



**SALIENCY-DRIVEN  
CONVOLUTIONAL NEURAL NETWORK  
FOR BRAIN TUMOR DETECTION**

**PAULO HENRIQUE DE CASTRO OLIVEIRA**

**DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**FACULDADE DE TECNOLOGIA**

**UNIVERSIDADE DE BRASÍLIA**

**UNIVERSIDADE DE BRASÍLIA  
FACULDADE DE TECNOLOGIA  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**SALIENCY-DRIVEN  
CONVOLUTIONAL NEURAL NETWORK  
FOR BRAIN TUMOR DETECTION**

**PAULO HENRIQUE DE CASTRO OLIVEIRA**

**Orientador: PROFA. DRA. MYLÈNE C.Q. FARIAS, ENE/UNB**

**DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA**

**PUBLICAÇÃO PPGENE.DM - XXX/AAAA  
BRASÍLIA-DF, 31 DE OUTUBRO DE 2022.**

**UNIVERSIDADE DE BRASÍLIA  
FACULDADE DE TECNOLOGIA  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**SALIENCY-DRIVEN  
CONVOLUTIONAL NEURAL NETWORK  
FOR BRAIN TUMOR DETECTION**

**PAULO HENRIQUE DE CASTRO OLIVEIRA**

DISSERTAÇÃO DE MESTRADO ACADÊMICO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA ELÉTRICA.

**APROVADA POR:**

Profa. Dra. Mylène C.Q. Farias, ENE/UnB  
Orientador

Prof. Dr. Cristiano Jacques Miosso Rodrigues Mendes, FGA/UnB  
Examinador interno

Profa. Dra. Fátima Nelsizeuma Sombra de Medeiros 3, CT/UFC  
Examinador externo

Dr. Pedro Garcia Freitas  
Examinador externo

**BRASÍLIA, 31 DE OUTUBRO DE 2022.**

## **FICHA CATALOGRÁFICA**

PAULO HENRIQUE DE CASTRO OLIVEIRA

**SALIENCY-DRIVEN CONVOLUTIONAL NEURAL NETWORK FOR BRAIN TUMOR DETECTION**

**2022xv, 147p., 201x297 mm**

(ENE/FT/UnB, Mestre, Engenharia Elétrica, 2022)

Dissertação de Mestrado - Universidade de Brasília

Faculdade de Tecnologia - Departamento de Engenharia Elétrica

## **REFERÊNCIA BIBLIOGRÁFICA**

PAULO HENRIQUE DE CASTRO OLIVEIRA (2022) SALIENCY-DRIVEN CONVOLUTIONAL NEURAL NETWORK FOR BRAIN TUMOR DETECTION. Dissertação de Mestrado em Engenharia Elétrica, Publicação xxx/AAAA, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 147p.

## **CESSÃO DE DIREITOS**

AUTOR: PAULO HENRIQUE DE CASTRO OLIVEIRA

TÍTULO: SALIENCY-DRIVEN CONVOLUTIONAL NEURAL NETWORK FOR BRAIN TUMOR DETECTION.

GRAU: Mestre ANO: 2022

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor se reserva a outros direitos de publicação e nenhuma parte desta dissertação de Mestrado pode ser reproduzida sem a autorização por escrito do autor.

---

PAULO HENRIQUE DE CASTRO OLIVEIRA

QRSW 8 BL. A4, SUDOESTE

# Agradecimentos

Agradeço inicialmente a Deus por sempre estar à frente nos meus caminhos. Agradeço à minha esposa pelo apoio incondicional ao longo dessa jornada. Apoio esse dado tanto nas celebrações das pequenas conquistas quanto nas aflições superadas. Agradeço à minha família pela crença neste trabalho realizado e pelo incentivo em continuar, mesmo diante das dificuldades. Aos meus pais, agradeço pela inspiração que tive em dar o melhor possível nas minhas decisões pessoais e profissionais.

Agradeço à professora Mylène, minha orientadora, pela oportunidade de trabalho conjunto e por me guiar nesta busca por explicações científicas desde o início do projeto. Devo também agradecer pela paciência e pela inspiração ao longo dessa jornada. Agradeço ao Daniel Silva Ferreira por todo apoio ao longo da escolha do tema e pelo repasse de conhecimento na área de aprendizado de máquinas e visão computacional. Por fim, agradeço à Universidade de Brasília e a toda a equipe do Departamento de Engenharia Elétrica por tornar viável este trabalho.

O processo de diagnóstico de tumores cerebrais realizado por médicos radiologistas com imagens de ressonância magnética (RM) pode ser, em algumas ocasiões, um procedimento demorado. Assim, uma análise rápida por meio de um sistema automatizado pode ajudar a melhorar as possibilidades de tratamento e otimizar os recursos hospitalares. Este trabalho propõe um método para a classificação de tumores cerebrais por meio da pré-seleção da região de tumor. Estimou-se a região de interesse com o uso de algoritmos de atenção visual e, em seguida, as regiões com maior nível de energia foram inseridas em uma rede neural que as classificou quanto à existência de tumor cerebral. Pré-selecionando a região de interesse, em vez de usar a imagem inteira, o *framework* teve uma acurácia de classificação final de 91,68%, 92,58%, 92,69% e 93,4% com os modelos Resnet18, Resnet34, VGG16 e Alexnet, respectivamente. Uma vez que o espaço dimensional da imagem de entrada é reduzido, as redes neurais são capazes de capturar detalhes adicionais das regiões do tumor durante a etapa de treinamento. Este estudo demonstra a importância dos mapas de saliência para a identificação de regiões tumorais em imagens de ressonância magnética. Adicionalmente, investigou-se o efeito de cinco artefatos nas métricas de classificação de uma rede neural: ruído Gaussiano, *blurring*, *ringing*, contraste e *ghosting*. Essa análise contou com a geração de 20 níveis de degradação com cada um dos cinco artefatos e o registro dos resultados de acurácia com cada um desses níveis. Concluiu-se que no cenário simulado com os artefatos de ruído Gaussiano, *blurring* e *ringing* houve um decréscimo exponencial das métricas de classificação, ao passo que as alterações gradativas de contraste e *ghosting*, diferentemente, geraram um comportamento aleatório dessas métricas.

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>1</b>
1.1	MOTIVAÇÃO E OBJETIVOS .....	2
1.2	PRODUÇÃO CIENTÍFICA E TÉCNICA .....	2
1.3	ORGANIZAÇÃO DA DISSERTAÇÃO .....	2
<b>2</b>	<b>FUNDAMENTOS .....</b>	<b>4</b>
2.1	ARQUITETURAS DAS REDES NEURAIS .....	4
2.1.1	ALEXNET .....	5
2.1.2	VGG16 .....	6
2.1.3	RESNET .....	7
2.2	ATENÇÃO VISUAL E ALGORITMOS DE PREDIÇÃO DE SALIÊNCIA .....	8
2.3	MÉTRICAS DE DESEMPENHO E DE QUALIDADE .....	11
<b>3</b>	<b>TUMOR DETECTION FRAMEWORK .....</b>	<b>15</b>
3.1	BRATS 2020 DATASET .....	15
3.2	FRAMEWORK PROPOSTO .....	17
3.2.1	GERAÇÃO DE MAPAS DE SALIÊNCIA .....	17
3.2.2	CLUSTERIZAÇÃO E RANQUEAMENTO DE REGIÕES .....	18
3.2.3	EXTRAÇÃO DA REGIÃO DE INTERESSE (ROI) .....	19
3.2.4	CLASSIFICAÇÃO POR MEIO DE CNNs .....	20
<b>4</b>	<b>GERAÇÃO DE ARTEFATOS COMUNS EM IMAGENS DE RESSONÂNCIA MAGNÉTICA .....</b>	<b>22</b>
4.1	ARTEFATOS COMUNS EM RESSONÂNCIA MAGNÉTICA .....	22
4.1.1	RUÍDO GAUSSIANO .....	23
4.1.2	BLURRING .....	24
4.1.3	RINGING .....	25
4.1.4	GHOSTING .....	25
4.1.5	CONTRASTE .....	26
<b>5</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>34</b>
5.1	COMPARAÇÃO DE PERFORMANCE ENTRE OS TIPO DE IMAGEM DE RESSONÂNCIA MAGNÉTICA .....	34

5.2	CORRELAÇÃO ENTRE REGIÕES SALIENTES E REGIÕES TUMORAIS .....	35
5.3	CLASSIFICAÇÃO DE TUMORES COM O <i>Framework</i> PROPOSTO .....	38
5.4	VANTAGENS E LIMITAÇÕES DO FRAMEWORK PROPOSTO .....	40
5.5	CLASSIFICAÇÃO DAS IMAGENS DEGRADADAS .....	41
5.5.1	RUÍDO GAUSSIANO .....	41
5.5.2	BLURRING .....	41
5.5.3	RINGING .....	44
5.5.4	GHOSTING .....	44
5.5.5	CONTRASTE .....	45
5.5.6	DISCUSSÃO .....	45
<b>6</b>	<b>CONCLUSÃO .....</b>	<b>49</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>51</b>

# LISTA DE FIGURAS

2.1	Arquitetura do modelo Alexnet. Imagem gerada por meio site: <draw.io>. ....	6
2.2	Arquitetura do modelo VGG16. Imagem gerada por meio site: <draw.io>. ....	7
2.3	Skip Connection Block ilustra a saída de uma camada composta pelos dados de entrada da camada anterior. Fonte: Artigo dos autores [He et al. ].....	7
2.4	Arquitetura do modelo Resnet34. Imagem gerada por meio site: <draw.io>. ...	8
2.5	Exemplos de imagens com objetos que se destacam devido aos diferentes atributos como: (a) cor, (b) forma, (c)-(d) orientação. Esses atributos são alguns que o cérebro humano utiliza na atenção seletiva <i>bottom-up</i> . Imagem gerada por meio site: <draw.io>. ....	8
2.6	Exemplo de captura dos movimentos oculares de um médico radiologista cerebral ao ser exibida uma imagem de ressonância magnética contendo tumor. Na imagem (a) está o especialista posicionado diante de um <i>Eye-Tracker</i> ; na (b) imagem de ressonância magnética e na (c) movimentos sacádicos e de fixação do especialista gerados pelo software <i>ViewPoint DataAnalysis</i> , da empresa [vie ].....	9
2.7	Matriz de confusão de um modelo de rede neural. Imagem gerada por meio site: <draw.io>. ....	12
3.1	Amostra de um corte axial do BRATS2020 dataset. Imagem obtida em [et al 2015] .....	16
3.2	Exemplos de cortes excluídos do dataset. Imagens do dataset exibidos por meio de simulação computacional.....	16
3.3	a) Corte axial do cérebro com presença de tumor e b) máscara correspondente binarizada com a região tumoralImagens do dataset exibidos por meio de simulação computacional. ....	17
3.4	<i>Framework</i> proposto para detecção de tumor cerebral. ....	17
3.5	(a) Amostra de corte original do scan MRI, (b) máscara do tumor, e mapas de saliência correspondentes: (c) CovSal, (d) FES, (e) GBVS-FES, (f) GBVS e (g) Itti. Todos os algoritmos foram obtidos do site de referência MIT saliency [MIT ]. ....	18
3.6	a) amostra de mapa GBVS, b) amostras de mapa FES e c) combinação linear GBVS-FES.....	18

3.7	(a) <i>scan</i> MRI, (b) mapa FES-GBVS, (c) mapa de saliência clusterizado com $k=13$ , (d) os clusteres pseudo-coloridos, (e) os seis clusteres com maior brilho mesclados, (f) máscara da ROI e g) ROI mostrada no mapa da saliência. ....	20
3.8	(a) <i>scan</i> MRI, (b) quadrante da ROI no mapa de saliência, (c) quadrante da ROI no <i>scan</i> MRI e (d) o recorte da ROI final. ....	20
4.1	Amostras de uma mesma imagem de teste original com os 20 níveis de ruído Gaussiano, em ordem crescente da esquerda para a direita, de cima para baixo.	29
4.2	Amostras de uma mesma imagem de teste original com os 20 níveis de <i>blurring</i> , em ordem crescente da esquerda para a direita, de cima para baixo. ....	30
4.3	Amostras de uma mesma imagem de teste original com os 20 níveis de <i>ringing</i> , em ordem crescente da esquerda para a direita, de cima para baixo. ....	31
4.4	Amostras de uma mesma imagem de teste original com os 20 níveis de <i>ghosting</i> , em ordem crescente da esquerda para a direita, de cima para baixo. ....	32
4.5	Amostras de uma mesma imagem de teste original com os 20 níveis de contraste, em ordem crescente da esquerda para a direita, de cima para baixo. ....	33
5.1	Porcentagem de imagens com tumor detectado versus número de <i>clusters</i> com maior intensidade de brilho selecionados. Valor de $k = 13$ . ....	36
5.2	Porcentagem de imagens com tumor detectado versus número de clusteres com maior intensidade de brilho selecionados. Valor de $k = 9$ . ....	36
5.3	Porcentagem de imagens com tumor detectado versus número de clusteres com maior intensidade de brilho selecionados. Valor de $k = 11$ . ....	37
5.4	Porcentagem de imagens com tumor detectado versus número de clusteres com maior intensidade de brilho selecionados. Valor de $k = 15$ . ....	37
5.5	Porcentagem de imagens com tumor detectado versus número de clusteres com maior intensidade de brilho selecionados. Valor de $k = 17$ . ....	38
5.6	<i>Scan</i> MRI e sua máscara de ativação (CAM) correspondente após a classificação realizada pela Resnet18 sem o uso do <i>framework</i> proposto (usando <i>scan</i> MRI completo). ....	39
5.7	<i>Scan</i> MRI e sua máscara de ativação (CAM) correspondente após a classificação realizada pela Resnet18 com o uso do <i>framework</i> proposto (usando o <i>scan</i> da ROI). ....	40
5.8	Amostra de uma imagem com tumor cuja área ocupa menos de 0.3% da área cerebral. ....	40
5.9	Resultados de classificação (acurácia, precisão, recall e F1 <i>score</i> ) de imagens com ruído Gaussiano com relação ao valor médio da MSE de cada nível de degradação. ....	42
5.10	Resultados de classificação (acurácia, precisão, recall e F1 <i>score</i> ) de imagens com <i>blurring</i> com relação ao valor médio da MSE de cada nível de degradação.	43
5.11	Resultados de classificação (acurácia, precisão, recall e F1 <i>score</i> ) de imagens com <i>ringing</i> com relação ao valor médio da MSE de cada nível de degradação.	44

- 5.12 Resultados de classificação (acurácia, precisão, recall e *F1 score*) de imagens com ghosting com relação ao valor médio da MSE de cada nível de degradação. 46
- 5.13 Resultados de classificação (acurácia, precisão, recall e *F1 score*) de imagens com contraste com relação ao valor médio da MSE de cada nível de degradação. 47

# LISTA DE TABELAS

4.1	Valores de variância utilizados para os diferentes níveis de ruído Gaussiano aplicado. ....	24
4.2	Valores de desvio padrão utilizados para gerar os diferentes níveis de <i>blurring</i> nas imagens de ressonância magnética.....	25
4.3	Valores dos raios de discos em imagens binárias utilizadas para gerar o efeito de <i>ringing</i> . ....	26
4.4	Valores dos parâmetros de intensidade que modelaram os 20 níveis de <i>ghosting</i> . ....	27
4.5	Valores dos parâmetros $\alpha$ e $\beta$ que modelaram os 20 níveis de contraste.....	28
5.1	Comparação de resultados de classificação entre diferentes tipos de imagens de ressonância: <i>Flair</i> , T1, T2 e T1ce. São exibidas as médias e desvios padrões de cada uma das métricas: acurácia, precisão, recall e <i>F1 score</i> . ....	35
5.2	Acurácia, precisão, recall e <i>F1 score</i> que são os resultados de classificação de para Resnet18, Resnet34, VGG16 e Alexnet sem o <i>framework</i> proposto, ou seja, usando a imagem de ressonância magnética completa. ....	38
5.3	Precisão, precisão, recall e <i>F1 score</i> que são os resultados de classificação de para Resnet18, Resnet34, VGG16 e Alexnet com o <i>framework</i> proposto, ou seja, usando a região de interesse extraída da imagem de ressonância magnética. ....	39
5.4	Resultados de classificação com relação ao valor médio da MSE de cada nível de degradação de ruído Gaussiano. ....	42
5.5	Resultados de classificação com relação ao valor médio da MSE de cada nível de degradação com <i>blurring</i> . ....	43
5.6	Resultados de classificação com relação ao valor médio da MSE de cada nível de degradação com <i>ringing</i> . ....	45
5.7	Resultados de classificação com relação ao valor médio da MSE de cada nível de degradação com <i>ghosting</i> .....	46
5.8	Resultados de classificação com relação ao valor médio da MSE de cada nível de degradação com Contraste. ....	47

# **LISTA DE CÓDIGOS FONTE**

# Capítulo 1

## Introdução

Câncer é o crescimento de células anormais no corpo humano. De acordo com o Instituto Nacional do Câncer [INCA ], entre 1,4% e 1,8% dos tumores malignos no mundo ocorrem no sistema nervoso central (SNC). Dentre esses casos, 88% dos tumores no SNC são diagnosticados no cérebro humano. Apesar de serem menos de 2% de todos os casos de câncer, os tumores cerebrais possuem altas taxas de mortalidade. Cânceres primários no cérebro correspondem à terceira maior causa de mortes entre homens entre 15 e 54 anos e à quarta maior entre mulheres entre 15 e 34 anos de idade [Kesari and Stiles 2006].

Há múltiplas razões para a ocorrência de alterações genéticas nas células humanas que aumentam o risco de surgimento de tumores, como por exemplo o trabalho em ambientes com campos magnéticos de baixa frequência, a exposição a raios X e gama, o contato cotidiano com arsênio, chumbo, mercúrio, óleo mineral etc. [INCA ]. Embora haja muitos estudos relacionados sobre a prevenção do câncer, ainda não há um padrão ou protocolo específico, aceito pela comunidade médica, que previna o surgimento de tumores no sistema nervoso central. Há, entretanto, um consenso na literatura no sentido de que a detecção precoce possibilita um maior número de tratamentos e uma maior chance de sobrevivência. No estudo de [Thomson and Forman 2009], foram analisados os efeitos do diagnóstico tardio nas taxas de sobrevivência de pacientes com câncer no Reino Unido. A análise verificou 26 tipos de câncer com uma recomendação de exames que permitiriam o diagnóstico precoce como uma das metas de política de saúde pública.

O exame dos tecidos cerebrais e a análise de imagens médicas são as principais formas de se diagnosticar a existência de tumores cerebrais [Bernstein and Berger 2014]. O diagnóstico médico se dá principalmente por meio de imagens de ressonância magnética (MRI) e de tomografia computadorizada (TC). MRIs são eficazes devido ao bom nível de contraste das imagens de tecidos finos e à sua alta resolução espacial, o que ajuda na detecção de diferentes tipos de doenças. Além disso, o escaneamento em máquinas de ressonância magnética é eficiente para a captura de uma imagem cerebral, uma vez que o procedimento não é invasivo e a radiação é não-ionizante [Pereira et al. 2016]. Uma vez diagnosticado o tumor, o tratamento é realizado por meio da sua remoção, seguido de sessões de radioter-

pia e/ou quimioterapia. Nessa etapa, há acompanhamento do neurocirurgião, oncologista e radioterapeuta, todos eles profissionais envolvidos no tratamento do paciente.

## 1.1 Motivação e Objetivos

A motivação deste estudo é a constatação da necessidade de um diagnóstico precoce de tumores cerebrais, que pode ser realizado utilizando-se um software computacional (*Computer Aided Diagnosis*). Uma forma eficiente de diagnóstico permitirá que médicos especialistas se dediquem a atividades mais complexas e que pacientes contem com mais possibilidades de tratamento.

O objetivo desta pesquisa é criar um sistema que classifica imagens 2D (bidimensionais) baseado em exame de ressonância magnética com e sem tumor. O trabalho utiliza redes convolucionais e mapas de saliência para seleção de Regiões de Interesse (ROIs) dos exames MRIs. Além disso, buscou-se analisar o impacto que alguns artefatos aplicados em imagens de ressonância magnética podem gerar em classificadores automáticos baseados em redes convolucionais.

## 1.2 Produção científica e técnica

O desenvolvimento desta dissertação resultou na publicação do artigo científico intitulado “Using a Saliency-Driven Convolutional Neural Network Framework for Brain Tumor Detection”, na 6th *International Conference on Medical and Health Informatics*, [de C. O. et al. 2022].

Um segundo artigo científico, “The Influence of Magnetic Resonance Imaging Artifacts on CNN-based Brain Cancer Detection Algorithms“, foi também produzido pelos autores [Farias et al. 2022] e se encontra sob análise de aceitação. Nesse estudo, foram gerados códigos computacionais e resultados que analisam a influência de diversas modalidades de ruídos em sistemas automáticos de diagnóstico.

## 1.3 Organização da dissertação

Para melhor compreensão do estudo realizado, esta dissertação foi organizada da seguinte forma: o Capítulo 2 apresenta os fundamentos teóricos da área de aprendizado de máquina, como conceito e aplicações. Além disso, uma descrição sucinta dos algoritmos de predição de saliência utilizados é apresentada, bem como das arquiteturas de redes neurais testadas. O Capítulo 3 apresenta a descrição do *dataset* utilizado e detalha cada etapa do *framework* proposto, que inclui a geração dos mapas de saliência, a clusterização dos mapas, a extra-

ção das ROIs e a classificação das imagens com utilização de redes neurais convolucionais (CNNs). O Capítulo 4 descreve o segundo estudo realizado, representado pela influência que uma imagem deteriorada por artefatos tem em sistemas de classificação automática. Foram simulados cinco diferentes tipos de artefatos, sendo apresentados os respectivos resultados de classificação. O Capítulo 5 apresenta os resultados comparativos entre a classificação que utiliza a imagem completa e a classificação proposta neste estudo, que utiliza mapas de saliência para extração de interesse. São mostradas também, quantitativamente e qualitativamente, as limitações do modelo proposto. O Capítulo 6 apresenta as conclusões e as perspectivas de estudos futuros.

# Capítulo 2

## Fundamentos

Esse capítulo descreve o conceito da área de aprendizado de máquinas e suas aplicações. A evolução desse campo de pesquisa levou ao surgimento das primeiras arquiteturas de redes neurais. De maneira semelhante, foi descrita a área de predição de saliência que decorreu de estudos da psicologia, os quais têm investigado a atenção visual humana. Alguns desses estudos levaram à elaboração de modelos computacionais de predição de saliência, dentre os quais, cinco serão apresentados neste capítulo: ITTI, GBVS, FES, Covsal e FES-GBVS. Por fim, são apresentadas as métricas utilizadas para a comparação de modelos de rede neural quanto ao desempenho de classificação. Além disso, são descritas métricas de qualidade de imagens como PSNR e MSE que quantificam a degradação dos *scans* de ressonância magnética.

### 2.1 Arquiteturas das Redes Neurais

A área de aprendizado de máquina surgiu em um contexto em que existiam problemas que não poderiam ser resolvidos imediatamente por um algoritmo tradicional, ou uma função, que recebe dados de entrada e tem um dado de saída específico [Alpaydm 2009]. No entanto, para esse mesmo problema existiam muitos dados disponíveis que poderiam ser utilizados de alguma forma para solucioná-lo. Assim, a partir de uma base de dados de entrada busca-se um algoritmo ou sistema específico que aprenda e realize uma tarefa de classificação de maneira automática, por exemplo.

No sistema de informações de uma farmácia, exemplificando, há milhares de registros de compras de usuários como: nome do cliente, produtos adquiridos, dinheiro gasto etc. O gerente poderia utilizar essas informações para estimar quem são os clientes específicos para um determinado produto. Assim, não se sabe o exato processo para a geração de informações comportamentais de um usuário; no entanto, sabe-se que os dados não são totalmente aleatórios. Há certos padrões nos dados de compras de cada cliente.

Com o desenvolvimento e o barateamento de recursos computacionais, os volumes de

dados em sistemas de armazenamento têm crescido exponencialmente, e muitos deles estão disponíveis para todas as pessoas por meio da internet. Esses dados serão úteis quando, a partir deles, se torne possível extrair informação e conhecimento. As áreas de aplicações de aprendizado de máquina são várias como por exemplo: previsão de consumo de energia elétrica em cidades, reconhecimento facial, análise de risco para fornecimento de créditos bancários para clientes, criação de sistemas *antispam*, etc.

Nesta seção será dado foco aos aspectos de aprendizado de máquina importantes para a pesquisa realizada. Mais especificamente, serão detalhadas as arquiteturas de redes neurais convolucionais CNN - do inglês *convolutional neural networks* utilizadas neste trabalho: a Alexnet, a VGG16 e a Resnet.

### 2.1.1 Alexnet

Até 2012, a maior parte dos *datasets* públicos era pequena e as redes neurais testadas eram menos profundas e especializadas. Entre os *datasets* existentes, pode-se citar NORB criado por [LeCun et al. ] e CALTECH-101/256, criado pelos autores [Fei-Fei et al. 2007] e [Griffin et al. 2022], respectivamente. Esses bancos de dados contêm dezenas de milhares de imagens. Com o passar do tempo, os recursos computacionais foram se tornando mais baratos e surgiram novos *datasets* públicos anotados, com milhões de amostras, como, por exemplo, o *LabelMe*[Russell et al. 2007] e o *ImageNet* cujos autores são [Deng et al. 2009]. *ImageNet* é um *dataset* com 15 milhões de imagens, com *labels* em alta resolução e com 22.000 categorias de imagens.

Esse contexto facilitou o surgimento de redes neurais mais profundas, pois, para a tarefa de reconhecimento, são necessários *datasets* maiores. Assim, surgiu o modelo de rede neural Alexnet, proposto no estudo desenvolvido por [Krizhevsky et al. 2017]. A rede possui cinco camadas convolucionais, seguidas de camadas de *max-pooling* e três camadas densas (*fully connected*). A última camada com softmax com 1.000 saídas, devido às 1.000 categorias dos *datasets* do ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Sua arquitetura se encontra detalhada na Figura 2.1.

Entre as principais contribuições desse estudo, pode-se citar:

- o treinamento de modelos de redes profundas com *datasets* *ilsvrc-2010* e *ilsvrc-2012*, criados por [Russakovsky et al. 2015], com obtenção de um dos melhores resultados;
- a implementação com o uso de otimização de múltiplas GPUs
- as técnicas propostas para evitar *overfitting*.

Mais especificamente, [Krizhevsky et al. 2017] propôs o uso da função de não-linearidade RELU, em vez da tradicional *sigmoid*. A função RELU é dada pela seguinte equação:

$$f(x) = \max(0, x) \quad (2.1)$$

Essa função permitiu um treino mais acelerado em 25%. Além disso, foi utilizada a função de normalização *Local Response Normalization*. Essa função se baseia nos mecanismos utilizados pela retina do olho humano, tendo efeitos no brilho da imagem. A normalização utilizada juntamente com mecanismos de *Dropout*, que consiste na desativação de um percentual de neurônios durante o treino, contribui para evitar o *overfitting* do modelo.

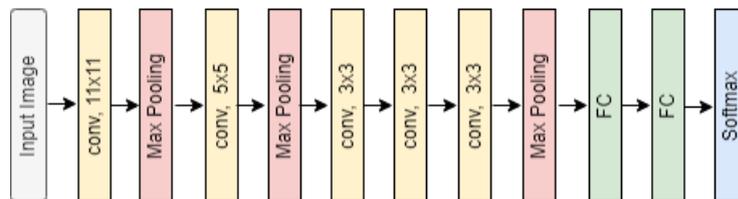


Figura 2.1: Arquitetura do modelo Alexnet. Imagem gerada por meio site: <draw.io>.

## 2.1.2 VGG16

A VGG16 foi originalmente proposta em artigo de [Simonyan and Zisserman 2014] em 2014. Nesse artigo, os autores investigaram a influência do aumento da quantidade de camadas para obtenção de melhores resultados. Essa abordagem foi possível devido aos filtros convolucionais pequenos (3 x 3), característica que tornou a função de decisão mais discriminativa e diminuiu o número de parâmetros da rede em aproximadamente 80%. O modelo possui 16 camadas com uma arquitetura uniforme, ou seja, há um padrão na sequência de camadas convolucionais e *pooling*, diferentemente da *Alexnet*, por exemplo. A Figura 2.2 ilustra com mais detalhes esse modelo.

O estudo realizado por [Simonyan and Zisserman 2014] comparou diferentes tamanhos de kernel para os filtros convolucionais: 1 x 1, 3 x 3 e 5 x 5. Os filtros 3 x 3 foram o que tiveram melhores resultados em uma rede com cinco camadas. Foi concluído também que o método de inicialização aleatória dos pesos tem relevância nos resultados de classificação, com piora das métricas. Ademais, o uso de imagens com tamanhos diferentes alcançou melhores resultados em relação ao treino de imagens com tamanhos fixos, ainda que no conjunto de testes as imagens tivessem tamanhos iguais. Mais especificamente, o uso *scale jittering* no treinamento ajuda a capturar características das imagens. Na avaliação em termos de profundidade, com até 19 camadas, foi alcançado o melhor desempenho no dataset ImageNet. Acima de 19 camadas, o modelo satura, ou seja, não há ganho significativo de desempenho.

A VGG16 usou a função RELU, que acentua a não-linearização da rede e evita o *vanishing gradient problem*, que diminui a velocidade de aprendizado. Ela foi testada no dataset do desafio ILSVRC com imagens categorizadas em 1.000 classes e divididas da seguinte forma: 1.3 milhões de imagens são utilizadas no treino, 50.000 imagens são utilizadas para validação e 100.000 imagens são utilizadas para teste. O modelo figurou dentre os que tiveram melhores resultados na competição ILSVRC-2012 e ILSVRC-2013, [Russakovsky et al. 2015].

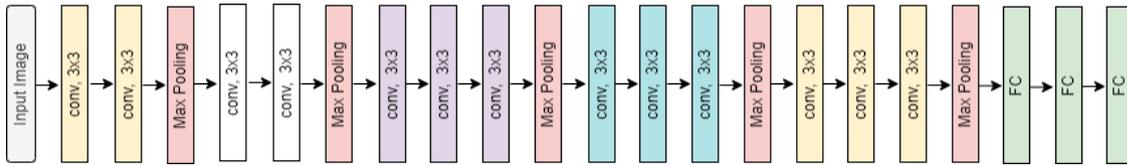


Figura 2.2: Arquitetura do modelo VGG16. Imagem gerada por meio site: <draw.io>.

### 2.1.3 Resnet

Redes muito profundas muitas vezes são muito difíceis de serem treinadas devido ao problema de *vanishing* e à explosão de gradientes. Com o aumento do número de camadas, o erro de treinamento inicialmente começa a diminuir, mas, com o tempo de treinamento, ele volta a aumentar.

[He et al. ] propuseram o uso de *Skip Connections*. Essas estruturas permitem usar a ativação de uma camada para alimentar a camada seguinte formando assim um bloco residual, como demonstrado na Figura 2.3. Dessa forma, as camadas são reformuladas durante o aprendizado, sendo afetadas em função dos parâmetros de entrada da rede, o que permitiu o treinamento de redes mais profundas e a eliminação do problema de *vanishing*. Dentre as principais vantagens dessa estratégia pode-se citar:

- As redes profundas residuais são fáceis de serem implementadas;
- O modelo pode ter ganhos de acurácia com o aumento de profundidade, sem gerar degradação no treinamento.

Dentre os experimentos do estudo, os autores mostraram que a *Resnet34*, ilustradas na Figura 2.4, é 2,8% melhor (em termos de acurácia) que a *Resnet18*. Com um erro de treinamento também menor, foi mostrado que mesmo sendo mais profunda, a *Resnet* teve uma convergência mais rápida e continua sendo um modelo com grande capacidade de generalização. A abordagem de resíduos tem efeito apenas com redes com mais de três camadas, do contrário, não são observadas vantagens.

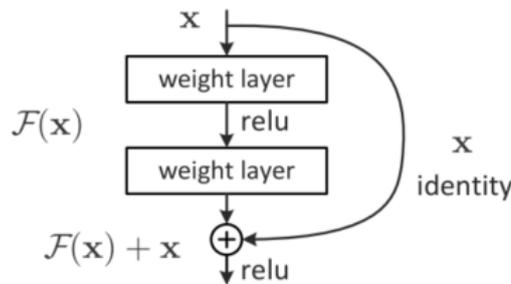


Figura 2.3: Skip Connection Block ilustra a saída de uma camada composta pelos dados de entrada da camada anterior. Fonte: Artigo dos autores [He et al. ]

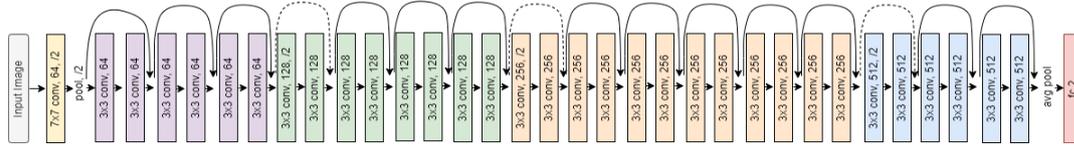


Figura 2.4: Arquitetura do modelo Resnet34. Imagem gerada por meio site: <draw.io>.

## 2.2 Atenção Visual e Algoritmos de Predição de Saliência

Atenção visual seletiva é a capacidade cerebral de processar informações importantes de uma cena. Ela ajuda o ser humano a lidar com o grande volume de informações que chegam pelo olhar humano. No livro publicado pelos autores [Liming Zhang 2013], a literatura classifica a atenção visual em dois tipos: *bottom-up* e *top-down*.

Os mecanismos *bottom-up* se baseiam nas características intrínsecas das imagens ou vídeos como brilho, contraste, cor, forma, movimento etc. Dessa forma, o cérebro humano identifica as regiões ou objetos mais relevantes. Por exemplo, os pontos de luz em uma noite escura, o movimento repentino de um objeto sobre a mesa ou imagens de flores vermelhas em uma grama verde são objetos que chamariam a atenção humana devido aos seus atributos. O mecanismo *bottom-up* mostra regiões candidatas que estão onde se encontram os objetos de interesse. De acordo com o estudo dos autores [Itti and Koch 2001], trata-se de uma atenção dirigida por estímulo visual que acontece com alta velocidade nas áreas iniciais do cérebro que compreendem: a retina, núcleo geniculado lateral e área visual V1 do córtex, [Liming Zhang 2013]. Na Figura 2.5, têm-se exemplo desses estímulos.

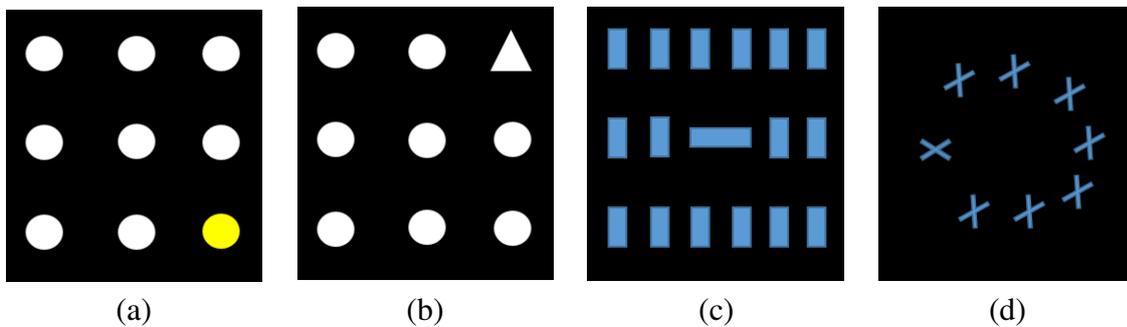


Figura 2.5: Exemplos de imagens com objetos que se destacam devido aos diferentes atributos como: (a) cor, (b) forma, (c)-(d) orientação. Esses atributos são alguns que o cérebro humano utiliza na atenção seletiva *bottom-up*. Imagem gerada por meio site: <draw.io>.

Os mecanismos de atenção *top-down* se baseiam em uma tarefa direcionada ao observador, na sua intenção, no seu conhecimento prévio sobre o estímulo e por fim no contexto de observação, como descrito por [Liming Zhang 2013]. Por exemplo, um engenheiro civil que analise a imagem da construção de um prédio, observará detalhes relacionados a estrutura de sustentação, lajes etc. Esse viés é ocasionado devido ao seu conhecimento prévio de engenharia que o faz focar em regiões diferentes de um observador leigo. De maneira genérica, a atenção *top-down* é aquela viesada pelo contexto, pelo conhecimento prévio, pela emoção

e pela atividade do observador. Na Figura 2.6, é mostrada a captura do olhar de um médico radiologista cerebral ao ser instruído para identificar a existência de tumor. Nessa imagem, há exemplo de mecanismos de atenção seletiva do tipo *top-down*.

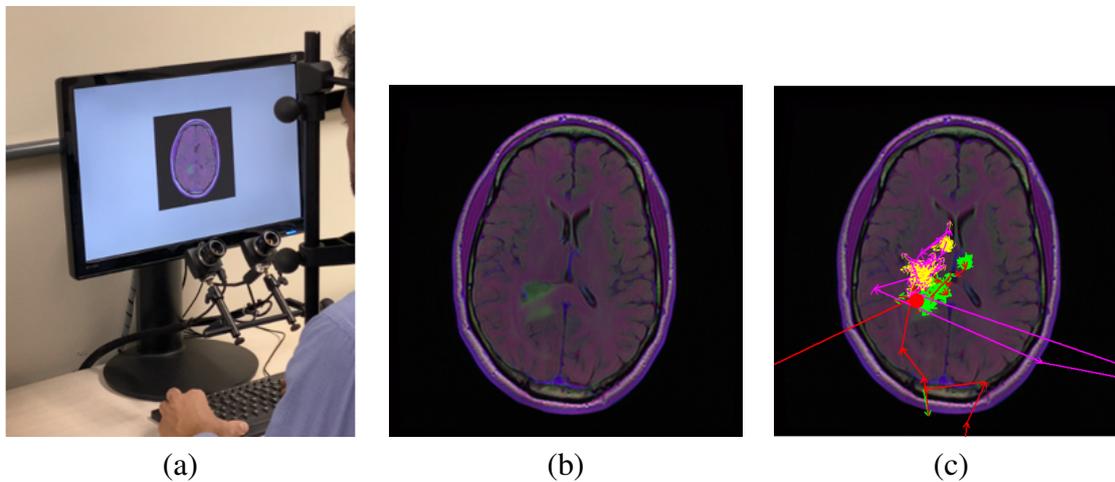


Figura 2.6: Exemplo de captura dos movimentos oculares de um médico radiologista cerebral ao ser exibida uma imagem de ressonância magnética contendo tumor. Na imagem (a) está o especialista posicionado diante de um *Eye-Tracker*; na (b) imagem de ressonância magnética e na (c) movimentos sacádicos e de fixação do especialista gerados pelo software *ViewPoint DataAnalysis*, da empresa [vie ].

O estudo da atenção visual se iniciou na área de psicologia comportamental com os estudos publicados por [Treisman and Gelade 1980] em que se denominou o termo *Fixation Prediction*. Desde então, cientistas trabalham no desenvolvimento de modelos computacionais que simulam a atenção visual para aplicações na área de engenharia, mesmo sabendo que nem todas as etapas do processo biológico de atenção visual estão compreendidas pela ciência.

Os autores [Ferreira et al. 2019b] investigaram 13 algoritmos de predição de saliência diferentes no estado da arte, incluindo modelos de aprendizagem, para identificar regiões clinicamente relevantes em imagens de lâminas de Papanicolau. Eles descobriram que a extração das regiões de interesse (ROI), por meio de métodos de previsão de saliência, permitiu classificar as áreas clínicas mais importantes em imagens de células cervicais, mesmo quando utilizados algoritmos *bottom-up*. Em outro trabalho, os autores [Ferreira et al. 2019a] propuseram um *framework* baseado em mapas de saliência para localizar e classificar células cervicais relevantes em três níveis: normal, baixo risco e alto risco de doenças. Os autores concluíram que os mapas de saliência aceleram a identificação de células relevantes dentro das quatro regiões mais notáveis, atingindo cerca de 90% de acurácia, reduzindo a influência de regiões da imagem sem interesse clínico.

[Nwe et al. 2020] combinaram rede neurais e predição de saliência em imagens de ressonância magnética. Eles usaram uma rede anisotrópica para prever os mapas de saliência de imagens de ressonância e uma segunda rede neural para segmentar os tumores cerebrais.

Já [Polec et al. 2017] propuseram um *framework* para auxiliar na detecção da esquizofrenia, que utiliza mapas subjetivos de saliência, ou seja, mapas capturados por um *Eye-Tracker*. Os padrões de movimento dos olhos capturados do paciente é então analisado estatisticamente para procurar quaisquer sinais ou características de esquizofrenia.

O estudo realizado por [Elazary and Itti 2008] avaliou a frequência com que os objetos considerados mais interessantes coincidem com as respectivas anotações realizadas por voluntários da pesquisa. Os autores utilizaram um *dataset* público, *LabelME* [Russell et al. 2007], que compreendia tanto imagens com cenas naturais estáticas quanto imagens dinâmicas (conjunto de imagens sequenciais). Nas imagens, qualquer usuário que as olhasse poderia marcar os objetos que julgasse genericamente mais interessantes no cenário. O objetivo desse estudo foi obter as anotações de 74.454 objetos que foram anotados por aproximadamente 24.863 pessoas. Com base em estudos prévios realizados pelos autores [Franconeri et al. 2005] e [Jonides and Yantis 1988] em que sugerem que características primitivas, como alterações de brilho na retina, podem guiar a atenção em direção a objetos interessantes, os autores [Elazary and Itti 2008] propuseram o algoritmo referenciado nesse estudo como Itti. Nele ficou demonstrado que as influências de baixo nível de uma imagem permitem inferir qual objeto pode ser julgado como interessante pelos observadores. Nesse estudo, uma amostra da imagem é inicialmente decomposta em mapas de características em diferentes escalas. Os domínios analisados para esses mapas são: intensidade, oponência das cores e orientações. Um total de 42 mapas de características são gerados e combinados em um mapa de saliência final. Por fim, o mapa final é normalizado por meio de uma matriz de polarização que pondera os *pixels* com um viés de centro. Entre os resultados encontrados, em 76% dos casos mais uma das três regiões salientes corresponderam à região de anotação do objeto.

O *Graph Based Visual Maps* (GBVS), proposto por [Harel et al. 2006], é um algoritmo *bottom-up* cujos mapas de saliência são gerados em duas etapas. Na primeira etapa, computam-se os mapas de ativação de cada um dos canais da imagem. Na segunda etapa, esses mapas são normalizados com ênfase nos pontos de maior visibilidade. Os mapas de ativação correspondem às regiões que de alguma maneira se distinguem das regiões vizinhas. Por meio de uma cadeia markoviana, foi criada uma métrica que computa a dissimilaridade entre uma região e as regiões vizinhas. O modelo superou o Itti & Kock, [Elazary and Itti 2008], quando testado em um *dataset* composto por imagens naturais e suas respectivas anotações (fixações oculares de observadores). Nesse cenário, obteve-se 98% de curva ROC, ao passo que o Itti obteve 84%. Esse modelo utilizou poder computacional de algoritmos de grafo para o alcance dos resultados. O conjunto de testes contou com 749 modificações em 108 imagens originais realizadas. Para a geração das anotações (*Ground Truth*) dessas imagens, três pessoas olharam cada imagem e os pontos de fixação foram capturados por meio de um *Eye-Tracker*. Essas fixações foram convoluídas com um kernel circular com decaimento constante correspondente ao decaimento da densidade do cone da retina. O estudo concluiu também a existência de um viés centralizado que ocorre devido à tendência das pessoas em

centralizar objetos em fotos. Além disso, propôs uma abordagem que utiliza múltiplas resoluções de uma mesma imagem, o que melhorou os resultados com a contrapartida de maior custo computacional.

Na pesquisa realizada por [Jampani et al. 2012], o GBVS foi utilizado para diagnosticar *Hard Exudates* e *Pneumocosis*.

O CovSal, descrito em estudo publicado por [Erdem and Erdem 2013] utiliza descritores estatísticos de segunda ordem que se baseiam no cálculo da covariância do vetor de características. Em cada canal da imagem, computa-se o descritor de segunda ordem que permite a integração com as *features* de outros canais. Esse foi considerado um bom modelo de atenção para imagens naturais. O algoritmo primeiramente calcula a saliência local, que é o quanto um pixel, em uma janela local, se distingue do que está ao redor. Esse descritor codifica estruturas locais e provê robustez e alto poder discriminativo entre regiões.

O *Fast and Efficient Saliency Detection* (FES), modelo desenvolvido por [Tavakoli et al. 2011], utiliza amostras esparsas da imagem original e estima a densidade do kernel. Trata-se de um algoritmo de fixação que utiliza o viés de centro, ou seja, assume que há uma probabilidade maior do ser humano focar nas regiões centrais de uma imagem baseando-se em evidências científicas. À época de sua criação, foram obtidos melhores resultados (curva ROC) em relação ao que havia até então como estado da arte. O modelo de dados do FES foi derivado a partir da fórmula de Bayes de probabilidade condicional e computa o pixel de saliência em uma matriz. O mapa de saliência final gerado também utiliza a abordagem de multiescalas, ou seja, computa-se o mapa de saliência em diferentes escalas e, em seguida, calcula-se um valor médio no mapa de saliência final gerado.

## 2.3 Métricas de Desempenho e de Qualidade

Neste estudo, o modelo de rede neural utilizada realizou uma classificação binária de imagens: com tumor e sem tumor. Se o resultado do teste é positivo, então a imagem contém tumor. Caso contrário, a imagem não contém tumor. Nesse contexto, uma das quatro situações irá ocorrer:

1. A imagem de um paciente que realmente tinha tumor e foi classificada como positiva. Isto é chamado de *True Positive (TP)*.
2. A imagem de um paciente que não continha tumor, mas que foi classificada como positiva. Isto é chamado de *False Positive (FP)*.
3. A imagem de um paciente que realmente tinha tumor e foi classificada como negativa. Isto é chamado de *False Negative (FN)*.

4. A imagem de um paciente que não continha tumor e que foi classificada como negativo. Isto é chamado de *True Negative (TN)*.

Os resultados de uma classificação podem ser organizados em uma matriz de confusão conforme mostrado na Figura 2.7.

		Predicted Label	
		Negative	Positive
True Label	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

Figura 2.7: Matriz de confusão de um modelo de rede neural. Imagem gerada por meio site: <draw.io>.

O melhor cenário para um classificador automático é que ocorram apenas casos do tipo *True Positive* e *True Negative*, ou seja, que o modelo classifique corretamente os casos em que há tumor cerebral e os casos em que não há tumor cerebral nas imagens de ressonância magnética. No entanto, para a maioria das aplicações é impossível obter erros absolutos igual a zero, como afirma o autor [Goodfellow et al. 2017].

E o pior dos cenários é o que ocorram mais casos do tipo *False Positive* e *False Negatives*. Quatro métricas foram aplicadas para avaliação dos modelos utilizados: acurácia, precisão, recall e *F1 score*.

A acurácia mede o percentual de casos classificados corretamente pelo modelo em relação ao total de casos, [Alpaydın 2009]. A acurácia é dada pela Equação 2.2.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad (2.2)$$

em que  $TP$  são os resultados true positives,  $TN$  true negatives,  $FP$  false positives e  $FN$  false negatives.

O *recall*, também referenciado com sensibilidade, é dado pela Equação 2.3 e equivale ao percentual de acertos nos casos classificados como positivos [Alpaydın 2009]. Quanto mais próximo de 1 é o recall, melhor o classificador. De maneira oposta, quanto mais próximo de zero, pior o classificador.

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (2.3)$$

A precisão é dada pela Equação 2.4 e equivale ao percentual de acertos nos casos classificados como positivos, [Alpaydın 2009]. Quanto mais próximo de 1 é a precisão, melhor o

classificador. De maneira oposta, quanto mais próximo de zero, pior o classificador.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.4)$$

Em aplicações biomédicas, o *recall* é uma métrica relevante, porque penaliza cenários em que se verifica falso-negativo, ou seja, casos em que o paciente apresenta uma doença, porém o diagnóstico é negativo. Uma métrica importante para comparar os modelos é o *F1 score*. Essa é a métrica adotada como referência nesse estudo para fins de comparação uma vez que leva em conta a precisão e o recall. *F1 score* é dado pela seguinte Equação 2.5

$$\text{F1 score} = \frac{2 \times \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (2.5)$$

Em muitos modelos de aprendizado de máquina, para obtenção de maior robustez estatística, utiliza-se o *k-fold*. Com o uso dessa técnica, um modelo específico realizará o processo de classificação do conjunto de testes *k* vezes. Nesse cenário, para cada uma das métricas descritas acima, calcula-se o valor médio e o desvio padrão das *k* simulações realizadas. Dessa forma, a métrica de acurácia mostrada na Equação 2.2, por exemplo, é dada pelo valor médio das acurácias de cada uma das simulações, conforme descrito na Equação 2.6.

$$\text{Accuracy} = \frac{\text{Accuracy } 1^{st} \text{ fold} + \dots + \text{Accuracy } k^{th} \text{ fold}}{k}. \quad (2.6)$$

A mesma analogia é aplicada no cálculo da precisão, recall, e F1 score.

O *k-fold* foi utilizado nessa pesquisa, e maiores detalhes serão apresentados na seção 3.2.4.

A segunda parte desse estudo analisou o impacto da degradação de imagens de ressonância magnética nas métricas de performance de uma rede neural. Essa degradação foi aferida por meio de duas métricas de qualidade, que compararam as imagens deterioradas com as imagens originais: MSE (*Mean Square Error*) e PSNR (*Peak Signal-to-Noise Ratio*). Ambas as métricas medem o quão próxima uma imagem degradada está da imagem original.

O MSE computa o erro acumulativo médio quadrático entre duas imagens. Quanto mais próximo de zero é o MSE, menos degradada está a imagem. Segundo [Gonzalez and Woods 2018], o MSE é calculado pela seguinte equação:

$$MSE = \sum_{m=1}^M \sum_{n=1}^N \frac{[f(x, y) - g(x, y)]^2}{M \times N}, \quad (2.7)$$

onde *M* representa o número de linhas e *N* o número de colunas da imagem.

O PSNR está relacionado ao MSE e, de acordo com [Gonzalez and Woods 2018], é

calculado utilizando a seguinte equação:

$$PSNR = 10 \times \log\left(\frac{L}{MSE}\right), \quad (2.8)$$

em que  $L$  corresponde ao número total de níveis de uma imagem. Quanto maior o PSNR, melhor é a qualidade da imagem deteriorada em relação à original. Como as imagens de escala cinza de ressonância utilizadas foram codificadas com oito bits de números inteiros positivos,  $R = 255$ .

# Capítulo 3

## Tumor Detection Framework

Nesse capítulo serão apresentadas todas as etapas do *framework* para a detecção de tumores cerebrais com o uso de mapas de saliência. Inicialmente, foi descrito o banco de dados original utilizado e quais adaptações foram feitas para o escopo desta pesquisa. Em seguida, é apresentado o fluxo de processos do modelo criado, que contempla: a geração de mapas de saliências, a clusterização e ranqueamento de regiões mais salientes dos mapas, a extração das regiões de interesse e por fim a classificação dessas regiões de interesse por meio de redes neurais convolucionais. Em cada fase do *framework*, são descritos e justificados os parâmetros e métodos utilizados.

### 3.1 BRATS 2020 Dataset

Os experimentos objeto deste trabalho foram conduzidos com utilização do banco de dados Multimodal Brain Tumor Segmentation Challenge 2020 (BRATS2020) criado por [et al 2015]. Esse dataset possui imagens de exames de ressonância magnética dos seguintes tipos: a) nativo (T1); b) ponderado em T1 pós-contraste (T1Gd); c) ponderado em T2 (T2) e d) fluido T2 volumes de recuperação de inversão atenuada (T2-FLAIR). Os arquivos das imagens estão no formato NifTI-1 e foram adquiridos de 371 pacientes, por meio de diferentes protocolos clínicos e diferentes modelos de scanners de 19 instituições.

Para cada paciente, há quatro exames de ressonância magnética 3D (tridimensional). As imagens foram segmentadas manualmente por até quatro avaliadores, seguindo-se o mesmo protocolo de anotação. As anotações compreendem o GD-enhancing (ET -etiqueta 4), o edema peritumoral (ED -etiqueta 2), o núcleo do tumor necrótico e o non-enhancing tumor core (NCR/NET —etiqueta 1). Do total de pacientes, 67 são classificados como Low Grade Gliomas (LGG), isto é, são casos em que o crescimento tumoral se dá de forma lenta. Os demais são classificados como High Grade Gliomas (HGG) que são mais agressivos e crescem rapidamente. Cada imagem 3D possui as seguintes dimensões:  $240 \times 240 \times 155$ . A Figura 3.1 mostra exemplos de cortes axiais de uma amostra do BRATS2020.

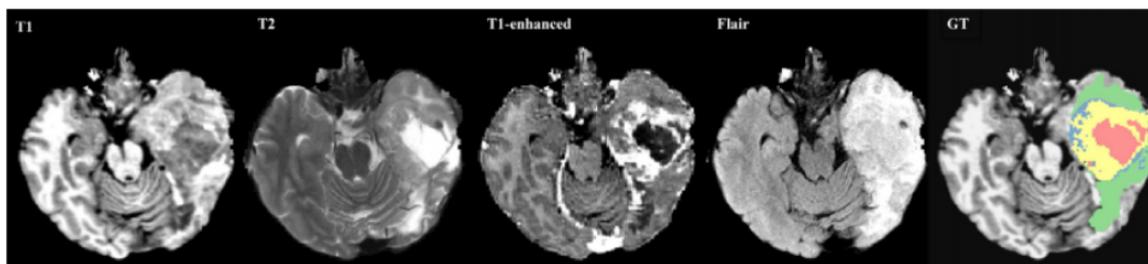


Figura 3.1: Amostra de um corte axial do BRATS2020 dataset. Imagem obtida em [et al 2015]

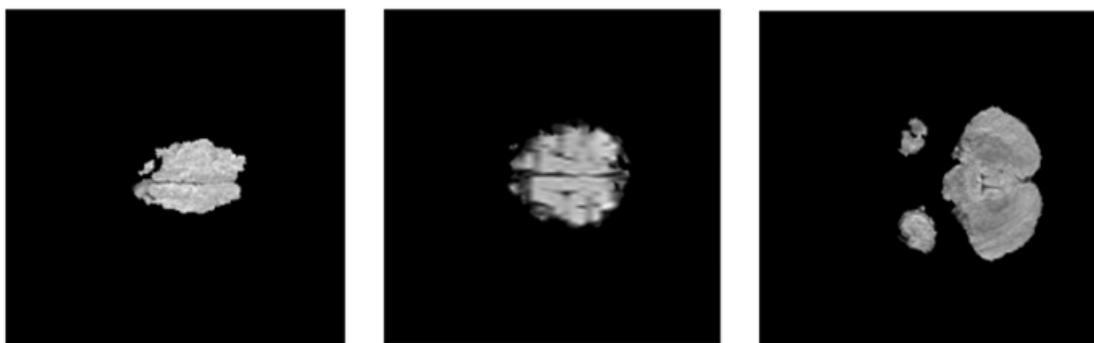


Figura 3.2: Exemplos de cortes excluídos do dataset. Imagens do dataset exibidos por meio de simulação computacional.

O banco de dados BRATS2020 é originalmente utilizado para testes e comparação de algoritmos de segmentação, ou seja, cada pixel é classificado individualmente em relação às anotações mencionadas. No entanto, o propósito deste estudo é classificar imagens bidimensionais quanto à existência ou não do tumor. Portanto, a partir dos arquivos do BRATS2020, foram realizadas adaptações para que ele fosse utilizado na análise de classificação binária entre imagens com tumor e imagens sem tumor.

Foram extraídos cortes 2D de todas as imagens 3D e foram eliminadas imagens próximas ao topo da cabeça e ao pescoço, porque, em geral, nessas imagens, a região cerebral se encontra reduzida. Além disso, esses cortes, de modo geral, contêm muitas informações (tecidos e estruturas anatômicas) não relevantes para classificação de tecidos tumorais na abordagem adotada neste estudo, como, por exemplo, tecidos ósseos. Isso está demonstrado na Figura 3.2.

As imagens foram redimensionadas para  $480 \times 480$  (com interpolação). Esse é um procedimento comum entre radiologistas neurológicos, e as dimensões  $480 \times 480$  estão na média de tamanho utilizado durante um diagnóstico. As máscaras referentes aos cortes foram binarizadas e permitiram anotar se a imagem continha ou não tumor, conforme demonstrado na Figura 3.3.

Após a extração dos cortes, foram selecionados aleatoriamente 8.099 *scans* 2D de imagens, dos quais 3.999 (49.4%) continham tumores cerebrais e 4.100 *scans* (50.6%) não continham, obtendo-se, portanto, um *dataset* balanceado. No Capítulo 5, foi demonstrado pos-

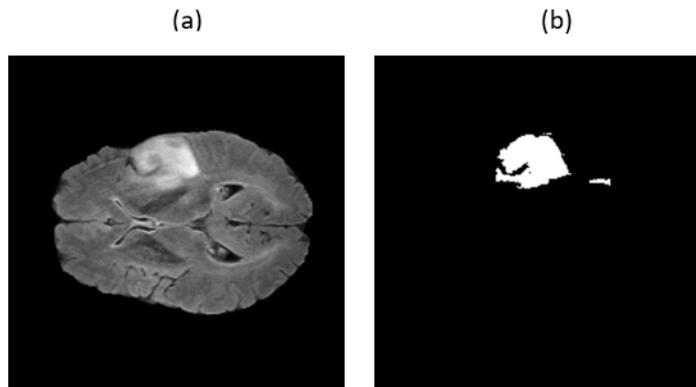


Figura 3.3: a) Corte axial do cérebro com presença de tumor e b) máscara correspondente binarizada com a região tumoral. Imagens do dataset exibidos por meio de simulação computacional.

teriormente que a escolha por imagens do tipo *Flair* foi correta, pois em um mesmo cenário de treinamento e teste, por meio de redes neurais, esse foi o tipo de imagem que obteve maior acurácia.

## 3.2 Framework proposto

O processo de detecção de tumor é composto de quatro etapas: geração dos mapas de saliência, clusterização e ranqueamento de regiões, extração da região de interesse e classificação por meio de rede neural. A Figura 3.4 mostra de maneira resumida esse processo.

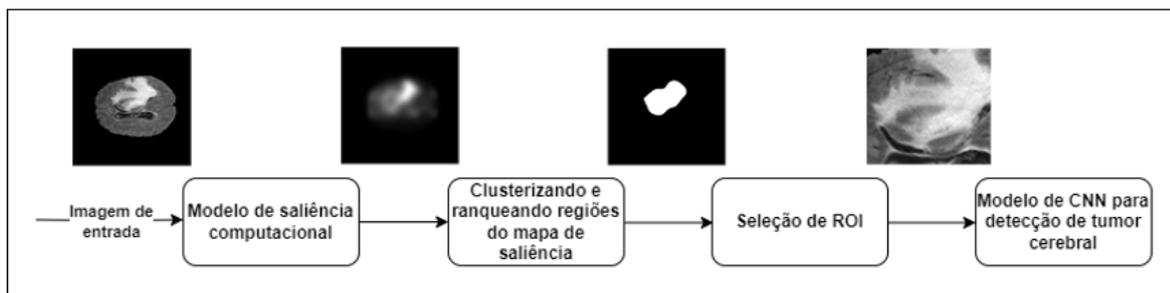


Figura 3.4: *Framework* proposto para detecção de tumor cerebral.

### 3.2.1 Geração de Mapas de Saliência

A geração dos mapas de saliência foi realizada por algoritmos *bottom-up*. Foram testados e em seguida comparados cinco algoritmos: GBVS, FES, CovSal, Itti e FES-GBVS. Os algoritmos GBVS, FES, CovSal e Itti foram descritos no capítulo anterior. Os mapas FES-GBVS foram gerados a partir da média aritmética dos mapas FES e GBVS que tiveram inicialmente os melhores resultados em termos de acerto da região tumoral. A Figura 3.5

mostra um exemplo de cada tipo de mapa de saliência e a Figura 3.6 apresenta um exemplo do mapa FES-GBVS. Visualmente foi possível constatar que as regiões mais salientes dos mapas correspondem, na maioria das vezes, às regiões tumorais das imagens de ressonância magnética. Essa observação levou à análise quantitativa de correlação demonstrada no Capítulo 5.

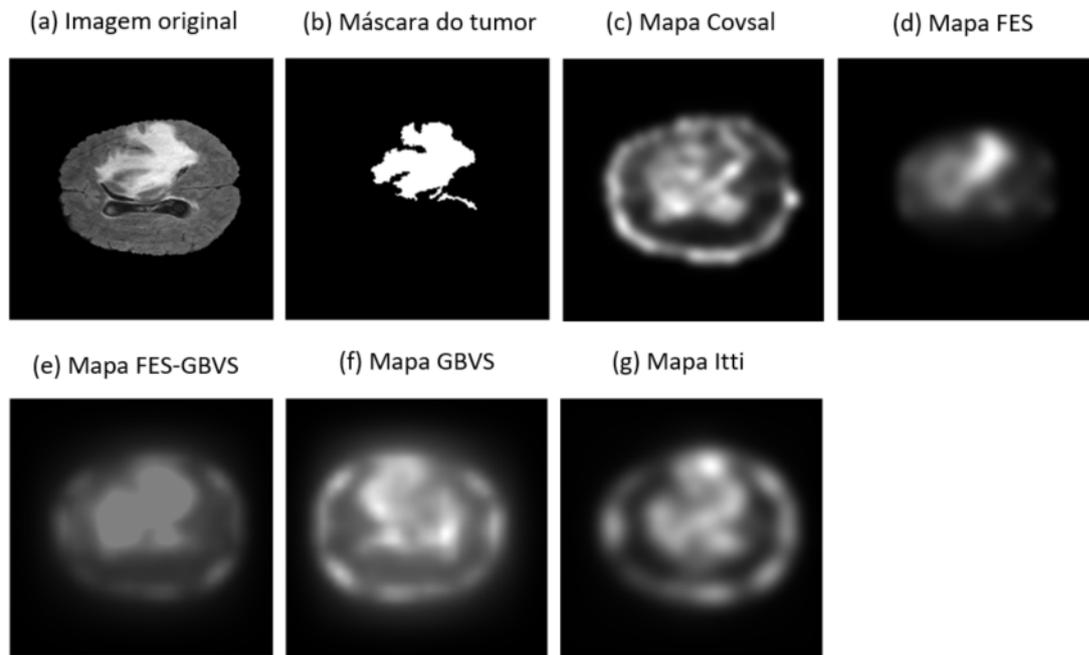


Figura 3.5: (a) Amostra de corte original do scan MRI, (b) máscara do tumor, e mapas de saliência correspondentes: (c) CovSal, (d) FES, (e) GBVS-FES, (f) GBVS e (g) Itti. Todos os algoritmos foram obtidos do site de referência MIT saliency [MIT ].

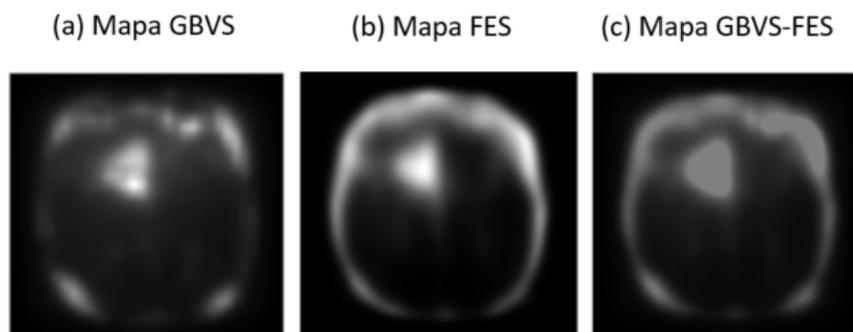


Figura 3.6: a) amostra de mapa GBVS, b) amostras de mapa FES e c) combinação linear GBVS-FES.

### 3.2.2 Clusterização e Ranqueamento de regiões

Com objetivo de identificar as regiões mais salientes, os mapas de saliência foram clusterizados com o algoritmo *k-means*, originalmente publicado por [Lloyd 1982]. Outras formas poderiam ser utilizadas para a identificação das regiões com brilho mais intenso, como, por

exemplo, janelas deslizantes. No entanto, o *k-means* se mostrou mais eficiente por capturar todas as regiões e também por ser um algoritmo consolidado pela literatura. O *k-means* é uma técnica clássica de clusterização que agrupa os pixels que possuem os valores próximos aos de um dos  $k$  centróides previamente escolhidos como referência. Trata-se de um algoritmo iterativo de duas etapas, que busca minimizar o erro de aproximação entre os valores de pixel e os valores dos centróides em um *cluster*. Na primeira etapa, que acontece em lotes, busca-se categorizar cada pixel da imagem em relação ao valor de centróide mais próximo.

A proximidade entre os pixels e o centróide é feita computando-se a distância euclidiana quadrática entre a amostra de pixel,  $x_m$ , e o valor dos centróides de cada *cluster*  $c_j$ . Os clusters  $c_j$  correspondem ao *j-ésimo cluster*, que varia de 1 a  $k$ . Após esses cálculos, a amostra de pixel é categorizada e os valores dos centróides de cada cluster são recomputados. Há várias técnicas para escolha de centróides. Neste estudo foi utilizada uma biblioteca com o algoritmo *k-means++*, que usa heurísticas para encontrar as sementes dos centróides [Arthur and Vassilvitskii 2007]. Essa técnica melhora o tempo de convergência em relação ao algoritmo Lloyd, dos autores [Lloyd 1982], além de gerar uma solução final de melhor qualidade.

A segunda etapa reclassifica os pixels caso a soma total das distâncias euclidianas em um *cluster* seja menor. Além disso, os centróides de um *cluster* são recomputados até que um mínimo global seja encontrado. Nesse caso, ocorrerá a convergência do algoritmo. Os mapas de saliência foram clusterizados com valor de  $k = 13$ . Os clusters são ranqueados de forma decrescente em relação aos valores dos centroids, de forma que o cluster  $K_1$  com valor de centroid  $C_1$ , tenha  $C_1 > C_2$ , que corresponde ao valor de centroid do cluster  $K_2$ . Dessa forma,  $C_1 > C_2 > C_3 \dots > C_N$ , onde  $N$  é a quantidade de clusters. A Figura 3.7, a seguir, mostra um exemplo de clusterização de uma amostra de mapa de saliência FES-GBVS. Quanto maior o valor de  $k$ , maior o número de clusters, maior o tempo de processamento de clusterização e menor o tamanho médio de cada *cluster*.

### 3.2.3 Extração da Região de Interesse (ROI)

Após a clusterização do mapa de saliência, as regiões com os maiores valores de centróide são mescladas. Dessa forma, as regiões mais intensas do mapa de saliência, ou seja, com maiores quantidades de pixels brancos concentrados, são identificadas e indicam as regiões que captariam maior atenção do olhar humano a partir das características mais primitivas, como por exemplo o contraste, a forma, a cor etc. Para selecionar a ROI, foi escolhida metade dos clusters, ou seja, os seis clusters com o nível de energia mais alto,  $K_1 \dots K_6$ . As escolhas do valor de  $k = 13$  e do número de regiões igual a 6, foram empíricas e conforme mostrado no Capítulo 5, buscou-se conciliar taxa de acertos de regiões tumorais e tamanho médio da ROI. Essas regiões salientes foram mescladas e, em seguida, traçou-se uma caixa delimitadora ao redor dessa região. A caixa delimitadora possui os seguintes parâmetros:  $x, y, h$  e  $w$ , em que  $x, y$  correspondem às coordenadas do canto superior direito do

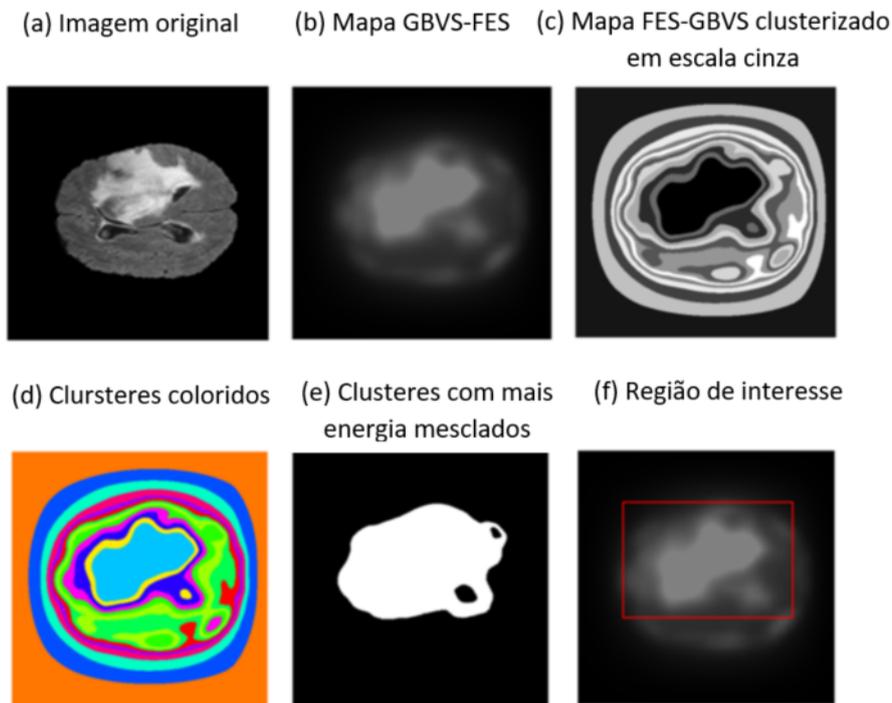


Figura 3.7: (a) *scan* MRI, (b) mapa FES-GBVS, (c) mapa de saliência clusterizado com  $k=13$ , (d) os clursteres pseudo-coloridos, (e) os seis clursteres com maior brilho mesclados, (f) máscara da ROI e g) ROI mostrada no mapa da saliência.

quadrante,  $h$  corresponde à altura e  $w$  corresponde à largura, conforme a Figura 3.8.

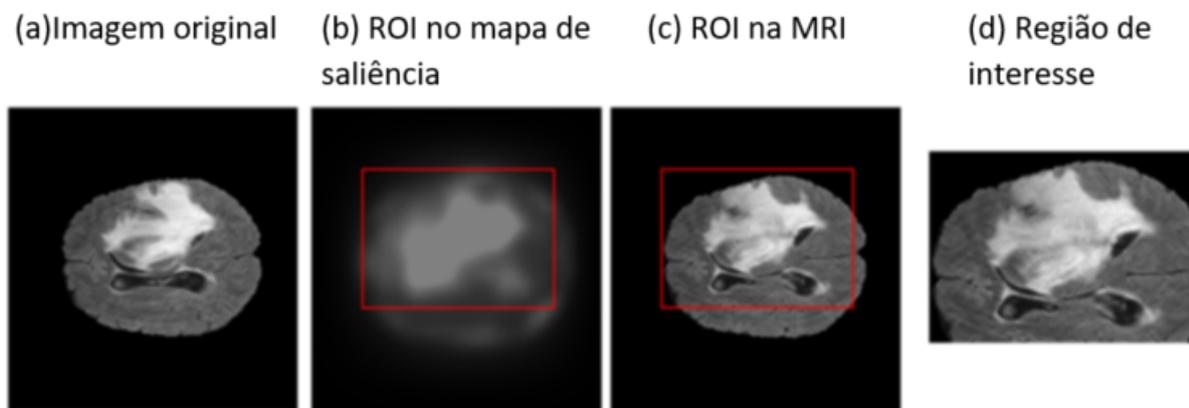


Figura 3.8: (a) *scan* MRI, (b) quadrante da ROI no mapa de saliência, (c) quadrante da ROI no *scan* MRI e (d) o recorte da ROI final.

### 3.2.4 Classificação por meio de CNNs

Para identificar se uma imagem tem um tumor, usa-se um sistema de classificação baseado em CNN. Mais especificamente, são consideradas quatro arquiteturas clássicas de CNNs: Resnet18, Resnet34, VGG16 e Alexnet. A codificação dos modelos usou bibliotecas *fastai2*, que são implementadas com a linguagem *Pytorch* [Paszke et al. 2017]. *Fastai* permite a criação e teste de modelos de redes neurais de maneira rápida, porque abstrai várias

complexidades do *Pytorch* [fast.ai].

Para realizar este treinamento, foram utilizados 80% das imagens para treinamento e 20% das imagens para teste. Como mencionado na Seção 3.1, cada um desses subconjuntos contém aproximadamente 50% das imagens com tumores e os 50% restantes são imagens sem tumores. Os pacientes do conjunto de treinamento, e conseqüentemente todas as imagens desses pacientes, são diferentes dos pacientes do conjunto de teste, para evitar a existência de imagens correlacionadas nos conjuntos de teste e de treino simultaneamente.

Para evitar qualquer viés estatístico na divisão do conjunto de imagens entre treino e teste, também se utilizou o *k-fold*. Com o *k-fold cross-validation*, um *dataset*  $D$  é dividido aleatoriamente em partes iguais,  $D_i$  com  $i = 1, 2, 3 \dots k$ . Nesse cenário, o modelo é treinado  $k$  vezes, e em cada uma das vezes um dos subconjuntos  $D_i$  é utilizado como conjunto de validação; os demais  $k - 1$  subconjuntos são utilizados para treino. Em geral, o valor de  $k$  utilizado varia entre 10 e 30 [Alpaydm 2009].

À medida que o valor de  $k$  aumenta, o número de treinos necessários também aumenta e dessa forma são obtidos estimadores mais robustos. Em contrapartida, os conjuntos de validação são menores e o custo computacional aumenta. No estudo realizado, foi utilizado o valor  $k = 10$  para subdivisão do conjunto de dados de treinamento. A estruturação do *dataset* com uso de imagens 2D permitiu uma grande quantidade de amostras e, como consequência, o uso do *k-fold*. Entretanto, em datasets públicos da área biomédica (com anotações de especialista) em geral são fornecidas poucas amostras.

Durante o treino das arquiteturas, foi usada a entropia cruzada como função de perda e o algoritmo *Adam* como otimizador. Foram considerados lotes de 64 imagens com taxa de aprendizado padrão de 0,002. Todas as arquiteturas avaliadas foram pré-treinadas, usando-se pesos importados do dataset *Imagenet*. A partir da última camada convolucional de cada modelo, a rede foi treinada usando-se o conjunto de dados específico do BRATS2020. Essa técnica é conhecida como *transfer-learning* e tem como vantagem fazer com que o modelo treinado generalize para amostras fora do *dataset*, ou seja, evita problema conhecido como *overfitting*.

# Capítulo 4

## Geração de Artefatos Comuns em Imagens de Ressonância Magnética

Neste capítulo é apresentado inicialmente o modelo matemático que descreve a degradação de imagens digitais. Cinco artefatos comuns em imagens de ressonância magnética foram simulados para analisar os impactos nas métricas de performance de uma rede neural. Esta rede classifica um *scan* de ressonância magnética quanto à existência ou não existência de tumor cerebral. Os artefatos testados foram: ruído Gaussiano, *blurring*, *ringing*, *ghosting* e contraste. Cada artefato foi conceituado segundo a literatura e, além disso, foram mostrados detalhes de suas implementações.

### 4.1 Artefatos Comuns em Ressonância Magnética

Durante a busca por *datasets* contendo imagens MR, notou-se em muitos casos a existência de artefatos que deterioram as imagens. Esses efeitos dificultam ou, muitas vezes, tornam inviável a análise visual de especialistas durante o diagnóstico. As principais fontes de degradação em imagens digitais podem estar no processo de aquisição ou no processo de transmissão das imagens. No processo de aquisição, questões como limitações operacionais dos sensores utilizados e temperatura ambiente poderão adicionar degradações (por exemplo, ruído eletrônico) nos sinais capturados. Quanto à transmissão de imagens, o canal poderá estar sob interferência eletromagnética de outras fontes. Um exemplo disso seria no caso de dados sendo transmitidos por uma rede *wireless* em um local com existência de motores elétricos ou microondas. Outras questões procedurais durante os exames também poderão interferir, como por exemplo, a movimentação da cabeça do paciente enquanto o scaneamento está sendo realizado ou a falta de perícia do técnico radiologista, que podem causar borramento. Mais especificamente, exemplos comuns em imagens de ressonância magnética são: ruído Gaussiano, *blurring*, *ghosting*, *ringing*, etc.

Com o avanço no desenvolvimento de ferramentas de suporte ao diagnóstico (CADs),

que usam ML e se baseiam em imagens (MRI, CT scans), viu-se neste estudo uma motivação no sentido de analisar os efeitos desses artefatos na acurácia dos modelos de redes neurais. Como as classificações realizadas neste estudo se basearam unicamente nas imagens MR 2D utilizadas como entrada para os modelos, foram investigados os efeitos que esses artefatos têm nas métricas de classificação. Essa investigação foi conduzida paralelamente ao principal tema desta dissertação. O experimento consistiu na simulação de cinco artefatos nas imagens de teste descritas na Seção 3.2.4. Essas degradações são as seguintes: ruído Gaussiano, *blurring*, *ringing*, contraste e *ghosting*.

Com base no modelo matemático apresentado no livro de [Gonzalez and Woods 2018], seja  $f(x, y)$  uma imagem digital submetida a uma degradação genérica dada por  $h(x, y)$ . Além disso, enquanto que  $n(x, y)$  é a degradação aditiva (por exemplo, o ruído). A imagem degradada  $g(x, y)$  é dada pela seguinte equação:

$$g(x, y) = (f * h)(x, y) + n(x, y), \quad (4.1)$$

em que  $(*)$  corresponde ao operador de convolução. No domínio da frequência, essa mesma expressão é dada pela seguinte equação:

$$G(u, v) = F(u, v)H(u, v) + N(u, v). \quad (4.2)$$

Para cada artefato, 20 níveis de degradação foram gerados. As imagens com cada nível de degradação de um artefato específico foram submetidas à classificação das arquiteturas CNN. Em seguida, as métricas de desempenho para a classificação, descritas na Seção 2.3, foram calculadas. Ao todo foram realizadas 100 simulações computacionais com diferentes cenários. As arquiteturas foram treinadas com as imagens originais. Entretanto, para analisar a robustez do algoritmo, testes foram realizados tanto com os originais quanto com as imagens degradadas.

### 4.1.1 Ruído Gaussiano

O ruído Gaussiano foi gerado a partir de pdf (*probability density function*) com distribuição Gaussiana, mostrado na Equação 4.3.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x - \mu)^2}{2\sigma^2} \quad (4.3)$$

Foi utilizada a função *random\_noise* da biblioteca *skimage.util*. Trata-se de uma biblioteca *open-source* da linguagem Python criada pelos autores [der Walt et al. 2014]. Uma semente pseudo-aleatória é utilizada para geração do ruído aditivo com distribuição Gaussiana. Além da imagem original, outro parâmetro dessa função é o valor de variância que altera a intensidade do ruído adicionado, ou seja, quanto maior a variância, maior a degradação da imagem. Os valores de variância utilizados para cada nível de degradação estão

mostrados na Tabela 4.1.

Nível de Ruído	Variância
1	0.26
2	0.39
3	0.58
4	0.82
5	1.11
6	1.46
7	1.88
8	2.38
9	2.96
10	3.62
11	4.37
12	5.23
13	6.19
14	7.25
15	8.4
16	9.75
17	11.18
18	12.75
19	14.46
20	16.32

Tabela 4.1: Valores de variância utilizados para os diferentes níveis de ruído Gaussiano aplicado.

Exemplos das imagens com os diferentes níveis de ruído podem ser observados na Figura 4.1.

### 4.1.2 Blurring

O *blurring* é uma suavização de variações locais de intensidades dos pixels. A imagem com *blurring* aparece borrada de maneira a não tonar visível, em algumas situações, os gradientes locais.

A geração desse efeito utilizou também filtros gaussianos aplicados nas direções horizontais e verticais das imagens. Mais especificamente, essa degradação foi gerada por meio da função *GaussianBlur* da biblioteca OpenCV, criada por [Bradski 2000]. Essa função realiza a convolução de matrizes 2D, nesse caso a imagem de ressonância, com kernel Gaussiano. O kernel é um filtro cujos pesos foram gerados por uma distribuição Gaussiana. Três parâmetros dessa função são: o tamanho do filtro dado pela variável *ksize*, valor da variável *sigmaX* e *sigmaY*. *SigmaX* corresponde ao desvio padrão da *pdf* Gaussiana na direção X e *sigmaY* corresponde ao desvio padrão da *pdf* Gaussiana na direção Y.

Os diferentes valores de *sigmaX* e *sigmaY* associados a cada nível de degradação podem ser visualizados na Tabela 4.2. Exemplos das imagens com os diferentes níveis de *blurring* podem observados na Figura 4.2.

Nível de Blur	SigmaX	SigmaY
1	0.3	0.3
2	0.91	0.91
3	1.53	1.53
4	2.14	2.14
5	2.76	2.76
6	3.37	3.37
7	3.99	3.99
8	4.61	4.61
9	5.22	5.22
10	5.84	5.84
11	6.45	6.45
12	7.07	7.07
13	7.68	7.68
14	8.30	8.30
15	8.92	8.92
16	9.53	9.53
17	10.15	10.15
18	10.76	0.76
19	11.38	11.38
20	12	12

Tabela 4.2: Valores de desvio padrão utilizados para gerar os diferentes níveis de *blurring* nas imagens de ressonância magnética.

### 4.1.3 Ringing

O *ringing* é um artefato que gera repetições transparentes e suavizadas das bordas de objetos nas imagens. Normalmente esse efeito acontece onde há transições de sinais e elas aparecem como múltiplas linhas alternadas, claras e escuras próximo às regiões de transições. Também conhecido como artefato de Gibbs [Budrys et al. 2018], o *ringing* acontece devido à natureza discreta dos dados e à limitação da amostragem, que deve ser finita, ao passo que a quantidade de componentes da operação de transformada de *Fourier* é infinita. Essas linhas são normalmente visíveis quando há mudanças abruptas e intensas, como no LCR-medula espinhal e na interface crânio-cérebro ou cérebro-background.

O efeito do *ringing* foi modelado de acordo com a equação  $G(u, v) = F(u, v) \times H(u, v)$ , onde  $H(u, v)$  corresponde a uma imagem binária, com um disco que simula uma perda de informações no processo de amostragem.

Exemplos das imagens com os diferentes níveis de *ringing* podem ser observados na Figura 4.3.

### 4.1.4 Ghosting

O *ghosting* é um ruído mostrado como uma ou mais reflexões repetidas do objeto principal em uma mesma imagem. Essas reflexões do objeto são borradas, transparentes e apare-

Nível de ringing	Raio do disco
1	98
2	94
3	90
4	85
5	81
6	76
7	72
8	67
9	63
10	58
11	54
12	50
13	45
14	41
15	36
16	32
17	27
18	23
19	18
20	14

Tabela 4.3: Valores dos raios de discos em imagens binárias utilizadas para gerar o efeito de *ringing*.

cem deslocadas na imagem de ressonância. Normalmente o *ghosting* é mais perceptível em regiões de *background* e pode ocorrer devido aos diferentes tipos de movimento do corpo do paciente durante o scaneamento, podendo incluir: fluxo sanguíneo, batimento cardíaco, movimentação da cabeça etc., como mostrado por [of Physicists in Medicine 2010]. Exemplos das imagens com os diferentes níveis de *ghosting* podem ser observados na Figura 4.4.

A função *Ghosting* usada faz parte da biblioteca TorchIO, criada por [Pérez-García et al. 2021], e oferece várias possibilidades de simulação do efeito como por exemplo: número de *ghosts*, valor de *restore*, eixo de replicação e intensidade. Para que visualmente fosse perceptível um nível crescente de degradação, todos os parâmetros tornaram-se valores fixos e o único parâmetro variável foi a intensidade. Dentre os parâmetros fixos, pode-se citar: número de *ghosts* = 5, eixo de referência 0, valor de *restore* = 0.02. Já a intensidade foi alterada para cada nível de degradação com valores mostrados na Tabela 4.4.

#### 4.1.5 Contraste

Os ajustes de contrastes são realizados por meio de transformações de pixel, ou seja, funções cujo valor de pixel de saída utilizam apenas o valor do pixel correspondente como parâmetro de entrada. O efeito do contraste foi modelado conforme a Equação 4.4. Os parâmetros  $\alpha > 0$  e  $\beta$  são frequentemente chamados de parâmetros de ganho e polarização; às vezes, diz-se que esses parâmetros controlam o contraste e o brilho, respectivamente

Nível de degradação	Intensidade
1	0.3
2	0.34
3	0.38
4	0.42
5	0.46
6	0.51
7	0.55
8	0.59
9	0.63
10	0.67
11	0.72
12	0.76
13	0.80
14	0.84
15	0.88
16	0.93
17	0.97
18	1.01
19	1.05
20	1.1

Tabela 4.4: Valores dos parâmetros de intensidade que modelaram os 20 níveis de ghosting.

[Szeliski 2022].

$$g(x, y) = \alpha f(x, y) + \beta \quad (4.4)$$

Os valores de alpha variam entre 0.09 e 0.945 e estão mostrados na Tabela 4.5. Os valores de beta são definidos de forma a se manter o histograma no centro. Em uma imagem em escala cinza, a escala de valores também é referenciada como *dynamic range*. Essa escala possui valores mínimos e valores máximos. O valor máximo da escala pode ser referenciado como valor de saturação e o valor mínimo pode ser referenciado como valor de ruído (nesse contexto). A diferença entre os valores máximos e mínimos de uma escala pode ser definida como contraste e a taxa de contraste (*contrast ratio*) como a razão entre essas duas quantidades. Exemplos das imagens com os diferentes níveis de contraste podem ser observados na Figura 4.5.

Nível de degradação por contraste	$\alpha$	$\beta$
1	0.95	6.04
2	0.9	11.80
3	0.86	17.56
4	0.81	23.32
5	0.77	29.08
6	0.72	34.84
7	0.67	40.60
8	0.63	46.36
9	0.59	52.12
10	0.54	57.88
11	0.50	63.64
12	0.45	69.40
13	0.41	75.16
14	0.36	80.92
15	0.32	86.68
16	0.27	92.44
17	0.22	98.20
18	0.18	103.96
19	0.14	109.72
20	0.09	115.48

Tabela 4.5: Valores dos parâmetros  $\alpha$  e  $\beta$  que modelaram os 20 níveis de contraste.

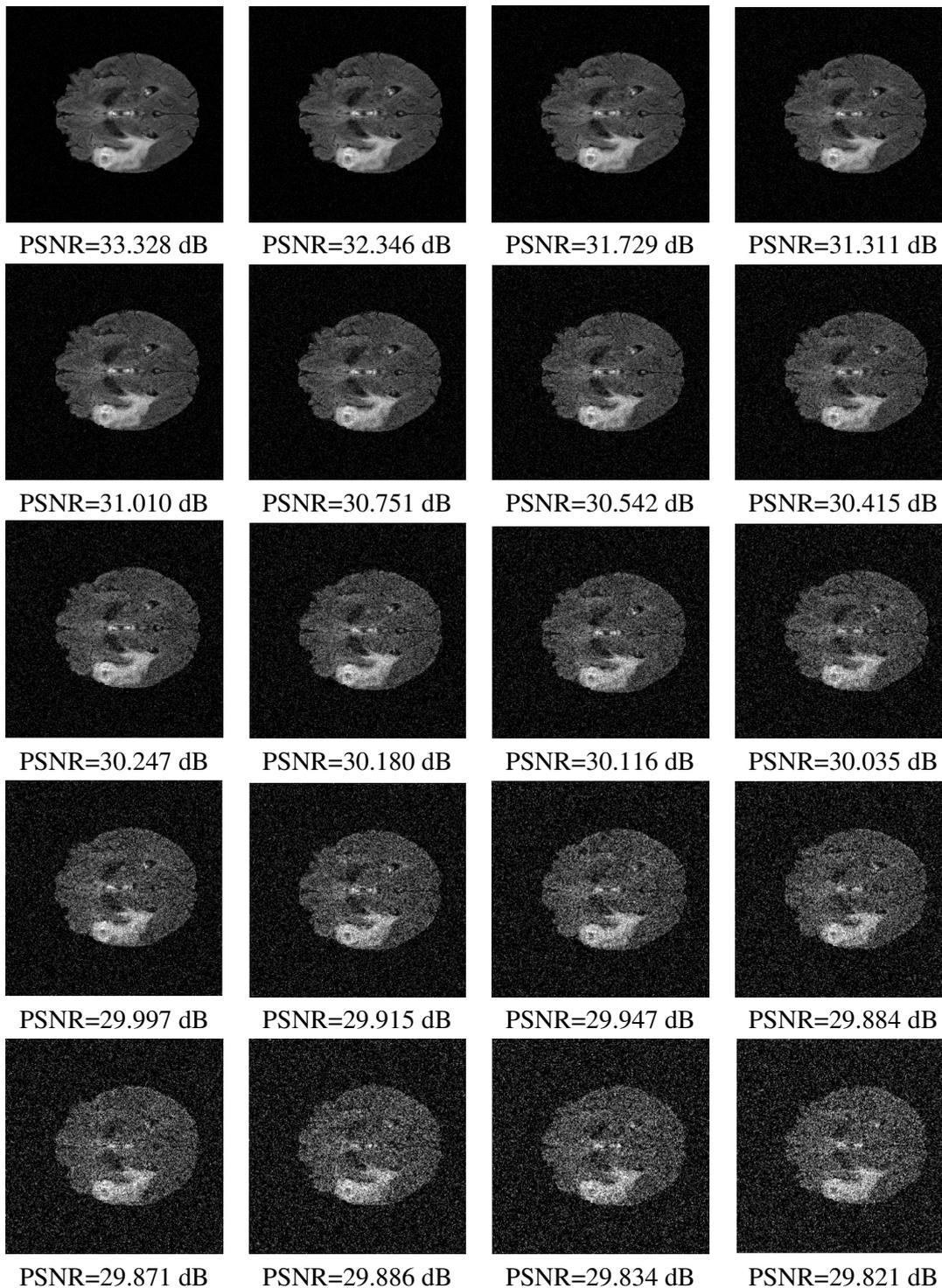


Figura 4.1: Amostras de uma mesma imagem de teste original com os 20 níveis de ruído Gaussiano, em ordem crescente da esquerda para a direita, de cima para baixo.

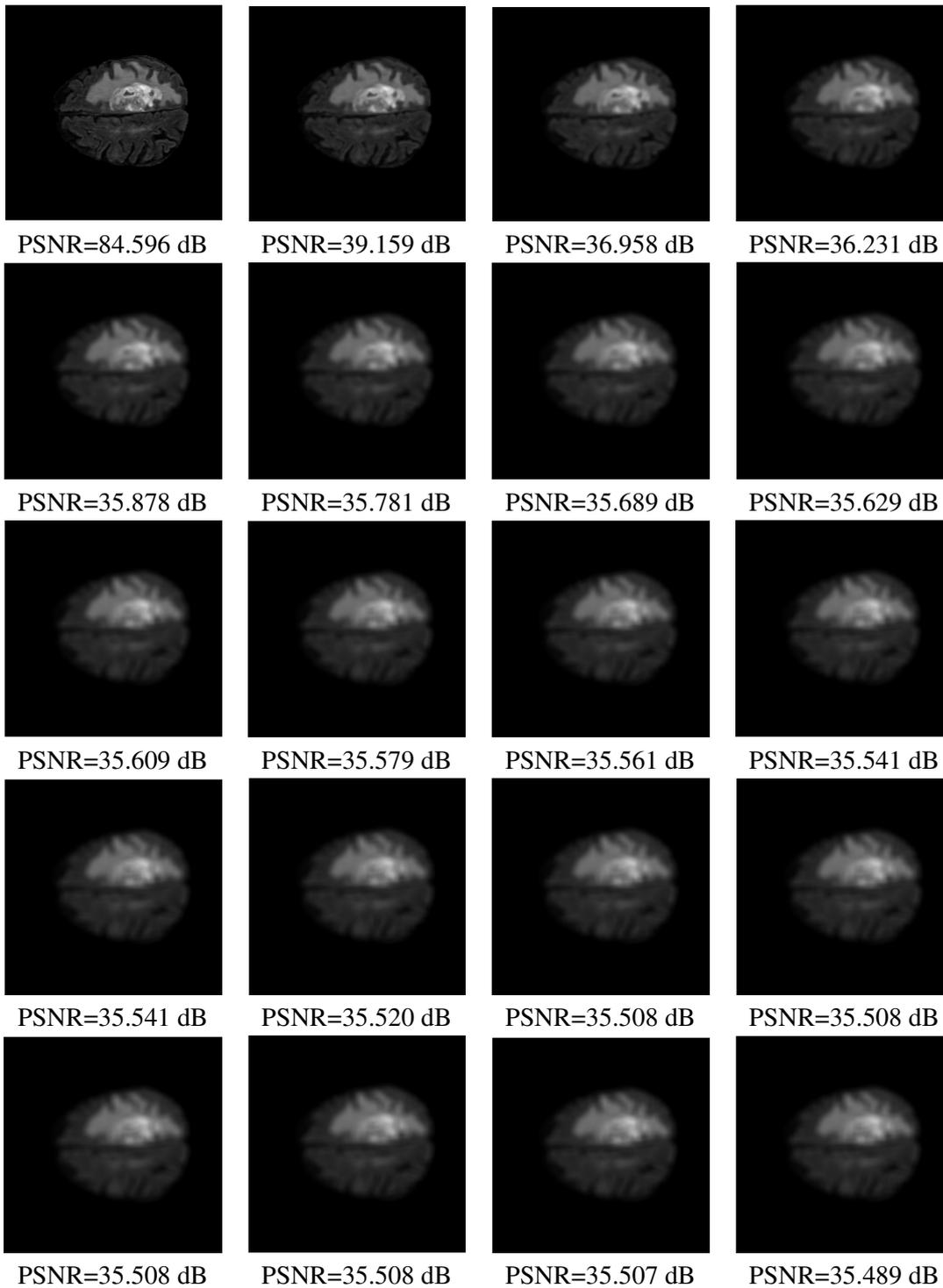


Figura 4.2: Amostras de uma mesma imagem de teste original com os 20 níveis de *blurring*, em ordem crescente da esquerda para a direita, de cima para baixo.

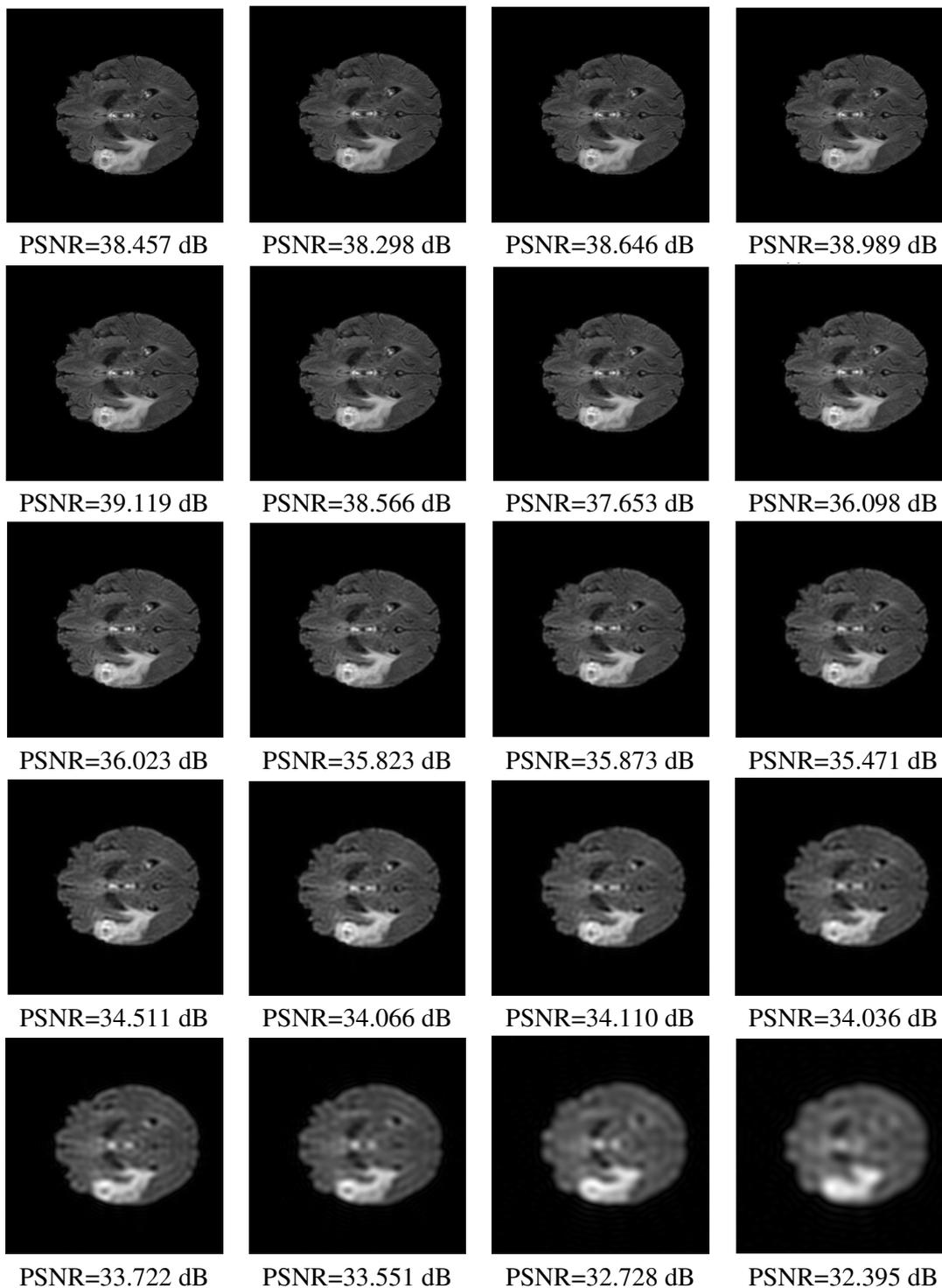


Figura 4.3: Amostras de uma mesma imagem de teste original com os 20 níveis de *ringing*, em ordem crescente da esquerda para a direita, de cima para baixo.

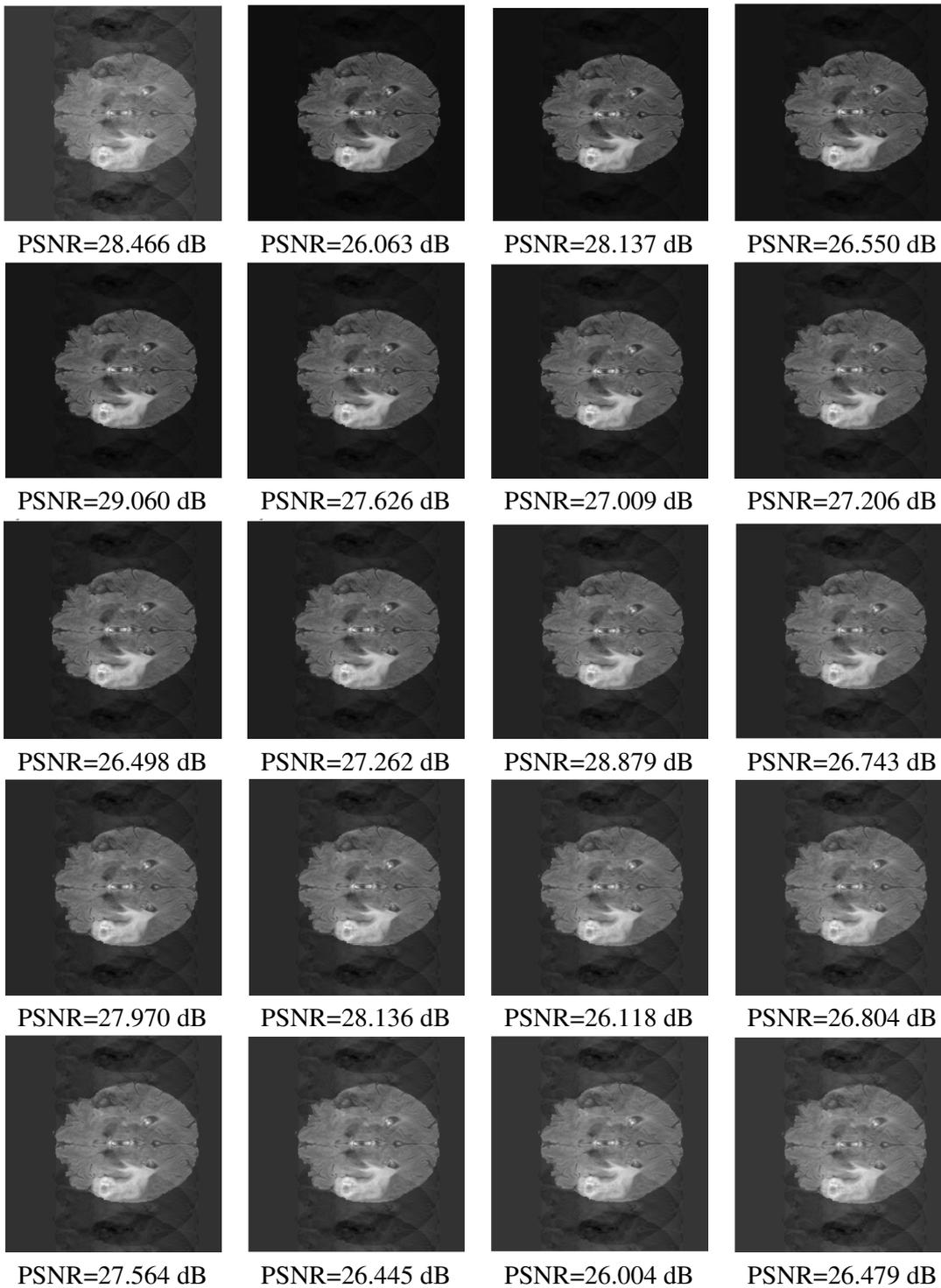


Figura 4.4: Amostras de uma mesma imagem de teste original com os 20 níveis de *ghosting*, em ordem crescente da esquerda para a direita, de cima para baixo.

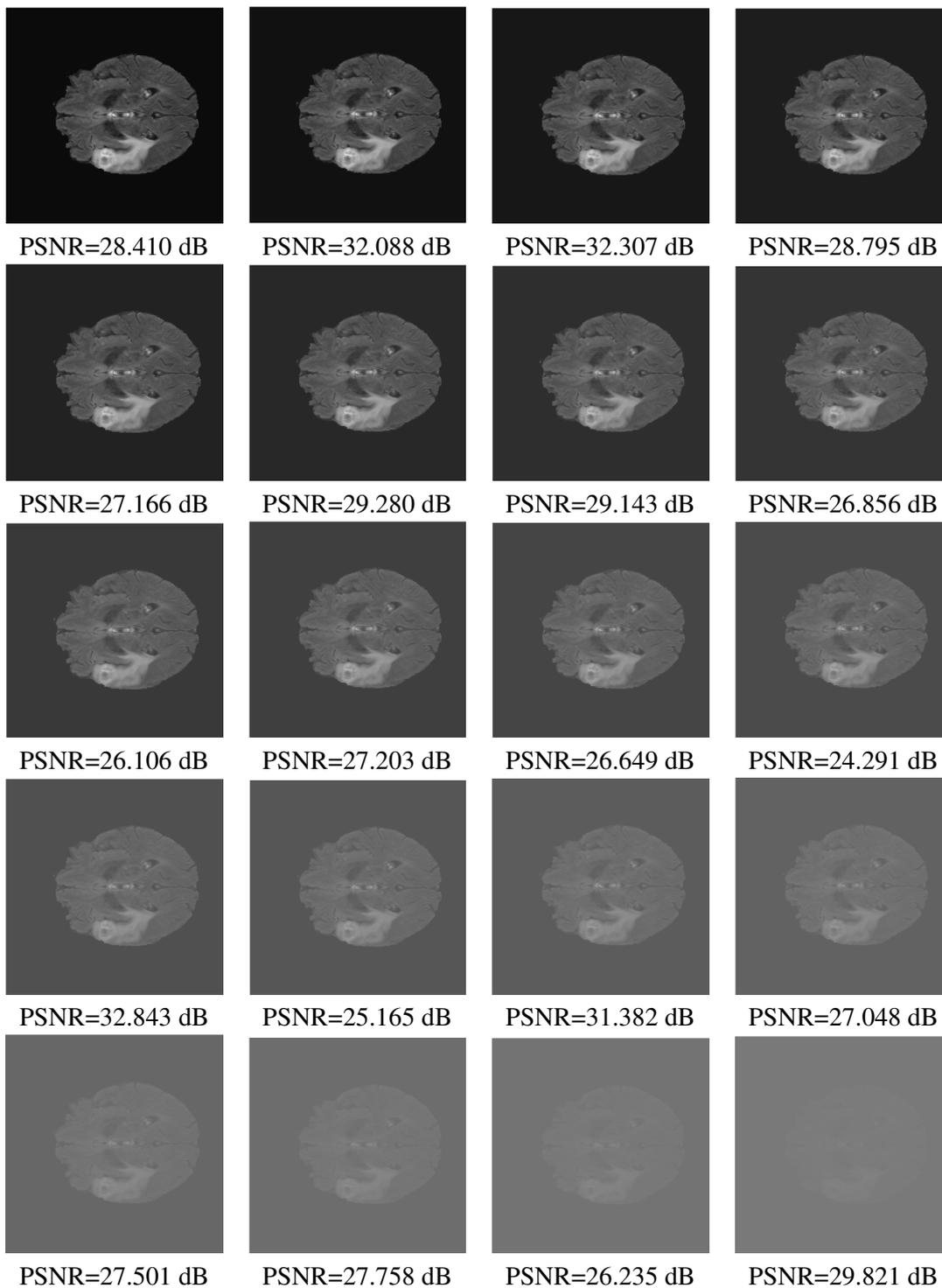


Figura 4.5: Amostras de uma mesma imagem de teste original com os 20 níveis de contraste, em ordem crescente da esquerda para a direita, de cima para baixo.

# Capítulo 5

## Resultados e Discussão

Neste capítulo são apresentados os principais resultados obtidos. Inicialmente, foram mostradas as métricas de performance para diferentes tipos de imagens de ressonância magnética. Em seguida, foram mostradas as análises quantitativas de correlação entre as regiões mais salientes dos mapas de saliência com as regiões tumorais. Ademais, foi realizada a comparação de resultados entre uma classificação convencional com imagens completas de ressonância magnética e uma classificação que utiliza o *framework* proposto no estudo. São mostradas as vantagens e limitações do modelo. Por fim, são mostradas as métricas de performance de redes neurais ao classificarem imagens degradadas pelos cinco artefatos discutidos no capítulo 4.

### 5.1 Comparação de performance entre os tipo de imagem de ressonância magnética

Conforme discutido na seção 3.1, o *dataset* criado possui quatro tipos de imagens de ressonância de cada paciente: T1, T2, T1ce e *Flair*. Para se ter um critério de decisão sobre qual tipo de imagem utilizar ao longo do estudo, foi realizada uma classificação binária quanto à existência ou à não existência de tumor com cada um dos tipos de imagens. Esta análise se deu com os mesmos parâmetros de simulação mostrados na Seção 3.2.4 com a rede neural *Resnet34*.

Foram utilizadas imagens completas como entrada da *Resnet34*. Outros modelos de rede neural poderiam ser utilizados, desde que as condições de simulações fossem as mesmas para cada um dos tipos de imagens comparados. A definição do modelo de rede neural utilizada é feita na criação do *dataloader*, objeto que configura os dados de entrada do modelo. A opção pelo modelo *Resnet34* se deu pela facilidade de já ser implementado como parte da biblioteca do *Fastai*.

A classificação inicial das imagens de teste com as mesmas amostras em todos os quatro

cenários mostrou que com imagens do tipo *Flair* os resultados obtidos eram melhores com valor de acurácia de 92% e *F1 score* de 92,5%, conforme mostrado na Tabela 5.1.

Dessa forma, a partir dos resultados mostrados na Tabela 5.1 optou-se pelo uso de imagens do tipo *Flair* nas demais simulações computacionais ao longo do estudo. Algumas abordagens utilizam os quatro tipos de imagens simultaneamente, alterando-se para esse propósito as configurações dos parâmetros de entrada da rede neural.

Tipo de MRI	Acurácia	Precisão	Recall	F1 score
<b>Flair</b>	<b>0.9217, std:0.0059</b>	<b>0.9077, std:0.0077</b>	<b>0.9420, std:0.0137</b>	<b>0.9245, std:0.0061</b>
T1	0.7575, std:0.0150	0.7845, std:0.0140	0.7513, std:0.0210	0.7221, std:0.0388
T2	0.8391, std:0.0144	0.8270, std:0.0117	0.8649, std:0.0333	0.8451, std:0.0163
T1ce	0.7334, std:0.0147	0.7563, std:0.0122	0.7276, std:0.0223	0.7028, std:0.0453

Tabela 5.1: Comparação de resultados de classificação entre diferentes tipos de imagens de ressonância: *Flair*, T1, T2 e T1ce. São exibidas as médias e desvios padrões de cada uma das métricas: acurácia, precisão, recall e *F1 score*.

## 5.2 Correlação entre Regiões Salientes e Regiões Tumorais

Conforme mostrado na Seção 3.2.1, em imagens que contêm tumor, foi constatado que as regiões mais brilhantes correspondiam às regiões tumorais. A literatura apresenta diferentes maneiras para se calcular quantitativamente essa correlação. No estudo realizado por [Elazary and Itti 2008], por exemplo, computou-se o centróide dos mapas de saliência e, em seguida, foi conferido o percentual de centróides de cada mapa, que se encontram dentro das regiões anotadas. No site MIT Saliency [MIT], são apresentadas algumas métricas, como AUC-JUD, SIM, NSS etc., que permitem a comparação entre diferentes algoritmos de predição de saliência.

Neste estudo, as regiões mais salientes de cada mapa foram clusterizadas e mescladas. Essas regiões mais salientes são posteriormente comparadas com as máscaras binárias originais das regiões tumorais. Essa comparação busca identificar se as regiões salientes estão parcial ou totalmente inseridas na região tumoral, permitindo o cálculo da correlação entre as regiões mais salientes e as regiões tumorais. Quanto maior o número de regiões salientes mescladas, maiores são as chances de esses clusters corresponderem a regiões tumorais, já que possuirão uma área maior na imagem. Em um cenário com valor de  $k$  igual a 13 no *k-means*, o total de correspondências entre as regiões salientes e as regiões tumorais é mostrado no gráfico de barras, evidenciado na Figura 5.1.

Quanto maior o valor de  $k$ , maior a quantidade de clusters, maior o tempo de processamento de clusterização do *k-means* e, também, menores os tamanhos médios de cada cluster. O critério de avaliação do melhor algoritmo dentre os comparados é o que permite acurácias acima de 90% com a menor quantidade de regiões salientes. Por meio desse critério, o FES-

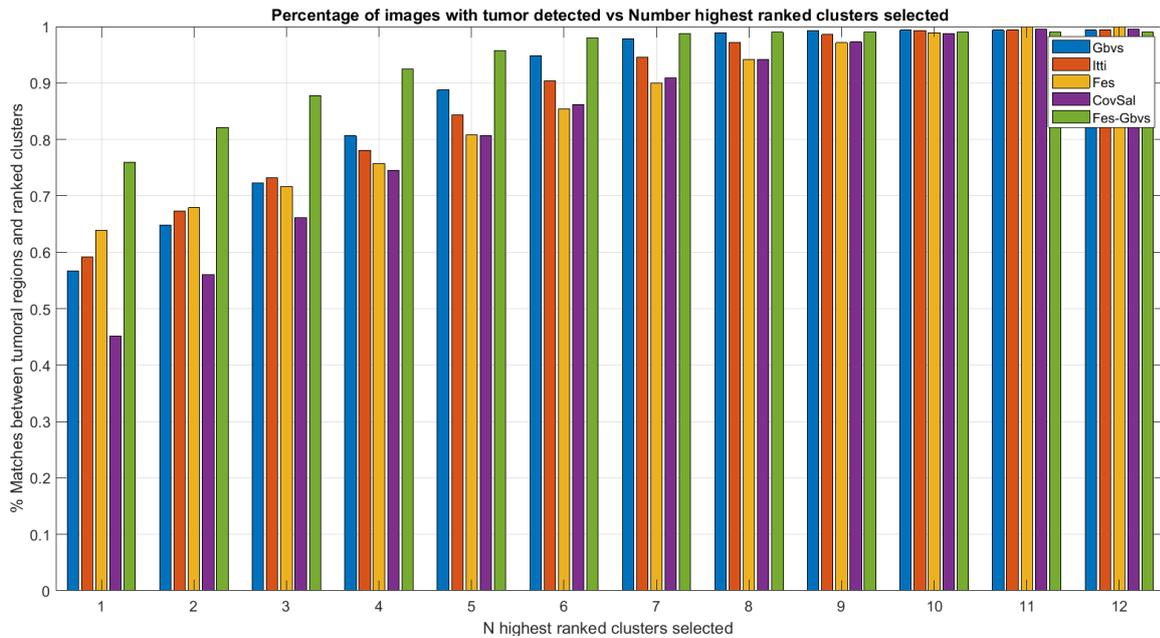


Figura 5.1: Porcentagem de imagens com tumor detectado versus número de *clusters* com maior intensidade de brilho selecionados. Valor de  $k = 13$ .

GBVS permitiu que em uma avaliação com seis regiões a taxa de acerto de regiões tumorais fosse de 98%, o melhor desempenho comparado em relação aos demais algoritmos. Sob o mesmo critério, o algoritmo com pior performance foi o CovSal, pois, para uma mesma quantidade de regiões salientes, o percentual de detecção de regiões tumorais foi o menor.

Diferentes quantidades de clusteres também foram testadas:  $k = 9, 11, 15, 17$  e os resultados podem ser visualizados nas Figuras 5.2 a 5.7, respectivamente.

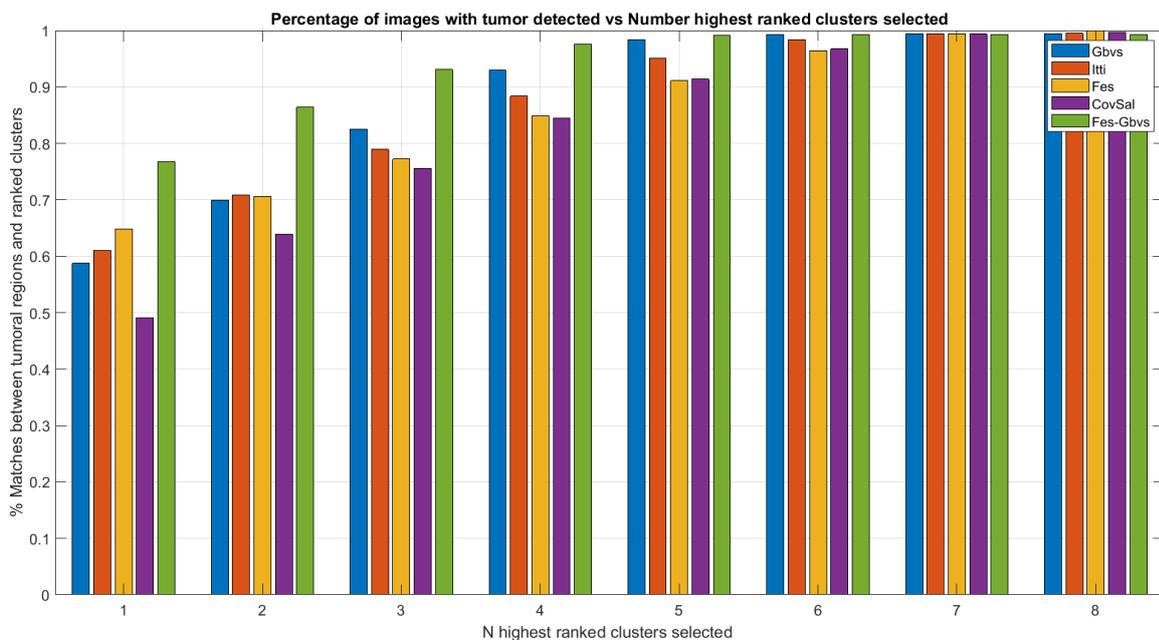


Figura 5.2: Porcentagem de imagens com tumor detectado versus número de clusteres com maior intensidade de brilho selecionados. Valor de  $k = 9$ .

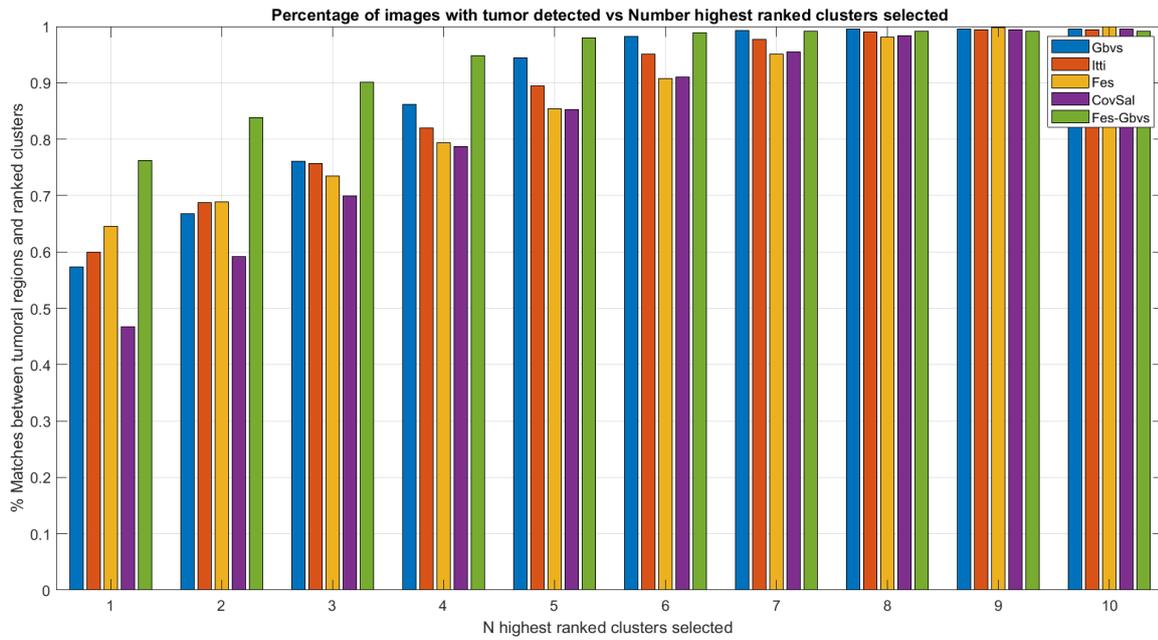


Figura 5.3: Porcentagem de imagens com tumor detectado versus número de clusters com maior intensidade de brilho selecionados. Valor de  $k = 11$ .

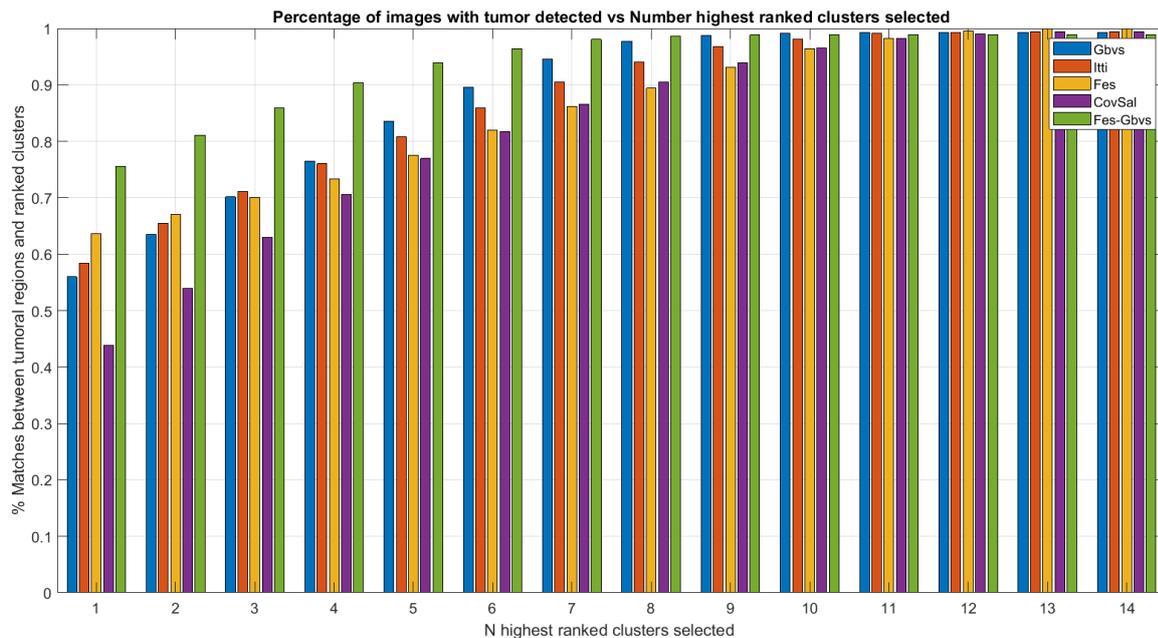


Figura 5.4: Porcentagem de imagens com tumor detectado versus número de clusters com maior intensidade de brilho selecionados. Valor de  $k = 15$ .

Em todos os cenários, com aproximadamente  $\lfloor \frac{k}{2} \rfloor$  regiões, o percentual de detecção de regiões tumorais foi acima de 90%. Observa-se também que em um dado momento, quando o número de clusters selecionados se aproxima do número total de clusters  $k$ , há uma saturação nos gráficos de barras no percentual de 100%. Isso é o esperado, pois à medida que mais regiões ranqueadas são mescladas, maior será a ROI que se aproximará do tamanho original da imagem de ressonância.

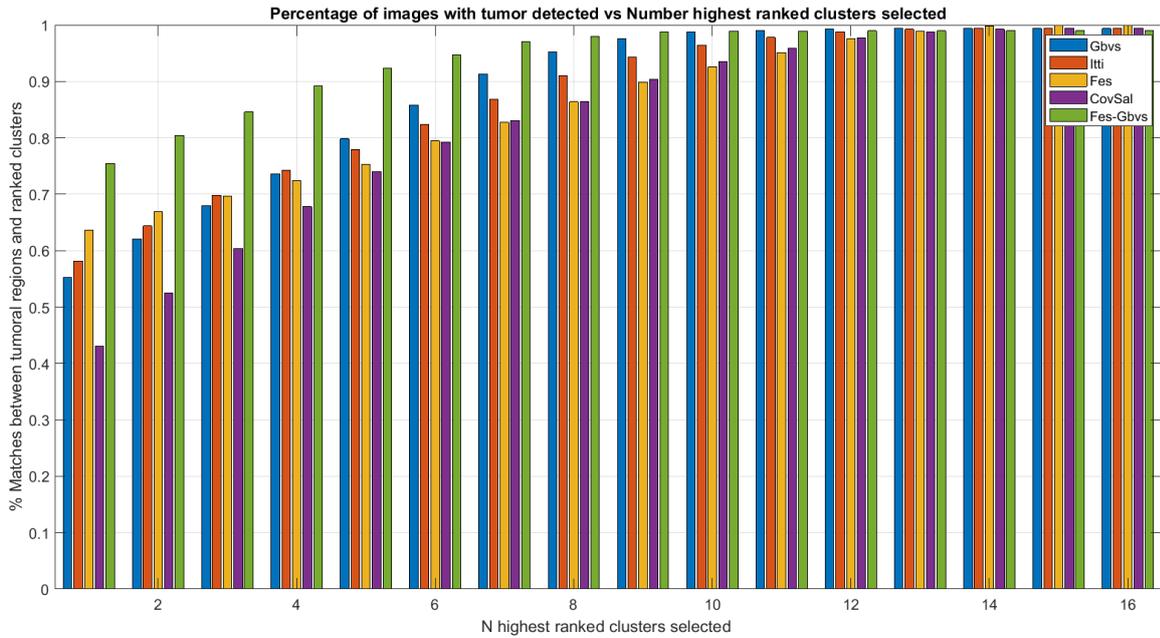


Figura 5.5: Porcentagem de imagens com tumor detectado versus número de clusteres com maior intensidade de brilho selecionados. Valor de  $k = 17$ .

### 5.3 Classificação de Tumores com o *Framework* Proposto

Os resultados de classificação por meio do uso do framework permitiram melhorar as taxas de acurácia na classificação quanto à existência de tumor com todos os modelos de CNNs testados.

Os parâmetros de simulação são os mesmos descritos na Seção 3.2.4. Os resultados da classificação final sem o framework proposto e com o framework proposto são mostrados nas Tabelas 5.1 e 5.2, respectivamente. Em cada uma das métricas mostradas, há o valor médio e o desvio padrão computados para os 10 conjuntos diferentes de amostras do conjunto de validação com o uso do  $k$ -fold. Dessa forma, a métrica de acurácia mostrada, por exemplo, é dada pela seguinte equação:

$$Accuracy = \frac{Accuracy\ 1^{st}\ fold + \dots + Accuracy\ 10^{th}\ fold}{10}. \quad (5.1)$$

A mesma analogia foi aplicada no cálculo da precisão, recall, e F1 score.

Modelo	Acurácia	Precisão	Recall	F1 score
Resnet18	0.9079, std:0.0099	0.9016, std:0.0078	0.9193, std:0.0219	0.9102, std:0.0106
Resnet34	0.9217, std:0.0059	0.9077, std:0.0077	0.9420, std:0.0137	0.9245, std:0.0061
<b>VGG16</b>	<b>0.925, std:0.008</b>	<b>0.9169, std:0.005</b>	<b>0.9394, std:0.021</b>	<b>0.9278, std:0.009</b>
Alexnet	0.9061, std:0.0077	0.8932, std:0.0092	0.9094, std:0.0078	0.9264, std:0.0160

Tabela 5.2: Acurácia, precisão, recall e F1 score que são os resultados de classificação de para Resnet18, Resnet34, VGG16 e Alexnet sem o *framework* proposto, ou seja, usando a imagem de ressonância magnética completa.

Modelo	Acurácia	Precisão	Recall	F1 score
Resnet18	0.9188, std:0.0064	0.8995, std:0.0092	0.9461, std:0.0100	0.9222, std:0.0061
Resnet34	0.9251, std:0.0093	0.9109, std:0.0072	0.9453, std:0.0142	0.9277, std:0.0093
VGG16	0.9271, std:0.0062	0.9085, std:0.0091	0.9300, std:0.0065	0.9528, std:0.0169
<b>Alexnet</b>	<b>0.9319, std:0.004</b>	<b>0.9117, std:0.006</b>	<b>0.9348, std:0.004</b>	<b>0.9592, std:0.010</b>

Tabela 5.3: Precisão, precisão, recall e F1 *score* que são os resultados de classificação de para Resnet18, Resnet34, VGG16 e Alexnet com o framework proposto, ou seja, usando a região de interesse extraída da imagem de ressonância magnética.

Utilizando-se as imagens completas (cenário sem o *framework* proposto), e tomando-se como métrica de referência o F1 *score*, a rede VGG16 foi a que teve melhor resultado com F1 *score* de 92,78%, enquanto a rede *Resnet18* foi a que teve o pior dos resultados, com F1 *score* de 91,02%. Utilizando-se os *crops* (cenário com o *framework* proposto), a rede *Alexnet* foi a que teve melhor resultado, com F1 *score* de 95,92%, enquanto a rede *Resnet18* obteve o pior dos resultados, com F1 *score* de 92,22%.

Quando a rede neural utiliza apenas regiões próximas às regiões tumorais, descartam outras regiões de uma imagem completa que não são relevantes para a classificação.

Em problemas de classificação por meio de CNNs, é importante que se verifique quais as regiões mais discriminativas nas imagens estão sendo utilizadas no processo de classificação. Essa análise é importante porque evita a implementação de modelos que tenham boas métricas, mas que utilizam características das imagens que sejam desconexas ao problema analisado. Essa verificação é feita por meio dos mapas de calor, em que as regiões com cores mais quentes mostram as regiões mais discriminativas utilizadas para a classificação, ou seja, onde os filtros convolucionais tiveram um peso maior. Nas Figuras 5.6 e 5.7, são mostrados os *heatmaps* para uma mesma amostra do conjunto de testes. Uma das amostras, Figura 5.6, corresponde a uma análise utilizando uma imagem completa, ao passo que na Figura 5.7 a amostra foi classificada por meio do *framework* proposto, ou seja, após a extração da ROI.

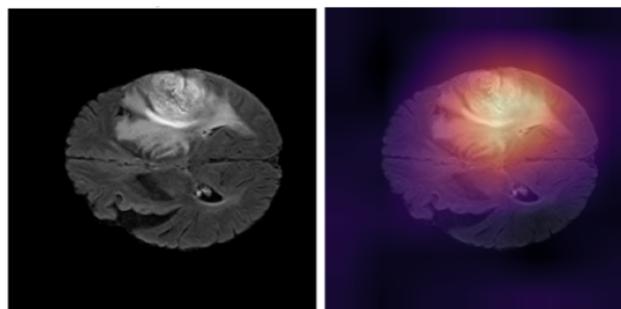


Figura 5.6: *Scan* MRI e sua máscara de ativação (CAM) correspondente após a classificação realizada pela Resnet18 sem o uso do *framework* proposto (usando *scan* MRI completo).

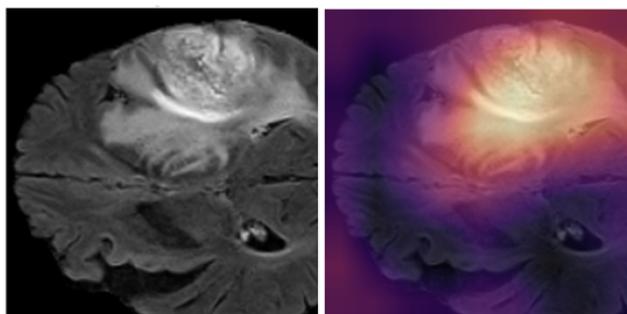


Figura 5.7: Scan MRI e sua máscara de ativação (CAM) correspondente após a classificação realizada pela Resnet18 com o uso do framework proposto (usando o *scan* da ROI).

## 5.4 Vantagens e limitações do Framework Proposto

Os *crops* submetidos às redes neurais foram dimensionados a partir de um valor máximo de *bounding box* computado, já que os lotes de uma rede neural devem ser quadráticos, com um mesmo tamanho de lado. Em várias amostras, os clusters mesclados com as regiões mais salientes tinham dimensões menores. Isso significa que se os *crops* originais fossem mantidos no conjunto de testes os resultados poderiam ser melhores, uma vez que haveria um foco maior na região tumoral e mais regiões não relevantes para a classificação seriam eliminadas.

No *framework* utilizado, tanto os modelos *bottom-up* quanto as redes neurais podem ser substituídos. Poderia ser utilizada, por exemplo, uma rede *Resnet101* que é mais profunda, ou poderia ser utilizado um modelo *top-down* como algoritmo de saliência. Os modelos *top-down* atualmente são os que correspondem ao estado da arte entre os algoritmos de predição de saliência. No entanto, o estudo falha em cenários em que as regiões tumorais são muito pequenas, conforme mostrado na Figura 5.8.

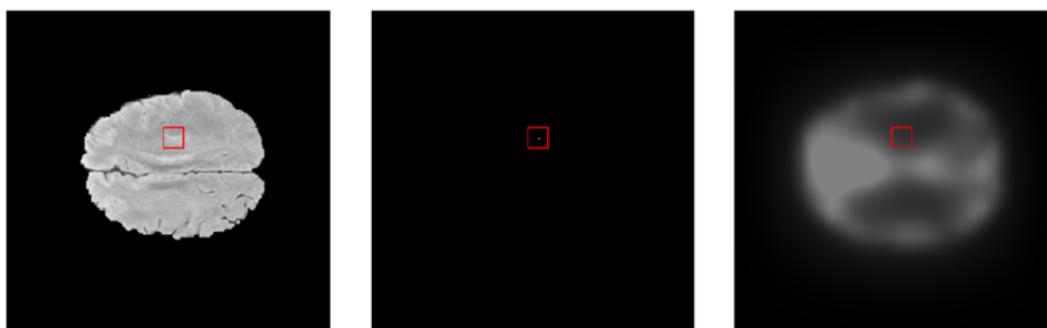


Figura 5.8: Amostra de uma imagem com tumor cuja área ocupa menos de 0.3% da área cerebral.

É importante observar que, nesse exemplo, é difícil perceber visualmente a existência de um tumor. Um radiologista neurológico nesse cenário provavelmente utilizaria outras imagens e exames para confirmá-la. No mapa de saliência gerado, a região tumoral também não é destacada. Dessa forma, quando o mapa é clusterizado, essa região tumoral fica fora da região de interesse e, portanto, é incorretamente classificado pela rede neural. Uma análise

quantitativa foi realizada para esse cenário, computando-se a área de todos os exemplos com tumores menores que não foram identificados nas regiões mais salientes dos mapas de saliência. Na média, as áreas desses tumores correspondiam a aproximadamente 0.3% da região cerebral de saliência.

## 5.5 Classificação das Imagens Degradadas

O valor médio  $\overline{\text{PSNR}}$  e  $\overline{\text{MSE}}$  para cada um dos níveis de degradação foi computado como a média dos valores de  $\overline{\text{PSNR}}$  de cada imagem do conjunto de teste (1628 imagens) com esse nível de degradação. Para cada um dos artefatos testados, a seleção dos 20 níveis de degradação buscou uma escala de valores cujo nível máximo de degradação pudesse gerar no classificador binário o pior resultado possível, ou seja, uma acurácia próxima a 50%.

Conforme mostrado nas Equações 2.2 a 2.4, uma métrica terá efeito na outra. Dessa forma, o comportamento obtido de acurácia, seja de incremento ou decremento; por exemplo, será semelhante ao comportamento no gráfico da precisão. Em todos os cenários, foi utilizado o mesmo modelo de rede neural, *Resnet34*. Os parâmetros de simulação foram os mesmos descritos na Seção 3.2.4. O modelo foi treinado com imagens originais sem nenhuma degradação. Já as imagens de testes foram degradadas.

### 5.5.1 Ruído Gaussiano

Com o nível 1 de degradação, a  $\overline{\text{PSNR}}$  é 33.42 dB e a acurácia correspondente é de 93.7%. Com o nível de degradação 20, a  $\overline{\text{PSNR}}$  é de 29.95 dB e a acurácia é de 53.4%, conforme se verifica na Tabela 5.4. Há um decaimento significativo da curva de acurácia com ruído Gaussiano. Mesmo com o nível máximo de ruído Gaussiano aplicado, visualmente é possível perceber a região tumoral. No último nível de degradação, a acurácia atinge valor de aproximadamente 53.4% o que torna a classificação das imagens aleatória. Visualmente, como mostrado na Figura 5.9, é possível ver que a partir do nível 17 de degradação os detalhes mais grosseiros da região central (*Cerebrospinal fluid*) se tornam menos perceptíveis.

### 5.5.2 Blurring

O 1º nível de *blurring* tem uma  $\overline{\text{PSNR}}$  de 70.11 dB e a acurácia correspondente de aproximadamente 92%, ao passo que o 20º nível tem uma  $\overline{\text{PSNR}}$  de 35.6 dB, com uma acurácia de 85.7%, conforme Tabela 5.5. O principal efeito do *blurring* é suavizar as zonas de gradiente o que dificulta que um especialista identifique os detalhes finos, como por exemplo, saber quais as regiões de edema ou quais as regiões necrosadas.

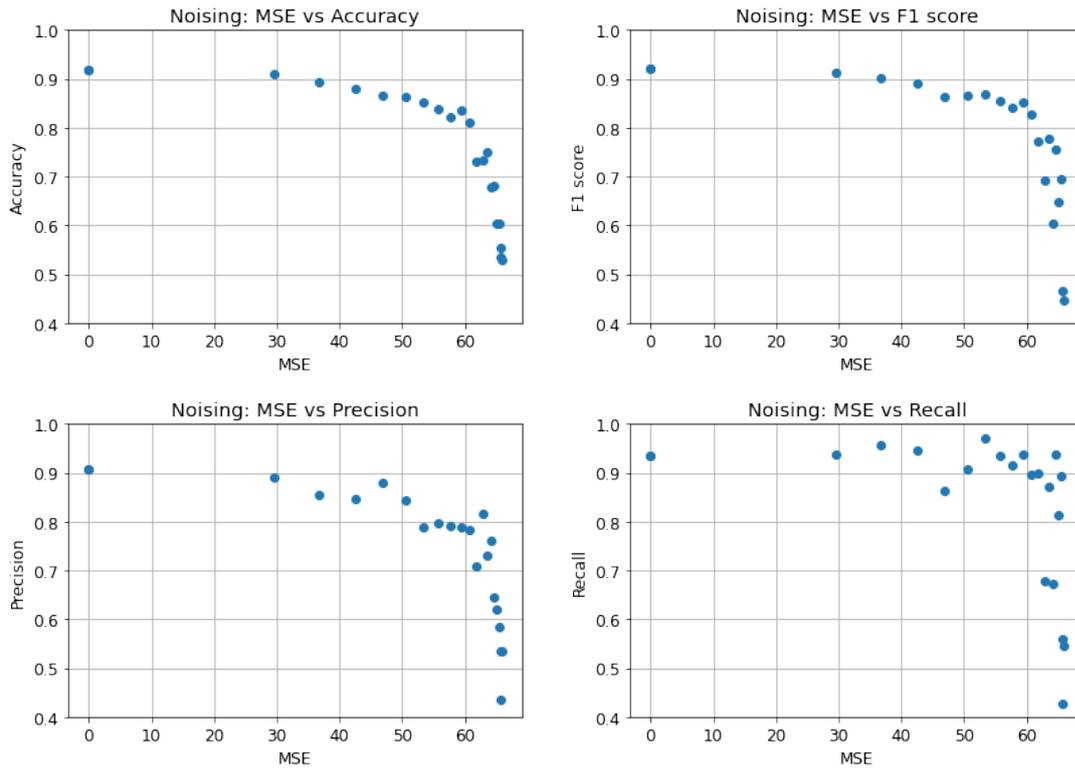


Figura 5.9: Resultados de classificação (acurácia, precisão, recall e F1 score) de imagens com ruído Gaussiano com relação ao valor médio da MSE de cada nível de degradação.

Níveis de ruído	PSNR	MSE	Acurácia	Precisão	Recall	F1 score
1	33.428	29.555	0.908	0.891	0.937	0.912
2	32.477	36.794	0.893	0.855	0.956	0.900
3	31.854	42.466	0.880	0.845	0.945	0.889
4	31.423	46.890	0.864	0.878	0.864	0.862
5	31.106	50.446	0.861	0.844	0.905	0.866
6	30.859	53.391	0.850	0.789	0.971	0.869
7	30.669	55.783	0.837	0.796	0.935	0.855
8	30.516	57.781	0.820	0.791	0.914	0.841
9	30.397	59.393	0.834	0.787	0.937	0.852
10	30.300	60.733	0.809	0.783	0.895	0.825
11	30.220	61.852	0.730	0.708	0.898	0.771
12	30.155	62.783	0.733	0.815	0.679	0.690
13	30.102	63.557	0.749	0.730	0.871	0.777
14	30.060	64.181	0.677	0.761	0.671	0.604
15	30.026	64.679	0.682	0.644	0.938	0.754
16	30.000	65.069	0.602	0.619	0.812	0.648
17	29.979	65.383	0.602	0.584	0.892	0.694
18	29.965	65.600	0.555	0.535	0.558	0.466
19	29.955	65.746	0.535	0.434	0.427	0.364
20	29.948	65.860	0.530	0.534	0.544	0.445

Tabela 5.4: Resultados de classificação com relação ao valor médio da MSE de cada nível de degradação de ruído Gaussiano.

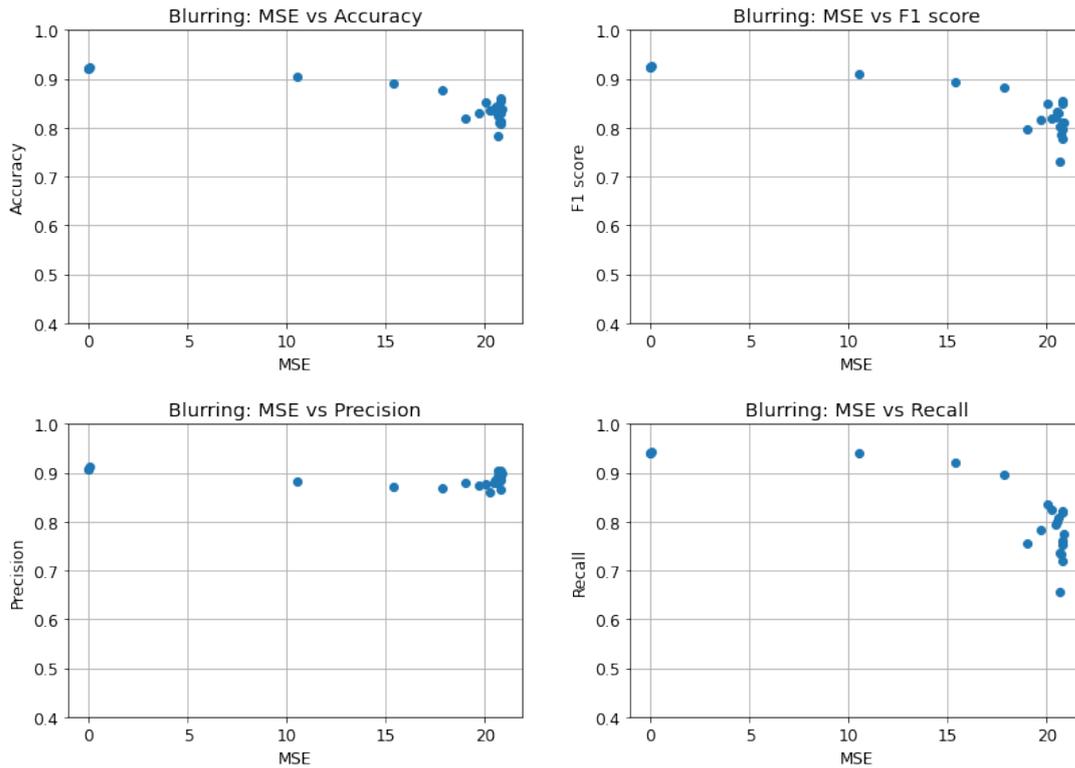


Figura 5.10: Resultados de classificação (acurácia, precisão, recall e F1 score) de imagens com *blurring* com relação ao valor médio da MSE de cada nível de degradação.

Níveis de Blur	PSNR	MSE	Acurácia	Precisão	Recall	F1 score
1	70.119	0.008	0.919	0.912	0.931	0.921
2	42.690	3.784	0.930	0.907	0.961	0.933
3	39.653	7.552	0.914	0.870	0.977	0.920
4	38.327	10.091	0.904	0.858	0.974	0.912
5	37.476	12.139	0.892	0.872	0.927	0.897
6	36.908	13.727	0.891	0.844	0.964	0.900
7	36.401	15.328	0.870	0.830	0.941	0.880
8	36.228	15.927	0.872	0.849	0.914	0.877
9	35.815	17.418	0.870	0.837	0.931	0.879
10	35.769	17.630	0.871	0.882	0.867	0.871
11	35.656	18.074	0.866	0.862	0.883	0.868
12	35.543	18.534	0.846	0.858	0.846	0.842
13	35.404	19.119	0.850	0.872	0.833	0.840
14	35.346	19.346	0.865	0.864	0.881	0.866
15	35.312	19.505	0.865	0.845	0.906	0.872
16	35.260	19.742	0.846	0.838	0.878	0.849
17	35.253	19.770	0.847	0.876	0.819	0.841
18	35.217	19.937	0.865	0.873	0.864	0.866
19	35.207	19.957	0.858	0.834	0.909	0.866
20	35.085	20.513	0.856	0.876	0.838	0.854

Tabela 5.5: Resultados de classificação com relação ao valor médio da MSE de cada nível de degradação com *blurring*.

### 5.5.3 Ringing

O 1º nível de *ringing* tem uma  $\overline{\text{PSNR}}$  de 39.5 dB com acurácia de 92.6 % ao passo que o 20º nível tem uma  $\overline{\text{PSNR}}$  de 32.5 dB e nível médio acurácia de 58.12%, conforme demonstrado na Tabela 5.6. A repetição das bordas sobrepostas gera um efeito final semelhante ao *blurring*. Nesse cenário, também há um decaimento da acurácia da rede neural à medida que o nível de degradação aumenta. O efeito do *ringing* é mais nítido nos últimos níveis nas regiões próximas às bordas da região cerebral. No último nível com um valor de  $\overline{\text{PSNR}}$  de 32.52 dB, conforme mostrado na Tabela 5.6, é mais fácil se perceber as repetições de linhas.

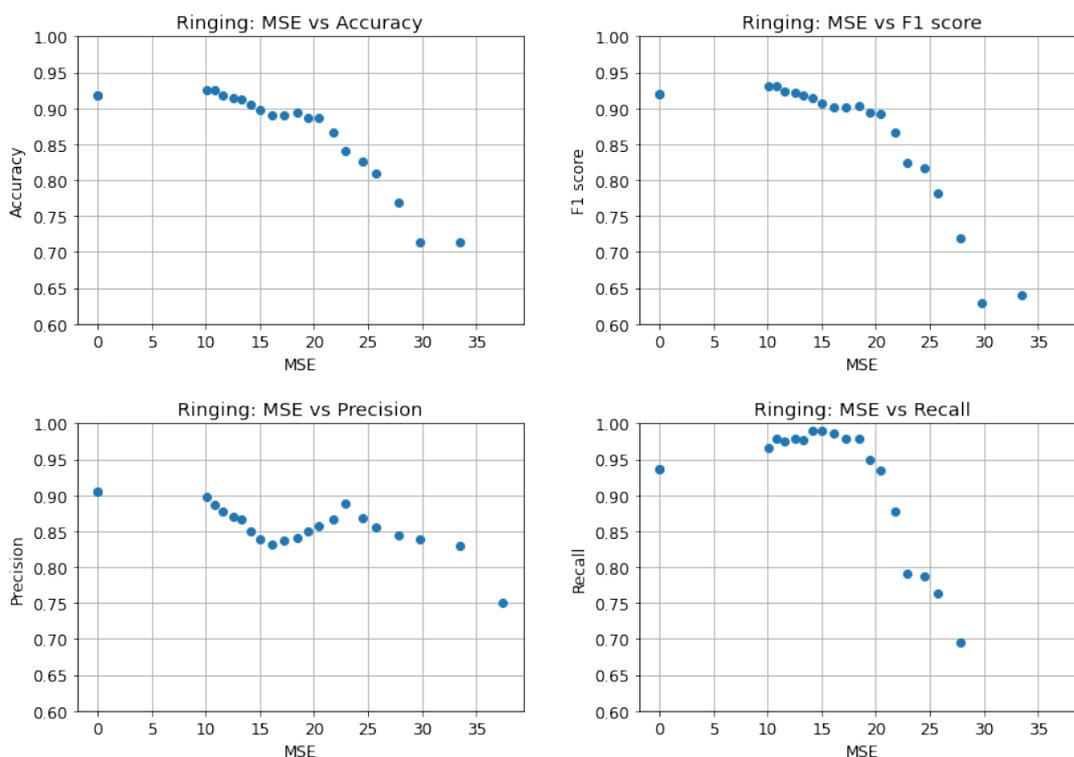


Figura 5.11: Resultados de classificação (acurácia, precisão, recall e F1 score) de imagens com *ringing* com relação ao valor médio da MSE de cada nível de degradação.

### 5.5.4 Ghosting

O 1º nível de *ghosting* tem uma  $\overline{\text{PSNR}}$  de 27.57 dB com nível de acurácia de 78.58%, ao passo que o 20º nível tem uma  $\overline{\text{PSNR}}$  de 27.574 dB com nível de acurácia de 85.2%, conforme se vê na Tabela 5.7. Diferentemente dos artefatos descritos anteriormente, o aumento dos níveis de *ghosting* teve como resultado valores aleatórios para os resultados de acurácia de classificação.

Em outras palavras, à medida que os reflexos ficam mais nítidos, os níveis de acurácia podem ficar maiores ou menores. Não se tem com clareza o efeito da degradação no processo de classificação. Além disso, as métricas de qualidade utilizadas nessa situação não

Nível de ringing	PSNR	MSE	Acurácia	Precisão	Recall	F1 score
1	39.509	10.087	0.926	0.896	0.966	0.930
2	39.055	10.891	0.925	0.887	0.979	0.930
3	38.634	11.619	0.917	0.876	0.975	0.923
4	38.174	12.529	0.914	0.870	0.978	0.921
5	37.833	13.278	0.911	0.867	0.977	0.918
6	37.440	14.186	0.905	0.850	0.988	0.914
7	37.105	15.031	0.897	0.838	0.990	0.907
8	36.691	16.176	0.890	0.831	0.986	0.902
9	36.348	17.233	0.890	0.836	0.977	0.900
10	35.971	18.463	0.894	0.841	0.978	0.903
11	35.699	19.401	0.886	0.849	0.948	0.894
12	35.430	20.423	0.885	0.857	0.933	0.892
13	35.102	21.778	0.866	0.866	0.876	0.867
14	34.842	22.909	0.839	0.889	0.791	0.823
15	34.510	24.444	0.825	0.867	0.786	0.816
16	34.251	25.738	0.808	0.854	0.763	0.781
17	33.882	27.775	0.768	0.845	0.694	0.719
18	33.552	29.795	0.714	0.839	0.553	0.629
19	33.023	33.429	0.713	0.830	0.555	0.640
20	32.516	37.443	0.581	0.750	0.261	0.356

Tabela 5.6: Resultados de classificação com relação ao valor médio da MSE de cada nível de degradação com *ringing*.

quantificam adequadamente esse incremento de deterioração. Os valores, tanto de  $\overline{\text{PSNR}}$  quanto de  $\overline{\text{MSE}}$ , podem ser maiores ou menores em relação ao nível de degradação anterior. Isso gera um comportamento aleatório nos gráficos mostrados na Figura 5.12.

### 5.5.5 Contraste

O 1º nível de contraste tem uma  $\overline{\text{PSNR}}$  de 28.29 dB com um nível de acurácia de 91.69% ao passo que o 20º nível tem uma  $\overline{\text{PSNR}}$  de 30.126 dB com nível de acurácia de 51.14%, conforme se verifica na Tabela 5.8. À medida que o contraste diminui, os níveis de acurácia diminuem também. No entanto, o efeito aleatório das métricas de qualidade,  $\overline{\text{PSNR}}$  e  $\overline{\text{MSE}}$ , geram um comportamento randômico no gráfico. Dessa forma, um ajuste adequado do contraste em uma imagem de ressonância pode levar a melhores resultados de classificação e ser utilizado como técnica de enhancement.

### 5.5.6 Discussão

As análises dos efeitos de degradação permitem identificar quais deles têm algum tipo de influência no uso de modelos automáticos de classificação. Dentre os artefatos com influência, pode-se verificar quais deles têm efeitos mais significativos. Este estudo, mostrou que o ruído Gaussiano, *blurring*, *ringing* e contraste impactam consideravelmente nos resultados de acurácia de classificação em imagens de ressonância magnética. Com relação ao efeito

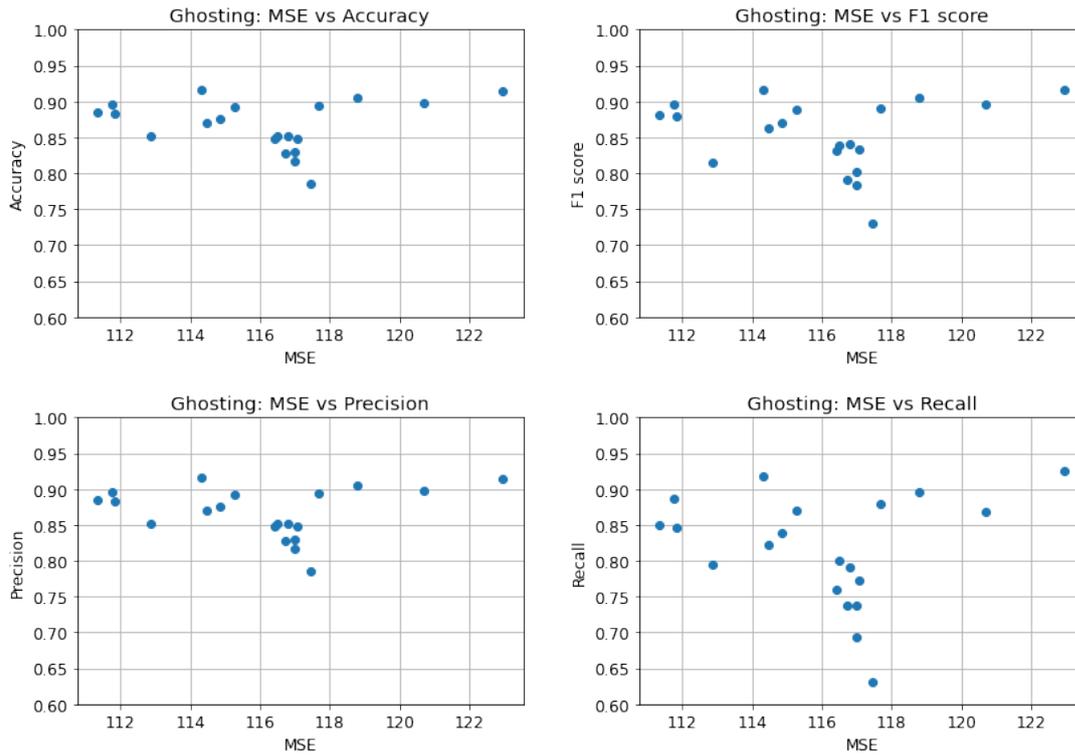


Figura 5.12: Resultados de classificação (acurácia, precisão, recall e F1 score) de imagens com ghosting com relação ao valor médio da MSE de cada nível de degradação.

Nível de Ghost	PSNR	MSE	Acurácia	Precisão	Recall	F1 score
1	27.56	117.449	0.785	0.933	0.63	0.729
2	27.764	114.309	0.915	0.916	0.917	0.916
3	27.394	122.959	0.913	0.909	0.924	0.915
4	27.458	120.691	0.897	0.926	0.868	0.895
5	27.506	118.786	0.904	0.917	0.895	0.904
6	27.538	117.693	0.893	0.911	0.879	0.89
7	27.634	115.274	0.891	0.917	0.869	0.887
8	27.767	111.762	0.895	0.907	0.886	0.894
9	27.769	111.334	0.884	0.918	0.849	0.88
10	27.749	111.84	0.882	0.92	0.845	0.878
11	27.716	112.865	0.851	0.911	0.795	0.814
12	27.638	114.834	0.875	0.912	0.838	0.87
13	27.653	114.475	0.87	0.918	0.821	0.862
14	27.568	116.436	0.847	0.929	0.759	0.83
15	27.571	116.476	0.852	0.905	0.8	0.837
16	27.553	117.008	0.828	0.915	0.737	0.802
17	27.561	116.999	0.816	0.933	0.693	0.782
18	27.558	117.071	0.848	0.918	0.772	0.832
19	27.577	116.719	0.827	0.914	0.737	0.79
20	27.574	116.801	0.852	0.908	0.79	0.84

Tabela 5.7: Resultados de classificação com relação ao valor médio da MSE de cada nível de degradação com *ghosting*.

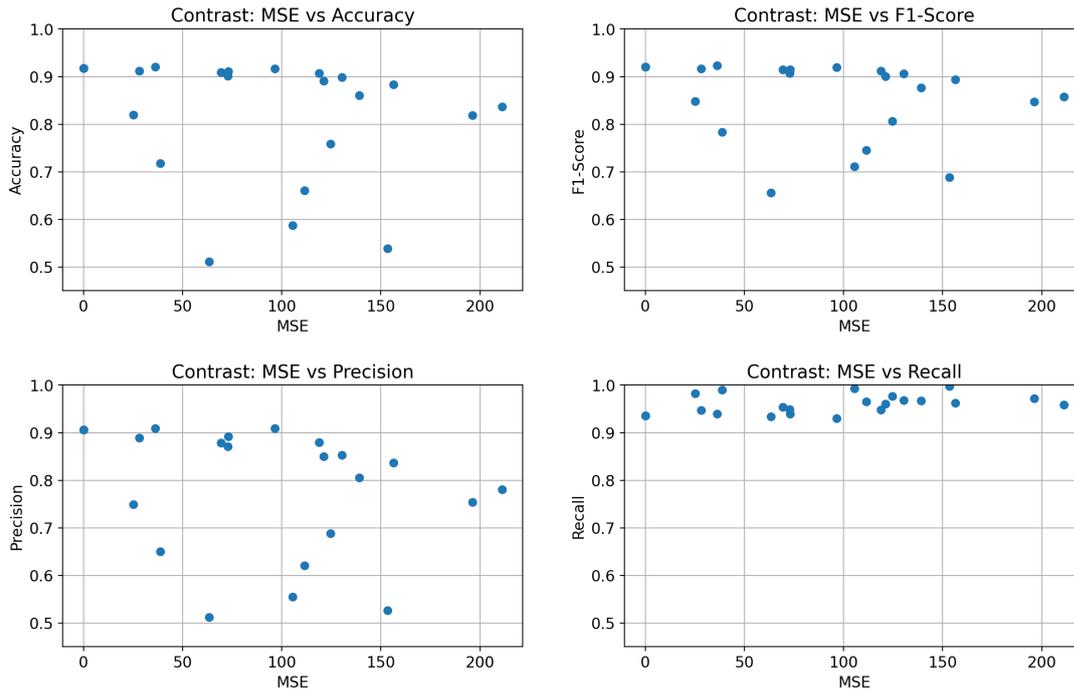


Figura 5.13: Resultados de classificação (acurácia, precisão, recall e F1 score) de imagens com contraste com relação ao valor médio da MSE de cada nível de degradação.

Nível de contraste	PSNR	MSE	Acurácia	Precisão	Recall	F1 score
1	28.292	96.548	0.916	0.909	0.929	0.919
2	32.592	36.316	0.920	0.908	0.939	0.923
3	33.801	28.097	0.912	0.889	0.946	0.916
4	29.524	72.971	0.910	0.891	0.939	0.914
5	27.389	118.869	0.907	0.879	0.948	0.912
6	29.738	69.381	0.909	0.878	0.954	0.914
7	29.524	72.784	0.901	0.870	0.949	0.907
8	26.982	130.422	0.898	0.852	0.968	0.906
9	26.191	156.447	0.883	0.836	0.962	0.893
10	27.304	121.065	0.891	0.850	0.960	0.900
11	26.703	139.011	0.860	0.804	0.966	0.876
12	24.891	211.018	0.836	0.780	0.958	0.857
13	34.255	25.255	0.819	0.749	0.981	0.848
14	25.209	196.092	0.818	0.754	0.971	0.847
15	32.304	38.697	0.718	0.650	0.990	0.783
16	27.177	124.607	0.758	0.688	0.976	0.806
17	27.664	111.384	0.660	0.620	0.965	0.745
18	27.904	105.394	0.587	0.555	0.993	0.711
19	26.277	153.304	0.538	0.526	0.997	0.688
20	30.126	63.275	0.511	0.511	0.933	0.655

Tabela 5.8: Resultados de classificação com relação ao valor médio da MSE de cada nível de degradação com Contraste.

do *ghosting*, não foi possível quantificar o impacto na classificação. Ademais, para *ghosting* e para o contraste, as métricas de qualidade  $\overline{\text{PSNR}}$  e  $\overline{\text{MSE}}$  não refletem exatamente o que é

observado visualmente. À medida que esses dois últimos efeitos se tornam mais severos, as métricas de  $\overline{\text{PSNR}}$  e  $\overline{\text{MSE}}$  não demonstram previsibilidade de crescimento ou decréscimo.

As análises também permitem concluir que uma técnica diferente de *enhancement* deveria ser aplicada a cada tipo de artefato, devido à natureza distinta entre os artefatos testados. Ainda em um modelo real, haveria a necessidade de outro modelo automático para identificação dos tipos de artefatos presentes em uma imagem, antes da aplicação de técnicas de melhoria da qualidade.

Além dos tópicos mencionados, a evolução deste estudo, ensejaria a verificação de métricas de qualidade sem referência em MRI.

# Capítulo 6

## Conclusão

Os estudos permitiram concluir que a partir de características primitivas como cor, brilho, forma, contraste etc., é possível extrair regiões tumorais de imagens 2D de ressonância magnética. Ou seja, algoritmos *bottom-up* capturam corretamente as diferenças histológicas entre tecidos cerebrais saudáveis e regiões de tumor.

Nos mapas de saliência gerados a partir de imagens que continham tumor, foi observado qualitativamente que as regiões de maior brilho na maior parte dos casos, correspondiam às regiões tumorais. A análise quantitativa mostrou que com as seis regiões mais salientes foram obtidas taxas de acerto de 98.02%, 86.2%, 85.42% e 90.42%, com os mapas FES-GBVS, GBVS, COVSAL, ITTI, e FES, respectivamente. Em geral, em todos os cenários com diferentes valores de  $k$  (número de clusters), observou-se um padrão em que com  $\lfloor \frac{k}{2} \rfloor$  regiões era possível se obter acurácias acima de 90% na detecção total ou parcial das regiões tumorais.

A seleção das regiões com maior brilho no mapa foi feita por meio do *k-means*. No entanto, este estudo deixa em aberto a possibilidade de uso de outros métodos de seleção de regiões, como por exemplo a combinação de filtros digitais para encontrar picos. De maneira análoga, o estudo propõe uma combinação em série de extração de ROI por meio de mapas de saliência e classificação utilizando-se redes neurais. Dessa forma, tanto os algoritmos de mapas de saliência quanto as redes neurais poderão ser alterados.

Na classificação final quanto à existência ou não de tumor, em todos os modelos de redes neurais, houve ganhos em termos de acurácia e de *F1 score* utilizando o *framework* proposto. Foi alcançado um resultado global de 93.19% de acerto com o modelo *Alexnet*. Esses resultados demonstram que as regiões pré-selecionadas pelos mapas de saliência eliminam regiões de uma imagem de ressonância não relevantes e/ou discriminativas para o processo de classificação. Além disso, eles permitem que a rede neural, por meio das operações de convolução, capture detalhes mais específicos do tumor.

O treino das redes neurais fez o uso da técnica de *transfer-learning*, em que os pesos utilizados nos neurônios foram importados a partir de outro modelo treinado com *dataset*

*Imagenet*. Apenas a partir da última camada convolucional é que os pesos foram treinados a partir das imagens do BRASTS2020. Esse cenário buscou tornar os contextos de simulações o mais semelhante possível e evitar que os diferentes *hyper-parâmetros* otimizados de um modelo em relação a outro pudessem criar algum tipo de viés na análise comparativa entre esses modelos. No entanto, os resultados poderiam ser melhores com o treino dos modelos em todas as camadas convolucionais. Essa análise, entretanto, ficará para estudos posteriores.

No entanto, em casos onde as regiões tumorais são muito pequenas, quase imperceptíveis aos olhos humanos, os mapas de saliência falham em adicionar essas regiões entre as mais salientes e, conseqüentemente, a classificação final realizada pela CNN também falhará. A análise quantitativa validou que em áreas tumorais menores que 0.3% das regiões cerebrais os mapas de saliência não destacam essas regiões.

Este estudo contribuiu para a análise da influência de artefatos em *scans* MR e para identificar o impacto desses artefatos nas métricas de classificação em uma rede neural. Foi demonstrado que determinados artefatos afetam significativamente a acurácia de modelos automáticos baseados em redes neurais.

Dentre as cinco formas de degradação testadas, o ruído Gaussiano, o *blurring*, o *ringing* e o contraste têm efeitos severos na acurácia das classificações. Outras formas de degradação como *ghosting*, por exemplo, gerou um perfil aleatório nas métricas de classificação à medida que os níveis de degradação se tornam mais acentuados.

Essa análise mostra um problema típico de imagens de diagnóstico e permite o entendimento sobre quais artefatos são mais significativos na degradação de precisão de modelos de classificação automática. Além disso, ela auxilia na priorização de técnicas de *enhancement* a serem aplicadas para melhoria da qualidade das imagens.

Em trabalhos futuros, pretende-se investigar modelos *top-down* que atualmente têm alcançado o estado da arte entre os métodos de predição de saliência. Nesses modelos, o conhecimento prévio do observador acerca dos estímulos visuais direciona o olhar para regiões específicas e esse viés se reflete nos modelos existentes. Muitos desses modelos *top-down* são baseados em redes neurais e são treinados a partir do registro de observação de especialistas com relação às imagens de estímulo.

# Referências Bibliográficas

- [vie ] Viewpoint eyetracker® by arrington research, inc. Software ViewPoint DataAnalysis.
- [Alpaydm 2009] Alpaydm, E. (2009). *Introduction to Machine Learning*. The MIT Press, 2nd edition.
- [Arthur and Vassilvitskii 2007] Arthur, D. and Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, page 1027–1035, USA. Society for Industrial and Applied Mathematics.
- [Bernstein and Berger 2014] Bernstein, M. and Berger, M. S. (2014). *Neuro-Oncology: The Essentials*. Thieme, 3rd edition.
- [Bradski 2000] Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- [Budrys et al. 2018] Budrys, T., Veikutis, V., Lukosevicius, S., Gleizniene, R., Monastyreckiene, E., and Kulakiene, I. (2018). Artifacts in magnetic resonance imaging: how it can really affect diagnostic image quality and confuse clinical diagnosis? *Journal of Vibroengineering*, 20(2):1202–1213.
- [de C. O. et al. 2022] de C. O., P. H., Farias, M. C., Ferreira, D. S., Krylov, A. S., and Ding, Y. (2022). Using a saliency-driven convolutional neural network framework for brain tumor detection. *6th International Conference on Medical and Health Informatics, Japan*.
- [Deng et al. 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.
- [der Walt et al. 2014] der Walt, S. V., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T. (2014). scikit-image: image processing in python. *PeerJ*, 2:e453.
- [Elazary and Itti 2008] Elazary, L. and Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, 8(3):3.

- [Erdem and Erdem 2013] Erdem, E. and Erdem, A. (2013). Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of Vision*, 13(4):11–11.
- [et al 2015] et al, B. H. M. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024.
- [Farias et al. 2022] Farias, M. C., de Castro Oliveira, P. H., Lopes, G. D. S., Miosso, C. J., and Lima, J. A. (2022). The influence of magnetic resonance imaging artifacts on cnn-based brain cancer detection algorithms. *Springer Nature*.
- [fast.ai ] fast.ai. fast.ai making neural nets uncool again. <https://www.fast.ai/>. Acesso em: 10-10-2021.
- [Fei-Fei et al. 2007] Fei-Fei, L., Fergus, R., and Perona, P. (2007). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1):59–70.
- [Ferreira et al. 2019a] Ferreira, D. S., Ramalho, G. L., Torres, D., Tobias, A. H., Rezende, M. T., Medeiros, F. N., Bianchi, A. G., Carneiro, C. M., and Ushizima, D. M. (2019a). Saliency-driven system models for cell analysis with deep learning. *Computer Methods and Programs in Biomedicine*, 182:105053.
- [Ferreira et al. 2019b] Ferreira, D. S., Ramalho, G. L. B., Medeiros, F. N. S., Bianchi, A. G. C., Carneiro, C. M., and Ushizima, D. M. (2019b). Saliency-driven system with deep learning for cell image classification. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE.
- [Franconeri et al. 2005] Franconeri, S. L., Hollingworth, A., and Simons, D. J. (2005). Do new objects capture attention? *Psychological Science*, 16(4):275–281.
- [Gonzalez and Woods 2018] Gonzalez, R. C. and Woods, R. E. (2018). *Digital Image Processing*. Pearson Education.
- [Goodfellow et al. 2017] Goodfellow, I., Bengio, J., and Courville, A. (2017). *Deep Learning*. MIT Press Ltd.
- [Griffin et al. 2022] Griffin, G., Holub, A., and Perona, P. (2022). Caltech 256.
- [Harel et al. 2006] Harel, J., Koch, C., and Perona, P. (2006). Graph-based visual saliency. In Schölkopf, B., Platt, J., and Hoffman, T., editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press.
- [He et al. ] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition.
- [INCA ] INCA. Instituto nacional de câncer (inca). <https://www.inca.gov.br/>. Acesso em: 13-05-2022.

- [Itti and Koch 2001] Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203.
- [Jampani et al. 2012] Jampani, V., Ujjwal, Sivaswamy, J., and Vaidya, V. (2012). Assessment of computational visual attention models on medical images. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing - ICVGIP '12*. ACM Press.
- [Jonides and Yantis 1988] Jonides, J. and Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, 43(4):346–354.
- [Kesari and Stiles 2006] Kesari, S. and Stiles, C. D. (2006). The bad seed: PDGF receptors link adult neural progenitors to glioma stem cells. *Neuron*, 51(2):151–153.
- [Krizhevsky et al. 2017] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90.
- [LeCun et al. ] LeCun, Y., Huang, F. J., and Bottou, L. Learning methods for generic object recognition with invariance to pose and lighting. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*. IEEE.
- [Liming Zhang 2013] Liming Zhang, W. L. (2013). *Selective Visual Attention: Computational Models and Applications*. IEEE COMPUTER SOC PR.
- [Lloyd 1982] Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137.
- [MIT ] MIT. Mit saliency benchmark. <http://saliency.mit.edu/home.html>. Acesso em: 04-03-2021.
- [Nwe et al. 2020] Nwe, T. L., Min, O. Z., Gopalakrishnan, S., Lin, D., Prasad, S., Dong, S., Li, Y., and Pahwa, R. S. (2020). Improving 3d brain tumor segmentation with predict-refine mechanism using saliency and feature maps. In *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE.
- [of Physicists in Medicine 2010] of Physicists in Medicine, A. A. A. (2010). Acceptance testing and quality assurance procedures for magnetic resonance imaging facilities. <https://aapm.org/pubs/reports/detail.asp?docid=101>. Acesso em: 10-06-2022.
- [Paszke et al. 2017] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Nips-w. In *Automatic differentiation in PyTorch*.

- [Pereira et al. 2016] Pereira, S., Pinto, A., Alves, V., and Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, 35(5):1240–1251.
- [Pérez-García et al. 2021] Pérez-García, F., Sparks, R., and Ourselin, S. (2021). TorchIO: A python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 208:106236.
- [Polec et al. 2017] Polec, J., Vargic, R., Csoka, F., Smolejova, E., Heretik, A., Bielikova, M., Svrcek, M., and Moro, R. (2017). Detection of schizophrenia spectrum disorders using saliency maps. In *2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT)*. IEEE.
- [Russakovsky et al. 2015] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252.
- [Russell et al. 2007] Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2007). LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1-3):157–173.
- [Simonyan and Zisserman 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.
- [Szeliski 2022] Szeliski, R. (2022). *Computer Vision: Algorithms and Applications*. Springer, 2nd edition.
- [Tavakoli et al. 2011] Tavakoli, H. R., Rahtu, E., and Heikkilä, J. (2011). Fast and efficient saliency detection using sparse sampling and kernel density estimation. In *Image Analysis*, pages 666–675. Springer Berlin Heidelberg.
- [Thomson and Forman 2009] Thomson, C. S. and Forman, D. (2009). Cancer survival in england and the influence of early diagnosis: what can we learn from recent EURO CARE results? *British Journal of Cancer*, 101(S2):S102–S109.
- [Treisman and Gelade 1980] Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136.