



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

ULRICH KAUÊ MENDES ALENCAR DA SILVA

CLASSIFICAÇÃO DE TUMOR CEREBRAL: um estudo comparativo entre Rede Neural Convolucional e Rede Neural Convolucional com mecanismo de atenção

DM 24/2024

BELÉM
2024

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**“CLASSIFICAÇÃO DE TUMORES CEREBRAIS: UM ESTUDO COMPARATIVO ENTRE
REDE NEURAL CONVOLUCIONAL E REDE NEURAL CONVOLUCIONAL COM
MECANISMO DE ATENÇÃO”**

AUTOR: ULRICH KAUÊ MENDES ALENCAR DA SILVA

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO
JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA
ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: 30/09/2024

BANCA EXAMINADORA:

Prof.^a Dr.^a Adriana Rosa Garcez Castro
(Orientadora – PPGEE/ITEC/UFPA)

Prof.^a Dr.^a Jasmine Priscyla Leite de Araújo
(Avaliadora Interna – PPGEE/ITEC/UFPA)

Prof. Dr. Orlando Shigueo Ohashi Júnior
(Avaliador Externo – UFPA)

VISTO:

Prof. Dr. Diego Lisboa Cardoso
(Coordenador do PPGEE/ITEC/UFPA)

ULRICH KAUE MENDES ALENCAR DA SILVA

CLASSIFICAÇÃO DE TUMOR CEREBRAL: um estudo comparativo entre Rede Neural Convolucional e Rede Neural Convolucional com mecanismo de atenção

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, do Instituto de Tecnologia da Universidade de Belém, da Universidade Federal do Pará, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de computação aplicada.

Orientador(a): Dr(a). Adriana Rosa Garcez Castro

DM 24/2024

BELÉM
2024

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

- M538c Mendes Alencar da Silva, Ulrich Kauê.
 CLASSIFICAÇÃO DE TUMOR CEREBRAL: um estudo
 comparativo entre Rede Neural Convolucional e Rede Neural
 Convolucional com mecanismo de atenção / Ulrich Kauê Mendes
 Alencar da Silva. — 2024.
 XIV, 68 f. : il. color.
- Orientador(a): Profª. Dra. Adriana Rosa Garcez Castro
 Dissertação (Mestrado) - Universidade Federal do Pará,
 Instituto de Tecnologia, Programa de Pós-Graduação em
 Engenharia Elétrica, Belém, 2024.
1. Tumores cerebrais. 2. imagem de ressonância
 magnética. 3. redes neurais convolucionais. 4. mecanismo de
 atenção. I. Título.

CDD 006.3

Dedico este trabalho ao meu falecido avô Antônio Moreira, que apoiou minha jornada com orgulho e que, infelizmente não pode estar presente para este momento.

AGRADECIMENTOS

Aos meus pais Marcelo e Cinthia por apoiarem meu desenvolvimento acadêmico desde a graduação em engenharia da computação, investindo tudo possível para manter a motivação e procura pelo meu auge. Em especial meu pai buscou me aconselhar e guiar em momentos de dificuldades e a minha mãe que buscou nunca me deixar esquecer que acima de estudante, engenheiro e filho ainda sou um ser humano que deve ter a humildade de saber que dificuldade existem para todos.

Ao meu irmão Yohan e minha namorada Pamella, por estarem sempre presentes nos momentos de felicidade e de dificuldade, disponibilizando-se para dar todo suporte cabível, seja no processo do desenvolvimento deste trabalho ou como forma de transpassar as adversidades da caminhada.

A minha orientadora Adriana Castro, que me deu esta oportunidade e acreditou novamente em meu potencial confiando em minha criatividade e perspicácia para desenvolver este trabalho de mestrado, trazendo ideias e se disponibilizando para auxiliar na tarefa de atingir o melhor resultado possível.

E pôr fim aos meus avós, em especial meu avô Antônio Moreira, falecido em junho de 2023, que sempre demonstrou enorme orgulho e felicidade transbordante pelos meus estudos e conquistas, deixo meu eterno agradecimento e o compromisso de sempre honrar sua memória.

“Machines as simple as thermostats can be said to have beliefs, and having beliefs seems to be a characteristic of most machines capable of problem solving performance.” (MCCARTHY, 1979, p. 3)

RESUMO

Os tumores cerebrais são doenças neurológicas com elevado potencial de impacto na vida dos indivíduos acometidos, requerendo um diagnóstico rápido e preciso por meio de exames complementares de imagem, como a ressonância magnética, que é considerada padrão-ouro nesse processo. Considerando a necessidade de um diagnóstico mais rápido, sistemas de classificação baseados em Aprendizado de Máquina vêm sendo desenvolvidos e dentro deste contexto essa dissertação, tem como objetivo apresentar um estudo comparativo entre uma Rede Neural Convolutacional (CNN) e uma CNN com mecanismo de atenção, desenvolvidas para a classificação de tumores cerebrais a partir de imagens de ressonância magnética. O estudo comparativo visa identificar o impacto do mecanismo de atenção no desempenho da CNN para classificação de tumores. Para desenvolvimento e avaliação dos modelos propostos foi utilizada uma base de dados pública, coletada do *website* Kaggle, e disponibilizada por Masoud Nickparvar, sendo esta composta por 7023 imagens de ressonâncias magnéticas cerebrais, segmentadas em quatro classes: glioma, meningioma, sem tumor e pituitário. Como resultado, a partir das métricas de desempenho obtidas, considerando a base de imagens usadas para teste em ambas as CNNs, observou-se uma melhora no desempenho da CNN após a introdução do mecanismo de atenção, onde a rede com esse mecanismo apresentou um aumento de 1.98% na métrica acurácia, 2.07% na métrica precisão, 2.18% na métrica sensibilidade e 1.72% na métrica F1-score em relação a CNN sem mecanismo de atenção, podendo ainda ser destacado os resultados obtidos em particular para a classe de tumor meningioma, visto que a CNN sem o mecanismo de atenção apresentou dificuldades para classificação desta classe e, após a integração do mecanismo de atenção o modelo obteve um aumento de acurácia de 6.54% para esta classe.

Palavras-chave: tumores cerebrais; imagem de ressonância magnética; redes neurais convolucionais; mecanismo de atenção.

ABSTRACT

Brain tumors are neurological diseases with a high potential impact on the lives of affected individuals, requiring a rapid and accurate diagnosis through complementary imaging tests, such as magnetic resonance imaging, which is considered the gold standard in this process. Considering the need for faster diagnosis, classification systems based on Machine Learning have been developed and within this context, this dissertation aims to present a comparative study between a Convolutional Neural Network (CNN) and a CNN with an attention mechanism, developed for the classification of brain tumors from magnetic resonance images. The comparative study aims to identify the impact of the attention mechanism on the performance of the CNN for tumor classification. For the development and evaluation of the proposed models, a public database was used, collected from the Kaggle website and made available by Masoud Nickparvar, which is composed of 7023 brain magnetic resonance images, segmented into four classes: glioma, meningioma, no tumor and pituitary. As a result, from the performance metrics obtained, considering the image base used for testing in both CNNs, an improvement in the CNN performance was observed after the introduction of the attention mechanism, where the network with this mechanism presented an increase of 1.98% in the accuracy metric, 2.07% in the precision metric, 2.18% in the sensitivity metric and 1.72% in the F1-score metric in relation to the CNN without the attention mechanism. It is also possible to highlight the results obtained in particular for the meningioma tumor class, since the CNN without the attention mechanism presented difficulties in classifying this class and, after the integration of the attention mechanism, the model obtained an accuracy increase of 6.54% for this class.

Key words: brain tumors; magnetic resonance images; convolutional neural networks; attention mechanism.

LISTA DE ILUSTRAÇÕES

Figura 1 – Cérebro dividido em seus lóbulos	21
Figura 2 – Ressonância magnética de tumor glioma	21
Figura 3 – Ressonância magnética de tumor meningioma	22
Figura 4 – Ressonância magnética de tumor hipofisário	22
Figura 5 – Máquina de ressonância magnética	24
Figura 6 – Exemplo de imagem de RM cerebral saudável	24
Figura 7 – Estrutura genérica de uma CNN	26
Figura 8 – Camada de convolução – Aplicação de um kernel em uma imagem	27
Figura 9 – Exemplo da operação de max-pooling	29
Figura 10 – Arquitetura do mecanismo <i>attention</i>	36
Figura 11 – Exemplo de Score-CAM para imagem de RM do conjunto de dados	38
Figura 12 – Fluxo de etapas do procedimento metodológico	41
Figura 13 – Imagens de RM de Glioma (1), Meningioma (2), Sem tumor (3) e Pituitário (4)	42
Figura 14 – Estrutura da CNN	44
Figura 15 – Estrutura da CNN com mecanismo de atenção	47
Figura 16 – Estrutura da camada de atenção	49
Figura 17 – Matriz de confusão da CNN	52
Figura 18 – Quatro exemplos de Score-CAM para CNN sem mecanismo de atenção	53
Figura 19 – Matriz de confusão da CNN com atenção	55
Figura 20 – Score-CAM da CNN para cada classe de tumores antes (a) e depois (b) do mecanismo de atenção	55
Figura 21 – Score-CAM para cada classe de tumores da CNN sem mecanismo de Atenção (a) e com mecanismo de atenção (b)	58

LISTA DE TABELAS

Tabela 1 – Resumo da divisão da base de imagens para treino e teste	42
Tabela 2 – Resumo da divisão da base de imagens para treino e validação	43
Tabela 3 – Parâmetros de treinamento e propriedades da CNN	45
Tabela 4 – Parâmetros de treinamento e propriedades da CNN com mecanismo de atenção	48
Tabela 5 – Métricas globais resultantes da CNN	51
Tabela 6 – Métricas resultantes da CNN para cada classe	51
Tabela 7 – Métricas globais da CNN com atenção	54
Tabela 8 – Métricas resultantes da CNN com atenção para cada classe	54
Tabela 9 – Comparação de métricas entre os modelos de redes convolucionais com e sem atenção	57
Tabela 10 – Comparação dos modelos com trabalhos da literatura atual	59

LISTA DE ABREVIATURAS E SIGLAS

ADAM	Adaptive moment estimation
AFF-VGG19	Attention Feature Fusion VGG19
CAM	Class activation map
CBAM	Convolutional block attention module
CNN	Redes neurais convolucionais
DCGAN	Deep Convolutional Generative Adversarial Networks
DCTN	Dual convolutional tumor network
FN	Falso negativo
FP	Falso positivo
GAM	Gradient awareness minimization
Grad-CAM	Gradient-weighted class activation mapping
IA	Inteligência Artificial
KNN	K-nearest neighbors
LSTM	Long-shot term memory
MAE	Mean Absolute error
MSE	Mean squared error
PET	Tomografia por emissão de pósitrons
PLN	Processamento de linguagem natural
PRCnet	Parallel Residual Convolutional Network
PRDCNN	Parallel deep convolutional neural network
ReLU	Rectified linear unit
RM	Ressonância magnética
RNN	Recurrent neural network
RMSprop	Root mean square propagation
SE	Squeeze and Excitation
SENet	Squeeze-and-Excitation Network
SGD	Stochastic gradient descent
SpCaNet	Spinal convolution attention network
SVM	Support vector machine
ViT	Visual Transformer
VN	Verdadeiro negativo
VP	Verdadeiro positivo

YOLOv7 You Only Look Once version 7

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivo geral e específicos	18
1.2	Estrutura do trabalho.....	18
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	Tumores cerebrais	20
2.1.1	Imagem por ressonância magnética	24
2.2	Redes neurais convolucionais.....	25
2.2.1	Camada convolucional	26
2.2.2	Camada de <i>pooling</i>	28
2.2.3	Camada de <i>dropout</i>	30
2.2.4	Camadas <i>flatten</i> e totalmente conectadas	30
2.2.5	Treinamento das redes neurais convolucionais.....	30
2.3	Mecanismo de atenção.....	32
2.3.1	Tipos de mecanismo atenção	33
2.3.1.1	<i>Hard attention</i>	33
2.3.1.2	<i>Soft attention</i>	34
2.3.1.3	<i>Self attention</i>	35
2.4	Mapa de ativação de classes	37
2.5	Estado da arte de classificação de tumores cerebrais baseados em CNN	38
3	METODOLOGIA.....	41
3.1	Mapa metodológico	41
3.2	Conjunto de dados de imagens de RM – Tratamento dos dados	41
3.3	Treinamento da CNN e CNN com mecanismo de atenção para classificação de tumores cerebrais	43
3.3.1	Definição e treinamento da CNN sem mecanismo de atenção	43
3.3.2	Definição e treinamento da CNN com mecanismo de atenção.....	46
3.4	Comparação de resultados dos modelos classificadores.....	49
4	RESULTADOS	51
4.1	Resultados para CNN sem mecanismo de atenção	51
4.2	Resultados para CNN com módulo de atenção	53
5	DISCUSSÃO DOS RESULTADOS	57
5.1	Comparação de resultados entre a CNN sem e com mecanismo de atenção.....	57

5.2	Comparação dos modelos com resultados da literatura.....	58
6	CONCLUSÃO	60
	REFERÊNCIAS.....	61

1 INTRODUÇÃO

A complexidade da anatomia do cérebro é alta e suas distintas partes são responsáveis por diferentes funções nervosas, sendo que qualquer parte do cérebro pode ser acometido com um tumor, totalizando mais de 120 tipos distintos, que são diretamente dependentes do tecido ou região onde se originam. Os tumores cerebrais, cancerígenos ou não, podem ser considerados um risco à vida humana, seja por causarem aumentos de pressão intracraniana (levando a bloqueio do fluxo dos fluídos cerebrais) ou pela possibilidade de disseminação para demais áreas do cérebro. Ademais, quando uma área específica do cérebro é afetada pelo tumor, o paciente pode apresentar perda parcial ou total das funções neurológicas associadas à esta determinada região.

Dadas as dificuldades de diferenciar tumores cerebrais e outras doenças de menor risco a vida humana, diversos testes e procedimentos podem ser realizados para sua detecção, como por exemplo exames neurológicos, tomografia computadorizada do crânio, tomografia por emissão de pósitrons do cérebro (PET) e imagens de ressonância magnética. Dentre estes, a ressonância magnética é o exame complementar padrão-ouro para o diagnóstico de tumores cerebrais e utiliza ímãs de alta potência para gerar figuras das estruturas do cérebro, e é comumente utilizado pela sua capacidade de apresentar o cérebro mais claramente que os demais exames de imagem.

Em vista da alta dificuldade no processo de identificação e classificação dos tumores cerebrais e a necessidade de um diagnóstico rápido, para evitar possíveis sequelas e minimizar o alto risco de vida atualmente, trabalhos vêm sendo apresentados na literatura com propostas de modelos computacionais baseados em Aprendizado Profundo para auxílio ao especialista no processo de classificação de tumores cerebrais através imagens, sendo que as Redes Neurais Convolucionais têm se destacado para este tipo de aplicação devido principalmente a sua capacidade de processar imagens digitais e extrair delas informações significativas para o processo de classificação. Em Das *et al.* [1] os autores propuseram um algoritmo combinando redes neurais convolucionais com aprendizado por transferência, chamado EfficientNet-B6, para classificação de tumores cerebrais baseado em imagens de ressonância magnética. O modelo utilizado nesse trabalho atingiu resultados com mais de 90% de sensibilidade, precisão, F1-score e acurácia de treino, teste e validação para o conjunto de dados utilizado. Assim, sendo indicado como significativamente melhor que modelos da literatura e preciso para categorização de tumores cerebrais. Em Rahman e Islam [2] os autores, buscando extrair características locais e globais das imagens de ressonância magnética do cérebro, levantaram

uma topologia denominada *parallel deep convolutional neural network* (PDCNN) que é composta por duas redes convolucionais simultâneas com diferentes tamanhos de janelas de processamento. O PDCNN foi avaliado com o uso de 3 conjuntos distintos de dados atingindo valores de acurácia maiores que 97% em cada e devido a sua capacidade de extrair características locais e globais das imagens foi considerado tanto preciso quanto eficiente na classificação de tumores no cérebro. Em Özkaraca *et al.* [3] os autores utilizaram conjuntos de dados de código aberto advindos do *website* Kaggle para treinar um modelo modificado de rede neural que conseguisse contornar as deficiências dos modelos VGG16, DenseNet e de redes convolucionais básicas na classificação de tumores cerebrais em imagens de ressonância magnética. Ao final do estudo, observou-se um aumento no desempenho de classificação e no tempo de processamento quando comparados a outros modelos da literatura. Em Farhan *et al.* [4], os autores desenvolveram um modelo intitulado *Parallel Residual Convolutional Network* (PRCnet) para análise de imagens de ressonância magnética do cérebro visando classificar tumores cerebrais. Esse modelo aglomerou diversas técnicas, como: camadas paralelas, filtros de diferentes tamanhos, conexões entre as camadas e entre outras. Em seguida, realizou-se o treinamento e teste do PRCnet em 2 conjuntos de dados obtendo acurácias de 94.77% para o conjunto A e 97.1% para o conjunto B. Por fim, as métricas alcançadas pelo modelo foram consideravelmente melhores que modelos estado-da-arte selecionados para comparação.

Nos últimos anos modelos computacionais baseados em aprendizagem profunda combinada com mecanismo de atenção têm atraído grande atenção na área de visão computacional, sendo que estes modelos vêm alcançando desempenhos mais avançados em várias áreas de diagnóstico médico baseado em imagens. Os mecanismos de atenção, inicialmente aplicados com sucesso na área de Processamento de Linguagem Natural, se baseiam na ideia de que humanos tendem a concentrar sua atenção em certas partes do espaço visual, para adquirir informações, em vez de ver a imagem como um todo, sendo que quando estes mecanismos são aplicados nas CNNs, podem permitir que a rede tenha atenção seletiva para características específicas extraídas das imagens de entrada, ou para combinações de características, sendo que estes mecanismos podem ser implementados para corresponder a camadas da rede, que irão combinar informações de camadas anteriores, permitindo que camadas mais profundas utilizem essas informações para o processo final de classificação das imagens de entrada.

No que diz respeito ao uso de CNN com mecanismos de atenção para classificação de tumores cerebrais, alguns trabalhos já vêm sendo apresentados na literatura, com estruturas específicas de mecanismos de atenção, apresentando bons resultados, como em Apostolopoulos

et al. [5] onde os autores apresentam a integração de modelos de atenção, redes convolucionais e blocos de fusão de características para o desenvolvimento de uma rede multicaminho robusta objetivando a extração de características de imagens médicas (ressonância magnética) para classificação de tumores cerebrais. Dando continuidade, o treinamento da rede denominada *Attention Feature Fusion VGG19* (AFF-VGG19) ocorreu a partir do uso de um conjunto de dados envolvendo classificação de demência, tumores e distúrbios cerebrais. Por fim, o AFF-VGG19 demonstrou resultados superiores em comparação a redes estado-da-arte, alcançando acurácias de 95.65% na distinção entre doenças de Parkinson e Alzheimer e de 94.97% na classificação de casos de demência. Em Liu e Yang [6] os autores propuseram um modelo melhorado do DenseNet-201 em conjunto com a arquitetura de *channel* attention, o *Squeeze-and-Excitation Network* (SENet). Dessa forma, a arquitetura de atenção é aplicada como um desvio conectando a camada convolucional à camada de transição. Além disso, utilizaram um método de aprendizado não-supervisionado para gerar expansões dos dados de imagens de ressonância magnéticas cerebrais, chamado *Deep Convolution Generative Adversarial Networks* (DCGAN). Por fim, o modelo criado apresentou melhorias na habilidade de generalização e na acurácia de classificação das imagens. Em Bodapati *et al.* [7] os autores extraíram as características locais de imagens médicas (especificamente ressonância magnética) de tumores cerebrais por blocos convolucionais de redes InceptionResNetV2 e Xception, sendo suas saídas vetorizadas usando técnicas de *pooling*. Ademais, um mecanismo de atenção foi proposto para aumento de foco em regiões tumorais que eventualmente auxiliaria na diferenciação dos tipos de tumores presentes no conjunto de dados. Ao final, estudos empíricos, realizados em dois conjuntos de dados, revelaram que a abordagem feita no estudo se mostrou superior na generalização das imagens e simples em questão de número de camadas ao ser comparada com modelos complexos da literatura atual.

Considerando a importância de desenvolvimento de modelos para classificação de tumores cerebrais e que os mecanismos de atenção vêm despertando um interesse significativo na comunidade de pesquisa, visto que, a partir dos resultados já apresentados na literatura, eles vêm trazendo melhorias relevantes no desempenho de arquiteturas de redes neurais convolucionais e, levando em conta ainda que, para aplicação em qualquer problema específico, não se tem ainda uma forma baseada em princípios para a escolha de um mecanismo de atenção específico, que leve a melhorias garantidas para uma determinada aplicação, esta dissertação tem então como objetivo geral realizar um estudo comparativo de aplicação de uma CNN sem mecanismo de atenção e de uma CNN com um mecanismo de atenção específico para classificação de tumores cerebrais, visando determinar o impacto deste mecanismo de

atenção no desempenho da rede neural convolucional baseada em imagens de ressonância magnética do cérebro, podendo o resultado deste estudo trazer indicações em relação a possibilidade de desenvolvimento de futuros modelos com melhor capacidade de classificação de tumores cerebrais, os quais poderão ser utilizados para auxílio ao diagnóstico dessas doenças por profissionais da área da saúde, podendo acarretar em um diagnóstico e início de tratamento precoce, trazendo maior qualidade de vida aos pacientes acometidos.

1.1 Objetivo geral e específicos

Este trabalho tem como objetivo geral apresentar os resultados do estudo comparativo realizado entre uma rede CNN sem e com um mecanismo de atenção específico para classificação de tumores cerebrais baseado em imagens de ressonância magnética, visando identificar o impacto do uso deste tipo mecanismo de atenção no desempenho da CNN.

Para alcançar o objetivo geral foram determinados os seguintes objetivos específicos:

- Selecionar um conjunto de dados público, com imagens de ressonância magnética, com representatividade de casos de tumores cerebrais, que apresentasse qualidade para desenvolvimento das CNNs para estudo comparativo;
- Desenvolver uma CNN para classificação de tumores cerebrais, que obtivesse resultados similares ou superiores a de trabalhos semelhantes da literatura atual;
- Implementar e adicionar um mecanismo de atenção específico à estrutura da CNN anterior, objetivando atingir desempenho similar ou superior ao apresentado na literatura atual (acima de 96% de acurácia);
- Obter métricas de desempenho global e por tipo de tumor para ambas CNNs desenvolvidas, considerando imagens não utilizadas durante fase de desenvolvimento das CNNs.
- Identificar e concluir a respeito do impacto do uso do mecanismo de atenção específico no desempenho da CNN.

1.2 Estrutura do trabalho

A estrutura deste trabalho é composta por mais 5 capítulos, além deste capítulo introdutório. O capítulo 2 apresenta o embasamento teórico necessário para o bom entendimento da proposta deste trabalho, sendo apresentada uma breve fundamentação teórica sobre tumores cerebrais, redes neurais convolucionais, mecanismo de atenção e mapas de ativação de classes, finalizando com o estado da arte sobre classificação de tumores cerebrais baseados em redes neurais convolucionais. O capítulo 3 apresenta a metodologia utilizada para

desenvolvimento dos modelos convolucionais propostos para classificação de tumores cerebrais. O capítulo 4 apresenta os resultados obtidos pelos dois modelos, buscando apresentar o desempenho alcançado por ambas as redes ao classificar tumores cerebrais, utilizando as métricas comparativas selecionadas, além de apresentar as matrizes de confusão e os mapas de ativação de classes para ambos os modelos. O capítulo 5 apresenta a discussão dos resultados, apresentando a comparação analítica entre as métricas, matrizes de confusão e mapa de ativação de classes dos modelos. Por fim, o capítulo 6 apresenta as conclusões obtidas a partir do estudo comparativo realizado.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo será apresentada a fundamentação teórica necessária para o bom entendimento da proposta deste trabalho, sendo que serão apresentados conceitos sobre Tumores cerebrais, Redes Neurais Convolucionais, Mecanismos de atenção e Mapas de ativação de classes. Ao final do capítulo será apresentado o Estado da Arte sobre classificadores de tumores cerebrais baseados em redes neurais convolucionais e redes neurais convolucionais com mecanismo de atenção.

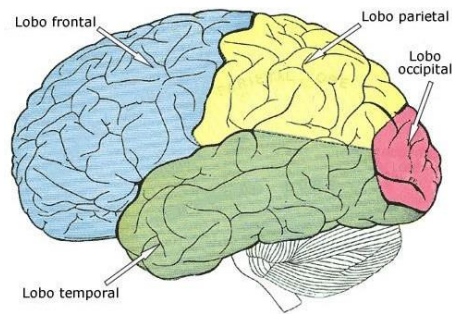
2.1 Tumores Cerebrais

O cérebro é o centro do sistema nervoso do corpo humano e serve para controlar suas funções, como habilidades motoras, visão, fome, entre outros [8]. Ele é composto por matéria cinza e matéria branca, as quais são formadas, respectivamente, pelos corpos celulares e pelos axônios dos neurônios (prolongamentos condutores de impulsos elétricos que transmitem informações para outras células), sendo que a matéria cinza é responsável por processar e interpretar a informação, enquanto a matéria branca transmite a informação para distintas partes do sistema nervoso.

Este órgão é dividido em dois hemisférios por um sulco profundo, chamado de fissura, e esses dois hemisférios comunicam-se através de uma densa camada de nervos denominada de corpo caloso [9]. Cada hemisfério é dividido em 4 lóbulos, como apresentado na Figura 1, sendo eles:

- Lóbulo temporal: Processa informações do paladar, audição e olfato, além de estar presente no armazenamento de memórias.
- Lóbulo occipital: Permite o reconhecimento de imagens através da conexão das imagens vindas dos olhos com imagens armazenadas na memória humana.
- Lóbulo parietal: Auxilia na interpretação de informações sensoriais, processando sabores, texturas e temperatura.
- Lóbulo frontal: Responsável por auxiliar o controle do pensamento, memória de curto prazo e movimento.

Figura 1 – Cérebro dividido em seus lóbulos

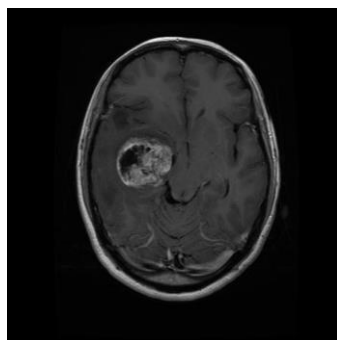


Fonte: [10]

O tumor cerebral é definido como um crescimento anormal de células em algum determinado tecido deste órgão [11] e encontra-se entre as anomalias mais severas que podem causar danos irreparáveis ao organismo humano, dependendo da sua intensidade e localização. Há mais de 120 tipos de tumores cerebrais, os quais afetam diferentes partes do cérebro humano, as quais apresentam funções específicas, ou seja, a região em que o tumor se desenvolve apresenta sua função prejudicada. De forma geral podem ser divididos em dois tipos: os tumores benignos, que são não cancerígenos e tendem a crescer lentamente e os tumores malignos, sendo cancerígenos e tendem a apresentar um rápido crescimento. Alguns dos tipos de tumores cerebrais são [11]:

- Gliomas: Estes tumores advêm das células gliais, responsáveis pela sustentação das células do tecido cerebral. Apesar de poderem ser benignos, a tendência é que esses tumores sejam de caráter maligno. A Figura 2 apresenta um exemplo de imagem de ressonância magnética de cérebro com tumor glioma.

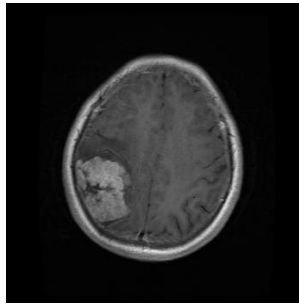
Figura 2 – Ressonância magnética de tumor glioma



Fonte: [12]

- Meningiomas: São tumores que se iniciam na membrana ao redor do cérebro, apesar de comumente serem benignos, podem apresentar casos malignos. A Figura 3 apresenta um exemplo de imagem de ressonância magnética de cérebro com tumor meningioma.

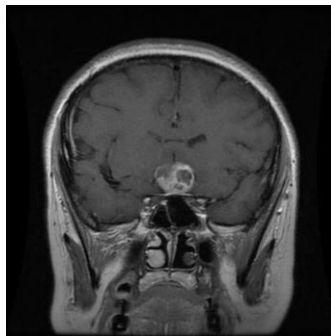
Figura 3 – Ressonância magnética de tumor meningioma



Fonte: [12]

- Tumores hipofisários: Neste caso são tumores que se iniciam em uma pequena glândula próxima a base do cérebro chamada hipófise ou glândula pituitária. Tumores hipofisários são benignos. A Figura 4 apresenta um exemplo de imagem de ressonância magnética de cérebro com tumor hipofisário.

Figura 4 – Ressonância magnética de tumor hipofisário



Fonte: [12]

Os sintomas de tumores cerebrais variam e como agravante podem ser confundidos como advindos de doenças menos graves, dificultando assim o seu diagnóstico. Dentre os diversos sintomas provocados por tumores cerebrais temos como exemplos dos mais comuns [11]:

- Pressão ou dores de cabeça, sendo piores pela manhã;
- Náuseas ou vômitos;
- Problemas de visão, como vista embaçada, visão duplicada ou perda de visão na região periférica;
- Perda de movimento ou sensibilidade em uma perna ou braço;
- Problemas de equilíbrio;
- Sensação de cansaço exacerbada;
- Problemas de fala;
- Mudanças de personalidade ou comportamento;
- Convulsões, especialmente se não houver histórico prévio;

- Problemas de memória;
- Tontura ou vertigem;

Devido à dificuldade de discernir tumores cerebrais de outras doenças, visto que apresentam um compartilhamento de sintomas, existem alguns testes e procedimentos para realizar a investigação e o diagnóstico de possíveis tumores cerebrais, dentre eles [13]:

- Exames neurológicos: Estes exames testam diferentes partes do cérebro com o intuito de verificar seu funcionamento. Nestes exames podem ser testados equilíbrio, reflexo, coordenação, visão, entre outras funções referentes ao cérebro. Por fim, caso uma ou mais dessas funções apresentem mal funcionamento é possível considerar como um indicador da região afetada pelo tumor, porém vale ressaltar que esses exames não realizam diagnóstico, ou seja, não detectam o tumor cerebral, mas auxiliam na busca do possível tumor.

- Tomografia computadorizada: Neste exame são utilizados feixes de radiação (raios-x) em diversos ângulos para produção de imagens transversais do cérebro, seus resultados são facilmente acessíveis, pois apresentam retorno rápido e é um dos primeiros exames feitos, caso haja a suspeita de um tumor cerebral. A tomografia pode detectar problemas ao redor ou no cérebro, porém caso aumente a suspeita de tumor cerebral, existe a necessidade de realizar uma imagem por ressonância magnética.

- Imagem por ressonância magnética: Este exame realiza o uso de ímãs para criar imagens e são geralmente utilizados para diagnosticar tumores cerebrais devido a sua capacidade de gerar imagens mais nítidas quando comparado a outros exames de imagem. Em geral, é injetado um agente de contraste em uma veia antes da realização do exame, pois este corante auxilia na nitidez da imagem e facilita a visualização de tumores menores. Vale salientar que existem tipos específicos de imagem por ressonância magnética, como a espectroscopia de ressonância magnética, utilizada para medição dos níveis de alguns químicos nas células tumorais, auxiliando na definição do tipo de tumor cerebral baseando-se na alta ou baixa concentração de desses químicos nessas células.

- Tomografia por emissão de pósitrons: A tomografia por emissão de pósitrons se diferencia da tomografia computadorizada, pois injeta um composto radioativo em uma veia. Esse composto circula pelo corpo através do sangue e se prende às células tumorais fazendo com que estas células se destaquem na imagem produzida pela máquina de tomografia por emissão de pósitrons. Vale ressaltar que tumores de crescimento lento podem não ser detectados neste exame, por isso ele apresenta maior potencial de ajuda com a detecção de tumores de crescimento rápido.

2.1.1 Imagem por ressonância magnética

Imagem por ressonância magnética (RM), é um exame capaz de produzir imagens detalhadas de inúmeras estruturas internas do corpo humano de forma não invasiva [14], a partir de grandes ímãs e ondas de rádio emitidas por uma máquina em formato cilíndrico, como apresentado na Figura 5. Este exame é capaz de gerar visualizações nítidas de estruturas ósseas, órgãos, músculos, entre outros.

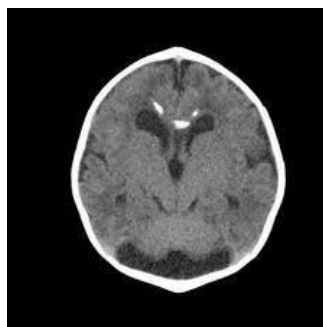
Figura 5 – Máquina de ressonância magnética



Fonte: [15]

A máquina de RM cria um forte campo magnético que obriga os átomos do corpo humano a um alinhamento, as ondas de rádio enviadas pela máquina alteram a posição desses átomos e após o desligamento das ondas os átomos tendem a retomar sua posição original liberando ondas de rádio nesse retorno. Estas ondas liberadas pelos átomos do corpo são processadas por um computador e convertidas em imagens da região analisada, vale salientar que este exame não emite radiação ionizada, ou seja, possibilita sua realização com maior frequência sem causar danos ao paciente e é mais preciso que tomografias computadorizadas para tecidos moles do corpo humano, como cérebro, nervos e músculos. Estes tecidos moles são mais nítidos em imagens de RM, do que em tomografias, sendo que nas imagens geradas do cérebro, como por exemplo da Figura 6, é possível diferenciar as matérias cinza e branca.

Figura 6 – Exemplo de imagem de RM cerebral saudável



Fonte: [12]

2.2 Redes neurais convolucionais

O aprendizado de máquina é uma subárea da Inteligência Artificial (IA) que trabalha com o desenvolvimento de algoritmos que possibilitam determinados modelos aprenderem padrões a partir de uma grande quantidade de dados.

O aprendizado de máquina pode ser dividido em 4 categorias, sendo elas [16, 17]:

- Aprendizado supervisionado: aprendizado que tem como principal característica o uso de conjuntos de dados rotulados para realização do treinamento dos modelos, tendo como objetivo, por exemplo, a previsão de valores ou classificação de dados.

- Aprendizado não-supervisionado: aprendizado que utiliza conjuntos de dados não rotulados durante fase de treinamento do modelo, buscando analisar e agrupar os dados, descobrindo padrões e agrupamentos, sem a intervenção humana.

- Aprendizado semissupervisionado: aprendizado que apresenta características do aprendizado supervisionado e não supervisionado sendo que utiliza um conjunto menor de dados rotulados para treinamento com o objetivo de extrair recursos de um conjunto de dados maior, não rotulado.

- Aprendizado por reforço: aprendizado que se diferencia dos demais por aprender a partir de tentativa e erro, sendo que neste caso resultados positivos são reforçados por meio de “recompensas” enquanto resultados negativos resultam em “prejuízo”.

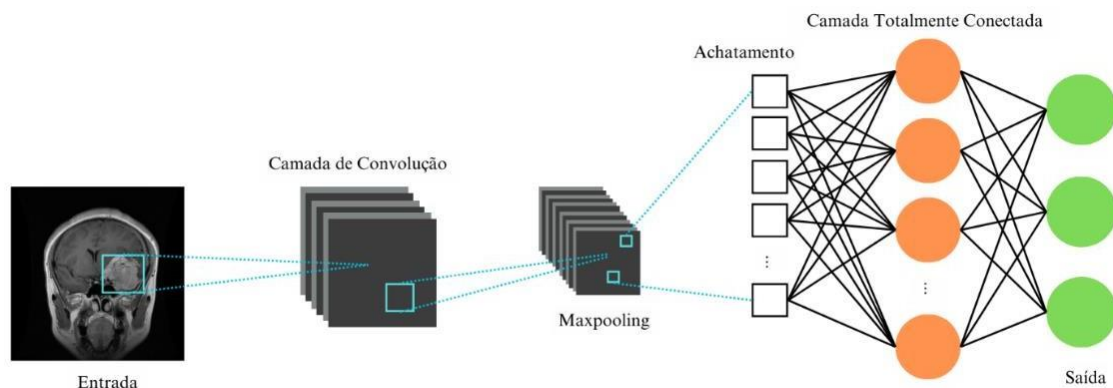
As Redes Neurais Artificiais, que são baseadas em algoritmos de aprendizado de máquina, são modelos computacionais que possibilitam aos computadores aprenderem com dados, utilizando-se de processos que mimetizam o cérebro humano e têm imenso potencial, sendo utilizadas em diversas aplicações [18].

Entre as diversas aplicações das redes neurais, a visão computacional, em particular, vem se destacando atualmente, sendo uma área onde há o enfoque em permitir que os computadores sejam capazes de processar, identificar e entender objetos em imagens e vídeos [19], podendo ser aplicada em inúmeros setores da vida humana, como por exemplo, no setor de manufatura e para o auxílio no diagnóstico de doenças a partir de exames de imagem. No contexto de auxílio ao diagnóstico médico, os modelos de redes neurais podem ser aplicados para análise de exames de imagem, aumentando a velocidade e a precisão dos diagnósticos de doenças, auxiliando os profissionais da área a identificarem anomalias de forma mais rápida, verificando algumas relações nos dados que os humanos não conseguem observar.

As Redes Neurais Convolucionais são um tipo específico de redes neurais que são o estado-da-arte quando se refere ao processamento de imagens no campo da visão

computacional. Essas redes apresentam como diferencial a utilização de camadas convolucionais para filtrar os dados de entrada em busca de detalhes ou informações úteis [20]. A estrutura típica de uma CNN, como apresentado na Figura 7, é composta principalmente por camadas de convolução, de *pooling*, de *dropout*, achatamento (*flatten*) e camadas totalmente conectadas.

Figura 7 – Estrutura genérica de uma CNN



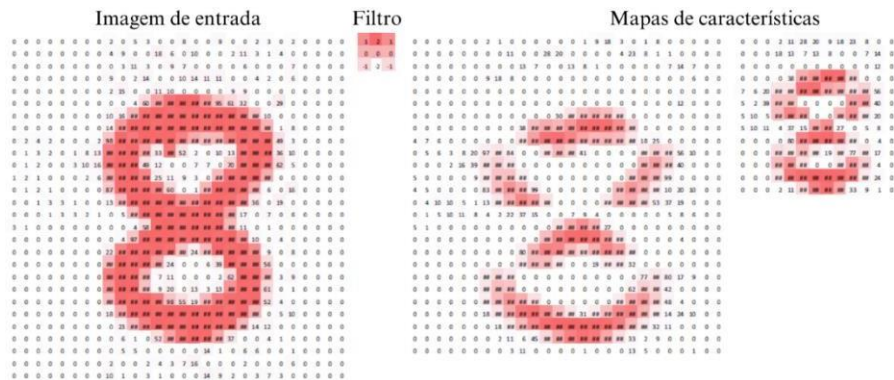
Fonte: Autor (2024)

Nas seções seguintes será apresentado com mais detalhes cada uma das camadas de uma CNN assim como será apresentado o processo de treinamento deste tipo de rede neural.

2.2.1 Camada convolucional

A camada convolucional realiza uma operação chamada convolução, como apresentado na Figura 8, que é definida como uma forma de mistura entre duas funções ou informações, ou seja, entre os dados de entrada e um filtro, chamado *kernel*, gerando um mapa de atributos [21, 22]. Basicamente considera-se um filtro buscando determinadas informações nos dados de entrada para geração do mapa de atributos, extraindo características e informações úteis da imagem, para realizar determinada tarefa, como por exemplo, focar no formato dos objetos ignorando o restante das informações.

Figura 8 – Camada de convolução - Aplicação de um *kernel* em uma imagem



Fonte: Adaptado de [23]

A camada convolucional utiliza alguns componentes para realização da operação de convolução, sendo eles [23, 24]:

- **Filtros (*kernel*):** Apresentam a função de detectar características na imagem. Uma multiplicação elemento-a-elemento entre o *kernel* (matriz) e um segmento dos dados de entrada (matriz) é realizada e a soma dos resultados dessas multiplicações é realizada para obter o valor de saída da convolução entre essas matrizes. Esses passos são repetidos até que o *kernel* tenha percorrido toda a imagem, gerando assim o mapa de característica final da camada convolucional, sendo que vale ressaltar que são esses filtros que carregam os pesos utilizados para realizar uma tarefa determinada, os quais são alterados iterativamente no treinamento da CNNs ao longo de suas épocas em busca do melhor *kernel* para a determinada tarefa e conjunto de dados (imagens).

- **Stride:** É a distância, em *pixels*, que ocorre o movimento do *kernel* sobre a matriz de entrada, com o objetivo de posicionar o campo de recepção dos dados, sendo que quanto maior o *stride* menor é a sobreposição dos filtros, o que culmina em uma diminuição no espaço dimensional da saída, podendo sacrificar a precisão do modelo, pois pode impactar diretamente na percepção de detalhes finos em uma imagem.

- **Padding:** É uma técnica utilizada para tratar do problema de quando o tamanho do filtro e da imagem não são compatíveis, sendo que de forma geral ele adiciona linhas e colunas nas extremidades das imagens, onde o mais tipicamente utilizado é o *zero padding*, que preenche essas linhas e colunas com zeros permitindo extrair informações de extremidades, pois o filtro consegue ser centralizado nos elementos dos cantos e extremidades da imagem, permitindo que sejam melhor considerados nas operações de convolução, controlando o tamanho da saída, pois ao adicionar esses zeros nas extremidades, a saída pode ser controlada, permitindo maior flexibilidade na construção das CNN. O *padding* pode mitigar a perda de

informação, pois essa técnica auxilia na redução da rápida perda de dimensionalidade espacial nos mapas de características, mitigando possíveis perdas de informação.

Um ponto crucial da operação de convolução é o compartilhamento de pesos, onde os filtros são aplicados ao longo de todas as posições da imagem, gerando características importantes para a operação de convolução, como [24]: o mantimento da invariância das características locais extraídas pelos filtros à medida que percorrem as posições da imagem e detectam padrões; a capacidade de identificar e aprender hierarquias espaciais de padrões de características através do *downsampling* em conjunto com a operação de *pooling*, resultando na captura de um campo de visão maior; e um aumento da eficiência do modelo pela viabilização da redução da quantidade de parâmetros a serem aprendidos quando em comparação com redes neurais totalmente conectadas.

Por fim, após a definição do mapa de atributos é aplicada uma função de ativação, geralmente não-linear, nos elementos do mapa para transformá-lo em dados compatíveis para entrada da próxima camada de convolução. A função de ativação *Rectified linear unit* (ReLU) é a mais utilizada, principalmente pela sua eficiência computacional (advinda da sua simplicidade) para o treinamento de modelos altamente complexos. A ReLU apresenta capacidade de gerar representações esparsas, devido a geração de zeros nas saídas quando as entradas são negativas, sendo que assim somente alguns neurônios das redes são ativados, o que pode gerar melhoria na eficiência da rede.

2.2.2 Camada de *pooling*

As camadas de *pooling* são utilizadas em casos em que se deseja reduzir a dimensionalidade dos dados de entrada, realizando uma concentração de informações, que é especialmente útil em uma CNN, visto que dados de entrada próximos apresentam informações semelhantes [23]. Essa camada tem como intuito a redução do gasto de recursos computacionais (memória) e número de parâmetros treináveis da rede, o que influencia diretamente na minimização do tempo de treinamento de CNNs. Ademais, permite correções básicas para variações como pequenas rotações e translações nas imagens devido ao afunilamento dos dados de entrada. Apesar disso, vale ressaltar que a área de ação dessa camada, caso muito grande, pode causar descarte de informações, prejudicando o desempenho da rede.

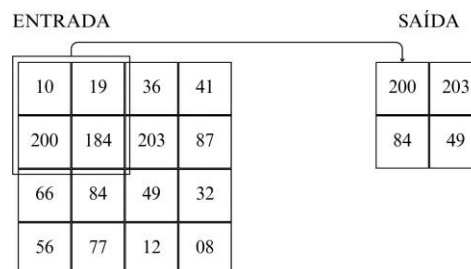
As camadas de *pooling* funcionam de forma semelhante as camadas convolucionais, onde cada neurônio apresenta uma ligação limitada a uma janela dos dados de entrada da camada e utilizando de conceitos das camadas convolucionais, como o *stride* e *padding*, percorre os dados de entrada para realizar a redução do número de parâmetros, a partir do

método selecionado para realização do *pooling*. Dentre os mais usados tipos de *pooling* tem-se [23, 25]:

- *Average pooling*: Esse método de *pooling* foi utilizado na primeira rede neural baseada em convolução e tem como cerne dividir a entrada em regiões retangulares não sobrepostas e gerar como saída a média computada dos valores presentes em cada região. Esse método tem como uma de suas vantagens a suavização das imagens, pois calculando-se a média dos valores, ocorre uma suavização das características, o que traz a rede menor sensibilidade a variações específicas dos valores;

- *Max pooling*: É a mais popular das operações e pode ser aplicada para reduzir a variabilidade nos mapas de características. Esse tipo de *pooling* extrai dos mapas de características o maior valor presente na janela dos dados e o utiliza para gerar sua saída final, conforme apresentado na Figura 9. Vale ressaltar que, ao escolher o valor máximo de uma região tem-se como consequência uma rede mais robusta para pequenas variações na posição das imagens, além de gerar uma melhor eficiência computacional, pois ao reduzir a dimensionalidade, a quantidade de cálculos realizados torna-se menor em camadas subsequentes;

Figura 9 – Exemplo da operação de *max-pooling*



Fonte: Autor (2024)

- *Mixed pooling*: É uma proposta híbrida, onde um valor λ é selecionado aleatoriamente sendo 0 ou 1, comportando-se como *average pooling* ($\lambda = 0$) ou como *max pooling* ($\lambda = 1$). Os resultados dos valores de λ são gravados para propagação direta e são utilizados durante o processo de retropropagação. Essa abordagem busca tirar proveito das vantagens distintas do *max pooling* e do *average pooling* objetivando a melhora da performance do modelo;

- *Global average pooling*: Esse tipo de *pooling* apresenta um tipo de *downsampling* extremo, onde em cada mapa de características é feita a média de todos os seus valores, sendo então transformados em *arrays* de dimensões 1x1, porém a profundidade dos mapas é mantida. Essa operação é mais aplicada anteriormente às camadas totalmente conectadas, objetivando reduzir o número de parâmetros para aprendizado e possibilitar às CNNs o processamento de entradas com tamanhos variáveis.

2.2.3 Camada *dropout*

O treinamento das CNNs apresenta a possibilidade de gerar modelos com baixa capacidade de generalização, sendo que para tratar este tipo de problema podem ser utilizadas as camadas de *dropout*. Essas camadas têm como função remover temporariamente e de forma aleatória alguns neurônios das camadas escondidas, seguindo uma probabilidade p para cada época [26]. Nesse sentido, para cada época de treinamento uma quantidade de neurônios das camadas escondidas, podendo ser uma camada convolucional ou totalmente conectada, é desconsiderada, seguindo a probabilidade p e em uma época seguinte esses neurônios podem ser reativados, tendo outros neurônios desconsiderados para tal época e assim sucessivamente.

2.2.4 Camadas *flatten* e totalmente conectadas

Finalizada a extração das características pelas camadas convolucionais, os mapas gerados podem ser utilizados como dados de entrada para as camadas totalmente conectadas. Entretanto para possibilitar a utilização desses mapas como entrada para uma camada totalmente conectada é preciso realizar o redimensionamento desses mapas bidimensionais para o formato vetorial unidimensional, sendo que para isso os mapas de características são alimentados à camada *flatten*.

Posteriormente à vetorização dos mapas, é responsabilidade das camadas totalmente conectadas realizar a tarefa de classificação ou regressão, sendo que seus parâmetros, como número de neurônios nas camadas, função de ativação, entre outros, são variados de acordo com a tarefa desempenhada pela CNN e os valores de pesos aprendidos durante processo de aprendizagem da CNN.

Para o caso da CNN sendo utilizada para um problema de classificação, a função de ativação geralmente utilizada nos neurônios da camada de saída é a função *softmax*. Essa função gera na sua saída uma distribuição de probabilidades de acordo com:

$$\sigma(X)_i = \frac{e^{X_i}}{\sum_{j=1}^K e^{X_j}} \quad (1)$$

onde X é o vetor de entrada, X_i é o i -ésimo elemento do vetor X , K é o número de elementos do vetor X e $\sigma(X)_i$ é a probabilidade de saída para o i -ésimo elemento de X .

2.2.5 Treinamento das redes neurais convolucionais

O treinamento das CNNs compreende um conjunto de etapas, que inicia com o processo de preparação dos dados de entrada. Quando se fala da preparação dos dados de entrada, uma CNN trabalha com imagens e essas imagens devem passar por uma fase de pré-processamento,

onde serão ajustadas com o objetivo de aumentar a capacidade de generalização do modelo proposto, sendo que dentre os ajustes a serem realizados tem-se a normalização, onde os tamanhos dos *pixels* são normalizados para que seus valores fiquem dentro de um mesmo alcance (0 a 1 ou -1 a 1, por exemplo). Além da normalização, os dados de entrada podem passar por um processo chamado de aumento dos dados, sendo que o aumento de dados pode ser realizado por meio de rotações, giros ou cortes nas imagens, sempre buscando maior diversidade no conjunto de treino. É importante destacar também a necessidade nesta fase da separação do conjunto de dados para treino, validação e teste dos modelos.

Após o tratamento dos dados de entrada é possível avançar para o estágio de definição da arquitetura do modelo, ou seja, definir quantas e quais camadas e funções serão utilizadas para desenvolvimento e treino da CNN, além de selecionar os parâmetros de cada camada. Diversos treinamentos deverão ser realizados em busca da melhor arquitetura sendo que para realizar o treinamento do modelo, alguns parâmetros relacionados ao algoritmo de treino devem ser testados em busca de resultados melhores de treinamento, sendo eles:

- Taxa de aprendizagem: controla a velocidade do aprendizado da rede em cada iteração, influenciando na descida do gradiente em busca do mínimo global da função a ser minimizada durante o treino das CNNs;
- Tamanho do lote: O tamanho do lote indica a quantidade de imagens utilizadas em uma iteração no processo de treinamento do modelo, influenciando diretamente na eficiência de treinamento e na performance do modelo;
- Número de épocas: Esse parâmetro refere-se ao número de passagens por todo o conjunto de treinamento, onde cada época consiste em um ciclo completo pelas imagens de treinamento.

Finalizada a definição da arquitetura da rede e dos parâmetros do algoritmo, é possível iniciar o treinamento, primeiramente pela propagação direta, na qual as imagens são passadas pela rede camada a camada, onde as camadas convolucionais adicionam filtros, as funções de ativação introduzem não-linearidades e as camadas de *pooling* realizam o *downsampling* dos dados. Chegada a última camada da rede, sua saída é achatada e é introduzida em camadas totalmente conectadas e ao final da última camada totalmente conectada sua saída é exposta a uma função de ativação (como a *softmax* para classificação) produzindo assim as previsões da CNN.

É válido expor aqui sobre as funções de perda, sendo que essas têm como propósito medir a diferença entre os resultados obtidos pela rede e seus rótulos reais, tendo características

específicas e aplicações específicas, podendo ser utilizadas para casos de regressão ou de classificação. Entre o conjunto de funções de perda geralmente utilizadas, têm-se:

- *Cross-entropy loss*: É comumente utilizada para tarefas de classificação em CNNs e mede a performance de modelos em que as saídas são probabilidades entre 0 e 1, sendo que seu valor aumenta quando a probabilidade prevista diverge do resultado real. Essa função pode ser dividida em binária e categórica sendo utilizada para classificações com saídas binárias e multiclasse, respectivamente. Por fim, apresenta grande importância e uso por trazer um bom alinhamento com a interpretação probabilística dos resultados, sendo útil para casos em que a confiança nas previsões é essencial e os gradientes são simples de calcular, tornando-a eficiente para uso em algoritmos de otimização baseados em gradientes descendente. A função de perda é calculada de acordo com:

$$Loss = - \sum_{i=1}^n t_i \log(p_i) \quad (2)$$

onde para n classes, t_i é o rótulo real e p_i é a probabilidade da função *softmax* para a i -ésima classe;

Após o cálculo do erro entre saída da rede e a saída desejada, é possível realizar a retropropagação do erro para ajustar os pesos da rede, buscando diminuir a diferença entre os rótulos reais e os previstos, aumentando assim a precisão da CNN. Para este processo de ajustes dos pesos buscando a minimização de uma função perda tem-se diversos algoritmos de otimização possíveis, dentre eles pode-se citar:

- *Adaptive moment estimation (ADAM)*: É um algoritmo de otimização avançado altamente utilizado no treinamento de CNNs devido a sua taxa de aprendizado adaptável e características de *momentum* que auxiliam na sua rápida e eficiente convergência. O ADAM é robusto para gradientes com muito ruído, sendo adequado para treinamento de redes complexas e com grandes conjuntos de dados e seus parâmetros padrões funcionam bem para uma grande gama de casos reduzindo a necessidade de extensivos ajustes. Apesar de necessitar de uma maior quantidade de memória e poder computacional, esse maior custo tende a ser justificado pelos benefícios trazidos pelo algoritmo ao treinamento e performance de CNNs. Por fim o ADAM é capaz de lidar com gradientes esparsos, por meio da sua taxa de aprendizagem adaptativa, o que é benéfico para certos tipos de dados e modelos.

2.3 Mecanismos de atenção

A atenção vem se tornando um conceito muito importante na área da aprendizagem profunda, onde mecanismos de atenção e de auto atenção tornaram-se notórios em tarefas de processamento de linguagem natural (PLN). Além da área de PLN, trabalhos vêm sendo

apresentados na literatura, com resultados promissores, com aplicação desses mecanismos em conjunto com as CNNs [27] em visão computacional, uma importante subárea da IA.

A atenção refere-se ao processo cognitivo de seleção, onde os humanos se concentram em alguns estímulos enquanto ignoram outros, apresentando a tendência de focar sua atenção em determinadas regiões do espaço visual, buscando extrair informações, em vez de considerar a imagem como um todo. A atenção melhora a representação dos estímulos sendo que ao focar em recursos essenciais e suprimir os desnecessários, nosso cérebro aumenta o poder de representação, reduzindo desta forma o ruído que tem de processar.

Os mecanismos de atenção são baseados exatamente neste processo biológico humano de focar seletivamente em estímulos buscando conseguir informações, sendo que a aplicação do mecanismo de atenção em conjunto com as CNNs possui como principal finalidade melhorar a capacidade destas redes de delimitar uma região da imagem com informações relevantes, ignorando as demais, utilizando para isto a análise de partes das imagens para gerar um entendimento maior do contexto geral para a problemática em que o modelo foi aplicado.

Para a classificação de tumores cerebrais a partir de imagens de RM, por exemplo, o mecanismo de atenção pode aumentar a capacidade da CNN de delimitar regiões das imagens com maior incidência de tumores, sendo que dessa forma a rede passa a focar em determinadas regiões na imagem para cada tipo de tumor trazendo, teoricamente, maior probabilidade de realizar uma classificação correta do tumor. É importante destacar que, para garantir compatibilidade com as CNNs, o mecanismo de atenção deve assumir a forma de uma camada que combina informações das camadas anteriores da rede, resultando em melhorias de desempenho.

2.3.1 Tipos de mecanismo de atenção

Apesar de manter o princípio lógico, o mecanismo de atenção apresenta diversas formas de ser implementado, divididas em 3 grupos maiores [28], sendo algumas formas mais comumente aplicadas para criação de modelos híbridos com as CNNs.

2.3.1.1 *Hard attention*

Esse tipo de atenção utiliza uma seção específica da imagem de entrada e funciona como um comutador, determinando se a rede prestará ou não atenção à seção selecionada. Para isso, esse modelo segmenta a imagem em blocos e simula uma votação entre eles. Utilizando-se de uma versão modificada do *self-attention*, a cada intervalo de tempo os blocos relevantes são

escolhidos e decisões são tomadas exclusivamente levando em consideração o subconjunto selecionado como importante.

Neste tipo de mecanismo o principal desafio é a sua não diferenciabilidade. Em redes neurais, existe uma dependência de métodos de gradiente no processo de treinamento, como o *backpropagation* (retropropagação). Apesar de seu bom funcionamento com funções contínuas, o *hard attention* toma decisões discretas quebrando essa continuidade e impedindo o uso direto dos gradientes. Em vista disso, o contorno desse problema se dá pelo uso de técnicas como algoritmos de aprendizado por reforço, que atribuem recompensas às escolhas feitas pela rede, trazendo melhoras nas decisões ao longo do tempo. Dessa forma, esses algoritmos podem ser utilizados para treinar o mecanismo de atenção levando em consideração que seguem uma lógica discreta, entretanto o aprendizado por reforço tende a ter uma variação alta, podendo tornar o treinamento mais lento e instável.

Apesar dos problemas relativos ao processo de aprendizado, o *hard attention* apresenta como principais benefícios:

- Eficiência computacional: Pela seleção de uma pequena porção dos dados para processamento, o *hard attention* pode ser muito mais eficiente, principalmente em aplicações que lidam com grandes volumes de dados, como imagens de alta resolução;
- Esparsidade natural: Visto que, em muitos casos, somente uma parte dos dados de entrada é realmente relevante para sua tarefa, o *hard attention* ignora essas partes irrelevantes, tornando o modelo mais focado e eficiente. Por exemplo, para reconhecimento de objetos em imagens, o fundo pode ser considerado irrelevante;
- Interpretabilidade: A sua clara tomada de decisão sobre quais partes da entrada são importantes pode trazer consigo uma maior facilidade de interpretação.

2.3.1.2 Soft attention

No *soft attention*, o modelo calcula as probabilidades para cada parte da entrada, indicando quanto de atenção deve ser direcionada para cada elemento, esses pesos ou probabilidades são aplicados a todas as partes relevantes, assim permitindo ao modelo considerar todas as informações, porém dando mais ênfase às partes que mais influenciam os resultados da tarefa em questão. Esse mecanismo tem como principais características:

- Diferenciável: Diferentemente do *hard attention*, é possível ao *soft attention* ser treinado com métodos padrão, como o *backpropagation*, assim facilitando a integração com variados modelos de redes neurais;

- Distribuição contínua: Por não tomar decisões binárias, este mecanismo distribui o foco suavemente permitindo ao modelo atendimento parcial de várias partes dos dados de entrada simultaneamente;

- Flexibilidade: Dada a distribuição do foco por toda a entrada, o *soft attention* apresenta maior robustez e flexibilidade, se saindo melhor em situações difíceis de determinar previamente as partes mais importantes.

O processo do mecanismo *soft attention* pode ser resumido em 3 etapas: primeiro, calcula-se uma pontuação para cada parte da entrada (por exemplo, cada pixel de uma imagem), sendo feito geralmente utilizando-se vetores de atenção e multiplicando-os com os estados ocultos da rede; segundo, por meio da função *softmax*, converte-se essas pontuações em probabilidades. As partes da entrada com pesos maiores são consideradas como mais relevantes; terceiro, aplica-se os pesos às respectivas partes da entrada, resultando em uma média ponderada, onde o resultado recebe maior influência das partes com maiores pesos. Assim, o *soft attention* apresenta os seguintes benefícios principais:

- Melhor performance: O desempenho das redes neurais é melhorado em tarefas complexas, como classificação de imagens, devido ao mecanismo permitir que o modelo foque em diversas partes dos dados de entrada;

- Facilidade de treinamento: A integração de maneira simples e eficiente em redes neurais mais tradicionais, visto que pode ser treinado com métodos como o gradiente descendente, que facilita o treinamento do *soft attention*;

- Generalização: Por não fazer suposições fortes sobre as partes da entrada e aprender diretamente dos dados quais partes são mais importantes, esse mecanismo tende a generalizar bem para variados tipos de entrada;

- Redução de erros: Devido a distribuição do foco por não depender de uma decisão única, como o *hard attention*, o *soft attention* pode ajudar a reduzir erros e mitigar possíveis falhas na captura de detalhes importantes dos dados.

2.3.1.3 Self attention

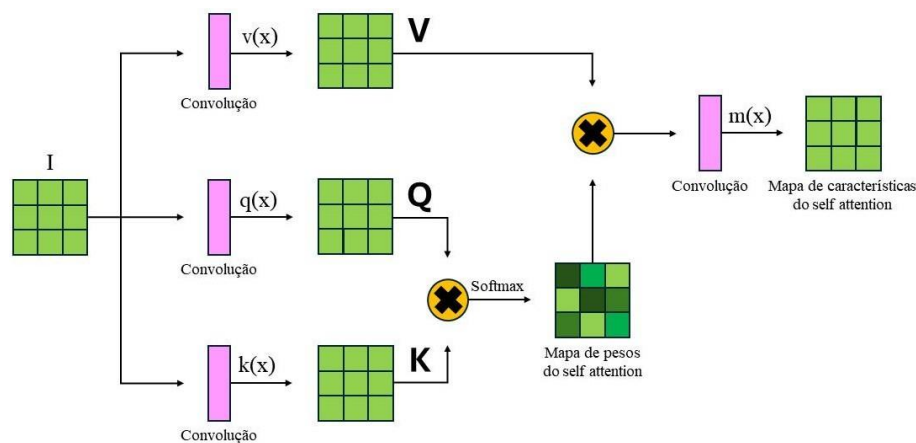
O mecanismo *self Attention*, utilizado no modelo *Transformer* [29] e o escolhido para ser usado em nosso trabalho, cria uma percepção diferente dos dados de entrada por meio da relação de diferentes posições dele. Apesar de ter sido inicialmente proposto em PLN, muitos autores sugerem sua aplicação (por meio do modelo *Transformer*) em tarefas de visão computacional, sendo desta forma a razão para escolha deste mecanismo para nosso trabalho.

O *self attention* combina informações usando os conceitos de *queries*, *keys* e *values* e é utilizado de uma forma que se cada *pixel* for tratado como uma variável aleatória no mapa de características e calcularmos a covariância entre os *pixels*, o valor das previsões desses nas imagens pode ser ajustado de acordo com a sua similaridade com outros *pixels* da mesma imagem. O mecanismo *self attention* apresenta as seguintes características principais:

- Dependências globais: O *self attention* permite que cada elemento “observe” todos os outros na sequência, capturando dependências de longo alcance de forma eficiente. Isso permite ao mecanismo contornar as limitações das redes *long short-term memory* (LSTM) ou *recurrent neural network* (RNN), que processam sequencialmente as entradas;
- Paralelização: Esse mecanismo processa simultaneamente todos os elementos da sequência, assim aumentando a eficiência computacional e a celeridade do treinamento;
- Escalabilidade: Adapta-se bem a diferentes tipos de dados, pela possibilidade de ser aplicado em sequências de variados tamanhos;

O *self attention* realiza o cálculo da atenção para cada elemento em relação ao demais da sequência de entrada. O mecanismo de atenção, apresentado na Figura 10, realiza o mapeamento dos dados de entrada em matrizes denominadas *queries* (Q), *keys* (K) e *values* (V). Após esse mapeamento, as matrizes *queries* e *keys* são multiplicadas e o resultado é passado por uma camada de *softmax*, assim gerando a matriz de pesos do mecanismo de atenção que indica quais regiões da imagem de entrada têm mais relevância para classificação. Por fim, as matrizes de *values* e de pesos de atenção são multiplicadas gerando o mapa de características do mecanismo.

Figura 10 – Arquitetura do mecanismo *attention*



Apesar das características principais do mecanismo mostrarem vantagens do *self attention*, ele apresenta limitações como: um custo computacional quadrático relacionado ao tamanho da sequência que, embora consiga capturar dependências globais, pode ser desafiador para dados sequenciais longos; alto consumo de memória para sequências muito grandes, pois o mecanismo requer o mantimento de informações sobre os dados da sequência.

2.4 Mapa de ativação de Classes

Explicar como as redes neurais de forma geral fazem suas previsões, sempre foi considerado uma caixa preta devido a sua complexidade e, dada essa falta de clareza, este tópico é de grande interesse pela comunidade de pesquisas. Em vista disso, muitos métodos vêm sendo desenvolvidos buscando facilitar o entendimento e oferecer maior explicação das previsões obtidas por modelos como as CNNs.

Um dos métodos mais difundidos é o mapa de ativação de classes (*class activation map* - CAM) que, conceitualmente, realiza a construção de mapas de calor nas imagens indicando as regiões consideradas pela CNN como mais pertinente para reconhecimento de uma categoria particular [30]. Em outras palavras, uma CAM pode auxiliar no processo de entendimento em relação a quais regiões da imagem influenciam na decisão, de um modelo de rede neural, para identificar ou classificar informações presentes em uma imagem.

Essa técnica funciona a partir do uso dos mapas de ativação gerados pela camada convolucional, geralmente a anterior à camada de classificação final, aplicando-se uma média ponderada usando os pesos da camada de classificação final, sendo assim calculado um mapa para cada classe presente no conjunto de dados.

Após o cálculo do CAM, o mapa gerado para a classe é redimensionado para a resolução da imagem original e sobreposto a ela. Assim, as regiões com cores mais quentes indicam onde o modelo dedicou mais foco para tomar a decisão na determinada classe.

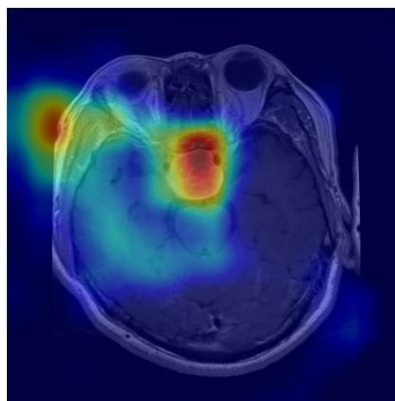
Os CAMs apresentam vantagens como aumentar a interpretabilidade da rede convolucional, possibilitando a visualização das partes da imagem utilizadas para a tomada de decisão, dessa forma revelando se o modelo está fazendo inferências baseando-se nas características corretas. Apesar disso, essa técnica requer o uso de uma arquitetura de rede específica (contendo camadas de convolução, por exemplo) para relacionar diretamente as características das imagens com a camada de saída, ademais o mapa gerado tende a ser menos detalhado que as imagens originais devido a *downsamplings* realizados nas camadas de convolução e de *pooling*.

Como forma de estender o CAM, foi desenvolvido o Score-CAM [31] que tem como principal diferença a não utilização direta dos pesos da rede para o cálculo da importância, oferecendo uma abordagem mais flexível para gerar mapas de ativação. Essa extensão, como o CAM, gera mapas de ativação da última camada de convolução e para cada mapa realiza a multiplicação com o valor original da imagem, gerando uma nova imagem ponderada. Em seguida, a nova imagem é passada pela rede e a saída é utilizada como uma pontuação representando a importância do mapa de ativação. Por fim o mapa final é uma combinação ponderada desses mapas, baseado nas pontuações obtidas.

O Score-CAM, quando comparado ao CAM apresenta as vantagens de não depender da estrutura da rede, assim podendo ser aplicado a um maior leque de conjuntos de rede, sem ser preso a uma estrutura padrão. Além dessa diferença, essa técnica usa a própria rede para cálculo da importância dos mapas de ativação, podendo acarretar representações mais precisas.

A Figura 11 apresenta um exemplo de um Score-CAM gerado para a CNN com o mecanismo de atenção, a partir de uma imagem de RM de tumor cerebral pituitário advinda do conjunto de teste dos dados utilizados neste estudo.

Figura 11 – Exemplo de Score-CAM para imagem de RM do conjunto de dados



Fonte: Autor (2024)

2.5 Estado da arte de classificação de tumores cerebrais baseados em CNN

Essa seção apresenta o estado-da-arte, considerando os últimos 5 anos, de trabalhos voltados para classificação de tumores cerebrais através de modelos computacionais baseados em CNNs e CNNs com mecanismo de atenção.

Em Kahn *et al.* [32], é proposta uma arquitetura de CNN baseada em blocos convolucionais para detecção de tumores cerebrais a partir de uma base de dados de RM multiclasse. Essa arquitetura apresentou robustez e eficiência na diferenciação dos tipos de tumores, além de ter demonstrado alta acurácia na sua avaliação em 3 conjuntos de dados

diversos, dessa forma obtendo resultados competitivos com modelos estado-da-arte como o VGG16, VGG19, ResNet50 e entre outros.

Remzan *et al.* [33], propuseram um modelo de CNN personalizado para classificação de tumores cerebrais utilizando um conjunto de imagens de RM. A partir disso, seus resultados foram considerados pelos cálculos de 4 métricas (acurácia, precisão sensibilidade e F1-score) e por motivos comparativos, variou-se o classificador da rede entre o tradicional *softmax*, *XGBoost*, *Support Vector Machines* (SVM), *Random Forest* e *K-Nearest Neighbors* (KNN). Ao final, foi possível notar que a rede com o classificador tradicional *softmax* obteve um desempenho melhor quando comparada aos demais, a qual se sobressaiu também ao ser analisada em relação a modelos já presentes na literatura.

Kaya [34] apresentou um modelo de CNN para classificação de imagens de RM de tumores cerebrais que se beneficiou da otimização dos seus hiperparâmetros a partir de processos gaussianos e bayesianos. Assim, utilizando-se de dois conjuntos de dados simultaneamente para treinamento, o modelo proposto entregou resultados satisfatórios em ambos os conjuntos de teste, tornando-o valioso para a detecção de tumores cerebrais.

Al-zoghby *et al.* [35] desenvolveram um modelo denominado *Dual Convolutional Tumor Network* (DCTN) com o objetivo de concatenar a arquitetura VGG-16 pré-treinada e uma CNN personalizada. Para realizar essa concatenação, é feita a média das saídas das duas redes e a concatenação de seus resultados para assim os dados serem apresentados para as camadas totalmente conectadas. Após feitos os testes da DCTN, foi concluído que a metodologia proposta obteve alta melhoria de resultados em relação às pesquisas já existentes na literatura.

Em Güler e Namli [36] é apresentado um estudo que realiza uma amálgama de arquiteturas (VGG, ResNet, DenseNet e SqueezeNet) e métodos de aprendizado de máquina (SVM, KNN, *Naive Bayes*, *Decision Tree*, *Linear Regression Analysis*) para classificação de imagens de tumores cerebrais. Mais especificamente, as arquiteturas são treinadas para extração de características das imagens e para classificação são utilizados os modelos de aprendizado citados com um método de aprendizagem em conjunto para seleção dentre eles do melhor para validação. Essa lógica de aplicação desses métodos e modelos destacou suas eficiências em classificar de forma precisa imagens de RM cerebrais trazendo possíveis melhorias nas capacidades de diagnóstico de tumores no cérebro.

Tüzün e Özdemir [37] trouxeram uma análise comparativa entre modelos de *deep learning* para doenças cerebrais utilizando como dados imagens de RM. Para essa análise foram implementados os modelos Efficientnet-b0 (99.54%), InceptionV3 (99.47%), Mobilenetv2

(98.93%) e GoogleNet (98.25%), sendo suas métricas estatísticas obtidas em ordem decrescente, dessa forma indicando sugestões úteis para tomadas de decisão sobre utilização de determinados modelos, em particular, trazendo vantagens para o diagnóstico precoce de doenças como os tumores cerebrais.

Apesar da grande distribuição do uso das CNNs para o problema de classificação de tumores cerebrais, nos últimos anos surgiram trabalhos na literatura indicando modelos híbridos das CNNs com adição do mecanismo de atenção, buscando trazer maior foco na análise das características dos dados como forma de aumentar sua performance.

Em Jun e Liyuan [38], é proposto um modelo híbrido que busca contornar a limitação das CNNs na identificação precisa dos pontos focais de lesões cerebrais a partir da adição do mecanismo de atenção. Nesse sentido, o mecanismo é utilizado para seleção das informações mais relevantes dos pontos focais, enquanto a rede convolucional multicaminho distribui os dados anteriormente à conversão dos resultados alcançando valores de acurácia que superam os apresentados por estudos anteriores que se utilizaram do mesmo conjunto de dados.

Abdusalomov *et al.* [39] apresentou um refinado modelo *You Only Look Once version 7* (YOLOv7) para classificação de tumores de glioma, meningioma e hipofisários. Em vista disso, para aumentar a captura de características das imagens foi introduzido um *Convolutional Block Attention Module* (CBAM) ao modelo YOLOv7. Por fim, o estudo concluiu que o modelo desenvolvido apresentava acurácia maior que modelos estado-da-arte anteriores para detecção de tumores.

Tang *et al.* [40], propuseram o *gradient awareness minimization* (GAM) para contornar a generalização insuficiente do SGD. Utilizando-se desse algoritmo em conjunto com o *positional attention convolution block*, o *relative self-attention transformer block* e o *intermitent fully connected layer* para desenvolvimento do *Spinal Convolution Attention Network* (SpCaNet). Esse método, dito como mais leve e eficiente em reconhecimento de tumores cerebrais, apresentava 1/3 do número de parâmetros dos modelos estado-da-arte comparativos e obteve bom desempenho na classificação dos tumores no cérebro.

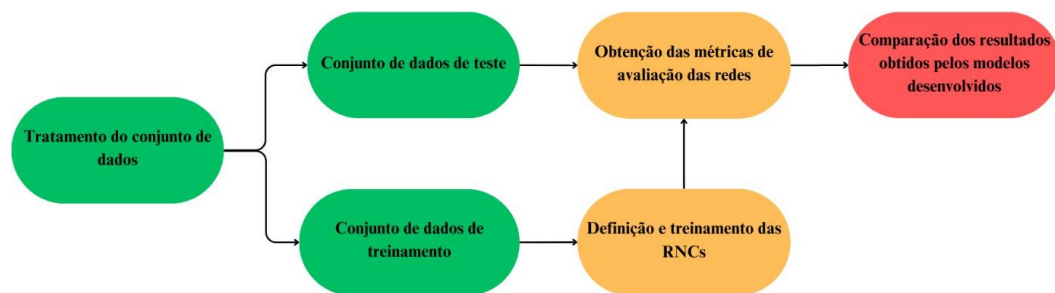
3 METODOLOGIA

Neste capítulo será apresentada a metodologia adotada para desenvolvimento dos modelos CNN e CNN com mecanismo de atenção propostos neste trabalho para classificação de tumores cerebrais a partir de imagem de ressonância magnética.

3.1. Mapa metodológico

Como procedimento metodológico, para desenvolvimento e comparação dos dois modelos de CNNs propostos neste trabalho, 3 etapas foram estabelecidas, conforme apresentado no diagrama da Figura 12, sendo elas: tratamento do conjunto de imagens de RM a ser utilizado para treinamento e testes dos dois modelos classificadores; definição e treinamento da CNN e CNN com mecanismo de atenção, com obtenção das métricas de avaliação dos modelos considerando a base de teste de imagens de RM e comparação dos resultados obtidos pelos modelos classificadores desenvolvidos.

Figura 12 – Fluxo de etapas do procedimento metodológico



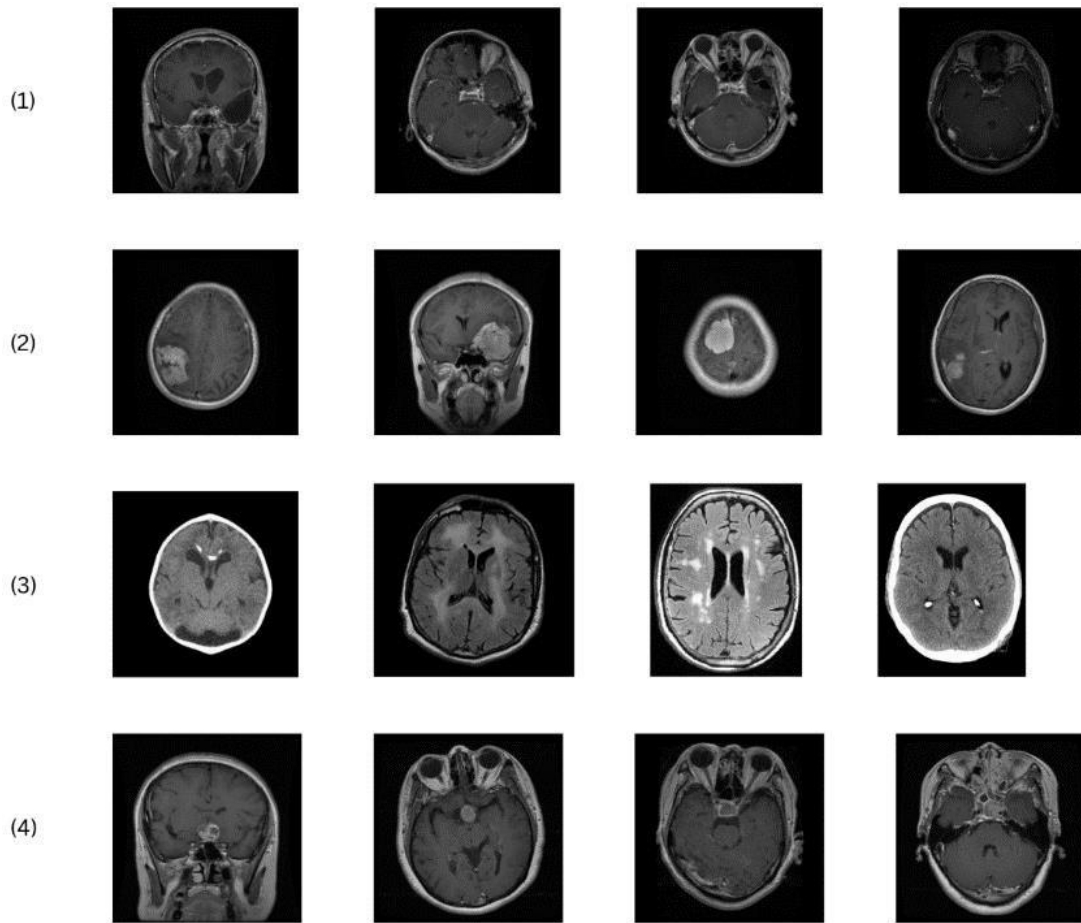
Fonte: Autor (2024)

A seguir cada uma das etapas do procedimento metodológico será apresentada com mais detalhes.

3.2 Conjunto de dados de imagens de RM – Tratamento dos dados

Para desenvolvimento dos classificadores de tumores cerebrais propostos neste trabalho foi utilizada a base de dados de imagens de Ressonância Magnética disponibilizada por Masoud Nickparvar [12], sendo um conjunto de imagens que é uma combinação dos conjuntos de imagens de Sartaj Buhvaji [41], Br35H [42] e Figshare [43]. A base contém um total de 7023 imagens de RM cerebrais, divididas em 4 classes, sendo elas: *glioma*, *meningioma*, *pituitário* e *sem tumor*, sendo que devido os conjuntos apresentarem imagens advindas de diferentes bases, os seus tamanhos são diferentes. A Figura 13 apresenta exemplos de imagens de RM da base de dados para as classes *glioma*, *meningioma*, *sem tumor* e *pituitário*.

Figura 13 – Imagens de RM de Glioma (1), Meningioma (2), Sem tumor (3) e Pituitário (4)



Fonte: Autor (2024)

A base de dados apresenta uma divisão, feita pelo autor do conjunto de dados, em uma proporção de aproximadamente 81.33% das imagens para treinamento e 18.67% das imagens para teste. O conjunto de treino possui 5712 imagens, sendo que 1595 são da classe *sem tumor*, 1457 da classe *pituitário*, 1339 da classe *meningioma* e 1321 imagens da classe *glioma*. O conjunto de teste contém 1311 imagens, divididas em 405 para a classe *sem tumor*, 300 para a classe *pituitário*, 306 para a classe *meningioma* e 300 imagens para a classe *glioma*. A tabela 1 resume a divisão da base de imagens para treino e teste.

Tabela 1 – Resumo da divisão da base de imagens para treino e teste

Conjunto de dados	Sem tumor	Pituitário	Meningioma	Glioma	Total
Treino	1595	1457	1339	1321	5712
Teste	405	300	306	300	1311

Fonte: Autor (2024)

A base de dados de treino original foi dividida em grupos para treino e validação, com uma porcentagem de 75% das imagens para treino e 25% para validação. O conjunto de treino então ficou composto com 4284 imagens, sendo 1196 da classe *sem tumor*, 1093 da classe

pituitário, 1004 da classe *meningioma* e 991 da classe *glioma* e o conjunto de validação ficou com 1428 imagens, sendo 399 da classe *sem tumor*, 364 da classe *pituitário*, 335 da classe *meningioma* e 330 da classe *glioma*. A tabela 2 resume a divisão da base de imagens para treino e validação.

Tabela 2 – Resumo da divisão da base de imagens para treino e validação

Conjunto de dados	Sem tumor	Pituitário	Meningioma	Glioma	Total
Treino	1196	1093	1004	991	4284
Validação	399	364	335	330	1428

Fonte: Autor (2024)

Alguns pré-processamentos foram aplicados nas imagens de treino e validação sendo elas, a normalização das imagens para média e desvio padrão de 0.5 e o redimensionamento das imagens, onde após diversos treinos com alguns tamanhos de imagens diferentes, os melhores resultados obtidos foram para o redimensionamento da imagem para 200x200 pixels. Para o conjunto de teste, foram feitas as mesmas normalizações e redimensionamento aplicados a base de treino

Para aumento da base de dados (data augmentation) de treino e validação, para melhoria da capacidade de generalização das redes, foram criadas imagens a partir de rotações e giros horizontais aleatórios nas imagens originais.

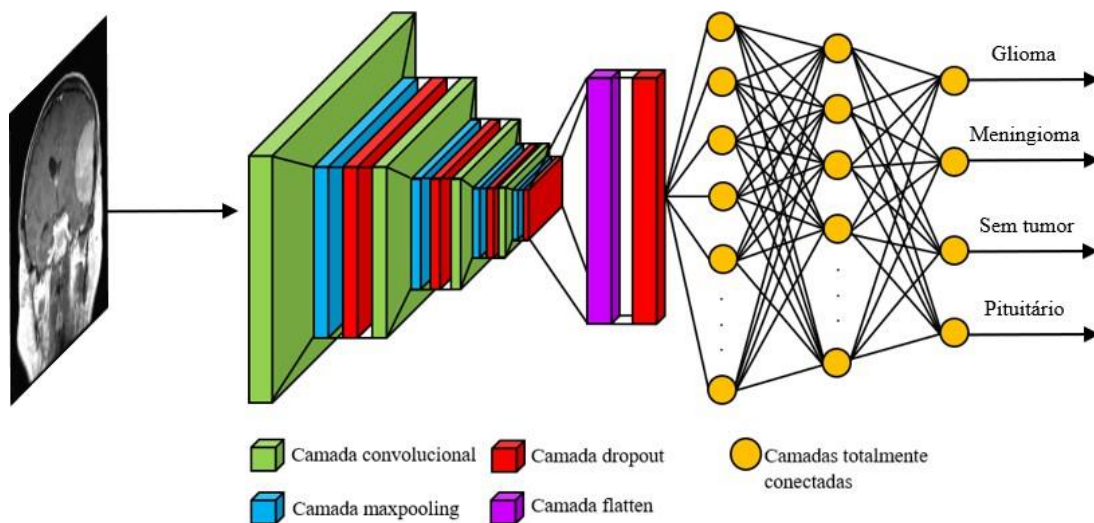
3.3 Treinamento da CNN e CNN com mecanismo de atenção para classificação de tumores cerebrais

Nesta etapa foi realizada a definição das estruturas das redes CNN e CNN com mecanismo de atenção, assim como foi realizada a etapa de treinamentos das redes.

3.3.1 Definição e treinamento da CNN sem mecanismo de atenção

A Figura 14 apresenta a estrutura da CNN selecionada após diversos treinamentos, onde foi variado o número de camadas convolucionais e camadas totalmente conectadas, buscando melhores resultados, considerando as métricas de avaliação na base de teste.

Figura 14 – Estrutura da CNN



Fonte: Autor (2024)

A CNN selecionada apresenta quatro camadas de convolução (Conv2d) com tamanho de filtros 4x4 na primeira e terceira camada e 5x5 na segunda e quarta camada, tendo valor de *strides* de 1 para a segunda e quarta camada e valor de *strides* 2 para a primeira e terceira camada, além de valores de *padding* de 0 para a quarta camada e 1 para as demais camadas. Associado às Conv2d, foi aplicada a função de ativação *Rectified Linear Unit* (ReLU).

Nas quatro camadas de *pooling* utilizou-se a função estatística de agregação *max-pooling* (MaxPool2d), sendo que a dimensão utilizada foi 4x4 na primeira camada, 5x5 na segunda e quarta camada e 3x3 na terceira camada, tendo os valores de *strides* 2 para a primeira camada e 1 para as demais camadas e os valores de *padding* de 1 para a primeira e quarta camada e o valor 0 para a segunda e terceira camada. Às camadas de *Dropout* foi atribuído o valor de 30% dos neurônios anulados randomicamente. Posteriormente, utilizou-se uma camada *Flatten* e então seguiu-se à implementação de 3 camadas totalmente conectadas (*Dense*), cujo número de neurônios em cada camadas foi de 1000 na primeira camada, 500 na segunda camada e 4 na terceira camada. Após as camadas *Dense* tem-se a camada de saída com 4 neurônios e funções de ativação *softmax*, onde cada neurônio indica a probabilidade de um tipo de tumor cerebral.

O algoritmo utilizado para treinamento da rede foi o ADAM e a função para avaliação de desempenho utilizada durante o treinamento foi a *cross-entropy loss*. Todos os treinamentos foram feitos com até 100 épocas, sendo utilizada uma função para armazenamento dos pesos da rede na época em que se obteve o menor erro nos dados de validação. Os parâmetros de treinamento e propriedades da rede proposta estão resumidos na tabela 3.

Tabela 3 – Parâmetros de treinamento e propriedades da CNN

Item	Descrição
Número de épocas	100
Tamanho dos minilotes de treinamento	28
Taxa de aprendizagem	0.0001
Algoritmo de treinamento	ADAM
Métrica de avaliação de desempenho do modelo	<i>Cross-entropy loss</i>
Número de camadas convolucionais	4
Tamanho dos filtros das camadas convolucionais	4x4 e 5x5
Função de ativação	ReLU
Número de camadas de <i>pooling</i>	4
Tamanho das janelas de <i>pooling</i>	3x3, 4x4 e 5x5
Número de camadas de <i>dropout</i>	5
Taxa de neurônios anulados nas camadas de <i>dropout</i>	30%
Número de camadas <i>Dense</i>	3
Tempo de treinamento	35 minutos

Fonte: Autor (2024)

Os dados de treinamento foram divididos em 28 minilotes, sendo que essa estratégia é utilizada para reduzir o esforço computacional do algoritmo que ajusta os pesos da CNN, além de tornar o treinamento mais rápido onde, ao utilizar minilotes, os parâmetros da rede são ajustados ao final da passagem dos dados contidos em cada minilote.

O pré-processamento dos dados, assim como a modelagem e treinamento das redes, foram feitos em um computador com processador AMD Ryzen 7 5800X 3.8 GHz, RAM de 64 GB DDR4, em *Python* 3.10.14, usando *Jupyter Notebook*. A ferramenta de aprendizado de máquina utilizada foi o *Pytorch*, além de *Sci-kit Learn* e outras bibliotecas básicas de *Python* usadas para processamento de dados e visualizações gráficas, como *Numpy*, *Matplotlib*, *Pandas* e *Seaborn*.

Diversos treinamentos da CNN foram realizados com pesos iniciais diferentes e as CNNs treinadas foram avaliadas, considerando a base de teste, para definição do melhor modelo, a partir das métricas: acurácia, precisão, sensibilidade e f1-score. Essas métricas se baseiam nos cálculos de acordo com a relação entre os resultados previstos e os valores reais e utilizam-se de conceitos como: verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN), onde:

- Verdadeiro positivo: Quando o modelo indica que a classe é positiva e a classificação real é positiva;

- Verdadeiro negativo: O modelo indica como resultado classe negativa e a classificação real é negativa;

- Falso positivo: Nessa situação o modelo resulta na classe positiva, mas a classificação real é negativa;

- Falso negativo: Esse caso determina quando o modelo resulta em classe negativa, mas a classificação real é positiva.

A métrica acurácia mede o percentual de acertos e indica o quão bem o modelo prevê corretamente as classes presentes nos dados, podendo ser calculada através de:

$$\text{acurácia} = \frac{VP+VN}{VP+VN+FP+FN} \quad (5)$$

A métrica sensibilidade (também chamada de *recall*) avalia a capacidade da CNN de detectar com sucesso resultados com classificação positiva. Essa métrica indica dentre os casos positivos no conjunto de dados quantos foram corretamente detectados pelo modelo podendo ser calculada através de:

$$\text{sensibilidade} = \frac{VP}{VP+FN} \quad (6)$$

A precisão avalia a quantidade de verdadeiro positivos com relação ao total de valores positivos obtidos pelo modelo de rede neural, ou seja, ela auxilia a identificar o quão bom é o desempenho da CNN para identificar corretamente instâncias positivas no conjunto de dados podendo ser calculada através de:

$$\text{precisão} = \frac{VP}{VP+FP} \quad (7)$$

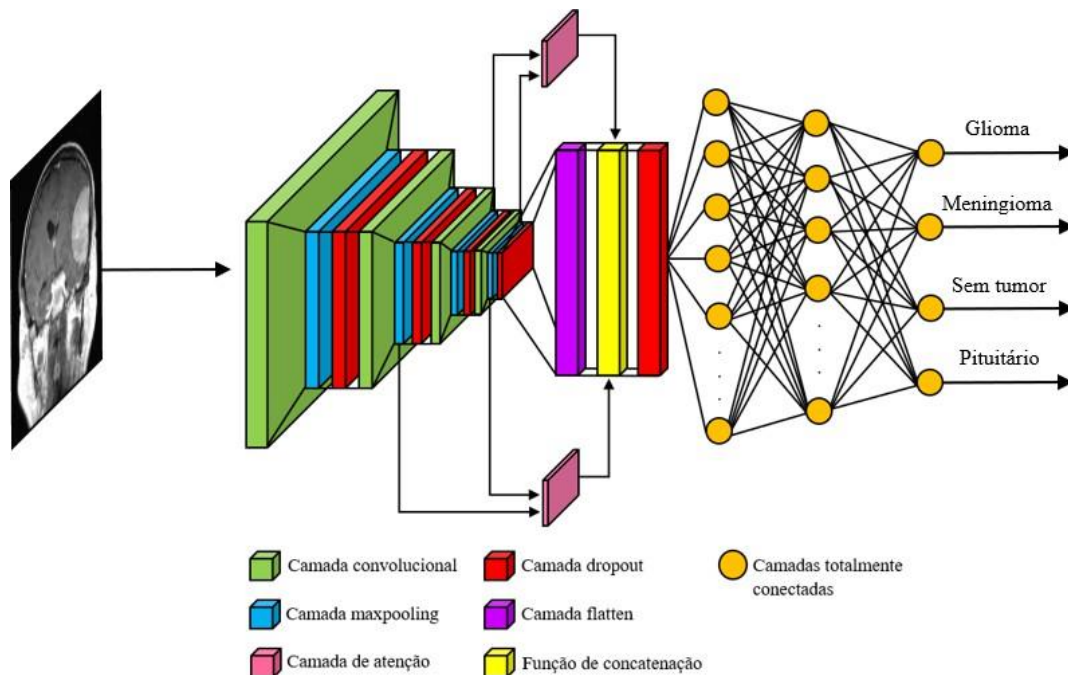
Por fim, a F1-score é uma média harmônica calculada com base na precisão e na sensibilidade sendo muito útil quando está sendo utilizado um conjunto de dados desbalanceado:

$$F1_score = 2 * \frac{\text{precisão} * \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}} \quad (8)$$

3.3.2 Definição e treinamento da CNN com mecanismo de atenção

A Figura 15 apresenta a estrutura da CNN com mecanismo de atenção utilizada para desenvolvimento do sistema classificador de tumores cerebrais a partir de imagens de RM.

Figura 15 – Estrutura da CNN com mecanismo de atenção



Fonte: Autor (2024)

O modelo se baseou na estrutura da CNN sem mecanismo de atenção apresentada na figura 14, sendo que duas camadas de atenção foram adicionadas à esta estrutura.

As duas camadas de atenção adicionadas utilizaram a saída da segunda, terceira e quarta camadas de *pooling* como entrada, sendo organizado para que a primeira camada de atenção realizasse seu processamento com informações provenientes da segunda e quarta camada de *pooling* e a segunda camada de atenção com informações da terceira e quarta camada de *pooling*. As saídas das camadas de atenção foram então concatenadas (união de dois ou mais elementos em uma sequência única) com as saídas da camada *Flatten*, conforme o conceito apresentado em [29] sobre atenção multicabeças, e o resultado desta concatenação foi passado por uma camada de *Dropout* com 30% de anulação randômica. Por fim, o vetor de dados resultante com tamanho 25184 foi utilizado como dado de entrada para 3 camadas totalmente conectadas (*Dense*), cujo número de neurônios foi definido de 1000 para a primeira camada, 500 para a segunda camada sendo que ao final, tem-se a camada de saída com 4 neurônios, com funções de ativação *softmax*, sendo que cada neurônio desta camada indica a probabilidade para um tipo de tumor cerebral.

Para o treinamento, assim como para a rede CNN sem mecanismo de atenção, o algoritmo utilizado foi o ADAM, tendo métrica de avaliação de desempenho utilizada durante o treinamento a *cross-entropy loss*. Todos os treinamentos foram realizados com até 300 épocas, sendo utilizada uma função para armazenamento dos pesos da rede na época em que se obteve

o menor erro nos dados de validação. Os parâmetros de treinamento e propriedades da rede proposta estão resumidos na tabela 4.

Tabela 4 – Parâmetros de treinamento e propriedades da CNN com mecanismo de atenção

Item	Descrição
Número de épocas	300
Tamanho dos minilotes de treinamento	28
Taxa de aprendizagem	0.00001
Algoritmo de treinamento	ADAM
Métrica de avaliação de desempenho do modelo	<i>Cross-entropy loss</i>
Número de camadas convolucionais	4
Tamanho dos filtros das camadas convolucionais	4x4 e 5x5
Função de ativação	ReLU
Número de camadas de <i>pooling</i>	4
Tamanho das janelas de <i>pooling</i>	3x3, 4x4 e 5x5
Número de camadas de <i>dropout</i>	5
Número de camadas de atenção	2
Taxa de neurônios anulados nas camadas de <i>dropout</i>	30%
Número de camadas <i>Dense</i>	3
Tempo de treinamento	1 hora e 27 minutos

Fonte: Autor (2024)

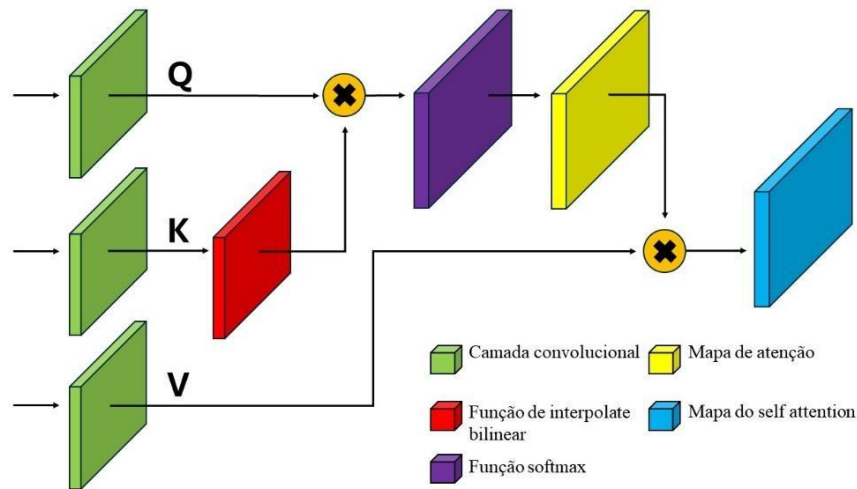
Para o estudo proposto definiu-se como mecanismo de atenção a ser utilizado em conjunto com a CNN o *self attention* utilizado nos modelos *Transformers* e apresentado no artigo intitulado “*Attention Is All You Need*”, pois este tipo de mecanismo já vem sendo utilizado em diversos estudos apresentados na literatura atual [44], indicando bons resultados para a área de processamento de imagens, além de apresentar melhor velocidade e uso do espaço de armazenamento que outros modelos de *attention* para grandes conjuntos de dados [29].

A estratégia utilizada para aplicação do mecanismo foi de utilizá-lo como uma camada da CNN, construindo-o com classes já existentes na biblioteca PyTorch, presentes no módulo *nn*, pois dessa forma o mecanismo poderia ser completamente compatível com a linguagem de programação, com a biblioteca e com o processamento via placa de vídeo.

Dessa forma, a nova camada denominada de “*AttentionLayer*”, é um conjunto sequencial de classes já existentes no PyTorch que realiza os passos do mecanismo como indicado no artigo dos *Transformers*.

A estrutura da camada de atenção utilizada na CNN está apresentada na Figura 16, sendo indicado quais camadas e funções da biblioteca PyTorch foram utilizadas para sua composição.

Figura 16 – Estrutura da camada de atenção



Fonte: Autor (2024)

De forma mais detalhada as camadas de convolução geram as matrizes *queries*, *keys* e *values* a partir das saídas das camadas de *pooling* da CNN. Em seguida, a matriz *keys* passa por uma função de *interpolate* bilinear buscando a compatibilidade de dimensões com as da matriz *queries* para assim realizar a multiplicação entre ambas. Posteriormente, o resultado obtido é passado por uma função *softmax* tendo como saída o mapa de atenção que contém os pesos das seções da imagem de entrada. Ao final, essa matriz de atenção é multiplicada com a matriz de *values*, assim alcançando o mapa final do *self attention* indicando as regiões mais relevantes da imagem.

Diversos treinamentos da CNN com mecanismo de atenção foram realizados com pesos iniciais diferentes e as CNNs treinadas foram avaliadas, considerando a base de teste, para definição do melhor modelo, a partir das métricas: acurácia, precisão, sensibilidade e f1-score.

3.4 Comparação de resultados dos modelos classificadores

Nesta etapa, com o propósito de comparar o desempenho dos modelos classificadores desenvolvidos e verificar se o uso do mecanismo de atenção pode trazer melhorias nos resultados da CNN para o problema específico de classificação de tumores cerebrais, foram utilizadas as 4 métricas obtidas a partir da base de imagens separadas para teste dos modelos, sendo elas: acurácia, precisão, sensibilidade e F1-score. As métricas foram obtidas de forma global, além de especificamente para cada tipo de classe de tumor cerebral presentes no conjunto de dados. Para cada modelo também foi analisada a matriz de confusão, com o objetivo de se ter uma visão geral do modelo quanto a quantidade de imagens corretamente classificadas.

Para uma análise mais detalhada e comparativa em relação a qual parte das imagens cada uma das CNNs treinadas, CNN e CNN com mecanismo de atenção, considera como mais relevantes para seu resultado de classificação, foi aplicada para algumas imagens da base de teste o CAM, sendo essa uma técnica utilizada para criar mapas de calor que destacam regiões consideradas como mais relevantes pela CNN para reconhecimento de uma determinada classe. O CAM foi aplicado na imagem gerada pela CNN a partir da última camada de *pooling* antes de ser entregue a camada *flatten* e pela CNN com mecanismo de atenção na última camada de convolução que constitui o mecanismo de atenção integrado à rede antes dos dados serem apresentados para camada *flatten*. O CAM foi implementado utilizando um módulo pronto disponibilizado no GitHub como uma biblioteca *Python* denominada de *TorchCam* [45] apresentando integração fácil com a biblioteca *Pytorch* utilizada para desenvolvimento dos modelos de CNN com e sem o mecanismo de atenção.

4 RESULTADOS

Neste capítulo serão apresentados os resultados obtidos dos treinamentos realizados das redes CNN e CNN com mecanismo de atenção para classificação de tumores cerebrais a partir de imagens de ressonância magnética.

Com os modelos classificadores definidos, uma análise comparativa dos resultados será realizada para indicação do melhor modelo, onde procura-se investigar se o mecanismo de atenção, especificamente para este caso de diagnóstico de tumores cerebrais baseados em imagens de ressonância magnética, pode ser utilizado para melhoria dos resultados dos classificadores desenvolvidos a partir de CNNs sem mecanismos de atenção.

4.1 Resultados para CNN sem mecanismo de atenção

A tabela 5 apresenta as métricas globais obtidas para a CNN sem mecanismo de atenção.

Tabela 5 – Métricas globais resultantes da CNN

Modelo	Acurácia	Precisão	Sensibilidade	F1-score
CNN	96.41%	96.24%	96.17%	96.61%

Fonte: Autor (2024)

É possível observar que os resultados, para todas as métricas, ficaram em torno de 96%, podendo ser considerado um resultado satisfatório quando comparado aos resultados de outros classificadores já apresentados na literatura atual, utilizando o mesmo conjunto de dados de teste, e ainda considerando a menor complexidade da CNN desenvolvida em relação aos demais.

Para uma análise mais completa dos resultados da CNN, a tabela 6 apresenta os valores das métricas calculadas por classe considerando base de teste.

Tabela 6 – Métricas resultantes da CNN para cada classe

CNN	Glioma	Meningioma	Sem tumor	Pituitário
Acurácia	95.33%	90.52%	99.51%	99.33%
Precisão	99.45%	98.98%	99.68%	98.21%
Sensibilidade	98.90%	97.76%	99.84%	98.84%
F1-score	99.18%	98.37%	99.76%	99.02%

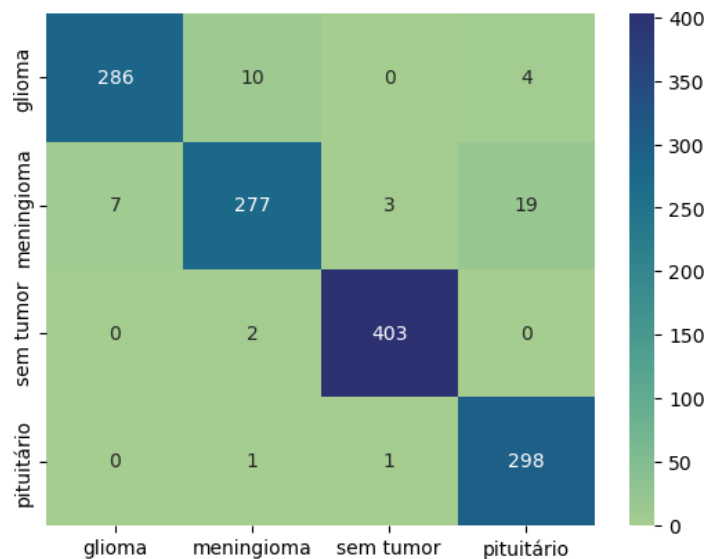
Fonte: Autor (2024)

Conforme pode ser observado na tabela 6, o modelo obteve resultados satisfatórios para todas as classes, principalmente no que diz respeito a métrica sensibilidade, que fornece exatamente a capacidade do classificador indicar um indivíduo para uma classe que realmente

ele pertence. Os melhores resultados, de forma geral, foram para as classes *sem tumor* e *pituitário* sendo que apresentou como melhor métrica a sensibilidade com porcentagem de 99.84% para ambos, indicando que o modelo teve maior facilidade para reconhecimento de casos saudáveis, sem a presença de tumores cerebrais. Além disso, foi possível notar que o modelo apresentou resultados de métricas menores para o caso de tumor *meningioma*, com o menor valor das métricas sendo a acurácia de 90.52%, mostrando que o modelo teve dificuldade na identificação deste tipo de tumor.

Para uma melhor análise dos resultados obtidos, a Figura 17 apresenta a matriz de confusão obtida, considerando a base de teste, onde se observa que o modelo conseguiu classificar corretamente 1264 imagens de um total de 1311, sendo que como já observado através da tabela 5, a rede CNN sem mecanismo de atenção apresentou melhores resultados de classificação para classe *sem tumor* com 403 imagens classificadas corretamente em um total de 405 imagens e maior dificuldade para o tipo de tumor *meningioma* com 277 imagens classificadas corretamente em um total de 306.

Figura 17 – Matriz de confusão da CNN



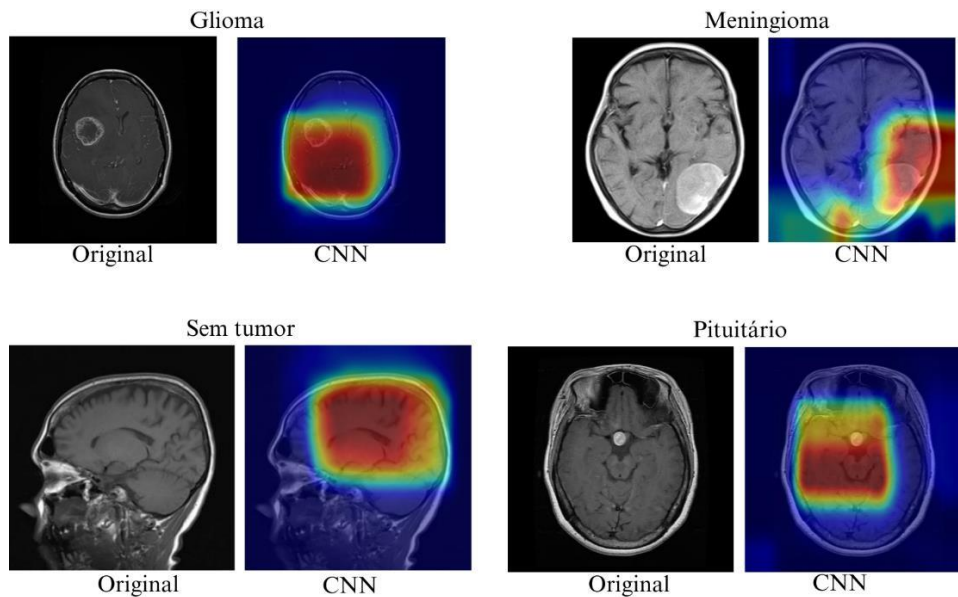
Fonte: Autor (2024)

Na matriz de confusão, também é possível observar que a CNN classificou incorretamente 19 casos como *pituitário* ao invés de *meningioma*. A análise comparativa das imagens geradas por RM, sugeriu que esses equívocos decorreram da interação entre os planos de incidência (crânio-caudal, sagital e axial) utilizados para a obtenção das imagens detalhadas das estruturas internas, a localização dos tumores e o tamanho dos *kernels* da rede. No plano crânio-caudal (visão de cima para baixo), o alinhamento do tumor presente nas meninges com a região hipofisária causou uma possível imprecisão da sua posição real, devido à natureza bidimensional da RM. No plano sagital (visão lateral) e axial (visão anterior ou posterior),

supõe-se que a proximidade dos tumores com a região hipofisária em conjunto com o tamanho dos *kernels* provocou um mal-entendimento da sua região de origem incitando um engano na classificação.

Para uma análise mais detalhada em relação a qual parte das imagens a CNN considera como mais relevante para sua classificação e, para futura comparação com os resultados da CNN com mecanismos de atenção, foi aplicada para algumas imagens da base de teste o Score-CAM. Essa técnica foi aplicada na última camada de convolução da CNN antes de ser entregue a camada *flatten*. A Figura 18 apresenta 4 exemplos de imagens geradas pelo Score-CAM para CNN sem mecanismo de atenção, considerando os 4 tipos de classes.

Figura 18 – Quatro exemplos de Score-CAM para CNN sem mecanismo de atenção



Fonte: Autor (2024)

A CNN buscou destacar as regiões de maior densidade no cérebro nas imagens de exemplo das classes *glioma*, *meningioma* e *pituitário*, pois são a representação do tumor nas imagens de RM apresentadas. Apesar disso, a rede não apresentou um foco especial para a imagem de exemplo da classe *sem tumor*, onde supõe-se que seja pela ausência deste na imagem de RM. Ademais, é válido informar que a rede obteve uma área grande de destaque, onde apesar da presença da massa tumoral nas imagens das classes *glioma*, *meningioma* e *pituitário*, ocorreu também o enfoque de regiões sem a presença de tumores.

4.2 Resultados para CNN com módulo de atenção

A tabela 7 apresenta as métricas globais obtidas para a CNN com mecanismo de atenção.

Tabela 7 – Métricas globais da CNN com atenção

Modelo	Acurácia	Precisão	Sensibilidade	F1-score
CNN com atenção	98.39%	98.31%	98.35%	98.33%

Fonte: Autor (2024)

É possível notar que os resultados ficaram em torno de 98% para todas as métricas, sendo um desempenho consideravelmente bom quando comparado aos modelos híbridos apresentados na literatura atual, avaliados no mesmo conjunto de dados de teste, considerando a menor complexidade da CNN com mecanismo de atenção desenvolvida em relação aos demais.

Para uma análise mais completa dos resultados da CNN com mecanismo de atenção, a tabela 8 apresenta as métricas calculadas por classe.

Tabela 8 – Métricas resultantes da CNN com atenção para cada classe

CNN com atenção	Glioma	Meningioma	Sem tumor	Pituitário
Acurácia	98.00%	97.06%	99.01%	99.33%
Precisão	99.69%	99.31%	99.85%	99.54%
Sensibilidade	99.54%	99.31%	99.69%	99.85%
F1-score	99.61%	99.31%	99.77%	99.69%

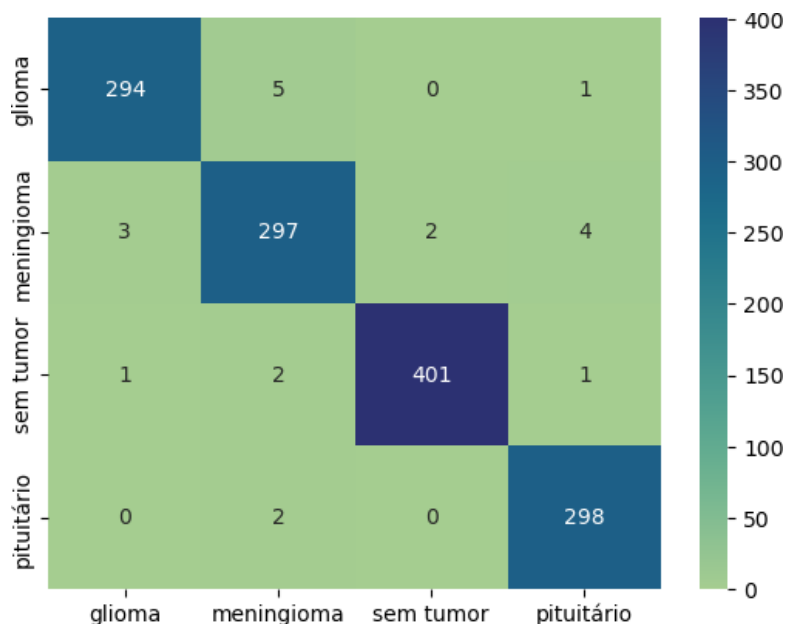
Fonte: Autor (2024)

Conforme pode ser observado na tabela 8, o modelo obteve resultados que podem ser considerados excelentes para todas as classes, principalmente no que diz respeito a métrica sensibilidade, que fornece exatamente a capacidade do classificador indicar um indivíduo para uma classe que realmente ele pertence e a métrica precisão que é útil quando se deseja evitar falsos positivos, nesse caso, evitar que pessoas saudáveis sejam diagnosticadas com tumor cerebral. Os melhores resultados, de forma geral, foram para as classes *sem tumor*, com a métrica precisão atingindo 99.85% e *pituitário* com a métrica sensibilidade atingindo também 99.85%, indicando que o modelo teve maior facilidade para reconhecimento de casos saudáveis e com tumores pituitários. Além disso, foi possível notar que o modelo apresentou resultados de métricas menores para o caso de tumor *meningioma*, com o menor valor das métricas sendo a acurácia de 97.06%.

Para uma melhor análise dos resultados obtidos, a Figura 19 apresenta a matriz de confusão obtida, considerando a base de teste, onde se observa que o modelo conseguiu classificar corretamente 1290 imagens de um total de 1311, sendo que como já observado através da tabela 8, a rede CNN com mecanismo de atenção apresentou melhores resultados de classificação para classe *sem tumor* com 401 imagens classificadas corretamente em um total

de 405 imagens e maior dificuldade para o tipo de tumor *meningioma* com 297 imagens classificadas corretamente em um total de 306.

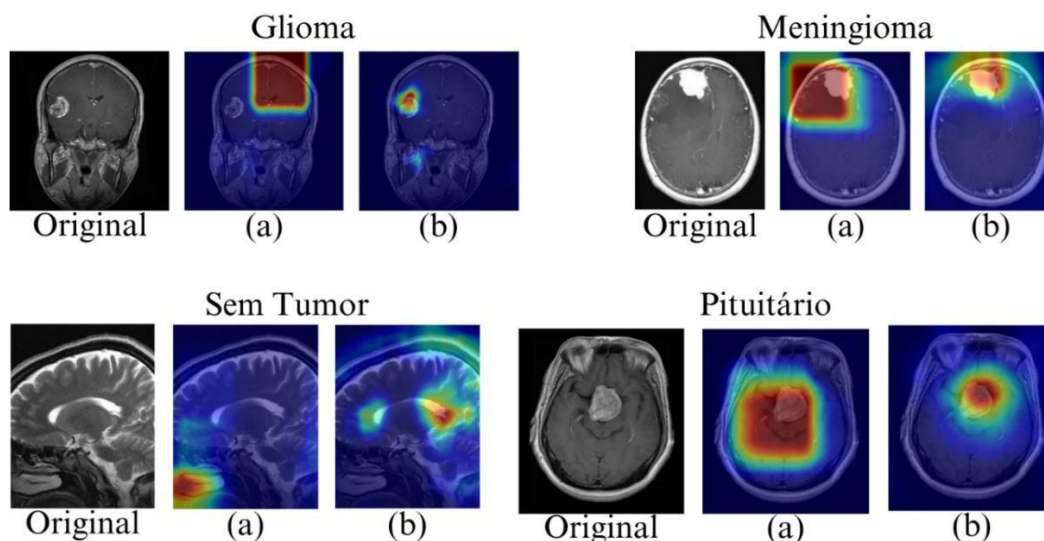
Figura 19 – Matriz de confusão da CNN com atenção



Fonte: Autor (2024)

Para análise mais detalhada em relação a qual parte das imagens a CNN com atenção considera como mais relevante para sua classificação e, para futura comparação com os resultados da CNN sem o mecanismo, foi aplicada para algumas imagens da base de teste o Score-CAM. Essa técnica foi aplicada na última camada de convolução do mecanismo de atenção da CNN antes de ser entregue à camada *flatten* e na última camada de convolução antes do mecanismo de atenção. A Figura 20 apresenta 4 exemplos de imagens geradas pelo Score-CAM para CNN com mecanismo de atenção, considerando os 4 tipos de classes.

Figura 20 – Score-CAM da CNN para cada classe de tumores antes (a) e depois (b) do mecanismo de atenção



Fonte: Autor (2024)

A CNN com mecanismo de atenção procurou destacar as regiões mais densas do cérebro nas imagens de exemplo das classes *glioma*, *meningioma* e *pituitário*, já que essas representam o tumor nas imagens de RM. No entanto, para a imagem de exemplo da classe *sem tumor*, a rede não apresentou um foco específico, o que provavelmente se deve à ausência de uma massa tumoral. Além disso, é possível destacar que após a adição do mecanismo de atenção a região de destaque diminuiu nas imagens das classes *glioma*, *meningioma* e *pituitário*, ignorando partes do cérebro que não constam a presença da massa tumoral.

5 DISCUSSÃO DOS RESULTADOS

Esse capítulo apresenta uma discussão sobre todos os resultados obtidos a partir das duas CNNs treinadas para classificação de tumores cerebrais baseadas em imagens de ressonância magnética. O principal objetivo da análise de resultados e discussão é verificar o impacto do mecanismo de atenção no desempenho da CNN visando identificar se o uso desse pode trazer melhorias nos resultados considerando as métricas de desempenho usadas para análise dos classificadores desenvolvidos. Será também realizada uma análise comparativa dos resultados obtidos a partir dos dois classificadores desenvolvidos com os resultados de classificadores já apresentados na literatura.

5.1 Comparação de resultados entre CNN sem e com mecanismo de atenção

A tabela 9 apresenta as métricas globais obtidas para a CNN sem mecanismo de atenção e com mecanismo de atenção. A partir dos resultados obtidos pode ser observado que a adição do mecanismo de atenção trouxe melhorias no desempenho da CNN. A CNN com mecanismo de atenção apresentou melhores resultados considerando todas as métricas, apresentado um aumento percentual de 1.98% de acurácia, 2.07% de precisão, 2.18% de sensibilidade e 1.72% de F1-score.

Tabela 9 – Comparação de métricas entre os modelos de redes convolucionais com e sem atenção

Modelo	Acurácia	Precisão	Sensibilidade	F1-score
CNN	96.41%	96.24%	96.17%	96.61%
CNN com atenção	98.39%	98.31%	98.35%	98.33%

Fonte: Autor (2024)

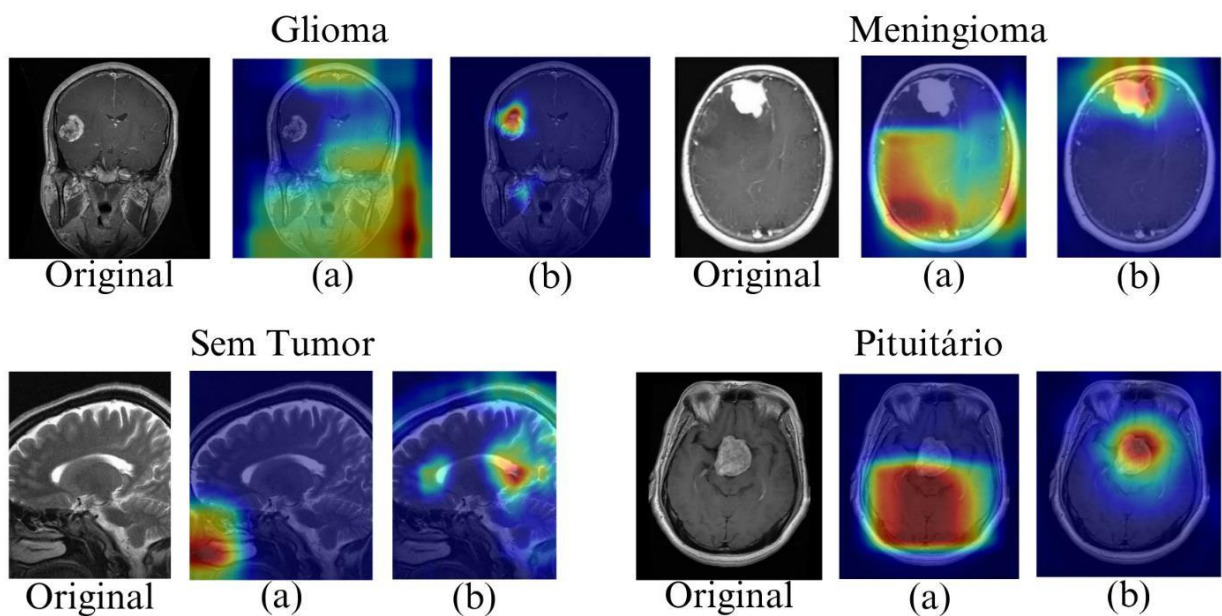
Além disso, ao realizar a comparação das tabelas 6 e 8, que nos apresentam as métricas por classe para CNN e CNN com mecanismos de atenção, foi possível verificar que o modelo obteve melhorias nas métricas comparativas após a adição do mecanismo de atenção. Dentre essas evoluções notadas é válido destacar as mudanças nas métricas da classe *meningioma*, visto que essa trouxe maior dificuldade de classificação para a CNN sem mecanismo de atenção. Ao analisar-se as métricas calculadas para os modelos desenvolvidos relacionadas ao caso de *meningioma*, foram identificados aumento percentuais de 9.54% de acurácia, 0.33% de precisão, 1.55% de sensibilidade e 0.94% de F1-score.

Considerando agora as matrizes de confusão dos dois modelos ocorreu um aumento de 26 imagens corretamente classificadas pelo modelo após a adição do mecanismo de atenção na sua estrutura, pode-se destacar melhorias nos resultados das classes de *glioma* e *meningioma*,

onde obteve-se um aumento de 286 para 294 imagens em *glioma* e de 277 para 297 em *meningioma*, apesar de que a classe *sem tumor* apresentou uma diminuição de duas imagens corretamente classificadas, o que confirmou os resultados das tabelas 5 e 6.

Dando continuidade ao processo comparativo através do exemplo da Figura 21, a análise dos Score-CAMs gerados para as redes desenvolvidas indica que a rede sem o mecanismo de atenção focou em uma área maior de informação da imagem, enquanto a rede com o mecanismo focou em regiões menores e mais específicas para cada tipo de tumor.

Figura 21 – Score-CAM para cada classe de tumores da CNN sem mecanismo de atenção (a) e com mecanismo de atenção (b)



Fonte: Autor (2024) atenção

A partir dos resultados comparativos obtidos, é possível concluir que a inclusão do mecanismo de atenção, na estrutura da CNN, trouxe melhorias em seu desempenho geral para a classificação de tumores cerebrais, dando maior foco nas regiões de interesse das imagens para classificação dos tumores específicos pela CNN, minimizando a dificuldade apresentada pelo modelo sem mecanismo de atenção para classificação de determinados casos na base de dados, como para a classe de tumor *meningioma*.

5.2 Comparação dos modelos com resultados da literatura

No contexto de comparar os resultados desse estudo com relação a modelos do estado-da-arte da literatura atual, uma pesquisa foi realizada em busca de trabalhos que utilizaram a mesma base de dados usada neste estudo. A tabela 8 apresenta os resultados obtidos pelos trabalhos selecionados sendo importante mencionar que uma comparação direta dos resultados

é válida visto que os trabalhos utilizaram a mesma base de dados de teste composta por 1311 imagens.

Tabela 10 – Comparação dos modelos com trabalhos da literatura atual

Modelos	Acurácia	Precisão	Sensibilidade	F1-score
ResNet-CBAM [46]	99.43%	98.70%	99.00%	99.00%
ConvAttenMixer [47]	97.94%	98.58%	95.27%	96.65%
CNN (proposta)	96.41%	96.24%	96.17%	96.61%
CNN com atenção (proposta)	98.39%	98.31%	98.35%	98.33%

Fonte: Autor (2024)

Em vista das métricas apresentados na tabela 10, podemos observar que nossos resultados são competitivos com resultados apresentados na literatura atual. Em [46] foi utilizada a rede *ResNet50*, que é uma CNN robusta e complexa, usada para aplicações de visão computacional e em [47] foram utilizados dois tipos de mecanismo de atenção com estrutura mais complexa que a utilizada neste estudo e dois tipos de operações de convolução. Assim, é possível afirmar que ambos modelos utilizaram estruturas de CNN mais complexas que a desenvolvida neste estudo, que contém apenas um tipo de mecanismo de atenção, onde o modelo de CNN com atenção proposto, apesar de mais simples e apresentando menor complexidade computacional apresentou resultados similares e até superiores aos dois modelos mais complexos, podendo então ser uma boa opção para o desenvolvimento de classificadores a serem utilizados em sistemas de diagnóstico inteligentes de tumores cerebrais para auxílio ao diagnóstico médico.

6 CONCLUSÃO

Tumores cerebrais são enfermidades complexas e de difícil diagnóstico, com a ressonância magnética sendo o exame complementar padrão-ouro para o seu diagnóstico clínico, visto que fornece imagens mais limpas do cérebro que outros exames de imagens.

Dentro deste contexto, o objetivo deste trabalho foi apresentar um estudo comparativo entre uma CNN com e sem mecanismo de atenção com o objetivo de analisar a influência do mecanismo de atenção *self attention*, utilizado nos modelos *Transformers*, nos resultados da classificação de tumores cerebrais por CNNs utilizando imagens de ressonância magnética do cérebro.

Após o treino e o teste de ambas as redes, baseando-se na avaliação das métricas comparativas calculadas no conjunto de teste, concluímos que o uso do mecanismo de atenção trouxe melhorias no desempenho do modelo, onde foi observado que o modelo com o mecanismo de atenção mostrou resultados superiores nas 4 métricas (acurácia, precisão, sensibilidade e F1-score). É importante salientar que o mecanismo de atenção torna o modelo mais complexo, dessa forma aumenta o tempo de processamento em comparação à CNN sem o mecanismo. Apesar dos resultados promissores desta dissertação para a tarefa em questão é importante salientar que as conclusões obtidas são limitadas à classificação de tumores cerebrais e ao conjunto de dados utilizados, ou seja, o estudo desenvolvido foi capaz de determinar a capacidade do mecanismo de atenção de aumentar o desempenho de redes convolucionais para este caso específico, sendo que para outros conjuntos de dados para classificação de tumores ou outras aplicações é importante a realização de estudos mais profundos para definir a capacidade do mecanismo de atenção levando em consideração as especificidades de cada tarefa ou problema de aplicação.

Finalmente, esses resultados trazem uma boa perspectiva para o desenvolvimento de modelos baseados em CNNs, abrindo novas possibilidades e espaço para estudos buscarem aumentos de desempenho por meio da adição de mecanismos de atenção.

Como forma de trabalhos futuros, é visível a implementação de CNNs com outras configurações de mecanismos de atenção para que uma análise mais detalhada possa ser feita sobre a influência de cada tipo de mecanismo nos resultados para o problema de classificação de tumores cerebrais.

REFERÊNCIAS

1. DAS, S. et al. Efficient Net-B6 model-based Transfer Learning for The Classification of Brain Tumors. **Proceedings of the 5th International Conference on Communication and Information Processing**, Outubro 2023. DOI 10.2139/ssrn.4607794. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4607794. Acesso em: 11 set 2024.
2. RAHMAN, T.; ISLAM, M. S. MRI brain tumor detection and classification using parallel deep convolutional neural networks. **Measurement: Sensors**, v. 26, n. 100694, Abril 2023. DOI 10.1016/j.measen.2023.100694. Disponível em: https://www.sciencedirect.com/science/article/pii/S2665917423000302?_cf_chl_tk=v8V4xlAKEMbRmhqL_354VYtsBgq2ZjvAptfUmyErTl8-1726419391-0.0.1.1-6036. Acesso em: 11 set 2024.
3. ÖZKARACA, O. et al. Multiple Brain Tumor Classification with Dense CNN Architecture Using Brain MRI Images. **Life**, v. 13, n. 2, Janeiro 2023. DOI 10.3390/life13020349. Disponível em: <https://www.mdpi.com/2075-1729/13/2/349>. Acesso em: 11 set 2024.
4. FARHAN, A. S.; KHALID, U. M. M. PRCnet: An Efficient Model for Automatic Detection of Brain Tumor in MRI Images. 2023. DOI 10.1101/2023.09.28.560042. Disponível em: <https://www.biorxiv.org/content/10.1101/2023.09.28.560042v1#page>. Acesso em: 11 set 2024.
5. IOANNIS D. Apostolopoulos; Sokratis Aznaouridis; Mpesi Tzani. An Attention-Based Deep Convolutional Neural Network for Brain Tumor and Disorder Classification and Grading in Magnetic Resonance Imaging. **Information**, v. 14, n. 3, Março 2023. DOI 10.3390/info14030174. Disponível em: <https://www.mdpi.com/2078-2489/14/3/174>. Acesso em: 11 set 2024.
6. LIU, M.; YANG, J. Image Classification of Brain tumor based on Channel Attention Mechanism. **Journal of Physics: Conference Series**, v. 2035, Setembro 2021. DOI 10.1088/1742-6596/2035/1/012029. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/2035/1/012029>. Acesso em: 11 set 2024.

7. BODAPATI, J. D. et al. Joint training of two-channel deep neural network for brain tumor classification. **Signal, Image and Video Processing**, v. 15, p. 753 – 760, Outubro 2020. DOI 10.1007/s11760-020-01793-2. Disponível em: <https://link.springer.com/article/10.1007/s11760-020-01793-2>. Acesso em: 12 set 2024.

8. JOHNS HOPKINS MEDICINE. **Brain Anatomy and How the Brain Works**. Disponível em: <https://www.hopkinsmedicine.org/health/conditions-and-diseases/anatomy-of-the-brain>. Acesso em: 20 maio 2024.

9. MAYO CLINIC. **How your brain works**. 2024. Disponível em: <https://www.mayoclinic.org/diseases-conditions/epilepsy/in-depth/brain/art-20546821>. Acesso em: 20 maio 2024.

10. OLIVEIRA, M. **Lobos Cerebrais**. Disponível em: <https://www.infoescola.com/anatomia-humana/lobos-cerebrais/>. Acesso em: 01 jul 2024.

11. MAYO CLINIC. **Brain tumor**. 2023. Disponível em: <https://www.mayoclinic.org/diseases-conditions/brain-tumor/symptoms-causes/syc-20350084>. Acesso em: 21 maio 2024.

12. NICKPARVAR, M. **Brain Tumor MRI Dataset**. 2021. Disponível em: <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>. Acesso em: 29 maio 2024.

13. MAYO CLINIC. **Brain tumor**. 2023. Disponível em: <https://www.mayoclinic.org/diseases-conditions/brain-tumor/diagnosis-treatment/drc-20350088>. Acesso em: 21 maio 2024.

14. JOHNS HOPKINS MEDICINE. **Magnetic Resonance Imaging (MRI)**. Disponível em: <https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/magnetic-resonance-imaging-mri>. Acesso em: 22 maio 2024.

15. MEDICINA DIAGNÓSTICA. **Inauguração da nova sala de Ressonância Magnética**. Disponível em: <https://www.medicinadiagnostica.com.br/novidades-clinica-diagnostico-por-imagem/inauguracao-da-nova-sala-de-ressonancia-magnetica>. Acesso em: 01 jul 2024.

16. GOOGLE CLOUD. **What is Machine Learning (ML)?** Disponível em: <https://cloud.google.com/learn/what-is-machine-learning>. Acesso em: 22 maio 2024.
17. IBM. **O que é aprendizado de máquina (ML)?** Disponível em: <https://www.ibm.com/br-pt/topics/machine-learning>. Acesso em: 22 maio 2024.
18. AMAZON. **What is a Neural Network?** Disponível em: <https://aws.amazon.com/what-is/neural-network/#:~:text=A%20neural%20network%20is%20a,that%20resembles%20the%20human%20brain>. Acesso em: 22 maio 2024.
19. MICROSOFT. **What is computer vision?** Disponível em: <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-computer-vision#object-classification>. Acesso em: 22 maio 2024.
20. MATHWORKS. **What is computer vision?** Disponível em: <https://www.mathworks.com/discovery/convolutional-neural-network.html>. Acesso em: 23 maio 2024.
21. NVIDIA. **Convolutional Neural Network (CNN).** Disponível em: <https://developer.nvidia.com/discover/convolutional-neural-network>. Acesso em: 23 maio 2024.
22. NVIDIA. **Deep Learning in a Nutshell: Core Concepts.** Disponível em: <https://developer.nvidia.com/blog/deep-learning-nutshell-core-concepts/>. Acesso em: 19 jun 2024.
23. YAMASHITA, R. et al. Convolutional neural networks: an overview and application in radiology. **Insights into Imaging**, v. 9, n. 4, p. 611 – 629, Junho 2018. DOI 10.1007/s13244-018-0639-9. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/29934920/>. Acesso em: 01 jul 2024.
24. IBM. **What are convolutional neural networks?** Disponível em: <https://www.ibm.com/topics/convolutional-neural-networks>. Acesso em: 23 maio 2024.

25. GHOLAMALINEZHAD, H.; KHOSRAVI, H. Pooling Methods in Deep Neural Networks, a Review. 2020. DOI 10.48550/arXiv.2009.07485. Disponível em: <https://arxiv.org/abs/2009.07485>. Acesso em: 01 jul 2024.

26. SRIVASTAVA, N. et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. **Journal of Machine Learning Research**, Toronto, v. 15, n. 56, p. 1929 – 1958, Junho 2014. Disponível em: <https://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf>. Acesso em: 19 jun 2024.

27. PAPANASTASIOU, G. et al. Is attention all you need in medical image analysis? A review. **IEEE Journal of Biomedical and Health Informatics**, v. 28, n. 3, p. 1398 – 1411, Janeiro 2024. DOI 10.1109/JBHI.2023.3348436. Disponível em: <https://ieeexplore.ieee.org/document/10376277>. Acesso em: 28 ago 2024.

28. NIU, Z.; ZHONG, G.; YU, H. 28. A review on the attention mechanism of deep learning. *Neurocomputing*, v. 452, p. 48 – 62, Abril 2021. DOI 10.1016/j.neucom.2021.03.091. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S092523122100477X>. Acesso em: 06 set 2024.

29. VASWANI, A. et al. Attention is All you Need. In: ANAIS, 2017, Long Beach. **Neural Information Processing Systems**. Long Beach: Curran Associated Inc., 2017. v. 30, p. 6000 – 6010. ISBN 978-1-51086096-4. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Acesso em: 24 maio 2024.

30. MINH, A. P. T. Overview of Class Activation Maps for Visualization Explainability. 2023. DOI 10.48550/arXiv.2309.14304. Disponível em: <https://arxiv.org/abs/2309.14304>. Acesso em: 09 set 2024.

31. WANG, H. et al. Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks. In: IEEE; CVF (org.). **IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops**. 2020. (IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops), p. 111 – 119. DOI 10.1109/CVPRW50498.2020.00020. Disponível em: <https://ieeexplore.ieee.org/document/9150840>. Acesso em: 09 set 2024.

32. KHAN, M. A.; PARK, H. A Convolutional Block Base Architecture for Multiclass Brain Tumor Detection Using Magnetic Resonance Imaging. **Electronics**, v. 13, n. 2, Janeiro 2024. DOI 10.3390/electronics13020364. Disponível em: <https://www.mdpi.com/2079-9292/13/2/364>. Acesso em: 09 set 2024.

33. REMZAN, N.; KARIM, T.; FARCHI, A. Automatic Classification of Preprocessed MRI Brain Tumors Images Using Deep Convolutional Neural Networks. **International Journal on “Technical and Physical Problems of Engineering”**, v. 15, n. 1, p. 68 – 73, Março 2023. Disponível em: https://www.researchgate.net/publication/371304477_AUTOMATIC_CLASSIFICATION_OF_PREPROCESSED_MRI_BRAIN_TUMORS_IMAGES_USING_DEEP_CONVOLUTIONAL_NEURAL_NETWORK. Acesso em: 09 set 2024.

34. KAYA, M. Bayesian Optimization-based CNN framework for Automated Detection of Brain Tumors. **Balkan Journal of Electrical & Computer Engineering**, v. 11, n. 4, p. 394 – 403, Outubro 2023. DOI 10.17694/bajece.1346818. Disponível em: <https://dergipark.org.tr/en/pub/bajece/issue/81549/1346818>. Acesso em: 09 set 2024.

35. AL-ZOGHBY, A. M. et al. Dual Deep CNN for Tumor Brain Classification. **Diagnostics**, v. 13, n. 12, Junho 2023. DOI 10.3390/diagnostics13122050. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/37370945/>. Acesso em: 09 set 2024.

36. GÜLER, M.; NAMLI, E. Brain Tumor Detection with Deep Learning Methods’ Classifier Optimization Using Medical Images. **Applied Sciences**, v. 14, n. 2, Janeiro 2024. DOI 10.3390/app14020642. Disponível em: <https://www.mdpi.com/2076-3417/14/2/642>. Acesso em: 09 set 2024.

37. TÜZÜN, B. N.; ÖZDEMİR, D. Classification of Brain Tumors with Deep Learning Models. **Journal of Scientific Reports-A**, n. 54, Setembro 2023. DOI 10.59313/jsr-a.1293119. Disponível em: <https://dergipark.org.tr/en/pub/jsr-a/issue/78798/1293119>. Acesso em: 09 set 2024.

38. JUN, W.; LIYUAN, Z. Brain Tumor Classification Based on Attention Guided Deep Learning Model. **International Journal of Computational Intelligence Systems**, v. 15, Maio 2022. DOI 10.1007/s44196-022-00090-9. Disponível em:

- <https://link.springer.com/article/10.1007/s44196-022-00090-9#citeas>. Acesso em: 10 set 2024.
39. ABDUSALOMOV, A. B.; MUKHIDDINOV, M.; WHANGBO, T. K. Brain Tumor Detection Based on Deep Learning Approached and Magnetic Resonance Imaging. **Cancers**, v. 15, n. 16, Agosto 2023. DOI [10.3390/cancers15164172](https://doi.org/10.3390/cancers15164172). Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10453020/>. Acesso em: 10 set 2024.
 40. TANG, C. et al. GAM-SpCaNet: Gradient awareness minimization-based spinal convolution attention network for brain tumor classification. *Journal of King Saud University Computer and Information Sciences*, v. 35, n. 2, p. 560 – 575, Fevereiro 2023. DOI [10.1016/j.jksuci.2023.01.002](https://doi.org/10.1016/j.jksuci.2023.01.002). Disponível em: <https://www.sciencedirect.com/science/article/pii/S1319157823000022?via%3Dihub>. Acesso em: 10 set 2024.
 41. BHUVAJI, S. **Brain Tumor Classification (MRI)**. 2020. 111 – 119 p. Disponível em: <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>. Acesso em: 29 maio 2024.
 42. HAMADA, A. **Br35H :: Detecção de Tumor Cerebral 2020**. 2020. 111 – 119 p. Disponível em: <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection?select=no>. Acesso em: 29 maio 2024.
 43. FIGSHARE. **brain tumor dataset**. 2017. Disponível em: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427. Acesso em: 29 maio 2024.
 44. BAFFOUR, A. A. et al. Spatial self-attention network with self-attention distillation for finegrained image recognition. **Journal of Visual Communication and Image Representation**, v. 81, Novembro 2021. DOI [10.1016/j.jvcir.2021.103368](https://doi.org/10.1016/j.jvcir.2021.103368). Disponível em: <https://www.sciencedirect.com/science/article/pii/S104732032100242X>. Acesso em: 10 set 2024.
 45. FRANÇOIS-GUILLAUME FERNANDEZ. **TorchCAM: class activation explorer**. 2020. Disponível em: <https://github.com/frgfm/torch-cam>. Acesso em: 11 set 2024.

46. OLADIMEJI, O. O.; IBITOYE, A. O. J. Brain tumor classification using ResNet50convolutional block attention module. **Applied Computing and Informatics**, Dezembro 2023. DOI 10.1108/ACI-09-2023-0022. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/ACI-09-2023-0022/full/html>. Acesso em: 28 ago 2024.
47. ALZHRANI, S. M. ConvAttenMixer: Brain tumor detection and type classification using convolutional mixer with external and self-attention mechanisms. **Journal of King Saud University - Computer and Information Sciences**, v. 35, n. 10, Dezembro 2023. DOI 10.1016/j.jksuci.2023.101810. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1319157823003646>. Acesso em: 28 ago 2024.