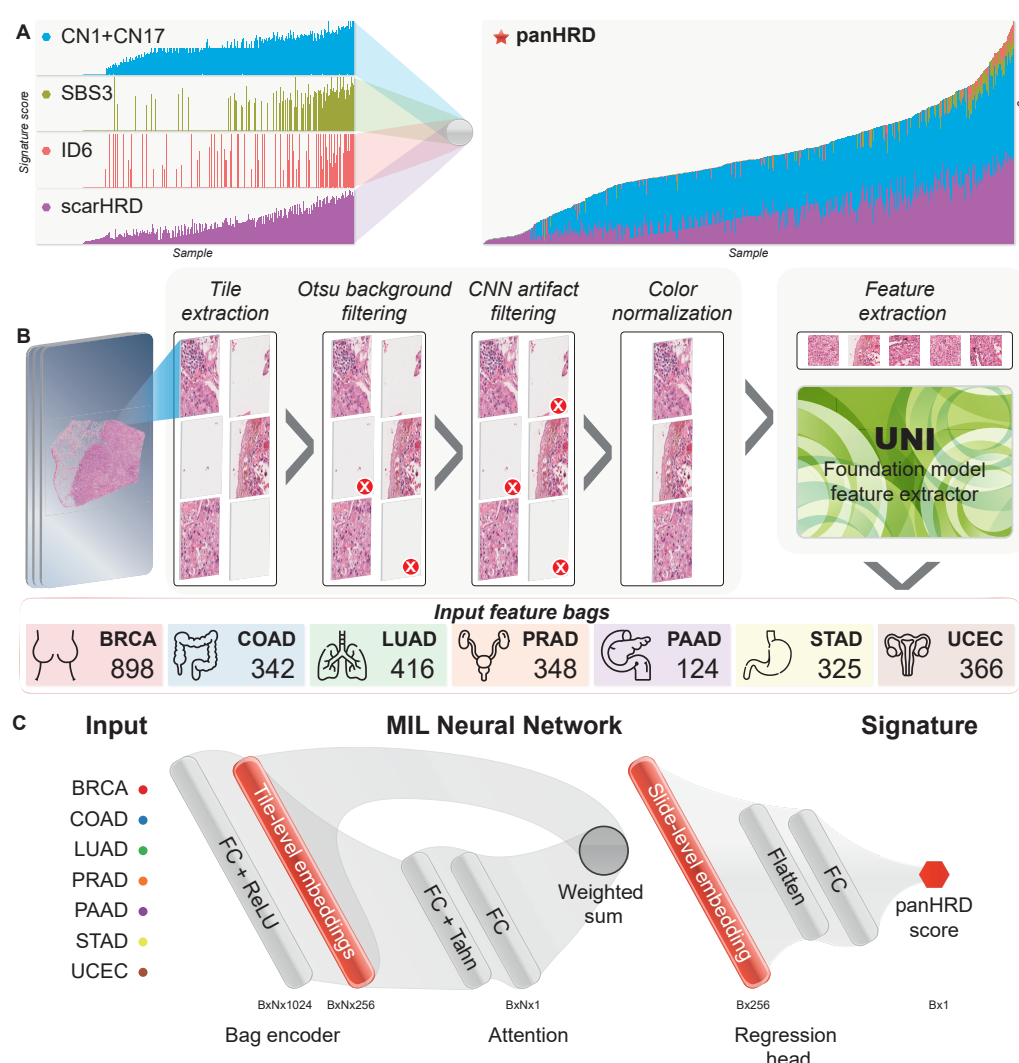


Figure 7. Workflow of the proposed method. (A) Our panHRD signature is generated by combining HRD-relevant signatures. (B) Slides across tumor types are divided into 256x256 patches, the background is removed, color-normalization is applied and features are extracted by the UNI model. (C) Architecture of the model proposed to predict the score.

Molecular data

In addition to the WSI data, molecular data were also obtained for the respective samples to be used in subsequent analyses, including the determination of HRD status and other associations with known HR-related molecular data. Specifically, further molecular data regarding TCGA-BRCA were obtained from Polak et al.¹³ (Figure 2). Together, these combined datasets allow for a more comprehensive understanding of the relationship between HRD scores inferred from morphological patterns and the underlying molecular mechanisms.

Variant calling in the AC-BRCA cohort

Somatic SNVs were called using an *in-house* pipeline following GATK's best practice guidelines as described previously⁴⁵. Briefly, the raw reads were aligned to the human reference genome GRCh38/hg38 using the Burrows–Wheeler aligner (BWA-MEM v0.7.17- r1188) with default settings. Next, all preprocessing steps were performed following the guidelines for best practices of the Genome Analysis Toolkit (GATK4)⁴⁶. Duplicated reads were removed using Picard (v2.22.8; <https://broadinstitute.github.io/picard/>) and base scores were recalibrated. Finally, somatic SNV calling was performed for the whole-exome data from analysis-ready BAM files using Mutect2 (v4.2) for tumor samples, followed by filtering with a panel of 169 unmatched non-tumor leukocyte samples (Panel of Normals). Extensive filtering was applied to remove samples with a low mapping quality, as well as strand and position biases and oxoguanine oxidative artifacts. Residual

germline mutations from the Genome Aggregation Database (<https://gnomad.broadinstitute.org/>) and the Online Archive of Brazilian Mutations (<http://abraom.ib.usp.br/>) were removed.

Mutational Signatures Estimation

Four different sets of mutational signatures were estimated for the TCGA-BRCA and AC-BRCA cohorts, given the availability of the data. For the TCGA-BRCA cohort, single base substitutions (SBS) were estimated from the COSMIC SBS V2 set, and small insertions and deletions (ID) with the COSMIC ID V3.4 set. Mutational data used for the estimation of both SBS and ID sets were retrieved from the MC3 dataset⁴⁷, and signature activities were estimated with sigProfiler⁴⁸. Copy number variation (CN) signature estimations were obtained from Steele et al.²⁵ and scarHRD scores from Loeffler et al.¹⁸. In the AC-BRCA cohort, the detection of SBS, CN, ID signatures, and scarHRD scores was performed as described by the respective method, based on the identified mutations as outlined in the previous section.

Among these signature sets, we are interested in signatures related to HRD, specifically SBS3, ID6, CN1, CN17, and the scarHRD. The SBS3 signature is a base substitution signature linked to defective homologous recombination DNA repair, characterized by its flat and featureless pattern. The ID6 signature is associated with homologous recombination-based DNA damage repair. CN1 and CN17 are defined by a 48-context copy number classification scheme, CN1 is primarily a signature associated with genomes with little or no copy number alteration and is also negatively associated with other signatures such as the scarHRD. CN17 is defined by segments of loss of heterozygosity (LOH) with TCNs ranging from 2 to 4, and heterozygous segments with TCNs ranging from 3 to 8, each segment being between 1 to 40 Mb in size²⁵. Finally, scarHRD evaluates the extent of copy number variations (CNVs), loss of heterozygosity (LOH), and large-scale state transitions (LSTs), which are similar genomic features associated with structural changes in the genome²⁶.

Aggregated score for HRD

Let $\mathcal{E} = \{\mathcal{E}_k\}$, $1 \leq k \leq K$, be the exposures to a set of k different subsets of biomarkers involved in HRD. The k th subset, $\mathcal{E}_k$, contains the exposures of each of i genome samples to m_k possible biomarkers of type k . More precisely, for each k let

$$\mathcal{E}_k = \{E_{k,b,i}\}, \quad 1 \leq b \leq m_k, \quad 1 \leq i \leq G,$$

where $E_{k,b,i}$ denotes the exposure of genome i to biomarker b of type k . In this study we consider four sets of different biomarkers for the definition of an HRD score: single base substitution mutational signatures, indel signatures, copy number signatures, and the scarHRD (index), more precisely,

$$\begin{aligned} \mathcal{E}_1 &= \{E_{1,b,i}\}, & b &\in \{\text{SBS1}, \dots, \text{SBS30}\}, \\ \mathcal{E}_2 &= \{E_{2,b,i}\}, & b &\in \{\text{ID1}, \dots, \text{ID23}\}, \\ \mathcal{E}_3 &= \{E_{3,b,i}\}, & b &\in \{\text{CN1}, \dots, \text{CN21}\}, \\ \mathcal{E}_4 &= \{E_{4,b,i}\}, & b &\in \{\text{scarHRD}\}. \end{aligned} \tag{1}$$

The exposure to each biomarker in the sets $\mathcal{E}_k$, $k = 1, 2, 3$, was estimated via non-negative matrix factorization methods. The estimated exposures are however not necessarily comparable, especially across the different k types. We consider therefore the following normalization,

$$\widehat{E}_{k,b,i} = \frac{E_{k,b,i}}{\sum_{b=1}^{m_k} E_{k,b,i}}, \quad \text{for } k = 1, 2, 3, \tag{2}$$

that is, we normalize the signature exposition in each sample, by calculating the proportion of the signature of interest by the total exposition of that sample. As such $\widehat{E}_{k,b,i} \in [0, 1]$ for all k, b and i . For $k = 4$, that is for the set constituted by the single biomarker scarHRD, we normalize with respect to the maximum observed value for all samples across the dataset

$$\widehat{E}_{4,\text{scarHRD},i} = \frac{E_{4,\text{scarHRD},i}}{\max_{1 \leq i \leq G} E_{4,\text{scarHRD},i}}. \tag{3}$$

We now focus on a small subset of the biomarkers in (1), known to be well correlated to HRD. These are defined as:

$$\mathcal{E}_1 = \{E_{1,\text{SBS3},i}\}, \quad \mathcal{E}_2 = \{E_{2,\text{ID6},i}\}, \quad \mathcal{E}_3 = \{E_{3,\text{CN1},i}, E_{3,\text{CN17},i}\}, \quad \mathcal{E}_4 = \{E_{4,\text{scarHRD},i}\}.$$

Now, for the subset $k = 3$ we combine the normalized exposures for the biomarker CN1 and CN17 as:

$$\widehat{E}_{3,\text{CN1CN17},i} = \frac{(1 - \widehat{E}_{3,\text{CN1},i}) + \widehat{E}_{3,\text{CN17},i}}{2},$$

because CN17 and CN1 are both associated with HRD, but CN1 exhibits a strong negative correlation to CN17. The set $\mathcal{E}_3$ is then re-defined as $\mathcal{E}_3 = \{E_{3,CN1CN17,i}\}$.

Next, we consider a regularization that takes into account the different number of genome samples of each type of tumor. Suppose there are T types of tumors and that in all $1 \leq i \leq G$ samples, G_t are from the type t , $1 \leq t \leq T$. For $\epsilon \in (0, 1)$ let

$$w_{k,b,t}(\epsilon) = \frac{1}{G_t} \sum_{i=1}^{G_t} 1\{\widehat{E}_{k,b,i} \geq \epsilon\}$$

where $1\{\widehat{E}_{k,b,i} \geq \epsilon\}$ is 1 if $\widehat{E}_{k,b,i} \geq \epsilon$ and 0 otherwise. The quantities $w_{k,b,t}(\epsilon)$ denote the relative fraction of samples in tumor of type t with exposures $\widehat{E}_{k,b,i}$ above ϵ . The addition of different weights for each tumor type can bias the scores and make them incomparable across different tumor types. To make sure all tumors remain on the same scale we re-scale the weights such that their sum for each tumor type is the same

$$\widehat{w}_{k,b,t}(\epsilon) = \frac{w_{k,b,t}(\epsilon)}{\sum_{k=1}^4 w_{k,b,t}(\epsilon)}$$

The HRD score for the i -th sample of tumor of type t is now computed as

$$s_{t,i}(\epsilon) = \sum_{k=1}^4 \sum_{b=1}^{m_k} \widehat{w}_{k,b,t}(\epsilon) \widehat{E}_{k,b,i}. \quad (4)$$

The analyses reported here were obtained with $\epsilon = 0.1$. This choice gave the best results in terms of the MSE (mean squared errors) and correlation for the values predicted by the regression model.

Whole-slide image (WSI) dataset

The image dataset used in this study was compiled from three different databases. The first set of WSIs was downloaded from The Cancer Genome Atlas (TCGA, gdc.cancer.gov), including 3,060 WSIs encompassing 2,819 patients across seven distinct tumor types. Specifically, the cancers analyzed included invasive breast carcinoma (TCGA-BRCA, $n = 898$), prostate adenocarcinoma (TCGA-PRAD, $n = 348$), pancreatic adenocarcinoma (TCGA-PAAD, $n = 124$), endometrial carcinoma (TCGA-UCEC, $n = 366$), colon adenocarcinoma (TCGA-COAD, $n = 342$), lung adenocarcinoma (TCGA-LUAD, $n = 416$) and stomach adenocarcinoma (TCGA-STAD, $n = 325$) (Figure 7). For external validation, we used two additional datasets: one from the Clinical Proteomic Tumor Analysis Consortium (CPTAC-BRCA, $n = 122$) and another from the A.C. Camargo Cancer Center (AC-BRCA, $n = 22$).

Slide preprocessing and feature extraction

Tiles of 256x256 pixels were extracted from each slide with a 50% overlap and 1 micrometer per pixel magnification. To filter the background tiles, we utilized the Otsu thresholding method⁴⁹ and a neural network trained to distinguish background tiles on histological images⁵⁰. Color normalization was performed using the method outlined by Macenko et al.⁵¹. We subsequently used the UNI model, a general-purpose self-supervised model for pathology trained more than 100 million images from more than 100,000 diagnostic H&E-stained WSIs spanning 20 major tissue types³², to extract a 1024-feature embedding vector from each color-normalized tile. This feature vector was then employed in subsequent steps to train the Multiple Instance Learning (MIL) model.

MIL Model Definition and Training

The model architecture was based on the regression MIL model proposed by El Nahhas et al.²⁰. Briefly, the model takes as input a bag of tiles for each patient's slide together with the observed panHRD score for that patient, computed as defined in Equation (4). Each tile is then represented by a 1024-feature vector encoded by the UNI model³², and an attention module estimates weights for each tile. The weighted sum of tile features is considered to form slide-level embedding and predict the panHRD score for each slide. Dropout⁵² layers with $p = 0.1$ were added to the attention module, and prior to the final slide-level embedding. The *Sigmoid* function was used as the final activation in order to limit the model output between 0 and 1 (Figure 7C).

The dataset was initially partitioned into 85% training and 15% test data. The training data were subsequently divided into five subsets using K-fold cross-validation, each containing a training set and a validation set. In cases where multiple slides were available from the same subject, the feature vectors from all the slides were aggregated into a single bag to produce a single inference per individual. We trained a MIL model for each fold to observe the impact of selecting different subsets of

training samples. Each model was trained with a batch size of 32 slides and a bag size of 4096 tiles, tiles were re-sampled every epoch when the slide had more than 4096 tiles and zero-padded when the slide had fewer, the padding was ignored by the attention module. The models were trained with the *AdamW* optimizer⁵³ using a fixed learning rate of 10^{-4} and the training ended when the validation loss stopped improving for 10 epochs. The mean squared error (MSE) was used as the loss function to optimize the regression model.

Actionable Tile Analysis

To visualize the actionable tiles across tumors, we examined the distribution of our model's attention and prediction scores on the WSI. For each fold, we selected the top 10 tiles with the highest attention scores from each slide in the test set. Next, we used the t-SNE method⁴¹ to project UNI's 1024-feature vector into a 2-dimensional space, allowing us to visualize the distinct tiles by their similarity. We applied a discretization method to divide the t-SNE 2D space into a 30x30 hexagonal grid, allowing us to analyze the distinct clusters of similar tiles using different criteria, such as the tumor type and predicted score. Figure 6A, 6B and Supplementary Figure 2 are colored by the mean score, and Figure 6C is aggregated by the label with the highest frequency in that cluster.

References

1. Stewart, M. D. *et al.* Homologous recombination deficiency: Concepts, definitions, and assays. *Oncologist* **27**, 167–174 (2022).
2. Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol* **26**, 52–64 (2016).
3. Beucher, A. *et al.* ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *EMBO J* **28**, 3413–3427 (2009).
4. Mateo, J. *et al.* DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. *N Engl J Med* **373**, 1697–1708 (2015).
5. Audeh, M. W. *et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet* **376**, 245–251 (2010).
6. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
7. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
8. Nguyen, L., W M Martens, J., Van Hoeck, A. & Cuppen, E. Pan-cancer landscape of homologous recombination deficiency. *Nat Commun* **11**, 5584 (2020).
9. Rempel, E. *et al.* Pan-cancer analysis of genomic scar patterns caused by homologous repair deficiency (HRD). *NPJ Precis Oncol* **6**, 36 (2022).
10. Popova, T. *et al.* Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. *Cancer Res* **72**, 5454–5462 (2012).
11. Birkbak, N. J. *et al.* Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. *Cancer Discov* **2**, 366–375 (2012).
12. Abkevich, V. *et al.* Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *Br J Cancer* **107**, 1776–1782 (2012).
13. Polak, P. *et al.* A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat Genet.* **49**, 1476–1486 (2017).
14. Ma, J., Setton, J., Lee, N. Y., Riaz, N. & Powell, S. N. The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. *Nat Commun* **9**, 3292 (2018).
15. Prosz, A. *et al.* Mutational signature-based identification of DNA repair deficient gastroesophageal adenocarcinomas for therapeutic targeting. *NPJ Precis Oncol* **8**, 87 (2024).
16. Echle, A. *et al.* Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer* **124**, 686–696 (2021).
17. Valieris, R. *et al.* Deep Learning Predicts Underlying Features on Pathology Images with Therapeutic Relevance for Breast and Gastric Cancer. *Cancers (Basel)* **12** (2020).

18. Loeffler, C. M. L. *et al.* Direct prediction of Homologous Recombination Deficiency from routine histology in ten different tumor types with attention-based Multiple Instance Learning: a development and validation study. *medRxiv* (2023).
19. Lazard, T. *et al.* Deep learning identifies morphological patterns of homologous recombination deficiency in luminal breast cancers from whole slide images. *Cell Rep Med* **3**, 100872 (2022).
20. El Nahhas, O. S. M. *et al.* Regression-based Deep-Learning predicts molecular biomarkers from pathology slides. *Nat Commun* **15**, 1253 (2024).
21. Bergstrom, E. N. *et al.* Deep learning artificial intelligence predicts homologous recombination deficiency and platinum response from histologic slides. *J. Clin. Oncol.* JCO2302641 (2024).
22. Krajewska, M., Fehrmann, R. S., de Vries, E. G. & van Vugt, M. A. Regulators of homologous recombination repair as novel targets for cancer treatment. *Front Genet.* **6**, 96 (2015).
23. Hoppe, M. M., Sundar, R., Tan, D. S. P. & Jeyasekharan, A. D. Biomarkers for Homologous Recombination Deficiency in Cancer. *J Natl Cancer Inst* **110**, 704–713 (2018).
24. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
25. Steele, C. D. *et al.* Signatures of copy number alterations in human cancer. *Nature* **606**, 984–991 (2022).
26. Sztupinszki, Z. *et al.* Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *NPJ Breast Cancer* **4**, 16 (2018).
27. M, G. & *et al.* The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).
28. Belli, C., Duso, B. A., Ferraro, E. & Curigliano, G. Homologous recombination deficiency in triple negative breast cancer. *The Breast* **45**, 15–21 (2019).
29. Stewart, M. D. *et al.* Homologous recombination deficiency: concepts, definitions, and assays. *The Oncol.* **27**, 167–174 (2022).
30. Takamatsu, S. *et al.* Utility of homologous recombination deficiency biomarkers across cancer types. *JCO Precis. Oncol.* (2022).
31. Nixon, M. J. *et al.* Pik3ca and map3k1 alterations imply luminal a status and are associated with clinical benefit from pan-pi3k inhibitor buparlisib and letrozole in er+ metastatic breast cancer. *npj Breast Cancer* **5**, DOI: [10.1038/s41523-019-0126-6](https://doi.org/10.1038/s41523-019-0126-6) (2019).
32. Chen, R. J. *et al.* Towards a general-purpose foundation model for computational pathology. *Nat Med* **30**, 850–862 (2024).
33. Tutt, A. *et al.* Carboplatin in brca1/2-mutated and triple-negative breast cancer brcaness subgroups: the tnt trial. *Nat. Medicine* **24**, 628–637, DOI: [10.1038/s41591-018-0009-7](https://doi.org/10.1038/s41591-018-0009-7) (2018).
34. Chopra, N. *et al.* Homologous recombination dna repair deficiency and parp inhibition activity in primary triple negative breast cancer. *Nat. Commun.* **11**, DOI: [10.1038/s41467-020-16142-7](https://doi.org/10.1038/s41467-020-16142-7) (2020).
35. Tutt, A. N. *et al.* Adjuvant olaparib for patients with brca1- or brca2-mutated breast cancer. *New Engl. J. Medicine* **384**, 2394–2405, DOI: [10.1056/nejmoa2105215](https://doi.org/10.1056/nejmoa2105215) (2021).
36. National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). The clinical proteomic tumor analysis consortium breast invasive carcinoma collection (cptac-brca) (2020).
37. Serra, H., Da Ines, O., Degroote, F., Gallego, M. E. & White, C. I. Roles of XRCC2, RAD51B and RAD51D in RAD51-independent SSA recombination. *PLoS Genet.* **9**, e1003971 (2013).
38. Mirza-Aghazadeh-Attari, M. *et al.* 53BP1: A key player of DNA damage response with critical functions in cancer. *DNA Repair (Amst.)* **73**, 110–119 (2019).
39. Chen, L. *et al.* Characterization of pik3ca and pik3r1 somatic mutations in chinese breast cancer patients. *Nat. Commun.* **9** (2018).
40. Reinhardt, K. *et al.* Pik3ca-mutations in breast cancer. *Breast Cancer Res. Treat.* **196**, 483–493 (2022).
41. van der Maaten, L. & Hinton, G. Visualizing data using t-sne. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
42. Surova, O. & Zhivotovsky, B. Various modes of cell death induced by DNA damage. *Oncogene* **32**, 3789–3797 (2013).
43. Keung, M. Y. T., Wu, Y. & Vadgama, J. V. PARP inhibitors as a therapeutic agent for homologous recombination deficiency in breast cancers. *J. Clin. Med.* **8**, 435 (2019).

44. McGrail, D. J. *et al.* Widespread brca1/2-independent homologous recombination defects are caused by alterations in rna-binding proteins. *Cell Reports Medicine* **4**, 101255, DOI: [10.1016/j.xcrm.2023.101255](https://doi.org/10.1016/j.xcrm.2023.101255) (2023).
45. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud*. 9781491975190 (O'Reilly Media, Inc., 2020).
46. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–303 (2010).
47. Ellrott, K. *et al.* Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst.* **6**, 271–281.e7 (2018).
48. Bergstrom, E. N. *et al.* SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC Genomics* **20**, 685 (2019).
49. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Syst. Man, Cybern.* **9**, 62–66 (1979).
50. Valieris, R. *et al.* Weakly-supervised deep learning models enable her2-low prediction from he stained slides. *Breast Cancer Res.* **26** (2024).
51. Macenko, M. *et al.* A method for normalizing histology slides for quantitative analysis (2009).
52. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
53. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization (2017).
54. Krug, K. *et al.* Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell* **183**, 1436–1456.e31 (2020).

Acknowledgments

The authors thank the pathology facility of the A.C.Camargo Cancer Center for the whole slide imaging scanning service and guidance. We would like to thank Paz Polak for valuable suggestions and critical review of the manuscript.

Author contributions statement

RV designed and implemented the AI-based algorithms. LM performed the explainability analysis and generated the figures for the article. AD and LR contributed to the data acquisition, preparation, and bioinformatics analysis. IS and MB contributed to the evaluation of slide images. GOS, LD, DNN, DMC, and EDN reviewed and edited the manuscript. DMC contributed to data curation and validation. ITS, RR, and RV developed the panHRD score. ITS contributed to conceptualization, funding acquisition, and supervised the study. ITS, RR and RV wrote the final version of the manuscript, on which all the authors provided comments. All authors agreed to the submission of this paper. The work reported in the paper has been performed by the authors unless specified in the text.

Data availability

The molecular data and H&E slides for the TCGA dataset are available at <https://portal.gdc.cancer.gov/>. The molecular biomarker data from TCGA-BRCA were obtained from Polak et al.¹³. Copy number variation (CN) signature estimations were obtained from Steele et al.²⁵ and scarHRD scores from Loeffler et al.¹⁸. The molecular data for the BRCA-CPTAC cohort⁵⁴ are available at <https://portal.gdc.cancer.gov/>, and the histopathology data is available at The Cancer Imaging Archive³⁶. The panHRD scores for TCGA data used in this work are available in Supplementary Table 3. Further details necessary to reanalyze the data in this paper can be provided by the lead contact upon request.

Code availability

All code used in training the models is available at <https://github.com/tojallab/reg-mil>, under the MIT license.

Funding

This work was partially supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), grant #22/12991-0.

Declaration of Competing Interests

All authors must disclose any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work.

Ethics Approval and Consent to Participate

The overall study was approved by the Ethics Committee of Fundação Antônio Prudente (#3028/20).