

LARA DE SOUZA MORENO

ANÁLISE RADIÔMICA DO CÂNCER COLORRETAL CORRELACIONADA COM  
MARCADORES TUMORAIS E DE BIÓPSIA LÍQUIDA

BRASÍLIA

2025

UNIVERSIDADE DE BRASÍLIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MÉDICAS FACULDADE DE  
CIÊNCIAS DA SAÚDE

LARA DE SOUZA MORENO

ANÁLISE RADIÔMICA DO CÂNCER COLORRETAL CORRELACIONADA COM  
MARCADORES TUMORAIS E DE BIÓPSIA LÍQUIDA

Dissertação apresentada  
como requisito para a obtenção do  
título de Mestre em Ciências da  
Saúde pelo Programa de Pós-  
Graduação em Ciências Médicas da  
Universidade de Brasília.

Orientador: Prof. Dr. Fábio Pittella Silva

Co-orientador: Prof. Dr. Tony Alexandre de Medeiros

BRASÍLIA

2025

## **LARA DE SOUZA MORENO**

Análise radiômica do câncer colorretal correlacionada com marcadores tumorais e de biópsia líquida

Dissertação apresentada como requisito para a obtenção do título de Mestre em Ciências da Saúde pelo Programa de Pós-Graduação em Ciências Médicas da Universidade de Brasília.

## **BANCA EXAMINADORA**

Prof. Dr. Fábio Pittella Silva – (Presidente)

Universidade de Brasília

Prof. Dr. Alexandre Sérgio de Araújo Bezerra

Universidade de Brasília

Prof. Dr. Sérgio Augusto Lopes de Souza

Universidade Federal do Rio de Janeiro

Prof. Dr. Rodrigo Haddad

Universidade de Brasília

## RESUMO

O câncer colorretal (CCR) é uma das neoplasias de maior impacto global, tanto em incidência quanto em mortalidade. A radiômica, ao extrair informações quantitativas de imagens médicas, tem emergido como uma ferramenta promissora para caracterizar a heterogeneidade tumoral de forma objetiva e potencialmente prever desfechos clínicos e biológicos. Este estudo teve como objetivo avaliar a capacidade de assinaturas radiômicas derivadas de tomografias computadorizadas (TC) em distinguir tumores T1–T2 de T3–T4, correlacionando esses achados com o perfil molecular determinado por sequenciamento tumoral e biópsia líquida (cfDNA/ctDNA).

Foram incluídos 53 pacientes (55 tumores) com diagnóstico confirmado de adenocarcinoma colorretal, sem tratamento prévio. As imagens de TC trifásica foram segmentadas manualmente na fase portal, e um total de 107 características radiômicas foram extraídas usando a biblioteca PyRadiomics, agrupadas em medidas de forma, primeira ordem e textura. O pré-processamento incluiu normalização (z-score), redução de dimensionalidade pelo análise de componentes principais (PCA) e aumento sintético de dados via SMOTE. Foram realizados 990 experimentos para predição de estágio T e 990 para estágio N, utilizando 11 algoritmos de aprendizado de máquina e 15 subconjuntos de features, totalizando 100 repetições por experimento.

A melhor performance para predição do estágio T foi obtida com o classificador Regressão Logística, utilizando apenas features de forma, normalizadas, aumentadas via SMOTE e reduzidas por PCA, apresentando sensibilidade média de 72,1%, especificidade de 68,0%, acurácia balanceada de 70,0% e AUC de 75,1%. Para predição do estágio N, o melhor desempenho foi obtido com o AdaBoost usando exclusivamente features de textura, normalizadas e balanceadas por SMOTE, com sensibilidade média de 74,2%, especificidade de 62,2% e AUC de 75,0%.

Para a análise da biópsia líquida obtemos o DNA livre circulante (cfDNA) dos pacientes no momento da realização da TC ao diagnóstico. O DNA circulante tumoral (ctDNA) foi identificado através do sequenciamento do cfDNA utilizando-se painel compreensivo com mais de 240 mutações distribuídas nos 14 genes com maior frequência de alterações no câncer colorretal. A quantificação do DNA circulante

tumoral (ctDNA) foi representada através da frequência alélica total (VAF) das mutações identificadas. Observamos variabilidade substancial na carga mutacional circulante (ctDNA/cfDNA), sem relação linear clara com o volume tumoral. Contudo, casos específicos apresentaram cargas mutacionais elevadas apesar de lesões pequenas, sugerindo perfis biológicos mais agressivos e liberação desproporcional de DNA tumoral.

Este estudo demonstra que assinaturas radiômicas derivadas de TC possuem potencial para auxiliar no estadiamento pré-operatório do CCR e para integração futura com biomarcadores moleculares. Entretanto, limitações como o tamanho reduzido da amostra, a natureza unicêntrica e a necessidade de validação externa restringem a generalização dos resultados. No futuro, espera-se incorporar dados genômicos e de biópsia líquida às assinaturas radiômicas desenvolvidas, com o objetivo de prever perfis genéticos e aprimorar a estratificação de risco na prática clínica.

**Palavras-chave:** Radiômica; Câncer colorretal; Tomografia computadorizada; Aprendizado de máquina; Biópsia líquida; ctDNA; Radiogenômica.

## ABSTRACT

Colorectal cancer (CRC) remains one of the most prevalent and lethal malignancies worldwide. Radiomics, by extracting quantitative descriptors from medical imaging, has emerged as a promising tool to objectively characterize tumor heterogeneity and potentially predict clinical and biological outcomes. This study aimed to evaluate the ability of radiomic signatures derived from computed tomography (CT) to distinguish early-stage (T1–T2) from advanced-stage (T3–T4) tumors and to correlate these imaging features with molecular profiles obtained through tumor sequencing and liquid biopsy (cfDNA/ctDNA).

A total of 53 patients (55 tumors) with histologically confirmed colorectal adenocarcinoma and no prior treatment were included. Portal venous–phase CT scans were manually segmented, and 107 radiomic features were extracted using the PyRadiomics library, including shape, first-order, and higher-order texture descriptors. Preprocessing steps comprised z-score normalization, dimensionality reduction using PCA, and synthetic data augmentation with SMOTE. A full experimental pipeline was conducted with 990 models for T-stage prediction and 990 for N-stage prediction, combining 15 feature subsets, 11 machine-learning algorithms, and 6 preprocessing strategies, each repeated 100 times.

For T-stage classification, the best performance was achieved by Logistic Regression using shape features that had been normalized, SMOTE-balanced, and reduced by PCA, yielding a mean sensitivity of 72.1%, specificity of 68.0%, balanced accuracy of 70.0%, and AUC of 75.1%. For N-stage prediction, the AdaBoost classifier using only texture features with normalization and SMOTE balancing outperformed other models, showing a mean sensitivity of 74.2%, specificity of 62.2%, and AUC of 75.0%.

For liquid biopsy analysis, circulating cell-free DNA (cfDNA) was obtained from patients at the time of diagnostic CT acquisition. Circulating tumor DNA (ctDNA) was identified through cfDNA sequencing using a comprehensive panel covering more than 240 mutations across the 14 genes most frequently altered in colorectal cancer. Quantification of ctDNA was represented by the total variant allele frequency (VAF) of the identified mutations. Substantial variability was observed in the circulating

mutational burden (ctDNA/cfDNA), with no clear linear relationship to tumor volume. However, specific cases exhibited high mutational burdens despite small lesion size, suggesting more aggressive biological profiles and disproportionate release of tumor-derived DNA.

This study demonstrates the potential of CT-based radiomic signatures to assist in preoperative staging of CRC and highlights the future value of integrating imaging features with molecular biomarkers. However, limitations such as small sample size, single-center design, manual segmentation, and the absence of external validation constrain broader applicability. Future directions include incorporating tumor genomic data and liquid biopsy results to enhance radiomic signatures and enable the prediction of molecular profiles, contributing to precision oncology in colorectal cancer.

**Keywords:** Radiomics; Colorectal cancer; Computed tomography; Machine learning; Liquid biopsy; ctDNA; Radiogenomics.

# Sumário

<b>1. INTRODUÇÃO</b>	<b>13</b>
<b>1.1. Parte 1: O Câncer Colorretal (CCR)</b>	<b>13</b>
1.1.1. Diagnóstico	14
1.1.2. Tratamento	17
<b>1.2. Parte 2: Radiômica no câncer colorretal</b>	<b>18</b>
1.2.1. Fundamentos e fluxo de trabalho	18
1.2.2. Padronização, reprodutibilidade e validação	19
1.2.3. Aplicações clínicas no câncer colorretal	20
1.2.4. Radiogenômica e integração molecular	21
1.2.5. Modelos baseados em inteligência artificial	22
1.2.6. Limitações e perspectivas futuras	22
<b>1.3. Parte 3: Biópsia líquida no câncer colorretal</b>	<b>23</b>
1.3.1. Fundamentos e princípios biológicos	23
1.3.2. DNA livre circulante (cfDNA) e DNA tumoral circulante (ctDNA)	24
1.3.3. Aplicações clínicas no câncer colorretal	25
1.3.4. Técnicas e desafios metodológicos	26
1.3.5. Perspectivas futuras e integração com medicina de precisão	27
<b>2. OBJETIVOS</b>	<b>28</b>
<b>2.1. Objetivo geral</b>	<b>28</b>
2.1.1. Objetivos específicos	28
<b>3. METODOLOGIA</b>	<b>29</b>
<b>3.1. Desenho do estudo</b>	<b>29</b>
<b>3.2. Aspectos éticos</b>	<b>29</b>
<b>3.3. População do estudo</b>	<b>29</b>
<b>3.4. Dados do estadiamento histopatológico</b>	<b>30</b>
<b>3.5. Aquisição das imagens radiológicas e análise subjetiva</b>	<b>31</b>
<b>3.6. Segmentação tumoral das imagens</b>	<b>31</b>
<b>3.7. Pré-processamento</b>	<b>32</b>
3.7.1. Extração das características radiômicas	32
3.7.2. Normalização das características	36
3.7.3. Redução da dimensionalidade dos dados	37
3.7.4. Aumento de dados	37
<b>3.8. Implementação e classificação</b>	<b>38</b>
<b>3.9. Avaliação da performance da assinatura</b>	<b>39</b>
<b>3.10. Análise do perfil molecular pela Biópsia Líquida</b>	<b>43</b>
3.10.1. Extração e quantificação do cfDNA do plasma	43
3.10.2. Preparação de bibliotecas de cfDNA e Sequenciamento de Nova Geração (NGS)	43
3.10.3. Análise bioinformática e frequência alélica	44
<b>3.11. Análise da carga mutacional no plasma correlacionada ao volume tumoral radiológico</b>	<b>44</b>
<b>4. Resultados</b>	<b>46</b>
<b>4.1. População do estudo</b>	<b>46</b>
<b>4.2. Volumetria tumoral</b>	<b>46</b>
<b>4.3. Assinatura radiômica</b>	<b>48</b>

4.4.	Correlação de dados com a biópsia líquida .....	54
5.	<b>Discussão</b> .....	57
5.1.	Assinatura radiômica .....	57
5.2.	Correlação de dados com a biópsia líquida .....	60
5.3.	Correlação radiômica e biópsia líquida .....	63
5.4.	Limitações .....	63
6.	<b>Conclusão</b> .....	64
7.	<b>Conclusão dos Objetivos Específicos</b> .....	64
8.	<b>Referências</b> .....	66

## LISTA DE TABELAS E FIGURAS

Figura 1: Incidência do CCR por estado brasileiro. Fonte: Revista Brasileira de Cancerologia, 2023 (3).	13
Figura 2: Carcinogênese do CCR. Fonte: Lecturio: Carcinoma Colorretal, 2025 (5).	14
Figura 3: Estadiamento do CCR por RM.	15
Figura 4: Fluxo de trabalho da análise radiômica. Fonte: ResearchGate, 2019 (31).	19
Figura 5: Utilidade clínica da biópsia líquida. Fonte: Biomedicines, 2021 (62).	24
Figura 6. Desenho do Estudo	29
Tabela 1: Grupos de pacientes separados por classificação T.	30
Figura 7. Segmentação da Região Tumoral	32
Tabela 2- Features não-texturais tradicionais extraídas de 41 pacientes.	33
Tabela 3- Features não-texturais de primeira-ordem extraídas de 41 pacientes.	33
Tabela 4 - Higher-order texture features extraídas de 41 pacientes.	34
Tabela 5. Subconjuntos de features radiômicas utilizados nos experimentos e o número de features radiômicas por classe em cada subconjunto.	42
Tabela 6. Características clínicas basais da coorte do estudo (n = 53 pacientes).	46
Figura 8. Distribuição do volume tumoral por Violin plot.	47
Figura 9. Distribuição do volume tumoral por Histograma.	47
Figura 10. Representação da segmentação do paciente CBL 38 e sua peça cirúrgica correspondente.	48
Figura 11. Representação da segmentação do paciente CBL 18 e peça cirúrgica correspondente.	48
Tabela 7. Avaliação dos diferentes cenários de pré-processamento: (i) features brutas, (ii) features normalizadas, (iii) features normalizadas e reduzidas por PCA, (iv) features normalizadas com aumento de dados via SMOTE, (v) features normalizadas e reduzidas por PCA seguidas de aumento de dados via SMOTE e (vi) features normalizadas com aumento de dados via SMOTE seguidas de redução por PCA. Os valores indicam a porcentagem de experimentos de predição do estágio T e do estágio N — entre um total de 165 experimentos em cada caso, abrangendo todas as combinações de 15 conjuntos de features e 11 classificadores — que foram considerados “bem-sucedidos”, ou seja, aqueles que combinaram sensibilidade média superior a 70% e especificidade média superior a 50%.	49
Tabela 8. Avaliação das classes de features radiômicas utilizadas para a predição dos estágios T e N. Os valores indicam a porcentagem de classificadores — entre um total de 11 — que foram considerados “bem-sucedidos”, isto é, que combinaram sensibilidade média superior a 70% e especificidade média superior a 50%.	50
Tabela 9. Avaliação dos subconjuntos que combinam diferentes classes de features radiômicas, mostrando também a porcentagem de classificadores (entre 11) que foram “bem-sucedidos” segundo o mesmo critério de desempenho (sensibilidade média >70% e especificidade média >50%).	51
Tabela 10. resultados da predição do estágio T obtidos com diferentes algoritmos de classificação. São exibidos os valores médios e os respectivos intervalos de confiança de 95% (todos expressos em porcentagem). Os resultados foram obtidos utilizando o conjunto de features shape14 em conjunto com a estratégia de pré-processamento (vi).	52
Tabela 11. Resultados da predição do estágio N obtidos com diferentes algoritmos de classificação. São exibidos os valores médios e os respectivos intervalos de confiança de 95% (todos expressos em porcentagem). Os resultados foram obtidos utilizando o conjunto de features texture75 em conjunto com a estratégia de pré-processamento (iv).	52
Figura 12. Matrizes de confusão para: (a) a predição do estágio T, utilizando o classificador de regressão logística aplicado ao conjunto de features shape14 com a estratégia de pré-processamento (vi); e (b) a predição do estágio N, utilizando o classificador AdaBoost aplicado ao conjunto de features texture75 com a estratégia de pré-processamento (iv).	53
Tabela 12. Dados individuais dos pacientes incluídos no estudo	54
Figura 13. Relação entre a concentração plasmática de cfDNA obtida no momento do diagnóstico e o volume tumoral segmentado nas imagens de tomografia computadorizada. Cada ponto representa um paciente. Observa-se uma tendência de	

correlação positiva fraca entre as variáveis, embora sem significância estatística ( $\rho = 0,32$ ; $p = 0,085$ ).....	56
Figura 14. Relação entre a frequência alélica do ctDNA (VAF%) e o volume tumoral segmentado nas imagens de tomografia computadorizada. Cada ponto representa um paciente. Não se observou correlação significativa entre as variáveis ( $\rho = 0,16$ ; $p = 0,42$ ) .....	56
Figura 15. Relação entre a carga mutacional ajustada e o volume tumoral segmentado nas imagens de tomografia computadorizada. Cada ponto representa um paciente. Não se observou correlação significativa entre as variáveis ( $\rho = -0,21$ ; $p = 0,276$ ).....	57

## **LISTA DE ABREVIATURAS**

TC – Tomografia computadorizada

CCR – Câncer colorretal

APC – Polipose adenomatosa coli

PACS – Sistema de arquivamento e comunicação de imagens

DICOM – Comunicação de Imagens Digitais em Medicina

RIS – Sistema de Informação de Radiologia

UDI – Unidade de Diagnóstico por Imagem

NRRD – Nearly Raw Raster Data

TCLE – Termo de Consentimento Livre e Esclarecido

### **Radiômica**

GLCM – Gray Level Co-occurrence Matrix

GLSZM – Gray Level Size Zone Matrix

GLRLM – Gray Level Run Length Matrix

NGTDM – Neighbouring Gray Tone Difference Matrix

GLDM – Gray Level Dependence Matrix

### **Pré-processamento / Modelagem**

PCA – Análise de Componentes Principais

SMOTE – Synthetic Minority Over-sampling Technique

### **Classificadores**

LR – Regressão Logística

NB – Naive Bayes

DT – Árvores de Decisão

ET – Extra Trees

RF – Random Forest

MLP – Multi-Layer Perceptron

AB – AdaBoost

GB – Gradient Boosting

LDA – Linear Discriminant Analysis

XGB – XGBoost

### **Métricas e termos de avaliação**

Acc – Acurácia

Se – Sensibilidade

Spe – Especificidade

Pre – Precisão

F1 – F1-Score

BCA – Acurácia balanceada

TP – Verdadeiros positivos

FP – Falsos positivos

TN – Verdadeiros negativos

FN – Falsos negativos

AUC – Área sob a curva ROC

NPV – Valor preditivo negativo

### **Genômica / Biópsia líquida**

cfDNA – DNA livre circulante

ctDNA – DNA tumoral circulante

MRD – Doença Residual Mínima

ddPCR – Droplet Digital Polymerase Chain Reaction

BEAMing – Beads, Emulsion, Amplification, Magnetics

NGS – Next-Generation Sequencing

VAF – Variant Allele Frequency

SNP – Single Nucleotide Polymorphism

MSI – Instabilidade de microssatélites

MSI-H – Instabilidade de microssatélites alta

MSS – Microssatélite estável

CNV – Copy Number Variation

LGR – Large Genomic Rearrangements

TMB – Tumor Mutational Burden

HRR – Homologous Recombination Repair

HRD – Homologous Recombination Deficiency

### **Plataformas e softwares**

VEP – Variant Effect Predictor

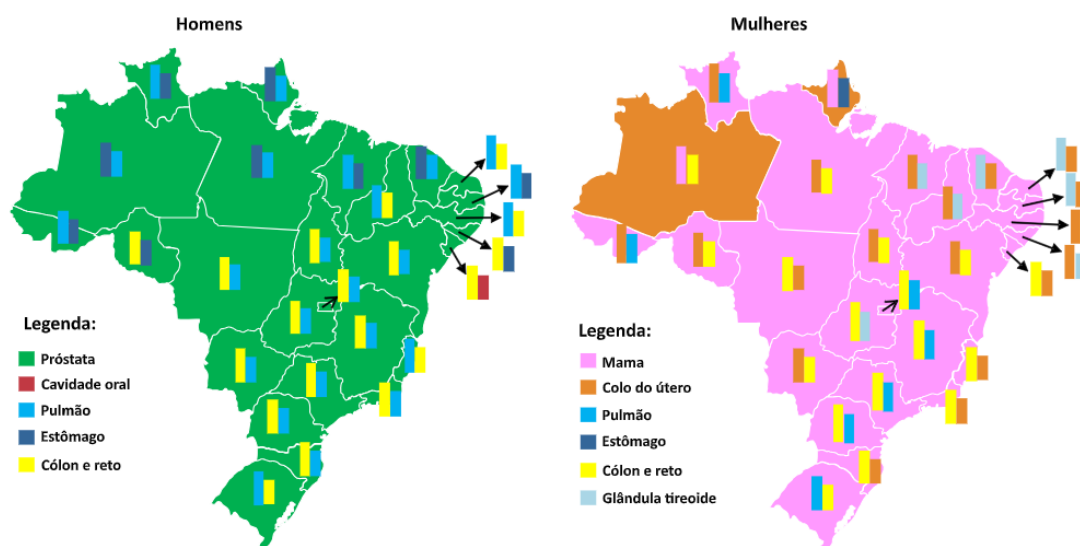
RQS – Radiomics Quality Score

IBSI – Image Biomarker Standardisation Initiative

## 1. INTRODUÇÃO

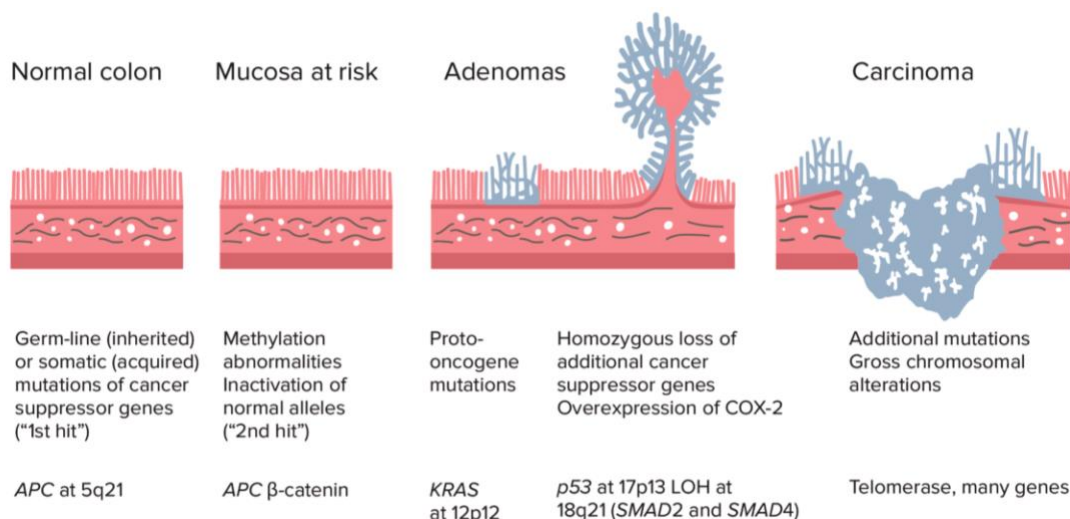
### 1.1. Parte 1: O Câncer Colorretal (CCR)

O câncer colorretal (CCR) é o terceiro tipo de neoplasia mais comumente diagnosticado e a terceira principal causa de morte relacionada ao câncer no mundo (1). Para o triênio 2023–2025, o Instituto Nacional de Câncer (INCA) estima cerca de 45.630 novos casos no Brasil (2).



**Figura 1: Incidência do CCR por estado brasileiro.**  
**Fonte: Revista Brasileira de Cancerologia, 2023 (3).**

Sua incidência está associada tanto a fatores ambientais — como dieta ocidental e sedentarismo — quanto a síndromes hereditárias, como a polipose adenomatosa familiar e a síndrome de *Lynch* (1,4).



**Figura 2: Carcinogênese do CCR.**

Fonte: Lecturio: Carcinoma Colorretal, 2025 (5).

Do ponto de vista molecular, a carcinogênese colorretal decorre de uma progressão tumoral ordenada, marcada por alterações genéticas e epigenéticas que transformam o epitélio normal em adenoma e, posteriormente, em carcinoma invasivo. Esse processo, conhecido como sequência adenoma–carcinoma, envolve mutações no gene *APC* (*Adenomatous Polyposis Coli*), seguidas por alterações em *KRAS*, *PIK3CA* e *TP53*, que modulam a proliferação celular, evasão apoptótica e instabilidade genômica. Três principais vias moleculares caracterizam o CCR: instabilidade cromossômica (CIN), instabilidade de microssatélites (MSI) e o fenótipo de hipermetilação de ilhas CpG (CIMP), frequentemente sobrepostos e com implicações prognósticas distintas(6).

A maioria dos casos é esporádica (≈85%), enquanto 10–15% têm base hereditária (7). Além disso, observa-se tendência crescente de diagnósticos em pacientes jovens, sobretudo em países de renda média e alta, o que tem impulsionado discussões sobre redução da idade de rastreamento(1,7). No Brasil, análises regionais recentes reforçam heterogeneidade de mortalidade e variações temporais que impactam estratégias locais de prevenção e cuidado(4,8).

### 1.1.1. Diagnóstico

O diagnóstico e o estadiamento do CCR envolvem uma combinação de métodos endoscópicos, laboratoriais e de imagem.

A colonoscopia é o padrão-ouro para rastreamento e diagnóstico, permitindo visualização direta da mucosa e ressecção de lesões precursoras. Em pacientes com

contraindicação, recusa ou risco aumentado de complicações, a colonoscopia virtual por tomografia computadorizada (CTC) constitui alternativa não invasiva com acurácia comparável à colonoscopia para detecção de pólipos  $\geq 10$  mm(4,9).

Entre os métodos laboratoriais, a pesquisa de sangue oculto nas fezes (PSOF) e os testes de DNA fecal têm papel consolidado na triagem populacional, embora apresentem menor sensibilidade em lesões planas (10).

O estadiamento radiológico é determinante para a conduta terapêutica. A tomografia computadorizada (TC) contrastada de tórax, abdome e pelve é o exame inicial de escolha, permitindo avaliar extensão locorregional e metástases hepáticas e pulmonares (11). A ressonância magnética (RM) é o método de escolha para o estadiamento do reto, fornecendo informações precisas sobre o grau de invasão tumoral (T), o envolvimento da fásia mesorretal e a presença de invasão venosa extramural (EMVI) — dados fundamentais para o planejamento cirúrgico e indicação de terapia neoadjuvante (12). O PET-CT tem papel complementar em casos de recidiva, doença metastática oculta ou achados inconclusivos nos exames convencionais.(13)

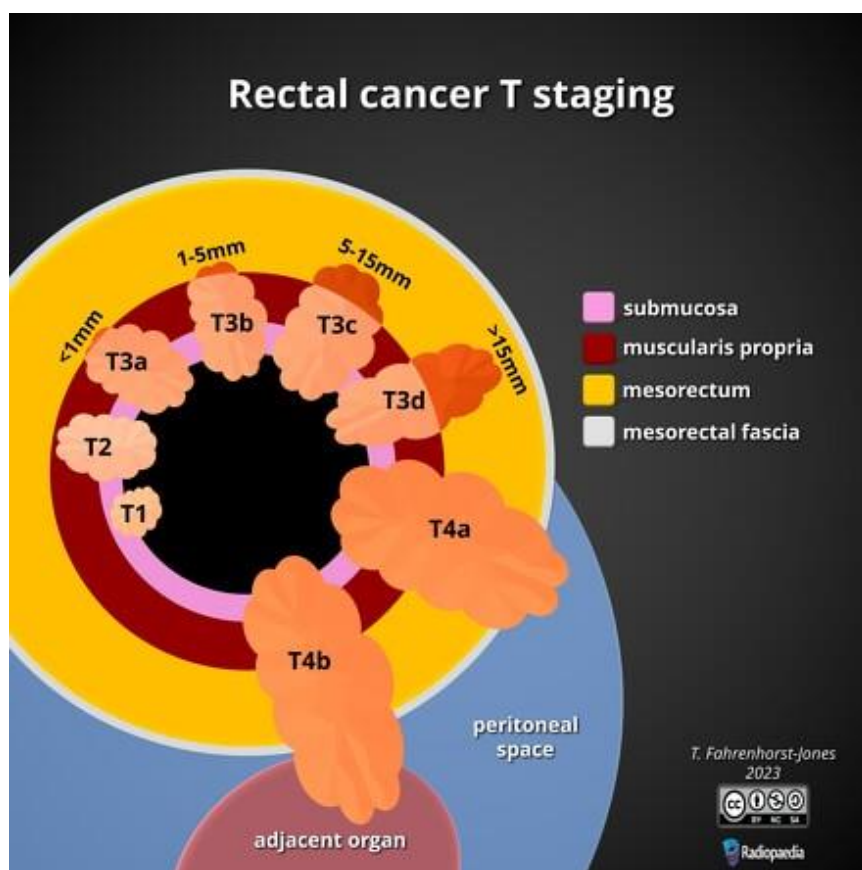


Figura 3: Estadiamento do CCR por RM.

Fonte: Radiopaedia Australia Pty Ltd, 2025 (14).

O estadiamento histopatológico baseia-se na classificação TNM do *American Joint Committee on Cancer* (AJCC) (15). O componente T reflete a profundidade de invasão tumoral na parede intestinal, variando desde tumores restritos à mucosa e submucosa (Tis–T1) até aqueles que invadem estruturas adjacentes (T4). O componente N é determinado pela presença e pelo número de linfonodos regionais comprometidos por metástases, enquanto o M indica a presença de metástases à distância (16). Além do TNM, a avaliação histopatológica inclui fatores prognósticos relevantes, como grau de diferenciação tumoral, invasão linfovascular e perineural, margens cirúrgicas e status de instabilidade de microssatélites (MSI), os quais contribuem para a estratificação prognóstica e definição de terapia adjuvante (16). Esses parâmetros influenciam diretamente as decisões sobre a abordagem cirúrgica, o uso de terapias adjuvantes ou neoadjuvantes e o prognóstico geral. Tumores T3–T4 geralmente apresentam comportamento mais agressivo, maior risco de recorrência e maior necessidade de terapia multimodal em comparação com lesões T1–T2.

O estadiamento clínico do câncer colorretal utiliza exames de imagem para estimar as categorias T, N e M e, a partir delas, classificar os estágios I a IV antes do tratamento (17). O estágio I clínico é sugerido quando a imagem demonstra tumor restrito à parede intestinal, sem sinais de invasão extramural ou linfonodos suspeitos(17). O estágio II clínico é caracterizado por invasão além da muscular própria, com extensão extramural ou comprometimento da serosa, mas sem evidência radiológica de linfonodos metastáticos ou metástases à distância(17). O estágio III clínico é definido pela presença de linfonodos regionais suspeitos, com base em critérios morfológicos e funcionais de imagem, independentemente da extensão tumoral local(17). O estágio IV clínico corresponde à identificação de metástases à distância em exames de estadiamento, como TC de tórax, abdome e pelve ou RM hepática, sendo determinante para a indicação de tratamento sistêmico(17).

Pacientes com doença em estágio III são amplamente reconhecidos como aqueles que mais se beneficiam da quimioterapia adjuvante (18), enquanto, no estágio II, o benefício é geralmente restrito a casos com fatores clínicos de risco e permanece tema de debate contínuo(19).

### 1.1.2. Tratamento

O manejo do CCR é multidisciplinar, incluindo cirurgia, quimioterapia, terapias-alvo, imunoterapia e, em casos selecionados, radioterapia. A escolha terapêutica depende do estágio, perfil molecular e condições clínicas do paciente.

#### 1.1.2.1. Cirurgia e tratamento locorregional

A ressecção cirúrgica completa (R0) é o tratamento de escolha para doença localizada (estágios I–III). No câncer de reto, a terapia neoadjuvante total (*Total Neoadjuvant Therapy*, TNT) — que combina quimiorradioterapia e quimioterapia sistêmica — tem se mostrado eficaz em melhorar o controle locorregional e aumentar as taxas de preservação esfinteriana (20). Em pacientes com doença metastática isolada, especialmente hepática, a ressecção de metástases pode alcançar sobrevida em 5 anos de até 40% em casos selecionados, desde que apoiada por avaliação radiológica de alta resolução e planejamento multidisciplinar (21).

#### 1.1.2.2. Quimioterapia e terapias-alvo

Nos estágios II de alto risco e III, a quimioterapia adjuvante baseada em fluoropirimidinas e oxaliplatino (FOLFOX ou CAPOX) permanece o padrão (11). Embora o benefício absoluto em sobrevida global para o estágio II seja modesto, recomenda-se individualização com base em fatores clínico-patológicos e diretrizes das sociedades médicas (11).

Na doença metastática, a seleção de esquemas depende do perfil genético (por exemplo, RAS/BRAF) e da localização tumoral. Tumores RAS/BRAF selvagens e localização esquerda podem se beneficiar de agentes anti-EGFR (cetuximabe, panitumumabe), enquanto tumores com mutações em RAS ou BRAF tendem a regimes baseados em anti-VEGF (bevacizumabe); quando indicado, a intensificação com FOLFOXIRI + bevacizumabe é opção em casos selecionados(22,23).

17/01/2026 20:01:00Evidências recentes demonstram que a arquitetura e o grau de maturação da vasculatura tumoral exercem influência decisiva sobre a resposta terapêutica em metástases hepáticas de câncer colorretal. Lesões com vasos imaturos e dependentes de angiogênese — como as de padrão histopatológico

desmoplásico (DHGP) — apresentam maior suscetibilidade aos agentes antiangiogênicos, refletindo uma redução significativa da densidade microvascular após o tratamento. Em contraste, metástases que utilizam co-optação vascular — como as de padrão de crescimento por substituição (RHGP) — mantêm vasos maduros e funcionais provenientes da microcirculação hepática, exibindo resistência a terapias anti-VEGF e menor necrose tumoral. Assim, a organização vascular tumoral, mais do que a simples densidade microvascular, emerge como determinante essencial da perfusão, da entrega de fármacos e, conseqüentemente, dos desfechos clínicos.(24)

### **1.1.2.3. Imunoterapia e terapias emergentes**

A introdução dos inibidores de checkpoint imunológico revolucionou o tratamento de pacientes com instabilidade de microssatélites alta (MSI-H) ou deficiência de reparo de DNA (dMMR), com taxas de resposta de 40–50% e sobrevidas prolongadas(25,26). Combinações e estratégias para tumores com estabilidade de microssatélites (MSS) permanecem em investigação. Novas abordagens incluem vacinas de neoantígenos, terapias celulares (CAR-T e TCR-T) e imunoterapia intratumoral, voltadas a modular o microambiente e ampliar a resposta imune antitumoral(27).

## **1.2. Parte 2: Radiômica no câncer colorretal**

### **1.2.1. Fundamentos e fluxo de trabalho**

A radiômica consiste na extração de características quantitativas de imagens médicas convencionais, convertendo dados visuais em parâmetros numéricos que descrevem a morfologia, a textura e a heterogeneidade tumoral de modo objetivo. Esse processo permite caracterizar o fenótipo tumoral invisível à avaliação qualitativa, oferecendo subsídios para diagnósticos, prognósticos e predições terapêuticas (28,29).

O fluxo de trabalho radiômico compreende quatro etapas principais: (i) aquisição e pré-processamento da imagem — com padronização de protocolos, normalização da intensidade e discretização do histograma; (ii) segmentação do volume de interesse (ROI) de forma manual, semiautomática ou automatizada; (iii)

extração das *features* (características radiômicas); e (iv) modelagem e validação de desempenho. As *features* radiômicas são geralmente classificadas em primeira ordem (*first order* - distribuição de intensidades), forma (*shape* - medidas geométricas), e textura (*higher order texture* - relações espaciais entre voxels), podendo ser complementadas por atributos transformados via *wavelet* ou *Laplacian-of-Gaussian*.(30)

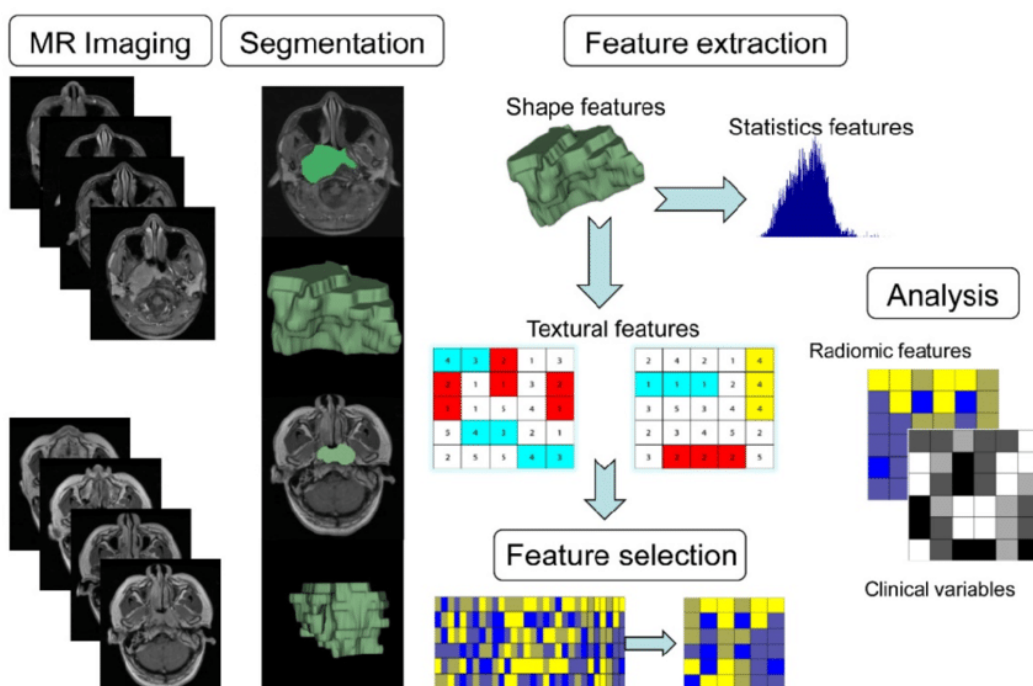


Figura 4: Fluxo de trabalho da análise radiômica.

Fonte: ResearchGate, 2019 (31).

Softwares padronizados, como o *PyRadiomics*, seguem as definições do *Image Biomarker Standardisation Initiative* (IBSI) para garantir reprodutibilidade entre plataformas (32). A seleção de *features* robustas geralmente é conduzida por coeficiente de correlação intraclasse ( $ICC > 0,80$ ), teste de variância e regularização penalizada (*Least Absolute Shrinkage and Selection Operator* – LASSO). A etapa final envolve algoritmos supervisionados, como *Random Forest*, *Support Vector Machine* (SVM), *XGBoost* e regressão logística multivariada, frequentemente avaliados por AUC (*Area Under the Curve*), acurácia e f-score.

### 1.2.2. Padronização, reprodutibilidade e validação

A heterogeneidade de protocolos de aquisição, reconstrução e segmentação representa um dos maiores desafios da radiômica. Variações de aparelhos, espessura

de corte, técnicas de reconstrução e contrastes diferentes podem alterar os valores das *features*. Por isso, recomenda-se detalhar os parâmetros de aquisição e empregar estratégias de normalização (z-score, re-amostragem isotrópica) (33).

A segmentação também introduz variabilidade inter- e intraobservador. Estudos em CCR sugerem que volumes tridimensionais delineados manualmente com validação por ICC oferecem melhor captura da heterogeneidade tumoral que amostras bidimensionais isolada. (34) Métodos automáticos baseados em U-Net (35) ou level set (36) tem sido explorados para reduzir tempo e subjetividade, mas ainda carecem de validação multicêntrica(37).

A reprodutibilidade das *features* deve ser testada antes da modelagem. A exclusão de variáveis instáveis reduz o risco de *overfitting* e melhora a generalização. Adicionalmente, o uso do *Radiomics Quality Score* (RQS) vem sendo incentivado para classificar a qualidade metodológica dos estudos (38). Modelos robustos devem incluir validação interna por *cross-validation* ou *bootstrapping* e validação externa em coortes independentes, relatando intervalos de confiança e curvas de decisão.

Por fim, a harmonização de imagens entre diferentes centros tem mostrado reduzir discrepâncias de distribuição de *features* e ampliar a portabilidade de modelos, passo crucial para futuras aplicações clínicas(33).

### **1.2.3. Aplicações clínicas no câncer colorretal**

#### **1.2.3.1. Estadiamento locorregional**

A radiômica tem sido aplicada para aprimorar o estadiamento locorregional do câncer colorretal, especialmente na diferenciação dos estágios T e N. Estudos utilizando ressonância magnética e tomografia computadorizada demonstraram que modelos baseados em *features* de textura e forma apresentam desempenho moderado a bom para essa finalidade. Dinapoli et al. mostraram que parâmetros texturais derivados da RM do reto podem auxiliar na distinção entre estágios T2 e T3 e na predição do comprometimento linfonodal(34). Resultados semelhantes foram observados por Horvat et al., que correlacionaram padrões texturais e qualitativos da RM com a extensão tumoral e o envolvimento da fáscia mesorretal (39). De modo complementar, Caruso et al. avaliaram a radiômica de TC em câncer de cólon não metastático e identificaram associações entre *features* de forma e heterogeneidade e

maior risco de extensão tumoral local(40).

### **1.2.3.2. Predição de mutações KRAS/NRAS/BRAF**

Diversos estudos demonstraram que assinaturas radiômicas extraídas de TC contrastada permitem prever o status mutacional de RAS e BRAF, com desempenho dependente do protocolo e do modelo. Yang et al. desenvolveram um modelo radiômico capaz de prever simultaneamente KRAS, NRAS e BRAF, alcançando AUC de 0,83 em validação externa(41). Xue et al. propuseram um nomograma clínico-radiômico para KRAS com AUC de 0,92, superior a modelos puramente radiômicos(42).

Negreros-Osuna et al. mostraram que parâmetros simples de textura, como heterogeneidade e desvio-padrão de intensidade, discriminam tumores BRAF-mutados e predizem prognóstico em 5 anos (43). Porto-Álvarez et al. integraram features radiômicas e descritores de imagem (espessura de parede, morfologia, realce) para prever KRAS/NRAS/BRAF em metástases hepáticas, obtendo melhor desempenho que modelos puramente quantitativos (44).

### **1.2.3.3. Aplicações em resposta terapêutica e prognóstico**

Em tumores retais localmente avançados tratados com quimiorradioterapia neoadjuvante, modelos de radiômica baseados em RM demonstraram alta eficácia na previsão de resposta patológica completa, alcançando uma AUC de 0,91 (45). *Features* de textura associadas à maior homogeneidade tumoral, como menor entropia na matriz de coocorrência de níveis de cinza (*GLCM-Entropy*), foram associadas a melhor prognóstico e maior sobrevida livre de progressão em pacientes com câncer colorretal, destacando o potencial da radiômica como ferramenta preditiva complementar para o manejo clínico personalizado.(46)

### **1.2.4. Radiogenômica e integração molecular**

A integração entre radiômica e genômica deu origem à radiogenômica, que busca correlacionar padrões de imagem com perfis genéticos e moleculares. No CCR, essa abordagem visa prever assinaturas mutacionais e marcadores de agressividade de forma não invasiva (47).

Estudos correlacionando features radiômicas com mutações KRAS, NRAS,

BRAF, TP53 e MSI-H sugerem que textura e forma refletem características biológicas do tumor(41,43,44). A análise de textura em tomografia computadorizada demonstrou potencial para prever o status mutacional de KRAS em tumores colorretais, alcançando AUC de até 0,82 em modelos baseados em *support vector machine*, evidenciando que características quantitativas de heterogeneidade tumoral podem refletir o perfil genético subjacente(48).

Modelos multimodais que integram características radiômicas derivadas de tomografia com dados genômicos e clínicos demonstraram desempenho superior na predição de prognóstico em pacientes com câncer colorretal, ressaltando o valor da abordagem integrativa para capturar a heterogeneidade tumoral e aprimorar a estratificação individual de risco (49).

#### **1.2.5. Modelos baseados em inteligência artificial**

A aplicação de inteligência artificial (IA) e aprendizado de máquina impulsionou a radiômica, permitindo lidar com grandes volumes de dados e identificar padrões complexos. Algoritmos supervisionados como *Random Forest*, *Gradient Boosting*, *SVM*, *Naive Bayes* e redes neurais profundas são amplamente utilizados(50–53).

Modelos de *deep learning*, especialmente redes neurais convolucionais (CNNs), são capazes de extrair automaticamente *features* hierárquicas. Em tomografias computadorizadas de metástases hepáticas de câncer colorretal, CNNs demonstraram capacidade de distinguir mutações KRAS e BRAF, com AUC superior a 0,85 (54).

O uso de técnicas de balanceamento de classes, como *SMOTE*, melhora desempenho em coortes desbalanceadas (55). Estratégias de explicabilidade — *SHAP values* e *feature importance ranking* — permitem interpretar decisões algorítmicas e identificar *features* relevantes.(56)

Apesar dos avanços, o uso clínico de IA ainda enfrenta barreiras regulatórias e reprodutibilidade limitada.(29,52)

#### **1.2.6. Limitações e perspectivas futuras**

As principais limitações atuais incluem heterogeneidade técnica entre centros, ausência de padronização universal de aquisição, variabilidade de segmentação e falta de validação prospectiva(38). Muitos modelos são derivados de coortes

retrospectivas pequenas. Iniciativas como IBSI e RQS promovem transparência metodológica e relatórios padronizados (30).

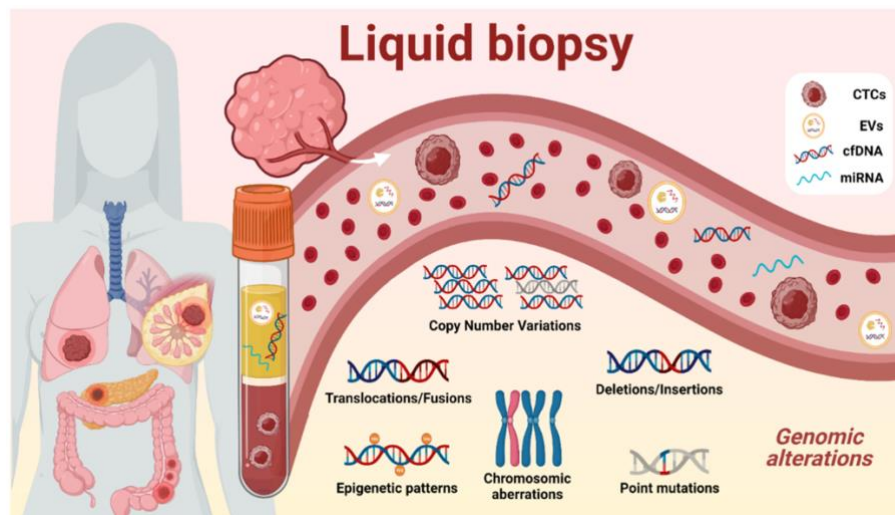
O futuro da radiômica aponta para a integração multidimensional de dados — combinando imagens, biomarcadores moleculares e dados clínicos — em modelos preditivos multimodais. Essa convergência se alinha à medicina de precisão e tem potencial para guiar decisões terapêuticas e monitorar resposta de forma não invasiva (57,58).

### **1.3. Parte 3: Biópsia líquida no câncer colorretal**

#### **1.3.1. Fundamentos e princípios biológicos**

A biópsia líquida representa uma das ferramentas mais promissoras da oncologia de precisão moderna, permitindo a caracterização molecular de tumores de forma minimamente invasiva. Diferentemente das biópsias teciduais, que captam apenas um fragmento estático da neoplasia, ela oferece uma visão dinâmica e abrangente da heterogeneidade tumoral e das alterações moleculares em tempo real(59,60).

O princípio baseia-se na detecção de componentes tumorais circulantes — como DNA tumoral circulante (ctDNA), células tumorais circulantes (CTCs), RNA livre, exossomos e plaquetas educadas por tumor (TEPs) — que são liberados na corrente sanguínea, urina, saliva ou outros fluidos biológicos (61). Esses elementos refletem as alterações genéticas e epigenéticas do tumor primário e de suas metástases, permitindo avaliar carga tumoral, resposta terapêutica e presença de doença residual mínima (MRD).



**Figura 5: Utilidade clínica da biópsia líquida.**  
**Fonte: Biomedicines, 2021 (62).**

Estudos recentes demonstram que a biópsia líquida tem papel crescente em rastreamento, diagnóstico precoce, prognóstico e monitoramento terapêutico no câncer colorretal (CCR), complementando e, em alguns contextos, superando as limitações da tomografia e da colonoscopia (63).

### 1.3.2. DNA livre circulante (cfDNA) e DNA tumoral circulante (ctDNA)

O DNA livre circulante (cfDNA) é composto por fragmentos de DNA extracelular liberados por células normais e neoplásicas através de apoptose, necrose ou secreção ativa. Sua descoberta precedeu a do ctDNA, sendo detectado em pequenas quantidades no plasma de indivíduos saudáveis. No contexto oncológico, uma fração específica desse material deriva das células tumorais, recebendo o nome de ctDNA (64,65).

O ctDNA carrega mutações somáticas idênticas às do tumor primário, incluindo alterações em KRAS, NRAS, BRAF, TP53 e genes de reparo de DNA. Sua proporção dentro do cfDNA total — chamada de fração tumoral — varia conforme o estágio da doença, carga tumoral e resposta ao tratamento. Em média, essa fração é inferior a 1% em tumores iniciais, mas pode ultrapassar 20% em neoplasias metastáticas (66).

A meia-vida curta do ctDNA (cerca de 1 a 2 horas) permite que sua detecção reflita rapidamente alterações na atividade tumoral e na eficácia terapêutica. Além disso, sua distribuição fragmentada, com tamanhos médios de 150–200 pares de bases, fornece pistas sobre a origem celular e o mecanismo de liberação (65,67).

A distinção entre cfDNA e ctDNA é essencial: enquanto o primeiro é onipresente e inespecífico, o ctDNA carrega assinaturas genômicas exclusivas do tumor, incluindo mutações pontuais, inserções/deleções, metilações aberrantes e instabilidade de microssatélites. Essa especificidade confere à biópsia líquida valor prognóstico e preditivo singular. (66–68)

### **1.3.3. Aplicações clínicas no câncer colorretal**

#### **1.3.3.1. Diagnóstico e rastreamento**

O uso do ctDNA e de outros biomarcadores circulantes na triagem de CCR tem crescido rapidamente. Testes comerciais de triagem para câncer colorretal, como o Cologuard (mt-sDNA) e o Epi proColon (methylated SEPT9 no sangue), demonstraram sensibilidades de aproximadamente 90–92 % para detecção de câncer colorretal, com especificidades próximas a 87–91%, superando o teste de sangue oculto e contribuindo para o rastreamento não invasivo (69–71).

Painéis baseados em mutações somáticas (KRAS, BRAF, PIK3CA) e metilação de SEPT9 e NDRG4 são atualmente os mais avaliados para rastreamento não invasivo (72).

#### **1.3.3.2. Prognóstico e estratificação de risco**

A detecção pós-operatória de ctDNA é um indicador forte de risco de recidiva precoce e de menor sobrevida em câncer colorretal, com estudos mostrando risco aumentado em ordem de 10-vezes para pacientes ctDNA positivos em comparação aos negativos.(73).

#### **1.3.3.3. Doença residual mínima (MRD)**

A detecção de MRD por ctDNA tem se tornado uma das aplicações mais revolucionárias da biópsia líquida. A ausência de ctDNA após cirurgia ou quimioterapia adjuvante correlaciona-se com excelente prognóstico, enquanto sua persistência indica presença de micrometástases e risco elevado de progressão (66,68)

Estudos prospectivos demonstram que a redução ou *clearance* do ctDNA durante o tratamento pode anteceder a resposta radiológica em vários meses (por exemplo, ≈ 3 a 6 meses), abrindo caminho para ajustes terapêuticos precoces (74).

#### 1.3.3.4. Predição de resposta e resistência terapêutica

A quantificação seriada de ctDNA durante o tratamento reflete a dinâmica tumoral em tempo real. A emergência de clones mutantes em KRAS, NRAS ou no domínio extracelular de EGFR pode sinalizar resistência secundária ao bloqueio de EGFR (como com cetuximabe ou panitumumabe) antes da progressão clínica evidente (75). Essa abordagem possibilita redirecionar terapias antes da progressão clínica evidente.

#### 1.3.4. Técnicas e desafios metodológicos

A análise da biópsia líquida depende de métodos altamente sensíveis para detecção de fragmentos raros de ctDNA.

As principais técnicas incluem:

- *Droplet Digital PCR* (ddPCR): quantificação absoluta de mutações específicas.
- BEAMing (*Beads, Emulsion, Amplification, Magnetics*): método híbrido PCR-digital para detecção de mutações conhecidas com alta precisão (77).
- NGS de alto desempenho: permite rastrear múltiplas mutações simultaneamente, detectando variantes com frequência alélica  $<0,1\%$  (78).
- Painéis de metilação: exploram assinaturas epigenéticas específicas de tumores colorretais, com destaque para genes SEPT9, SDC2 e NDRG4(79,80).

A escolha do método depende do objetivo clínico — ddPCR e BEAMing são ideais para o monitoramento de mutações específicas e detecção de doença residual mínima (MRD), oferecendo alta sensibilidade e precisão em alvos conhecidos (81).

Em contrapartida, NGS e painéis de metilação proporcionam cobertura mais ampla, permitindo o rastreamento simultâneo de múltiplas mutações ou assinaturas epigenéticas, sendo mais indicados para detecção precoce, estratificação molecular e caracterização abrangente do perfil tumoral (82).

Os principais desafios metodológicos (83,84) incluem:

- Variabilidade pré-analítica (tipo de tubo, tempo de centrifugação e armazenamento);
- Interferência de cfDNA não tumoral;

- Limites de detecção em tumores iniciais ou após cirurgia curativa;
- Necessidade de padronização entre laboratórios e validação prospectiva.

### **1.3.5. Perspectivas futuras e integração com medicina de precisão**

A biópsia líquida está no centro da transição para a oncologia personalizada, integrando genômica, transcriptômica e imagem funcional. A tendência atual é o desenvolvimento de modelos multimodais, que combinam dados de ctDNA, radiômica e variáveis clínicas para aprimorar a predição de resposta e prognóstico(85,86).

A aplicação de algoritmos de inteligência artificial em dados genômicos de cfDNA tem permitido identificar padrões complexos de fragmentação e metilação característicos de diferentes tipos tumorais, alcançando acurácia superior a 90% na determinação do tecido de origem (87).

Embora desafios de custo, sensibilidade e regulamentação ainda limitem sua adoção rotineira, há perspectiva crescente de que a biópsia líquida seja incorporada aos algoritmos clínicos do CCR nos próximos anos, tanto no contexto diagnóstico quanto no monitoramento de MRD e terapia-alvo.

Nessa conjuntura da medicina de precisão, este projeto propõe a correlação entre assinaturas radiômicas e um painel genético abrangente de mutações em uma coorte de pacientes com câncer colorretal tratados na rede pública (SUS-HUB), integrando ainda achados moleculares obtidos por biópsia líquida e análise tecidual.

## **2. OBJETIVOS**

### **2.1. Objetivo geral**

Investigar as principais alterações presentes em imagens de tomografias computadorizadas, extrair as suas características radiômicas e comparar com o perfil genético tumoral e marcadores de biópsia líquida.

#### **2.1.1. Objetivos específicos**

- Correlacionar as características radiômicas extraídas das imagens com os dados de estadiamento anatomopatológico para investigar a capacidade da radiômica em identificar padrões associados à agressividade tumoral;
- Analisar os dados do sequenciamento do DNA livre circulante (cfDNA) obtidos por painel gênico para biópsia líquida;
- Quantificar a frequência alélica das mutações (carga mutacional) dos tumores e do cfDNA a fim de se comparar o perfil molecular do câncer de cada indivíduo do estudo;
- Comparar as informações radiológicas com a carga mutacional;
- Estabelecer potenciais biomarcadores de imagem não invasivos para auxiliar no manejo clínico do CCR.

### 3. METODOLOGIA

#### 3.1. Desenho do estudo

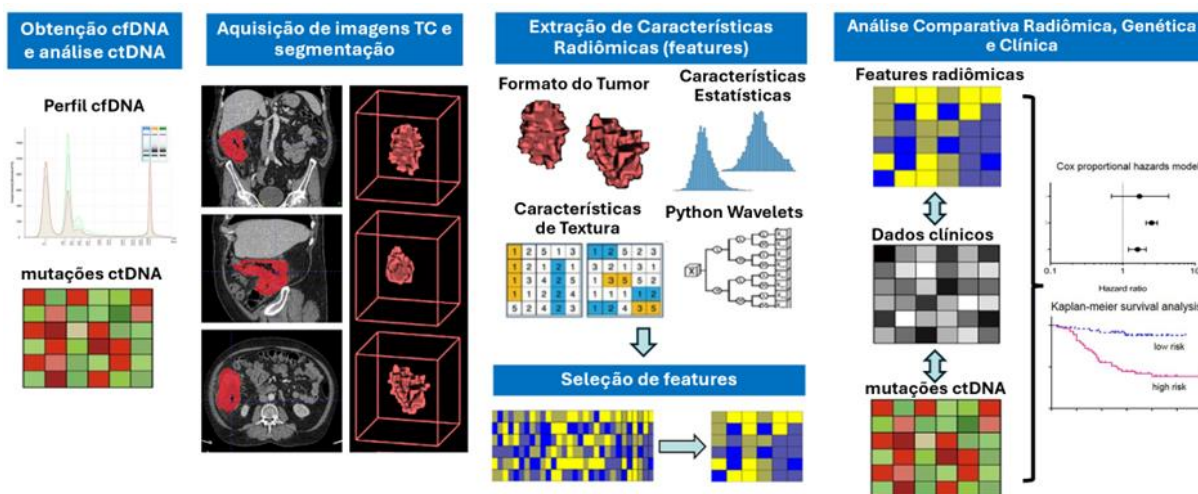


Figura 6. Desenho do Estudo

#### 3.2. Aspectos éticos

O presente estudo é parte de projeto aprovado no Comitê de Ética em Pesquisa (CEP) da Faculdade de Ciências da Saúde da Universidade de Brasília sob CAAE número 65822222.4.0000.0030. Os pacientes foram convidados a participar do estudo assinando termo de consentimento livre esclarecido (TCLE), aprovado no CEP. Apenas os pacientes que aceitaram participar foram incluídos no estudo.

#### 3.3. População do estudo

Trata-se de um estudo populacional observacional prospectivo no qual foi feita a análise radiológica e molecular de 53 pacientes, todos portadores de câncer colorretal, que não foram submetidos a tratamento prévio. Os dados radiológicos foram coletados a partir da aquisição de imagens de tomografia computadorizada. As imagens passaram por análise subjetiva para confecção do laudo e posteriormente foram avaliadas, segmentadas e processadas por softwares específicos.

Os dados moleculares foram gerados pela quantificação do DNA livre circulante (cfDNA) e sequenciamento do DNA tumoral (ctDNA) dos pacientes do estudo.

#### Critérios de inclusão:

- Diagnóstico confirmado de adenocarcinoma;
- Ausência de tratamento oncológico prévio (cirurgia, quimioterapia ou radioterapia);
- Disponibilidade de exames de tomografia computadorizada de abdome completo (trifásico) e de boa qualidade;
- Disponibilidade de material biológico adequado;
- Assinatura do Termo de Consentimento Livre e Esclarecido (TCLE).

#### **Crítérios de exclusão:**

- Realização prévia de qualquer tratamento oncológico antes da coleta de dados e imagens;
- Imagens radiológicas incompletas (sem contraste) ou realizadas em serviços externos ao HUB;
- Insuficiência de material biológico para as análises genéticas propostas;
- Tratamento cirúrgico em hospital externo ao HUB (anatomopatológico indisponível para correlação);
- Recusa em participar do estudo ou não assinatura do TCLE.

### **3.4. Dados do estadiamento histopatológico**

Para correlação com os dados radiológicos foram obtidos os dados do estadiamento TNM histopatológico (15) dos pacientes submetidos à intervenção cirúrgica, emitidos pelo departamento de patologia do HUB/UnB. Desse modo, foram divididos em grupos (labels) para correlação, como por exemplo grupo A de pacientes com classificação “T” I e II e grupo B “T” III e IV ou grupo C com “N” positivo e grupo D “N” negativo.

**Tabela 1: Grupos de pacientes separados por classificação T.**

ESTADIAMENTO T	Grupo A: I e II	Grupo B: III e IV
Número de pacientes	11	42 (44 com 2 sincrônicos)

ESTADIAMENTO N	Grupo A: negativo	Grupo B: positivo
Número de pacientes	20	33 (35 com 2 sincrônicos)

Fonte: Autoria própria

### 3.5. Aquisição das imagens radiológicas e análise subjetiva

As imagens tomográficas foram adquiridas nos tomógrafos com multidetectores (Toshiba Aquilion 64 canais e Philips MX EVO2 16 canais) do Hospital Universitário de Brasília (HUB/UnB). Foram incluídas imagens do tórax e abdome total para estadiamento inicial dos pacientes, por meio de protocolo trifásico, com fase sem contraste seguida por administração intravenosa de meio de contraste iodado por bomba injetora automática, sendo posteriormente feitas as fases arterial e portal venosa.

As imagens foram acessadas inicialmente a partir do sistema de arquivamento e comunicação de imagens (PACS) no formato DICOM (Comunicação de Imagens Digitais em Medicina) e laudadas pelo RIS (Sistema de Informação de Radiologia) das estações de trabalho (*workstations*) da Unidade de Diagnóstico por Imagem (UDI) do HUB, disponibilizando, então, o laudo radiológico com estadiamento inicial dos pacientes, que evidenciava uma avaliação inicial da extensão do tumor, sua relação anatômica com as estruturas adjacentes e se haviam metástases regionais ou à distância, possibilitando o estadiamento inicial radiológico.

Cada imagem tomográfica é composta por *voxels* (*volume elements*), que representam unidades tridimensionais de informação com altura, largura e profundidade, sendo o análogo tridimensional dos píxeis bidimensionais. O tamanho de *voxel* é definido pelas dimensões do *field of view* (FOV), matriz de aquisição e espessura de corte, e sua unidade de medida é o milímetro cúbico (mm<sup>3</sup>) (88). A partir da segmentação da lesão e contagem dos *voxels* incluídos na região de interesse (ROI), é possível estimar de forma precisa o volume tumoral, parâmetro importante para caracterização da agressividade tumoral e que compõe as *features* morfológicas da análise radiômica.

### 3.6. Segmentação tumoral das imagens

Num segundo momento, as fases portais dos exames, anonimizadas, foram processadas pelo *software ITK-snap 4.0* para segmentação da região tumoral, sem conhecimento prévio do estadiamento histopatológico, evidenciando o volume tumoral e gerando arquivos no formato *Nearly Raw Raster Data (NRRD)* para serem posteriormente processados pela biblioteca *PyRadiomics*.

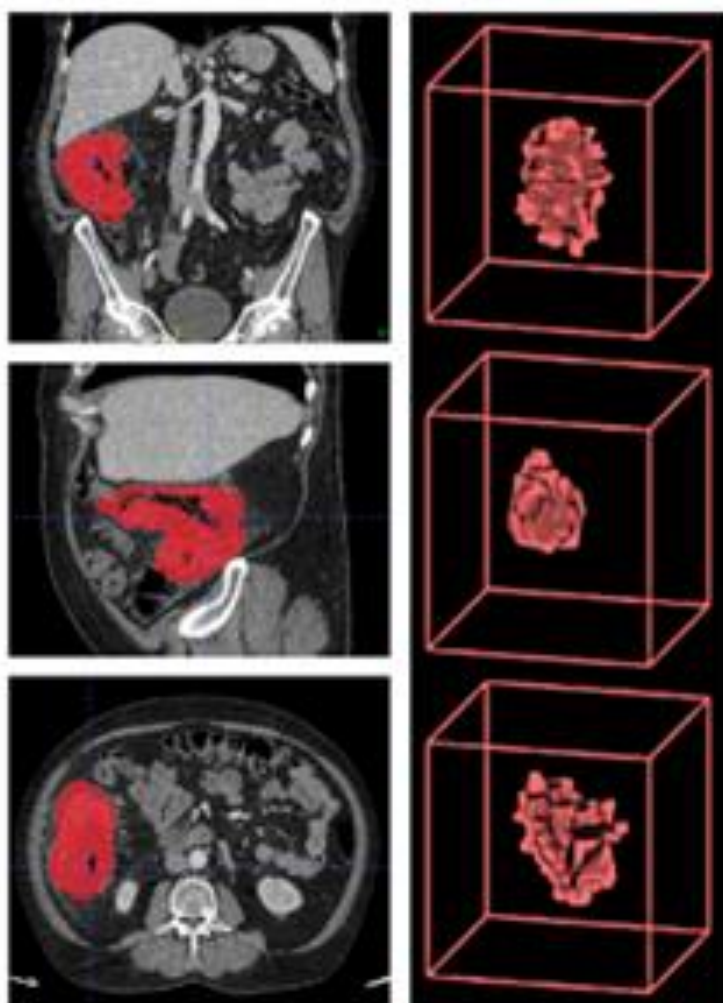


Figura 7. Segmentação da Região Tumoral

### 3.7. Pré-processamento

#### 3.7.1. Extração das características radiômicas

Os arquivos NRRD foram processados pela biblioteca *PyRadiomics*, uma plataforma de código aberto, que permite extrair um grande número de características

radiômicas (*features*) de imagens médicas, as quais são analisadas por algoritmos de aprendizado de máquina e correlacionadas com os dados clínicos elegíveis.

Extraíu-se 107 *features* radiômicas, sem filtros ativados, relacionadas a três classes: Não-texturais tradicionais baseadas em forma, estatísticas de primeira-ordem e "texturais de alta ordem" ("*higher-order texture features*"). Essas *features* podem evidenciar forma e volume tumoral, características específicas dos *voxels* da região delimitada, interrelação de pixels e heterogeneidade da textura tumoral.

Das *features* não-texturais tradicionais baseadas em forma, foram geradas 14 *features* (tabela 2), que incluem descritores do tamanho tridimensional e da forma da região de interesse, ou seja, características do formato do tumor, como esfericidade e volume (89).

**Tabela 2- *Features* não-texturais tradicionais extraídas de 41 pacientes.**

<i>Features</i> não-texturais tradicionais baseadas em forma
<i>original_shape_Elongation</i>
<i>original_shape_Flatness</i>
<i>original_shape_LeastAxisLength</i>
<i>original_shape_MajorAxisLength</i>
<i>original_shape_Maximum2DDiameterColumn</i>
<i>original_shape_Maximum2DDiameterRow</i>
<i>original_shape_Maximum2DDiameterSlice</i>
<i>original_shape_Maximum3DDiameter</i>
<i>original_shape_MeshVolume</i>
<i>original_shape_MinorAxisLength</i>
<i>original_shape_Sphericity</i>
<i>original_shape_SurfaceArea</i>
<i>original_shape_SurfaceVolumeRatio</i>
<i>original_shape_VoxelVolume</i>

*Fonte: Zwaneburg et al., 2016*

Das *features* estatísticas de primeira-ordem, foram obtidas 18 (tabela 3). Elas descrevem a distribuição das intensidades de *voxel* dentro da região da imagem definida por meio de métricas comumente usadas e básicas, como por exemplo média, mediana e variância(89).

**Tabela 3- *Features* não-texturais de primeira-ordem extraídas de 41 pacientes.**

*Features* estatísticas de primeira-ordem

<i>original_firstorder_10Percentile</i>
<i>original_firstorder_90Percentile</i>
<i>original_firstorder_Energy</i>
<i>original_firstorder_Entropy</i>
<i>original_firstorder_InterquartileRange</i>
<i>original_firstorder_Kurtosis</i>
<i>original_firstorder_Maximum</i>
<i>original_firstorder_MeanAbsoluteDeviation</i>
<i>original_firstorder_Mean</i>
<i>original_firstorder_Median</i>
<i>original_firstorder_Minimum</i>
<i>original_firstorder_Range</i>
<i>original_firstorder_RobustMeanAbsoluteDeviation</i>
<i>original_firstorder_RootMeanSquared</i>
<i>original_firstorder_Skewness</i>
<i>original_firstorder_TotalEnergy</i>
<i>original_firstorder_Uniformity</i>
<i>original_firstorder_Variance</i>

Fonte: Zwaneburg et al., 2016

Além disso, foram obtidas 75 *features* de *higher-order*, que consistem em cinco grupos (tabela 4). Do grupo *Gray Level Co-occurrence Matrix (GLCM)*, foram extraídas 24 *features*, que são resultados de uma matriz da região de interesse segmentada. Do grupo *Gray Level Size Zone Matrix (GLSZM)* foram obtidas 16 *features*, que quantificam as zonas de nível de cinza em uma imagem. Ainda, 16 *features* foram extraídas do grupo *Gray Level Run Length Matrix (GLRLM)*, que quantifica execuções de nível cinza, definidas como o comprimento em número de pixels consecutivos que têm o mesmo valor de nível de cinza. Além disso, foram obtidas 5 *features* do *Neighbouring Gray Tone Difference Matrix (NGTDM)*, que quantifica a diferença entre um valor de cinza e o valor médio de cinza de seus vizinhos dentro da distância. Por fim, 14 *features* do *Gray Level Dependence Matrix (GLDM)*, que quantifica as dependências de nível de cinza em uma imagem (89).

Tabela 4 - *Higher-order texture features* extraídas de 41 pacientes.

<i>GLCM (Gray-level co-occurrence matrix)</i>	<i>GLDM (Gray Level Dependence Matrix)</i>	<i>GLRLM (Gray Level Run Length Matrix)</i>	<i>GLSZM (Gray Level Size Zone)</i>	<i>NGTDM (Neighbouring Gray Tone)</i>
---	--	---	-------------------------------------	---------------------------------------

				Difference Matrix)
original_glc m_Autocorr elation	original_gldm_Depe ndenceEntropy	original_glrlm_Gra yLevelNonUniform ity	original_glszm_Gr ayLevelNonUnifor mity	original_n gtdm_Bus yness
original_glc m_ClusterPr ominence	original_gldm_Depe ndenceNonUniformit y	original_glrlm_Gra yLevelNonUniform ityNormalized	original_glszm_Gr ayLevelNonUnifor mityNormalized	original_n gtdm_Co arseness
original_glc m_ClusterS hade	original_gldm_Depe ndenceNonUniformit yNormalized	original_glrlm_Gra yLevelVariance	original_glszm_Gr ayLevelVariance	original_n gtdm_Co mplexity
original_glc m_ClusterT endency	original_gldm_Depe ndenceVariance	original_glrlm_Hig hGrayLevelRunE mphasis	original_glszm_Hi ghGrayLevelZone Emphasis	original_n gtdm_Co ntrast
original_glc m_Contrast	original_gldm_GrayL evelNonUniformity	original_glrlm_Lon gRunEmphasis	original_glszm_La rgeAreaEmphasis	original_n gtdm_Str ength
original_glc m_Correlati on	original_gldm_GrayL evelVariance	original_glrlm_Lon gRunHighGrayLev elEmphasis	original_glszm_La rgeAreaHighGray LevelEmphasis	
original_glc m_Differenc eAverage	original_gldm_HighG rayLevelEmphasis	original_glrlm_Lon gRunLowGrayLev elEmphasis	original_glszm_La rgeAreaLowGrayL evelEmphasis	
original_glc m_Differenc eEntropy	original_gldm_Large DependenceEmphas is	original_glrlm_Lo wGrayLevelRunE mphasis	original_glszm_Lo wGrayLevelZoneE mphasis	
original_glc m_Differenc eVariance	original_gldm_Large DependenceHighGr ayLevelEmphasis	original_glrlm_Ru nEntropy	original_glszm_Si zeZoneNonUnifor mity	
original_glc m_Id	original_gldm_Large DependenceLowGra yLevelEmphasis	original_glrlm_Ru nLengthNonUnifor mity	original_glszm_Si zeZoneNonUnifor mityNormalized	
original_glc m_Idm	original_gldm_LowG rayLevelEmphasis	original_glrlm_Ru nLengthNonUnifor mityNormalized	original_glszm_S mallAreaEmphasi s	
original_glc m_Idmn	original_gldm_Small DependenceEmphas is	original_glrlm_Ru nPercentage	original_glszm_S mallAreaHighGray LevelEmphasis	
original_glc m_Idn	original_gldm_Small DependenceHighGr ayLevelEmphasis	original_glrlm_Ru nVariance	original_glszm_S mallAreaLowGray LevelEmphasis	
original_glc m_lmc1	original_gldm_Small DependenceLowGra yLevelEmphasis	original_glrlm_Sh ortRunEmphasis	original_glszm_Zo neEntropy	
original_glc m_lmc2		original_glrlm_Sh ortRunHighGrayL evelEmphasis	original_glszm_Zo nePercentage	

<i>original_glc m_InverseV ariance</i>		<i>original_glrIm_Sh ortRunLowGrayLe velEmphasis</i>	<i>original_glszm_Zo neVariance</i>	
<i>original_glc m_JointAver age</i>				
<i>original_glc m_JointEne rgy</i>				
<i>original_glc m_JointEntr opy</i>				
<i>original_glc m_MCC</i>				
<i>original_glc m_Maximu mProbability</i>				
<i>original_glc m_SumAver age</i>				
<i>original_glc m_SumEntr opy</i>				
<i>original_glc m_SumSqu ares</i>				

*Fonte: Zwaneburg et al., 2016*

A partir das diferentes classes de características radiômicas, foram desmembrados 15 datasets para classificação, que consistem nos grupos “shape”, “firstorder”, “glcm”, “gldm”, “glrlm”, “glszm”, “ngtdm”, “textura combinado”, shape *features* combinadas com *first-order* ou *texture*; e o conjunto completo de *features*.

### **3.7.2. Normalização das características**

Diferentes características radiômicas vêm em diferentes ordens de magnitude. A normalização dos dados é uma etapa essencial no pré-processamento de características radiômicas, uma vez que essas variáveis podem apresentar escalas muito distintas. Essa heterogeneidade pode comprometer o desempenho de algoritmos de aprendizado de máquina, especialmente aqueles sensíveis à magnitude dos valores das variáveis. Estudos comparativos demonstraram que diferentes métodos de escalonamento — incluindo normalização e padronização — influenciam significativamente a performance e a estabilidade dos modelos preditivos (90). A

padronização, também conhecida como *z-score normalization*, foi utilizada neste estudo. Esse procedimento assegura que todas as características contribuam de forma equilibrada para o treinamento dos modelos preditivos, prevenindo que atributos de maior magnitude dominem o processo de aprendizado.

O método de padronização foi aplicado a cada conjunto de recursos, de modo que cada um deles tenha média zero e desvio padrão unitário (91,92). Os recursos normalizados são calculados como  $z = (x - \mu) / \sigma$ , onde  $x$  são os recursos não normalizados e  $\mu$  e  $\sigma$  são a média e o desvio padrão do conjunto de recursos correspondente, respectivamente (93).

### 3.7.3. Redução da dimensionalidade dos dados

A Análise de Componentes Principais (PCA) foi utilizada para a redução de dimensionalidade, que consiste em projetar os dados de características de um espaço de alta dimensionalidade para um subespaço de menor dimensionalidade, mantendo a maior variabilidade possível dos dados. PCA é um método estatístico que identifica as direções (componentes principais) ao longo das quais os dados possuem a maior variância. O objetivo é transformar as variáveis originais em um novo conjunto de variáveis não correlacionadas, chamadas de componentes principais, ordenados pela importância de cada um na explicação da variância dos dados. (94)

Ao selecionar apenas os primeiros componentes com maior variância explicada, o PCA permite reduzir ruído, mitigar redundâncias e melhorar a eficiência computacional dos modelos. Além disso, a técnica contribui para evitar overfitting, especialmente em cenários com alta dimensionalidade e número reduzido de amostras (95).

### 3.7.4. Aumento de dados

O aumento de dados (*data augmentation*) utilizando a técnica SMOTE (*Synthetic Minority Over-sampling Technique*) foi aplicado para balancear a distribuição das classes e mitigar o viés relacionado ao desbalanceamento amostral. Essa abordagem cria exemplos sintéticos da classe minoritária por meio de combinações lineares entre os vizinhos mais próximos, o que contribui para melhorar a generalização do modelo e reduzir o sobreajuste. Em estudos radiômicos, essa estratégia é particularmente útil diante de amostras limitadas e classes desiguais —

situação frequente em coortes clínicas — favorecendo o desempenho de classificadores supervisionados (55,96).

### **3.8. Implementação e classificação**

A assinatura radiômica consiste em um conjunto de características quantitativas extraídas de imagens médicas, que serão utilizadas para identificar padrões associados ao comportamento biológico do tumor (97). Tais assinaturas permitem representar aspectos da morfologia, textura e intensidade tumoral de forma objetiva, sendo empregadas em modelos de predição e estratificação de risco.

A utilização de múltiplos algoritmos de aprendizado supervisionado permite avaliar a robustez da assinatura radiômica diante de diferentes paradigmas de modelagem, como métodos probabilísticos, baseados em distância, árvores de decisão e redes neurais. Essa abordagem comparativa favorece a seleção dos modelos com melhor desempenho geral e aplicabilidade clínica, considerando a variabilidade dos dados em cenários reais.

Os experimentos de aprendizado de máquina foram realizados utilizando o scikit-learn (versão 1.5.1, <https://scikit-learn.org>). Cada experimento foi executado 100 vezes, e foram registrados os valores médios dos resultados, bem como os intervalos de confiança de 95% (IC).

Em cada iteração de cada experimento, os 55 tumores segmentados foram divididos aleatoriamente em um conjunto de treinamento (75% dos casos) e um conjunto de validação (25% dos casos). Nessa etapa, foi utilizado o validador cruzado estratificado *ShuffleSplit* por classe, a fim de preservar a proporção de amostras de cada classe.

A proposta para a classificação dos grupos utilizou diferentes algoritmos de aprendizado de máquina para identificar padrões discriminativos entre as características radiômicas. O algoritmo K-Nearest Neighbors (KNN) atribui um rótulo a uma nova amostra com base na classe predominante entre seus k vizinhos mais próximos, determinados por uma medida de distância (53).

Além do KNN, empregou-se a Regressão Logística (*Logistic Regression*, LR), um modelo estatístico amplamente utilizado por estimar a probabilidade de uma amostra pertencer a uma classe específica por meio da função sigmoide, sendo

particularmente útil para problemas binários (98). O classificador Naive Bayes (NB) aplica o Teorema de Bayes sob a suposição de independência entre os atributos, sendo amplamente utilizado para problemas de classificação probabilística (99).

Também foram utilizadas Árvores de Decisão (*Decision Tree*, DT), que realizam divisões sucessivas nos dados com base em critérios de entropia ou ganho de informação (100). O classificador *Extra Trees* (ET) é uma variação das árvores de decisão que introduz maior aleatoriedade na escolha dos pontos de divisão, contribuindo para uma melhor generalização do modelo (101). O *Random Forest* (RF) é um conjunto de múltiplas árvores de decisão treinadas em subconjuntos aleatórios dos dados, combinando suas previsões para reduzir o risco de *overfitting*(50).

Além disso, empregou-se a *Multi-Layer Perceptron* (MLP), uma rede neural artificial composta por múltiplas camadas de neurônios interconectados, capaz de modelar relações complexas nos dados (102). O *AdaBoost* (AB) é um método de *boosting* que combina classificadores fracos para formar um modelo robusto (103), enquanto o *Gradient Boosting* (GB) constrói árvores de decisão sequenciais, minimizando o erro residual de modelos anteriores (51).

Aplicou-se, ainda, o *Linear Discriminant Analysis* (LDA), que busca encontrar uma combinação linear de variáveis independentes que melhor separa as classes (104). Por fim, o *XGBoost* (XGB) é uma poderosa técnica baseada em *gradient boosting*, otimizada para desempenho e eficiência computacional, sendo amplamente utilizada em competições de aprendizado de máquina(105).

Esses algoritmos englobam tanto classificadores lineares (por exemplo, a regressão logística, que assume separação linear entre as classes) quanto classificadores de conjunto (ensemble) mais complexos (como random forest e gradient boosting), que combinam múltiplos modelos de baixo desempenho individual para capturar relações não lineares e de ordem superior (106).

Essa diversidade de abordagens permitiu comparar modelos mais simples e interpretáveis com métodos baseados em conjuntos mais poderosos, frequentemente mais adequados para dados de alta dimensionalidade.

### **3.9. Avaliação da performance da assinatura**

A avaliação da performance dos modelos de classificação foi realizada com base em métricas estatísticas consolidadas e recomendadas para estudos com dados desbalanceados (103–105). O desempenho dos modelos foi avaliado utilizando sensibilidade (Se), especificidade (Spe), acurácia balanceada (BA), área sob a curva ROC (AUC), precisão (Pre), valor preditivo negativo (NPV), F1-score e acurácia (Acc).

Tais métricas oferecem diferentes perspectivas sobre a performance do modelo: a acurácia reflete a proporção global de classificações corretas, enquanto sensibilidade e especificidade são cruciais em contextos clínicos, nos quais falsos negativos ou positivos têm impactos distintos (110,111). O F1-score, sendo a média harmônica entre precisão e sensibilidade, é particularmente útil em bases de dados desbalanceadas (112). A acurácia balanceada representa a média entre sensibilidade e especificidade, corrigindo distorções provocadas por desequilíbrio entre as classes (113). Já a área sob a curva ROC (AUC) quantifica a capacidade global do modelo em distinguir corretamente entre as classes, independentemente do ponto de corte escolhido (114).

Cada uma dessas métricas é calculada da seguinte forma:

**AUC (Area Under the Curve):** Mede a capacidade do modelo em discriminar entre classes, correspondendo à área sob a curva ROC. Valores próximos de 1 indicam excelente desempenho, enquanto 0,5 representa desempenho aleatório.

$$AUC = \int_0^1 Se(1 - Spe) d(1 - Spe)$$

**Acurácia (Acc):** Mede a proporção de previsões corretas em relação ao total de amostras, sendo calculada como:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

**Sensibilidade (Se) ou Recall:** Avalia a capacidade do modelo de identificar corretamente os casos positivos, definida como:

$$Se = \frac{TP}{TP + FN}$$

**Especificidade (Spe):** Mede a capacidade do modelo de identificar corretamente os casos negativos, calculada por:

$$Spe = \frac{TN}{TN + FP}$$

**Precisão (Pre):** Indica a proporção de previsões positivas que realmente pertencem à classe positiva, dada por:

$$Pre = \frac{TP}{TP + FP}$$

**F1-Score (F1):** É a média harmônica entre a precisão e a sensibilidade, balanceando ambas as métricas:

$$F1 = 2 \times \frac{Pre \times Se}{Pre + Se}$$

**Acurácia Balanceada (BCA):** Considera o equilíbrio entre sensibilidade e especificidade, útil para conjuntos de dados desbalanceados. Sua fórmula é:

$$BCA = \frac{Se + Spe}{2}$$

Onde TP, FP, TN e FN correspondem ao número de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos, respectivamente. Todas as métricas foram expressas em porcentagens, obtidas multiplicando seus valores por 100%.

A validação foi realizada por meio de repetidas amostragens hold-out com 100 iterações, e os intervalos de confiança (ICs) foram calculados utilizando a distribuição

t de Student, que permite avaliar a estabilidade do desempenho dos classificadores e reduzir a variabilidade introduzida por divisões específicas do conjunto de dados (115).

Os experimentos foram conduzidos com diferentes subconjuntos das *features* extraídas (Tabela AA), incluindo a avaliação de cada conjunto de forma independente, a combinação de *features* de forma com *features* de primeira ordem ou de textura, o conjunto combinado de *features* de textura, e o conjunto completo de *features*.

set \ class	First Order	Shape (3D)	GLCM	GLSZM	GLRLM	NGTDM	GLDM
firstorder18	18	-	-	-	-	-	-
shape14	-	14	-	-	-	-	-
glcm24	-	-	24	-	-	-	-
glrlm16	-	-	-	16	-	-	-
glszm16	-	-	-	-	16	-	-
ngtdm5	-	-	-	-	-	5	-
gldm14	-	-	-	-	-	-	14
shapefirstorder32	18	14	-	-	-	-	-
shapeglcm38	-	14	24	-	-	-	-
shapeglrlm30	-	14	-	16	-	-	-
shapeglszm30	-	14	-	-	16	-	-
shapengtdm19	-	14	-	-	-	5	-
shapegldm28	-	14	-	-	-	-	14
texture75	-	-	24	16	16	5	14
fullset107	18	14	24	16	16	5	14

**Tabela 5. Subconjuntos de features radiômicas utilizados nos experimentos e o número de features radiômicas por classe em cada subconjunto.**

Foram realizadas análises comparativas entre diferentes cenários de pré-processamento, com o objetivo de avaliar o efeito aditivo de cada etapa sobre o desempenho da classificação:

- (i) features sem pré-processamento;
- (ii) features normalizadas;
- (iii) features normalizadas e reduzidas por PCA;
- (iv) features normalizadas com aumento de dados via SMOTE;
- (v) features normalizadas e reduzidas por PCA, seguidas de aumento via SMOTE; e
- (vi) features normalizadas com aumento via SMOTE, seguidas de redução por PCA.

O número de dimensões no subespaço reduzido pelo PCA foi de 3, exceto nos experimentos com o conjunto combinado de features de textura (texture75) e com o

conjunto completo de features (fullset107), nos quais foram utilizadas 5 dimensões.

Essa abordagem permitiu verificar o impacto individual de cada etapa no desempenho global dos modelos, contribuindo para a identificação da melhor configuração para uso clínico da assinatura radiômica desenvolvida.

### **3.10. Análise do perfil molecular pela Biópsia Líquida**

#### **3.10.1. Extração e quantificação do cfDNA do plasma**

A extração do DNA livre circulante (cfDNA) foi realizada no Laboratório de Patologia Molecular do Câncer da Faculdade de Saúde da Universidade de Brasília (Cancerlab), a partir de plasma sanguíneo previamente centrifugado em duas etapas para remoção de células e detritos celulares. Foram utilizados os kits MagMAX™ Cell-Free Total Nucleic Acid Kit e MagMAX™ Cell-Free DNA Isolation Kit (Thermo Fisher Scientific), conforme as recomendações do fabricante. Após a extração, o cfDNA foi quantificado por fluorimetria com o Qubit 3 Fluorometer e o kit Qubit dsDNA High Sensitivity, e avaliado quanto à integridade e tamanho médio por eletroforese automatizada no sistema Agilent TapeStation 2200 com o kit D5000 High Sensitivity. Essas técnicas permitem identificar fragmentos com tamanho médio entre 150–200 pb, característicos de cfDNA tumoral, e avaliar a qualidade para preparação de bibliotecas.

#### **3.10.2. Preparação de bibliotecas de cfDNA e Sequenciamento de Nova Geração (NGS)**

A construção das bibliotecas foi realizada com o kit Oncomine™ cfDNA Assay (Thermo Fisher Scientific), contendo mais de 200 mutações distribuídas em 14 genes, apropriado para a análise de mutações em genes relevantes no câncer colorretal a partir de DNA livre circulante. A quantificação das bibliotecas foi feita com o Ion Library TaqMan™ Quantitation Kit, e o controle de qualidade foi realizado por meio da Agilent TapeStation 2200 com o kit D1000 High Sensitivity. As bibliotecas qualificadas foram carregadas no chip 530 utilizando o equipamento Ion Chef™. O sequenciamento foi conduzido no sistema Ion GeneStudio™ S5 Plus, com profundidade média de 50.000×, cobrindo regiões-alvo com alta uniformidade e sensibilidade para variantes

com frequência alélica de até 0,1% (116).

### **3.10.3. Análise bioinformática e frequência alélica**

A análise dos dados foi iniciada na plataforma Torrent Suite™ (Thermo Fisher), com alinhamento das leituras ao genoma de referência hg19 e controle de qualidade inicial. Os arquivos de extensão “.bam” foram enviados à plataforma Ion Reporter™ para processamento no *workflow* OncoPrint™ Liquid Biopsy w1.9. Foram incluídas na análise apenas as amostras com pelo menos 3.300.000 leituras, cobertura molecular >3.500×, ≥90% das leituras no hotspot-alvo e ≥90% de uniformidade. Variantes foram consideradas quando apresentaram ao menos duas contagens moleculares e frequência alélica ≥0,065%. A frequência alélica molecular foi definida como a proporção de fragmentos mutantes em relação ao total de fragmentos moleculares. Variantes germinativas com frequência >40% ou classificadas como SNPs foram excluídas. A anotação das variantes foi realizada por meio da ferramenta Variant Effect Predictor (VEP) com base nos bancos gnomAD e COSMIC, sendo consideradas apenas aquelas classificadas como somáticas patogênicas ou confirmadas (117,118).

### **3.11. Análise da carga mutacional no plasma correlacionada ao volume tumoral radiológico**

Inicialmente comparamos a associação entre o quantitativo total de DNA livre circulante (cfDNA) recuperado de cada paciente no momento do diagnóstico, com os volumes tumorais obtidos por imagem. Em seguida, investigamos a associação entre a carga mutacional do DNA circulante de origem tumoral (ctDNA) e os volumes tumorais obtidos por imagem, investigada com o objetivo de explorar seu potencial valor prognóstico. A carga mutacional do ctDNA foi definida como a concentração total de mutações específicas do câncer, obtida por meio de sequenciamento de nova geração (NGS) a partir da biópsia líquida.

De caráter experimental, apesar da característica dinâmica do cfDNA, investigamos se a diluição do ctDNA ajustado à quantidade de cfDNA poderia se correlacionar ao volume tumoral, o que chamamos de “carga mutacional ajustada”. Essa medida foi calculada pela fórmula:

Carga mutacional ajustada = VAF (frequência alélica da mutação) / concentração de cfDNA (ng/mL) / volume tumoral (mL).

O volume tumoral foi determinado a partir da segmentação manual nas imagens de tomografia computadorizada (CT), utilizando o software ITK-SNAP.

## 4. Resultados

### 4.1. População do estudo

Um total de 53 pacientes com 55 tumores foram incluídos na análise final. A idade mediana foi de 62 anos (intervalo: 40–87), e 31 pacientes (58%) eram do sexo feminino. Todos os pacientes realizaram estadiamento pré-operatório por tomografia computadorizada (TC), e nenhum recebeu quimio ou radioterapia neoadjuvante. Com base na análise histopatológica, 11 pacientes (21%) foram classificados como T1–T2 (4 T1, 7 T2) e 42 pacientes (79%) como T3–T4 (28 T3, 14 T4). Em relação ao estadiamento linfonodal, 20 pacientes (38%) eram N0 e 33 (62%) eram N+. Não foram observadas diferenças significativas na distribuição de idade ou sexo entre os grupos.

**Tabela 6. Características clínicas basais da coorte do estudo (n = 53 pacientes).**

<b>Característica</b>	<b>Valor</b>
<b>Idade (anos)</b>	62,6 ± 9,9 (intervalo: 40–87)
<b>Sexo, n (%)</b>	
Feminino	31 (58%)
Masculino	22 (42%)
<b>Estadiamento T, n (%)</b>	
T1–T2	11 (21%)
T3–T4	42 (79%)
<b>Estadiamento N, n (%)</b>	
Negativo (N0)	20 (38%)
Positivo (N1–N2)	33 (62%)

### 4.2. Volumetria tumoral

Os exames tomográficos foram segmentados pelo software ITK-SNAP, que gerou os valores de volumetria tumoral de cada paciente a partir da contagem dos pixels da região de interesse selecionada manualmente. Os dados foram gerados em mm<sup>3</sup> e tabelados em cm<sup>3</sup> para melhor correlação clínica posteriormente. A média dos volumes foi de 54,2 cm<sup>3</sup> e mediana de 35,73 cm<sup>3</sup>, com menor valor de 2,25 cm<sup>3</sup> e maior de 389,1 cm<sup>3</sup>.

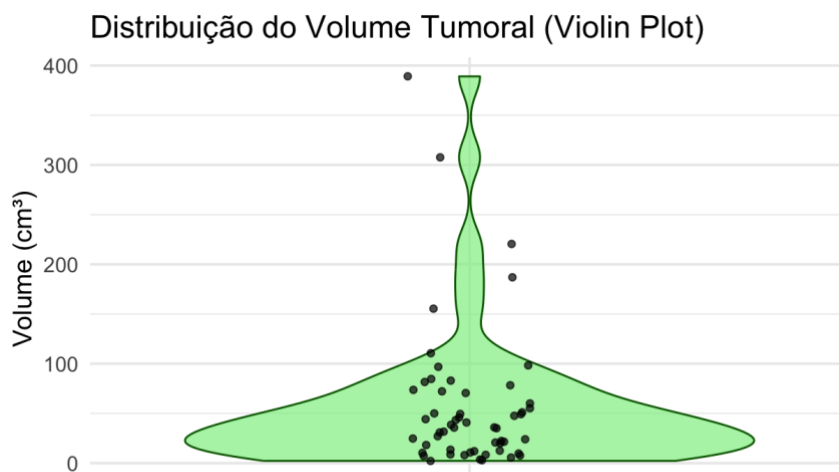


Figura 8. Distribuição do volume tumoral por Violin plot.

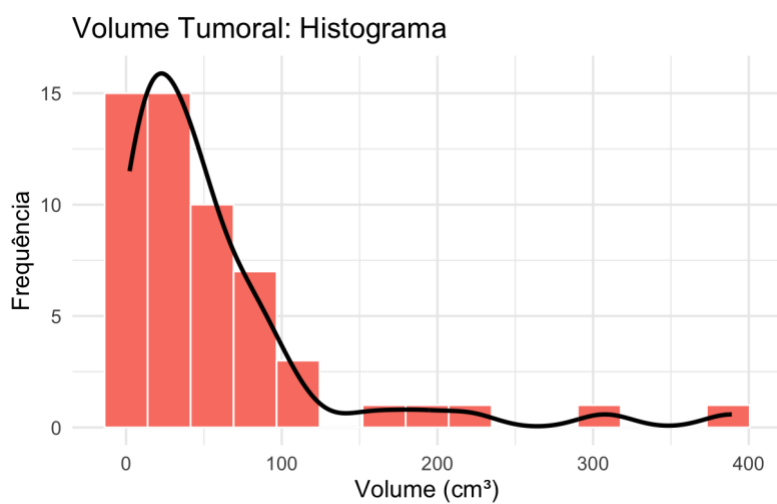


Figura 9. Distribuição do volume tumoral por Histograma.

Fonte: Autoria própria.



Figura 10. Representação da segmentação do paciente CBL 38 e sua peça cirúrgica correspondente.

Fonte: Autoria própria.

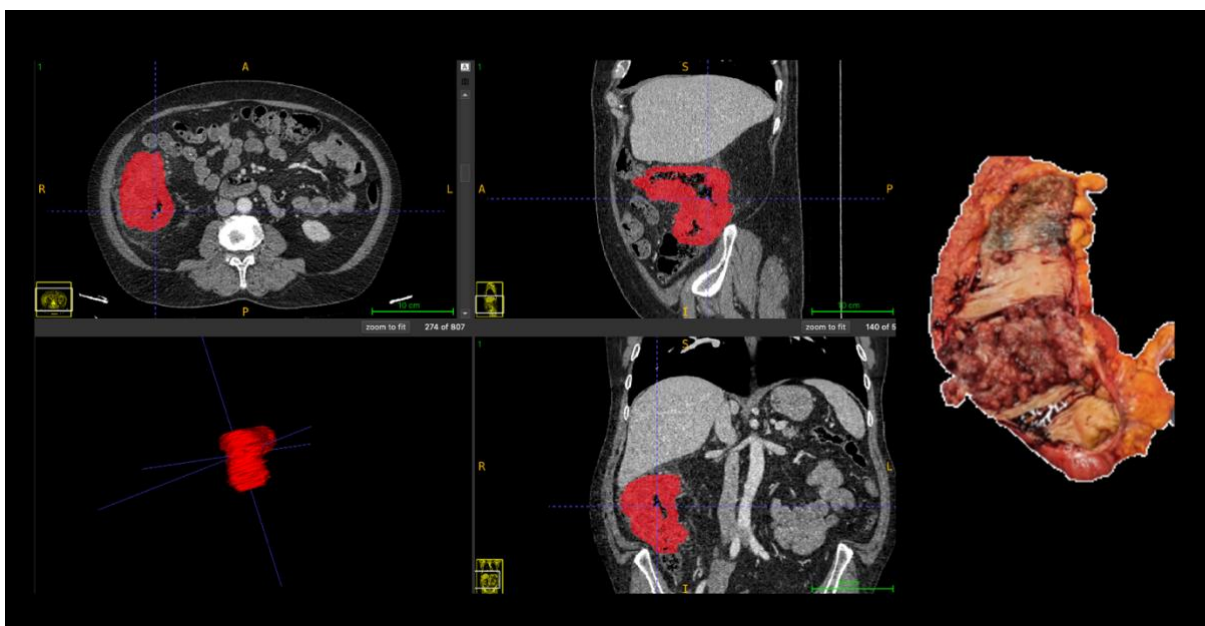


Figura 11. Representação da segmentação do paciente CBL 18 e peça cirúrgica correspondente.

Fonte: Autoria própria.

### 4.3. Assinatura radiômica

Para a predição do estágio T (T1–T2 vs. T3–T4), foi realizado um total de 990 experimentos, abrangendo todas as combinações de 15 subconjuntos de *features*

radiômicas, 6 estratégias de pré-processamento e 11 classificadores ( $15 \times 6 \times 11 = 990$ ). Em cada experimento, foram executadas 100 repetições, e os resultados médios foram analisados. De forma semelhante, 990 experimentos (com 100 repetições) também foram conduzidos para a predição do estágio N (N0 vs. N+).

Devido ao desequilíbrio entre classes, a maioria dos experimentos apresentou alta sensibilidade média; no entanto, apenas uma fração deles exibiu equilíbrio razoável entre sensibilidade e especificidade. Por simplicidade, foram rotulados como “bem-sucedidos” os experimentos que combinaram sensibilidade média superior a 70% e especificidade média superior a 50%, métricas escolhidas empiricamente para escolha de dados mais relevantes.

A **Tabela 7** apresenta a avaliação das diferentes estratégias de pré-processamento.

Na predição do estágio T, os melhores resultados foram obtidos com features normalizadas com aumento de dados via SMOTE seguido de redução por PCA (26 de 165 experimentos foram “bem-sucedidos”). Na predição do estágio N, os melhores resultados ocorreram com features normalizadas com aumento de dados via SMOTE, sem redução por PCA (37 de 165 experimentos foram “bem-sucedidos”). A redução por PCA mostrou-se benéfica na predição do estágio T, mas prejudicial na predição do estágio N. Devido ao forte desequilíbrio de classes, o aumento de dados pelo SMOTE foi fundamental em ambos os cenários.

**Tabela 7. Avaliação dos diferentes cenários de pré-processamento: (i) features brutas, (ii) features normalizadas, (iii) features normalizadas e reduzidas por PCA, (iv) features normalizadas com aumento de dados via SMOTE, (v) features normalizadas e reduzidas por PCA seguidas de aumento de dados via SMOTE e (vi) features normalizadas com aumento de dados via SMOTE seguidas de redução por PCA. Os valores indicam a porcentagem de experimentos de predição do estágio T e do estágio N — entre um total de 165 experimentos em cada caso, abrangendo todas as combinações de 15 conjuntos de features e 11 classificadores — que foram considerados “bem-sucedidos”, ou seja, aqueles que combinaram sensibilidade média superior a 70% e especificidade média superior a 50%.**

<b>preprocessing</b>	<b>T staging</b>	<b>N staging</b>
raw	0.6	5
normalized	0.6	5
norm-PCA	1	4
norm-SMOTE	7	<b>22</b>
norm-PCA-SMOTE	10	10
norm-SMOTE-PCA	<b>16</b>	6

As Tabelas 8 e 9 apresentam a taxa de “sucesso” dos experimentos realizados com os diferentes subconjuntos de features radiômicas. Na predição do estágio T, utilizando a estratégia de pré-processamento (vi), os melhores resultados foram obtidos empregando apenas *features* baseadas em forma, sendo que 8 dos 11 classificadores foram considerados “bem-sucedidos”. A combinação de *features* de forma com *features* de primeira ordem e/ou de textura não trouxe benefícios adicionais.

Por outro lado, na predição do estágio N, utilizando a estratégia de pré-processamento (iv), os melhores resultados foram alcançados com o uso de apenas *features* de textura — por exemplo, 7 dos 11 classificadores foram “bem-sucedidos” utilizando apenas *features* GLCM, e 6 dos 11 classificadores apresentaram sucesso ao empregar todas as features de textura. A combinação de *features* de textura com *features* de forma, feita experimentalmente para testar combinação de classes, mostrou-se prejudicial ao desempenho.

**Tabela 8. Avaliação das classes de features radiômicas utilizadas para a predição dos estágios T e N. Os valores indicam a porcentagem de classificadores — entre um total de 11 — que foram considerados “bem-sucedidos”, isto é, que combinaram sensibilidade média superior a 70% e especificidade média superior a 50%.**

feature set	T staging*	N staging**
firstorder18	0	9
shape14	<b>73</b>	0
glcm24	0	<b>64</b>
glrlm16	0	45
glszm16	0	45
ngtdm5	9	9
gldm14	9	27
* pre-processing: norm-SMOTE-PCA		
** pre-processing: norm-SMOTE		

Tabela 9. Avaliação dos subconjuntos que combinam diferentes classes de features radiômicas, mostrando também a porcentagem de classificadores (entre 11) que foram “bem-sucedidos” segundo o mesmo critério de desempenho (sensibilidade média >70% e especificidade média >50%).

feature set	T staging*	N staging**
shapefirstorder32	18	0
shapeglcm38	9	18
shapeglrlm30	9	27
shapeglszm30	18	9
shapengtdm19	<b>73</b>	0
shapegldm28	18	18
texture75	0	<b>54</b>
fullset107	0	9
* pre-processing: norm-SMOTE-PCA		
** pre-processing: norm-SMOTE		

A Tabela 10 apresenta os resultados da predição do estágio T obtidos com cada classificador, utilizando o conjunto de features shape14 e a estratégia de pré-processamento (vi). Destaca-se o desempenho do classificador de regressão logística (LR), que apresentou sensibilidade média de 72,1%, especificidade média de 68,0%, acurácia balanceada média de 70,0% e AUC média de 75,1%; a matriz de confusão correspondente está mostrada na Figura 7(a).

Além disso, observou-se que a regressão logística apresentou maior incidência de “sucesso” em comparação com outros classificadores, quando considerados diferentes subconjuntos de features e/ou outras estratégias de pré-processamento (dados não mostrados).

Tabela 10. resultados da predição do estágio T obtidos com diferentes algoritmos de classificação. São exibidos os valores médios e os respectivos intervalos de confiança de 95% (todos expressos em porcentagem). Os resultados foram obtidos utilizando o conjunto de features shape14 em conjunto com a estratégia de pré-processamento (vi).

classifier	Se	Spe	BA	AUC
LR	72 (69-75)	<b>68 (63-73)</b>	<b>70 (67-73)</b>	<b>75 (72-78)</b>
KNN	75 (72-78)	62 (57-66)	68 (66-71)	71 (68-74)
NB	70 (68-73)	59 (53-64)	65 (62-67)	70 (66-74)
DT	80 (77-83)	52 (47-58)	66 (63-69)	66 (63-69)
ET	<b>86 (84-88)</b>	47 (42-53)	67 (64-70)	70 (66-73)
RF	85 (83-88)	49 (44-55)	67 (64-70)	68 (65-72)
MLP	81 (78-84)	55 (50-61)	68 (65-71)	73 (69-76)
AB	77 (74-80)	51 (45-57)	64 (61-67)	69 (65-72)
GB	76 (73-79)	51 (45-56)	63 (60-66)	69 (65-72)
LDA	65 (62-68)	<b>68 (63-73)</b>	67 (64-70)	73 (70-77)
XGB	83 (80-85)	55 (50-60)	69 (66-72)	67 (64-71)

A Tabela 11 apresenta os resultados da predição do estágio N obtidos com cada classificador, utilizando o conjunto de features texture75 e a estratégia de pré-processamento (iv). Destaca-se o desempenho do classificador AdaBoost (AB), que apresentou sensibilidade média de 74,2%, especificidade média de 62,2%, acurácia balanceada média de 68,2% e AUC média de 75,0%; a matriz de confusão correspondente está mostrada na Figura 7(b).

Além disso, observou-se que o AdaBoost apresentou maior incidência de “sucesso” em comparação com outros classificadores, quando considerados diferentes subconjuntos de features e/ou outras estratégias de pré-processamento (dados não mostrados).

Tabela 11. Resultados da predição do estágio N obtidos com diferentes algoritmos de classificação. São exibidos os valores médios e os respectivos intervalos de confiança de 95% (todos expressos em porcentagem). Os resultados foram obtidos utilizando o conjunto de features texture75 em conjunto com a estratégia de pré-processamento (iv).

classifier	Se	Spe	BA	AUC
LR	61 (58-65)	46 (41-50)	54 (51-56)	55 (52-58)
KNN	71 (67-74)	60 (55-64)	65 (63-68)	70 (68-73)
NB	46 (42-49)	<b>69 (64-73)</b>	57 (55-60)	62 (59-65)
DT	68 (65-71)	50 (45-55)	59 (56-62)	59 (56-62)
ET	<b>81 (78-83)</b>	52 (48-57)	66 (64-69)	72 (70-75)
RF	75 (72-78)	54 (50-58)	65 (62-67)	72 (69-75)
MLP	69 (66-72)	45 (40-50)	57 (54-59)	60 (57-63)
AB	74 (71-77)	62 (58-67)	<b>68 (66-71)</b>	<b>75 (72-78)</b>
GB	71 (68-74)	57 (52-61)	64 (61-67)	70 (67-73)
LDA	64 (61-68)	53 (49-57)	59 (56-61)	59 (56-63)
XGB	74 (71-77)	55 (50-59)	64 (62-67)	74 (71-77)

(a)

		Predicted		
Total: 14		T3-T4: 8.89	T1-T2: 5.11	
Actual	T3-T4: 11	TP: 7.93	FN: 3.07	Se: 72.1%
	T1-T2: 3	FP: 0.96	TN: 2.04	Spe: 68.0%
		Pre: 89.9%	NPV: 39.9%	Acc: 71.2%

(b)

		Predicted		
Total: 14		N+: 8.57	N0: 5.43	
Actual	N+: 9	TP: 6.68	FN: 2.32	Se: 74.2%
	N0: 5	FP: 1.89	TN: 3.11	Spe: 62.2%
		Pre: 78.8%	NPV: 57.3%	Acc: 69.9%

Figura 12. Matrizes de confusão para: (a) a predição do estágio T, utilizando o classificador de regressão logística aplicado ao conjunto de features shape14 com a estratégia de pré-processamento (vi); e (b) a predição do estágio N, utilizando o classificador AdaBoost aplicado ao conjunto de features texture75 com a estratégia de pré-processamento (iv).

Metodologicamente, nosso pipeline aderiu aos princípios estabelecidos da pesquisa radiômica, incorporando pré-processamento padronizado (normalização z-score), redução de dimensionalidade para mitigar redundância entre características (PCA) e estratégias de reamostragem para lidar com desbalanceamento de classes (SMOTE). A reprodutibilidade dos resultados entre classificadores e estratégias de validação reforçam a robustez e estabilidade da abordagem.

#### 4.4. Correlação de dados com a biópsia líquida

A Tabela 12 apresenta os dados individuais dos 29 pacientes, com dados de imagem e genéticos completos, incluídos na análise da carga mutacional ajustada, com as respectivas frequências alélicas (VAF%), concentrações plasmáticas de cfDNA, volumes tumorais segmentados nas imagens de tomografia computadorizada e valores de carga mutacional ajustada.

Os valores de VAF variaram entre 0% e 2,47%, refletindo ampla heterogeneidade na proporção de fragmentos mutados circulantes entre os pacientes. As concentrações de cfDNA oscilaram de 2,3 a 30,8 ng/mL, enquanto os volumes tumorais variaram expressivamente de 2,9 a 389,1 mL, evidenciando a diversidade morfológica da amostra.

De forma experimental, investigamos se a diluição do ctDNA ajustado à quantidade de cfDNA poderia se correlacionar ao volume tumoral, o que chamamos de “carga mutacional ajustada”. A carga mutacional ajustada, calculada pela razão entre VAF, cfDNA e volume tumoral, apresentou valores predominantemente baixos, com amplitude entre  $4,78 \times 10^{-8}$  e  $1,80 \times 10^{-3}$ . A maioria dos pacientes apresentou valores inferiores a  $10^{-5}$ , indicando baixa densidade mutacional circulante relativa ao tamanho tumoral e à concentração total de cfDNA.

Observou-se, ainda, que os maiores valores de carga mutacional ajustada ocorreram em casos com tumores de menor volume e/ou menor concentração plasmática de cfDNA. O paciente CBL24 apresentou o valor mais elevado da coorte ( $1,80 \times 10^{-3}$ ), destacando-se por associar um volume tumoral reduzido (5,6 mL) e frequência alélica elevada (2,47%). De modo semelhante, os pacientes CBL31 (2,9 mL) e CBL36 (3,6 mL) também exibiram valores relativamente altos de carga ajustada ( $8,0 \times 10^{-5}$  e  $8,8 \times 10^{-5}$ , respectivamente), mesmo com frequências alélicas mais modestas.

**Tabela 12. Dados individuais dos pacientes incluídos no estudo**

Paciente	Frequência alélica (%)	cfDNA (ng/mL)	Volume tumoral (mL)	Carga mutacional ajustada
CBL4	0.26	11.96	43.36	$4.99 \times 10^{-6}$
CBL6	0.00	2.34	12.43	0.00

CBL13	0.23	10.65	83.03	$2.61 \times 10^{-6}$
CBL14	0.52	5.21	12.60	$7.87 \times 10^{-5}$
CBL18	0.00	30.75	389.10	0.00
CBL22	0.54	5.74	18.14	$5.20 \times 10^{-5}$
CBL24	2.47	2.46	5.58	$1.80 \times 10^{-3}$
CBL28	0.06	13.43	8.27	$5.06 \times 10^{-6}$
CBL31	0.16	6.80	2.90	$8.00 \times 10^{-5}$
CBL32	0.11	3.74	50.01	$6.00 \times 10^{-6}$
CBL36	0.21	6.54	3.60	$8.77 \times 10^{-5}$
CBL37	0.77	7.99	30.80	$3.13 \times 10^{-5}$
CBL38	0.00	7.86	7.70	0.00
CBL40	0.00	5.18	8.80	0.00
CBL42	0.29	9.56	49.53	$6.18 \times 10^{-6}$
CBL45	1.80	25.67	35.15	$1.99 \times 10^{-5}$
CBL47	1.78	5.99	110.50	$2.69 \times 10^{-5}$
CBL50	0.00	8.84	13.52	0.00
CBL53	0.28	26.61	31.67	$3.38 \times 10^{-6}$
CBL56	0.00	7.23	78.33	0.00
CBL62	0.00	4.34	24.66	0.00
CBL72	1.33	7.27	73.73	$2.49 \times 10^{-5}$
CBL75	0.24	5.53	35.73	$1.21 \times 10^{-5}$
CBL79	0.36	7.65	20.07	$2.33 \times 10^{-5}$
CBL84	0.04	24.82	307.60	$4.78 \times 10^{-8}$
CBL85	1.15	5.06	55.29	$4.11 \times 10^{-5}$
CBL92	0.17	6.80	60.11	$4.28 \times 10^{-6}$
CBL97	0.00	5.01	21.46	0.00
CBL98	1.00	6.97	155.30	$9.20 \times 10^{-6}$

**Valores apresentados de frequência alélica (VAF%), concentração plasmática de cfDNA, volume tumoral segmentado por ITK-SNAP e carga mutacional ajustada (em notação científica). O valor da carga mutacional foi calculado segundo a fórmula:  $VAF / (cfDNA \times \text{volume tumoral})$ .**

Devido à assimetria das distribuições de cfDNA e ctDNA, optou-se pela aplicação do teste de correlação de Spearman para avaliar a relação entre esses marcadores circulantes e o volume tumoral. A avaliação entre o volume tumoral e a concentração plasmática de cfDNA encontra-se ilustrada na Figura 8. Não foi observada correlação estatisticamente significativa entre essas variáveis ( $\rho = 0,32$ ;  $p = 0,085$ ), embora tenha sido identificada uma tendência de associação positiva fraca.

A relação entre o volume tumoral e a fração alélica do ctDNA (VAF) está apresentada na Figura 9. De forma semelhante ao cfDNA, não foi identificada correlação significativa entre o volume tumoral e o VAF ( $\rho = 0,16$ ;  $p = 0,42$ ),

sugerindo que a quantidade de DNA tumoral circulante não refletiu diretamente o tamanho anatômico do tumor na amostra analisada.

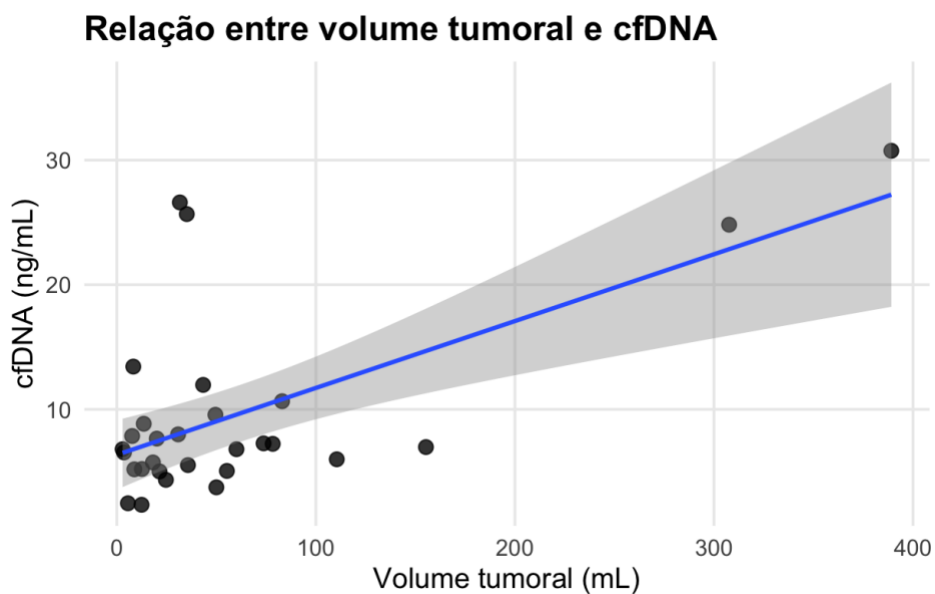


Figura 13. Relação entre a concentração plasmática de cfDNA obtida no momento do diagnóstico e o volume tumoral segmentado nas imagens de tomografia computadorizada. Cada ponto representa um paciente. Observa-se uma tendência de correlação positiva fraca entre as variáveis, embora sem significância estatística ( $\rho = 0,32$ ;  $p = 0,085$ ).

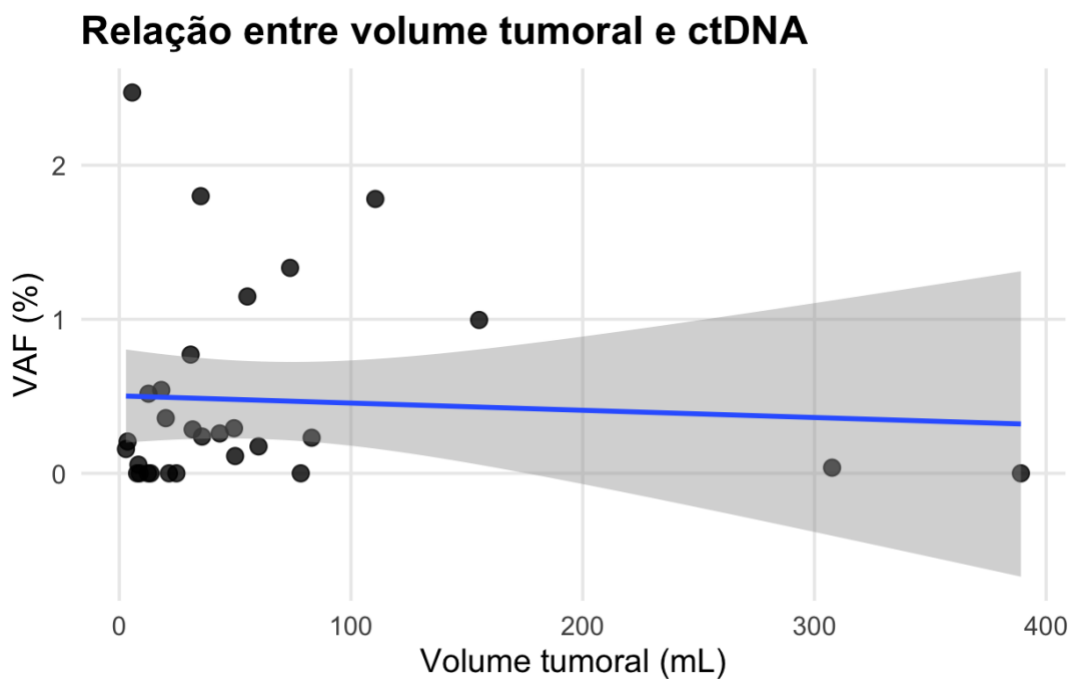
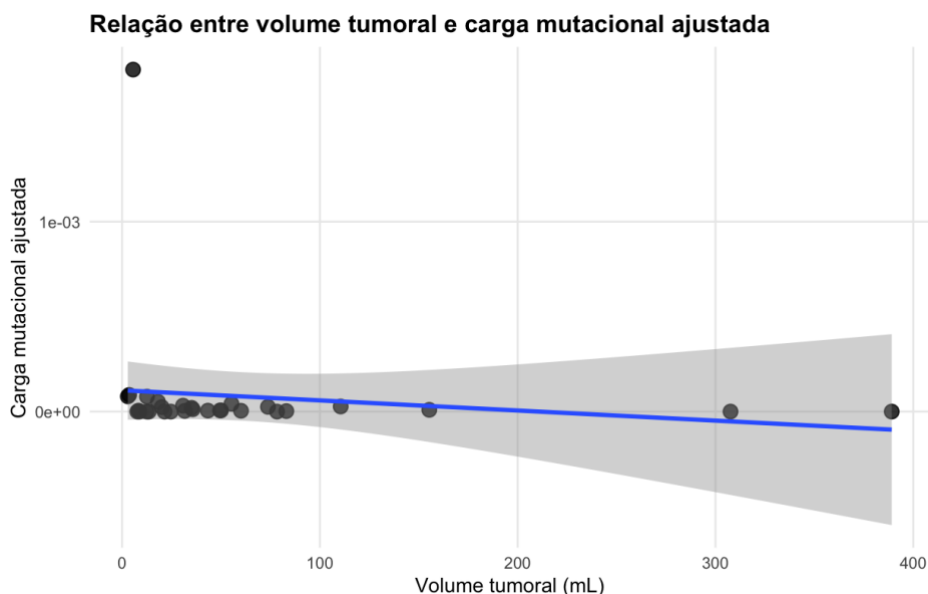


Figura 14. Relação entre a frequência alélica do ctDNA (VAF%) e o volume tumoral segmentado nas imagens de tomografia computadorizada. Cada ponto representa um paciente. Não se observou correlação significativa entre as variáveis ( $\rho = 0,16$ ;  $p = 0,42$ ).

A análise de correlação entre o volume tumoral e a carga mutacional ajustada não demonstrou associação significativa ( $\rho = -0,21$ ;  $p = 0,276$ ), revelando uma tendência negativa fraca e não consistente entre as variáveis (Figura 10).



**Figura 15.** Relação entre a carga mutacional ajustada e o volume tumoral segmentado nas imagens de tomografia computadorizada. Cada ponto representa um paciente. Não se observou correlação significativa entre as variáveis ( $\rho = -0,21$ ;  $p = 0,276$ ).

## 5. Discussão

### 5.1. Assinatura radiômica

Este estudo demonstrou que a radiômica baseada em TC pode fornecer informações clinicamente relevantes para o estadiamento pré-operatório do câncer colorretal, particularmente na diferenciação entre tumores T1–T2 e T3–T4. Após a aplicação de normalização, PCA e SMOTE em características baseadas em forma, o melhor desempenho foi obtido com o classificador LR, que apresentou sensibilidade média de 0,721, especificidade de 0,680, acurácia balanceada de 0,70 e AUC de 0,751. Outros classificadores, incluindo KNN (AUC = 0,708) e LDA (AUC = 0,732), exibiram desempenho intermediário, enquanto DT (AUC = 0,662) e XGB (AUC = 0,62) apresentaram menor capacidade discriminativa. Importante destacar que modelos de conjunto, como o classificador ET, atingiram a maior sensibilidade (0,86), porém às custas de redução na especificidade (0,473). Esses achados ressaltam a capacidade

das características radiômicas de capturar assinaturas morfológicas de invasão tumoral. Ao quantificar atributos macroscópicos como irregularidade, alongação e complexidade de superfície, tais características podem refletir indiretamente os padrões microscópicos de infiltração da parede e de ruptura do contorno que os patologistas utilizam para definir as categorias T. Essa correspondência sugere que a radiômica baseada em forma pode atuar como um correlato imagético de invasão histopatológica, reforçando seu potencial como biomarcador não invasivo para o estadiamento pré-operatório e o planejamento terapêutico no câncer colorretal.

A predição do acometimento linfonodal (N0 vs. N+) alcançou desempenho comparável ao do estadiamento T. Após aplicação de normalização e SMOTE em características texturais, o melhor desempenho foi obtido com o classificador AB, que atingiu sensibilidade média de 0,742, especificidade de 0,622, acurácia balanceada de 0,682 e AUC de 0,75. Classificadores de conjunto, como ET (AUC = 0,723) e RF (AUC = 0,721), também apresentaram resultados robustos, reforçando a reprodutibilidade das assinaturas radiômicas entre diferentes tarefas. Notavelmente, a especificidade aumentou ligeiramente em comparação à análise do estadiamento T, indicando melhor equilíbrio entre predições verdadeiramente positivas e verdadeiramente negativas. A avaliação das *features* demonstrou que matrizes derivadas de textura, como GLCM, GLRLM e GLSZM, foram as que mais contribuíram para a predição linfonodal, enquanto as características de forma predominaram na análise do estadiamento T. Esse padrão sugere que descritores texturais capturam a heterogeneidade intratumoral — incluindo variações na densidade celular, necrose e composição estromal — que os patologistas associam a agressividade biológica e disseminação linfática. Assim, a radiômica baseada em textura poderia atuar como um correlato quantitativo da complexidade microambiental tumoral, complementando as características de forma, que refletem principalmente a invasão local. Apesar disso, como a segmentação foi limitada ao tumor primário, sem delineamento direto dos linfonodos, modelos futuros que incorporem regiões peritumorais ou linfonodais poderão aprimorar ainda mais o desempenho classificatório.

Estudos anteriores que alcançaram resultados superiores na predição linfonodal, com AUCs variando de 0,75 a 0,85, frequentemente empregaram abordagens de segmentação direta ou multirregional. Liu et al. demonstraram que combinar radiômica intratumoral e peritumoral baseada em RM com dados clínicos

melhorou a predição de metástase linfonodal no câncer retal, atingindo AUC de 0,83 na coorte de validação(119). De forma semelhante, Yuan et al. mostraram que um modelo clínico-radiômico combinado, integrando características intratumorais e peritumorais derivadas de TC, alcançou AUC de 0,83 na predição nodal em carcinoma retal(120). Eresen et al. também relataram alto desempenho na classificação do status linfonodal do câncer de cólon quando os linfonodos foram individualmente segmentados e analisados usando um modelo de extreme gradient boosting (XGBoost), atingindo AUC de 0,83, acurácia de 0,83, sensibilidade de 0,81 e especificidade de 0,85 (121). Embora nossas *features* tenham sido extraídas exclusivamente do tumor primário, o modelo ainda assim alcançou desempenho discriminativo comparável para o estadiamento N, sugerindo que a heterogeneidade intratumoral, por si só, é capaz de capturar parte do sinal biológico associado à disseminação linfonodal.

Nossos achados são consistentes com outros estudos baseados em TC. Huang et al. desenvolveram e validaram um nomograma radiômico combinando características derivadas de TC com fatores clínicos para predição pré-operatória de metástase linfonodal em câncer colorretal, relatando C-index de 0,736 na coorte de treinamento e 0,778 na validação externa (122). Zhang et al. demonstraram que uma assinatura radiômica baseada em TC foi capaz de discriminar adenocarcinomas colorretais de alto e baixo grau, atingindo AUC de 0.812 no conjunto de treinamento, 0.735 na validação e até 0.895 em subgrupos de câncer de reto (123). Em geral, esses trabalhos sustentam a robustez da radiômica baseada em TC para caracterização do tumor primário.

A radiômica baseada em RM tem sido investigada de forma mais extensa no câncer retal, frequentemente reportando resultados superiores tanto para estadiamento T quanto N. Por exemplo, Patanè et al. desenvolveram um modelo radiômico baseado em RM ponderada em T2 que atingiu AUC de 0,82 para estadiamento T (124). Outras análises baseadas em histogramas ponderados em T2 demonstraram desempenho moderado para a predição de metástase linfonodal, com AUCs variando entre 0,65 e 0,75, conforme relatado por Yang et al.(125). Em comparação com esses estudos em RM, nosso modelo baseado em TC apresentou desempenho semelhante para o estadiamento T e capacidade discriminativa comparável para a predição nodal, apesar de utilizar exclusivamente segmentação

tumoral. Esse resultado reforça o potencial das *features* radiômicas em capturar marcadores relevantes tanto de invasão local quanto de disseminação linfonodal, mesmo sem delineamento explícito dos linfonodos.

Do ponto de vista clínico, a discriminação acurada do estágio T é crucial. O superestadiamento de tumores T2 pode expor pacientes a radioquimioterapia desnecessária e à sua morbidade associada, enquanto o subestadiamento de tumores T3 aumenta o risco de ressecção incompleta e recidiva. Em nossa análise, os modelos radiômicos alcançaram AUC média de 0,751, sensibilidade de 0,721 e especificidade de 0,68, indicando bom desempenho na identificação de doença avançada (T3–T4). Esses achados sugerem que a radiômica pode complementar métodos convencionais de estadiamento, particularmente em centros onde a RM não é amplamente disponível ou em câncer de cólon, onde a TC permanece o principal método de imagem.

Além disso, a predição do acometimento linfonodal alcançou acurácia comparável, com o classificador AB atingindo AUC de 0,750. Isso sugere que *features* radiômicas derivadas exclusivamente do tumor podem codificar informações biologicamente significativas tanto sobre invasão local quanto sobre disseminação linfonodal.

Adicionalmente, a integração com biomarcadores moleculares surgiu como uma direção promissora: Lubner et al. demonstraram que características texturais derivadas da TC em metástases hepáticas de câncer colorretal correlacionaram-se com o status mutacional de KRAS e com desfechos clínicos (126). Essas estratégias radiogenômicas reforçam o potencial de características de imagem como preditores não invasivos da biologia tumoral. Apesar do importante potencial, essas abordagens permanecem amplamente investigacionais no câncer colorretal e requerem validação em coortes maiores e prospectivas.

## **5.2. Correlação de dados com a biópsia líquida**

A carga mutacional ajustada, calculada pela razão entre VAF, cfDNA e volume tumoral, apresentou valores predominantemente baixos, indicando que a maior parte dos pacientes exibia baixa densidade mutacional circulante relativa ao tamanho tumoral e à quantidade total de cfDNA. Observou-se, no entanto, que alguns casos

isolados apresentaram valores elevados dessa métrica, mesmo em tumores de menor volume e/ou em cenários com baixa concentração plasmática de cfDNA. Esses achados sugerem que, embora a maioria dos tumores pequenos liberem quantidades reduzidas de DNA tumoral na circulação, um subconjunto pode exibir liberação proporcionalmente maior, possivelmente refletindo diferenças biológicas individuais, como maior atividade proliferativa, vascularização aumentada ou instabilidade genômica.

Esse comportamento heterogêneo também foi refletido nas análises de correlação entre marcadores circulantes e volume tumoral. Não foi identificada correlação significativa entre o volume tumoral e a concentração plasmática de cfDNA ( $\rho = 0,32$ ;  $p = 0,085$ ), embora tenha sido observada uma tendência positiva fraca. De maneira semelhante, a frequência alélica do ctDNA (VAF) não apresentou correlação significativa com o volume tumoral ( $\rho = 0,16$ ;  $p = 0,42$ ). Esses resultados demonstram que, em nível populacional, o tamanho anatômico das lesões não se associa de forma consistente à quantidade de cfDNA ou ctDNA detectada na circulação, reforçando a natureza não linear da liberação de material genético tumoral. Além disso, o cfDNA é um biomarcador intrinsecamente dinâmico e suscetível a variações transitórias desencadeadas por inflamação sistêmica, infecções, exercício físico, comorbidades crônicas e condições metabólicas, além de diferenças pré-analíticas de coleta e processamento (126). Esses fatores podem contribuir para a dispersão dos valores encontrados e atenuar possíveis relações entre volume tumoral e biomarcadores circulantes.

Esses achados são compatíveis com os dados da literatura sobre o comportamento do ctDNA em tumores sólidos, incluindo o câncer colorretal. Diehl et al. demonstraram que os níveis de ctDNA refletem principalmente a dinâmica tumoral, acompanhando reduções após ressecção e aumentos relacionados à recidiva, o que reforça o papel do turnover celular na quantidade de DNA tumoral circulante (65). De forma complementar, Wan et al. ressaltaram que tumores pequenos, porém biologicamente mais ativos — com maior proliferação, vascularização ou instabilidade genômica — podem liberar proporções elevadas de ctDNA, não necessariamente relacionadas ao tamanho anatômico(127).

Estudos de fragmentômica também demonstram que a liberação de cfDNA é bem heterogênea entre indivíduos. Cristiano et al. mostraram que perfis de

fragmentação e a quantidade total de cfDNA tumoral variam amplamente entre tumores com volume semelhante, refletindo diferenças intrínsecas de biologia tumoral (87). Mouliere et al. reforçaram que tanto a quantidade quanto o padrão de fragmentos de cfDNA dependem de mecanismos associados à agressividade tumoral, morte celular e remodelação epigenética (128). Esses dados sustentam a ausência de correlação linear entre volume tumoral e carga mutacional ajustada observada na presente coorte.

Outro ponto relevante é que, em tumores localizados, os níveis de ctDNA tendem a ser extremamente baixos. Reinert et al. demonstraram que uma proporção significativa de pacientes com câncer colorretal estágio I–III apresenta ctDNA indetectável mesmo com métodos de alta sensibilidade (68). Revisões recentes — como as de Sánchez-Herrero et al. e Masfarré et al. — também destacam que tumores localizados frequentemente liberam frações inferiores a 0,01% do cfDNA total, resultando em VAFs muito baixos e cargas mutacionais reduzidas (129,130). A predominância de valores muito baixos observada na presente coorte é, portanto, compatível com padrões já estabelecidos para doença localizada.

A heterogeneidade observada na carga mutacional ajustada também é consistente com estudos que avaliam a relação entre cfDNA e carga tumoral metabólica. Nygaard et al. e Winther-Larsen et al. demonstraram que, embora exista associação entre cfDNA e atividade tumoral (PET/CT), essa relação é frequentemente imperfeita e sujeita a grande variabilidade entre os indivíduos (131,132). Esses achados reforçam que a liberação de ctDNA é modulada por múltiplos fatores biológicos, e não apenas pelo volume anatômico.

Cabe ressaltar, entretanto, uma limitação conceitual relacionada ao cálculo da “carga mutacional ajustada”. Embora utilizada de forma exploratória, essa métrica depende da divisão do ctDNA pelo cfDNA total, que, por sua vez, é influenciado por múltiplos fatores não tumorais. Como o cfDNA é altamente dinâmico e sujeito a amplas variações decorrentes de comorbidades, inflamação, exercício físico e condições sistêmicas, ele não representa um denominador biologicamente estável ou relacionado exclusivamente ao tumor. Portanto, o ctDNA isoladamente — expresso pela frequência alélica (VAF) ou pela simples detecção de mutações — é a métrica mais adequada para comparação com volume tumoral e com características radiômicas. Dessa forma, a carga mutacional ajustada deve ser interpretada com

cautela, sendo considerada apenas uma análise exploratória que pode gerar hipóteses, mas não substitui parâmetros moleculares consolidados.

Em conjunto, os resultados demonstram que a liberação de cfDNA e ctDNA é modulada por um conjunto complexo de fatores biológicos e sistêmicos que não se refletem de forma direta no volume radiológico. Tanto o ctDNA isolado quanto a métrica exploratória de carga mutacional ajustada parecem captar aspectos da biologia tumoral — como proliferação, instabilidade genômica e eficiência de liberação de fragmentos — que não são evidenciados apenas pela avaliação anatômica. Assim, essas abordagens podem oferecer informações complementares ao volume tumoral e às características radiômicas, especialmente em cenários de doença localizada, nos quais o ctDNA é escasso e a caracterização molecular enfrenta desafios técnicos importantes.

### **5.3. Correlação radiômica e biópsia líquida**

Embora as análises radiômica e molecular tenham sido realizadas separadamente, uma das maiores contribuições potenciais deste estudo é demonstrar que a integração de ambas as abordagens pode oferecer um modelo não invasivo e multimodal de avaliação da doença. Tal integração pode permitir o desenvolvimento de ferramentas preditivas mais robustas, com maior valor clínico para estratificação de risco, decisão terapêutica e monitoramento.

### **5.4. Limitações**

Diversas limitações devem ser reconhecidas. Primeiro, este foi um estudo de centro único, com uma amostra relativamente pequena e desbalanceada, o que pode restringir a generalização dos achados. Segundo, a segmentação tumoral foi realizada manualmente, introduzindo potencial dependência do observador, apesar da utilização de software padronizado. Terceiro, nossa análise foi limitada a características radiômicas intratumorais extraídas exclusivamente do tumor primário, sem segmentação direta dos linfonodos. Por fim, a validação externa em múltiplas instituições e plataformas de imagem permanece essencial antes que esses modelos possam ser aplicados na prática clínica.

## 6. Conclusão

Este estudo demonstrou que características radiômicas extraídas de tomografia computadorizada são capazes de prever com boa acurácia o estadiamento T e N no câncer colorretal, utilizando apenas o tumor primário como volume de interesse. Assinaturas baseadas em forma foram mais eficazes para o estadiamento T, enquanto características texturais se destacaram na predição do acometimento linfonodal. Paralelamente, a análise da carga mutacional ajustada do cfDNA mostrou-se uma métrica complementar, refletindo a atividade tumoral e a heterogeneidade biológica de forma independente do volume tumoral.

A convergência entre radiômica e biópsia líquida representa um passo em direção à oncologia de precisão, oferecendo ferramentas não invasivas que podem melhorar a estratificação de risco e orientar decisões terapêuticas de forma individualizada. Este estudo teve como principal objetivo validar um pipeline metodológico de análise radiômica aplicável em coortes clínicas reais. Como próximo passo, pretendemos incorporar os dados moleculares — derivados de biópsia líquida e sequenciamento tecidual — às assinaturas radiômicas, com o objetivo de prever o perfil genético tumoral por meio de imagens, consolidando uma abordagem radiogenômica de predição não invasiva.

Para tradução clínica efetiva, são necessários estudos multicêntricos com maior poder estatístico, validação externa e integração de dados multimodais em modelos robustos de predição.

## 7. Conclusão dos Objetivos Específicos

- **Correlação entre características radiômicas e estadiamento anatomopatológico**

As características radiômicas demonstraram capacidade moderada de discriminar tumores iniciais de tumores localmente avançados, bem como de contribuir para a predição do comprometimento linfonodal. Diferentes classes de features mostraram desempenhos distintos, sugerindo que a radiômica pode captar aspectos complementares da agressividade tumoral.

- **Análise do sequenciamento do DNA livre circulante (cfDNA)**

A análise do cfDNA por painel gênico foi factível e evidenciou heterogeneidade molecular entre os pacientes avaliados.

- **Quantificação da frequência alélica das mutações (carga mutacional)**

A quantificação da frequência alélica revelou ampla variabilidade interindividual na carga mutacional circulante, indicando que a liberação de ctDNA pode variar independentemente do tamanho tumoral ou do estágio anatômico.

- **Comparação entre informações radiológicas e carga mutacional**

Não foi observada correlação linear significativa entre os parâmetros radiológicos, incluindo o volume tumoral, e a carga mutacional circulante. Entretanto, alguns casos apresentaram cargas mutacionais elevadas mesmo em tumores de pequeno volume.

- **Estabelecimento de potenciais biomarcadores de imagem não invasivos**

Os achados sugerem que assinaturas radiômicas derivadas de tomografia computadorizada apresentam potencial como biomarcadores de imagem não invasivos para auxiliar no estadiamento e na estratificação de risco do câncer colorretal, sendo necessária validação adicional em coortes maiores e multicêntricas.

## 8. Referências

1. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. *CA Cancer J Clin.* 2023;73(3):233–54.
2. Câncer IN de. Estimativa 2023: incidência de câncer no Brasil. Rio de Janeiro, RJ: Instituto Nacional De Câncer; 2023. 160 p.
3. Santos M de O, Lima FC da S de, Martins LFL, Oliveira JFP, Almeida LM de, Cancela M de C. Estimativa de Incidência de Câncer no Brasil, 2023-2025. *Rev Bras Cancerol.* 6 de fevereiro de 2023;69(1):e-213700.
4. Pires BB, Pereira LM, Silva Filho CO, Albuquerque MPA, Rocha Filho OMDS, Barbosa NB, et al. CÂNCER COLORRETAL E A IMPORTÂNCIA DA PREVENÇÃO E DIAGNÓSTICO PRECOCE NA PROMOÇÃO DA SAÚDE. Em: *Promoção da Saúde: conceito, estratégia e prevenção em pesquisa [Internet].* 1º ed. Editora Científica Digital; 2023 [citado 24 de outubro de 2025]. p. 85–97. Disponível em: <http://www.editoracientifica.com.br/articles/code/230513128>
5. Carcinoma Colorretal | Concise Medical Knowledge [Internet]. [citado 19 de dezembro de 2025]. Disponível em: <https://www.lecturio.com/pt/concepts/carcinoma-colorretal/>
6. Ogino S, Goel A. Molecular Classification and Correlates in Colorectal Cancer. *J Mol Diagn JMD.* janeiro de 2008;10(1):13–27.
7. Zinkeng A, Taylor FL, Cheong SH, Song H, Merchant JL. Early Onset Colorectal Cancer: Molecular Underpinnings Accelerating Occurrence. *Cell Mol Gastroenterol Hepatol.* 5 de novembro de 2024;19(2):101425.
8. Neto AM de M, Silva LM, Muniz AFN, Diniz TP. Tendência da mortalidade por Câncer Colorretal no estado do Piauí e Brasil no período de 2016 a 2020. *Braz J Health Rev.* 28 de julho de 2023;6(4):16100–22.
9. Pickhardt PJ. Imaging and Screening for Colorectal Cancer with CT Colonography. *Radiol Clin North Am.* novembro de 2017;55(6):1183–96.
10. Imperiale TF, et al. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med.* 2014;370(14):1287–1297. - Pesquisa Google [Internet]. [citado 24 de outubro de 2025]. Disponível em: <https://www.nejm.org/doi/full/10.1056/NEJMoa1311194>
11. Benson AB, Venook AP, Adam M, Chang G, Chen YJ, Ciombor KK, et al. Colon Cancer, Version 3.2024, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Netw JNCCN.* junho de 2024;22(2 D):e240029.
12. Beets-Tan RGH, Lambregts DMJ, Maas M, Bipat S, Barbaro B, Curvo-Semedo L, et al. Magnetic resonance imaging for clinical management of rectal cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting. *Eur Radiol.* abril de 2018;28(4):1465–75.
13. Rodríguez-Fraile M, Cózar-Santiago MP, Sabaté-Llobera A, Caresia-Aróstegui AP, Delgado Bolton RC, Orcajo-Rincon J, et al. FDG PET/CT in colorectal cancer. *Rev Espanola Med Nucl E Imagen Mol.* 2020;39(1):57–66.
14. Yang N. Radiopaedia. [citado 19 de dezembro de 2025]. Rectal cancer (staging) | Radiology Reference Article | Radiopaedia.org. Disponível em:

<https://radiopaedia.org/articles/rectal-cancer-staging>

15. Weiser MR. AJCC 8th Edition: Colorectal Cancer. *Ann Surg Oncol*. junho de 2018;25(6):1454–5.
16. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA Cancer J Clin*. março de 2017;67(2):93–9.
17. So JS, Cheong C, Oh SY, Lee JH, Kim YB, Suh KW. Accuracy of Preoperative Local Staging of Primary Colorectal Cancer by Using Computed Tomography: Reappraisal Based on Data Collected at a Highly Organized Cancer Center. *Ann Coloproctology*. outubro de 2017;33(5):192–6.
18. Twelves C, Wong A, Nowacki MP, Abt M, Burris H, Carrato A, et al. Capecitabine as adjuvant treatment for stage III colon cancer. *N Engl J Med*. 30 de junho de 2005;352(26):2696–704.
19. Dienstmann R, Salazar R, Tabernero J. Personalizing colon cancer adjuvant therapy: selecting optimal treatments for individual patients. *J Clin Oncol Off J Am Soc Clin Oncol*. 1º de junho de 2015;33(16):1787–96.
20. Bahadoer RR, Dijkstra EA, van Etten B, Marijnen CAM, Putter H, Kranenbarg EMK, et al. Short-course radiotherapy followed by chemotherapy before total mesorectal excision (TME) versus preoperative chemoradiotherapy, TME, and optional adjuvant chemotherapy in locally advanced rectal cancer (RAPIDO): a randomised, open-label, phase 3 trial. *Lancet Oncol*. janeiro de 2021;22(1):29–42.
21. Bismuth H, Adam R, Lévi F, Farabos C, Waechter F, Castaing D, et al. Resection of nonresectable liver metastases from colorectal cancer after neoadjuvant chemotherapy. *Ann Surg*. outubro de 1996;224(4):509–20; discussion 520-522.
22. Van Cutsem E, Cervantes A, Adam R, Sobrero A, Van Krieken JH, Aderka D, et al. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol Off J Eur Soc Med Oncol*. agosto de 2016;27(8):1386–422.
23. esmo-clinical-practice-guideline-metastatic-colorectal-cancer [Internet]. [citado 24 de outubro de 2025]. Disponível em: <https://www.esmo.org/guidelines/esmo-clinical-practice-guideline-metastatic-colorectal-cancer>
24. Lazaris A, Amri A, Petrillo SK, Zoroquiain P, Ibrahim N, Salman A, et al. Vascularization of colorectal carcinoma liver metastasis: insight into stratification of patients for anti-angiogenic therapies. *J Pathol Clin Res*. 14 de abril de 2018;4(3):184–92.
25. André T, Shiu KK, Kim TW, Jensen BV, Jensen LH, Punt C, et al. Pembrolizumab in Microsatellite-Instability–High Advanced Colorectal Cancer. *N Engl J Med*. 2 de dezembro de 2020;383(23):2207–18.
26. Overman MJ, Lonardi S, Wong KYM, Lenz HJ, Gelsomino F, Aglietta M, et al. Durable Clinical Benefit With Nivolumab Plus Ipilimumab in DNA Mismatch Repair-Deficient/Microsatellite Instability-High Metastatic Colorectal Cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 10 de março de 2018;36(8):773–9.
27. Zhu Y, Li X, Chen T, Wang J, Zhou Y, Mu X, et al. Personalised neoantigen-based

- therapy in colorectal cancer. *Clin Transl Med.* 3 de novembro de 2023;13(11):e1461.
28. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp.* 14 de novembro de 2018;2(1):36.
  29. Shur JD, Doran SJ, Kumar S, Ap Dafydd D, Downey K, O'Connor JPB, et al. Radiomics in Oncology: A Practical Guide. *Radiogr Rev Publ Radiol Soc N Am Inc.* outubro de 2021;41(6):1717–32.
  30. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *Radiology.* maio de 2020;295(2):328–38.
  31. ResearchGate [Internet]. [citado 19 de dezembro de 2025]. Figure 1: Workflow of the radiomics analysis. (1) Acquisition of... Disponível em: [https://www.researchgate.net/figure/Workflow-of-the-radiomics-analysis-1-Acquisition-of-high-quality-magnetic-resonance\\_fig1\\_334606028](https://www.researchgate.net/figure/Workflow-of-the-radiomics-analysis-1-Acquisition-of-high-quality-magnetic-resonance_fig1_334606028)
  32. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 1º de novembro de 2017;77(21):e104–7.
  33. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology.* maio de 2020;295(2):328–38.
  34. Dinapoli N, Casà C, Barbaro B, Chiloiro GV, Damiani A, Di Matteo M, et al. Radiomics for rectal cancer. *Transl Cancer Res.* agosto de 2016;5(4):424–31.
  35. Weenen E van. *evanweenen/u-net* [Internet]. 2025 [citado 26 de outubro de 2025]. Disponível em: <https://github.com/evanweenen/u-net>
  36. Jiang X, Zhang R, Nie S. Image Segmentation Based on Level Set Method. *Phys Procedia.* 1º de janeiro de 2012;33:840–5.
  37. Raju ASN, Venkatesh K, Gatla RK, Konakalla EP, Eid MM, Titova N, et al. Colorectal cancer detection with enhanced precision using a hybrid supervised and unsupervised learning approach. *Sci Rep.* 25 de janeiro de 2025;15(1):3180.
  38. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* dezembro de 2017;14(12):749–62.
  39. Horvat N, Veeraraghavan H, Pelossof RA, Fernandes MC, Arora A, Khan M, et al. Radiogenomics of Rectal Adenocarcinoma in the Era of Precision Medicine: A Pilot Study of Associations between Qualitative and Quantitative MRI Imaging Features and Genetic Mutations. *Eur J Radiol.* abril de 2019;113:174–81.
  40. Caruso D, Polici M, Zerunian M, Del Gaudio A, Parri E, Giallorenzi MA, et al. Radiomic Cancer Hallmarks to Identify High-Risk Patients in Non-Metastatic Colon Cancer. *Cancers.* 15 de julho de 2022;14(14):3438.
  41. Yang L, Dong D, Fang M, Zhu Y, Zang Y, Liu Z, et al. Can CT-based radiomics signature predict KRAS/NRAS/BRAF mutations in colorectal cancer? *Eur Radiol.* maio de 2018;28(5):2058–67.

42. Xue T, Peng H, Chen Q, Li M, Duan S, Feng F. Preoperative prediction of KRAS mutation status in colorectal cancer using a CT-based radiomics nomogram. *Br J Radiol.* 1º de junho de 2022;95(1134):20211014.
43. Negreros-Osuna AA, Parakh A, Corcoran RB, Pourvaziri A, Kambadakone A, Ryan DP, et al. Radiomics Texture Features in Advanced Colorectal Cancer: Correlation with BRAF Mutation and 5-year Overall Survival. *Radiol Imaging Cancer.* setembro de 2020;2(5):e190084.
44. Porto-Álvarez J, Cernadas E, Aldaz Martínez R, Fernández-Delgado M, Huelga Zapico E, González-Castro V, et al. CT-Based Radiomics to Predict KRAS Mutation in CRC Patients Using a Machine Learning Algorithm: A Retrospective Study. *Biomedicines.* 29 de julho de 2023;11(8):2144.
45. Liao Z, Luo D, Tang X, Huang F, Zhang X. MRI-based radiomics for predicting pathological complete response after neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a systematic review and meta-analysis. *Front Oncol.* 10 de março de 2025;15:1550838.
46. Badic B, Hatt M, Durand S, Jossic-Corcus CL, Simon B, Visvikis D, et al. Radiogenomics-based cancer prognosis in colorectal cancer. *Sci Rep.* 5 de julho de 2019;9(1):9743.
47. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: bridging imaging and genomics. *Abdom Radiol N Y.* junho de 2019;44(6):1960–84.
48. Taguchi N, Oda S, Yokota Y, Yamamura S, Imuta M, Tsuchigame T, et al. CT texture analysis for the prediction of KRAS mutation status in colorectal cancer via a machine learning approach. *Eur J Radiol.* setembro de 2019;118:38–43.
49. Wang Y, Ma LY, Yin XP, Gao BL. Radiomics and Radiogenomics in Evaluation of Colorectal Cancer Liver Metastasis. *Front Oncol.* 2021;11:689509.
50. Breiman L. Random Forests. *Mach Learn.* 1º de outubro de 2001;45(1):5–32.
51. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat.* outubro de 2001;29(5):1189–232.
52. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer Oxf Engl 1990.* março de 2012;48(4):441–6.
53. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 3 de junho de 2014;5(1):4006.
54. Shi R, Chen W, Yang B, Qu J, Cheng Y, Zhu Z, et al. Prediction of KRAS, NRAS and BRAF status in colorectal cancer patients with liver metastasis using a deep artificial neural network based on radiomics and semantic features. *Am J Cancer Res.* 1º de dezembro de 2020;10(12):4513–26.
55. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res.* 1º de junho de 2002;16:321–57.
56. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. Em: *Advances in Neural Information Processing Systems [Internet].* Curran Associates, Inc.;

- 2017 [citado 26 de outubro de 2025]. Disponível em:  
[https://papers.nips.cc/paper\\_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html](https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html)
57. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. fevereiro de 2016;278(2):563–77.
  58. Jia LL, Zhao JX, Zhao LP, Tian JH, Huang G. Current status and quality of radiomic studies for predicting KRAS mutations in colorectal cancer patients: A systematic review and meta-analysis. *Eur J Radiol*. janeiro de 2023;158:110640.
  59. Ma L, Guo H, Zhao Y, Liu Z, Wang C, Bu J, et al. Liquid biopsy in cancer: current status, challenges and future prospects. *Signal Transduct Target Ther*. 2 de dezembro de 2024;9(1):336.
  60. Di Meo A, Bartlett J, Cheng Y, Pasic MD, Yousef GM. Liquid biopsy: a step forward towards precision medicine in urologic malignancies. *Mol Cancer*. 14 de abril de 2017;16(1):80.
  61. Patelli G, Lazzari L, Crisafulli G, Sartore-Bianchi A, Bardelli A, Siena S, et al. Clinical utility and future perspectives of liquid biopsy in colorectal cancer. *Commun Med*. 24 de abril de 2025;5(1):137.
  62. Palacín-Aliana I, García-Romero N, Asensi-Puig A, Carrión-Navarro J, González-Rumayor V, Ayuso-Sacido Á. Clinical Utility of Liquid Biopsy-Based Actionable Mutations Detected via ddPCR. *Biomedicines*. 28 de julho de 2021;9(8):906.
  63. Nagayama S, Low SK, Kiyotani K, Nakamura Y. Precision Medicine for Colorectal Cancer with Liquid Biopsy and Immunotherapy. *Cancers*. 25 de setembro de 2021;13(19):4803.
  64. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*. 19 de fevereiro de 2014;6(224):224ra24.
  65. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med*. setembro de 2008;14(9):985–90.
  66. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci Transl Med*. 6 de julho de 2016;8(346):346ra92.
  67. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 23 de fevereiro de 2018;359(6378):926–30.
  68. Reinert T, Henriksen TV, Christensen E, Sharma S, Salari R, Sethi H, et al. Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. *JAMA Oncol*. 1º de agosto de 2019;5(8):1124–31.
  69. Stürzlinger H, Conrads-Frank A, Eisenmann A, Invansits S, Jahn B, Janzic A, et al. Stool DNA testing for early detection of colorectal cancer: systematic review using the HTA Core Model® for Rapid Relative Effectiveness Assessment. *GMS Ger Med Sci*. 23 de junho de 2023;21:Doc06.
  70. Dolatkah R, Dastgiri S, Jafarabadi MA, Abdolahi HM, Somi MH. Diagnostic accuracy of

- multitarget stool DNA testing for colorectal cancer screening: A systematic review and meta-analysis. *Gastroenterol Hepatol*. dezembro de 2022;45(10):753–66.
71. Gupta S. Screening for colorectal cancer. *Hematol Oncol Clin North Am*. junho de 2022;36(3):393–414.
  72. Chen M, Zhao H. Next-generation sequencing in liquid biopsy: cancer screening and early detection. *Hum Genomics*. 1º de agosto de 2019;13(1):34.
  73. Chen G, Peng J, Xiao Q, Wu HX, Wu X, Wang F, et al. Postoperative circulating tumor DNA as markers of recurrence risk in stages II to III colorectal cancer. *J Hematol Oncol*. *Hematol Oncol*. 17 de maio de 2021;14(1):80.
  74. Kim S, Lim Y, Kang JK, Kim HP, Roh H, Kim SY, et al. Dynamic changes in longitudinal circulating tumour DNA profile during metastatic colorectal cancer treatment. *Br J Cancer*. 1º de setembro de 2022;127(5):898–907.
  75. Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G, et al. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat Med*. julho de 2015;21(7):795–801.
  76. Nikanjam M, Kato S, Kurzrock R. Liquid biopsy: current technology and clinical applications. *J Hematol Oncol*. *Hematol Oncol*. 12 de setembro de 2022;15(1):131.
  77. Menon V, Brash DE. Next-Generation Sequencing Methodologies To Detect Low-Frequency Mutations: “Catch Me If You Can”. *Mutat Res Rev Mutat Res*. 2023;792:108471.
  78. Müller D, Györfy B. DNA methylation-based diagnostic, prognostic, and predictive biomarkers in colorectal cancer. *Biochim Biophys Acta BBA - Rev Cancer*. 1º de maio de 2022;1877(3):188722.
  79. Chen J, Sun H, Tang W, Zhou L, Xie X, Qu Z, et al. DNA methylation biomarkers in stool for early screening of colorectal cancer. *J Cancer*. 28 de agosto de 2019;10(21):5264–71.
  80. Vessies DCL, Greuter MJE, van Rooijen KL, Linders TC, Lanfermeijer M, Ramkisoensing KL, et al. Performance of four platforms for KRAS mutation detection in plasma cell-free DNA: ddPCR, Idylla, COBAS z480 and BEAMing. *Sci Rep*. 15 de maio de 2020;10(1):8122.
  81. Liao H, Li H. <p>Advances in the Detection Technologies and Clinical Applications of Circulating Tumor DNA in Metastatic Breast Cancer</p>. *Cancer Manag Res*. 18 de maio de 2020;12:3547–60.
  82. Kim H, Park KU. Clinical Circulating Tumor DNA Testing for Precision Oncology. *Cancer Res Treat Off J Korean Cancer Assoc*. abril de 2023;55(2):351–66.
  83. Cisneros-Villanueva M, Hidalgo-Pérez L, Rios-Romero M, Cedro-Tanda A, Ruiz-Villavicencio CA, Page K, et al. Cell-free DNA analysis in current cancer clinical trials: a review. *Br J Cancer*. fevereiro de 2022;126(3):391–400.
  84. Magbanua MJM, Li W, van 't Veer LJ. Integrating Imaging and Circulating Tumor DNA Features for Predicting Patient Outcomes. *Cancers*. janeiro de 2024;16(10):1879.
  85. Cherukuri SP, Kaur A, Goyal B, Kukunoor HR, Sahito AF, Sachdeva P, et al. Artificial

- Intelligence-Enhanced Liquid Biopsy and Radiomics in Early-Stage Lung Cancer Detection: A Precision Oncology Paradigm. *Cancers*. 29 de setembro de 2025;17(19):3165.
86. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature*. junho de 2019;570(7761):385–9.
  87. Shafiq-ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E. Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci Rep*. 12 de julho de 2018;8(1):10545.
  88. Zwanenburg A, Leger S, Vallières M, Löck S, Initiative for. Image biomarker standardisation initiative - feature definitions. 21 de dezembro de 2016;
  89. Ahsan MM, Mahmud MAP, Saha PK, Gupta KD, Siddique Z. Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies*. setembro de 2021;9(3):52.
  90. Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Sebastopol, CA: O'Reilly Media; 2017.
  91. Erwin Kreyszig. *Advanced Engineering Mathematics*. 4th Edition. New York: John Wiley & Sons; 1979.
  92. Medeiros TA, Saraiva Junior RG, Cassia G de S e, Nascimento FA de O, Carvalho JLA de. Classification of 1p/19q Status in Low-Grade Gliomas: Experiments with Radiomic Features and Ensemble-Based Machine Learning Methods. *Braz Arch Biol Technol*. 2023;66:e23230002.
  93. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *Lond Edinb Dublin Philos Mag J Sci*. 1º de novembro de 1901;2(11):559–72.
  94. Mehrabinezhad A, Teshnehlab M, Sharifi A. A comparative study to examine principal component analysis and kernel principal component analysis-based weighting layer for convolutional neural networks. *Comput Methods Biomech Biomed Eng Imaging Vis*. 31 de dezembro de 2024;12(1):2379526.
  95. He H, Garcia EA. Learning from Imbalanced Data. *IEEE Trans Knowl Data Eng*. setembro de 2009;21(9):1263–84.
  96. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 3 de junho de 2014;5(1):4006.
  97. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. Third edition. Hoboken, New Jersey: Wiley; 2013.
  98. Irina Rish. *An Empirical Study of the Naive Bayes Classifier*. Yorktown Heights, NY: IBM Thomas J. Watson Research Center; 2001 nov. Report No.: RC22230 (W0111-022).
  99. Quinlan JR. Induction of decision trees. *Mach Learn*. 1º de março de 1986;1(1):81–106.
  100. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 1º de abril de 2006;63(1):3–42.

101. Gardner MW, Dorling SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos Environ*. 1º de agosto de 1998;32(14):2627–36.
102. Freund Y, Schapire RE. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J Comput Syst Sci*. 1º de agosto de 1997;55(1):119–39.
103. Fisher RA. The Use of Multiple Measurements in Taxonomic Problems. *Ann Eugen*. 1936;7(2):179–88.
104. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Em: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]*. New York, NY, USA: Association for Computing Machinery; 2016 [citado 27 de outubro de 2025]. p. 785–94. (KDD '16). Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939785>
105. Mienye ID, Sun Y. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*. 1º de janeiro de 2022;10:99129–49.
106. Owusu-Adjei M, Ben Hayfron-Acquah J, Frimpong T, Abdul-Salaam G. Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems. *PLOS Digit Health*. 30 de novembro de 2023;2(11):e0000290.
107. Erickson BJ, Kitamura F. Magician’s Corner: 9. Performance Metrics for Machine Learning Models. *Radiol Artif Intell*. maio de 2021;3(3):e200126.
108. Mohd Haniff NS, Ng KH, Kamal I, Mohd Zain N, Abdul Karim MK. Systematic review and meta-analysis on the classification metrics of machine learning algorithm based radiomics in hepatocellular carcinoma diagnosis. *Heliyon*. 14 de agosto de 2024;10(16):e36313.
109. Tharwat A. Classification assessment methods. *Appl Comput Inform*. 30 de julho de 2020;17(1):168–92.
110. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 1: sensitivity and specificity. *BMJ*. 11 de junho de 1994;308(6943):1552.
111. Sasaki, Yutaka. *The truth of the F-measure*. Manchester, UK: University of Manchester, School of Computer Science; 2007.
112. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The Balanced Accuracy and Its Posterior Distribution. Em: *2010 20th International Conference on Pattern Recognition [Internet]*. 2010 [citado 27 de outubro de 2025]. p. 3121–4. Disponível em: <https://ieeexplore.ieee.org/document/5597285>
113. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. abril de 1982;143(1):29–36.
114. Wickenberg-Bolin U, Göransson H, Fryknäs M, Gustafsson MG, Isaksson A. Improved variance estimation of classification performance via reduction of bias caused by small sample size. *BMC Bioinformatics*. 13 de março de 2006;7(1):127.
115. Dasari A, Morris VK, Allegra CJ, Atreya C, Benson AB, Boland P, et al. ctDNA applications and integration in colorectal cancer: an NCI Colon and Rectal–Anal Task Forces whitepaper. *Nat Rev Clin Oncol*. dezembro de 2020;17(12):757–70.

116. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn JMD*. janeiro de 2017;19(1):4–23.
117. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 6 de junho de 2016;17(1):122.
118. Liu X, Yang Q, Zhang C, Sun J, He K, Xie Y, et al. Multiregional-Based Magnetic Resonance Imaging Radiomics Combined With Clinical Data Improves Efficacy in Predicting Lymph Node Metastasis of Rectal Cancer. *Front Oncol*. 2020;10:585767.
119. Yuan H, Xu X, Tu S, Chen B, Wei Y, Ma Y. The CT-based intratumoral and peritumoral machine learning radiomics analysis in predicting lymph node metastasis in rectal carcinoma. *BMC Gastroenterol*. 16 de novembro de 2022;22(1):463.
120. Eresen A, Li Y, Yang J, Shangguan J, Velichko Y, Yaghmai V, et al. Preoperative assessment of lymph node metastasis in Colon Cancer patients using machine learning: a pilot study. *Cancer Imaging Off Publ Int Cancer Imaging Soc*. 25 de abril de 2020;20(1):30.
121. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 20 de junho de 2016;34(18):2157–64.
122. Huang X, Cheng Z, Huang Y, Liang C, He L, Ma Z, et al. CT-based Radiomics Signature to Discriminate High-grade From Low-grade Colorectal Adenocarcinoma. *Acad Radiol*. outubro de 2018;25(10):1285–97.
123. Patanè V, Atripaldi U, Sansone M, Marinelli L, Del Tufo S, Arrichiello G, et al. MRI-based radiomics for preoperative T-staging of rectal cancer: a retrospective analysis. *Int J Colorectal Dis*. 2025;40(1):174.
124. Yang L, Liu D, Fang X, Wang Z, Xing Y, Ma L, et al. Rectal cancer: can T2WI histogram of the primary tumor help predict the existence of lymph node metastasis? *Eur Radiol*. dezembro de 2019;29(12):6469–76.
125. Lubner MG, Stabo N, Lubner SJ, del Rio AM, Song C, Halberg RB, et al. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. *Abdom Imaging*. outubro de 2015;40(7):2331–7.
126. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*. abril de 2017;17(4):223–38.
127. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*. abril de 2017;17(4):223–38.
128. Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med*. 7 de novembro de 2018;10(466):eaat4921.

129. Sánchez-Herrero E, Serna-Blasco R, Robado de Lope L, González-Rumayor V, Romero A, Provencio M. Circulating Tumor DNA as a Cancer Biomarker: An Overview of Biological Features and Factors That may Impact on ctDNA Analysis. *Front Oncol.* 2022;12:943253.
130. Masfarré L, Vidal J, Fernández-Rodríguez C, Montagut C. ctDNA to Guide Adjuvant Therapy in Localized Colorectal Cancer (CRC). *Cancers.* 8 de junho de 2021;13(12):2869.
131. Nygaard AD, Holdgaard PC, Spindler KLG, Pallisgaard N, Jakobsen A. The correlation between cell-free DNA and tumour burden was estimated by PET/CT in patients with advanced NSCLC. *Br J Cancer.* 21 de janeiro de 2014;110(2):363–8.
132. Winther-Larsen A, Demuth C, Fledelius J, Madsen AT, Hjorthaug K, Meldgaard P, et al. Correlation between circulating mutant DNA and metabolic tumour burden in advanced non-small cell lung cancer patients. *Br J Cancer.* agosto de 2017;117(5):704–9.