

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
PRÓ-REITORIA DE PESQUISA, PÓS-GRADUAÇÃO E INOVAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Adriell Gomes Marques

**Nova Abordagem para Auxílio ao Diagnóstico Médico em Imagens de
Tumores Cerebrais Através de Fusão de Dados e Modelos de
Linguagem de Grande Escala**

Fortaleza – CE

30 de Outubro de 2025

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

M357n Marques, Adriell Gomes Marques.

Nova Abordagem para Auxílio ao Diagnóstico Médico em Imagens de Tumores Cerebrais Através de Fusão de Dados e Modelos de Linguagem de Grande Escala : Dissertação / Adriell Gomes Marques Marques. - 2025.

105 f. : il. color.

Dissertação (Mestrado) - Instituto Federal do Ceará, Mestrado em Ciência da Computação, Campus Fortaleza, 2025.

Orientação: Prof. Dr. Carlos Maurício Jaborandy de Mattos Dourado Júnior.

Coorientação: Prof. Dr. Luís Fabricio de Freitas Souza.

1. Detecção e Segmentação Profunda. 2. Imagem por Ressonância Magnética. 3. Células Computacionais Inteligentes. 4. Aprendizado Profundo. 5. Ajuste Fino. I. Título.

CDD 004

Adriell Gomes Marques

Nova Abordagem para Auxílio ao Diagnóstico Médico em Imagens de Tumores Cerebrais Através de Fusão de Dados e Modelos de Linguagem de Grande Escala

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Linha de Pesquisa: Processamento Digital de Sinais e Imagens

Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPI)

Programa de Pós-Graduação em Ciência da Computação (PPGCC)

Fortaleza – CE

30 de Outubro de 2025

Agradecimentos

Gostaria de agradecer à minha família, que sempre me deu bons conselhos, lições de vida, boas conversas e motivos para seguir em frente; especialmente à minha mãe, Alcilene Gomes dos Rêz, que sempre me auxiliou e é um exemplo de dedicação e humildade; Ao meu padrasto, Antônio Henrique de Oliveira Silva, que me preparou para a vida e para as dificuldades e alegrias que ela nos traz; E á minha avó, Antônia Maria Gomes dos Rêz, que sempre me deu ajuda, abrigo, cuidado, e buscou me dar bons conselhos em momentos difíceis.

Quero agradecer também à minha namorada e companheira de vida, Ana Karoline dos Santos Silva, que me suporta todos os dias, é minha rede de apoio, sempre me incentivou e me deu forças para enfrentar os desafios, mostrando que, apesar das adversidades, não há motivo para não buscar a felicidade e fazer rir até mesmo quando a vida nos empurra.

Agradeço ao meu orientador, Prof. Dr. Carlos Maurício Jaborandy de Mattos Dourado Júnior, por ter aceitado me orientar e por sua paciência, generosidade e apoio nos momentos mais desafiadores.

Também sou grato à banca examinadora pelo aceite do convite e pelo aprendizado que esse processo certamente me proporcionará.

Em suma, sou grato a todos que fizeram parte da minha trajetória, oferecendo vivências, experiências, conhecimentos, risadas, tristezas e todo o universo de emoções que a vida nos traz. Agradeço pelas oportunidades que agarrei e pela força de me reerguer e lutar pelo meu futuro a cada dia, pelos altos e baixos deste percurso e por ter chegado até aqui, com minhas cicatrizes e hematomas.

Se a vida não faz sentido, que busquemos o nosso, neste grande rolar de pedras sobre a colina...

*“Deixe tudo acontecer a você, Beleza e terror, Apenas continue... Nenhum sentimento é
final...”*

(Rainer Maria Rilke.)

Resumo

Os tumores cerebrais trazem alto impacto clínico e social, sobretudo devido às dificuldades associadas ao seu diagnóstico precoce. A complexidade do diagnóstico dos tumores decorre de fatores como a heterogeneidade intrínseca deste tipo de patologia, que pode variar em localização, extensão e padrões histológicos, o que dificulta a definição de critérios diagnósticos uniformes. Para contornar esses desafios, diversos métodos inovadores a partir dos exames de ressonância magnética e radiômica têm sido empregados. O método proposto neste estudo se utiliza de aprendizado profundo para detectar tumores cerebrais com o *framework* “*You only look once*” (YOLO), e um processo de ajuste fino no final da camada de rede usando células computacionais sintéticas capazes ajustar a região detectada, aprimorando assim, a segmentação das bordas do tumor cerebral. Além disso, o método utiliza um *pipeline* de classificação que combina um conjunto de classificadores e extratores aliados à busca de hiperparâmetros ótimos pelo método de *grid search*, para encontrar a melhor combinação e os melhores hiperparâmetros para o conjunto de dados. O método obteve resultados satisfatórios com altos índices de acurácia para detecção de regiões lesionadas pelo tumor, para segmentação de tumores cerebrais, e para a classificação de presença dos mesmos. O tempo de segmentação foi inferior a 1 segundo, superando o estado da arte e demonstrando o potencial do método proposto. A nova abordagem propõe o uso de duas bases de dados: A base de dados de segmentação adotada foi a TCGA/LGG; A base de dados de classificação adotada foi a BraTS2020, para classificar a presença de tumores em imagens de ressonância magnética, bem como o tempo de vida do paciente. As etapas de segmentação e classificação são validadas mediante comparação com a literatura, com trabalhos que utilizaram o mesmo conjunto de dados em mesmas métricas. Por fim, o método adota uma arquitetura de modelo de linguagem de grande escala (LLM) para tumores cerebrais capaz de gerar um pré-diagnóstico por meio dos dados clínicos classificados, e pode ser usada em sistemas para auxiliar o diagnóstico por imagem médica. Como contribuição, este estudo emprega novos modelos de detecção combinados com um método inovador baseado em processamento digital de imagens para aprimorar as métricas de segmentação, bem como o uso de Fusão de Dados, combinando as diferentes modalidades de ressonância magnética de tumores para aprimorar o desempenho da classificação. O estudo também utiliza modelos de LLM para refinar o pré-diagnóstico obtido após a classificação. Assim, este estudo desenvolveu um método de Diagnóstico Assistido por Computador (CAD) por meio de IA com PDI, CNN e LLM.

Palavras-chave: Detecção e Segmentação Profunda, Imagem por Ressonância Magnética, Células Computacionais Inteligentes, Aprendizado Profundo, Ajuste Fino

Abstract

Brain tumors have a significant clinical and social impact, particularly due to the challenges associated with early diagnosis. The complexity of tumor diagnosis stems from factors such as the intrinsic heterogeneity of this type of pathology, which can vary in location, extent, and histological patterns, making it difficult to define uniform diagnostic criteria. To overcome these challenges, several innovative methods based on magnetic resonance imaging and radiomics have been employed. The method proposed in this study leverages deep learning to detect brain tumors using the “You Only Look Once” (YOLO) framework, followed by a fine-tuning process at the end of the network layer employing synthetic computational cells capable of adjusting the detected region, thereby refining the segmentation of the brain tumor’s borders. In addition, the method uses a classification pipeline that combines an ensemble of classifiers and feature extractors with a grid search for optimal hyperparameter selection to find the best model combination and settings for the dataset. The method achieved satisfactory results with high accuracy rates for lesion detection, brain tumor segmentation, and tumor presence classification. Segmentation time was under one second, surpassing the state of the art and demonstrating the potential of the proposed approach. The new methodology proposes the use of two datasets: the TCGA/LGG dataset for segmentation, and BraTS2020 for classifying tumor presence in magnetic resonance images as well as patient survival time. The segmentation and classification stages are validated via comparison with the literature, using studies that employed the same datasets and metrics. Finally, the method adopts a large language model (LLM) architecture for brain tumors capable of generating a preliminary diagnosis based on the classified clinical data, and can be integrated into systems to assist medical image diagnosis. As a contribution, this study employs novel detection models combined with an innovative image processing-based approach to enhance segmentation metrics, as well as data fusion that integrates different MRI modalities of tumors to improve classification performance. The study also uses LLMs to refine the preliminary diagnosis obtained after classification. Thus, this work proposes an AI-based Computer-Aided Diagnosis (CAD) method combining image processing, CNNs, and LLMs.

Keywords: Deep Detection and Segmentation, Magnetic Resonance Imaging, Intelligent Computational Cells, Deep Learning, Fine-Tuning

Lista de ilustrações

Figura 1 – Amostras do Dataset TCGA.	34
Figura 2 – Amostras do conjunto de dados BraTs2020.	35
Figura 3 – Representação da Arquitetura do YOLOv8 de forma resumida.	37
Figura 4 – Demonstração de resultados pelos métodos de convex Hull e Concave Hull.	43
Figura 5 – Exemplo de arquitetura CNN, modelo LeNET.	52
Figura 6 – Arquitetura VGG	53
Figura 7 – Arquitetura DenseNet	54
Figura 8 – Arquitetura geral Llama3	57
Figura 9 – Representação das classes possíveis para os pixels de métrica de segmentação.	58
Figura 10 – Representação das classes possíveis para as predições, exemplo de classificação binária.	60
Figura 11 – Ilustração Geral da metodologia desenvolvida neste estudo, com Detecção baseada no YOLOv8, treinado sob o Dataset TCGA; Segmentação baseada no <i>Divisible Cell Segmentation</i> , Extração de atributos com arquiteturas CNN, Classificação com pipeline de combinação e de busca de hiperparâmetros ótimos, treinado com o dataset BraTS2020; Seguida pela construção de pré diagnóstico textual com LLM.	63
Figura 12 – Metodologia proposta, Etapa 1 - Detecção de tumor cerebral em imagem de ressonância magnética, Etapa 2 - Segmentação do tumor cerebral, Etapa 3 - Classificação dos atributos do tumor cerebral, Etapa 4 - Resultados e pré-diagnóstico formalizados por LLM.	65
Figura 13 – Resultados da inicialização automática para detecção de tumores cerebrais em imagens de ressonância magnética craniana. Em (A), as imagens originais são exibidas, enquanto em (B), as regiões detectadas podem ser observadas através da sobreposição das caixas delimitadoras.	77
Figura 14 – Apresenta o processo de segmentação de tumores cerebrais realizado pelo método <i>Divisible Cell Segmentation</i> . As imagens ilustram o movimento de células computacionais, dividindo-se para detectar as bordas do tumor cerebral conforme se movem cada uma em sua direção específica de acordo com o seu vetor de direção. A última imagem mostra o resultado da segmentação pelo método, que inclui conectar os pontos centrais das células na borda do tumor através do <i>Concave Hull</i> , formando a segmentação tumoral.	79

Figura 15 – Exibe o resultado da segmentação usando o método <i>Divisible Cell Segmentation</i> em diferentes imagens de ressonância magnética do crânio. A coluna (A) mostra as ressonâncias magnéticas como imagens de entrada, e a coluna (B) mostra os resultados da segmentação da região da borda da lesão pelo método proposto, como imagens de saída. . . .	80
Figura 16 – Resultado de cada etapa da metodologia proposta: Em (a) imagem em tons de cinza; (b) detecção; (c) segmentação; (d) classificação; (e) pré-diagnóstico com LLM.	87
Figura 17 – Ilustração gráfica da tabela de comparação de métricas 11.	88
Figura 18 – Comparação gráfica da precisão dos métodos propostos e dos métodos da literatura, na tabela 13.	90
Figura 19 – Comparação gráfica da precisão dos métodos propostos e dos métodos de última geração para a tabela 14.	92

Lista de tabelas

Tabela 1	– Resumo dos trabalhos discutidos na seção de Trabalhos Relacionados deste estudo. A tabela contém informações sobre os autores, arquitetura aplicada, base de dados utilizada, proposta de método, resultados e possíveis sugestões de melhoria.	33
Tabela 2	– Tabela de range de hiperparâmetros adotados no Grid Search para os classificadores no banco de dados Brats2020.	51
Tabela 3	– Resultados do YOLOv8 na detecção com máscara binária de tumores cerebrais. A tabela mostra os resultados do experimento com base na validação cruzada.	75
Tabela 4	– Resultado da segmentação do método <i>Divisible Cell segmentation</i> , inicializado com auxílio do YOLOv8.	78
Tabela 5	– Tempo de segmentação do método proposto, com <i>Divisible Cell Segmentation</i> aplicado sobre o conjunto de teste.	78
Tabela 6	– Resultados da classificação para as combinações de extratores e classificadores testados em relação ao tempo de sobrevivência, ou <i>Survival Days</i> , para o dataset Brats2020.	82
Tabela 7	– Resultados da classificação para a combinação de extratores e classificadores por idade, para o banco de dados Brats2020.	83
Tabela 8	– Resultados da classificação para a combinação de extratores e classificadores para classe de presença/ausência de tumores cerebrais na imagem, para o banco de dados Brats2020.	84
Tabela 9	– Resultados da classificação da melhor combinação extrator/classificador para os marcadores de detecção (<i>Detection</i>), tumor inteiro (<i>Whole Tumor</i>), núcleo necrótico/tumor não realçador (<i>NCR/NET</i>), edema peritumoral (<i>ED</i>) e pixels de fundo (<i>Background</i>) no banco de dados BraTS2020.	85
Tabela 10	– Tabela comparativa do método de segmentação abordado (<i>Divisible Cell Segmentation</i>) com métodos encontrados na literatura que também se utilizava da base TCGA.	86
Tabela 11	– Comparação entre o método proposto, e os métodos desenvolvidos por (NASCIMENTO et al., 2023a).	88
Tabela 12	– resultados do teste estatístico de Kolmogorov-Smirnov aplicados na comparação entre o método proposto neste estudo e os métodos desenvolvidos por (NASCIMENTO et al., 2023a).	89

Tabela 13 – Comparação com trabalhos encontrados no estado da arte para classificação do tempo de sobrevivência de pacientes para o banco de dados de validação Brats2020.	90
Tabela 14 – Comparação com o estado da arte para classificação binária (presença) de tumor cerebral no banco de dados BraTS2020.	91

Lista de abreviaturas e siglas

OMS	Organização Mundial da Saúde
CAD	Diagnóstico Auxiliado por Computador (<i>Computer-Aided Diagnosis</i>)
PDI	Processamento Digital de Imagens
PARZEM	Janelamento de Parzem
CNN	Rede Neural Convolucional (<i>Convolutional Neural Network</i>)
RNAs	Redes Neurais Artificiais
R-CNN	Redes Neurais Convolucionais baseadas em Regiões
ROI	Região de Interesse
ACC	Acurácia (<i>Accuracy</i>)
DICE	Coefficiente de Dice (<i>Dice Coefficient</i>)
SEN	Sensibilidade (<i>Sensitivity</i>)
SPE	Especificidade (<i>Specificity</i>)
PPV	Valor Preditivo Positivo (<i>Positive Predictive Value</i>)
NPV	Valor Preditivo Negativo (<i>Negative Predictive Value</i>)
MCC	Coefficiente de Correlação de Matthews (<i>Matthews Correlation Coefficient</i>)
JAC	Índice de Jaccard (<i>Jaccard Index</i>)
TP	Verdadeiro Positivo
FP	Falso Positivo
TN	Verdadeiro Negativo
FN	Falso Negativo
YOLOv8	<i>You Only Look One Level</i>
YOLOv8l	<i>YOLOv8 Large</i>
YOLOv8m	<i>YOLOv8 Medium</i>

YOLOv8n	<i>YOLOv8 Nano</i>
YOLOv8s	<i>YOLOv8 Small</i>
YOLOv8x	<i>YOLOv8 Extra Large</i>
UFCA	Universidade Federal do Cariri
UFC	Universidade Federal do Ceará
IFCE	Instituto Federal de Educação, Ciência e Tecnologia do Ceará
PPGCC	Programa de Pós-Graduação em Ciência da Computação
ROC	Curva Característica de Operação do Receptor (<i>Receiver Operating Characteristic</i>)
AUC	Área Sob a Curva (<i>Area Under the Curve</i>)
WGAN	Rede Generativa Adversarial de Wasserstein (<i>Wasserstein GAN</i>)
Mask R-CNN	Redes Neurais Convolucionais baseadas em Regiões com Máscara
MLP	Perceptron de Múltiplas Camadas (<i>Multi-Layer Perceptron</i>)
SVM	Máquina de Vetores de Suporte (<i>Support Vector Machine</i>)
KNN	K Vizinhos Mais Próximos (<i>K-Nearest Neighbors</i>)
RF	Floresta Aleatória (<i>Random Forest</i>)
NB	Naive Bayes
DICOM	Comunicação e Imagens Digitais em Medicina (<i>Digital Imaging and Communications in Medicine</i>)
CT	Tomografia Computadorizada (<i>Computed Tomography</i>)
INPE	Instituto Nacional de Pesquisas Espaciais
INMET	Instituto Nacional de Meteorologia

Lista de símbolos

ω	Letra grega Ômega minúsculo, indica a operação de máximo local no detector de fronteira difuso.
Ω	Letra grega Ômega maiúsculo, indica a operação de mínimo local no detector de fronteira difuso.
Π	Letra grega Pi Maiúscula, denominada Produto ou Produtório, no conceito matemático, refere-se à multiplicação.
Φ	Letra grega Fi, equivale ao conjunto de parâmetros definidos pela i-ésima componente da mistura.
Θ	Letra grega Teta, representa a medida de um determinado vetor, sendo um conjunto de parâmetros da mistura.
θ	Letra grega teta, referência ao ângulo.
Σ	Notação Sigma, somatório ou somatória, significa a soma de termos.
π	Símbolo Pi, representa uma proporção numérica definida pela relação entre o perímetro de uma circunferência e seu diâmetro.
ϕ	Letra grega Fi (variante), correspondente a uma função de densidade Gaussiana.
α	Letra grega Alpha, conotação angular de início ou inicialização.
β	Letra grega Beta, conotação ponderação ao ângulo.
$\int$	Símbolo Integral, representa um sinal que indica a integração e o resultado de integrar uma diferencial ou derivada.
μ	Letra grega Mi (Pronúncia: Miú), refere-se ao peso específico de um coeficiente.
∂	Símbolo "del", usado para distinguir derivadas parciais das derivadas ordinárias de uma variável.
$\cap$	Símbolo Intersecção, corresponde à intersecção entre conjuntos.
ω	Letra grega Ômega minúsculo, indica o mínimo local.
Ω	Letra grega Ômega maiúsculo, indica a composição do mínimo com o máximo local.

ν	Letra grega Nu, indica o limiar da força balão.
∇	Símbolo grego Nabla, denomina o operador diferencial.
$\in$	Símbolo que indica a relação de pertinência.

Sumário

1	INTRODUÇÃO	18
1.1	Motivação para a Dissertação	23
1.2	Descrição do Problema	24
1.3	Objetivo Geral e Específicos	25
1.3.1	Objetivo Geral	25
1.3.2	Objetivos Específicos	26
1.4	Produção científica	27
2	TRABALHOS RELACIONADOS	28
3	MATERIAIS E MÉTODOS	34
3.1	Base de dados - TCGA / BraTS2020	34
3.2	Frameworks de Detecção	35
3.2.1	YOLOv8	36
3.3	Técnicas de Processamento Digital de Imagem	40
3.3.1	<i>Divisible Cell Segmentation</i>	40
3.3.2	Conexões de Coordenadas de Pontos – <i>Concave Hull</i>	43
3.3.3	Data Fusion - Interpolação das Modalidades de RM	44
3.4	Classificadores de Aprendizado de Máquina	45
3.4.1	KNN	46
3.4.2	Naive Bayes	47
3.4.3	MLP	48
3.4.4	Random Forest	48
3.4.5	SVM	49
3.4.6	Grid Search	50
3.5	Redes Neurais Convolucionais	51
3.5.1	VGG	52
3.5.2	DenseNet	54
3.6	Modelos de Processamento de Linguagens natural em larga escala (LLMs)	55
3.6.1	Modelo Llama 3	56
3.7	Métricas de Avaliação	58
3.7.1	Métricas de Segmentação	58
3.7.2	Métricas de classificação	59
3.8	Teste estatístico de Kolmogorov-Smirnov	61

4	METODOLOGIA	63
5	RESULTADOS E DISCUSSÃO	75
5.1	Inicialização da abordagem	75
5.2	Segmentação	76
5.3	Classificação	78
5.4	Validação de Resultados (Tumor Cerebral) - Comparação com a	
	Literatura	86
5.4.1	Comparação - Segmentação	86
5.4.2	Comparação - Classificação	90
6	CONCLUSÃO	93
	REFERÊNCIAS	95

1 Introdução

Tumores Cerebrais são uma proliferação anormal e descontrolada de células localizadas no tecido do sistema nervoso central, e podem ser classificados como primários, originados no cérebro ou medula espinhal, ou secundários, gerados a partir do processo de metástase de cânceres localizados em outras regiões corporais que se alastraram pelo corpo (RAHMAT et al., 2021). Dentre os tumores primários, os que se destacam são: Glioblastoma Multiforme (POUYAN et al., 2025), Astrocitoma Anaplásico (OROS; PULYK; JR, 2024) e o Oligodendroglioma (SKOURAS et al., 2025), variando em capacidade de crescimento e evolução. Embora os tumores cerebrais sejam menos comuns que os demais tipos de câncer, sua localização no sistema nervoso central costuma impor uma série de desafios em seu diagnóstico e terapia, sendo necessário muitas vezes uma abordagem multidisciplinar para aumentar as chances e a qualidade de vida do paciente (KHALIGHI et al., 2024).

Os tumores cerebrais são classificados como malignos ou benignos com base em seu potencial de crescimento e capacidade de invadir o tecido cerebral adjacente (HANIF et al., 2017). Tumores cerebrais malignos apresentam crescimento rápido, invasão local e potencial de disseminação para outras regiões do cérebro, enquanto os tumores cerebrais benignos geralmente crescem em um ritmo mais lento e permanecem localizados (PATEL, 2020). De acordo com a literatura científica, o glioblastoma multiforme (GBM) é o tipo mais prevalente e agressivo de tumor cerebral maligno. Conforme observado por (WU et al., 2021), “sendo o tumor cerebral primário mais comum em adultos e tendo geralmente uma taxa de sobrevida média de apenas 15 meses”. Outros tumores cerebrais malignos incluem astrocitoma anaplásico e oligodendroglioma anaplásico (SALLEMI; NJEH; LEHERICY, 2015; WIJETHILAKE et al., 2021).

De acordo com a revisão de literatura de Ricard et al. (2013), “os meningiomas são os tumores cerebrais primários mais comuns, com uma incidência de aproximadamente 2,2 por 100.000 pessoas a cada ano.” Esses tumores são normalmente benignos, mas podem crescer e comprimir o tecido cerebral, levando a sintomas neurológicos (RICARD et al., 2012).

A identificação do tumor em estágios mais iniciais e o diagnóstico preciso do seu tipo são essenciais para um tratamento eficaz, com menor ocorrência de efeitos colaterais (BHIMAVARAPU; CHINTALAPUDI; BATTINENI, 2024). A ressonância magnética (RM) é a principal ferramenta diagnóstica utilizada para detectar tumores cerebrais por parte dos médicos especialistas, permitindo a visualização detalhada do tecido cerebral e dos tumores. De acordo com a revisão bibliográfica de (SMIRNIOTOPOULOS et al., 2007), “a RM é um dos métodos de exame de maior escolha para a captura de imagens de tumores

cerebrais, graças à sua superior capacidade de caracterização anatômica e patológica não invasiva”.

Desta forma, as ressonâncias magnéticas produzem imagens com intensidades de pixel variadas que facilitam a compreensão de diferentes estruturas nas varreduras, auxiliando no diagnóstico médico e na tomada de decisões (MANUWEERA et al., 2024).

É importante ressaltar que quando trata-se de tumor, cada caso é único e o tratamento deve ser individualizado de acordo com as características patológicas e fisiológicas do paciente (MASUCCI et al., 2024). A literatura médica estabelece parâmetros analíticos baseados em exames de ressonância magnética e tomografia computadorizada (TC), proporcionando amplo conhecimento por meio do estudo clínico e do diagnóstico (BOMPOU et al., 2024). Com os avanços da tecnologia computacional, diversos métodos são utilizados para auxiliar no diagnóstico por imagem.

Um dos principais desafios na identificação de tumores cerebrais por meio de ressonância magnética é a dificuldade em distinguir certas bordas delimitantes entre o tecido tumoral do tecido saudável, dificultando a diferenciação exata de até onde se dá o tumor e o tecido saudável dentro da massa cerebral (YANKEELOV et al., 2016). Isso pode interferir diretamente no diagnóstico, gerando perdas de precisão e impactando negativamente no acompanhamento da progressão da patologia no paciente, haja vista que a variação tumoral, a localização regional e a textura são fatores importantes a serem levados em consideração para o diagnóstico (WU et al., 2017).

Outro problema comum que pode impactar na qualidade do diagnóstico é a presença de artefatos de imagem que podem atrapalhar na delimitação do tumor cerebral: os artefatos podem surgir a partir de vários fatores, incluindo movimentação do paciente durante a aquisição da imagem, falhas técnicas no equipamento de ressonância magnética ou uso inadequado de agentes de contraste de imagem (ZAITSEV; MACLAREN; HERBST, 2015). Esses artefatos podem resultar em diagnósticos incorretos e gerar confusão entre os profissionais de saúde (WU et al., 2017).

Além disso, a divergência entre observadores na interpretação das imagens pode contribuir para os desafios na identificação de tumores cerebrais em ressonância magnética: Diferenças na experiência dos profissionais de saúde e nas interpretações subjetivas das imagens podem levar a diagnósticos variados para o mesmo paciente (GROWCOTT et al., 2020). Essa questão é particularmente crítica em casos em que a detecção precoce do tumor é essencial para um tratamento eficaz (PROVENZALE; ISON; DELONG, 2009).

Também é importante enfatizar que a identificação de um tumor cerebral em uma ressonância magnética é uma tarefa complexa que deve ser realizada por profissionais de saúde experientes e bem treinados (R; V; GULUWADI, 2024). A interpretação das imagens exige uma análise integrada de múltiplos aspectos: Como a localização e o volume

do tumor, as características do tecido adjacente, padrões de realce, presença de edema e sinais de invasão, entre outros. (MARTUCCI et al., 2023).

A adoção de métodos de segmentação e classificação tumoral, em colaboração com o médico especialista, é fundamental para determinar o tratamento adequado e a expectativa de vida do paciente com maior eficácia (RANJBARZADEH et al., 2023). Ao obter acesso ao tipo de tumor, bem como às suas dimensões e estado atual, o médico especialista pode definir a condição do paciente com maior precisão, e até mesmo avaliar a progressão da patologia (KRITHIGA; GEETHA, 2021).

O uso de técnicas computacionais, como aprendizado de máquina e aprendizado profundo, é essencial para auxiliar o diagnóstico médico por meio de sistemas de diagnóstico auxiliado por computador (CAD) (WANG et al., 2024). Essas abordagens permitem identificar anomalias precocemente e potencializam a análise médica por meio da extração, categorização e segmentação precisa de regiões de interesse (LEMING et al., 2023).

A segmentação de imagens é uma técnica complexa que envolve a divisão de uma imagem em regiões distintas com características específicas (MINAEE et al., 2021). Em imagens médicas, a segmentação pode ser aplicada para separar tumores de tecido saudável ou para delinear órgãos na imagem, para o planejamento cirúrgico (SUN et al., 2019). De acordo com (GONZALEZ; WOODS, 1992), a segmentação fornece uma representação quantitativa das lesões, tornando-se um recurso crucial para diagnóstico preciso e monitoramento terapêutico.

A classificação de imagens é outra tarefa crítica em sistemas de visão computacional, pois envolve a atribuição de um rótulo a uma imagem com base em suas características (GALIĆ et al., 2023). Diversos modelos são capazes de extrair características e classificar a imagem de forma independente (GIUDICI; GRAMEGNA; RAFFINETTI, 2023). O diagnóstico de um paciente pode ser definido com mais precisão com o auxílio de atributos específicos presentes na imagem, fornecendo ao médico especialista informações adicionais para o planejamento do tratamento e/ou cirurgia (LIU et al., 2019). Entre esses atributos estão as medidas das estruturas intracranianas, a morfologia e o estado da lesão e informações clínicas complementares provenientes do histórico do paciente (AYADI et al., 2021).

O surgimento de modelos de Deep Learning cada vez mais robustos oferece uma nova perspectiva para a área médica, com a aplicação de novas redes neurais e modelos de processamento de linguagem natural empregados como ferramentas para auxiliar no diagnóstico médico (RANJBARZADEH et al., 2023). Essas tecnologias ainda dependem de modelos CNN como uma espécie de espinha dorsal para a análise de imagens médicas, visto que as camadas internas desses modelos envolvem transformações, modificações, redimensionamento e etapas de pré-processamento do domínio da imagem que geralmente aprimoram a região de interesse (HAN et al., 2024). No entanto, recursos profundos também

introduzem inúmeros atributos indistintos e redundantes ao processo de classificação, que, se não forem devidamente tratados, podem impactar negativamente o desempenho do modelo (NIZAMANI et al., 2023).

Na literatura, existem inúmeros métodos de segmentação adaptados para tumores cerebrais (KRISHNAPRIYA; KARUNA, 2023). Modelos mais antigos, puramente baseados em processamento digital de imagens (PDI), foram superados por estudos mais recentes focados na segmentação de tumores cerebrais em conjuntos de dados de ressonância magnética usando CNNs, como arquiteturas aprimoradas de U-Net, nnU-Net e ensembles envolvendo ambos (RANJBARZADEH et al., 2023). A adoção desses modelos pode proporcionar alto desempenho de segmentação em diferentes modalidades de ressonância magnética devido aos seus processos operacionais. Por exemplo, o estudo de (NIZAMANI et al., 2023) propôs modelos baseados em variações de U-Net com uma estrutura sistemática de pré e pós-processamento com o objetivo de abordar a alta complexidade de recursos profundos. Os autores utilizaram o conjunto de dados BraTS2020 para validação dos resultados. O modelo alcançou precisões médias acima de 99% na segmentação, demonstrando a eficiência das CNNs com ajustes de processamento. Vale a pena investigar se a criação de métodos de pré ou pós-processamento baseados em PDI na segmentação tem o potencial de melhorar o desempenho com tempo de processamento adicional mínimo em comparação com modelos inteiramente baseados em CNNs.

Ainda assim, é possível encontrar novos métodos de segmentação baseados no processamento digital de imagens sendo propostos na literatura, que se aplicados corretamente, conseguem exprimir resultados tão satisfatórios quanto, ou até superiores, aos de métodos baseados em CNN para atividades específicas (SOUZA et al., 2023). O estudo de (SOUZA, 2023) é um exemplo de casos como este. O método proposto pelo autor é computacionalmente inspirado no processo de divisão celular mitótica da biologia, na qual as propostas Células Sintéticas Computacionais desenvolvidas realizam divisão exponencial durante o processo de movimentação no interior da imagem. Quando corretamente inicializado, o método é capaz de detectar, ao longo de várias iterações de divisão e movimentação, as bordas da região de interesse. Ao encontrar o limiar da região de interesse com a região de fundo, as células sintéticas computacionais estacionam, e as coordenadas de sua localização são armazenadas em um vetor de pontos, que podem ser conectados para formar o polígono da região de interesse com a transformada de Hough, realizando assim o processo de segmentação. Em seus resultados, os autores fizeram teste do método para segmentação de AVC e pulmão, obtendo acurácias acima de 90%.

A classificação de tumores cerebrais na literatura, no geral, também se baseia fortemente no uso de redes convolucionais profundas, capazes de extrair características complexas para representação de imagens de ressonância magnética e superar os métodos clássicos de classificação (AKTER et al., 2024). Geralmente, esses modelos empregam

métodos auxiliares, como aprendizado de transferência (JAIN; JAIN, 2023), para combinar diferentes modelos CNN ou mesmo Fusão de Dados (JANARDHAN *et al.*, 2024), visando refinar descritores específicos ao consolidar informações complementares de diferentes bases. Métodos como o proposto por Jain *et al.* (JAIN; JAIN, 2023) foram desenvolvidos com base nesses conceitos. Os autores propuseram um novo método de classificação de tumores para imagens de ressonância magnética usando o conjunto de dados BraTS, alavancando o aprendizado de transferência onde as redes MobileNetV2, InceptionV3 e ResNet50 funcionam como classificadores. As características profundas, que consistem em representações hierárquicas aprendidas automaticamente, e capturam desde texturas até estruturas semânticas, são refinadas e classificadas por uma camada densa personalizada, otimizada para realçar distinções críticas entre tumores. O modelo alcançou precisões de classificação acima de 98%. Vale a pena investigar se a adoção da fusão de dados combinada com a aprendizagem de transferência tradicional poderia gerar melhorias ainda mais significativas na classificação, auxiliando na redução de estruturas complexas e características indistintas presentes em atributos profundos.

O uso de modelos de processamento de linguagem natural e visual como ferramentas para auxiliar no diagnóstico médico pode trazer benefícios em termos de pontos de vista e interpretações clínicas obtidas, bem como na análise de características específicas que podem contribuir diretamente para o diagnóstico. Tais características poderiam ser sutis e passar despercebidas em análises menos experientes ou menos completas da patologia (GOH *et al.*, 2024). Estudos como o de Van *et al.* (VAN; VERMA; WU, 2024), que utilizaram modelos de linguagem visual (VLMs) para obter pré-diagnóstico a partir de imagens médicas, demonstram a potencial aplicabilidade desses modelos. Embora não superem os modelos de visão computacional em tarefas de segmentação ou classificação, podem funcionar como chatbots especializados, adicionando informações suplementares (MENG *et al.*, 2024). Espera-se que a fusão dos outputs de modelos robustos de visão computacional com a capacidade de raciocínio contextual de LLMs gere interpretações enriquecidas, combinando evidência visual com inferência semântica.

Baseando-se nos problemas levantados, na complexidade e variabilidade dos tumores e na exigência de uma identificação automática e eficiente, além da existência de métodos capazes de capturar e rotular múltiplas características tumorais. Este estudo apresenta uma abordagem baseada em aprendizado profundo e ajuste fino para a detecção, segmentação e classificação de tumores cerebrais em imagens de ressonância magnética (RM) utilizando uma metodologia que inclui processamento digital de imagens. Dentre os aspectos mais relevantes do estudo, o modelo proposto visa aprimorar as métricas de segmentação por meio da adoção de técnicas inovadoras da literatura para detecção e segmentação. Além disso, a adoção de aprendizado de transferência combinado com extratores CNN e classificadores clássicos, bem como o uso de duas bases de dados pode gerar potenciais melhorias no desempenho da classificação.

Neste trabalho, propomos uma nova abordagem para análise de tumores cerebrais combinando segmentação via células computacionais, classificação de imagens de ressonância magnética por meio de métodos híbridos de transferência de conhecimento e um modelo LLM para aprimorar o pré-diagnóstico. O método integra detecção, segmentação, extração de features e classificação para gerar rótulos médicos informativos sobre o quadro clínico do paciente. Além disso, dois conjuntos de dados complementares de ressonância magnética de tumores cerebrais são utilizados para detecção, segmentação e classificação dentro do método. Os atributos classificados são então empregados para gerar um pré-diagnóstico com a contextualização e refinamento providos por um LLM. Os resultados das etapas de segmentação e classificação são validados por meio da avaliação frente ao estado da arte em trabalhos análogos que empregam os mesmos datasets para o mesmo problema.

As principais contribuições deste artigo são apresentadas da seguinte forma:

- Aplicação do inovador método de segmentação de imagens médicas baseado no conceito de células sintéticas computacionais, combinadas com a detecção do framework YOLO.
- Desenvolvimento de pipeline de extração com extratores CNN e classificação com extratores *Machine Learning* através do processo de *Transfer Learning*.
- Desenvolvimento de uma ampla metodologia que busca detectar, segmentar e classificar o Tumor cerebral em imagens de MRI.
- Aplicação de modelo de processamento de Linguagem Natural na resposta da classificação, com elaboração de resposta na formatação de pré diagnóstico textual.

1.1 Motivação para a Dissertação

A adesão do médico especialista às diversas tecnologias de auxílio ao diagnóstico médico (CAD) podem trazer benefícios como: o aumento da acurácia diagnóstica, com identificação de padrões sutis fora do alcance do olho humano, o ganho da produtividade e eficiência em atividades de alto grau de repetição, como triagens de imagens e pré anotação de dados; padronização e consistência de dados através de minimização da variabilidade intra e interobservador; suporte á decisão através do levantamento ordenado dos dados clínicos, auxílio no acompanhamento clínico e comunicação com o paciente.

Na literatura, existem diversos métodos computacionais com boa performance para aplicações clínicas como detecção, segmentação e classificação de imagens médicas. Alguns modelos, embora tenham boa performance, são evidentemente dependentes da intervenção humana para inicialização ou parametrização de acordo com a base aplicada, sendo considerados métodos semi-automáticos.

Com o advento de novos métodos computacionais, cada vez mais modelos voltados

para aplicações na área médica, de alta eficácia, são encontrados na literatura: Métodos computacionais que demandam quantidades de imagens cada vez menores, com menor tempo de treinamento, mas mantendo um alto grau de acurácia, além de uma maior facilidade de uso, personalização e configuração para bases clínicas.

Ainda assim, no que diz respeito à melhora da performance e acurácia desses métodos, existem desafios a serem superados: muitos advindos das condições da base de dados utilizada, como qualidade das capturas, confiabilidade das anotações clínicas e equilíbrio entre os diferentes tipos de classes.

Todos esses aspectos fazem necessário não somente o conhecimento à cerca das particularidades da base de dados aplicada, mas também do conhecimento de métodos adicionais de pré ou pós processamento adequados de acordo com as características da base, aplicados com finalidade de ampliar o resultado de performance obtida pelos modelos.

No contexto do tumor cerebral, existem desafios a serem superados por estes modelos computacionais robustos no que diz respeito à delimitação da região afetada pela patologia na imagem, diferenciação da benignidade do tumor, e até mesmo a contagem exata da área lesionada. Embora a Ressonância Magnética tenha variações de modalidade, que mudem a perspectiva de visualização da patologia na imagem, os métodos implementados devem ser preparados para lidar com variações sutis da massa cinzenta, se adaptando às condições da imagem e até a modalidade de captura, para melhor delimitar a região segmentada e ter os atributos extraídos.

Vale destacar também as novas possibilidades de aplicação clínica advindas do surgimento de modelos cada vez mais robustos de Processamento de linguagem natural, que são capazes de processar enormes quantidades de dados médicos da literatura, para trazer análises e ponderações adicionais levando em consideração os dados clínicos obtidos, visando auxiliar o médico especialista no planejamento clínico, acompanhamento e tomada de decisões.

Esta dissertação tem como foco explorar e aprimorar abordagens de detecção, segmentação e classificação do Tumor Cerebral em imagens de RM. Com o objetivo de fornecer uma metodologia completa, que busca delimitar a região do tumor, segmentar e classificar regiões internas tumorais, como o Núcleo necrótico e não realçado.

1.2 Descrição do Problema

A detecção, segmentação e classificação de tumores cerebrais em imagens de Ressonância magnética, são etapas diferentes, mas geralmente interconectadas em um sistema de auxílio ao diagnóstico médico, cada uma com desafios intrínsecos.

A detecção e segmentação são comumente afetadas por variabilidade morfológica

do tumor na imagem, com diferentes formatos, texturas, intensidade de cor e localizações no exame. Desta forma a definição exata da localização, bem como a delimitação precisa das bordas irregulares do tumor, e sua sutil fronteira com a massa cerebral saudável, são desafios à lidar na implementação de um sistema que busca eficiência de detecção e segmentação. Isto é, o sistema deve estar preparado para as diferentes condições de captura das imagens de entrada, mantendo o grau de confiabilidade das detecções.

Embora haja uma ampla gama de modelos capazes de realizar a segmentação do tumor de maneira totalmente automática. Vale ressaltar que existem métodos inovadores de segmentação, baseados em processamento digital de imagens, que podem vir á obter uma boa performance quando bem aplicados e inicializados na imagem, respondendo bem até mesmo para casos mais complexos de segmentação. A segmentação pode inclusive ser realizada mediante a cooperação entre diferentes métodos, na qual um realiza a segmentação inicial e o outro atua como *fine-tuning* nas bordas.

A classificação do tumor, geralmente tem sua eficácia diretamente relacionada com a segmentação, pois é comum os atributos serem extraídos da região de interesse delimitada para melhorar a representabilidade do atributos extraídos. Isso torna vital a etapa de seleção de extratores de atributo adequados, que sejam capazes de extrair atributos que capturem as peculiaridades clínicas das classes á serem classificadas.

A parametrização adequada de modelos também se mostra um desafio constante na aplicação de métodos clássicos e de *Machine Learning*. Geralmente costuma-se demandar conhecimento prévio sobre a influência relativa dos parâmetros no desempenho do modelo, além de saber quais parâmetros podem afetar mais a performance em detrimento de outros. Frequentemente empregam-se técnicas de otimização de hiperparâmetros para automatizar a busca pela configuração que maximiza o desempenho na base de dados, reduzindo a necessidade de ajuste manual.

1.3 Objetivo Geral e Específicos

1.3.1 Objetivo Geral

Formular, avaliar e discutir um método de detecção, segmentação e classificação automática de tumores cerebrais em imagens de RM, baseado na integração de conceitos como: Detecção automática da patologia com YOLOv8; Segmentação da patologia pelo método das células sintéticas computacionais; Seleção de Melhor combinação de Extrator/Classificador com aplicação Grid Search; Geração de pré diagnóstico clínico textual através da aplicação de modelo de processamento de Linguagem Natural. Espera-se que o modelo seja capaz de detectar, segmentar, classificar e elaborar um pré diagnóstico clínico a partir da imagem de RM de entrada, com um nível adequado de acurácia.

De maneira geral, a proposta do trabalho busca mitigar a dependência direta do especialista ao automatizar etapas centrais como a detecção, segmentação, extração de atributos, classificação e geração de um pré-diagnóstico clínico textual, de forma que o sistema entregue uma avaliação inicial consistente e explicável a partir da imagem de RM.

1.3.2 Objetivos Específicos

1. Configurar e treinar o *framework* YOLOv8 para detectar o Tumor Cerebral nas imagens de RM, visando altos índices de acurácia, devido ao refinamento da arquitetura, facilidade de implementação e melhorias realizadas nesta versão do *framework*.
2. Adaptar e aplicar o algoritmo de Células Sintéticas Computacionais desenvolvido por (SOUZA et al., 2023) na segmentação a partir da etapa anterior de detecção, visando a melhora das métricas de avaliação de segmentação.
3. Construir um pipeline de Classificação que combine um conjunto delimitado de extratores de características de CNN, com um conjunto delimitado de classificadores *Machine Learning*. A comunicação realizada mediante *transfer-learning*, com uso de métodos de busca de hiperparâmetros ótimos.
4. Aplicar diferentes bases dados de Tumor Cerebral em conjunto para o desenvolvimento da metodologia, para as etapas de detecção, segmentação e classificação.
5. Integração do resultado da etapa de classificação com modelos de processamento de Linguagem Natural, visando a adequação da resposta do sistema ao formato de pré diagnóstico clínico objetivo.
6. Proporcionar uma discussão quantitativa e qualitativa à cerca dos resultados obtidos em cada uma das etapas chaves da metodologia proposta, elencando perspectivas de interpretação causal, bem como as métricas, funcionamento dos modelos selecionados, resposta à pré processamento e comportamento esperado dos métodos.

1.4 Produção científica

Durante este projeto de mestrado, os seguintes trabalhos científicos foram aceitos e publicados, a saber:

- (SIBGRAPI 2024) – *New Approach Generative AI Melanoma Data Fusion for Classification in Dermoscopic Images with Large Language Model* – 1º Autor – Qualis A3 (Eng IV), A3 (Comp);
- (Frontiers in Communications and Networks 2024) – *Health of Things Melanoma Detection System—detection and segmentation of melanoma in dermoscopic images applied to edge computing using deep learning and fine-tuning models* – 2º Autor – Qualis C (Eng IV), C (Comp);
- (Computerized Medical Imaging and Graphics 2025) – ***A novel generative model for brain tumor detection using magnetic resonance imaging*** – 2º Autor – Qualis A1 (Eng IV), A2 (Comp);
- (ISDA 2023) – *SDA-Detection Melanoma: Deep Approach System for Detection and Segmentation in Melanoma Images using Fine-Tuning* – 4º Autor – Qualis B1 (Eng IV), B1 (Comp);
- (ISDA 2023) – *ALPR System Perspective Adjustment: New Automatic License Plate Recognition Approach for Brazilian Mercosur Model Vehicle Plates* – 2º Autor – Qualis B1 (Eng IV), B1 (Comp);
- (SBESC 2024) – *Wearable Stroke Alert System – New Health of Things Approach Based on Generative AI and Datafusion for Real-Time Stroke Monitoring* – 1º Autor – Qualis B1 (Eng IV), B1 (Comp).

2 Trabalhos Relacionados

A segmentação de tumores cerebrais em imagens de ressonância magnética é uma tarefa crítica na medicina, essencial para o diagnóstico precoce, tratamento e prognóstico do paciente. Pesquisadores como (ALLAH; SARHAN; ELSHENNAWY, 2023) desenvolveram uma nova variação profunda da U-Net, denominada Edge U-Net, para segmentação de tumores cerebrais em imagens de ressonância magnética. O método proposto integra informações sobre as bordas da lesão para melhorar a acurácia da segmentação, empregando uma estrutura codificadora-decodificadora inspirada na U-Net, combinada com um Bloco de Orientação de Bordas (EGB) e uma nova função de perda que combina perdas de pixels e bordas. O modelo obteve melhorias significativas, como demonstrado por seus resultados, atingindo pontuações de Dice acima de 90% para segmentação de gliomas. O conjunto de dados utilizado foi compilado no estudo de (CHENG et al., 2015). Em relação às limitações do estudo, o modelo ainda enfrenta desafios clássicos na segmentação tumoral, como variação de intensidade e clareza das bordas da lesão em relação ao tecido saudável, que poderiam ser melhor abordados por meio da aplicação de métodos de pré-processamento ou *ajuste fino*.

O estudo de (RAMMURTHY; MAHESH, 2022) propôs uma técnica de otimização utilizando a "Otimização de Harris Hawks para Baleias" (*Whale Harris Hawks Optimization* - WHHO) para detectar tumores cerebrais em imagens de ressonância magnética. A WHHO proposta foi projetada integrando o Algoritmo de Otimização de Whale (*Whale Optimization Algorithm* - WOA) e o Algoritmo de Otimização de Harris Hawks (*Harris Hawks Optimization Algorithm* - HHO), com a detecção de tumores realizada utilizando o *Deep CNN* treinado com WHHO. Para validar o método desenvolvido, os autores segmentaram tumores do banco de dados BRATS 2014, alcançando uma precisão máxima de 81,6% e uma especificidade máxima de 79,1%, demonstrando assim o potencial do método para a detecção de imagens médicas. No entanto, estudos que incorporam métodos de pré-processamento baseados em processamento digital de imagens podem aprimorar significativamente o modelo proposto, enfatizando as características distintivas do método utilizado.

Outro trabalho relevante é o de (TUSTISON et al., 2010), que propôs uma abordagem para segmentação de tumores cerebrais em imagens de ressonância magnética com base em uma combinação de métodos de registro, difusão e segmentação. O método proposto foi avaliado em diversos bancos de dados, incluindo o BRATS, e demonstrou resultados promissores, alcançando uma acurácia média de 88,6% na segmentação de tumores cerebrais.

(ÖZKARACA et al., 2023) propõem uma abordagem simétrica para segmentação de tumores cerebrais em imagens de ressonância magnética por meio da integração de um Bloco de Atenção Tripla (*Triple Attention Block* - TAB) e módulos residuais profundos, visando aprimorar a precisão e a eficiência do modelo. A metodologia desenvolvida pelos autores inclui a substituição de módulos convencionais da arquitetura CNN por unidades residuais para melhor mitigar problemas de borda na imagem, seguida pela aplicação do TAB para fundir características locais e globais. O conjunto de dados escolhido para a aplicação está disponível no Kaggle e inclui os desafios BraTS2018 e BraTS2019. Na segmentação, o modelo obteve pontuações Dice de 88,25%, 90,66% e 85,89% para a segmentação do tumor inteiro, núcleo do tumor e modalidades de tumor aprimorado, respectivamente. Vale ressaltar que o uso de conjuntos de dados mais atualizados, bem como uma etapa adicional de testes estatísticos, poderiam aprimorar a discussão qualitativa sobre o desempenho do modelo.

(DIPU; SHOHAN; SALAM, 2021) apresentaram duas abordagens baseadas em *Deep Learning* para a detecção e classificação de tumores cerebrais no banco de dados BRATS 2018. A primeira abordagem utilizou a técnica *You Only Look Once* (YOLOv5) para identificar tumores nas imagens, demonstrando eficácia e rapidez na detecção. A segunda abordagem empregou a biblioteca *FastAi* (HOWARD; GUGGER, February, 2020), conhecida por sua versatilidade, alcançando uma acurácia de 95,78% na segmentação de tumores cerebrais, superando o *YOLOv5*, que obteve 85,95%. Esses resultados ressaltam a eficiência dessas técnicas de *Deep Learning* em aplicações médicas, particularmente no diagnóstico de tumores cerebrais, e destacam áreas potenciais para futuras investigações e melhorias metodológicas.

(TRIPATHI; VERMA; SHARMA, September, 2021) desenvolveram um método inovador e automático de aprendizado profundo para a quantificação precisa de regiões tumorais em imagens de ressonância magnética cerebral. Este método se destaca na delimitação de regiões tumorais irregulares, preservando a fidelidade de bordas e pixels. A técnica incorpora elementos avançados, como conexões residuais internas, normalização cruzada (CCN) e a Unidade Retificadora Linear Paramétrica (PRELU) para aprimorar o desempenho da rede. O modelo, denominado CCN-PR-seg-net, foi validado por radiologistas em relação a segmentações manuais, demonstrando eficiência e acurácia.

O estudo recente de (NASCIMENTO et al., 2023b) apresentou um modelo inovador e automatizado baseado na Internet das Coisas (IoT) para classificar, detectar e segmentar tumores cerebrais em imagens de ressonância magnética. Utilizando a rede Detectron2 e técnicas de ajuste fino. O modelo alcançou 93,10% de acurácia de classificação com a combinação DenseNet201 + SVM RBF e uma satisfatória acurácia de 99,41% na segmentação tumoral. Embora eficiente, o modelo se beneficiaria de uma validação e comparação mais amplas com outras arquiteturas de CNN para aprimorar sua aplicabilidade

em ambientes clínicos. Esse método ressalta a crescente relevância da tecnologia na medicina, fomentando futuras inovações no diagnóstico e tratamento de condições médicas complexas.

O trabalho de (SOUZA et al., 2023) concentra-se no desenvolvimento de um modelo inovador para detecção e segmentação de imagens médicas baseado na teoria de células computacionais sintéticas. O modelo é capaz de movimentar células computacionais dentro da região de AVC Hemorrágico e Pulmão em imagens de TC. Ao final do processo de movimentação, as células posicionadas na borda da região formam um polígono fechado correspondente à máscara de segmentação binária pelo método de transformada de Hough. Os resultados mostraram uma precisão média de segmentação de 99,82%, indicando excelente desempenho. A incorporação de modelos de inicialização alternativos além do Detectron2, bem como a aplicação de métodos de busca para refinar os parâmetros de segmentação, podem aprimorar ainda mais o desempenho do modelo.

(HAVAIEI et al., 2017) propuseram um método para segmentar tumores cerebrais em imagens de ressonância magnética usando Redes Neurais Convolucionais (CNNs) no conjunto de dados BRATS 2013. O trabalho alcançou uma acurácia média de 93,4%, demonstrando o potencial significativo da técnica para diagnóstico por imagem médica. No entanto, o estudo destacou que a eficácia do modelo é altamente dependente de parâmetros de pré-processamento, o que pode comprometer a precisão em diversos cenários.

Considerando a importância de classificar a presença ou ausência de um tumor em exames de ressonância magnética, (IQBAL et al., 2024) desenvolveram um método nessa área. A abordagem proposta combina atributos espaciais profundos com outras características de imagem, integrando o ResNet à fusão de atributos, resultando em um modelo denominado FusionNet. A rede atingiu uma acurácia de 97,53%, demonstrando um forte desempenho na classificação binária de tumores cerebrais. A aplicação de técnicas avançadas de balanceamento pode aprimorar ainda mais o desempenho do modelo.

Também focado na classificação binária, o modelo proposto por (GAUTAM; AHLAWAT; MITTAL, 2022) teve como objetivo classificar a presença de tumores em exames de ressonância magnética (RM), especificamente utilizando o conjunto de dados BraTS2020. O método proposto extraiu atributos por meio de manipulações de domínio e dimensionais da imagem, seguidas de redução de dimensionalidade usando PCA. Os atributos foram classificados por um conjunto de classificadores, incluindo SVM, Árvore de Decisão e Naive Bayes. O *pipeline* desenvolvido pelos autores alcançou uma acurácia média de 96%, demonstrando boa eficácia na extração de atributos. Notavelmente, a aplicação de técnicas de pré e pós-processamento às imagens pode aprimorar as características do tumor, potencialmente melhorando o desempenho de classificação do modelo.

Considerando o potencial de extração de informações vitais para um diagnóstico aprimorado por meio da classificação de imagens tumorais, (PATEL et al., 2021) empregaram a arquitetura U-Net para segmentar e classificar as regiões internas dos tumores. A

CNN foi modificada com um bloco de kernel seletivo para ajustar o campo de atenção da rede. Para a segmentação, foram obtidas pontuações médias de Dice acima de 80%, e para a classificação do tempo de sobrevivência, foi obtida uma acurácia de 49,5%.

(AKBAR; FATICHAH; SUCIATI, 2020) desenvolveram uma abordagem semelhante baseada no uso do MobileNet combinado com um modelo linear de predição de sobrevida. Os autores utilizaram variações do MobileNet, como V1 e V2, com camadas congeladas ou descongeladas. Seus resultados mostraram uma precisão média de 40,2% para a classificação do tempo de sobrevida dos pacientes.

(AGRAVAT; RAVAL, 2020) propuseram um modelo de segmentação e classificação utilizando uma rede neural 3D totalmente convolucional (FCNN 3D) com uma estrutura codificadora-decodificadora de três camadas. O modelo segmentou com sucesso as regiões internas dos tumores e classificou o tempo de sobrevida dos pacientes. A acurácia média para a classificação do tempo de sobrevida no banco de dados de validação BraTS2020 foi de 51,7%, o que representa um bom desempenho para uma tarefa de classificação tão desafiadora. Esse desempenho poderia ser ainda mais aprimorado com a aplicação de técnicas de ajuste fino ao modelo.

(TABATABAEI; REZAEI; ZHU, 2023) propõem uma abordagem híbrida para a classificação de tumores cerebrais em imagens de ressonância magnética em seu estudo. O método desenvolvido integra uma rede baseada em Transformer com unidades de autoatenção (SAU) e um ResNet aprimorado com *transfer learning*, denominado iResNet. A arquitetura implementada realiza a fusão bidirecional de características locais e globais durante a extração de atributos. Esse processo demonstrou uma melhoria significativa na capacidade de classificação do modelo, alcançando uma acurácia média de 99,13% no conjunto de dados de teste, utilizando um conjunto de dados de 3.064 imagens de ressonância magnética divididas em treinamento (60%), validação (20%) e teste (20%). A metodologia emprega estratégias de regularização, ajuste fino e processamento paralelo, reduzindo o sobreajuste e otimizando o desempenho para classes desafiadoras, como gliomas e meningiomas. Vale ressaltar que a aplicação de uma etapa preliminar de segmentação direcionada à região de interesse antes da classificação pode aprimorar o desempenho de extração/classificação de atributos específicos.

(PEI; MURAT; COLEN, 2021) empregaram uma arquitetura UNet modificada para segmentação semântica de imagens, denominada srUNet. O modelo foi utilizado para segmentar e classificar as regiões internas de tumores no banco de dados BraTS2020, bem como para classificar o tempo de sobrevida dos pacientes com base em atributos extraídos das imagens. Os resultados mostraram uma acurácia de 43% na classificação do tempo de sobrevida, que pode ser aprimorada por meio de técnicas de ajuste fino e pré-processamento do banco de dados.

(ARUMUGAM et al., 2024), em seu estudo, também apresentam um método

desenvolvido para a detecção e classificação precisa de tumores cerebrais em imagens de ressonância magnética. A abordagem emprega a otimização do Agente de Odor Crossover (CSA) para ajuste de hiperparâmetros do MLP, incluindo taxa de aprendizado, número de camadas e neurônios. Os autores validaram o método usando três conjuntos de dados distintos. Os resultados demonstraram-se satisfatórios, com uma taxa de acurácia de 98,56%, superando abordagens anteriores em métricas como pontuação F1, sensibilidade e especificidade. Uma comparação com outros métodos de otimização de parâmetros poderia ampliar ainda mais a discussão sobre a eficácia do CSA.

Portanto, inspirado nos trabalhos discutidos nesta seção, bem como nas limitações identificadas em seus estudos e nos desafios metodológicos levantados durante a criação dos modelos de cada autor, este estudo propõe o desenvolvimento de um método automático para detecção, segmentação e classificação de tumores cerebrais em imagens de ressonância magnética. O método proposto combina técnicas de detecção automática de Deep Learning, uma abordagem inovadora baseada em processamento digital de imagens capaz de emular células computacionais sintéticas para segmentação tumoral em imagens de ressonância magnética, e um *pipeline* de aprendizado de transferência para extração e classificação de características, seguida pela construção de um pré diagnóstico textual com auxílio de LLM.

A Tabela 1 apresenta o estado da arte, os trabalhos abordados nesta seção, listando os autores, arquiteturas, abordagens, conjuntos de dados utilizados, operação realizada em relação à classificação e segmentação e resultados. Dentre a comparação dos trabalhos pesquisados e o método proposto, destacam-se: a abordagem de detecção, segmentação e classificação em um único *pipeline* desenvolvido e automatizado, enquanto os artigos abordados, no geral, realizam apenas uma ou outra ação; o uso de modelos LLM para gerar um pré-diagnóstico formalizado com base nos atributos extraídos, etapa que poderia ser abordada por alguns dos trabalhos abordados; combinação de extratores CNN e classificadores clássicos com transferência de aprendizado; e uso duas bases de dados distintas, visando a melhora do desempenho da classificação.

Tabela 1 – Resumo dos trabalhos discutidos na seção de Trabalhos Relacionados deste estudo. A tabela contém informações sobre os autores, arquitetura aplicada, base de dados utilizada, proposta de método, resultados e possíveis sugestões de melhoria.

Artigo	Arquitetura	Base de Dados	Proposta de Método	Resultados	Aprofundamento-Sugerido
(RAMMURTHY; MAHESH, 2022)	WHHO + DeepCNN	BRATS 2014	Detectar tumores cerebrais com WHHO e DeepCNN.	81,6% de acurácia	Pré-processamento aprimorado.
(TUSTISON et al., 2010)	Registro, difusão e segmentação	BRATS	Segmentar tumores usando métodos combinados.	88,6% de acurácia	Validação extensiva.
(DIPU; SHOHAN; VERMA; SALAM, 2021)	YOLOv5 + FastAi	BRATS 2018	Detectar e classificar tumores com YOLOv5 e FastAi.	95,78% de acurácia (FastAi)	Sugestão de investigação adicional.
(TRIPATHI; SHARMA, September, 2021)	CCN-PR-seg-net	Diversas bases	Quantificar regiões tumorais usando componentes avançados.	Alta eficiência na quantificação tumorais	Refinamentos na segmentação.
(NASCIMENTO et al., 2023b)	Detectron2 + DenseNet201	Diversas	Classificar, detectar e segmentar tumores com IoT.	99,41% de eficácia na segmentação	Mais validações.
(SOUZA et al., 2023)	Células computacionais sintéticas	TC de AVC e pulmão	Detectar e segmentar usando células computacionais sintéticas.	99,82% de acurácia	Poderia estar outros modelos.
(IQBAL et al., 2024)	FusionNet	BRATS 2020	Classificar tumores com fusão de características.	97,53% de acurácia	Adesão de técnicas de balanceamento.
(GAUTAM; AHLAWAT; MITTAL, 2022)	PCA + SVM, Decision Tree, Naive Bayes	BRATS 2020	Classificar presença de tumores com PCA e classificadores.	96% de acurácia	Melhor destaque das características.
(PATEL et al., 2021)	U-Net modificada	BRATS 2020	Segmentar e classificar tumores com U-Net modificada.	Dice > 80% (segmentação), 49,5% (tempo de sobrevivência)	Baixa acurácia na predição de sobrevivência.
(AKBAR; FATICHAH; SUCIATI, 2020)	MobileNet + modelo linear de sobrevivência	BRATS 2020	Prever tempo de sobrevivência com MobileNet.	40,2% de acurácia (sobrevivência)	Baixa acurácia na predição de sobrevivência.
(AGRAVAT; RAVAL, 2020)	CNN 3D completa	BRATS 2020	Segmentar e classificar tempo de sobrevivência com CNN 3D.	51,7% de acurácia (sobrevivência)	Sugestão de uso de ajuste fino.
(PEI; MURAT; COLEN, 2021)	srUNet	BRATS 2020	Segmentar e classificar tumores e tempo de sobrevivência com srUNet.	43% de acurácia (sobrevivência)	Sugestão de Ajuste fino e pré-processamento.
(ALLAH; SARHAN; ELSHEN-NAWY, 2023)	Edge U-Net	Conjunto Cheng	Segmentar tumores cerebrais usando informação de borda e EGB.	Dice > 90% para gliomas	Variação de intensidade e bordas pouco nítidas.
(ÖZKARACA et al., 2023)	TAB + módulos residuais	BRATS 2018, BRATS 2019	Melhorar segmentação com blocos de atenção tripla (TAB).	Dice: 88,25%, 90,66%, 85,89%	Discussão qualitativa limitada.
(TABATABAEI; REZAEI; ZHU, 2023)	Transformer + iResNet	Customizado (3064 imagens)	Classificar tumores com fusão bidirecional de características.	99,13% de acurácia	Sugestão de segmentação prévia para classificação.
(ARUMUGAM et al., 2024)	CSA-MLP	Múltiplas bases	Detectar e classificar tumores usando MLP otimizada por CSA.	98,56% de acurácia	Comparação limitada de métodos de otimização.

3 Materiais e Métodos

3.1 Base de dados - TCGA / BraTS2020

Para este estudo, dois conjuntos de dados foram utilizados para segmentação. O primeiro é o TCGA/LGG - Segmentation Dataset, que contém 110 pacientes da coleção de gliomas de baixo grau do The Cancer Genome Atlas (TCGA), e o segundo é o Brain Tumor Retrieval (BraTS 2020). Neste último conjunto de dados as imagens são obtidas a partir de fatias de imagens de ressonância magnética com contraste cerebral (CE-MRI). O conjunto de dados BraTS2020 consiste em 3.064 fatias de 233 pacientes, incluindo 708 meningiomas, 1.426 gliomas e 930 tumores hipofisários, conforme disponibilizado pelo (TRIPATHI; VERMA; SHARMA, September, 2021). A Figura 1 mostra algumas imagens do conjunto de dados TCGA.

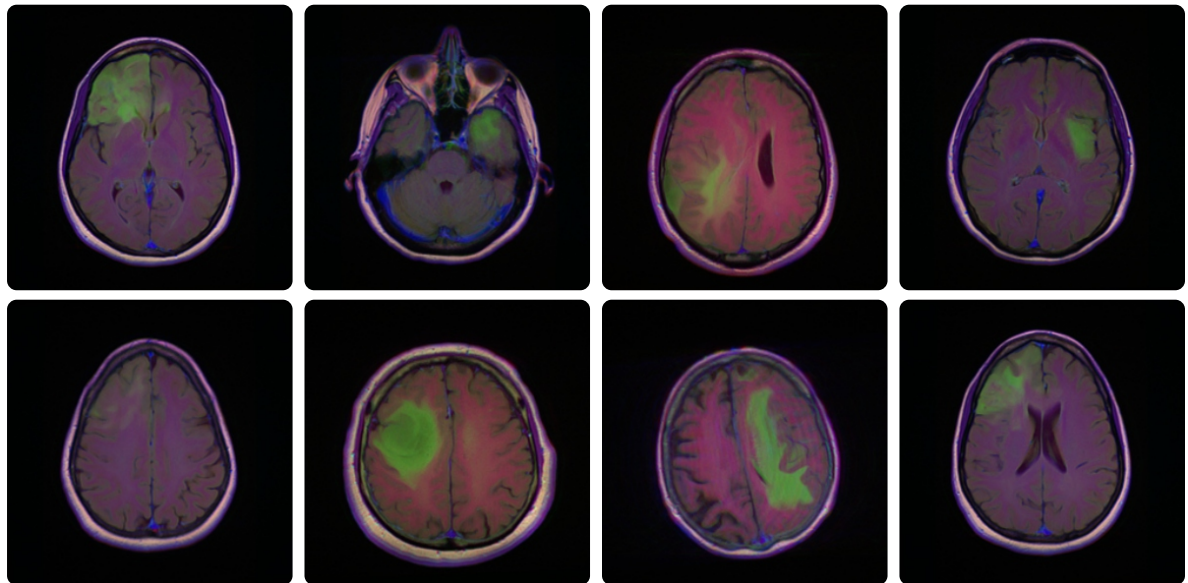


Figura 1 – Amostras do Dataset TCGA.

Para a classificação, foi utilizado o banco de dados BraTS2020, um dos principais bancos de dados para segmentação de tumores cerebrais em imagens de ressonância magnética, como parte de uma série de desafios anuais do BraTS, que visam promover pesquisa e inovação em algoritmos para segmentação e classificação automática de tumores cerebrais. O banco de dados inclui não apenas o padrão ouro de segmentação definido por médicos especialistas, mas também conta com informações relacionadas ao paciente e ao tumor, contendo 8 classes, como:

- o tempo estimado de sobrevida dos pacientes após o diagnóstico (*Survival Days*)
- idade (*Age*)

e outras 5 classes proveniente da contagem de regiões necróticas e contrastadas do tumor no momento do exame (HENRY et al., 2021):

- detecção (*Detection*)
- tumor total (*Whole Tumor*)
- núcleo tumoral necrótico e não realçado (*NCR/NET*)
- edema peritumoral (*ED*)
- e região de fundo (*Background*)

A figura 2 mostra algumas imagens do conjunto de dados BraTS2020. Vale ressaltar que o dataset possui diferentes modalidades do mesmo exame, sendo elas:

- T1 (nativa)
- T1Gd (T1 ponderada com contraste, também chamada T1-CE)
- T2 (ponderada em T2)
- T2-FLAIR (Fluid Attenuated Inversion Recovery)

A classe de idade é dada em anos, a classe de tempo de sobrevida é dada em dias, enquanto que as demais 5 classes provenientes das regiões do tumor e de fundo, são dadas em contagem de pixels.

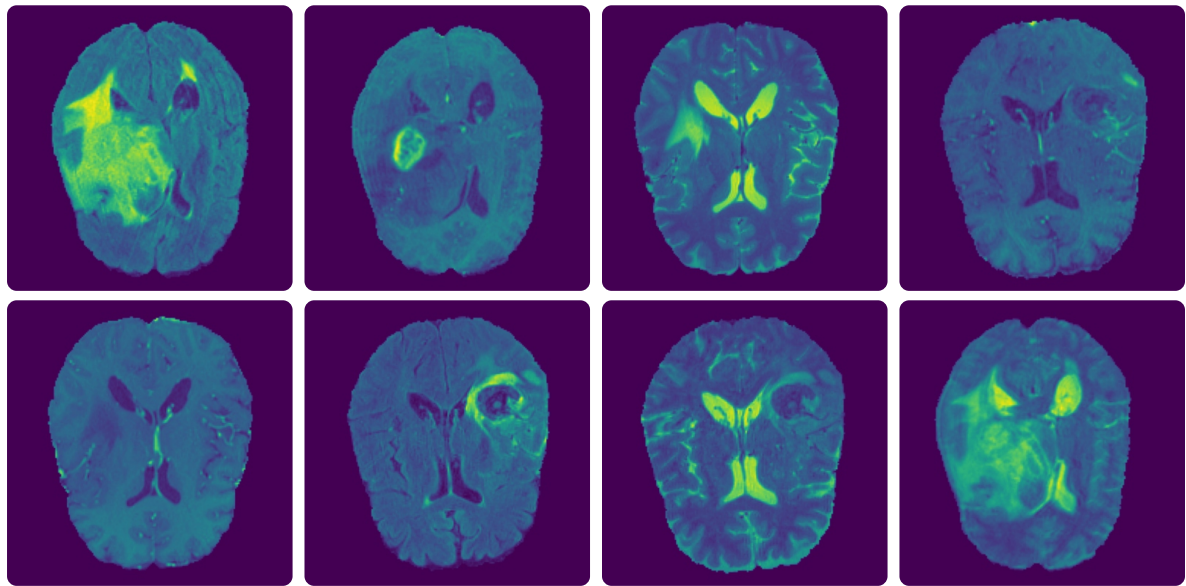


Figura 2 – Amostras do conjunto de dados BraTs2020.

3.2 Frameworks de Detecção

Os *frameworks* de detecção de objetos tem se mostrado cada vez mais presentes na literatura atual (MITTAL, 2024). Muitos destes modelos se utilizam de arquiteturas CNNs, realizando detecção e/ou segmentação dos objetos de interesse na imagem de acordo com as configurações durante a etapa de treinamento (LI et al., 2020).

Dentre os *frameworks* existentes, tem-se o Detectron2, desenvolvido pelo *Facebook AI Research* (FAIR) (WU et al., 2023), sendo uma plataforma modular baseada em pytorch para a implementação de algoritmos de visão computacional. O *framework* é capaz de realizar a detecção de objetos e segmentação de instâncias (MITTAL, 2024).

O *Tensorflow Object Detection API* também faz parte deste grupo, sendo parte integrante do *TensorFlow*, mantido pelo *Google Brain* (ABADI et al., 2016). Este *framework* oferece uma ampla gama de modelos pré treinados e ferramentas de treinamento, avaliação e exportação de modelos de detecção de objetos, além de ter compatibilidade com diferentes formatos de entrada de pesos, com integração nativa ao *TensorFlow* (SINGH et al., 2021).

Neste estudo, optou-se pelo *framework* YOLO (*You only Look Once*) (REDMON et al., 2016), devido á sua ampla documentação de configuração e uso, baixa latência de inferência e desempenho para cenários com menor amostragem (REDMON et al., 2016). Além disso, o YOLO possui implementação oficial via biblioteca *ultralytics*, com ampla documentação para o treinamento dos modelos, ajustes de hiperparâmetros e exportação de formatos (REDMON et al., 2016). Neste estudo, optou-se pelo YOLOv8 por oferecer um bom equilíbrio entre facilidade de aplicação, velocidade de treinamento e desempenho robusto, além de incorporar refinamentos em relação a versões anteriores: como cabeças *decoupled*, *label assignment* dinâmico, *defaults* de treinamento melhor calibrados que versões anteriores, aprimoramentos nas augmentações e maior estabilidade na convergência do modelo, que tendem á elevar sua acurácia sem aumentar excessivamente a complexidade.

Ademais, a implementação oficial (via *ultralytics*) disponibiliza pesos pré-treinados, pipeline de *fine-tuning* e boas ferramentas de exportação, que reduzem o atrito experimental. Versões mais novas do YOLO, embora potencialmente incrementais em performance e consequentemente em métricas, tendem a ser mais pesadas e exigem preparação da base de dados ligeiramente diferente em relação á versão 8, o que tende á aumentar o custo computacional e de engenharia para a mesma finalidade de detecção, tornando o YOLOv8 a escolha pragmática dentro das limitações de capacidade de treinamento disponíveis.

3.2.1 YOLOv8

O *framework* YOLOv8 (*You Only Look Once*, versão 8), representa uma das evoluções recentes do modelo de detecção de objetos e uma única etapa, em vez de separar o processo em dois estágios (proposta de regiões e refinamentos), como arquiteturas baseadas em R-CNN. Desta forma, para uma dada imagem de entrada, o YOLOv8 processa, em uma única passagem de inferência, a posição, tamanho da caixa delimitadora, a confiança de presença de objeto e a probabilidade das classes presentes (REDMON et al., 2016). A Figura 3 ilustra de maneira geral a arquitetura do YOLO.

Em suma, a arquitetura do *framework* YOLOv8 pode ser dividida em 3 blocos

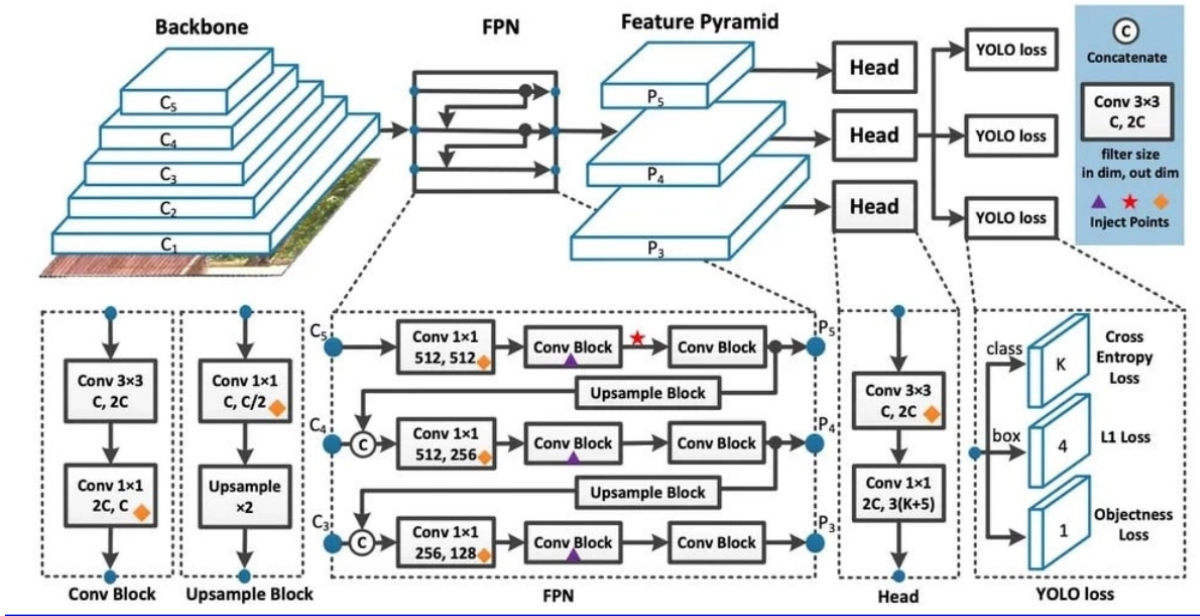


Figura 3 – Representação da Arquitetura do YOLOv8 de forma resumida.

Fonte: Imagem adaptada de yolov8.org.

principais, sendo o *Backbone*, *Neck* e *Detection Head*, com mais um bloco adicional de segmentação, denominado *Segmentation Head* (TERVEN; CÓRDOVA-ESPARZA; ROMERO-GONZÁLEZ, 2023).

O bloco **BackBone** da arquitetura do YOLOv8 é responsável por aplicar a extração de características, de maneira a receber a imagem de entrada e transformá-la em um mapa de atributos, a partir dos padrões visuais variados, são geradas várias camadas. Na primeira camada, aplica-se a operação Focus, que realoca blocos espaciais 2×2 para a dimensão de canais, aumentando a profundidade do tensor e compactando detalhes sem elevar significativamente o custo computacional. O objetivo é enriquecer a retenção de detalhes sem aumentar significativamente o custo computacional requerido. Cada uma das camadas de atributos transformadas pelo *Backbone* segue a sequência (YOLOv8.org, 2025):

$$\text{Convolução} \rightarrow \text{BatchNorm} \rightarrow \text{SiLU (Swish)}, \quad (3.1)$$

Na qual cada convolução extrai filtros locais, a normalização (*BatchNorm*) busca estabilizar os valores e a ativação SiLU tenta garantir o melhor fluxo de gradiente. À medida em que a rede aprofunda suas camadas, a resolução espacial dos mapas diminui e o número de canais aumenta. Desta forma, as camadas mais rasas capturam bordas, texturas e cores, enquanto as camadas mais profundas normalmente representam conceitos semânticos, como contornos e formatos (YOLOv8.org, 2025).

Apesar de ter um grande número de camadas, a arquitetura do CSP (*Cross Stage Partial*) empregada no *backbone* divide o fluxo de atributos em dois caminhos, e os re-agrega posteriormente, de forma a reduzir redundâncias de gradiente, e diminuir a quantidade

total de parâmetros, se causar prejuízo na qualidade de suas representações (YOLOv8.org, 2025).

O bloco **Neck** da arquitetura do YOLOv8 representa a fusão multiescala, na qual os mapas de características produzidos no bloco de *Backbone*, com diferentes resoluções espaciais, são combinados de maneira a formar um mapa único chamado C2f-PAN. Em suma, algumas camadas de atributos operam com mapas reduzidos para detecção de objetos grandes, e outras camadas, operam em mapas maiores para detecção de objetos médios ou pequenos (YOLOv8.org, 2025).

Em suma, durante o processo de fusão multiescala do bloco **Neck**, os fluxos de *bottom-up* e *top-down* são aplicados e fundido no mapa C2f-PAN. O fluxo *top-down* atua para mapas com maior profundidade, de alto nível semântico e baixa resolução, sendo projetados para resoluções maiores, semelhante ao conceito aplicado no *Feature Pyramid Network* (FPN); Já o fluxo *bottom-up* atua para mapas com menor profundidade, resolução mais elevada, conseqüentemente uma maior retenção de detalhes espaciais, retornando informações para as camadas profundas segundo a lógica *Path Aggregation Network* (PAN) (YOLOv8.org, 2025).

O último bloco de detecção da arquitetura YOLOv8 é o **Detection Head**, responsável pela geração das caixas delimitadoras de detecção a partir do mapa de características gerado pelo *Neck*. Este último bloco divide a imagem em uma grade de $S \times S$ células, sendo o valor de S variado de acordo com a camada de saída aplicada. Em cada célula, existem as respectivas coordenadas (i, j) de cada um dos três níveis de escala (maior resolução, intermediária, menor resolução). O modelo produz um vetor de dimensão $5 + C$, na qual C representa o número de classes existentes na respectiva célula. Os cinco primeiros componentes do $5 + C$ representam $\{t_x, t_y, t_w, t_h, t_o\}$. Sendo t_x e t_y representam a posição exatado centro do objeto de detecção dentro da célula; t_w e t_h representam os valores de largura e altura da caixa delimitadora na célula; t_o representa o log de objetividade da detecção, ficando entre 0 e 1 ao passar pelo $\sigma(t_o)$, oque significa a confiança da detecção á cerca da existência do objeto detectado. Os outros C componentes do $5 + C$ são logs brutos, correspondentes á distribuição de probabilidade de cada possível classe, após a aplicação de *softmax* sobre os mesmos (YOLOv8.org, 2025).

Para a conversão das coordenadas do $5 + C$ em coordenadas absolutas na imagem, uma vez que os valores são inicialmente obtidos a partir do seccionamento da imagem em uma grade $S \times S$, emprega-se o seguinte procedimento de normalização (YOLOv8.org, 2025).

Para a obtenção das coordenadas absolutas de localização, ou deslocamento, x_{pred} e y_{pred} , é aplicada a função *sigmoide* nas coordenadas t_x e t_y da grade (YOLOv8.org, 2025):

$$x_{\text{pred}} = \sigma(t_x) + i, \quad y_{\text{pred}} = \sigma(t_y) + j, \quad (3.2)$$

Na qual (i, j) correspondem às coordenadas da célula. A largura e altura w_{pred} e h_{pred} , podem ser obtidas com aplicação de função exponencial em t_w e t_h (YOLOv8.org, 2025):

$$w_{\text{pred}} = a_w \exp(t_w), \quad h_{\text{pred}} = a_h \exp(t_h), \quad (3.3)$$

Na qual (a_w, a_h) são âncoras iniciais, que podem ser definidas manualmente ou calculadas automaticamente, como por exemplo através do método K-means sobre as dimensões dos objetos do conjunto de treinamento, aprendidas pelo *framework* durante esta etapa (YOLOv8.org, 2025).

O coeficiente de objetividade, obtido por $\sigma(t_o)$, passa a servir como uma pontuação, ou *score*, que determina a probabilidade de haver um objeto de interesse dentro da caixa delimitadora, e as probabilidades de classe ($p(c)$), são obtidas pela normalização dos C s do $5 + C$ (YOLOv8.org, 2025):

$$p(c) = \text{softmax}(\{p_c\}). \quad (3.4)$$

Após concluída a geração das caixas delimitadoras e seus respectivos *scores*, para os objetos de interesse detectados na imagem, é aplicado método Non-Maximum Suppression (NMS) para eliminar caixas delimitadoras redundantes, que estão significativamente sobrepostas, mantendo apenas as caixas delimitadoras com um determinado limiar de *score* (YOLOv8.org, 2025).

Em síntese, o *framework* YOLOv8 executa em sua arquitetura as três funções essenciais, sendo a Extração de atributos pelo bloco **BackBone**; A fusão multiescala para geração do mapa de características, obtendo contexto e detalhes, pelo bloco **Neck**; E a predição direta das caixas delimitadoras e suas classes pelo bloco **Detection Head**. Esse processo ocorre de maneira rápida e precisa pelo *framework*, após o devido treinamento (TERVEN; CÓRDOVA-ESPARZA; ROMERO-GONZÁLEZ, 2023).

Vale ressaltar ainda, que o YOLOv8 é hábil para realizar a detecção do objeto pela construção do mapa binário, que podem ser utilizado na segmentação, por uma camada adicional de **Segmentation Head**, que recebe os mesmos mapas de características do *Neck* e utiliza camada de convolução seguidas por operações de *upsampling* para a construção de mapas de alta resolução, na qual cada ponto do mapa de resolução gerado indica a probabilidade do pixel pertencer ao objeto detectado, de maneira a formar a máscara binária da região de interesse, e pode ser utilizada na segmentação da mesma (TERVEN; CÓRDOVA-ESPARZA; ROMERO-GONZÁLEZ, 2023).

O YOLOv8 conta com diferentes versões de arquitetura, como a *nano*, *small*, *medium*, *large* e *xlarge*. Com cada versão variando em mAP, velocidade de operação e quantidade de parâmetros, sendo indicados de acordo com a situação, dada o balanço entre precisão de detecção e velocidade de resposta/treinamento. A implementação oficial está disponível na biblioteca Ultralytics, que simplifica o treinamento, avaliação e inferência, incluindo suporte a GPUs e exportação para formatos como ONNX e TensorRT (TERVEN; CÓRDOVA-ESPARZA; ROMERO-GONZÁLEZ, 2023).

3.3 Técnicas de Processamento Digital de Imagem

A segmentação é uma etapa fundamental em qualquer sistema de visão computacional que visa computacional na qual se deseja realizar alguma análise posterior, como reconhecimento de padrões, medição de áreas internas, detecção de anomalias e extração/classificação de atributos (MINAEE et al., 2021). Em suma, a segmentação consiste no particionamento da imagem em regiões homogêneas, ou na separação da região de interesse na imagem da região de fundo, isto é, o resto da mesma (ZHU et al., 2016).

Para fins de entendimento, as técnicas de segmentação podem ser agrupadas de acordo com seu princípio de funcionamento, como: Segmentação baseada em Limiarização (HU et al., 2024), que busca separar regiões por meio de critérios de intensidade de nível de cinza; Segmentação baseada em Bordas (CSURKA et al., 2022), que exploram descon- tinuidades de intensidades de nível de cinza e buscam detectar contornos com operadores e filtros; Segmentação baseada em regiões (DAGHER; ABBOUD, 2024), que propagam regiões crescentes iterativamente a partir de uma semente inicial; Segmentação baseada em agrupamento (DAGHER; ABBOUD, 2024), que agrupam pixels com características seme- lhantes em espaço, intensidade ou textura; Segmentação baseada em Contornos (BARBU; TĂNASE, 2025), que ajustam as bordas da região baseada em métodos de level set ou snakes deformáveis, dentre outros tipos de segmentação. Muitos destes modelos estão presentes, ou são implementados com o auxílio da biblioteca OpenCV (ADUSUMALLI et al., 2021).

A metodologia de segmentação pode ser baseada em princípios matemáticos, físicos, de mudança de domínio, e até mesmo em princípios da natureza, como no caso do *Divisible Cell Segmentation* (SOUZA, 2023), o método recente, abordado na segmentação deste estudo.

3.3.1 *Divisible Cell Segmentation*

O método *Divisible Cell Segmentation*, proposto por (SOUZA, 2023), aplica os conceitos da biologia da divisão celular mitótica por meio das Células Sintéticas Computacionais aplicadas no interior da região de interesse. O método foi inicialmente testado em

imagens de ensaio, conceitualmente elaboradas. Mas ao testar em bases de segmentação de AVC e Pulmão, os resultados foram satisfatórios.

O método necessita de auxílio para inicialização, com a aplicação da célula sintática computacional em alguma localização interna da região de interesse a ser segmentada.

No *Divisible Cell Segmentation*, a caixa delimitador pode ser dada como (Bb), definindo a imagem de trabalho $\text{IMg}(p)$, que corresponde á área interna da caixa na imagem total, expresso como:

$$\text{IMg}(p) = I_i \cup \text{Bb}, \quad (3.5)$$

Na qual I_i representa o corte definido pela caixa delimitadora imagem do exame (SOUZA, 2023).

O centróide é então obtido a partir das coordenadas da caixa delimitadora formada, através da equação:

$$\mathbf{C} = \left(\frac{x_{\min} + x_{\max}}{2}, \frac{y_{\min} + y_{\max}}{2} \right), \quad (3.6)$$

Em que $\mathbf{C}$ representa o centroide, $(x_{\min}, y_{\min})$ e $(x_{\max}, y_{\max})$ definem os cantos da Bb (SOUZA, 2023).

$\mathbf{C}$ é normalmente o ponto de aplicação da primeira célula sintética computacional, C_i . As células sintéticas computacionais irão posteriormente ser geradas por processo de divisão, e se desloração em direções distintas, em busca do que seria o limiar da região de interesse com a região de fundo na imagem, dentro da região definida pela caixa delimitadora inicial da inicialização do método (SOUZA, 2023).

em suma, a ideia central do algoritmo é a propagação de inúmeras células sintéticas a partir de C_i . Cada célula possui seu vetor de direção $\mathbf{V}(D)$, que aponta para direções distintas, guiando a movimentação da célula para as prováveis fronteiras de borda de segmentação. A equação abaixo simplifica este processo de movimento iterativo (SOUZA, 2023):

$$C_i \leftarrow C_i + \mathbf{V}(D), \quad (3.7)$$

O $\mathbf{V}(D)$ pode ser ajustado para cobrir diferentes direções, variando de forma automática no processo de geração de uma nova célula sintética computacional. A movimentação dá-se continuidade até o momento em que o nível de cinza da localização atual da célula ultrapassa um determinado valor de limiar $\mathbf{V}(L)$, por limite superior ou inferior ao valor estabelecido, dependendo da configuração (SOUZA, 2023).

$$I_i(x_i, y_i) \leq \text{VL} \quad \text{ou} \quad I_i(x_i, y_i) \geq \text{VL} \quad (3.8)$$

Em que $I_i(x_i, y_i)$ indica o valor de nível de cinza do pixel localizado na região de coordenadas x_i, y_i .

Vale ressaltar que as células sintéticas computacionais possuem função de Divisão Celular **DC** recursiva em um dado intervalo de tempo:

$$DC = C_i + \mathbf{V}(D) \quad (3.9)$$

Na qual o processo de Divisão Celular **DC** gera novas células sintéticas computacionais a partir da célula inicial C_i , com seu respectivo vetor de direção $\mathbf{V}(D)$. Δ é um valor de pequena magnitude que espalha as células filhas geradas pela divisão, C_{i_1} e C_{i_2} , que herdam parte da direção da célula progenitora, com pequeno ajuste angular de $\pm\theta$, que garante que as células irão em direções distintas. A divisão ocorre até o momento em que a célula atinge o critério de parada, $\lambda(C_i)$, obtido pelo encontro com $\mathbf{V}(L)$ (SOUZA, 2023):

$$\lambda(C_i) = \begin{cases} 1, & \text{se } I_i(x_i, y_i) \leq VL \quad \text{ou} \quad I_i(x_i, y_i) \geq VL, \\ 0, & \text{caso contrário,} \end{cases} \quad (3.10)$$

Desta forma, uma célula só prossegue em sua movimentação, caso $\lambda(C_i) = 0$. E Quando $\lambda(C_i) = 1$, a célula sintética computacional em questão estaciona na imagem, encerra seu processo de divisão, marca sua posição pelas coordenadas (x_i, y_i) como um ponto integrante da fronteira da região de interesse a ser segmentada. Ao preencher um polígono fechado, as demais células existentes em movimentação e busca do valor limiar são descartadas, obtendo somente a posição das células de borda. O resultado esperado é uma espécie de "pontilhado" da região de interesse (SOUZA, 2023).

Por fim, para a conexão dos pontos, existem algumas técnicas de fechamento do mesmo, que buscam delimitar as linhas de segmentação a partir de um ponto com o seu mais próximo. O método aplicado inicialmente por (SOUZA, 2023) para conexão dos pontos de coordenadas do polígono "pontilhado" foi desenvolvido com base na transformada de Hough (MUKHOPADHYAY; CHAUDHURI, 2015). Mas obteve dificuldades em situações de conexão de pontos mais completos, como em superfícies irregulares ou "rugosas", com alta densidade de pontos próximos. Mas métodos como a transformada de concave Hull poderiam vir a cumprir esta tarefa de fechamento e obtenção do polígono fechado a partir dos pontos (SOUZA, 2023).

Em suma, o *divisible cell segmentation* exprime na saída uma imagem binária com o fundo preto (pixels com valor 0), e apenas os pontos provenientes das posições de células sintéticas computacionais de borda são brancos (pixels com valor 255). E essa imagem pontilhada é então utilizada para formar a máscara binária de segmentação através de métodos de conexão entre os pontos de maneira a formar o polígono fechado. Métodos baseados na transformada de Hough, Hull convexo, Hull côncavo e demais modelos de conexão de pontos, podem ser implementados para realização desta construção do polígono.

3.3.2 Conexões de Coordenadas de Pontos – *Concave Hull*

Os métodos de Hull são técnicas fundamentadas na Geometria computacional, utilizadas para determinar a "casca" que envolve um determinado conjunto de pontos no espaço (PREPARATA; SHAMOS, 2012). O *Convex Hull* consiste na construção do polígono convexo de menor área que contém todos os pontos do espaço, tentando garantir que qualquer segmento de reta que ligue dois pontos do Hull permaneça inteiramente dentro dessa região, geralmente através dos algoritmos de Graham Scan, Jarvis March e Quickhull (PREPARATA; SHAMOS, 2012).

Já o *concave Hull* se baseia na fronteira mais ajustada aos dados, de maneira a atribuir cavidades e formas não convexas que acompanham a distribuição dos pontos, esse método geralmente se baseia por aproximação via α Shapes ou pelos K vizinhos mais próximos (EDELBRUNNER; MÜCKE, 1994; MOREIRA; SANTOS, 2007). A Figura 4 Apresenta a diferença entre os resultados para o Convex hull e Concave hull aplicados em um mesmo conjunto de pontos.

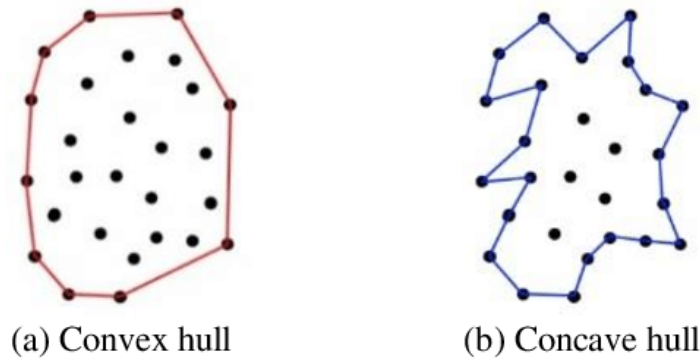


Figura 4 – Demonstração de resultados pelos métodos de convex Hull e Concave Hull.

Fonte: Imagem Adaptada de (YAHYA et al., 2017).

Para a conexão do pontos de contornos na imagem binária de pontos, o método realiza a detecção de segmentos lineares e curvos dominantes através da transformada de Hough para uma estimativa global de orientação, que facilita a montagem de segmentos preliminares de contorno (HASSANEIN et al., 2015).

Em seguida, o fechamento inicial do polígono é feito através do Casco Convexo (*convex hull*), na qual busca-se construir o menor polígono convexo H_c a partir do conjunto de pontos P da imagem. Esse processo é representado por:

$$\text{Conv}(P) = \bigcap \{C \subseteq \mathbb{R}^2 \mid C \text{ é convexo e } P \subseteq C\}. \quad (3.11)$$

Na qual o operador $\cap$ realiza a interseção de todos os conjuntos convexas $C \subseteq \mathbb{R}^2$ que contém P , resultando no menor conjunto convexo que engloba P (MOUNT, 2021).

Após a criação do polígono convexo, utiliza-se o conceito de *Alpha Shape* para a criação do polígono côncavo, na qual as cavidades internas são ajustadas para seguir a

geometria espacial do polígono. Seja $\alpha > 0$ um parâmetro de concavidade, define-se um conjunto filtrado de triângulos a partir da Triangulação de Delaunay (LUO; MI; TAO, 2021), $DT(P)$, do conjunto P :

$$\mathcal{T}_\alpha = \{t \in DT(P) \mid R(t) \leq \frac{1}{\alpha}\}, \quad (3.12)$$

Em que $R(t)$ é o raio do círculo circunscrito ao triângulo t . Para o parâmetro α , o casco côncavo é a fronteira do complexo simplicial resultante:

$$H_\alpha(P) = \partial\left(\bigcup_{t \in \mathcal{T}_\alpha} t\right). \quad (3.13)$$

À medida em que o parâmetro α diminui, os triângulos com raio maior são descartados de maneira a permitir que os detalhes e cavidades mais profundas, na formação do polígono de casco côncavo (LUO; MI; TAO, 2021). Para valores maiores de α , obtêm-se novamente o casco convexo:

$$\lim_{\alpha \rightarrow \infty} H_\alpha(P) = \text{Conv}(P). \quad (3.14)$$

Em suma, o casco convexo gera o menor polígono convexo que engloba todos os pontos P no plano, enquanto o casco côncavo relaxa a propriedade de convexidade do polígono, buscando a captura dos seus recortes internos, buscando representar mais fielmente a superfície interna da forma do objeto (MOUNT, 2021).

Para atenuação de irregularidades, e supressão de ruídos na conexão entre os pontos, é possível aplicar métodos de suavização baseada em *Kernels*, atribuindo pesos w_i a cada ponto p_i dentro do conjunto P , promovendo uma possível distribuição mais uniforme de vértices ao longo do casco côncavo $H_\alpha(P)$ (GUIBAS, 2021).

3.3.3 Data Fusion - Interpolação das Modalidades de RM

Para a fusão das modalidades T1-CE, T2, FLAIR e T1 em uma única imagem interpolada, o foi-se desenvolvido o pipeline que busca extrair a fatia, empilhar canas, selecionar e interpolar os mesmos, seguido por um processo de normalização.

Para cada volume tridimensional $M_i(x, y, z)$ da modalidade $i \in \{1, 2, 3, 4\}$ (T1-CE, T2, FLAIR, T1), busca-se extrair a fatia de índice k :

$$M_i^k(x, y) = M_i(x, y, k), \quad (x, y) \in \Omega \subset \mathbb{Z}^2. \quad (3.15)$$

Em seguida, aplica-se o empilhamento de canais, construindo um tensor multicanal via *early fusion*:

$$C(x, y, i) = M_i^k(x, y), \quad i = 1, 2, 3, 4, \quad (3.16)$$

Então, aplica-se a máscara binária de segmentação $S(x, y, z)$:

$$S^k(x, y) = S(x, y, k) \in \{0, 1\}, \quad (3.17)$$

e aplica-se:

$$C'(x, y, i) = C(x, y, i) \cdot S^k(x, y). \quad (3.18)$$

Em seguida, é realizada a seleção de canais, descartando-se o quarto canal para adaptar o resultado á entrada esperada pela CNN:

$$D(x, y, i) = C'(x, y, i), \quad i = 1, 2, 3. \quad (3.19)$$

Após isto, é então aplicado o processo de Interpolação espacial para re-amostrar a imagem D na frade de pixels alvo $\{(u_p, v_q)\}$. Usa-se o operador bilinear $\mathcal{I}$:

$$\widetilde{D}(u_p, v_q, i) = \mathcal{I}[D(\cdot, \cdot, i)](u_p, v_q), \quad (3.20)$$

onde

$$\mathcal{I}[f](u, v) = \sum_{m=0}^1 \sum_{n=0}^1 f(\lfloor u \rfloor + m, \lfloor v \rfloor + n) w_m(u - \lfloor u \rfloor) w_n(v - \lfloor v \rfloor), \quad (3.21)$$

com $w_0(t) = 1 - t$ e $w_1(t) = t$. Por fim, aplica-se a função de normalização ϕ com o auxilio da biblioteca Keras para cada canal:

$$F(u_p, v_q, i) = \phi(\widetilde{D}(u_p, v_q, i)). \quad (3.22)$$

Gerando o tensor resultante

$$F : \{(u_p, v_q)\} \times \{1, 2, 3\} \rightarrow \mathbb{R} \quad (3.23)$$

que é então a imagem multicanal final usada pelos extratores.

Em termos de operadores compostos, o pipeline pode ser resumido como sendo:

$$F = \phi \circ \mathcal{I} \circ \Pi_3 \circ M_{\text{stack}} \circ M_{\text{slice}}, \quad (3.24)$$

onde:

- M_{slice} extrai M_i^k ;
- M_{stack} empilha as quatro modalidades;
- Π_3 seleciona os três primeiros canais;
- $\mathcal{I}$ reamostra espacialmente;
- ϕ normaliza.

3.4 Classificadores de Aprendizado de Máquina

A aprendizagem de máquina, ou *Machine Learning*, é o campo da inteligência artificial que estuda sobre algoritmos capazes de aprender padrões e auxiliar na tomada de decisões a partir de um conjunto de dados, sem necessariamente serem programados para cada uma dessas tarefas ([SARKER, 2021](#)).

Em linhas gerais, a aprendizagem de máquina pode ser categorizada sob três paradigmas, sendo a Aprendizagem supervisionada, quando o algoritmo recebe a amostra na forma de classes y e atributos x , sendo treinado para representar $f(x) \cong y$ (SARKER, 2021); A Aprendizagem Não Supervisionada, que trabalha apenas com os atributos x , buscando agrupar os dados sem que haja rótulos definidos (SARKER, 2021); E a Aprendizagem por Reforço, que consiste na interação de um agente com um determinado ambiente, na qual o mesmo irá aprender a política de ações para maximizem o retorno acumulado de resposta a partir de recompensas ou punições no comportamento do agente (SHAKYA; PILLAI; CHAKRABARTY, 2023).

Neste estudo, foram utilizados os classificadores de Aprendizagem Supervisionada: KNN (K-vizinhos mais próximos), MLP (perceptron multicamadas), Naive Bayes, Random Forest e SVM (máquina de vetores de suporte) (NAYEEM et al., 2021). Os classificadores serão utilizados para realização de combinações entre modelos de extração e classificação, afim de se obter a melhor combinação comparativa para com a base de dados abordada em relação às métricas de classificação adotadas neste estudo.

3.4.1 KNN

O classificador K-Vizinhos mais próximos (KNN), é um classificador não paramétrico de aprendizado supervisionado que busca atribuir a classe de uma amostra de entrada com base nas classes de suas K amostras de treino mais próximas no espaço de atributos através do uso de métrica de distância entre as amostras (ANAVA; LEVY, 2016). Como a distância de Minkowski de ordem P :

$$d_p(x_i, x_j) = \left(\sum_{m=1}^n |x_{im} - x_{jm}|^p \right)^{1/p}, \quad (3.25)$$

Na qual, quando $p = 2$, a fórmula se transforma em distância Euclidiana, comumente usada na aplicação do K-vizinhos mais próximos (ANAVA; LEVY, 2016):

$$d(x_i, x_j) = \sqrt{\sum_{m=1}^n (x_{im} - x_{jm})^2}. \quad (3.26)$$

Sendo o conjunto dos k vizinhos mais próximos da nova amostra x :

$$N_k(x) = \{j : x_j \text{ está entre os } k \text{ menores } d_p(x, x_j)\}. \quad (3.27)$$

Sendo a predição da classe $\hat{y}$ da nova x definida por votação majoritária:

$$\hat{y}(x) = \arg \max_{c \in \mathcal{Y}} \sum_{j \in N_k(x)} \mathbf{1}(y_j = c), \quad (3.28)$$

Em que $\mathbf{1}(\cdot)$ representa a função de indicação, e $\mathcal{Y}$ o conjunto de classes existentes (ANAVA; LEVY, 2016). Adicionalmente, é possível aplicar ponderação inversa á distância para o

conjunto de dados de teste:

$$\hat{y}(x) = \arg \max_{c \in \mathcal{Y}} \sum_{j \in N_k(x)} w_j \mathbf{1}(y_j = c), \quad w_j = \frac{1}{d_p(x, x_j)}. \quad (3.29)$$

Desta forma, a escolha de k equilibra o viés e variância, com valores pequenos de k reduzindo o viés, mas aumentando a variância, e valores maiores de k fazendo o inverso. Existe uma regra empírica que recomenda a adoção de k como sendo $k \approx \sqrt{N}$, em função do número de amostras de treinamento N . Mas é possível aplicar métodos de busca de hiper parâmetros para alcançar o valor ideal de K de acordo com as características do Dataset abordado (ANAVA; LEVY, 2016).

3.4.2 Naive Bayes

O classificador *Naive Bayes* é um classificador paramétrico de aprendizado supervisionado que se baseia no teorema de Bayes, pautada na hipótese ingênua de independência condicional entre os atributos de uma amostra (PAJILA et al., 2023). Desta forma, dada uma amostra qualquer $x = (x_1, x_2, \dots, x_n)$, o objetivo do classificador é estimar a classe C_k de maneira a maximizar a probabilidade posterior:

$$P(C_k | x) = \frac{P(C_k) P(x | C_k)}{P(x)}. \quad (3.30)$$

de acordo com a hipótese da dependência condicional dos atributos:

$$P(x | C_k) = \prod_{i=1}^n P(x_i | C_k). \quad (3.31)$$

Então, é possível descrever a regra de decisão do *Naive Bayes* como sendo:

$$\hat{C} = \arg \max_k P(C_k) \prod_{i=1}^n P(x_i | C_k) = \arg \max_k \left[\log P(C_k) + \sum_{i=1}^n \log P(x_i | C_k) \right]. \quad (3.32)$$

Na qual, $P(C_k)$ representa a probabilidade *a priori* da classe C_k , $P(x_i | C_k)$ representa a verossimilhança do atributo x_i na classe C_k e $P(x)$ representa a probabilidade marginal, que é constante para todas as classes, sendo desprezada no processo de decisão (PAJILA et al., 2023).

Existem versões do *Naive Bayes*, adaptadas para lidar com as diferentes naturezas dos dados, como o Gaussian NB: $P(x_i | C_k) = \frac{1}{\sqrt{2\pi\sigma_{k,i}^2}} \exp\left(-\frac{(x_i - \mu_{k,i})^2}{2\sigma_{k,i}^2}\right)$, cujo x_i é dado como contínuo e com distribuição normal. E o Bernoulli NB: $P(x_i | C_k) = p_{k,i}^{x_i} (1 - p_{k,i})^{1-x_i}$, que é indicado para atributos binários (JOHN; LANGLEY, 2013).

Em síntese, o *Naive Bayes* é um classificador simples e eficiente, muito aplicado para cenários de alta dimensionalidade, mesmo que a suposição de independência ingênua do algoritmo seja muitas vezes não satisfeita de maneira estrita (PAJILA et al., 2023).

3.4.3 MLP

O classificador Perceptron Multicamadas, ou *MultiLayer Perceptron* - MLP, é também um classificador paramétrico de aprendizado supervisionado, construído a partir de uma cadeia de camadas de neurônios artificiais, sendo normalmente a arquitetura composta por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída (JOHN; LANGLEY, 2013). Em seu funcionamento, cada neurônio de uma dada camada l realiza uma combinação linear de ativações da camada anterior, seguida de uma função de ativação $\phi(\cdot)$:

$$\begin{aligned} z^{(l)} &= W^{(l)} a^{(l-1)} + b^{(l)}, \\ a^{(l)} &= \phi(z^{(l)}), \end{aligned} \quad (3.33)$$

Na qual $W^{(l)} \in \mathbb{R}^{d_l \times d_{l-1}}$ representa a matriz de pesos da camada l , $b^{(l)} \in \mathbb{R}^{d_l}$ representa o vetor de vieses da camada l , $a^{(l-1)}$ representa o vetor de ativações da camada anterior ($a^{(0)} = x$) e $\phi(\cdot)$ representa a função de ativação, podendo ser Relu, Sigmoide ou Tanh (JOHN; LANGLEY, 2013).

Após a propagação direta (*feed-forward*), é possível obter o vetor de escores $z^{(L)}$ na camada de saída L . Ao aplicar a função *softmax* sobre esta saída, é possível obter as estimativas de probabilidade para cada classe:

$$\hat{C} = \arg \max_k \text{softmax}(z^{(L)})_k = \arg \max_k a_k^{(L)}. \quad (3.34)$$

O treinamento desse classificador consiste no ajuste dos parâmetros $\theta = \{W^{(l)}, b^{(l)}\}_{l=1}^L$ das camadas, com finalidade de minimizar a entropia cruzada entre as previsões e as classes verdadeiras y de uma dada amostra de entrada:

$$J(\theta) = - \sum_{k=1}^K y_k \log(a_k^{(L)}). \quad (3.35)$$

A atualização dos parâmetros θ é feita pelo princípio de *backpropagation* e gradiente descendente:

$$\theta \leftarrow \theta - \eta \nabla_{\theta} J(\theta), \quad (3.36)$$

Na qual η representa a taxa de aprendizado do classificador (JOHN; LANGLEY, 2013).

Em suma, o MLP é um classificador com a quantidade de parâmetros definida de acordo com o número de camadas e neurônios da arquitetura. Esse classificador, se bem parametrizado e configurado, é capaz de modelar relações complexas entre os dados com boa eficácia (JOHN; LANGLEY, 2013).

3.4.4 Random Forest

O *Random Forest* é um classificador não paramétrico de aprendizagem supervisionada que se baseia no princípio da árvore de decisão, construindo uma "floresta" de árvores

de decisão T , sendo essas independentes entre si, $h_1, \dots, h_T$. Cada árvore é treinada em uma amostragem distinta do conjunto total de dados de treino $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$. Desta forma, em cada nó de cada árvore de decisão, isto é, apenas um determinado subconjunto aleatório de dados de m das p variáveis é considerado para uma divisão, para reduzir a correlação entre as árvores e buscar melhorar a generalização deste classificador (LOUPPE, 2014).

Após treinado, o método realiza regressão via predição agregada com média simples:

$$\hat{y}(x) = \frac{1}{T} \sum_{t=1}^T h_t(x). \quad (3.37)$$

Para a classificação, o *Random Forest* usa votação por maioria:

$$\hat{y}(x) = \arg \max_{c \in \mathcal{Y}} \sum_{t=1}^T \mathbf{1}(h_t(x) = c). \quad (3.38)$$

O classificador também se utiliza de atenuação da variância:

$$\text{Var}[\hat{y}(x)] = \rho \sigma^2 + \frac{1 - \rho}{T} \sigma^2, \quad (3.39)$$

Na qual σ^2 representa a variância individual de cada árvore e ρ representa a correlação média entre os pares de árvore. Desta forma, à medida em que a floresta T cresce, o segundo termo tende a zero, o que reduz erros na generalização do classificador (LOUPPE, 2014).

Além disso, para cada variável j tem sua importância quantificada quando se soma os ganhos de impureza, ΔI , em todos os nós na qual j foi utilizado:

$$\text{Imp}(j) = \sum_{t=1}^T \sum_{\substack{\text{nó } n \in h_t: \\ \text{split em } j}} \Delta I(n). \quad (3.40)$$

Desta forma, o princípio de funcionamento do *Random forest*, busca o tornar mais eficiente a ruído, sendo capaz de realizar intenções não lineares, sem que haja necessidade de ajuste fino de muitos hiper parâmetros para isso (LOUPPE, 2014).

3.4.5 SVM

O classificador de vetores de suporte, ou *Support Vector Machine* - SVM, é um classificador não paramétrico de aprendizagem supervisionada que busca traçar um hiperplano ótimo para separar as classes, com a maior margem possível entre ambas, no espaço de atributos. Para melhor visualizar esta divisão, o classificador geralmente aplica uma transformação ϕ no espaço dos atributos via operações com *Kernel* (BISHOP; NASRABADI, 2006). Para casos lineares suaves, o problema primal seria representado da seguinte forma:

$$\min_{w,b,\{\xi_i\}} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^n \xi_i, \quad \text{sujeito a} \quad y_i(w^\top \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, n, \quad (3.41)$$

Na qual $C > 0$ controla o equilíbrio entre as margens e penalização de erros ξ_i (BISHOP; NASRABADI, 2006). Através do Lagrangeano dual, é possível obter:

$$\max_{\{\alpha_i\}} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j), \quad 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^n \alpha_i y_i = 0, \quad (3.42)$$

Na qual $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$ representa a função *kernel* (BISHOP; NASRABADI, 2006). A decisão de classe para uma nova amostra qualquer x é dada por:

$$f(x) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b\right). \quad (3.43)$$

Dentre os kernels comumente aplicados no SVM, estes são:

- Linear: $K(x, x') = x^\top x'$,
- Polinomial: $K(x, x') = (x^\top x' + r)^d$,
- RBF (Gaussiano): $K(x, x') = \exp(-\gamma \|x - x'\|^2)$.

Os vetores de suporte são as amostras cujo $\alpha_i > 0$. O hiperplano ótimo busca maximizar a margem $2/\|w\|$. E os ajustes do valor de C , do tipo e parâmetros do kernel aplicado, permitem controlar a complexidade e capacidade de generalização e atuação do modelo (BISHOP; NASRABADI, 2006).

3.4.6 Grid Search

A maioria dos algoritmos de aprendizado de máquina possui hiperparâmetros configuráveis para ajustar sua classificação ao conjunto de dados abordado. Esses hiperparâmetros são selecionados manualmente, empiricamente ou por meio de algoritmos de busca, como a busca em grade (ALIBRAHIM; LUDWIG, 2021), que seleciona automaticamente os melhores hiperparâmetros de um intervalo de acordo com o conjunto de dados abordado. A Tabela 2 mostra os intervalos adotados para cada classificador.

Em suma, dado um conjunto p de hiperparâmetros, é possível definir para cada hiperparâmetro um conjunto discreto de valores, representados por (BISCHL et al., 2023):

$$\Theta_j = \{\theta_{j,1}, \theta_{j,2}, \dots, \theta_{j,n_j}\}, \quad j = 1, \dots, p. \quad (3.44)$$

Então, o espaço total de combinações destes hiperparâmetros, que varia com base na sua dimensionalidade, é dado por (BISCHL et al., 2023):

$$\Theta = \Theta_1 \times \Theta_2 \times \dots \times \Theta_p. \quad (3.45)$$

Tabela 2 – Tabela de range de hiperparâmetros adotados no Grid Search para os classificadores no banco de dados Brats2020.

Classificador	Parâmetros de Busca	Intervalos de Busca
Naive Bayes	-	-
MLP	Camadas ocultas	[2, 1001]
Nearest Neighbors	Número de vizinhos	1, 3, 5, 7, 9, 11, 13, 15
Random Forest	Profundidade máxima	6, Ilimitada
	Bootstrap	True, False
	Criterion	Gini, Entropia
SVM Linear	Parâmetro de Regularização (C)	$2^x x \in [-5, 15]$
SVM	Grau	3, 5, 7, 9
Polynomial	Parâmetro de Regularização (C)	$2^x x \in [-5, 15]$
SVM RBF	Gamma	$2^x x \in [-15, 3]$
	Parâmetro de Regularização (C)	$2^x x \in [-5, 15]$

Para cada $\theta = (\theta_1, \dots, \theta_p) \in \Theta$, o *Grid Search* avalia via validação cruzada K -fold a métrica de interesse J :

$$\bar{J}(\theta) = \frac{1}{K} \sum_{k=1}^K J(M_{\text{treino}}^{(k)}(\theta), D_{\text{val}}^{(k)}). \quad (3.46)$$

Então, é possível selecionar o hiperparâmetro ótimo de acordo com a métrica de interesse (BISCHL et al., 2023):

$$\theta^* = \arg \max_{\theta \in \Theta} \bar{J}(\theta), \quad (3.47)$$

E então, o modelo é treinado definitivo com todos os dados com θ^* .

3.5 Redes Neurais Convolucionais

O Aprendizado de Máquina, ou *Deep Learning*, é um subcampo do Aprendizado de Máquina que busca aplicar conceitos computacionais avançados, como as redes neurais artificiais profundas, para aprender automaticamente as representações de auto nível a partir do conjunto de dados brutos (BENGIO et al., 2009). As múltiplas camadas hierárquicas de processamento das redes neurais aplicam transformações não lineares, podendo ser convoluções, ativações ou normalizações, e alimentam as camadas seguintes (CHEN et al., 2020). Desta forma, as camadas iniciais normalmente extraem características de baixo nível da imagem de entrada, como bordas e textura, enquanto as camadas mais profundas combinam esses atributos extraídos, convertendo-os em conceitos complexos,

como objetos ou padrões (ZHAO; SONG; ERMON, 2017). O treinamento pode ser feito por retro propagação de erro, com ajuste de milhões de parâmetros em simultâneo na busca da minimização da função de custo, oque geralmente permite ao modelo generalizar tarefas diversas sem a necessidade de engenharia manual de características para tal (SHUKLA; CHOUDHARY, 2022).

A Figura 5 apresenta como exemplo a arquitetura LeNET, desenvolvida por Yann LeCun em 1988, sendo uma das redes pioneiras no uso de camadas de convolução com filtros de ordem 5×5 e passo 1. Além disso, esta arquitetura teve papel fundamental no reconhecimento de dígitos manuscritos (LECUN et al., 2002).

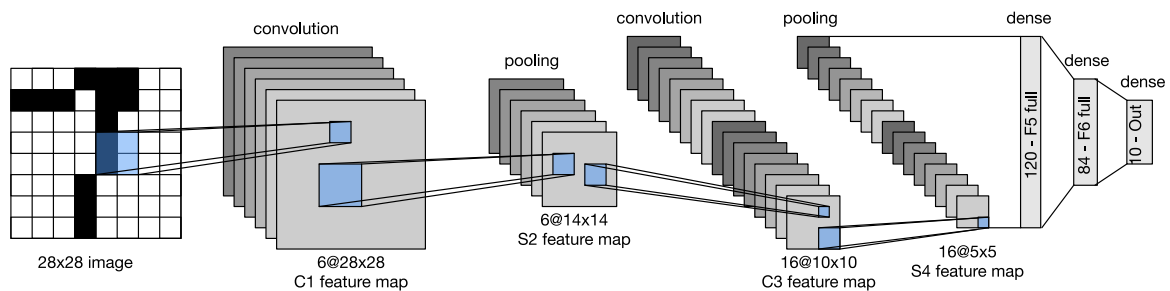


Figura 5 – Exemplo de arquitetura CNN, modelo LeNET.

Fonte: Imagem adaptada de Lab9.

As redes neurais convolucionais, CNN, são uma classe de arquiteturas *Deep Learning* projetadas para processar dados estruturados no formato de grades regulares (YAMASHITA et al., 2018). De forma resumida, cada camada convolucional da arquitetura da CNN realiza uma função específica sobre a imagem, podendo ser um conjunto de filtros, ou convoluções, com a finalidade de se gerar mapas de ativação capazes de capturar padrões na imagem, seja textura, bordas ou formas (ZAFAR et al., 2022). Após a geração dos mapas de ativação, geralmente são aplicadas operações de *pooling*, que buscam reduzir a resolução dos mapas de ativação para ampliar a invariância á pequenas translações, e reduzir a dimensionalidade (ZHAO et al., 2024).

As CNNs são capazes de aprender hierarquias de características com um conjunto de parâmetros por filtro, graças ao compartilhamento de pesos á conectividade local (YAMASHITA et al., 2018). Seu funcionamento e capacidade de lidar com os dados as tornam candidatas ideais em muitos sistemas de visão computacional que buscam classificar e segmentar imagens de forma semântica (GARCIA-GARCIA et al., 2017).

3.5.1 VGG

A VGG 16 é uma CNN advinda de uma família de arquiteturas desenvolvidas pelo *Visual Geometry Group*, da universidade de Oxford. Esse modelo em particular é composto por 13 camadas convolucionais de filtros 3×3 , organizadas em cinco blocos e intercalados por camadas de *pooling* 3×3 . Após as camadas convolucionais, há 3 camadas totalmente

conectadas, e uma camada de saída *softmax*. Desta forma, esta arquitetura, com auxílio do uso consistente de filtragem, é capaz de realizar a extração gradual de características hierárquicas (SIMONYAN; ZISSERMAN, 2014). A Figura 6 ilustra uma representação da Arquitetura da VGG.

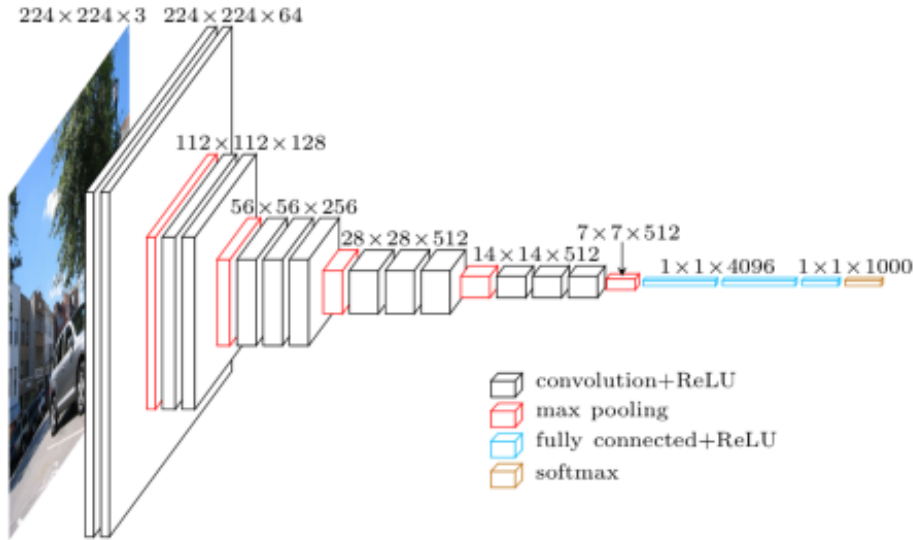


Figura 6 – Arquitetura VGG
Fonte: Imagem adaptada de Lab9.

A operação de convolução com soma de Bias em uma das camadas convolucionais do VGG-16 pode ser descrita de forma expandida da seguinte maneira:

$$z_{i,j,k}^{(l)} = \sum_{m=1}^F \sum_{n=1}^F \sum_{c=1}^{C^{(l-1)}} W_{m,n,c,k}^{(l)} \cdot a_{i+m-1,j+n-1,c}^{(l-1)} + b_k^{(l)} \quad (3.48)$$

Na qual $z_{i,j,k}^{(l)}$ representa o valor de saída antes da ativação na posição (i, j) do canal k . $W^{(l)}$ representa os filtros convolucionais da camada l , $a^{(l-1)}$ representam as ativações da camada anterior e $b^{(l)}$ representa os vieses (RYU, 2024). Após a soma dos bias, aplica-se a função de ativação ReLU:

$$a_{i,j,k}^{(l)} = \text{ReLU} \left(z_{i,j,k}^{(l)} \right) \quad (3.49)$$

Em notação vetorial simplificada, temos:

$$a^{(l)} = \text{ReLU} \left(W^{(l)} * a^{(l-1)} + b^{(l)} \right) \quad (3.50)$$

Na qual $*$ representa a convolução 2D, $W^{(l)}$ representa os filtros da camada l , $b^{(l)}$ representam os vieses e $a^{(l)}$ representa o mapa de ativação resultante (RYU, 2024).

A dimensão do mapa de saída de cada respectiva convolução é dada por:

$$H^{(l)} = \frac{H^{(l-1)} - F + 2P}{S} + 1, \quad W^{(l)} = \frac{W^{(l-1)} - F + 2P}{S} + 1, \quad (3.51)$$

Na qual $H^{(l-1)}, W^{(l-1)}$ representa a altura e largura da entrada, $F = 3$ representa o tamanho do filtro, $P = 1$ é o *padding* e $S = 1$ o *stride* (ZHANG et al., 2023).

Após o quinto filtro do bloco de *pooling*, as ativações são *flattened*, formando um vetor de características $\phi(x) \in \mathbb{R}^D$, que alimenta as camadas densas:

$$z = V \phi(x) + c, \quad \hat{y}_k = \text{softmax}(z)_k = \frac{\exp(z_k)}{\sum_{j=1}^C \exp(z_j)}, \quad (3.52)$$

Na qual $V \in \mathbb{R}^{C \times D}$ e $c \in \mathbb{R}^C$ representam os pesos e vieses da camada final C classes, e $\hat{y}$ representa o vetor de probabilidades preditas (CHOWDHURY, 2024).

Na estrutura de *Transfer Learning*, é possível aproveitar o vetor $\phi(x)$ extraído pela VGG-16 nas primeiras camadas, substituindo apenas as camadas densas finais por um classificador $g(V \phi(x) + c)$. Dessa forma, é possível obter bom desempenho de classificação, ao passo em que a quantidade de parâmetros necessárias é reduzida drasticamente durante a etapa de treino (SIMONYAN; ZISSERMAN, 2014).

3.5.2 DenseNet

A DenseNet é uma CNN de uma família de arquiteturas de redes caracterizadas por suas conexões densamente conectadas, na qual camada recebe a concatenação de todas as ativações das camadas anteriores como entrada, e passa sua ativação como saída. Esse processo busca melhorar a propagação de gradientes, e promover a reutilização dos atributos de maneira mais eficiente (HUANG et al., 2017). A Figura 7 ilustra uma representação da Arquitetura da DenseNet.

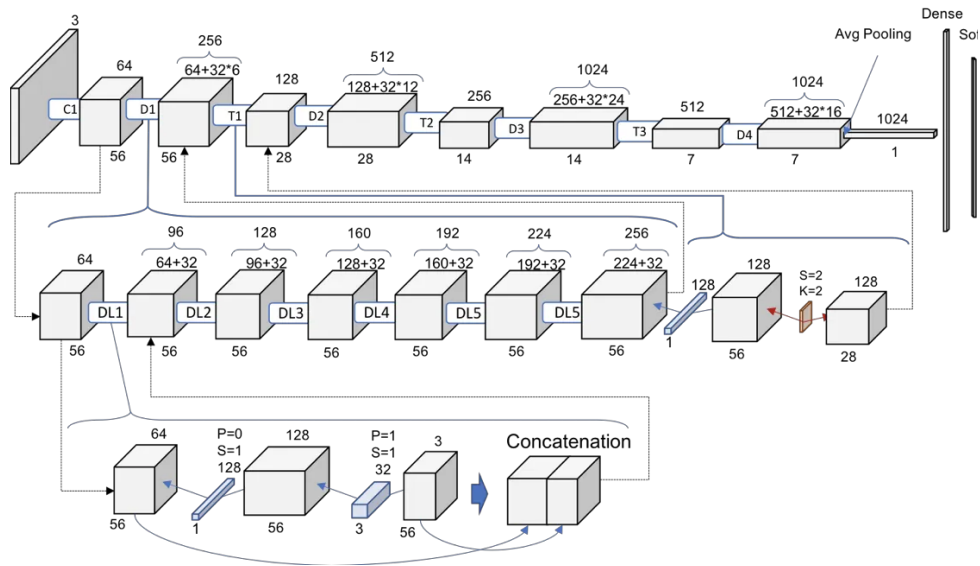


Figura 7 – Arquitetura DenseNet
Fonte: Imagem adaptada de [medium](#).

De maneira formal, toma-se $a^{(l)} \in \mathbb{R}^{d_l \times H_l \times W_l}$ como sendo o mapa de ativação da l -ésima camada (HUANG et al., 2017). Desta forma:

$$a^{(l)} = H_l\left([a^{(0)}, a^{(1)}, \dots, a^{(l-1)}]\right), \quad (3.53)$$

Na qual $[\cdot]$ denota concatenação ao longo da dimensão de canais:

$$H_l(X) = \text{Conv}(\text{ReLU}(\text{BN}(X))), \quad (3.54)$$

Sendo BN a normalização, ReLU e a convolução sendo aplicados em série (HUANG et al., 2017). Durante a aplicação, cada camada adiciona um número fixo k de mapas *growth rate*, de forma que o total de canais na camada l é dado por:

$$d_l = d_0 + k \cdot l, \quad (3.55)$$

Sendo d_0 o número de canais da camada inicial (HUANG et al., 2017).

Uma camada de transição é inserida entre os blocos densos para que seja controlado o crescimento dimensional e realização de *down-sampling*:

$$x' = \text{Pool}_{2 \times 2}(\text{Conv}_{1 \times 1}(x)), \quad (3.56)$$

Na qual $\text{Conv}_{1 \times 1}$ reduz o número de canais, via fator de compressão $\theta \in (0, 1]$, e o $\text{Pool}_{2 \times 2}$ realiza o *pooling*, *stride* 2 (HUANG et al., 2017).

Ao final do último bloco denso da arquitetura, aplica-se um *global average pooling* e uma camada totalmente conectada para se realizar a classificação:

$$\hat{y} = \text{softmax}(W \text{GAP}(a^{(L)}) + b). \quad (3.57)$$

Dessa forma, essa estrutura compacta busca maior eficiência de conexões, reduzindo o número de parâmetros e favorecendo o fluxo de informações ao longo da rede (ZHOU et al., 2022).

3.6 Modelos de Processamento de Linguagens natural em larga escala (LLMs)

Os modelos de Linguagem em Larga escala, ou *Large Language Models* - LLM, são redes neurais baseadas geralmente na arquitetura *Transformer*, que são treinadas com enormes quantidades de dados textuais para resolver diversas tarefas de processamento de linguagem natural (VASWANI et al., 2017). Os LLM, em seu funcionamento, aplicam tokenização para as palavras de entrada, transformando-as em um vetor de *embeddings* $e_i \in \mathbb{R}^d$, aplicando um mecanismo de auto-atenção para computar dependências contextuais:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V, \quad (3.58)$$

Na qual $Q = EW^Q$, $K = EW^K$ e $V = EW^V$ representam as projeções dos embeddings $E = [e_1, \dots, e_n]$ e d_k é a dimensão das chaves. O mecanismo de atenção permite que cada token incorpore informações de todos os outros no mesmo nível do *Transformer* (BAHDANAU; CHO; BENGIO, 2014).

O processo de treinamento das LLMs busca minimizar a perda de entropia cruzada sobre a tarefa de previsão do próximo token. Dada uma distribuição verdadeira y , e uma distribuição $\hat{y}$ predita pelo *transformer*, sobre um vocabulário de entrada de tamanho V , a função de perda para cada posição do vocabulário é dado por:

$$\mathcal{L}_{\text{CE}} = - \sum_{i=1}^V y_i \log(\hat{y}_i). \quad (3.59)$$

As LLM também empregam *positional encodings* para realizar a diferenciação da ordem dos tokens:

$$\text{PE}(p, 2i) = \sin\left(\frac{p}{10000^{2i/d}}\right), \quad \text{PE}(p, 2i + 1) = \cos\left(\frac{p}{10000^{2i/d}}\right), \quad (3.60)$$

Na qual p representa a posição do token na sequência e i representa o índice do componente de *embedding* (RUSH, 2018).

Em suma, a combinação do processo de tokenização das palavras de entrada, uso combinado de *embeddings*, auto-atenção e empilhamento de blocos *Transformer*, permite aos modelos LLMs a captura de relações semânticas e contextuais entre as palavras para gerar textos coerentes e contextualizados com base na entrada (BROWN et al., 2020).

3.6.1 Modelo Llama 3

O Llama 3 faz parte da terceira geração de Grandes Modelos de Linguagem da Meta, lançado em abril de 2024. A arquitetura do Llama 3 possui variantes com o número de parâmetros, e segue o padrão *transformer* de decodificador, introduzindo um mecanismo *Grouped-Query Attention* (GQA) com a finalidade de reduzir o custo de atenção durante a inferência e tornar o modelo mais escalável para cenários de aplicação (AI, 2024b). A Figura 8 ilustra de maneira geral a arquitetura do Llama3.

O modelo foi alimentado com mais de 15 trilhões de tokens de texto público durante a etapa de pré-treinamento, um equivalente á cerca de 7 vezes o volume usado no pré treinamento de sua versão anterior (AXIOS, 2024). Os dados de treino contam com heurística, remoção de conteúdo inadequado e deduplicação semântica para maximização de qualidade dos dados (AI, 2024b).

O Llama 3 foi projetado com suporte de janela de contexto de cerca de 8 000 tokens na configuração base, mas versões de instrução sintonizadas, como o Llama 3.1, estendem esse limite para cerca de 128 000 tokens, sendo avaliados em oito idiomas com *benchmarks* em 30 línguas (AI, 2024a).

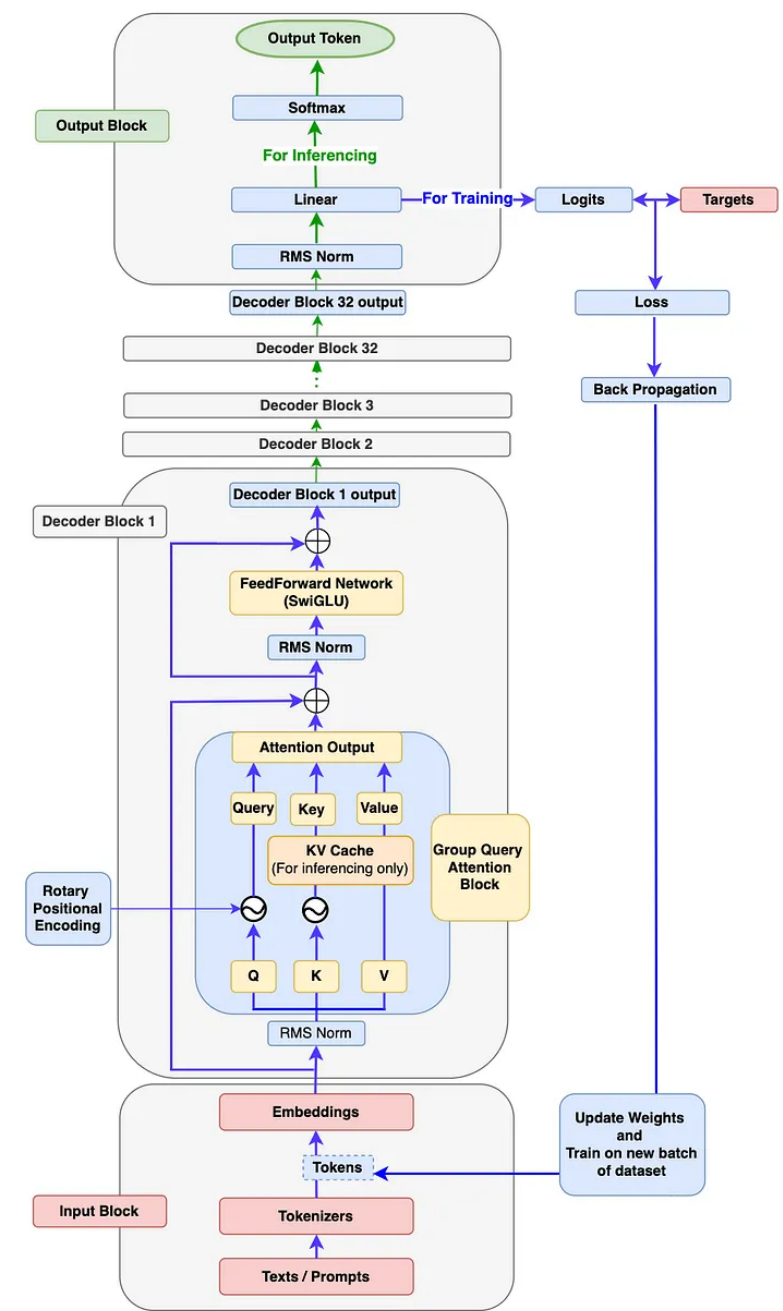


Figura 8 – Arquitetura geral Llama3
 Fonte: Imagem adaptada de [medium](#).

O modelo encontra-se disponível *Open source* diretamente no [repositório da Meta AI](#) e no [Hugging Face Hub](#), e pode ser consumido também via serviços de nuvem, como o [AWS Bedrock](#), o que favorece sua adoção em pesquisas na comunidade de IA ([SERVICES, 2024](#)).

3.7 Métricas de Avaliação

Para avaliar os resultados obtidos pelos métodos propostos neste estudo para as diferentes etapas propostas, afim de se obter a melhor combinação que será utilizada para novos dados de entrada em futuras aplicações em produção, foram abordadas métricas comumente utilizadas em diversos estudos para averiguação da segmentação e classificação. Para a segmentação, foram adotadas as métricas: Acurácia (ACC), Valor Preditivo Negativo (VPN), Valor Preditivo Positivo (VPP), Sensibilidade (SEN), Especificidade (ESP); e para a segmentação, foram adotadas as métricas: Acurácia (ACC), Precisão (PRE), Recall (REC), e F1-Score (F1) (TAHA; HANBURY, 2015).

3.7.1 Métricas de Segmentação

As Métricas de segmentação são geralmente obtidas a partir da comparação entre as máscaras binárias de segmentação obtida como resultado do modelo, e a máscara de segmentação Padrão Ouro, ou *Ground Truth* - GT.

A comparação entre as máscaras ocasiona em uma comparação entre pixels, nas quais os pixels da máscara segmentada são classificados de acordo com sua relação com o GT, conforme observado na Figura 9.

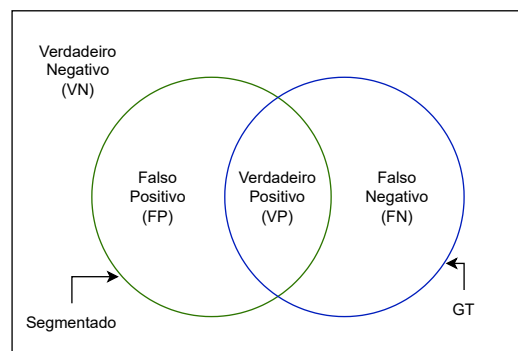


Figura 9 – Representação das classes possíveis para os pixels de métrica de segmentação.

Fonte: Imagem adaptada de [university of LouisVille](#).

As quatro classes de pixels contabilizadas depois da comparação são:

- Verdadeiro Positivo (VP): número de pixels (ou instâncias) corretamente classificados como pertencentes à classe de interesse (região alvo)
- Falso Positivo (FP): número de pixels (ou instâncias) classificados como pertencentes à classe de interesse, mas que na verdade são de fundo (ou outra classe).
- Verdadeiro Negativo (VN): número de pixels (ou instâncias) corretamente classificados como não pertencentes à classe de interesse (fundo).

- Falso Negativo (FN): número de pixels (ou instâncias) que pertencem à classe de interesse, mas foram incorretamente classificados como fundo.

A precisão é uma métrica de classificação e está diretamente relacionada ao número de pixels que foram segmentados corretamente em relação ao número total de pixels segmentados em uma imagem (TAHA; HANBURY, 2015). Sua fórmula é dada por:

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.61)$$

A métrica Valor Preditivo Negativo (VPN) indica a proporção de pixels identificados corretamente como pertencentes à região de fundo em relação ao número total de pixels identificados como tal (TAHA; HANBURY, 2015):

$$VPN = \frac{VN}{VN + FN} \quad (3.62)$$

A métrica Valor Preditivo Positivo (VPP) avalia a proporção de pixels identificados corretamente como regiões de interesse em relação ao número total de pixels identificados como tal (TAHA; HANBURY, 2015):

$$VPP = \frac{VP}{VP + FP} \quad (3.63)$$

Sentitividade (SEN), também conhecida como Recall, é uma métrica de classificação responsável por avaliar a proporção de pixels corretamente segmentados como pertencentes à região de interesse entre todos os pixels que realmente pertencem a esse grupo. Na classificação, a métrica avalia o número de rótulos corretamente classificados como verdadeiros positivos entre os pixels efetivamente classificados nesses grupos (TAHA; HANBURY, 2015). Sua fórmula é dada por:

$$SEN = \frac{VP}{VP + FN} \quad (3.64)$$

ESP (Especificidade) é a proporção de pixels identificados corretamente como pertencentes à região de fundo em relação ao número total de pixels reais na região de fundo (TAHA; HANBURY, 2015):

$$ESP = \frac{VN}{VN + FP} \quad (3.65)$$

3.7.2 Métricas de classificação

As métricas de classificação são calculadas de maneira semelhante, com as quatro classes contabilizadas, mas não a partir dos pixels, como na segmentação, e sim na

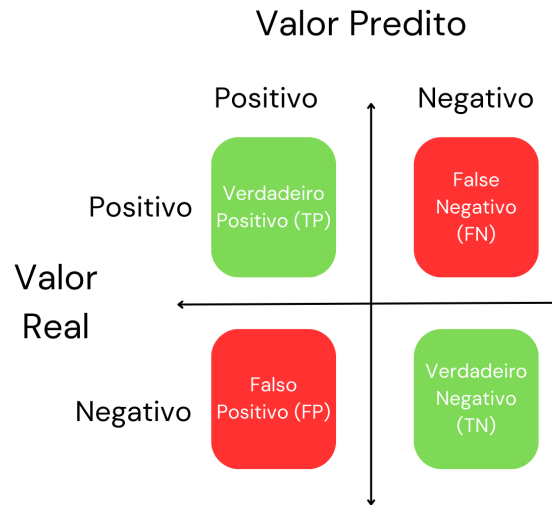


Figura 10 – Representação das classes possíveis para as predições, exemplo de classificação binária.

Fonte: Imagem adaptada de [Medium](#).

comparação entre as classes preditas pelos modelos com as classes reais dos dados de teste, conforme observado na Figura 10.

As quatro classes de predições de teste realizadas na comparação são:

- Verdadeiro Positivo (VP): número de predições corretamente classificados como pertencentes à classe de interesse.
- Falso Positivo (FP): número de predições classificadas como pertencentes à classe de interesse, mas que na verdade não são.
- Verdadeiro Negativo (VN): número de pixels predições corretamente classificados como não pertencentes à classe de interesse.
- Falso Negativo (FN): número de predições que pertencem à classe de interesse, mas foram incorretamente classificados como de outra classe.

A Acurácia (ACC), na classificação, é uma métrica de classificação diretamente relacionada ao número de predições que foram corretamente realizada, dentre o número total de predições ([TAHA; HANBURY, 2015](#)):

$$ACC = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.66)$$

A Precisão (PRE), na classificação, busca quantificar a fração total de predições positivas que de fato são verdadeiras positivas ([TAHA; HANBURY, 2015](#)):

$$PRE = \frac{VP}{VP + FP} \quad (3.67)$$

O Recall (REC), na classificação, busca avaliar a capacidade do modelo de prever corretamente todas as instâncias positivas, sendo esta métrica, uma proporção de verdadeiros positivos dentre todos os casos realmente pertencentes à classe positiva (TAHA; HANBURY, 2015):

$$\text{REC} = \frac{VP}{VP + FN} \quad (3.68)$$

A métrica F1 Score, é uma média harmônica entre a Precisão e o Recall, na busca de oferecer uma medida de equilíbrio entre as métricas, e lidar com situações de desbalanceamento de classes (TAHA; HANBURY, 2015):

$$\text{F1} = 2 \times \frac{\text{PRE} \times \text{REC}}{\text{PRE} + \text{REC}} \quad (3.69)$$

3.8 Teste estatístico de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov é um procedimento não paramétrico para quantificar a similaridade entre as distribuições subjacentes de duas amostras, baseado na maior diferença entre suas distribuições acumuladas empíricas. O teste compara a distribuição empírica das amostras com a distribuição teórica, geralmente uma distribuição normal. A hipótese nula é que as duas distribuições são idênticas, enquanto a hipótese alternativa é que elas são diferentes. O resultado do teste é uma estatística de teste que representa a diferença máxima entre as duas distribuições (PRÆSTGAARD, Semptember, 1995).

A hipótese nula do teste de Kolmogorov-Smirnov (H_0) afirma que não há diferença entre as distribuições cumulativas das duas amostras. A hipótese alternativa (H_1), por sua vez, é que as distribuições diferem em pelo menos um ponto. A decisão sobre a hipótese nula é tomada comparando a estatística do teste com um valor crítico que depende do tamanho das amostras e do nível de significância desejado. A rejeição da hipótese nula indica evidência estatística de que as distribuições das duas amostras são diferentes (PRÆSTGAARD, Semptember, 1995).

Sejam duas amostras ordenadas de maneira crescente, $X = \{x_1, x_2, \dots, x_{n_1}\}$ e $Y = \{y_1, y_2, \dots, y_{n_2}\}$, para cada ponto x , calcula-se a diferença absoluta entre as funções de distribuição acumuladas empíricas das duas amostras da seguinte forma:

$$F_1(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{1}_{\{x_i \leq x\}}, \quad F_2(x) = \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{1}_{\{y_j \leq x\}}, \quad (3.70)$$

Na qual $\mathbf{1}_{\{\cdot\}}$ representa a função indicadora e a estatística de teste D é definida como sendo a maior diferença absoluta entre ambas as funções:

$$D = \max_x |F_1(x) - F_2(x)| \quad (3.71)$$

Desta forma, a estatística representa a maior diferença vertical entre ambas as distribuições acumuladas. E quanto maior for D , maior será a evidência contra a hipótese nula H_0 (SMIRNOV, 1939).

Para permitir a comparação com valores críticos tabelados, sobretudo para amostras de tamanhos diferentes, a estatística pode ser normalizada da seguinte forma:

$$D^* = D \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad (3.72)$$

Na qual o valor de D^* é comparado com um valor crítico D_α , que varia de acordo com o nível de significância α adotado no teste. Desta forma, se $D^* > D_\alpha$, a hipótese nula H_0 é rejeitada, indicando que as distribuições são estatisticamente diferentes (SMIRNOV, 1939).

4 Metodologia

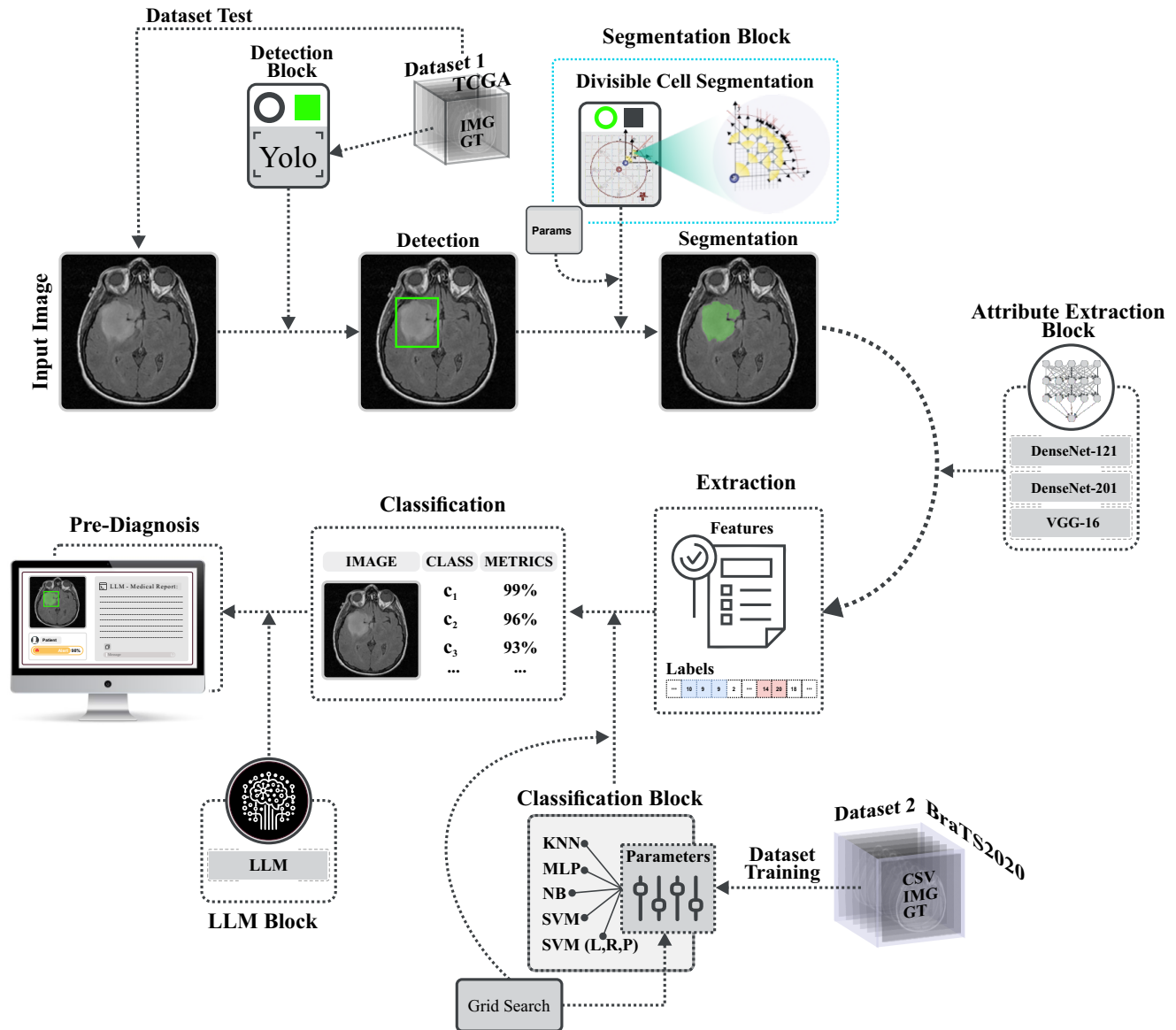


Figura 11 – Ilustração Geral da metodologia desenvolvida neste estudo, com Detecção baseada no YoloV8, treinado sob o Dataset TCGA; Segmentação baseada no *Divisible Cell Segmentation*, Extração de atributos com arquiteturas CNN, Classificação com pipeline de combinação e de busca de hiperparâmetros ótimos, treinado com o dataset BraTS2020; Seguida pela construção de pré diagnóstico textual com LLM.

Fonte: O próprio Autor.

Neste capítulo, é apresentada a metodologia do método proposto: uma nova abordagem com inicialização e detecção baseada em *Deep Learning* para a detecção; Segmentação de tumores cerebrais em imagens médicas de ressonância magnética, seguido pelo processo de classificação com um pipeline de seleção de melhor combinação de Extrator e classifica-

dor com busca de hiperparâmetros para classificação. O novo método, composto por células computacionais sintéticas, visa identificar as bordas do tumor, segmentando-o por meio de conexões celulares entre pixels, chamadas Células Computacionais Inteligentes (ICC). Este método é inspirado no processo biológico de citocinese, onde uma célula biológica se divide, multiplicando-se em outras células. Além disso, um pipeline de extração e classificação combinando aprendizado de transferência com busca em grade foi desenvolvido para esta aplicação, culminando em um diagnóstico sofisticado assistido por LLM por meio dos rótulos classificados. A Figura 11 representa o resumo do fluxo desenvolvido neste estudo, cada etapa desenvolvida será melhor discutida, com auxílio da Figura 12, que busca trazer mais detalhes às etapas.

Para desenvolver a aplicação, os classificadores e a busca em grade foram implementados utilizando as bibliotecas Python Scikit Learn (KRAMER; KRAMER, 2016), enquanto os extratores CNN foram desenvolvidos com o auxílio da biblioteca TensorFlow (DEVELOPERS, 2022). As operações de ajuste e leitura de dados foram realizadas utilizando as bibliotecas Numpy (HARRIS et al., 2020) e Pandas (REBACK et al., 2020), e a leitura de imagens foi realizada utilizando a biblioteca OpenCV (ADUSUMALLI et al., 2021). Tanto o Scikit Learn quanto o TensorFlow utilizam GPU para otimizar os processos. O framework YoloV8 foi implementado com auxílio da biblioteca *Ultralytics* (RAHIM; YUAN; BARABADY, 2025). O modelo LLM abordado foi o Llama3 (AI, 2024b).

SiDSECNLP – Sistema de Detecção, Segmentação, Extração e Classificação + NLP

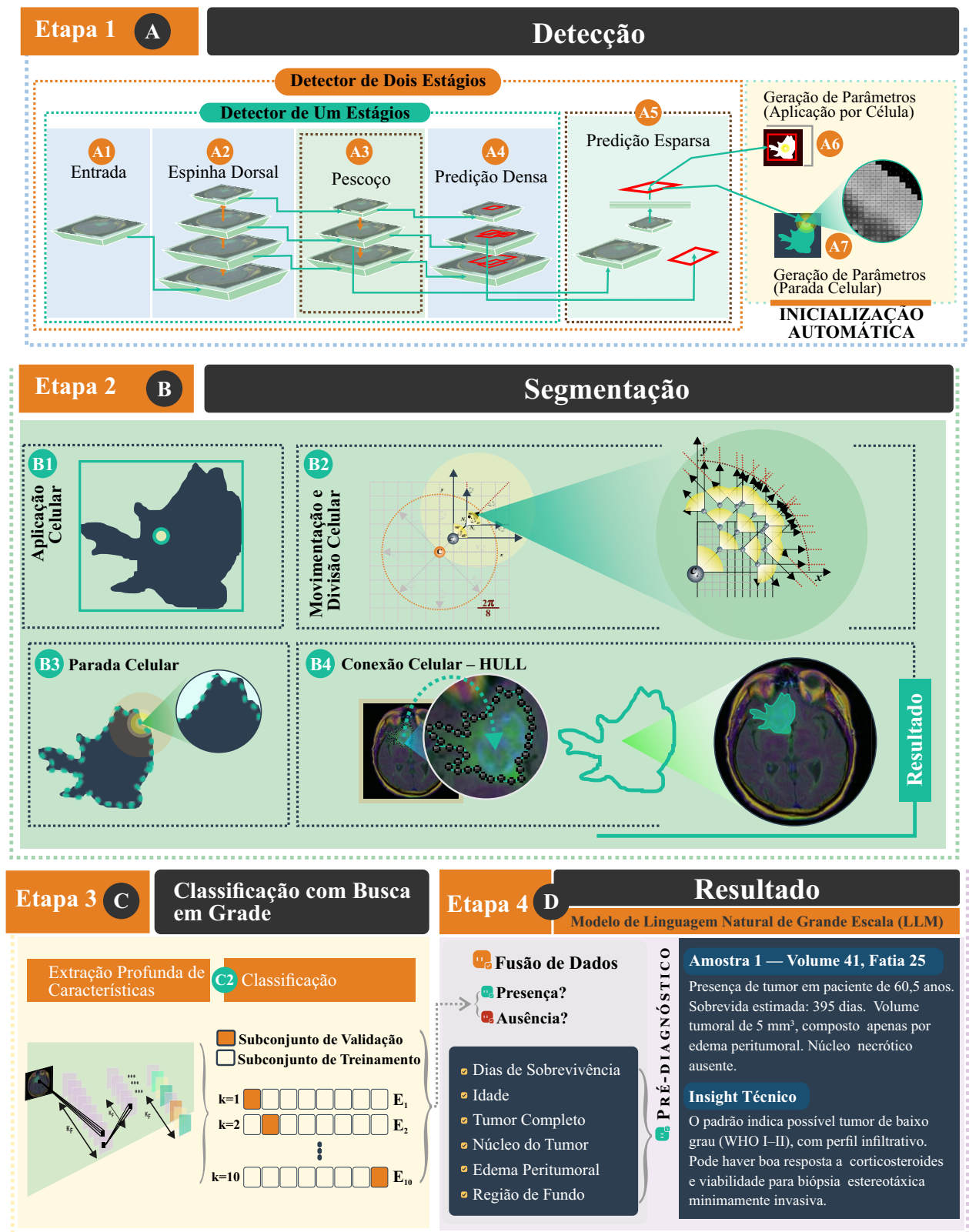


Figura 12 – Metodologia proposta, Etapa 1 - Detecção de tumor cerebral em imagem de ressonância magnética, Etapa 2 - Segmentação do tumor cerebral, Etapa 3 - Classificação dos atributos do tumor cerebral, Etapa 4 - Resultados e pré-diagnóstico formalizados por LLM.

A Figura 12 apresenta a metodologia proposta. Em (A1), a imagem média é inserida; Em (A2), os atributos profundos são extraídos; Em (A3), as regiões de interesse são propostas; Em (A4), as regiões de interesse são combinadas; em (A5), a região de interesse é detectada e a caixa delimitadora definitiva é gerada; Em (A6), o método de segmentação é iniciado. Em (A7), as células presentes nas bordas da região de interesse auxiliam na formação do polígono, formando a máscara de segmentação binária. A Etapa 1B ilustra o processo que ocorre entre (A6) e (A7).

Em (B1), a célula semente do método é aplicada à região de interesse localizada dentro da caixa delimitadora de detecção; em (B2), as células sintéticas computacionais se movimentam e se dividem dentro da região da lesão na imagem; em (B3), as células que atingem a borda do tumor na imagem são estacionadas; finalmente, em (B4), as conexões celulares entre as células localizados nas bordas são feitas através do método *Convex Hull*, ilustrado como resultado de (B4). A máscara binária de segmentação é então aplicada na terceira etapa, na qual foi montada uma estrutura de *pipeline* de extração/classificação com busca de hiperparâmetros. Em C1 ocorre a extração com as arquiteturas CNN. Em C2 ocorre a classificação com modelos de *machine learning*. Os atributos classificados pela melhor combinação de extrator/classificador e hiperparâmetros é então utilizado para obtenção dos atributos clínicos da imagem que serão utilizados para montar o pré diagnóstico textual com LLM no estágio 4.

Etapa 1 (A): Detecção - *Framework* YOLOv8 - a etapa 1 (A) diz respeito à aplicação e o funcionamento da arquitetura YOLOv8 na detecção do tumor nas imagens de ressonância magnética. O modelo foi treinado para construir a caixa delimitadora da região, e a máscara binária de detecção, que serão cruciais para as demais etapas do método proposto.

Etapa 1 (A1): Uso de Imagens Médicas de Ressonância Magnética - Nesta etapa, é apresentado o processo de treinamento do YoloV8 a partir do Dataset TCGA, foram utilizadas cerca de 3.064 imagens. Não foi aplicado pré processamento ou *data augmentation* além da conversão da imagem para tons de cinza. A divisão padrão dos *conjuntos de dados* é de 80% para teste e 20% para treinamento, foi utilizada validação cruzada com 3 folds para avaliação dos resultados de detecção, e os parâmetros de treinamento foram: Máx. 1.000 iterações, batch size = 2 imagens e 2 workers. Esta etapa corresponde a (A1) na estagio 1A da Figura 12.

Etapa 1 (A2): Backbone / Extração de Características Nesta etapa, a imagem passa por processo de redimensionamento, normalização e augmentations internas. Em seguida, o *backbone* da arquitetura do Yolo, com variação de **CSPDarknet** e blocos residuais com função de ativação SiLU. Cada respectivo bloco CSP busca combinar as convoluções de ordem 3x3 e 1x1 para extração das representações enquanto reduz a duplicação de gradientes e melhora a eficiência computacional. Cada bloco residual pode

ser representado como:

$$\mathbf{y} = \mathbf{x} + F(\mathbf{x}; W), \quad (4.1)$$

Na qual $\mathbf{x}$ representa o tensor de entrada, F as camadas convolucionais do bloco, e W seus pesos. Esta etapa é representada em (A2) na Figura 12.

Estágio 1 (A3): Cabeça Anchor-Free / Proposição de Regiões O Yolov8 se utiliza de uma metodologia *anchor-free head*, que opera sobre uma grade de células a partir de uma grade de células no mapa de características extraído. Em cada célula i , o *framework* busca prever os deslocamentos $\Delta x, \Delta y, \Delta w, \Delta h$ da caixa delimitadora, com *objectness score* p_{obj} e probabilidades de classe p_{cls} . Esta etapa é representada em (A3) na Figura 12.

Estágio 1 (A4): Neck / Combinação de Multiescala O *Neck* da arquitetura YOLO é construído a partir da combinação de FPN (*Feature Pyramid Network*) e PAN (*Path Aggregation Network*), de maneira a fundir os mapas de características de diferentes resoluções através dos Fluxos *top-down* e *botto-up*.

No fluxo *top-down*, as características semânticas profundas são propagadas para os níveis mais rasos. Já o fluxo *bottom-up* busca detalhes de alta resolução para refinar informações semânticas. Desta forma, a combinação desses mecanismos busca melhorar a detecção de objetos de tamanhos variados, para garantir o enriquecimento multiescala antes da etapa da definição da caixa delimitadora. Esta etapa é representada em (A4) na Figura 12.

Etapa 1 (A5): Cabeça de Detecção (Head) A cabeça de detecção é formada por um conjunto de três ramificações, sendo a Regressão de caixas ($\Delta x, \Delta y, \Delta w, \Delta h$); a *Objectness* (BCE Loss) e a Classificação de classe (BCE Loss). A função de perda total combina a entropia cruzada para a probabilidade e existência do objeto e classificação, e também $\mathcal{CIoULoss}$, para refinamento da caixa delimitadora:

$$\mathcal{L} = \lambda_{\text{obj}} \mathcal{L}_{\text{obj}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}} + \lambda_{\text{box}} \mathcal{L}_{\text{CIoU}}. \quad (4.2)$$

Por fim, aplica-se uma supressão de não-máximos baseada na *Distance-IoU* (DIoU-NMS), com o objetivo de eliminar caixas redundantes e manter apenas a caixa dentro do índice de confiança. Esta etapa é representada em (A5) na Figura 12.

Etapa 1 (A) (A6): Geração de Parâmetros de Aplicação Celular - Nesta etapa dá-se início ao processo de segmentação do tumor de dentro da Caixa delimitadora. É definido um ponto de partida inicial para a célula, que será utilizado na Etapa 1B para a aplicação do modelo de segmentação. Normalmente é utilizado o centroide da caixa delimitadora, mas em casos de regiões mais disformes, propõe-se uma metodologia baseada na Inicialização do Método da Árvore Quadrante para encontrar um ponto de partida ótimo, para os casos em que o centroide da caixa delimitadora não é capaz de inserir a

célula sintética computacional na região do tumor. Esta etapa da aplicação celular pode ser vista na Figura 12, Etapa 1 (A), estágio A6. Vale ressaltar que métodos de *Seed Selection* baseados em convolução e limiarização, aliado à detecção de componentes conectados também seria um ótimo método de correção da localização da semente inicial dentro da caixa delimitadora. A máscara de detecção do YOLO é utilizada como padrão ouro para avaliação se a região selecionada para inicialização da célula sintética computacional está no interior da região de interesse, a mesma também é utilizada no gradiente de valor de parâmetro de parada da célula.

Etapa 1 (A) (A7): Geração de Parâmetros de Parada Celular Adaptativos

– Após a detecção da região de interesse por meio da inicialização automática com o YOLO, é realizada a geração de parâmetros de parada adaptativos para células sintéticas computacionais. Essas células iniciarão o processo de segmentação assim que a aplicação iniciar no estágio (B1), conforme ilustrado no estágio 2 (B) (A7), na Figura 12.

O processo adaptativo para gerar o parâmetro de parada celular baseia-se no limiar da região circundante da máscara gerada pela inicialização automática, a máscara binária de detecção, na imagem de entrada original. Um processo de gradiente morfológico é aplicado a essa região, correspondente à área circundante da região de interesse, utilizando um elemento estruturante quadrado 5×5 para gerar uma imagem de mapa binário. Esse processo auxilia na obtenção dos valores limiares do que se entende pela região interna da área de interesse para inicialização, e também o parâmetro de parada de borda das células computacionais sintéticas.

Em suma, a Etapa 1 (A) refere-se à abordagem de inicialização automática para obtenção de parâmetros de segmentação e de parada. Na imagem 12, (A1) representa a imagem de entrada dentro da abordagem de inicialização do método; (A2) ilustra o processo de extração de características por meio de convoluções da imagem de entrada com relação à arquitetura do *framework* YOLO; Em (A3), é apresentada a Rede de Proposta de Região, onde é gerada uma proposta de região candidata para a imagem médica; em (A4), ilustra-se o processo de extração de características da região de interesse; em (A5), realiza-se a etapa de detecção de regiões na imagem médica mediante caixa delimitadora e máscara binária de detecção para obtenção de Parâmetros de segmentação. Em (A6), gera-se o parâmetro de aplicação para a célula computacional sintética a ser segmentada e, por fim, em (A7), realiza-se a geração do parâmetro de parada ótimo das bordas da região de interesse. A próxima seção, na Etapa 2 (B), abordará a cerca do funcionamento da abordagem de segmentação de imagens médicas aplicada neste estudo, o método *Divisible Cell Segmentation*, proposto por (SOUZA, 2023).

Etapa 2 (B): Segmentação — Método *Divisible Cell Segmentation* - A

Etapa 2 (B) da metodologia proposta neste estudo aborda a respeito da segmentação do Tumor Cerebral na imagem de RM. O *Divisible Cell Segmentation* baseia-se em células

sintéticas computacionais capazes de se dividir e multiplicar até a borda do objeto de interesse, preenchendo toda a região interna da mesma através de um enchame de pontos, com o objetivo de segmentar a região de interesse por meio de conexões celulares, formando um polígono fechado.

Etapa 2 (B1) - Aplicação Celular (Célula Sintética Computacional) -

A célula sintética computacional é um objeto que contém informações relacionadas à **variante de direção**, representada por V_D , à **Divisão Celular**, representada por D_C , e ao **parâmetro de parada**, representado por P . O valor atribuído a P é obtido a partir do processo de inicialização automática adaptativa e armazenado como V_L , o respectivo **valor limiar**.

Após a inicialização do método e a etapa de aplicação celular, representada pela Equação 4.3, onde o centroide, representado por C na expressão, recebe a célula c_i com variante de direção V_D e o parâmetro de parada P de acordo com seu respectivo índice dentro do espaço bidimensional. Isso corresponde aos agrupamentos de c_i , células posicionadas de acordo com o centroide a ser percorrido na matriz no ponto I , que corresponde ao ponto inicial da célula na inicialização do método. A aplicação da semente do método, bem como sua movimentação, podem ser observadas em B1 e B2 da Figura 12 Etapa 2B.

$$C \leftarrow c_i(V_D, P(V_L)) = \sum_i^{-1} \text{ onde } c_i(x_i, y_i) \text{ em } I \quad (4.3)$$

Etapa 2 (B2) - Movimento e Divisão Celular - (Célula Computacional Inteligente (CCI)) - O vetor de variante de direção, que guia os passos de movimento da célula, é definido pela distância de um passo nas coordenadas x e y, passo x e y, respectivamente. Neste caso, considere a função de coordenadas que será analisada pela célula c_i .

Sejam Bbx e Bby a largura e a altura da caixa delimitadora, respectivamente. O vetor V_D é definido como:

$$V_D \leftarrow f_{div}(c_{ix}, c_{iy}); * \quad (4.4)$$

tal que f_{div} é uma função recursiva, c_{ix} e c_{iy} são as coordenadas do ponto inicial c_i , e $*$ é o ângulo que define a direção do movimento da célula. A função f_{div} é definida como:

$$f_{div}(x_k, y_k) = \begin{cases} \emptyset, \text{ se } \begin{matrix} x_k + step_x > Bbx \text{ e} \\ y_k + step_y > Bby \end{matrix} \\ (x_k, y_k) \cup f_{div}(x_k + step_x, y_k + step_y) \end{cases} \quad (4.5)$$

Etapa 2 (B3) - Segmentação (Parâmetro de Parada) - O parâmetro de parada P é definido pela máscara já detectada durante a inicialização da abordagem automática. Isso envolve o cálculo dos pixels nas bordas da região de interesse, onde a função m-vetorial (x, y) é aplicada sobre a localização da delimitação da borda da máscara na imagem de entrada, conforme mostrado na Etapa 2B (B3) na Figura 12.

Etapa 2 (B4) - Segmentação (Conexões via Células Computacionais Sintéticas) - Ao final do processo de movimentação das células sintéticas computacionais sobre a região de interesse, inicia-se o processo de segmentação, utilizando a técnica de conexão via células computacionais sintéticas. Isso envolve a criação da imagem binária de pontos. Essa imagem é formada por uma imagem escura com um conjunto de pontos brancos posicionados na imagem, cujas coordenadas foram obtidas a partir das células estacionadas na borda da região de interesse

A imagem binária é usada para a aplicação da fase final do método: a conexão dos centros celulares, iniciando o processo de construção do plano de parâmetros das células. A equação 4.6 representa o processo de fechamento da região de segmentação por meio da conexão dos centros celulares.

$$\begin{aligned}
 &\forall c \subset B | P \rightarrow H(\rho, \theta) \\
 &H(\rho, \theta) = \sum_{k=1}^n w(k) \delta(\rho - x_i \cos(\theta) - y_i \sin(\theta)) \\
 &\quad \text{onde} \\
 &\quad k = c \\
 &w(k) = \exp\left(\frac{-d^2(k, i, j)}{2\sigma^2}\right) \\
 &\quad \text{onde} \\
 &d(k, i, j) = \min(d(i_k, j), d(j_k, i)) \\
 &\quad \text{de modo que} \\
 &\quad k \neq i \text{ e } k \neq j
 \end{aligned} \tag{4.6}$$

Onde B representa a borda da região de interesse, cujas células c são condicionadas pelo parâmetro de parada P e representadas por K . $H(\rho, \theta)$ representa o espaço de parâmetros da transformação, onde ρ é a distância perpendicular da reta à origem e θ é o ângulo que a reta forma com o eixo x . $\sum_{k=1}^n$ representa a soma de todos os pontos celulares centrais na imagem que são candidatos a formar uma reta. O peso $w(k)$ está associado a cada ponto k , representando a confiança dos pontos celulares centrais em pertencer a uma reta. Ele é calculado com base na distância $d^2(k, i, j)$ entre o ponto celular k e os pontos celulares i e j mais próximos a ele. Quanto menor a distância, maior o peso. A função delta $\delta(\rho - x_i \cos(\theta) - y_i \sin(\theta))$ indica se um ponto (x_i, y_i) na imagem contribui para a reta representada pelos parâmetros ρ e θ . A distância $d(k, i, j)$ é a distância entre o ponto celular central k e a reta definida pelos pontos i e j . As restrições $k \neq i$, $k \neq j$ garantem que i e j sejam diferentes de k (ILLINGWORTH; KITTLER, september. 1987).

O processo de conexão de pontos celulares localizados nas bordas da região tumoral é facilitado pelo algoritmo Convex Hull, um conceito fundamental em geometria computacional e matemática discreta, capaz de descrever um envelope convexo com o menor número de linhas que contêm os pontos no plano. Em resumo, o algoritmo Convex Hull encontra o menor polígono convexo conectando os pontos de um conjunto quando aplicado às bordas externas.

O processo de conexão de pontos celulares localizados nas bordas da região tumoral é facilitado pelo algoritmo Convex Hull, um conceito fundamental em geometria computacional e matemática discreta, capaz de descrever um envelope convexo com o menor número de linhas contendo os pontos no plano. Em resumo, o convex hull encontra o menor polígono convexo conectando os pontos de um conjunto quando aplicado às bordas externas. A equação 4.7 ilustra o processo de construção do polígono de segmentação fechado.

$$CH(P) = \left\{ \begin{array}{l} \{(x, y) \mid x = \sum_{i=1}^n \lambda_i x_i, y = \sum_{i=1}^n \lambda_i y_i, \\ \text{onde} \\ \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1 \end{array} \right. \quad (4.7)$$

Em que, $CH(P)$ representa o conjunto de pontos no casco convexo. (x, y) são as coordenadas dos pontos no casco convexo. $\sum_{i=1}^n \lambda_i x_i$ expressa a coordenada x de um ponto no casco convexo como uma combinação linear convexa das coordenadas x dos pontos no conjunto P . Similarmente, $\sum_{i=1}^n \lambda_i y_i$ expressa a coordenada y de um ponto no casco convexo como uma combinação linear convexa das coordenadas y dos pontos em P . $\lambda_i \geq 0$ garante que os coeficientes na combinação linear não sejam negativos, e $\sum_{i=1}^n \lambda_i = 1$ garante que a soma dos coeficientes seja igual a 1, significando uma combinação convexa ponderada.

Assim, o método proposto de conexões via pontos celulares conecta pontos de diferentes curvaturas por meio de linhas, gerando um contorno baseado em células computacionais inteligentes em seus pontos de parada dentro da matriz. Este objeto é considerado a máscara de segmentação binária do modelo, que é usada para extrair métricas de comparação que são então comparadas com a verdade básica (GT).

Estágio 3 (C1) - Extração de Características Profundas No estágio 3 (C) da Figura 12, ocorre o processo de classificação do tumor cerebral, com base na segmentação do estágio 2 das imagens de ressonância magnética. Este estágio envolve dados relacionados à classificação da idade do paciente, tempo de sobrevivência e contagem de pixels das regiões do tumor com realce de GD, edema peritumoral e núcleo tumoral necrótico e sem realce.

Para tanto, o modelo proposto combina extratores CNN e classificadores de aprendizado de máquina, visando obter a melhor combinação para a abordagem proposta. As

combinações são realizadas em conjunto com o método de busca em grade para definir os hiperparâmetros de alguns classificadores, conforme ilustrado na Tabela 2. Os classificadores utilizados incluem KNN, MLP, NaiveBayes, RandomForest, SVM linear, SVM polinomial e SVM RBF. Assim, o estágio 3 do modelo proposto é um pipeline de extração e classificação de imagens utilizando aprendizado de transferência, no qual os atributos profundos são classificados por classificadores de aprendizado de máquina. A validação cruzada foi utilizada para validar os resultados e gerar médias e desvios. Foi-se aplicado Grid Search para encontrar os melhores hiperparâmetros dos classificadores

Estágio 3 (C2) - Classificação

O processo de classificação ocorre por meio da combinação de dois bancos de dados de tumores, LGG-TCGA e BraTS2020. Nesse processo, o banco de dados utilizado para detecção (LGG-TCGA) auxilia na obtenção da região de interesse a partir da segmentação da região tumoral, tornando a extração de atributos mais direcionada, com o objetivo de otimizar a tarefa de classificação.

Vale ressaltar que as classes de tempo de sobrevivência, Idade e presença/ausência (detecção binária) foram utilizadas como parâmetro para seleção da melhor combinação extrator/classificador. Ambas as classes são afetadas pela segmentação, sendo a idade e o tempo de sobrevivência duas classes consideradas difíceis pela literatura, e a classe de detecção binária um caso mais fácil. Desta forma, a melhor combinação da comparação do pipeline é definido pela ponderação entre os casos mais difíceis e o mais fácil.

Em relação às outras classes além da idade, o tumor com realce de DG refere-se à região correspondente do tecido tumoral realçada pelo contraste à base de gadolínio, o edema peritumoral refere-se ao inchaço ao redor do tumor e o núcleo tumoral necrótico e sem realce refere-se ao tecido tumoral morto e fragmentos tumorais não realçados pelo contraste. Todas as classificações são derivadas de análises importantes realizadas pelo médico especialista.

O pipeline seleciona a melhor combinação de extrator classificador mediante comparação combinatória exaustiva, com a métrica de acurácia sendo o critério de seleção para a melhor combinação. O grid search é aplicado antes do treinamento definitivo dos classificadores, visando ampliar sua performance para com o dataset.

Vale ressaltar também que o método foi proposto de forma a não distinguir as diferentes modalidades de RM existentes no BraTS2020, sendo todas elas utilizadas tanto nas etapas de treino, como de teste dos classificadores na forma de uma imagem interpolada. O objeto foi obter a versão mais generalizada possível, através do processo de *dataFusion*, com a região do tumor exaltada em relação à região de fundo.

Etapa 4 (D) - Resultado do Modelo de linguagem de grande escala

Na etapa 4 (D) da Figura 12, um modelo de linguagem de grande porte (LLaMA3)

é utilizado em conjunto com os atributos classificados da etapa anterior para exposição dos resultados na forma de pequeno relatório formal de pré-diagnóstico, buscando ser claro e informativo. O modelo é previamente contextualizado para receber as entradas de classificação, e modelos de documentos clínicos de domínio público, bem como artigos e revistas voltados para a área médica de tumor cerebral.

Para a geração das saídas narrativas do LLM, primeiro é necessário que o mesmo associe cada contador de pixels aos seus respectivos componentes tumorais, através do dicionário de rótulos do metadata presente no dataset BraTS2020. Sendo por exemplo, `label1_pxl_cnt` → Núcleo Necrosado e Não Realçado (NCR/NET); `label2_pxl_cnt` → Edema Peritumoral (DE).

Em seguida, ocorre a converção da contagem de pixels da classe em volume (mm³) por meio de fórmula:

$$V_{\text{classe}} = N_{\text{px}} \times \text{pixel_spacing}_x \times \text{pixel_spacing}_y \times \text{slice_thickness}, \quad (4.8)$$

Na qual, em datasets isotópicos como o BraTS, é possível assumir $V_{\text{voxel}} = 1$, logo $1 \text{ pixel} = 1 \text{ mm}^3$.

Logo o volume total do tumor em cada fatia é obtido pela soma das subregiões:

$$V_{\text{total}} = \sum_{\text{classes de tumor}} V_{\text{classe}}. \quad (4.9)$$

De forma paralela, as demais classes, como idade e tempo de sobrevivência são obtidos também da classificação, e combinados com as áreas convertidas, forma o texto de resultado, conforme dado o exemplo abaixo:

- Idade: uma casa decimal (“60,5 anos”).
- Sobrevivência: inteiro (“395 dias”).
- Volume por região e total: inteiro em mm³ (“5 mm³”, “538 mm³”, “0 mm³” quando ausente).

O LLM preenche então, o seguinte trecho padronizado:

"Para a imagem de ressonância magnética da fatia {slice} do volume {volume_id}, foi detectada presença de tumor. A idade do paciente é aproximadamente {age} anos e o tempo de sobrevivência estimado após o diagnóstico é de {survival_days} dias. O volume total de tumor nesta fatia é de {V_total} mm³, dos quais {V_NCR} mm³ correspondem ao Núcleo Necrosado e Não Realçado (NCR/NET) e {V_DE} mm³ correspondem ao Edema Peritumoral (DE)."

Informações adicionais também podem vir a ser agregadas por parte do que o modelo consegue sintetizar pelo material dado no contexto em relação aos valores limiares classificados, mas estas informações ainda possuem cunho informacional, visando agregar o

ponto de vista do médico especialista, como o exemplo abaixo para os limiares do exemplo acima:

"O fato de todo o volume tumoral estar representado apenas por edema, sem evidência de necrose ou realce, sugere um perfil mais infiltrativo e possivelmente de baixo grau (WHO I-II). Nesses casos, a barreira hematoencefálica já se altera o suficiente para permitir acúmulo de líquido, mas sem formação de tecido necrótico, o que costuma indicar um crescimento mais lento e melhor resposta a corticosteroides para controle de pressão intracraniana. Além disso, o pequeno volume (5 mm^3) favorece intervenções minimamente invasivas, como biópsia estereotáxica guiada, reduzindo riscos e permitindo um diagnóstico histológico preciso antes de qualquer tratamento cirúrgico ou adjuvante."

5 Resultados e Discussão

Nesta secção são apresentados resultados da detecção, segmentação e classificação de tumores cerebrais em imagens de ressonância magnética (RM), conforme as etapas da metodologia proposta. Para a validação da segmentação e classificação, os resultados obtidos foram comparados com métodos encontrados na literatura que se utilizem da mesma base de dados, e mesmas métricas de avaliação.

Todo o processo de aprendizado profundo utiliza o sistema operacional Linux, distribuição Ubuntu 22.04, 32 GB de RAM, processador AMD Ryzen 5 5600, 3,5 GHz (4,4 GHz Max Turbo), cache de 35 MB, AM4 e placa de vídeo NVIDIA GeForce RTX 3060 (com 12 GB de memória). A implementação da detecção foi feita utilizando a linguagem Python, versão 3.11.1, utilizando as bibliotecas OpenCV v4.1.0, sci-kit-learn v0.20.2, v.1.14.0 e Keras v.2.2.4.

5.1 Inicialização da abordagem

Com a finalidade de avaliar os resultados de performance do YOLOv8 no que diz respeito à sua detecção por máscara, cuja performance é fundamental para obtenção de parâmetros de parada da segmentação, foi-se aplicado o experimento de avaliação da máscara mediante K-fold, com 3 folds, os resultados são expressos na tabela 3. Cada fold foi construído a partir de partições aleatórias dos dados de treino e validação; o desempenho final de detecção, porém, foi reportado sobre o conjunto de teste completo.

Tabela 3 – Resultados do YOLOv8 na detecção com máscara binária de tumores cerebrais. A tabela mostra os resultados do experimento com base na validação cruzada.

Método	ACC(%)	SEN(%)	ESP(%)	VPP(%)	VPN(%)
YOLOV8	99.87 ± 0.18	98.55 ± 2.45	99.89 ± 0.03	98.57 ± 2.33	99.02 ± 0.13
Validação-Cruzada					
Fold 1	99.81 ± 0.23	98.37 ± 2.10	99.69 ± 0.58	98.17 ± 1.88	99.01 ± 0.88
Fold 2	99.91 ± 0.13	98.37 ± 1.66	98.23 ± 1.67	98.88 ± 2.03	99.55 ± 0.39
Fold 3	99.86 ± 0.33	98.88 ± 0.23	99.91 ± 0.02	98.88 ± 1.07	98.88 ± 0.22

A tabela 3 apresenta o valor de ACC com valor obtido de 99.87%, demonstrando a eficácia da abordagem de inicialização automática do método para detecção inteligente da região do tumor no cérebro em imagem de RM. O modelo aparenta ter bons resultados em distinguir corretamente imagens livres de tumores, obtendo a métrica VPN em torno de 99.02%, relacionados a amostras negativas, isso demonstra o quão foi eficaz na identificação das regiões de fundo. Este fator, se aplica também a métrica de ESP que atingiu 99.89%, consolidando um bom desempenho da abordagem. Tanto a métrica de SEN quanto a

métrica de VPP obtiveram resultados acima de 98%, tendo em vista a complexidade de variações de intensidades nos *pixels* nas imagens de RM, na qual assemelham-se a partes da massa cefálica.

As métricas de PPV apresentadas na Tabela 3 estão relacionadas ao grau de complexidade em detectar com precisão essas regiões. No entanto, alcançar um ACC de 99,87% justifica a acurácia dos verdadeiros positivos em relação aos pixels encontrados em imagens de ressonância magnética, concluindo uma excelente detecção da região de interesse delineada pela caixa delimitadora na região tumoral. Após esse processo, novas soluções são apresentadas para o problema de segmentação, que é abordado na subseção 5.2. A validação cruzada foi realizada separando-se o conjunto de dados em três folds, cada um com 77 pacientes, totalizando 233 pacientes. Conduzimos três sessões de treinamento/teste, cada uma com dois folds para treinamento e um fold para teste, conforme apresentado na Tabela 3. Analisando os resultados obtidos em cada sessão de treinamento/teste com diferentes folds, é possível identificar a semelhança entre os folds utilizados no treinamento (66%) e no teste (33%) dos pacientes presentes no conjunto de dados.

Nas imagens da Figura 13, é possível identificar a precisão da detecção da região afetada pelo tumor cerebral. Na coluna (A), são apresentadas diversas imagens de RM do crânio para comparação visual da área lesionada na substância cinzenta do cérebro, nas quais os tumores exibem diferentes tamanhos e formatos no exame. Vale ressaltar que, para cada forma capturada na RM, existe um nível correspondente de desafio de segmentação da região, já que há diferentes contornos e gradientes de textura na imagem onde a lesão está presente — evidenciados pelo contraste da região de interesse nas imagens da coluna (A) — representam desafios para a segmentação de tumores cerebrais em variados tipos de RM.

Neste contexto, na coluna (B) da Figura 13, são apresentadas as detecções relacionadas à primeira fase da metodologia deste estudo. Estas detecções são capazes de delinear com precisão as regiões lesionadas de tumores cerebrais na massa cerebral através de caixas delimitadoras.

5.2 Segmentação

A segmentação do método *Divisible Cell Segmentation* foi avaliada mediante comparação da máscara binária de segmentação produzida pelo modelo com o padrão ouro, mas em uma etapa única.

A Tabela 4 apresenta diferentes valores obtidos na segmentação da região do tumor cerebral. O método DAS-Brain Tumor aprestou 99.80 de ACC e valores métricos de SEN, ESP, VPP e VPN acima de 98%, o que demonstra uma performance satisfatória para o modelo em relação ao dataset de segmentação abordado, o TCGA.

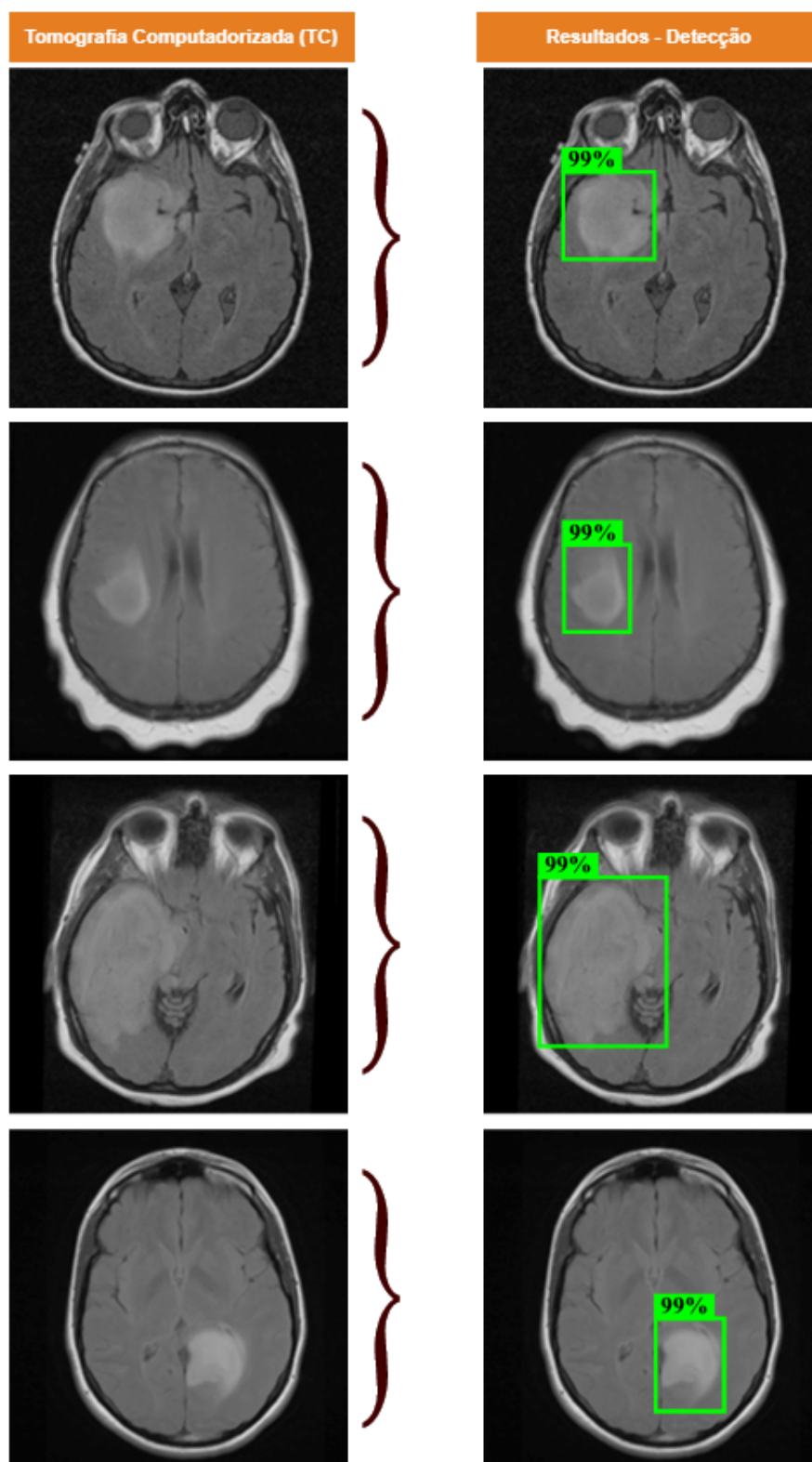


Figura 13 – Resultados da inicialização automática para detecção de tumores cerebrais em imagens de ressonância magnética craniana. Em (A), as imagens originais são exibidas, enquanto em (B), as regiões detectadas podem ser observadas através da sobreposição das caixas delimitadoras.

A Figura 14 apresenta o resultado do processo de segmentação usando o método *Divisible Cell Segmentation* em uma imagem da região cerebral com tumor. A segmentação

Tabela 4 – Resultado da segmentação do método *Divisible Cell segmentation*, inicializado com auxílio do YOLOv8.

Método	ACC(%)	SEN(%)	ESP(%)	VPP(%)	VPN(%)
Divisible Cell	99.80 ± 0.68	98.74 ± 1.06	99.79 ± 0.25	98.55 ± 2.30	99.78 ± 0.70

inicia-se em 1, onde o centroide é calculado com base nos parâmetros de detecção e na caixa delimitadora obtidos na etapa A5 de detecção da região do tumor com o auxílio do YOLOv8. Em 2, a aplicação realiza a primeira divisão celular, gerando as células computacionais inteligentes profundas iniciais. Nos passos de 3 a 7, é realizado o movimento e a divisão das células, que cresce exponencialmente sobre a região tumoral. As coordenadas das células localizadas nas bordas da área da lesão são usadas para construir a máscara de segmentação binária através do *Concave Hull*, conforme mostrado em 8. A imagem final representa o resultado da segmentação segundo o método proposto.

Na Figura 15, é possível analisar que, em diferentes imagens com diferentes tipos de tumores, bem como diferentes formas, os tumores foram segmentados com eficácia pelo método. Na imagem em questão, os resultados demonstram a relativa eficiência do avanço das células subdividindo-as até a borda do objeto de interesse, conforme os parâmetros setados. O método trouxe resultados satisfatórios para o contexto da aplicação por se tratar de células sintéticas computacionais, têm como objetivo alcançar as bordas dos tumores e regiões de definições imprecisas após definido o linear do critério de parada.

Um fator importante na avaliação de um modelo de segmentação é o tempo, pois impacta diretamente na aplicabilidade de um modelo, dependendo do contexto. Considerando isso, o tempo de segmentação foi calculado e apresentado na Tabela 5.

Tabela 5 – Tempo de segmentação do método proposto, com *Divisible Cell Segmentation* aplicado sobre o conjunto de teste.

Método	Tempo(s)
Método Proposto (<i>DAS - Brain Tumor</i>)	0.798 ± 1.55

5.3 Classificação

A base de classificação, BraTS2020, contém as classes em um arquivo de metadados à parte, com indicação para as diferentes fatias de imagens de RM. O conjunto de extrator/classificador, bem como os hiperparâmetros ótimos do GridSearch, foram calculados para a classe de tempo de sobrevida, ou *Survival Days*, pois esta foi considerada uma das classes mais interessantes e desafiadoras do estudo, embora seja uma classe notavelmente mais dificultosa de se obter boa performance, até mesmo na literatura, com a média de acurácias ficando em torno de 50%.

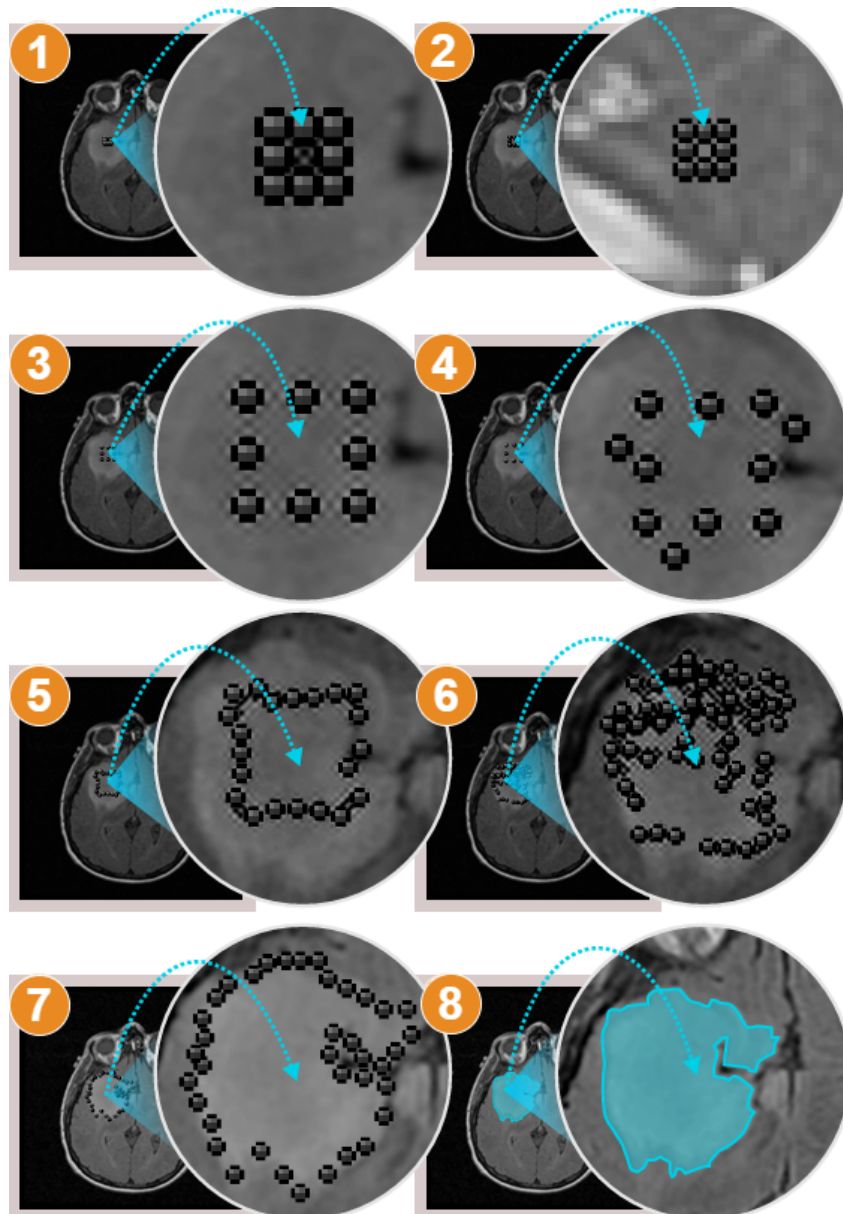


Figura 14 – Apresenta o processo de segmentação de tumores cerebrais realizado pelo método *Divisible Cell Semgnetation*. As imagens ilustram o movimento de células computacionais, dividindo-se para detectar as bordas do tumor cerebral conforme se movem cada uma em sua direção específica de acordo com o seu vetor de direção. A última imagem mostra o resultado da segmentação pelo método, que inclui conectar os pontos centrais das células na borda do tumor através do *Concave Hull*, formando a segmentação tumoral.

Outro fator que vale a pena ser ressaltado é que este estudo não buscou realizar uma distinção performática da classificação com base nas diferentes modalidades de RM existentes na base BraTS2020 por questões de escopo e de generalização. O pipeline proposto foi projetado para lidar com todas as modalidades do dataset mediante o pré processamento com *DataFusion* que gera uma imagem interpolada das mesmas, na qual cada modalidade busca exaltar características diferentes da patologia.

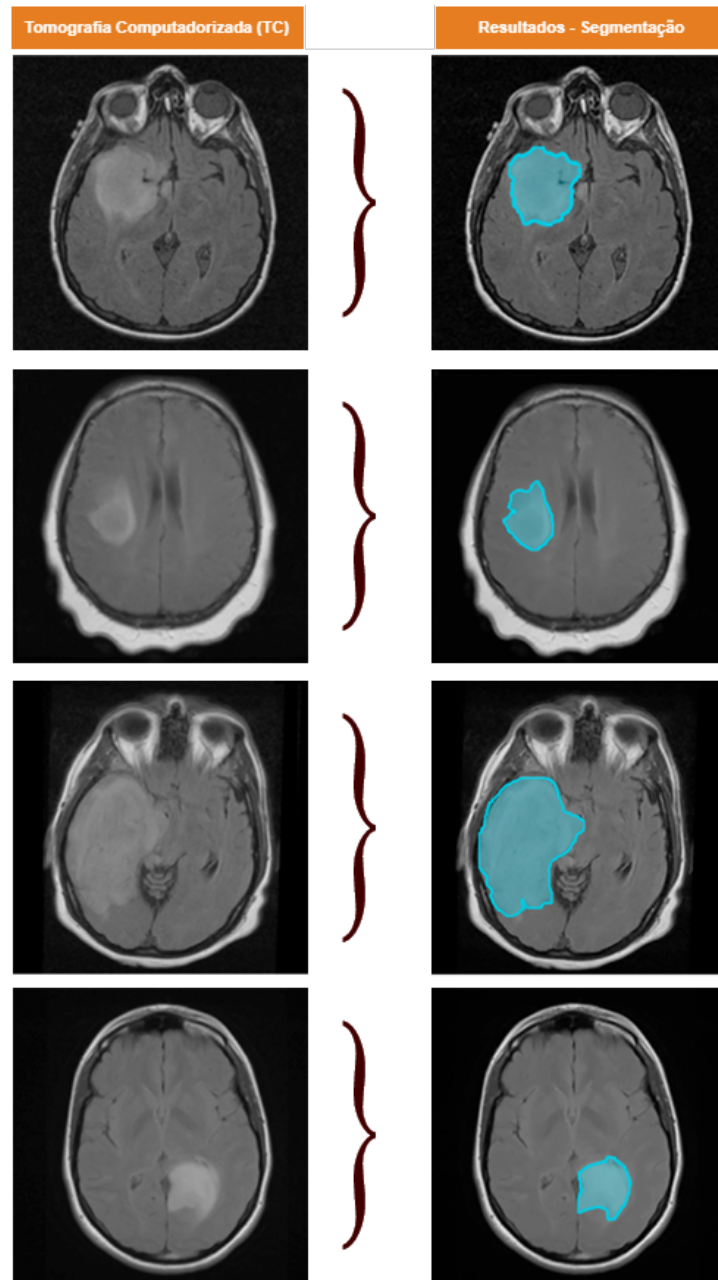


Figura 15 – Exibe o resultado da segmentação usando o método *Divisible Cell Segmentation* em diferentes imagens de ressonância magnética do crânio. A coluna (A) mostra as ressonâncias magnéticas como imagens de entrada, e a coluna (B) mostra os resultados da segmentação da região da borda da lesão pelo método proposto, como imagens de saída.

A tabela 6 apresenta os resultados da classificação para o conjunto de extratores e classificadores adotados para o banco de dados Brats2020 com o rótulo *Survival Days*.

Observa-se que a melhor combinação extrator/classificador para o rótulo de tempo de sobrevivência foi o DenseNet201 com SVM_RBF, cuja acurácia média obtida foi de 53,92%. Esse valor indica um resultado relativamente satisfatório para a métrica em questão, dada a dificuldade de sua classificação, cujo tempo de sobrevivência estimado leva em consideração uma série de fatores, como tamanho da lesão, áreas internas necróticas

e edemaciadas e homogeneidade da região. A precisão da combinação foi em torno de 61,42%, indicando uma margem relativamente boa de predições corretas em detrimento das predições totais. O recall foi em torno de 53.81%, e o F1-score foi de 53.42%. É notável que são resultados considerados insatisfatórios no que diz respeito á possibilidades de aplicações clínicas, cujos valores geralmente são em torno de 90%. Embora os resultados tenham sido considerados insatisfatórios em termos de aplicações clínica, vale ressaltar a estabilidade do desvio padrão, que oscila entre 2.5% e 3.5% para todas as métricas.

Acredita-se que ao contrário das demais classes, cujos resultados se saíram melhores, a classe de tempo de sobrevida poderia se beneficiar mais ao inserir as regiões de fundo na etapa de extração, cuja massa cinzenta do restante do cérebro poderia vir á trazer mais elementos discriminantes, oque será analisado em estudos futuros, com um comparativo de fatores metodológicos que influenciam os resultados. Todavia, os hiperparâmetros indicados para a busca em grade para o classificador SVM RBF da combinação DenseNet201 + SVM RBF foram: **gamma**: 0,001 e **“C”**: 10.

O DenseNet201 é considerado uma das versões mais robusta da família desta arquitetura, pelo que indica na literatura, com mais camadas, conseqüentemente, mais parâmetros, sendo adotado em tarefas de classificação mais complexas, com maior demanda por detalhes nos atributos. O classificador SVM com kernel RBF, por outro lado, tem plena capacidade de resolver problemas não lineares com sua mudança de domínio no espaço de atributos causada pelo kernel. Aliado á busca de hiperparâmetros ótimos, esta variação do SVM pode ser capaz de boa performance se bem configurado.

A tabela 6 também mostra os tempos médios de treino e classificação para cada combinação extrator-classificador discutida. O menor Tempo de Treino foi obtido pela combinação do VGG16 + Naive Bayes, com 0,23 segundos, o que pode ser interpretado pela menor complexidade arquitetural do método VGG16, gerando atributos profundos menos complexos, aliado ao princípio de funcionamento do Naive Bayes, que aplica a estatística probabilística sobre as amostras com um rápido tempo de resposta. O menor tempo de predição foi obtido pela combinação do DenseNet201 + MLP, com 0,01 segundos. Isso pode ser explicado pela rápida saída do modelo MLP para classificação após a configuração dos pesos, ao custo de um tempo de treino maior para a seleção dos mesmos.

Em todos os casos, o Randomforest é sempre o classificador com maior tempo de treinamento, pois durante esta etapa, o mesmo busca essencialmente construir dezenas ou centenas de árvores de decisão, sendo cada árvore gerada recursivamente e escolhendo em cada nó. E como os atributos profundos gerado pelos modelos DenseNet são mais complexos que os do VGG, é notável a diferença de tempo de treino, sendo 1074.69 ± 9.34 segundos para o DenseNet121, 1376.31 ± 3.99 segundos para o DenseNet201 e 495.73 ± 2.71 segundos para o VGG16.

Vale ressaltar que os tempos foram baixos, apesar da complexidade do modelo e

grande quantidade de dados do Dataset BraTs2020 devido às configurações da máquina, que buscou realizar os processos em GPU, além de que os tempos nas tabelas foram obtidos com treinamento depois do devido processo de obtenção de hiperparâmetros com Grid Search.

Tabela 6 – Resultados da classificação para as combinações de extratores e classificadores testados em relação ao tempo de sobrevivência, ou *Survival Days*, para o dataset Brats2020.

Extrator	Classificador	Acurácia (%)	Precisão (%)	Recall (%)	F1-Score (%)	Tempo de Treino (s)	Tempo de Predição (s)
DenseNet121	KNN	34.05 ± 2.40	42.28 ± 4.18	34.24 ± 2.68	33.54 ± 2.99	0.51 ± 0.16	0.03 ± 0.01
	MLP	48.93 ± 3.15	54.26 ± 3.76	48.55 ± 3.53	47.64 ± 3.21	28.20 ± 0.65	0.01 ± 0.00
	NaiveBayes	23.50 ± 2.14	30.28 ± 2.72	23.53 ± 3.24	23.08 ± 2.42	0.42 ± 0.01	0.12 ± 0.07
	RandomForest	48.93 ± 2.23	59.09 ± 3.51	48.57 ± 3.40	48.53 ± 3.16	1074.69 ± 9.34	0.05 ± 0.01
	SVM_Linear	49.88 ± 3.23	55.59 ± 4.38	48.94 ± 3.90	48.45 ± 3.47	18.91 ± 2.31	0.27 ± 0.01
	SVM_Polynomial	47.64 ± 3.31	53.90 ± 4.37	46.82 ± 4.22	46.39 ± 3.77	43.12 ± 5.08	0.27 ± 0.02
	SVM_RBF	48.17 ± 2.45	57.86 ± 2.84	47.27 ± 2.97	47.63 ± 2.68	62.65 ± 6.19	0.53 ± 0.04
DenseNet201	KNN	36.60 ± 1.83	43.82 ± 2.29	36.49 ± 2.98	35.37 ± 2.23	1.63 ± 0.21	0.09 ± 0.03
	MLP	53.85 ± 2.32	59.91 ± 2.24	54.16 ± 3.38	53.25 ± 2.44	33.31 ± 1.30	0.01 ± 0.00
	NaiveBayes	25.19 ± 2.86	33.82 ± 5.29	25.32 ± 4.08	24.91 ± 3.77	0.92 ± 0.16	0.18 ± 0.01
	RandomForest	51.07 ± 2.78	60.00 ± 3.57	50.90 ± 4.05	50.41 ± 3.77	1376.31 ± 3.99	0.05 ± 0.00
	SVM_Linear	54.18 ± 2.78	60.88 ± 1.95	54.06 ± 3.43	53.63 ± 2.65	38.21 ± 3.66	0.70 ± 0.03
	SVM_Polynomial	53.69 ± 3.00	61.56 ± 2.59	53.68 ± 3.85	53.46 ± 3.37	87.00 ± 7.91	0.73 ± 0.05
	SVM_RBF	53.92 ± 2.58	61.42 ± 2.80	53.81 ± 3.16	53.42 ± 2.93	144.69 ± 10.47	2.14 ± 0.01
VGG16	KNN	32.86 ± 2.16	40.17 ± 3.60	32.45 ± 2.99	31.61 ± 2.67	0.26 ± 0.00	0.01 ± 0.00
	MLP	44.20 ± 3.06	49.62 ± 3.32	43.74 ± 3.50	42.91 ± 3.40	52.14 ± 1.74	0.01 ± 0.00
	NaiveBayes	15.80 ± 2.74	21.66 ± 3.66	15.74 ± 3.70	15.68 ± 3.09	0.23 ± 0.00	0.05 ± 0.00
	RandomForest	47.47 ± 3.01	56.18 ± 1.90	47.15 ± 3.94	46.76 ± 3.36	495.73 ± 2.71	0.05 ± 0.00
	SVM_Linear	43.47 ± 2.09	50.63 ± 2.98	42.09 ± 2.60	42.26 ± 2.27	13.01 ± 0.84	0.21 ± 0.00
	SVM_Polynomial	34.28 ± 2.08	46.10 ± 2.69	34.38 ± 3.08	35.33 ± 2.72	28.02 ± 0.08	0.20 ± 0.00
	SVM_RBF	46.78 ± 2.14	55.99 ± 3.63	46.32 ± 4.51	46.19 ± 3.41	38.87 ± 0.03	0.38 ± 0.01

A tabela 7 apresenta os resultados obtidos para a classificação da idade do paciente para o banco de dados Brats2020. A melhor combinação para o rótulo de idade também foi o extrator DenseNet 201, com o classificador SVM_RBF, com uma acurácia média de 54,25%, a precisão em torno de 60.17% , o recall em torno de 53.95% e o f1-score em torno de 52.79%. Essa semelhança entre a idade e o tempo de sobrevivência pode indicar estabilidade de desempenho para essa combinação de extrator/classificador para o banco de dados Brats2020 e os rótulos classificados, pelo menos no que diz respeito á essas classes, Mas os dados são considerados insatisfatórios para aplicações clínicas para essas classes, cuja confiabilidade necessita ser alta.

Acredita-se que a idade do paciente também seja outra classe que se beneficiaria se a imagem de entrada do extrator não fosse somente a região da lesão segmentada, e sim toda a massa cinzenta do paciente, oque poderia gerar mais elementos discriminantes na etapa de classificação, como a Espessura cortical, e as áreas de massa cinzenta e branca. Mas como ambas as classes não tem tanta importância clinica como as demais relacionadas ás área do tumor, que acredita-se serem beneficiadas, bem como, a complexidade dos atributos profundo gerados tende á ser menor ao extrair somente a região segmentada, o estudo optou por aderir apenas á segmentação da região como entrada da da extração para todas as classes.

O menor Tempo de Treino do rótulo de idade foi obtido novamente pela combinação dos métodos VGG16 com NaiveBayes, oque pode ser explicado pelo menor número de

valores de classe possíveis, aliado ao princípio de funcionamento do NaiveBayes, enquanto o menor tempo de classificação foi obtido pelo método MLP, com 0,01 segundos, o que também corrobora na confirmação da velocidade do MLP após o treinamento mais lento.

Tabela 7 – Resultados da classificação para a combinação de extratores e classificadores por idade, para o banco de dados Brats2020.

Extrator	Classificador	Acurácia (%)	Precisão (%)	Recall (%)	F1-Score (%)	Tempo de Treino (s)	Tempo de Predição (s)
DenseNet121	KNN	32.30 ± 2.10	39.88 ± 2.87	31.94 ± 2.60	31.18 ± 2.20	0.55 ± 0.22	0.02 ± 0.00
	MLP	48.96 ± 2.47	53.51 ± 2.97	48.38 ± 3.05	47.40 ± 2.65	29.31 ± 1.52	0.01 ± 0.00
	NaiveBayes	24.03 ± 2.38	30.71 ± 2.14	24.45 ± 3.16	23.60 ± 2.26	0.42 ± 0.01	0.09 ± 0.00
	RandomForest	49.79 ± 2.01	57.04 ± 2.38	49.35 ± 3.11	48.45 ± 2.62	1113.15 ± 5.45	0.05 ± 0.00
	SVM_Linear	50.02 ± 3.54	55.44 ± 4.26	49.21 ± 4.07	48.71 ± 3.74	18.22 ± 1.90	0.28 ± 0.01
	SVM_Polynomial	47.67 ± 2.87	54.34 ± 3.95	47.02 ± 3.94	46.60 ± 3.53	39.05 ± 0.48	0.29 ± 0.00
	SVM_RBF	47.74 ± 2.50	57.34 ± 3.41	46.77 ± 3.18	47.36 ± 2.82	60.82 ± 6.49	0.52 ± 0.04
DenseNet201	KNN	34.98 ± 2.66	42.32 ± 3.66	34.84 ± 3.21	33.90 ± 2.64	1.53 ± 0.17	0.07 ± 0.01
	MLP	54.25 ± 2.05	60.17 ± 1.75	53.95 ± 3.24	52.79 ± 2.81	32.51 ± 1.21	0.01 ± 0.00
	NaiveBayes	26.45 ± 2.42	34.05 ± 3.41	26.77 ± 3.73	25.74 ± 3.05	0.77 ± 0.02	0.17 ± 0.00
	RandomForest	51.40 ± 3.12	60.06 ± 2.70	51.43 ± 3.59	50.82 ± 3.22	1413.55 ± 5.84	0.05 ± 0.00
	SVM_Linear	54.41 ± 2.92	60.76 ± 1.80	54.06 ± 3.75	53.41 ± 2.96	37.08 ± 4.02	0.69 ± 0.01
	SVM_Polynomial	54.12 ± 2.71	60.97 ± 2.49	53.74 ± 3.71	53.44 ± 3.26	90.75 ± 7.65	0.70 ± 0.01
	SVM_RBF	54.25 ± 2.28	60.86 ± 2.95	53.98 ± 2.88	53.38 ± 2.57	137.76 ± 10.43	1.84 ± 0.03
VGG16	KNN	31.44 ± 3.57	39.06 ± 4.20	31.24 ± 4.56	30.47 ± 3.79	0.26 ± 0.01	0.01 ± 0.00
	MLP	44.76 ± 3.06	49.97 ± 2.50	44.58 ± 3.31	43.84 ± 2.77	52.18 ± 1.53	0.01 ± 0.00
	NaiveBayes	15.17 ± 2.79	20.91 ± 3.52	14.81 ± 3.59	15.01 ± 3.09	0.24 ± 0.01	0.05 ± 0.00
	RandomForest	48.13 ± 2.60	55.05 ± 2.67	48.21 ± 3.81	47.15 ± 3.07	510.55 ± 2.53	0.05 ± 0.00
	SVM_Linear	43.54 ± 1.08	50.95 ± 1.85	42.84 ± 1.71	42.71 ± 1.09	13.23 ± 0.81	0.21 ± 0.01
	SVM_Polynomial	34.48 ± 3.59	45.38 ± 3.47	33.98 ± 4.17	34.77 ± 3.49	23.75 ± 0.06	0.22 ± 0.01
	SVM_RBF	45.72 ± 1.87	55.90 ± 2.72	45.20 ± 3.13	45.71 ± 2.82	32.39 ± 0.62	0.34 ± 0.02

No geral, tanto para a idade, como para o tempo de sobrevivência do paciente, as combinações não se mostraram tanto eficazes, seja pelo uso de apenas a região do tumor segmentada para extração, ou pela própria natureza desafiadora de delimitação de ambas as classes. A arquitetura VGG16 parece ter se saído um pouco pior em relação à DenseNet, com uma média de todas as métricas relativamente menores, em torno de 1%.

Já dentre os classificadores, O KNN e o NaiveBayes parecem ter as piores performances para essas classes, isto provavelmente se dando por algum tipo de baixo poder discriminativo dessas classes para com os atributos somado à dificuldade dos modelos de lidar com distribuições mais complexas, com muitas possibilidades de valores. O MLP, random Forest e SVM parecem ter uma capacidade maior de distinção, embora as métricas estejam entre 50 - 60%. Os desvios padrões de todos os métodos para ambas as classes é relativamente alto, sendo de 1.5 - 4%, o que indica alta variação de resultado de acordo com a imagem, e reforça a possibilidade dos dados estarem tendo baixa capacidade de representatividade das classes devido à segmentação da região de interesse.

A Tabela 8 ilustra os resultados da classificação para a classe Ausência/Presença, no banco de dados BraTS2020, cujo objetivo é detectar a presença do tumor na imagem, uma das classes mais difundidas na literatura, correspondendo à classificação binária, e segunda detecção da lesão na imagem, o que pode vir a confirmar o verdadeiro positivo da primeira etapa de detecção com *framework* YOLOv8. A melhor combinação extrator/classificador para detecção binária da presença do tumor na imagem foi novamente o DenseNet201 com o classificador SVM com Kernel Radial, com uma acurácia de 98,48%, Precisão de 98.05%, Recall de 97.96% e F1-score de 98.01%, o que indica relativa facilidade da

combinação extrator/classificador em discriminar a presença do tumor nas imagens. o desvio padrão é menor que 0.5%, oque indica que para todas as imagens de teste, o modelo teve performance estável.

A pior combinação de extrator/classificador foi para o VGG16 com naiveBayes, sendo a única combinação com acurácia abaixo de 50% para esta classe. No geral o VGG16 apresenta a menor performance de extração em relação do DenseNet. O KNN, para a classe de preseça/ausencia de tumor, obteve boa performance para todos os extratores, contrariando sua baixa performance para as classes de idade e tempo de sobrevida, oque pode ser explicado pela menor complexidade da detecção binária do tumor, frente aos multiplos valores possíveis de idade e sobrevida.

Tabela 8 – Resultados da classificação para a combinação de extratores e classificadores para classe de presença/ausência de tumores cerebrais na imagem, para o banco de dados Brats2020.

Extrator	Classificador	Acurácia (%)	precisão (%)	Recall (%)	F1-Score (%)	Tempo de Treino (s)	Tempo de Predição (s)
DenseNet121	KNN	97.98 ± 0.55	97.91 ± 0.58	97.95 ± 0.56	97.93 ± 0.57	1.04 ± 0.14	0.06 ± 0.01
	MLP	96.96 ± 0.19	96.91 ± 0.23	96.84 ± 0.19	96.88 ± 0.20	23.92 ± 1.14	0.01 ± 0.00
	NaiveBayes	79.88 ± 1.20	80.21 ± 1.28	80.97 ± 1.09	79.80 ± 1.25	0.06 ± 0.01	0.01 ± 0.00
	RandomForest	95.74 ± 0.34	95.77 ± 0.33	95.47 ± 0.43	95.61 ± 0.38	456.43 ± 3.23	0.02 ± 0.00
	SVM_Linear	93.86 ± 0.96	93.68 ± 1.03	93.74 ± 0.94	93.70 ± 0.98	16.44 ± 1.37	0.08 ± 0.00
	SVM_Polynomial	97.22 ± 0.53	97.19 ± 0.49	97.09 ± 0.63	97.14 ± 0.56	29.12 ± 2.32	0.10 ± 0.01
DenseNet201	SVM_RBF	97.60 ± 0.42	97.55 ± 0.43	97.53 ± 0.42	97.54 ± 0.42	53.47 ± 5.35	0.26 ± 0.02
	KNN	98.02 ± 0.24	98.02 ± 0.39	97.91 ± 0.37	97.96 ± 0.38	2.07 ± 0.37	0.11 ± 0.02
	MLP	97.16 ± 0.38	97.03 ± 0.41	97.14 ± 0.42	97.09 ± 0.41	25.51 ± 1.35	0.01 ± 0.00
	NaiveBayes	79.10 ± 1.38	79.95 ± 1.39	80.54 ± 1.21	79.06 ± 1.40	0.11 ± 0.01	0.01 ± 0.00
	RandomForest	95.32 ± 0.58	95.32 ± 0.61	95.32 ± 0.64	95.04 ± 0.68	560.27 ± 10.56	0.02 ± 0.01
	SVM_Linear	95.54 ± 0.56	95.50 ± 0.69	95.34 ± 0.54	95.41 ± 0.61	29.92 ± 1.53	0.11 ± 0.01
VGG16	SVM_Polynomial	97.58 ± 0.58	97.51 ± 0.59	97.52 ± 0.63	97.51 ± 0.61	90.04 ± 1.87	0.16 ± 0.00
	SVM_RBF	98.48 ± 0.28	98.05 ± 0.29	97.96 ± 0.31	98.01 ± 0.30	194.14 ± 1.38	1.10 ± 0.04
	KNN	97.38 ± 0.52	97.34 ± 0.51	97.27 ± 0.58	97.31 ± 0.54	0.51 ± 0.01	0.03 ± 0.00
	MLP	80.02 ± 1.46	79.66 ± 1.66	79.81 ± 2.24	79.55 ± 1.81	10.09 ± 0.35	0.01 ± 0.00
	NaiveBayes	49.16 ± 1.28	69.42 ± 1.44	56.00 ± 0.60	42.59 ± 1.17	0.03 ± 0.01	0.01 ± 0.00
	RandomForest	95.98 ± 0.84	96.13 ± 0.86	95.61 ± 0.97	95.84 ± 0.91	120.13 ± 0.61	0.03 ± 0.00
	SVM_Linear	84.18 ± 1.91	83.78 ± 2.02	84.49 ± 1.82	83.96 ± 1.99	10.22 ± 1.07	0.08 ± 0.00
	SVM_Polynomial	79.00 ± 1.26	78.68 ± 1.28	79.36 ± 1.08	78.76 ± 1.30	30.97 ± 3.21	0.17 ± 0.01
	SVM_RBF	80.56 ± 1.53	80.30 ± 1.57	81.05 ± 1.40	80.36 ± 1.58	42.97 ± 3.29	0.36 ± 0.01

Tabela 9 – Resultados da classificação da melhor combinação extrator/classificador para os marcadores de detecção (*Detection*), tumor inteiro (*Whole Tumor*), núcleo necrótico/tumor não realçador (*NCR/NET*), edema peritumoral (*ED*) e pixels de fundo (*Background*) no banco de dados BraTS2020.

Extrator	Classificador	Classes	Acurácia(%)
DenseNet201	SVMRBF	<i>Detection</i>	98.48 ± 0.28
		<i>Whole Tumor</i>	67.56 ± 1.03
		<i>NCR/NET</i>	54.78 ± 0.98
		<i>ED</i>	75.63 ± 0.72
		<i>Background</i>	97.73 ± 0.28

Para sintetizar a performance da melhor combinação em relação às demais classes abordadas para o dataset BraTS2020, as classes de tumor inteiro, núcleo necrótico/tumor sem realce (NCR/NET), edema peritumoral (ED) e pixels de fundo foram resumidas na tabela 9, que mostra os resultados da classificação da melhor combinação do pipeline na classificação.

A maior acurácia obtida foi para detecção foi obviamente para a classe de presença de tumor na imagem, com uma acurácia média de 98,48%, indicando facilidade na detecção da patologia. Isso se deve ao pré-processamento aplicado ao banco de dados de regiões tumorais BraTS2020, que reúne as modalidades flair, t1, t1c1 e t2 em uma única imagem, utilizando o método de interpolação em um processo de *DataFusion*. Isso facilita a detecção do tumor, que aparece em alto contraste na imagem.

A maior acurácia obtida foi para detecção foi obviamente para a classe de presença de tumor na imagem, com uma acurácia média de 98,48%, indicando facilidade na detecção da patologia. Para a classe tumoral completa, representando o núcleo tumoral e a região de edema articular, a acurácia média obtida foi de 67,56%. Para a classe NCR/NET, correspondente ao tecido tumoral morto e às partes do tumor não realçadas após a administração de contraste, a acurácia média obtida foi de 54,78%. Para a classe ED, correspondente à região de edema ao redor do tumor, causada pela infiltração de células tumorais ou resposta inflamatória do cérebro, um tecido sutilmente alterado no exame, a acurácia média obtida foi de 75,65%. Para a classe de fundo, que diz respeito a todos os outros pixels da imagem que não correspondem ao tumor, a acurácia média obtida foi de 97,73%.

Os valores menores de acurácia para as classes ED e NCR/NET podem se dar por condições de contraste e textura mais sutis, ou Variabilidade morfológica e volumétrica advinda dessas classes, o que as tornam naturalmente mais difíceis de discernir. O *whole tumor*, como é constituído da soma dessas partes, também tem a acurácia afetada.

No geral, o modelo proposto pela combinação do extrator/classificador DenseNet201 com o SVM RBF obteve resultado satisfatório na classificação da presença de

tumor, tecidos edematosos, região central do tumor e região de fundo, indicando ampla capacidade do modelo em classificar tais atributos em imagens médicas, e embora as classes de idade e tempo de sobrevivência tenham tido resultados mais baixos, o conjunto de todas essas classes podem fornecer pontos de vista valiosos para os médicos especialistas, sendo o sistema projetado para fornecer esses atributos a partir da imagem, um sistema CAD. Obviamente, a aplicação clínica do método teria de vir com novas etapas de testes de caso, experimentações e ajustes.

A Figura 16 ilustra de maneira resumida, o resultado de cada etapa da metodologia proposta: em (a) a imagem de entrada é ilustrada em tons de cinza; em (b) o tumor é detectado na imagem. Em (c) ocorre a segmentação. Em (d) os atributos são classificados, com a presença do tumor confirmada por meio da detecção e classificação. Em (e) ocorre o pré-diagnóstico com LLM, por meio dos atributos classificados.

5.4 Validação de Resultados (Tumor Cerebral) - Comparação com a Literatura

Nesta seção, os resultados obtidos pelos modelos desenvolvidos neste estudo para segmentação e classificação de tumores cerebrais por ressonância magnética são comparados com trabalhos encontrados na literatura.

5.4.1 Comparação - Segmentação

Tabela 10 – Tabela comparativa do método de segmentação abordado (*Divisible Cell Segmentation*) com métodos encontrados na literatura que também se utilizaram da base TCGA.

Método	Treino	ACC(%)
Método Proposto (<i>Divisible Cell</i>)	300 iterações (YOLODetect)	99.80 ± 0.09
CDS (Detectron2) (NASCIMENTO et al., 2023a)	100000 iterações	99.41 ± 0.62
Seg-Net (TRIPATHI; VERMA; SHARMA, September, 2021)	50 épocas	96.193 ± -
CCN-PR-Seg-Net (TRIPATHI; VERMA; SHARMA, September, 2021)	50 épocas	98.65 ± -
Seg-Net (TRIPATHI; VERMA; SHARMA, September, 2021)	100 épocas	97.304 ± -
U-Net (TRIPATHI; VERMA; SHARMA, September, 2021)	100 épocas	98.287 ± -
CCN-PR-Seg-Net (TRIPATHI; VERMA; SHARMA, September, 2021)	100 épocas	98.961 ± -

A tabela 10 trás a comparação entre o método proposto, *Divisible cell segmentation* com inicialização pelo YOLOv8, treinado com no máximo 300 iterações, e os métodos encontrados na literatura, de ([TRIPATHI; VERMA; SHARMA, September, 2021](#)), que fizeram uso de mesma base. O estudo do autor buscou realizar a segmentação com as redes Seg-Net e CCN-PR-Seg-Net em duas etapas de teste, mudando a quantidade de épocas do treinamento e usando a U-net com 100 épocas, o dataset usado foi o TCGA.

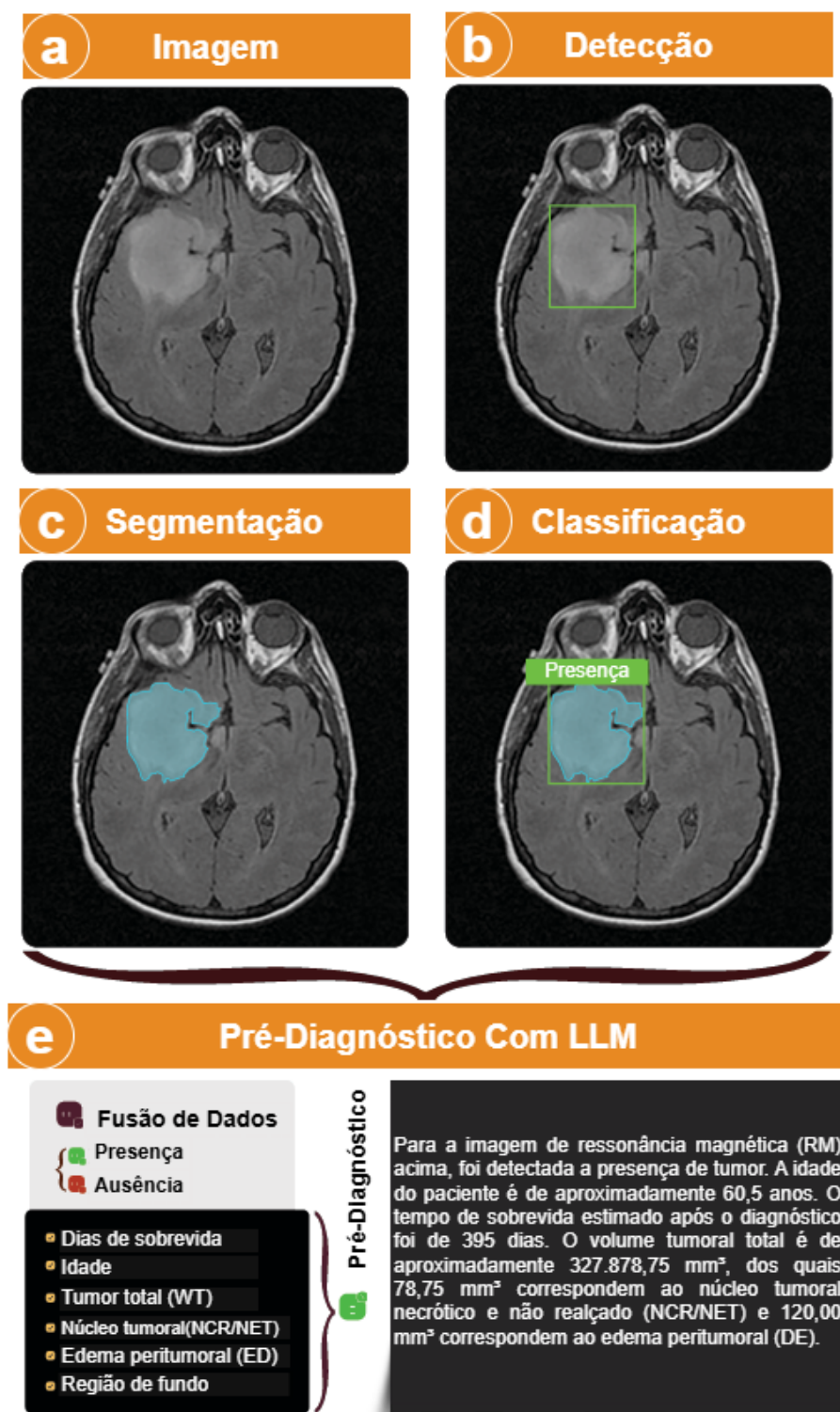


Figura 16 – Resultado de cada etapa da metodologia proposta: Em (a) imagem em tons de cinza; (b) detecção; (c) segmentação; (d) classificação; (e) pré-diagnóstico com LLM.

O método proposto neste estudo superou os modelos Seg-net e CCN-PR-Seg-net de 50 épocas em 3,637% e 1,15%, respectivamente. Já para as 100 épocas de treinamento, o modelo proposto superou os modelos Seg-Net, U-Net e CCN-PR-Seg-Net em 2,496%,

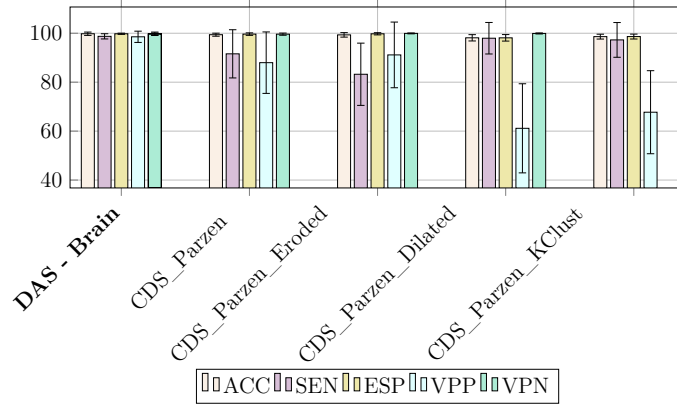


Figura 17 – Ilustração gráfica da tabela de comparação de métricas 11.

1,513% e 0,839%, respectivamente. Esses resultados comparativos demonstram o potencial do método proposto no que diz respeito à segmentação, mesmo que de maneira mais superficial.

Embora a abordagem proposta tenha sido apenas levemente superior em termos de obtenção de resultados, alguns questionamentos são levantados. Nesse sentido, nota-se que os métodos baseados em CNN usados para comparação não apresentam a métrica de desvio-padrão. Esse dado é importante pois informa a dispersão da distribuição dos dados. Então, mesmo o método *Divisible Cell*-YOLOv8 sendo superior ao *CCN-PR-Seg-net* em relação à acurácia, destaca-se que o desvio padrão desse método não é informando, logo, não se pode avaliar de forma profunda e eficaz qual método é mais eficiente em termos de estabilidade de performance para as amostras de teste, mas avalia-se a equivalência entre ambos os resultados apresentados pelos autores.

Outro ponto a destacar é que o método proposto, foi treinado apenas uma vez para esta comparação, com o limite de 300 iterações, enquanto os outros métodos necessitaram de duas sessões de treinamento, com 50 e 100 épocas. Obviamente um estudo mais aprofundado sobre as arquiteturas CNN, ou adoção de métodos de *fine-tuning* diversificados, poderiam aumentar significativamente a eficácia das redes abordadas por (TRIPATHI; VERMA; SHARMA, September, 2021).

Tabela 11 – Comparação entre o método proposto, e os métodos desenvolvidos por (NASCIMENTO et al., 2023a).

Método	ACC(%)	SEN(%)	ESP(%)	VPP(%)	VPN(%)
Método Proposto (<i>Divisible Cell</i>)	99.80 ± 0.68	99.74 ± 1.06	99.79 ± 0.25	98.55 ± 2.30	99.78 ± 0.70
CDS_Parzen	99.41 ± 0.62	91.59 ± 9.85	99.65 ± 0.51	87.98 ± 12.56	99.74 ± 0.37
CDS_Parzen_Eroded	98.61 ± 1.44	83.24 ± 12.72	99.78 ± 0.45	91.15 ± 13.42	99.53 ± 0.45
CDS_Parzen_Dilated	98.14 ± 1.29	97.97 ± 6.44	98.13 ± 1.33	61.15 ± 18.21	99.94 ± 0.18
CDS_Parzen_KClust	98.62 ± 0.99	97.28 ± 7.10	98.65 ± 1.00	67.73 ± 16.97	99.92 ± 0.21

A Tabela 11 aborda diferentes valores de métricas para o modelo proposto neste estudo e os métodos encontrados na literatura de (NASCIMENTO et al., 2023a), que utilizou o mesmo conjunto de dados para testes.

([NASCIMENTO et al., 2023a](#)) também buscou conciliar a detecção e segmentação realizada através de CNN, sendo o modelo aplicado, o *Detectron2*, seguido por uma gama de métodos de *fine-tuning*, baseados no Janelamento de Parzen, operações morfológicas de erosão, dilatação e clusterização sobre a máscara binária de detecção do *Detectron2*.

Com relação a acurácia, embora relativamente equiparados, o método proposto obteve a maior acurácia da comparação, sendo 0,39% maior que o CDS_parzen. A sensibilidade do método proposto foi maior em 0,77% o que o método CDS_Parzen_Dilated, o método com a maior sensibilidade. A Especificidade do método proposto foi praticamente equiparada á do CDS_Parzen_Eroded em termos de média, sendo maior em 0,01. A VPP do método proposto foi a maior, sendo maior em 7,4% que o método CDS_Parzen_Eroded. E o VPN do método proposto foi menor que a VPN dos métodos CDS_Parzen_Dilated e CDS_Parzen_KClust, sendo menor em 0,16% e 0,14% respectivamente.

Em suma, ([NASCIMENTO et al., 2023a](#)) obteve diferentes características ao modificar o método de fine tuning, sendo alguns métodos mais eficazes em detectar os verdadeiros positivos, como o CDS_Parzen_Dilated e CDS_Parzen_KClust, e outros em identificar os verdadeiros negativos, como o CDS_Parzen. Ainda assim, em termos de desvio padrão, o método proposto neste estudo se mostrou mais estável, com todas as métricas mostrando um desvio menor que 1,5%.

Ambos os modelos, o proposto e o comparado, buscaram avaliar a eficácia da segmentação da região que circunda as bordas do tumor de maneira mais completa, se utilizando de mais métricas além da acurácia, o que busca trazer uma análise performática mais aprofundada do ponto de vista dos resultados.

Tanto o melhor método de ([NASCIMENTO et al., 2023a](#)) bem como o proposto neste estudo, obtiveram as métricas de Especificidade (ESP) e Valor Preditivo Negativo (VPN) acima de 98%, confirmando tanto a eficácia dos métodos em diferenciar a região de fundo da região de interesse, mas também na confiabilidade de definir o que seria a região de fundo na imagem. O gráfico 17 ilustra os resultados da tabela 11, é perceptível uma maior estabilidade métrica para o modelo proposto, tendo uma maior estabilidade mesmo para as métricas de sensibilidade e VPP, o que confirma a hipótese de uma maior consistência de pixels verdadeiros positivos em detrimento dos falsos verdadeiros.

Tabela 12 – resultados do teste estatístico de Kolmogorov-Smirnov aplicados na comparação entre o método proposto neste estudo e os métodos desenvolvidos por ([NASCIMENTO et al., 2023a](#)).

Modelos	ACC(%)	ESP(%)	SEN(%)	VPP(%)	VPN(%)
<i>DivisibleCell CDS_Parzen</i>	V	x	x	x	x
<i>DivisibleCell CDS_Parzen_Eroded</i>	x	x	x	x	x
<i>DivisibleCell CDS_Parzen_Dilated</i>	x	x	x	x	x
<i>DivisibleCell CDS_Parzen_KClust</i>	x	x	x	x	x

Levando em consideração a equivalência estatística entre algumas das métricas para o método proposto e os métodos de (NASCIMENTO et al., 2023a), foi-se realizado o teste estatístico de Kolmogorov-Smirnov afim de se avaliar a equivalência estatística entre os vetores de métricas utilizados para gerar as médias e desvio de ambos os modelos.

A tabela 12 ilustra o resultado do desse teste para o método proposto por este estudo em comparação direta com os vetores métricos obtidos a partir do trabalho de (NASCIMENTO et al., 2023a). Pode-se observar que as acurácias do CDS_parzen e do método proposto são estatisticamente equivalentes, o que indica uma relativa similaridade com a distribuição dos resultados métricos no vetor de atributos, com sua homogeneidade. Os demais casos indicam diferença estatística, o que representa que os vetores de atributos que geraram as médias e os desvios são diferentes em termos estatísticos, de forma á buscar um maior respaldo ao indicar superioridade métrica entre os modelos.

5.4.2 Comparação - Classificação

Tabela 13 – Comparação com trabalhos encontrados no estado da arte para classificação do tempo de sobrevida de pacientes para o banco de dados de validação Brats2020.

Autor	Extrator	Classificador	Acurácia(%)
Proposto	DenseNet201	SVM_RBF	53.92 $\pm$ 2.58
(PATEL et al., 2021)	U-Net Architecture	U-Net Architecture	49.5 $\pm$ -
(AKBAR; FATICHAH; SUCIATI, 2020)	Modified MobileNet	Modified MobileNet	40.2 $\pm$ -
(AGRAVAT; RAVAL, 2020)	3D Fully CNN	3D Fully CNN	51.7 $\pm$ -
(PEI; MURAT; COLEN, 2021)	3D srUNet	3D srUNet	43.0 $\pm$ -

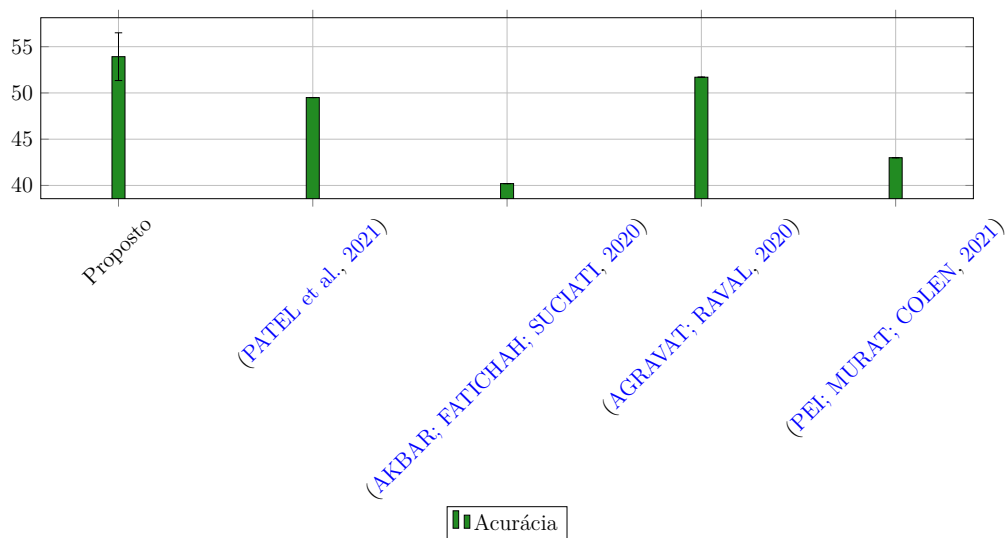


Figura 18 – Comparação gráfica da precisão dos métodos propostos e dos métodos da literatura, na tabela 13.

A tabela 13 apresenta a comparação entre a melhor combinação extrator/classificador adotada e os métodos propostos no estado da arte que se propuseram a realizar a classifi-

cação do tempo de sobrevida dos pacientes para o banco de dados Brats2020. O gráfico 19 ilustra visualmente os valores de comparação da tabela.

O modelo proposto, cuja combinação adotada foi o DenseNet201 + SVM RBF, obteve a maior acurácia, sendo 2,22% maior (53,92 versus 51,7) do que o segundo melhor modelo da comparação, a 3D Fully CNN de (AGRAVAT; RAVAL, 2020), que tem um bom rendimento para a extração de texturas. Isso indica o potencial de classificação do modelo proposto frente à comparação com a literatura, que se utilizou de arquiteturas robustas para realizar a mesma classificação. Vale ressaltar que o método busca extrair e classificar sobre a segmentação da região tumoral na tentativa de reduzir a complexidade dos atributos profundos extraídos, o que pode diminuir a eficácia do modelo sobre a classe de tempo de sobrevida e idade do paciente, e melhorar a eficácia na classificação das classes que envolvem a distinção das regiões tumorais.

Em suma, o que pode ser observado a partir dos resultados da tabela é que classificar o tempo de sobrevivência do paciente é uma tarefa complexa, como pode ser visto tanto pelas métricas abaixo de 50% encontradas na literatura, quanto pelos métodos, que geralmente são baseados na extração de atributos com CNN. Espera-se que os atributos profundos sejam capazes de extrair características com mais poder discriminativo para essa classe em questão.

Destaca-se, dessa forma, a nítida melhora métrica obtida pelo modelo proposto, que buscou aliar a extração de atributos profundos com a utilização de modelos CNN, ao bom desempenho de classificadores clássicos quando bem parametrizados, cujos melhores parâmetros para a base foram adotados com base em mecanismos de busca.

Tabela 14 – Comparação com o estado da arte para classificação binária (presença) de tumor cerebral no banco de dados BraTS2020.

Autor	Extrator	Classificador	Acurácia(%)
Proposto	DenseNet201	SVMRBF	98.48 ± 0.28
(IQBAL et al., 2024)	FusionNet	FusionNet	97.53 ± 0.24
(GAUTAM; AHLAWAT; MITTAL, 2022)	PCA	SVMRBF	96.0 ± -
(GAUTAM; AHLAWAT; MITTAL, 2022)	PCA	DecisionTree	96.0 ± -
(GAUTAM; AHLAWAT; MITTAL, 2022)	PCA	NaiveBayes	88.0 ± -
(GAUTAM; AHLAWAT; MITTAL, 2022)	PCA	SVMLinear	95.0 ± -

A tabela 14 compara o modelo proposto e outros trabalhos na literatura que também utilizaram do BraTS2020 para classificação binária de tumores na imagem, ou seja, sua presença/ausência. O gráfico 19 representa visualmente os valores comparados na tabela.

O modelo proposto, composto por DenseNet201 e SVM RBF, obteve a maior acurácia, com 98,48%, sendo 0,95% maior (98,58 versus 97,53) em relação ao segundo maior índice de acurácia, a FusionNet de (IQBAL et al., 2024), cujo modelo adotado foi totalmente automático.

Vale destacar o bom desempenho na utilização de extratores profundos, ou técnicas de processamento na imagem na geração de atributos para a classificação, aliado á método clássicos de *Machine Learning*, que também foi aplicado por (GAUTAM; AHLAWAT; MITTAL, 2022).

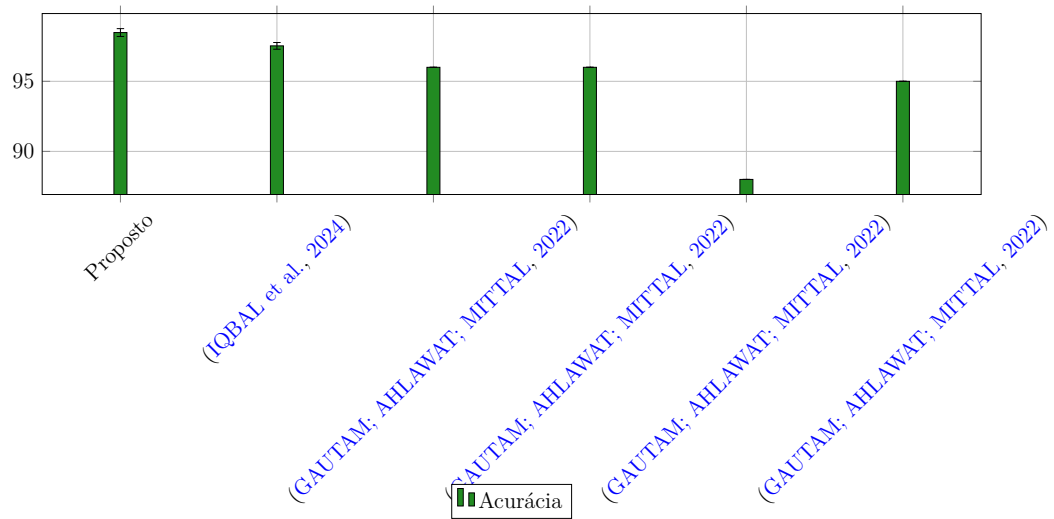


Figura 19 – Comparação gráfica da precisão dos métodos propostos e dos métodos de última geração para a tabela 14.

6 Conclusão

Este estudo apresentou um método totalmente automático para detecção e segmentação de tumores cerebrais em imagens de ressonância magnética, combinando uma etapa inicial de identificação de regiões suspeitas com o framework YOLOv8 e um refinamento de segmentação baseado em Células Computacionais Inteligentes (ICC). A proposta inovadora integrou ainda extração de atributos profundos via CNNs e classificadores clássicos otimizados por Grid Search, além de um módulo de pré-diagnóstico textual baseado em um modelo de LLM, com o objetivo de entregar um sistema end-to-end que abrange desde a localização até a interpretação clínica das lesões com um pré diagnóstico textual objetivo.

Os resultados obtidos demonstraram desempenho satisfatório: A etapa de detecção alcançou acurácia média de 99,87%, garantindo alta confiabilidade na identificação das regiões tumorais, enquanto a segmentação com o método *Divisible Cell Segmentation* atingiu 99,80% de acurácia, com tempo médio inferior a 0,8 s por imagem. Na fase de classificação, o modelo DenseNet201+SVM-RBF obteve 98,48% de acurácia na detecção binária de presença tumoral, e 53,92% na classificação da classe de tempo de sobrevivência, superando os métodos comparados na literatura que se propuseram às mesmas classes.

Do ponto de vista prático, os objetivos propostos foram integralmente atendidos: o framework YOLOv8 foi configurado e treinado para detectar tumores cerebrais com alta acurácia, conforme previsto; em seguida, o algoritmo de Células Sintéticas Computacionais ampliou os resultados de segmentação, elevando as métricas de SEN, VPP e VPN a níveis de estado da arte; o pipeline de classificação reuniu extratores de características de CNN e classificadores de machine learning, com a busca de hiperparâmetros otimizados por Grid Search, garantindo robustez e interpretabilidade nas predições; a metodologia foi validada em múltiplas bases de dados de tumor cerebral, comprovando sua capacidade de generalização; por fim, os resultados de classificação foram integrados a um módulo de LLM com o objetivo de produzir pré-diagnósticos textuais de forma autônoma e contextualizada no formato clínico esperado. A análise quantitativa e qualitativa desse estudo inicia a discussão á cerca de sua aplicação em ecossistemas de Health of Things (HoT), reduzindo o tempo de laudo e minimizando a variabilidade interobservador.

Reconhecem-se, porém, limitações a serem elencadas em trabalhos futuros: Aprimoramento do estágio inicial de detecção com técnicas alternativas de inicialização para casos de tumores periféricos, soluções mais eficazes que o YOLOv8; Teste e comparação de outros métodos de conexão entre os pontos celulares para formar uma melhor máscara de segmentação a partir do mapa de pontos do *divisible cell segmentation*, através de métodos como o *match encode* ;Incorporar mais extratores e mais classificadores no processo de

pipeline de classificação/extração e explorar LLMs mais profundos para enriquecer o pré-diagnóstico com recomendações e alertas adaptados às condições clínicas, além de análises mais concisas na capacidade de abstração e análise de delírio do modelo. Essas evoluções podem ampliar ainda mais a discussão á cerca da aplicabilidade do método, bem como trazer análises mais robustas de possibilidade de aplicação para uma patologia tão severa quanto o tumor cerebral.

Vale ressaltar também que incorporar médicos especialistas em etapas futuras de avaliação do sistema traria ganhos substanciais: além de validar clinicamente os resultados obtidos (por meio de estudos com leitores às cegas e análise de concordância interobservador), os profissionais da saúde podem auxiliar a identificar casos-limite, fornecer *feedbacks* qualitativos para refinamento das heurísticas e ajustar os limiares de confiança nas diferentes etapas do método com base em relevância clínica. A participação direta dos clínicos também fortalece a confiança dos usuários finais, facilita a adoção em fluxos de trabalho reais em ambientes clínicos, aperfeiçoa a rotulação dos dados e cria um pipeline mais crível do ponto de vista de aplicação no mundo real que equilibra desempenho automatizado e julgamento médico, acelerando inclusive a aceitação regulatória.

Referências

- ABADI, M. et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016. Citado na página 36.
- ADUSUMALLI, H. et al. Face mask detection using opencv. In: *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. Tirunelveli, India: IEEE, 2021. p. 1304–1309. Citado 2 vezes nas páginas 40 e 64.
- AGRAVAT, R. R.; RAVAL, M. S. 3d semantic segmentation of brain tumor for overall survival prediction. In: SPRINGER. *International MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2020. p. 215–227. Citado 4 vezes nas páginas 31, 33, 90 e 91.
- AI, M. *meta-llama/Llama-3.1-8B*. 2024. Hugging Face. Versões Llama 3.1 instrução-sintonizada suportam janela de contexto de 128 000 tokens e 8 idiomas. Disponível em: <<https://huggingface.co/meta-llama/Llama-3.1-8B>>. Citado na página 56.
- AI, M. *meta-llama/Meta-Llama-3-8B*. 2024. Hugging Face. Variantes de 8 B e 70 B parâmetros; usa Grouped-Query Attention (GQA) para escalabilidade durante inferência. Disponível em: <<https://huggingface.co/meta-llama/Meta-Llama-3-8B>>. Citado 2 vezes nas páginas 56 e 64.
- AKBAR, A. S.; FATICHAH, C.; SUCIATI, N. Modified mobilenet for patient survival prediction. In: SPRINGER. *International MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2020. p. 374–387. Citado 3 vezes nas páginas 31, 33 e 90.
- AKTER, A. et al. Robust clinical applicable cnn and u-net based algorithm for mri classification and segmentation for brain tumor. *Expert Systems with Applications*, Elsevier, v. 238, p. 122347, 2024. Citado na página 21.
- ALIBRAHIM, H.; LUDWIG, S. A. Hyperparameter optimization: Comparing genetic algorithm against grid search and bayesian optimization. In: IEEE. *2021 IEEE Congress on Evolutionary Computation (CEC)*. Kraków, Poland: IEEE, 2021. p. 1551–1559. Citado na página 50.
- ALLAH, A. M. G.; SARHAN, A. M.; ELSHENNAWY, N. M. Edge u-net: Brain tumor segmentation using mri based on deep u-net model with boundary information. *Expert Systems with Applications*, Elsevier, v. 213, p. 118833, 2023. Citado 2 vezes nas páginas 28 e 33.
- ANAVA, O.; LEVY, K. k*-nearest neighbors: From global to local. *Advances in neural information processing systems*, v. 29, 2016. Citado 2 vezes nas páginas 46 e 47.
- ARUMUGAM, M. et al. Crossover smell agent optimized multilayer perceptron for precise brain tumor classification on mri images. *Expert Systems with Applications*, Elsevier, v. 238, p. 121453, 2024. Citado 2 vezes nas páginas 31 e 33.
- AXIOS. *Meta releases new AI models and updates AI assistant*. 2024. Axios. “Llama 3 has been pre-trained on over 15 trillion tokens”. Disponível em:

<<https://www.axios.com/2024/04/18/meta-ai-llama-3-artificial-intelligence>>. Citado na página 56.

AYADI, W. et al. Deep cnn for brain tumor classification. *Neural processing letters*, Springer, v. 53, p. 671–700, 2021. Citado na página 20.

BAHDANAU, D.; CHO, K.; BENGIO, Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014. Citado na página 56.

BARBU, T.; TĂNASE, G. Level-set based active contour segmentation technique using a nonlinear second-order diffusion-based model. *Discrete and Continuous Dynamical Systems-S*, Discrete and Continuous Dynamical Systems-S, p. 0–0, 2025. Citado na página 40.

BENGIO, Y. et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, Now Publishers, Inc., v. 2, n. 1, p. 1–127, 2009. Citado na página 51.

BHIMAVARAPU, U.; CHINTALAPUDI, N.; BATTINENI, G. Brain tumor detection and categorization with segmentation of improved unsupervised clustering approach and machine learning classifier. *Bioengineering*, Multidisciplinary Digital Publishing Institute, v. 11, n. 3, p. 266, 2024. Citado na página 18.

BISCHL, B. et al. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, v. 13, n. 2, p. e1484, 2023. Citado 2 vezes nas páginas 50 e 51.

BISHOP, C. M.; NASRABADI, N. M. *Pattern recognition and machine learning*. Nova York, USA: Springer, 2006. v. 1. 15–738 p. Citado 2 vezes nas páginas 49 e 50.

BOMPOU, E. et al. Comparative evaluation of ct and mri in the preoperative staging of colon cancer. *Scientific Reports*, Nature Publishing Group UK London, v. 14, n. 1, p. 17145, 2024. Citado na página 19.

BROWN, T. et al. Language models are few-shot learners. *Advances in neural information processing systems*, v. 33, p. 1877–1901, 2020. Citado na página 56.

CHEN, M. et al. Towards understanding hierarchical learning: Benefits of neural representations. *Advances in Neural Information Processing Systems*, v. 33, p. 22134–22145, 2020. Citado na página 51.

CHENG, J. et al. Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PloS one*, Public Library of Science San Francisco, CA USA, v. 10, n. 10, p. e0140381, 2015. Citado na página 28.

CHOWDHURY, M. *Padding and Strides in CNN*. 2024. Medium. 5 min read; accessed 18 June 2025. Disponível em: <<https://medium.com/@minhazc.engg/padding-and-strides-in-cnn-58dc56493887>>. Citado na página 54.

CSURKA, G. et al. Semantic image segmentation: Two decades of research. *Foundations and Trends® in Computer Graphics and Vision*, Now Publishers, Inc., v. 14, n. 1-2, p. 1–162, 2022. Citado na página 40.

- DAGHER, I.; ABOUD, E. Combining contour-based and region-based in image segmentation. *F1000Research*, v. 12, p. 1312, 2024. Citado na página 40.
- DEVELOPERS, T. Tensorflow. *Zenodo*, 2022. Citado na página 64.
- DIPU, N. M.; SHOHAN, S. A.; SALAM, K. Deep learning based brain tumor detection and classification. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TECHNOLOGIES - CONIT. 2021, Hubli, India: **IEEE**, 2021. p. 1–6. Citado 2 vezes nas páginas 29 e 33.
- EDELSBRUNNER, H.; MÜCKE, E. P. Three-dimensional alpha shapes. *ACM Transactions On Graphics (TOG)*, ACM New York, NY, USA, v. 13, n. 1, p. 43–72, 1994. Citado na página 43.
- GALIĆ, I. et al. Machine learning empowering personalized medicine: A comprehensive review of medical image analysis methods. *Electronics*, MDPI, v. 12, n. 21, p. 4411, 2023. Citado na página 20.
- GARCIA-GARCIA, A. et al. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017. Citado na página 52.
- GAUTAM, S.; AHLAWAT, S.; MITTAL, P. Binary and multi-class classification of brain tumors using mri images. *Int. J. Exp. Res. Rev.*, v. 29, p. 1–9, 2022. Citado 4 vezes nas páginas 30, 33, 91 e 92.
- GIUDICI, P.; GRAMEGNA, A.; RAFFINETTI, E. Machine learning classification model comparison. *Socio-Economic Planning Sciences*, Elsevier, v. 87, p. 101560, 2023. Citado na página 20.
- GOH, E. et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Network Open*, American Medical Association, v. 7, n. 10, p. e2440969–e2440969, 2024. Citado na página 22.
- GONZALEZ, R. C.; WOODS, R. E. *Instructor’s manual for digital image processing*. Reading, MA: Addison-Wesley, 1992. Citado na página 20.
- GROWCOTT, S. et al. Inter-observer variability in target volume delineations of benign and metastatic brain tumours for stereotactic radiosurgery: results of a national quality assurance programme. *Clinical Oncology*, Elsevier, v. 32, n. 1, p. 13–25, 2020. Citado na página 19.
- GUIBAS, L. *Introduction to Alpha Shapes*. 2021. Stanford Graphics Lab. Acessado em junho de 2025. Disponível em: <https://graphics.stanford.edu/courses/cs268-11-spring/handouts/AlphaShapes/as_fisher.pdf>. Citado na página 44.
- HAN, Q. et al. Dm-cnn: Dynamic multi-scale convolutional neural network with uncertainty quantification for medical image classification. *Computers in biology and medicine*, Elsevier, v. 168, p. 107758, 2024. Citado na página 20.
- HANIF, F. et al. Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment. *Asian Pacific journal of cancer prevention: APJCP*, v. 18, n. 1, p. 3, 2017. Citado na página 18.

- HARRIS, C. R. et al. Array programming with numpy. *Nature*, Nature Publishing Group UK London, v. 585, n. 7825, p. 357–362, 2020. Citado na página 64.
- HASSANEIN, A. S. et al. A survey on hough transform, theory, techniques and applications. *arXiv preprint arXiv:1502.02160*, 2015. Citado na página 43.
- HAVAEI, M. et al. Brain tumor segmentation with deep neural networks. *Medical image analysis*, Elsevier, v. 35, p. 18–31, 2017. Citado na página 30.
- HENRY, T. et al. Brain tumor segmentation with self-ensembled, deeply-supervised 3d u-net neural networks: a brats 2020 challenge solution. In: SPRINGER. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6*. Cham, Switzerland: Springer, 2021. p. 327–339. Citado na página 35.
- HOWARD, J.; GUGGER, S. Fastai: A layered api for deep learning. *Information*, MDPI, Basel, Switzerland, v. 11, n. 2, p. 108, February, 2020. Citado na página 29.
- HU, P. et al. A multi-level thresholding image segmentation algorithm based on equilibrium optimizer. *Scientific Reports*, Nature Publishing Group UK London, v. 14, n. 1, p. 29728, 2024. Citado na página 40.
- HUANG, G. et al. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. Honolulu, HI, USA: IEEE, 2017. p. 4700–4708. Citado 2 vezes nas páginas 54 e 55.
- ILLINGWORTH, J.; KITTLER, J. The adaptive hough transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, New York, USA, n. 5, p. 690–698, september. 1987. Citado na página 70.
- IQBAL, S. et al. Hybrid deep spatial and statistical feature fusion for accurate mri brain tumor classification. *Frontiers in Computational Neuroscience*, Frontiers Media SA, v. 18, p. 1423051, 2024. Citado 4 vezes nas páginas 30, 33, 91 e 92.
- JAIN, S.; JAIN, V. Novel approach to classify brain tumor based on transfer learning and deep learning. *International Journal of Information Technology*, Springer, v. 15, n. 4, p. 2031–2038, 2023. Citado na página 22.
- JANARDHAN, K. et al. Mri fusion for glioma grade classification. In: IEEE. *2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. Bangalore, India: IEEE, 2024. p. 1–6. Citado na página 22.
- JOHN, G. H.; LANGLEY, P. Estimating continuous distributions in bayesian classifiers. *arXiv preprint arXiv:1302.4964*, 2013. Citado 2 vezes nas páginas 47 e 48.
- KHALIGHI, S. et al. Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment. *NPJ precision oncology*, Nature Publishing Group UK London, v. 8, n. 1, p. 80, 2024. Citado na página 18.
- KRAMER, O.; KRAMER, O. Scikit-learn. *Machine learning for evolution strategies*, Springer, p. 45–53, 2016. Citado na página 64.

- KRISHNAPRIYA, S.; KARUNA, Y. A survey of deep learning for mri brain tumor segmentation methods: Trends, challenges, and future directions. *Health and Technology*, Springer, v. 13, n. 2, p. 181–201, 2023. Citado na página 21.
- KRITHIGA, R.; GEETHA, P. Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review. *Archives of Computational Methods in Engineering*, Springer, v. 28, n. 4, p. 2607–2619, 2021. Citado na página 20.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Ieee, v. 86, n. 11, p. 2278–2324, 2002. Citado na página 52.
- LEMING, M. J. et al. Challenges of implementing computer-aided diagnostic models for neuroimages in a clinical setting. *NPJ Digital Medicine*, Nature Publishing Group UK London, v. 6, n. 1, p. 129, 2023. Citado na página 20.
- LI, T. et al. Survey and application of target detection algorithms based on deep learning. *World Scientific Research Journal*, Boya Century Publishing Limited, v. 6, n. 8, p. 76–85, 2020. Citado na página 35.
- LIU, Z. et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics*, v. 9, n. 5, p. 1303, 2019. Citado na página 20.
- LOUPPE, G. *Understanding random forests: From theory to practice*. Tese (Doutorado) — Universite de Liege (Belgium), 2014. Citado na página 49.
- LUO, Y.; MI, Z.; TAO, W. Deepdt: Learning geometry from delaunay triangulation for surface reconstruction. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press, 2021. v. 35, n. 3, p. 2277–2285. Citado na página 44.
- MANUWEERA, T. et al. Barriers and facilitators for participation in brain magnetic resonance imaging (mri) scans in cancer research: a feasibility and acceptability analysis. *Research Square*, p. rs–3, 2024. Citado na página 19.
- MARTUCCI, M. et al. Magnetic resonance imaging of primary adult brain tumors: state of the art and future perspectives. *Biomedicines*, MDPI, v. 11, n. 2, p. 364, 2023. Citado na página 20.
- MASUCCI, M. et al. Bridging the divide: a review on the implementation of personalized cancer medicine. *Journal of Personalized Medicine*, MDPI, v. 14, n. 6, p. 561, 2024. Citado na página 19.
- MENG, X. et al. The application of large language models in medicine: A scoping review. *Iscience*, Elsevier, v. 27, n. 5, 2024. Citado na página 22.
- MINAEE, S. et al. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 44, n. 7, p. 3523–3542, 2021. Citado 2 vezes nas páginas 20 e 40.
- MITTAL, P. A comprehensive survey of deep learning-based lightweight object detection models for edge devices. *Artificial Intelligence Review*, Springer, v. 57, n. 9, p. 242, 2024. Citado 2 vezes nas páginas 35 e 36.

- MOREIRA, A.; SANTOS, M. Y. Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points. INSTICC Press, 2007. Citado na página 43.
- MOUNT, D. *Lecture 2: Convex Hulls in the Plane*. 2021. CMSC 754, University of Maryland. Disponível em: <<https://www.cs.umd.edu/class/fall2021/cmsc754/Lects/lect02-hulls.pdf>>. Citado 2 vezes nas páginas 43 e 44.
- MUKHOPADHYAY, P.; CHAUDHURI, B. B. A survey of hough transform. *Pattern Recognition*, Elsevier, v. 48, n. 3, p. 993–1010, 2015. Citado na página 42.
- NASCIMENTO, J. J. D. C. et al. New health of things approach to classification and detection of brain tumors using transfer learning for segmentation in imr images. In: IEEE. *2023 International Joint Conference on Neural Networks (IJCNN)*. Gold Coast, QLD, Australia: IEEE, 2023. p. 1–8. Citado 5 vezes nas páginas 10, 86, 88, 89 e 90.
- NASCIMENTO, J. J. D. C. et al. New health of things approach to classification and detection of brain tumors using transfer learning for segmentation in imr images. In: *2023 International Joint Conference on Neural Networks (IJCNN)*. Gold Coast, QLD, Australia: IEEE, 2023. p. 1–8. Citado 2 vezes nas páginas 29 e 33.
- NAYEEM, M. J. et al. Prediction of hepatitis disease using k-nearest neighbors, naive bayes, support vector machine, multi-layer perceptron and random forest. In: IEEE. *2021 international conference on information and communication technology for sustainable development (ICICT4SD)*. Pune, India: IEEE, 2021. p. 280–284. Citado na página 46.
- NIZAMANI, A. H. et al. Advance brain tumor segmentation using feature fusion methods with deep u-net model with cnn for mri data. *Journal of King Saud University-Computer and Information Sciences*, Elsevier, v. 35, n. 9, p. 101793, 2023. Citado na página 21.
- OROS, M.; PULYK, O.; JR, M. O. Comprehensive review of high-grade astrocytoma grade iii: Anaplastic astrocytoma—prevalence, pathology, treatment, and recurrence. *Proceeding of the Shevchenko Scientific Society. Medical Sciences*, v. 73, n. 1, 2024. Citado na página 18.
- ÖZKARACA, O. et al. Multiple brain tumor classification with dense cnn architecture using brain mri images. *Life*, MDPI, v. 13, n. 2, p. 349, 2023. Citado 2 vezes nas páginas 29 e 33.
- PAJILA, P. B. et al. A comprehensive survey on naive bayes algorithm: Advantages, limitations and applications. In: IEEE. *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*. Tiruchirappalli, India: IEEE, 2023. p. 1228–1234. Citado na página 47.
- PATEL, A. Benign vs malignant tumors. *JAMA oncology*, American Medical Association, v. 6, n. 9, p. 1488–1488, 2020. Citado na página 18.
- PATEL, J. et al. Segmentation, survival prediction, and uncertainty estimation of gliomas from multimodal 3d mri using selective kernel networks. In: SPRINGER. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II* 6. Cham, Switzerland: Springer, 2021. p. 228–240. Citado 3 vezes nas páginas 30, 33 e 90.

- PEI, L.; MURAT, A.; COLEN, R. Multimodal brain tumor segmentation and survival prediction using a 3d self-ensemble resnet. In: SPRINGER. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I* 6. Cham, Switzerland: Springer International Publishing, 2021. p. 367–375. Citado 3 vezes nas páginas 31, 33 e 90.
- POUYAN, A. et al. Glioblastoma multiforme: insights into pathogenesis, key signaling pathways, and therapeutic strategies. *Molecular Cancer*, Springer, v. 24, n. 1, p. 58, 2025. Citado na página 18.
- PRÆSTGAARD, J. T. Permutation and bootstrap kolmogorov-smirnov tests for the equality of two distributions. *Scandinavian Journal of Statistics*, JSTOR, New York, USA, p. 305–322, September, 1995. Citado na página 61.
- PREPARATA, F. P.; SHAMOS, M. I. *Computational geometry: an introduction*. New York, NY: Springer Science & Business Media, 2012. Citado na página 43.
- PROVENZALE, J. M.; ISON, C.; DELONG, D. Bidimensional measurements in brain tumors: assessment of interobserver variability. *American Journal of Roentgenology*, American Roentgen Ray Society, v. 193, n. 6, p. W515–W522, 2009. Citado na página 19.
- R, M. T.; V, V. K.; GULUWADI, S. Enhancing brain tumor detection in mri images through explainable ai using grad-cam with resnet 50. *BMC medical imaging*, Springer, v. 24, n. 1, p. 107, 2024. Citado na página 19.
- RAHIM, A.; YUAN, F.; BARABADY, J. An ultralytics yolov8-based approach for road detection in snowy environments in the arctic region of norway. *Computers, Materials & Continua*, v. 83, n. 3, 2025. Citado na página 64.
- RAHMAT, R. F. et al. Classification of primary and secondary brain tumor using extreme learning machine. In: IEEE. *2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA)*. Medan, Indonesia: IEEE, 2021. p. 101–105. Citado na página 18.
- RAMMURTHY, D.; MAHESH, P. Whale harris hawks optimization based deep learning classifier for brain tumor detection using mri images. *Journal of King Saud University-Computer and Information Sciences*, Elsevier, v. 34, n. 6, p. 3259–3272, 2022. Citado 2 vezes nas páginas 28 e 33.
- RANJBARZADEH, R. et al. Brain tumor segmentation of mri images: A comprehensive review on the application of artificial intelligence tools. *Computers in biology and medicine*, Elsevier, v. 152, p. 106405, 2023. Citado 2 vezes nas páginas 20 e 21.
- REBACK, J. et al. pandas-dev/pandas: Pandas 1.0. 5. *Zenodo*, 2020. Citado na página 64.
- REDMON, J. et al. You only look once: Unified, real-time object detection. In: IEEE. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE, 2016. p. 779–788. Citado na página 36.
- RICARD, D. et al. Primary brain tumours in adults. *The Lancet*, Elsevier, v. 379, n. 9830, p. 1984–1996, 2012. Citado na página 18.

- RUSH, A. M. The annotated transformer. In: *Proceedings of workshop for NLP open source software (NLP-OSS)*. Melbourne, Australia: Association for Computational Linguistics, 2018. p. 52–60. Citado na página 56.
- RYU, E. K. *Chapter 3: Convolutional Neural Networks*. 2024. *Mathematical Foundations of Deep Neural Networks*, Spring 2024, Department of Mathematical Sciences, Seoul National University. Accessed: 2025-06-18. Disponível em: <<https://ernestryu.com/courses/DNN/chapter3.pdf>>. Citado na página 53.
- SALLEMI, L.; NJEH, I.; LEHERICY, S. Towards a computer aided prognosis for brain glioblastomas tumor growth estimation. *IEEE transactions on nanobioscience*, IEEE, v. 14, n. 7, p. 727–733, 2015. Citado na página 18.
- SARKER, I. H. Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, Springer, v. 2, n. 3, p. 160, 2021. Citado 2 vezes nas páginas 45 e 46.
- SERVICES, A. W. *Meta Llama 3 models are now available in Amazon Bedrock*. 2024. AWS Blog. Disponibiliza Llama 3 na Amazon Bedrock para inferência como serviço gerenciado. Disponível em: <<https://aws.amazon.com/blogs/aws/metlas-llama-3-models-are-now-available-in-amazon-bedrock/>>. Citado na página 57.
- SHAKYA, A. K.; PILLAI, G.; CHAKRABARTY, S. Reinforcement learning algorithms: A brief survey. *Expert Systems with Applications*, Elsevier, v. 231, p. 120495, 2023. Citado na página 46.
- SHUKLA, V.; CHOUDHARY, S. Deep learning in neural networks: an overview. *deep learning in visual computing and signal processing*, Apple Academic Press, p. 29–53, 2022. Citado na página 52.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado 2 vezes nas páginas 53 e 54.
- SINGH, G. et al. Web-page interfaced real-time object detection using tensorflow. In: IEEE. *2021 3rd international conference on advances in computing, communication control and networking (ICAC3N)*. Greater Noida, India: IEEE, 2021. p. 1439–1441. Citado na página 36.
- SKOURAS, P. et al. Oligodendroglioma of the hippocampus: A case report and systematic review on therapeutic approaches of oligodendroglioma after who 2021 classification. *Pharmaceuticals*, MDPI, v. 18, n. 3, p. 349, 2025. Citado na página 18.
- SMIRNIOTOPOULOS, J. G. et al. Patterns of contrast enhancement in the brain and meninges. *Radiographics*, Radiological Society of North America, v. 27, n. 2, p. 525–551, 2007. Citado na página 18.
- SMIRNOV, N. V. Estimate of deviation between empirical distribution functions in two independent samples. *Bulletin Moscow University*, v. 2, n. 2, p. 3–16, 1939. Citado na página 62.
- SOUZA, L. F. d. F. Divisible cell-segmentation: uma nova abordagem fine-tuning adaptativa para segmentação em tomografias computadorizadas por meio de células computacionais sintéticas. 2023. Citado 5 vezes nas páginas 21, 40, 41, 42 e 68.

- SOUZA, L. F. D. F. et al. Divisible cell-segmentation: A new approach for stroke detection and segmentation in ct scans using deep learning and fine-tuning. In: IEEE. *2023 International Joint Conference on Neural Networks (IJCNN)*. Gold Coast, QLD, Australia: IEEE, 2023. p. 1–9. Citado 4 vezes nas páginas [21](#), [26](#), [30](#) e [33](#).
- SUN, L. et al. Brain tumor segmentation and survival prediction using multimodal mri scans with deep learning. *Frontiers in neuroscience*, Frontiers Media SA, v. 13, p. 810, 2019. Citado na página [20](#).
- TABATABAEI, S.; REZAEI, K.; ZHU, M. Attention transformer mechanism and fusion-based deep learning architecture for mri brain tumor classification system. *Biomedical Signal Processing and Control*, Elsevier, v. 86, p. 105119, 2023. Citado 2 vezes nas páginas [31](#) e [33](#).
- TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, BioMed Central, v. 15, n. 1, p. 1–28, 2015. Citado 4 vezes nas páginas [58](#), [59](#), [60](#) e [61](#).
- TERVEN, J.; CORDOVA-ESPARZA, D.-M.; ROMERO-GONZÁLEZ, J.-A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine learning and knowledge extraction*, MDPI, v. 5, n. 4, p. 1680–1716, 2023. Citado 3 vezes nas páginas [37](#), [39](#) e [40](#).
- TRIPATHI, S.; VERMA, A.; SHARMA, N. Automatic segmentation of brain tumour in mr images using an enhanced deep learning approach. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Taylor & Francis, Oxfordshire, United Kingdom, v. 9, n. 2, p. 121–130, September, 2021. Citado 5 vezes nas páginas [29](#), [33](#), [34](#), [86](#) e [88](#).
- TUSTISON, N. J. et al. N4itk: improved n3 bias correction. *IEEE transactions on medical imaging*, IEEE, v. 29, n. 6, p. 1310–1320, 2010. Citado 2 vezes nas páginas [28](#) e [33](#).
- VAN, M.-H.; VERMA, P.; WU, X. On large visual language models for medical imaging analysis: An empirical study. In: IEEE. *2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. Washington, DC, USA: IEEE, 2024. p. 172–176. Citado na página [22](#).
- VASWANI, A. et al. Attention is all you need. *Advances in neural information processing systems*, v. 30, 2017. Citado na página [55](#).
- WANG, S. et al. Interactive computer-aided diagnosis on medical image using large language models. *Communications Engineering*, Nature Publishing Group UK London, v. 3, n. 1, p. 133, 2024. Citado na página [20](#).
- WIJETHILAKE, N. et al. Glioma survival analysis empowered with data engineering—a survey. *Ieee Access*, IEEE, v. 9, p. 43168–43191, 2021. Citado na página [18](#).
- WU, G. et al. Sparse representation-based radiomics for the diagnosis of brain tumors. *IEEE transactions on medical imaging*, IEEE, v. 37, n. 4, p. 893–905, 2017. Citado na página [19](#).
- WU, W. et al. Glioblastoma multiforme (gbm): An overview of current therapies and mechanisms of resistance. *Pharmacological Research*, Elsevier, v. 171, p. 105780, 2021. Citado na página [18](#).

- WU, Y. et al. *Detectron2*. 2023. Disponível em: <<https://github.com/facebookresearch/detectron2>>. Acessado em: 02 de fevereiro de 2023. Citado na página 36.
- YAHYA, Z. et al. A concave hull based algorithm for object shape reconstruction. *Int. J. Inf. Technol. Comput. Sci.*, v. 9, p. 1–9, 2017. Citado na página 43.
- YAMASHITA, R. et al. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, Springer, v. 9, p. 611–629, 2018. Citado na página 52.
- YANKEELOV, T. E. et al. Quantitative imaging in cancer clinical trials. *Clinical Cancer Research*, American Association for Cancer Research, v. 22, n. 2, p. 284–290, 2016. Citado na página 19.
- YOLOv8.org. *YOLOv8 Architecture Explained: Key Features*. 2025. <<https://yolov8.org/yolov8-architecture-explained/>>. Acessado em junho de 2025. Citado 3 vezes nas páginas 37, 38 e 39.
- ZAFAR, A. et al. A comparison of pooling methods for convolutional neural networks. *Applied Sciences*, MDPI, v. 12, n. 17, p. 8643, 2022. Citado na página 52.
- ZAITSSEV, M.; MACLAREN, J.; HERBST, M. Motion artifacts in mri: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, Wiley Online Library, v. 42, n. 4, p. 887–901, 2015. Citado na página 19.
- ZHANG, A. et al. *Padding and Strides*. 2023. Chapter 6.3 in *Dive into Deep Learning*. Online; accessed 18 June 2025. Disponível em: <https://d2l.ai/chapter_convolutional-neural-networks/padding-and-strides.html>. Citado na página 53.
- ZHAO, S.; SONG, J.; ERMON, S. Learning hierarchical features from deep generative models. In: PMLR. *International Conference on Machine Learning*. Sydney, Australia: PMLR, 2017. p. 4091–4099. Citado na página 52.
- ZHAO, X. et al. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, Springer, v. 57, n. 4, p. 99, 2024. Citado na página 52.
- ZHOU, T. et al. Dense convolutional network and its application in medical image analysis. *BioMed Research International*, Wiley Online Library, v. 2022, n. 1, p. 2384830, 2022. Citado na página 55.
- ZHU, H. et al. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *Journal of Visual Communication and Image Representation*, Elsevier, v. 34, p. 12–27, 2016. Citado na página 40.