

# Uma Arquitetura Multimodal Profunda Para a Predição de Recorrência de Tumores em Volumes de Ressonância Magnética

Gabriel Souza<sup>1</sup>, Alexei M. C. Machado<sup>1,2</sup>

<sup>1</sup>Departamento de Ciência da Computação  
Pontifícia Universidade Católica de Minas Gerais (PUC Minas)  
30.535-901 – Belo Horizonte – MG – Brasil

<sup>2</sup>Departamento de Anatomia e Imagem – Universidade Federal de Minas Gerais  
Av Alfredo Balena 190 – 30130-100 – Belo Horizonte – MG – Brasil

[gabrielvbsouza17@gmail.com](mailto:gabrielvbsouza17@gmail.com), [alexeimcmachado@gmail.com](mailto:alexeimcmachado@gmail.com)

**Resumo.** *A predição de recorrência de tumores tratados por radioterapia é um problema difícil, principalmente em cenários onde se analisam conjuntos de dados desbalanceados. Este estudo apresenta uma solução baseada em redes profundas para o problema, que explora a integração multimodal de informações clínicas e demográficas com as imagens anatômicas e do mapa de aplicação de radiação, reunindo-as em um único volume de dois canais para ampliar o contexto oferecido ao modelo e aumentar sua capacidade de generalização. Experimentos conduzidos com o dataset Brain-TR-GammaKnife demonstraram a superioridade da abordagem multimodal integrada, aumentando de forma estatisticamente significativa os valores de acurácia, sensibilidade e especificidade na predição de recorrência de tumores cerebrais.*

**Abstract.** *Predicting recurrence of tumors treated with radiotherapy is a difficult problem, especially in scenarios where unbalanced datasets are analyzed. This study presents a deep network-based solution to the problem, which explores the multimodal integration of clinical and demographic information with anatomical images and radiation application maps, bringing them together in a single two-channel volume to broaden the context offered to the model and increase its generalization capacity. Experiments conducted with the Brain-TR-GammaKnife dataset demonstrated the superiority of the integrated multimodal approach, significantly increasing the accuracy, sensitivity, and specificity values for the prediction of brain tumor recurrence.*

## 1. Introdução

Um desafio importante na área de Inteligência Artificial (IA) é conseguir prever eventos futuros, analisando dados históricos. Em áreas como o diagnóstico médico assistido por computador, é extremamente valioso poder prever a evolução de uma condição ou a resposta de um paciente a um tratamento, utilizando apenas os dados clínicos e imagens disponíveis. Especificamente no tratamento de tumores cerebrais, poder identificar precocemente quais pacientes têm maior risco de reincidência é fundamental para personalizar o acompanhamento e ajustar estratégias de ação terapêutica. Entretanto, essa detecção é dificultada pela semelhança entre alterações naturais decorrentes do tratamento e novos focos tumorais [Tang and Teoh 2023, Mohana Ropa et al. 2024].

A previsão de quais pacientes ou lesões apresentam maior probabilidade de recorrência é um problema não resolvido na prática. Modelos tradicionais mostram baixa sensibilidade para identificar casos de recidiva, especialmente em dados desbalanceados. Um exemplo disso pode ser visto no estudo realizado por [Wang et al. 2023], onde uma rede neural convolucional (CNN) é usada para prever recorrências tumorais, alcançando uma acurácia de 0,90 e uma especificidade de 0,89, mas com uma sensibilidade de apenas 0,10. Essa baixa sensibilidade dificulta a identificação da recorrência, pois muitos casos positivos podem ser classificados como negativos.

Em vista disso, este trabalho objetiva propor uma solução multimodal baseada em aprendizado profundo para a predição de recorrência tumoral em pacientes submetidos à radiocirurgia. O estudo utiliza o dataset Brain-TR-GammaKnife, que combina imagens de MRI da Região de Interesse (ROI), volumes de dose de radiação aplicados e informações clínicas dos pacientes. Por meio de ajustes específicos na etapa de preparação e balanceamento dos dados, busca-se identificar padrões mais complexos e sutis associados à recorrência tumoral. Dessa maneira, objetiva-se a concepção de uma solução multimodal que integre, de forma espacialmente alinhada, as informações da ROI, o mapa da dose de radiação aplicada e dados clínicos do paciente, permitindo que o modelo identifique padrões mais relevantes para a tarefa de predição de recorrência tumoral.

## 2. Trabalhos Relacionados

Abordagens baseadas em aprendizado de máquina são amplamente utilizadas na classificação de tumores cerebrais. O estudo de [Jia and Chen 2020] propôs um sistema que integra segmentação, extração de características e classificação por Support Vector Machine (SVM) para automatizar a identificação de tumores. Foram aplicadas técnicas de pré-processamento para melhorar a qualidade das imagens e extrair atributos como energia, média, entropia e contraste. Entretanto, a extração manual de características é trabalhosa e dependente da experiência do especialista, o que limita a escalabilidade e a generalização desses modelos.

Técnicas de aprendizado profundo permitem a extração automática de características diretamente das imagens, eliminando a necessidade de extração manual. Redes neurais convolucionais (CNNs) têm apresentado desempenho superior aos métodos tradicionais de detecção e classificação de tumores cerebrais. Nesse contexto, [Tang and Teoh 2023] investigaram o uso da arquitetura ResNet18, demonstrando a viabilidade de adaptar redes pré-treinadas em ImageNet para essa tarefa. Avaliado em um dataset público, o modelo apresentou resultados promissores, superando outros modelos em especificidade (0,90) e precisão (0,89), indicando bom equilíbrio na identificação de casos positivos e negativos.

Mais recentemente, [Hossain et al. 2023] propuseram uma arquitetura multimodal CNN dual-encoder 3D que combina dados de MRI com informações clínicas, como o metabolismo da glicose tumoral ao longo do tempo. O método alcançou acurácia de 0,74 na diferenciação entre recorrência tumoral e efeitos do tratamento, evidenciando o potencial da integração de múltiplas modalidades de dados. Contudo, o estudo utilizou apenas 35 amostras, o que limita a generalização dos resultados.

O trabalho de [Wang et al. 2023] apresentou o *dataset* Brain-TR-GammaKnife e um *framework* de aprendizado profundo como linha de base para predição de recorrência.

Embora a proposta evidencie o potencial da análise quantitativa de imagens e dados clínicos para apoiar decisões médicas, os resultados iniciais foram limitados. Estudos recentes buscaram aprimorar essas técnicas. [Mohana Ropa et al. 2024] propuseram um *framework* híbrido baseado em CNNs para automatizar a detecção e classificação de tumores cerebrais. A combinação de técnicas de pré-processamento, como a normalização de intensidade e filtragem Gaussiana, com aprendizado automático de características proporcionou maior precisão e robustez, embora o problema da recorrência de tumores não fosse abordado claramente. De forma semelhante, [Oviya et al. 2024] apresentaram um modelo U-Net 3D aprimorado para segmentação tumoral, incorporando mecanismos de atenção e conexões residuais. Avaliado no desafio BraTS 2020, o modelo alcançou coeficiente de Dice e índice de Jaccard superiores a 0.8 para cada sub-região tumoral, evidenciando o potencial de arquiteturas avançadas para apoiar o planejamento terapêutico, o monitoramento da doença e a previsão de desfechos clínicos.

### 3. Materiais

Para este estudo, foi utilizado o *dataset* de domínio privado e restrito do The Cancer Imaging Archive, Brain-TR-GammaKnife de [Wang et al. 2023], uma coleção de dados desenvolvida para a predição da recorrência de tumores cerebrais após tratamento com radioterapia Gamma Knife. Esse conjunto de dados integra arquivos de MRI de 47 pacientes com metástases cerebrais através de imagens 3D de lesões obtidas após o tratamento, as quais são classificadas em estáveis, quando o tamanho se manteve o mesmo ou reduziu, ou recorrente, quando o tumor progrediu após o tratamento. É importante notar que, embora haja 47 pacientes únicos, alguns podem ter mais de um exame em anos diferentes, o que resulta na quantidade de 244 grupos de exames. Sendo assim, o *dataset* é composto por:

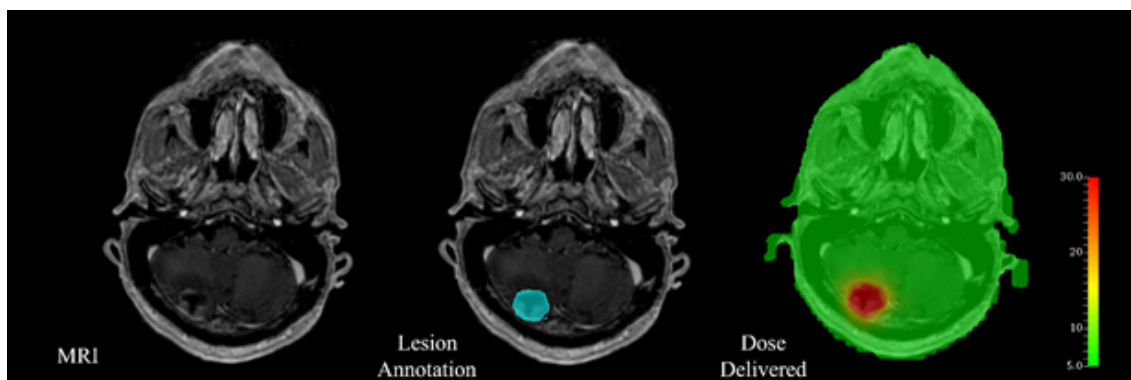
**Dados Clínicos e demográficos:** As informações clínicas dos pacientes contêm atributos relacionados tanto às características da lesão quanto ao histórico de tratamento e dados demográficos. Esses registros oferecem detalhes sobre localização anatômica, parâmetros de aplicação da radioterapia, diagnóstico inicial e condição da lesão observada na imagem, além de dados pessoais básicos.

**Imagens de Ressonância Magnética:** fornecidas com resoluções variando entre  $256 \times 256$  e  $512 \times 512$  pixels. Cada exame é composto por uma quantidade diferente de planos axiais (200 em média) e as imagens são em tons de cinza com 8 bits por pixel. Além disso, há uma imagem 3D correspondente com os recortes das Regiões de Interesse (ROIs) identificadas, classificadas, recortadas e rotuladas pela região da lesão para cada um dos pacientes.

**Imagens da Aplicação da Radioterapia:** para cada exame de MRI, está disponível uma imagem 3D que representa a quantidade de dose da radioterapia Gamma Knife aplicada em cada região específica do tumor. Essa imagem possui as mesmas dimensões da ROI correspondente, podendo ser sobreposta para melhor análise.

A Fig. 1 apresenta um exemplo de exame do *dataset* de [Wang et al. 2023], mostrando uma imagem de ressonância magnética de um paciente, a região de interesse destacada e o respectivo mapa de dose de radiação aplicada. O conjunto original apresenta um extremo desbalanceamento, com 221 amostras da classe estável e 23 da classe recorrente. A classe minoritária corresponde a apenas 14 pacientes únicos, sendo que as demais

amostras recorrentes são registros de exames repetidos, realizados em anos diferentes ou para distintas manifestações tumorais no mesmo indivíduo. O artigo original que apresenta o *dataset* menciona a aplicação de rotações nas imagens como forma de buscar o balanceamento, mas não descreve claramente os parâmetros ou a nova proporção obtida.



**Figura 1. Exemplo de imagem do dataset**

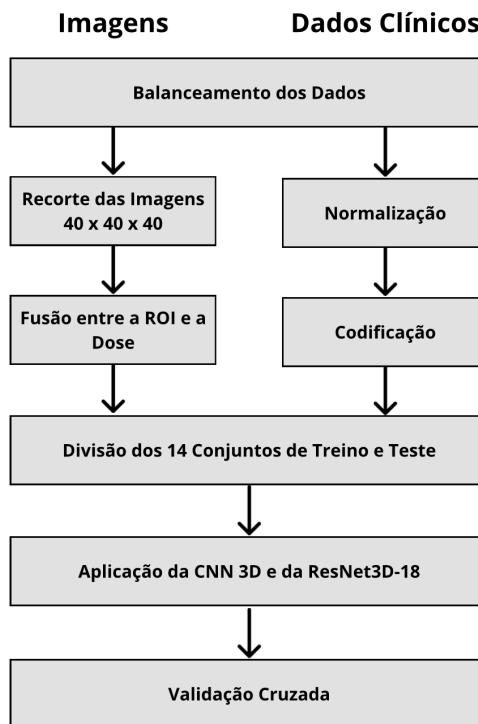
Para replicar de forma mais fiel à abordagem original, foram realizadas rotações nos três planos das imagens da ROI e nas respectivas imagens de dose de radiação nos ângulos de 45°, 90°, 135°, 180°, 225°, 270° e 315° graus. Cada rotação gerou uma nova amostra volumétrica correspondente. Após esse processo, o número de amostras recorrentes passou de 23 para 184, enquanto a classe estável manteve suas 221 amostras originais, reduzindo significativamente o desbalanceamento inicial entre as classes.

#### **4. Métodos**

A solução proposta neste estudo baseia-se na ResNet3D-18 [Raj et al. 2025], uma variação tridimensional da ResNet, adaptada para receber volumes como entrada. Essa versão possui 18 camadas de convolução 3D organizadas em blocos residuais, capazes de capturar padrões espaciais, preservando informações volumétricas. O modelo é treinado com um grande conjunto de exames médicos volumétricos, como MRI, CT (Tomografia Computadorizada) e PET (Tomografia por Emissão de Pósitrons) [Tang and Teoh 2023]. O pré-treinamento fornece ao modelo conhecimento prévio sobre estruturas anatômicas e padrões patológicos, acelerando o aprendizado e melhorando a capacidade de generalização, mesmo sob limitação de amostras.

Foi necessário realizar etapas de processamento dos dados para adequá-los como entrada nos modelos utilizados neste estudo: balanceamento entre classes, recorte e padronização das imagens, codificação e normalização dos dados clínicos para a aplicação de validação cruzada. As etapas são descritas na Fig. 2.

Com o objetivo de otimizar o processamento e reduzir a quantidade de informação irrelevante, foi implementado um procedimento de recorte cúbico com tamanho de  $40 \times 40 \times 40$  voxels centrado na área tumoral, através da identificação dos voxels mais claros (relacionados à lesão), tanto na imagem anatômica quanto na dose de radiação. O tamanho do recorte possibilitou manter uma compatibilidade com o procedimento adotado no artigo original, favorecendo a comparação direta entre os modelos.



**Figura 2. Etapas do Método.**

Para permitir que o modelo explorasse simultaneamente as informações anatômicas da lesão e os dados referentes à aplicação da radioterapia, foi realizada a fusão entre a imagem volumétrica da ROI e a imagem correspondente da dose de radiação. Cada volume tridimensional da ROI e seu respectivo volume da dose foram empilhados como canais distintos, formando um único tensor 4D no formato (40, 40, 40, 2), no qual o primeiro canal armazena os valores de intensidade da ROI e o segundo canal armazena os valores da dose de radiação aplicada naquela região. Essa organização fornece ao modelo informações complementares em uma mesma entrada, permitindo identificar padrões resultantes da interação entre a estrutura anatômica do tumor e o tratamento recebido.

Apesar de a fusão constituir parte essencial da contribuição deste trabalho, todas as imagens originais com ROI e dose separadas foram preservadas. Isso foi necessário para replicar os procedimentos descritos no artigo original, que utilizava apenas uma modalidade por vez como entrada da rede CNN 3D. Dessa forma, foi possível comparar diretamente o desempenho das arquiteturas nos cenários unimodais e multimodais, quantificando o impacto real da integração das modalidades na tarefa de predição de recorrência tumoral, o que representa um avanço no estudo proposto originalmente por [Wang et al. 2023]. Os dados clínicos e demográficos disponibilizados no *dataset* incluem idade, sexo, diagnóstico, tempo de tratamento, características da lesão e parâmetros da radioterapia.

A avaliação dos modelos foi realizada por meio de validação cruzada de 14 dobras, organizada por paciente para evitar que exames de um mesmo indivíduo aparecessem simultaneamente nos conjuntos de treino e teste. Ao final de cada dobra, foram calculadas as métricas de acurácia, sensibilidade e especificidade, bem como seus intervalos de

confiança, possibilitando a aplicação de testes de significância estatística (Student's t).

## 5. Experimentos

O procedimento experimental foi planejado para permitir a comparação entre diferentes arquiteturas de redes neurais tridimensionais e estratégias de integração das modalidades de dados disponíveis no *dataset*, sempre incorporando também as variáveis clínicas no processo de classificação. Dessa forma, a saída dos módulos convolucionais de cada arquitetura, constituindo o espaço latente extraído, foi concatenada ao vetor de dados clínicos e demográficos processado por um ramo denso, formando a entrada final para a camada de classificação binária (recorrente ou estável).

Foram definidos dois cenários para avaliação em todas as arquiteturas: (1) Modalidades Separadas, onde a ROI e a dose foram processadas individualmente como canais executados e analisados separadamente e depois concatenadas com os dados clínicos; e (2) Modalidades Concatenadas: onde a ROI e a dose foram combinadas em um único volume de dois canais, permitindo a extração conjunta de padrões desde as primeiras camadas convolucionais e depois concatenadas com os dados clínicos.

### 5.1. O modelo de CNN 3D

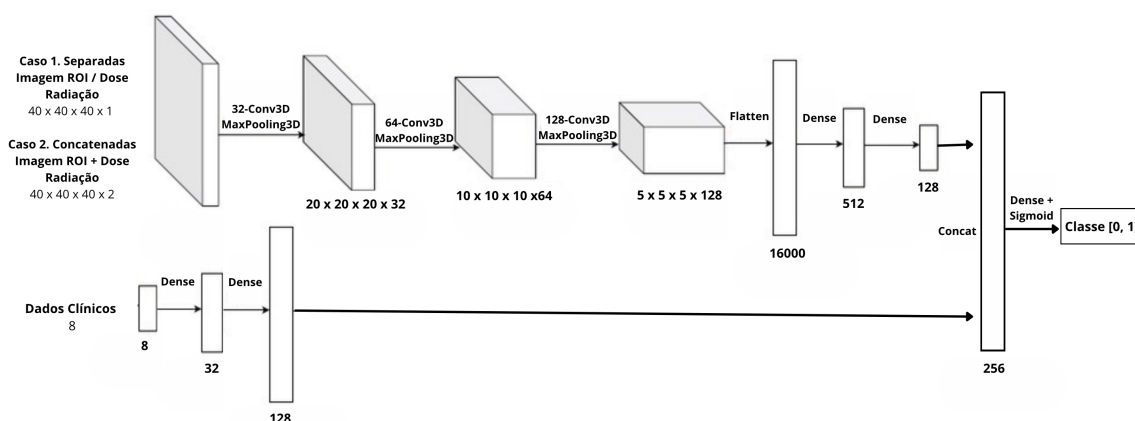
A primeira arquitetura testada, representada na Fig. 3 foi uma CNN 3D baseada na estrutura disponibilizada no repositório oficial do *dataset* [Sio1 2023], com adaptações para o formato de entrada utilizado neste trabalho. A rede é composta por três blocos convolucionais sequenciais com 32, 64 e 128 filtros, cada um com kernel  $3 \times 3 \times 3$  e ativação ReLU, seguido de *max pooling* e *batch normalization*. Ao final das convoluções, aplica-se *global average pooling*, seguido por uma camada densa de 512 unidades com ativação ReLU e *dropout* de 0,3, gerando um vetor visual de 128 unidades. Vale ressaltar que foram utilizadas 20 épocas e o otimizador Adam com taxa de aprendizado inicial de  $10^{-4}$ , tal como no código do artigo original [Wang et al. 2023].

Paralelamente, o vetor clínico foi processado por um perceptron multicamadas (MLP) composto por uma camada de 32 unidades com ativação ReLU e outra de 128 unidades com ativação ReLU. As saídas do módulo visual e do módulo clínico foram concatenadas e passadas para uma camada final de uma unidade com ativação sigmoide, produzindo a probabilidade de recorrência.

Nos experimentos do cenário 1, a rede processou as ROIs e doses em entradas separadas. No cenário 2, as duas modalidades foram integradas em um único volume com dois canais, fornecido à entrada da rede no lugar dos ramos separados. O restante da arquitetura e do processamento dos dados clínicos permaneceu idêntico.

### 5.2. O modelo de ResNet3D-18 pré-treinada pela MedicalNet

Os mesmos dois cenários também foram avaliados com o modelo da ResNet3D-18 disponibilizada pelo projeto *MedicalNet* [Chen et al. 2019], um repositório público mantido pela Tencent voltado ao processamento de imagens volumétricas médicas. No contexto da MedicalNet, o pré-treinamento consiste em expor a rede a um grande número de volumes médicos completos, obtidos de bases públicas de imagens clínicas, para que ela aprenda a extrair características espaciais e padrões visuais comuns a estruturas anatômicas e a variações patológicas.



**Figura 3. Arquitetura da CNN 3D proposta em [Wang et al. 2023]**

Ao aproveitar uma ResNet3D-18 pré-treinada como *backbone*, a solução proposta neste trabalho aproveitou pesos já ajustados para reconhecer formas anatômicas e padrões tridimensionais complexos, típicos de exames reais de pacientes. Isso permitiu que, mesmo sendo adaptado para a nova tarefa de predição de recorrência tumoral a partir da ROI, dose e dados clínicos, a rede precisasse treinar apenas as camadas finais, reduzindo o tempo de treinamento, mitigando o risco de sobreajuste e aumentando a capacidade de generalização a partir de um conjunto de dados limitado. Dessa forma, as camadas iniciais foram congeladas e apenas os parâmetros do último bloco convolucional (`layer4`) e da camada totalmente conectada (`fc`) foram liberados para atualização, configurando um *fine-tuning*. A camada final original foi substituída por:

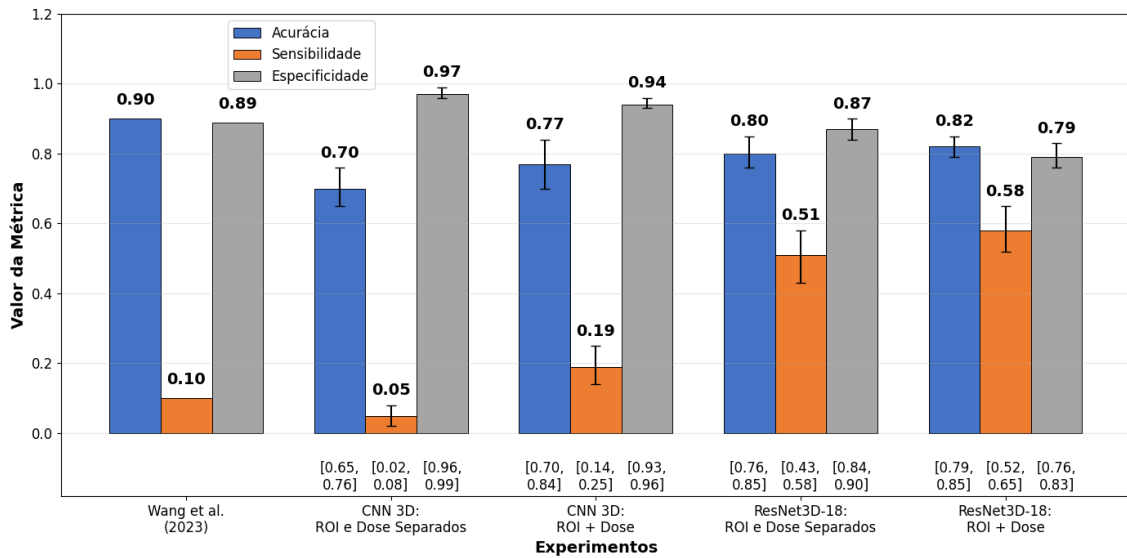
- Uma camada linear de 512 unidades com ativação ReLU;
- *Dropout* de 0.5 para reduzir sobreajuste;
- Saída final de características visuais com 512 unidades.

Em paralelo, o vetor clínico foi processado por um MLP com uma camada de 32 unidades e outra de 128 unidades, ambas com ativação ReLU. As saídas da ResNet (512 unidades) e do MLP clínico (128 unidades) foram concatenadas para formar um vetor de 640 características, passado por uma camada linear de uma unidade com ativação sigmoide para classificação binária.

O treinamento utilizou o otimizador Adam com taxa de aprendizado inicial de  $10^{-4}$  e função de perda por Entropia Cruzada. Além disso, para uma comparação justa com a CNN 3D, foram utilizadas também 20 épocas no treinamento. Seguindo o protocolo de validação cruzada por paciente descrito na etapa da Metodologia, cada iteração avaliou todo o conjunto de exames de um paciente recorrente no teste, acompanhado por pacientes estáveis com número equivalente de amostras, assegurando equilíbrio e evitando *data leakage*. Em ambos os cenários, os vetores visuais extraídos foram unidos às representações clínicas antes da classificação final, permitindo ao modelo explorar relações combinadas entre anatomia tumoral, distribuição da dose e características clínicas.

## 6. Resultados

Os resultados obtidos apresentam uma evolução consistente do desempenho dos modelos à medida que são aplicadas estratégias de integração multimodal e arquiteturas mais



**Figura 4. Métricas de acurácia (azul), sensibilidade (laranja) e especificidade (cinza) obtidas pelo trabalho original de Wang et al, e pelos 4 cenários experimentais propostos, com respectivos intervalos de confiança.**

sofisticadas. Para avaliar não apenas os valores médios, mas também a estabilidade das métricas ao longo das 14 dobras da validação cruzada por paciente, foram calculados intervalos de confiança de 0,95 para acurácia, sensibilidade e especificidade em cada cenário. Esses intervalos refletem a variabilidade entre diferentes partições do conjunto de dados e fornecem uma medida da confiabilidade das estimativas obtidas.

Na Fig. 4, observa-se que o modelo do trabalho original de [Wang et al. 2023] atingiu uma acurácia elevada de 0,90 e especificidade de 0,89, porém com uma sensibilidade extremamente baixa de 0,10, indicando que a maioria dos casos de recorrência não foi identificada. Esse desempenho é típico de cenários com forte desbalanceamento de classes, nos quais o modelo tende a privilegiar a classe majoritária para maximizar a acurácia.

Nos experimentos realizados com a CNN 3D simples no primeiro cenário, com ROI e dose processadas separadamente, a acurácia foi de 0,70 (IC [0,65, 0,76]), sensibilidade de 0,05 (IC [0,02, 0,08]) e especificidade de 0,97 (IC [0,96, 0,99]). O intervalo extremamente estreito da sensibilidade, centrado em valores muito baixos, indica que o modelo é consistentemente incapaz de detectar recorrências, ainda que a especificidade seja alta e estável. A acurácia apresenta intervalo relativamente estreito, mas com valor baixo em comparação ao artigo original. Vale ressaltar que esse experimento foi uma tentativa de replicação do experimento original de [Wang et al. 2023], contudo os resultados inferiores sugerem que o pré-processamento realizado pelos autores diferiu significativamente do empregado neste trabalho. Tal questão não pôde ser contornada, uma vez que o código original do artigo não foi disponibilizado.

Ao se integrarem as modalidades ROI e dose em uma mesma entrada no segundo cenário da CNN 3D, a sensibilidade aumentou para 0,19 (IC [0,14, 0,25]), acompanhada de pequena redução na especificidade para 0,94 (IC [0,93, 0,96]) e aumento da acurácia para 0,77 (IC [0,70, 0,84]). O intervalo da sensibilidade permanece estreito, mas centrado

em um valor quase quatro vezes superior ao cenário anterior, demonstrando um ganho consistente e estável proporcionado pela abordagem multimodal. O intervalo ligeiramente mais largo da acurácia reflete maior variação entre as dobras, decorrente de mudanças na capacidade de detectar recorrências em diferentes pacientes, mas sem comprometer a interpretação geral de melhoria no desempenho.

A utilização da arquitetura ResNet3D-18 pré-treinada pela MedicalNet trouxe ganhos substanciais em todas as métricas. No primeiro cenário com a ResNet3D-18, com modalidades separadas, a acurácia foi de 0,80 (IC [0,76, 0,85]), sensibilidade de 0,51 (IC [0,43, 0,58]) e especificidade de 0,87 (IC [0,84, 0,90]). O aumento expressivo na sensibilidade, sendo mais de dez vezes superior ao observado na CNN 3D, pode ser atribuído ao pré-treinamento em grandes bases de imagens volumétricas médicas (MRI, CT, PET), que permitiu à rede extrair características espaciais e texturais mais complexas, mesmo de modalidades tratadas separadamente. O intervalo de confiança da sensibilidade, embora mais largo que nos cenários com CNN, permanece em valores consistentemente acima de 0,40, demonstrando que, apesar de alguma variabilidade entre pacientes, o modelo mantém capacidade de detectar recorrências. A queda da especificidade para 0,87, acompanhada de um intervalo relativamente estreito, reflete o aumento de falsos positivos decorrente da maior detecção de casos verdadeiros, *trade-off* esperado e aceitável quando se prioriza a sensibilidade.

Além disso, vale pontuar que parte dessa melhoria pode estar associada à própria natureza arquitetural das redes residuais. O uso de *skip connections* nas ResNets facilita o fluxo de informação e de gradiente ao longo das camadas, reduzindo o problema do *vanishing gradient* em redes mais profundas. Isso permite que a rede retenha informações relevantes de níveis iniciais de processamento e as combine com representações mais abstratas geradas em camadas posteriores, resultando em descritores mais ricos e discriminativos. No contexto da predição de recorrência tumoral, essa capacidade de preservar e combinar características em diferentes níveis pode ter contribuído significativamente para identificar padrões sutis, especialmente na classe minoritária, elevando a sensibilidade sem comprometer de forma significativa a acurácia geral.

O cenário multimodal com a ResNet3D-18 apresentou os melhores resultados gerais: acurácia de 0,82 (IC [0,79, 0,85]), sensibilidade de 0,58 (IC [0,52, 0,65]) e especificidade de 0,79 (IC [0,76, 0,83]). A combinação de modelo com capacidade de representar padrões volumétricos e a fusão das modalidades desde a entrada potencializou a identificação de recorrentes. Os intervalos de sensibilidade e acurácia são estreitos e centrados em valores elevados, demonstrando estabilidade e consistência do modelo, mesmo com o aumento de casos de recorrência tumoral nos conjuntos de teste. A especificidade, embora ligeiramente inferior aos cenários com CNN, permanece em nível aceitável, refletindo um equilíbrio adequado entre detectar verdadeiros positivos e evitar falsos positivos. Mesmo a sensibilidade não atingindo valores tão elevados quanto a especificidade, ela representa um grande avanço em comparação ao modelo original, demonstrando que a integração multimodal associada à arquitetura pré-treinada é eficaz.

É importante destacar que a diminuição das acurácias em comparação ao artigo original está diretamente relacionada ao protocolo de validação empregado. Neste trabalho, a validação cruzada por paciente foi configurada para aumentar a proporção de casos da classe minoritária nos conjuntos de teste, elevando a dificuldade da tarefa e reduzindo

o valor absoluto da acurácia. Assim, a acurácia aqui obtida é uma métrica mais realista e equilibrada, evitando o viés da predominância da classe majoritária observado no trabalho original.

A análise conjunta dos valores médios e dos intervalos de confiança reforça três padrões importantes:

- Especificidades altas com ICs estreitos indicam que todos os modelos são capazes de identificar consistentemente a classe majoritária;
- Sensibilidades baixas com ICs estreitos nos cenários com CNN mostram consistência na falha de detectar recorrências, evidenciando limitações estruturais da arquitetura;
- Aumento da sensibilidade e da acurácia acompanhado de ICs estreitos nos cenários com ResNet3D-18 confirma que os ganhos são consistentes entre diferentes pacientes, especialmente quando a multimodalidade é empregada.

## 7. Conclusão

Este estudo apresentou uma solução multimodal para a predição de recorrência tumoral baseada em redes profundas que combinam imagens de ressonância magnética, mapas de dose de radioterapia, informações clínicas e demográficas. Os resultados obtidos demonstram que a utilização de dados multimodais, combinando de forma espacialmente alinhada às informações da ROI e do mapa de dose de radiação, fornece ao modelo um contexto adicional capaz de revelar padrões mais complexos e relevantes para a detecção de recorrências. A integração dessas duas modalidades contribuiu para um aumento expressivo da sensibilidade com impacto mínimo sobre a especificidade, evidenciando o potencial da abordagem para melhorar a detecção da classe minoritária. No melhor cenário avaliado, a sensibilidade atingiu 0,58, representando um aumento de aproximadamente 5,8 vezes em relação ao estudo original que apresentou o *dataset*, com acurácia de 0,82 e especificidade de 0,79, demonstrando um equilíbrio adequado entre as métricas.

A adoção de *backbones* pré-treinados em grandes bases médicas volumétricas, como a ResNet3D-18, revelou vantagens substanciais na capacidade de generalização e no reconhecimento de padrões sutis associados à recorrência. A solução proposta apresentou desempenho superior à CNN 3D, principalmente na sensibilidade, o que pode ser atribuído à combinação de convoluções tridimensionais, conexões residuais e pesos ajustados previamente em imagens médicas reais. A análise dos intervalos de confiança confirmou que os ganhos observados são consistentes e robustos entre diferentes partições de pacientes.

Também foi observada a importância de protocolos de avaliação que promovam equilíbrio na distribuição das classes nos conjuntos de teste. A implementação da validação cruzada por paciente, incluindo proporcionalmente mais casos recorrentes, gerou métricas mais representativas da performance real dos modelos, evitando que a acurácia fosse inflada pela predominância da classe estável. Esse tipo de avaliação garante uma análise mais justa e rigorosa, especialmente em cenários com forte desbalanceamento de dados.

Adicionalmente, os resultados demonstram que a inclusão de informações clínicas do paciente, como histórico médico, idade, tipo de diagnóstico e características do tratamento, complementa os dados visuais e amplia significativamente o escopo da análise

da rede. Esse conjunto de variáveis tabuladas fornece contexto adicional que contribui para que o modelo compreenda melhor o quadro clínico e identifique padrões que não seriam observáveis apenas nas imagens. A combinação de dados clínicos com dados volumétricos se revela, assim, um importante diferencial na tarefa de predição de recorrência, aumentando o potencial da abordagem para aplicações reais em ambientes clínicos.

Em síntese, a integração multimodal, o uso de arquiteturas pré-treinadas, a aplicação de métodos de avaliação equilibrados e a utilização de variáveis clínicas demonstraram-se estratégias promissoras para o desafio da predição de recorrência tumoral em imagens médicas tridimensionais. Embora desenvolvidas e validadas neste domínio, essas práticas possuem caráter geral e podem ser adaptadas para outros cenários que envolvam dados volumétricos, classes desbalanceadas e múltiplas fontes de informação. Limitações relacionadas ao tamanho relativamente reduzido do *dataset* e à ausência de validação externa indicam a necessidade de avaliações futuras em conjuntos de dados maiores e mais diversificados, bem como testes em outras tarefas de classificação para confirmar a transferibilidade das técnicas propostas.

## 8. Trabalhos Futuros

Diversas direções podem ser exploradas para ampliar e aprofundar os resultados obtidos neste estudo. Um primeiro aspecto é o aprimoramento da sensibilidade dos modelos, que, embora tenha apresentado um crescimento expressivo em relação ao trabalho original, ainda pode ser elevada por meio de estratégias adicionais de aumento de dados sintéticos e técnicas avançadas de aumento de dados específicas para imagens volumétricas. Métodos como transformações elásticas, simulações de ruído típico de MRI e variações controladas de intensidade podem contribuir para reduzir o número de falsos negativos e aumentar a capacidade de detectar casos recorrentes, sobretudo em contextos de desbalanceamento extremo.

Paralelamente, a utilização de conjuntos de dados maiores e mais diversificados representa uma linha essencial de investigação. A coleta de novos volumes médicos ou a integração de diferentes bases relacionadas ao tema permitiria avaliar a robustez e a capacidade de generalização dos modelos diante de maior variabilidade de casos clínicos, protocolos de exame e equipamentos de aquisição. Isso seria particularmente valioso para validar a transferibilidade das arquiteturas e técnicas propostas para diferentes contextos geográficos e institucionais.

O estudo também pode ser expandido para abranger novas tarefas de classificação, explorando como as arquiteturas propostas se comportam na predição de outras doenças, como lesões degenerativas, acidentes vasculares ou doenças neurológicas progressivas. Tal abordagem ajudará a verificar a generalidade das técnicas empregadas e sua aplicabilidade em diferentes domínios clínicos e computacionais, reforçando a independência das soluções em relação ao contexto específico.

No campo arquitetural, é pertinente testar modelos mais profundos e amplamente reconhecidos, como a ResNet3D-50, que possui maior capacidade de extração de características devido ao número adicional de camadas, bem como avaliar outras arquiteturas como a DenseNet3D, EfficientNet3D e redes equipadas com módulos de atenção. Essas arquiteturas podem potencialmente melhorar o foco em regiões específicas da imagem,

otimizar a integração multimodal e elevar ainda mais a sensibilidade sem comprometer a especificidade.

Investigações futuras poderiam, ainda, considerar aprimoramentos nos protocolos de avaliação, como o *stratified cross-validation* ou validação em bases externas independentes, assegurando métricas ainda mais robustas e generalizáveis. Adicionalmente, seria importante explorar técnicas de explicabilidade, como mapas de ativação ou métodos de atribuição de características, permitindo que as decisões dos modelos sejam mais transparentes e interpretáveis em ambientes clínicos. Isso aumentaria a confiança dos profissionais de saúde no uso dessas ferramentas como suporte à decisão diagnóstica e terapêutica.

**Agradecimentos** — Alexei M. C. Machado agradece o apoio financeiro do CNPq e da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) projetos APQ-02753-24 e APQ-06556-24.

## Referências

- Chen, S., Ma, K., and Zheng, Y. (2019). Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:1904.00625*.
- Hossain, T., Qureshi, Z., Jayakumar, N., Muttikkal, T. E., Patel, S., Schiff, D., Zhang, M., and Kundu, B. (2023). Multimodal deep learning to differentiate tumor recurrence from treatment effect in human glioblastoma. In *IEEE Int Symp on Biomedical Imaging*, pages 1–4.
- Jia, Z. and Chen, D. (2020). Brain tumor identification and classification of mri images using deep learning techniques. *IEEE Access*, page 1.
- Mohana Ropa, Y., Moon, K. R., and Singh, T. (2024). Deep learning-based hybrid framework utilizing opencv and cnn for automated brain tumor detection and mri image classification. In *Int Conf on Sustainable Expert Systems*, pages 1674–82.
- Oviya, I. R., Vishnu, S., Anusha, J. S., Chatiyode, V., Priyanga, S., and Charan, K. K. S. (2024). Enhanced brain tumor segmentation using unet3d: A deep learning perspective. In *Int Conf on Computing Communication and Networking Technologies*, pages 1–7.
- Raj, J. R. F., Presskila, X. A., Krishnan, R. S., Sarala, R., Isaac, R. A., and Soundiraraj, N. (2025). 3d resnet-18 based deep convolutional network for automated stroke detection in volumetric ct scans. In *International Conference on Inventive Computing and Informatics (ICICI)*, pages 01–08.
- Siol, M. S. U. (2023). Brain mri model implementation. [https://github.com/siolmsstate/brain\\_mri](https://github.com/siolmsstate/brain_mri).
- Tang, M. C. S. and Teoh, S. S. (2023). Brain tumor detection from mri images based on resnet18. In *Int Conf on Information Systems and Computer Networks*, pages 1–5.
- Wang, Y., Duggar, W., Caballero, D., Thomas, T., Adari, N., Mundra, E., and Wang, H. (2023). A brain mri dataset and baseline evaluations for tumor recurrence prediction after gamma knife radiotherapy. *Sci Data*, 10:785.