CBIS'24
XX Congresso Brasileiro de Informática em Saúde
08/10 a 11/10 de 2024 - Belo Horizonte/MG - Brasil

Assessing attention mechanisms' impact on automatic brain tumor classification

Avaliando o impacto de mecanismos de atenção na classificação automática de tumores cerebrais

Evaluando el impacto de mecanismos de atención en la clasificación automática de tumores cerebrales

Caio dos Santos Felipe[1], Thatiane Alves Pianoschi Alva[2], Carla Diniz Lopes Becker[2]

1 Undergraduate Student, Federal University of Health Sciences of Porto Alegre – UFCSPA, Porto Alegre (RS), Brazil.
2 Ph.D., Federal University of Health Sciences of Porto Alegre – UFCSPA, DECESA, Porto Alegre (RS), Brazil.

Corresponding author: Caio dos Santos Felipe
E-mail: caio.felipe@ufcspa.edu.br

Abstract

Objective: To compare a conventional convolutional neural network model and its attention-enhanced counterpart. **Method:** We trained both models on the same dataset images of gliomas, meningiomas, pituitary adenomas, and non-tumorous images; then, we compared both models using interpretable approaches, highlighting the regions used for their predictions. **Results:** Our analysis found that the attention-enhanced model focused more on tumor regions, with 99% accuracy. **Conclusion:** The outcome of this research underscores the importance of continued exploration into advanced neural network features to elevate the standards of diagnostic accuracy and efficiency in medical practice.
Keywords: Brain Tumor; Deep Learning; Convolutional Neural Network

Resumo

Objetivo: Comparar um modelo convencional de rede neural convolucional e sua versão melhorada com atenção. **Método:** Treinamos ambos os modelos no mesmo conjunto de dados contendo imagens de gliomas, meningiomas, adenomas pituitários e imagens não tumorais; em seguida, comparamos os modelos usando abordagens interpretáveis,

destacando as regiões usadas para suas previsões. **Resultados:** Nossa análise descobriu que o modelo com realce de atenção focou mais nas regiões tumorais, com 99% de acurácia. **Conclusão:** O resultado desta pesquisa sublinha a importância da exploração contínua de características avançadas de redes neurais para elevar os padrões de precisão diagnóstica e eficiência na prática médica.
Descritores: Tumor Cerebral; Aprendizado Profundo; Redes Neurais Convolucionais

Resumen

Objetivo: Comparar un modelo convencional de red neuronal convolucional y su versión mejorada con atención. **Método:** Entrenamos ambos modelos en el mismo conjunto de datos que contiene imágenes de gliomas, meningiomas, adenomas pituitarios e imágenes no tumorales; luego, comparamos los modelos utilizando enfoques interpretativos, destacando las regiones utilizadas para sus predicciones. **Resultados:** Nuestro análisis descubrió que el modelo con realce de atención se centró más en las regiones tumorales, con precisión del 99%. **Conclusión:** El resultado de esta investigación subraya la importancia de la exploración continua de características avanzadas de redes neuronales para elevar los estándares de precisión diagnóstica y eficiencia en la práctica médica.
Descriptores: Tumor Cerebral; Aprendizaje Profundo; Redes Neuronales Convolucionales

Introduction

The advent of deep learning technologies, particularly Convolutional Neural Networks (CNNs), has transformed the landscape of medical image analysis, yielding unprecedented advancements in the automated detection and classification of diseases. Among these, the application of CNNs for brain tumor classification is a critical area of exploration due to its implications for enhancing diagnostic accuracy and efficiency, thereby influencing patient management and treatment outcomes. Despite the notable successes, the inherent complexity and high variability of brain tumor imaging pose substantial challenges that can impair the performance of conventional CNN architectures. (1)

In this context, attention mechanisms emerge as a promising enhancement to CNN architectures, aiming to bolster their performance by enabling a focused analysis of salient image regions and features crucial for accurate classification. By dynamically allocating computational resources to the most informative parts of an image, attention mechanisms aspire to refine the model's feature extraction capabilities, thereby improving the specificity and sensitivity of the classification process.

This paper examines how attention mechanisms can augment the efficacy of CNNs in brain tumor classification, focusing on comparing an attention-augmented CNN model and a standard CNN architecture. In conducting this thorough comparative analysis, the evaluation criteria extend beyond traditional performance indicators to incorporate Explainable Artificial Intelligence (XAI) methods.

Alongside Gradient-weighted Class Activation Mapping (Grad-CAM), we will also employ the Local Interpretable Model-agnostic Explanations (LIME) technique. LIME will be used to evaluate the model focusing on interpretability, providing insights into the model's decision-making process by approximating it locally with an interpretable model. Grad-CAM provides visual explanations for decisions made by CNNs, highlighting the specific regions within the input images that contribute most significantly to the model's predictions [2, 3]. Grad-CAM has been applied across multiple disciplines and is extensively used as a primary method to enhance explainability and transparency in CNN models for image classification, with notably predominant usage in the healthcare sector [4, 5]. This dual approach allows for a more comprehensive understanding of model behavior, facilitating a deeper exploration into the strengths and limitations of the attention mechanisms employed.

The principal contribution of this research lies in its in-depth analysis of the role of attention mechanisms in enhancing CNN models for brain tumor classification. By presenting empirical evidence on the performance benefits of these mechanisms, the paper aims to shed light on their potential to refine and advance the capabilities of deep learning models in medical image analysis, demonstrating their practical applicability and effectiveness in a healthcare setting.

Related Work

In recent advancements, Alzahrani SM [6] proposed an attention-based CNN model incorporating Squeeze-and-Excitation (SE) blocks, achieving a notable accuracy of 97.94% applied to the same ”Brain Tumor MRI Dataset” [7] referenced in this article. In another study, Jun W *et al*. [8] also contributed to the field by proposing an attention-based CNN that employs a Dual-Attention Network Architecture, achieving an accuracy of 98.6%.

It is important to notice that visualization techniques such as Grad-CAM or LIME received limited attention in the works previously proposed, highlighting a gap in developing fully interpretable models for real-world medical applications, where understanding model decision processes is crucial.

Materials and Methods

A CNN is a type of artificial neural network optimized for processing and analyzing grid-like data, such as images, speech signals, or any data with spatial or temporal patterns [1]. The architecture of a CNN consists of several layers, including convolutional layers, pooling layers, and fully connected layers, each serving a specific function to help the network learn hierarchical features from input data [1]. Multiple techniques, including attention mechanisms, are currently being developed and tested to enhance the performance and interpretability of CNNs.

At its core, an attention mechanism in deep learning mimics the human ability to focus on specific parts of an input while ignoring others. In the specific context of brain tumor classification, focusing on tumor regions within Magnetic Resonance Imaging (MRI) images can enhance the model's diagnostic performance, ensuring predictions based on the most relevant information from vast and complex datasets.

In the pursuit of harnessing the full potential of attention mechanisms for enhancing the diagnostic capabilities of CNNs in medical imaging, this paper explores explicitly implementing the SE technique as the attention mechanism of choice. The SE technique represents a sophisticated approach to adaptively recalibrating channel-wise feature responses in CNNs, effectively allowing the model to emphasize informative features while suppressing less useful ones [9]. By integrating SE blocks into the CNN architecture, the

model can conduct a more focused analysis of the input images, dynamically adjusting the importance of each channel based on the global information contained within the feature maps [9].

Dataset Overview and Preprocessing

Our study leveraged the "Brain Tumor MRI Dataset" [7] comprising 7023 images, categorized into a training set with 5712 images and a testing set of 1311 images. We sorted the images into four categories: glioma, meningioma, notumor, and pituitary, based on the tumor type or absence thereof.

We identified inaccuracies in the glioma category labeling during the data preparation phase. Mislabeled images were replaced with accurately labeled ones from the "Figshare Brain Tumor Classification" dataset [10], ensuring the integrity of our data. Post-adjustment, the training dataset included 1140 glioma images (20.61%), 1339 meningioma images (24.21%), 1595 notumor images (28.84%), and 1457 pituitary images (26.34%), achieving a balanced mix for model training.

The study ultimately utilized 5531 images for training and 1297 for testing. We earmarked a subset comprising 20% of the training images as a validation set. All images underwent resizing to a standard resolution of 128x128 pixels and normalization to scale pixel values between 0 and 1, enhancing model training efficiency and stability. We also applied the Contrast Limited Adaptive Histogram Equalization (CLAHE) technique to enhance the contrast of the images [11].

Data Balance and Leakage

Initial data analysis revealed class distribution imbalances, prompting the implementation of class-specific weighting during model training: glioma (1.21), meningioma (1.03), notumor (0.86), and pituitary (0.94).

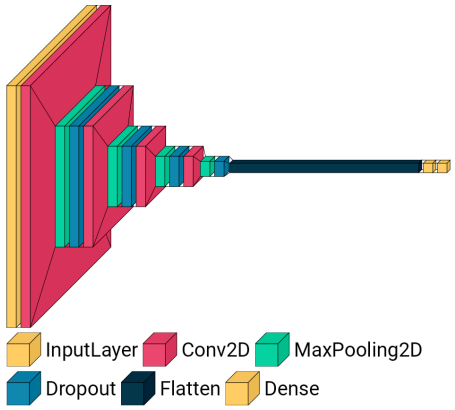
We shuffled data in validation folds to foster model generalization and mitigate bias before initiating training. Additionally, we segregated the dataset into training and validation subsets before normalization to preclude data leakage. We further adopted K-Fold Cross-Validation, safeguarding against the contamination of the training dataset by

the validation data, especially during normalization, thereby ensuring the robustness and validity of our model's performance evaluation.

Base Model Architecture

Our model's design incorporates an architecture consisting of four convolutional blocks, each engineered to process images through a sequence of layers adept at handling various multidimensional data formats. An illustrative depiction of this architecture is presented in Figure 1.

Figure 1 – Visual representation of the base model's architecture.



These convolutional layers are standardized with 3x3 kernels and utilize the ReLU activation function to ensure non-linearity. The configuration of filters and dropout rates across the blocks progresses as follows: We equipped the initial layer with 16 filters and a 20% dropout rate; the subsequent layer comprises 32 filters with a 25% dropout rate, followed by a third layer containing 64 filters at a 30% dropout rate; we designed the fourth layer with 128 filters and a 35% dropout rate. The model leverages fully connected layers for the classification component, with the initial layer hosting 32 neurons utilizing ReLU activation and L2 regularization to combat overfitting. A subsequent layer with four neurons adopts softmax activation to facilitate multi-class classification, incorporating L2 regularization for model robustness.

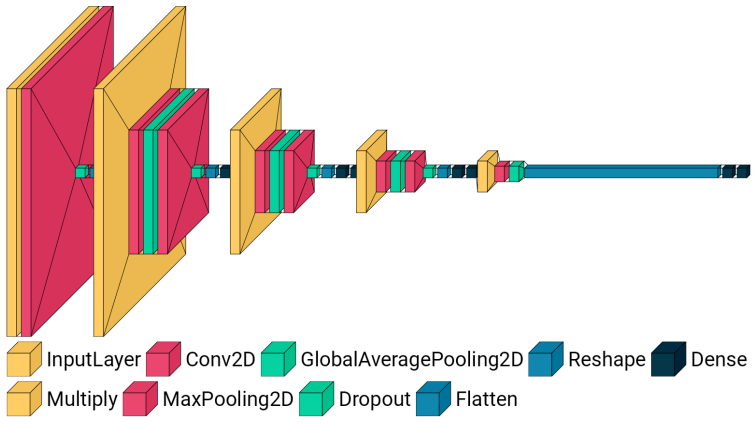
We trained the model for 100 epochs and compiled it using the Adam optimizer with a default learning rate of 0.001. We monitored model convergence using accuracy and sparse categorical cross-entropy as the metrics for the loss function.

Enhanced Model Architecture with Attention Mechanism

To explore the impact of attention mechanisms on model performance, we extended the base model architecture by incorporating SE layers, thus creating an enhanced model variant. This modification maintains the core structure of the original model, ensuring a fair and direct comparison between the standard and attention-augmented architectures.

The enhanced model mirrors the foundational architecture of the standard model. Distinctively, within each convolutional block of the enhanced model, we introduce SE layers. These layers apply an attention mechanism that performs channel-wise feature recalibration, enabling the model to assign adaptive importance to each channel based on the global information extracted from the feature maps. The comprehensive model architecture is visually articulated in Figure 2.

Figure 2 – Visual representation of the attention model's architecture.



Results and Discussions

The comparative analysis of the base model and the attention-enhanced neural network reveals distinct performances across various metrics. Both models exhibited high overall accuracy rates on the test set during validation, with the base model achieving 98% and the attention model slightly outperforming it at 99%.

In terms of precision by class, both models demonstrated exceptional precision, with the base model recording 100% for glioma, 99% for meningioma, 97% for notumor, and 97% for pituitary. The attention model matched or exceeded these results, particularly showing an improvement in identifying notumor cases with a precision rate of 99%, which indicates the attention mechanism's enhanced focus on relevant features for more accurate class predictions. Performance metrics, including precision, recall, F1 score, and accuracy, are detailed in Table 1.

Table 1 – Classification reports for each model, categorized by class.

model		precision	recall	f1-score	support
base model	*glioma*	1.00	1.00	1.00	286
	meningioma	0.99	0.94	0.96	306
	notumor	0.97	1.00	0.99	405
	pituitary	0.97	0.99	0.98	300
	accuracy			0.98	1297
	macro avg	0.98	0.98	0.98	1297
	weighted avg	0.98	0.98	0.98	1297
attention model	*glioma*	1.00	1.00	1.00	286
	meningioma	0.99	0.95	0.97	306
	notumor	0.99	1.00	0.99	405
	pituitary	0.97	0.99	0.98	300
	accuracy			0.99	1297
	macro avg	0.99	0.99	0.99	1297
	weighted avg	0.99	0.99	0.99	1297

The attention model correctly classified 286 glioma, 292 meningioma, 403 no tumor, and 298 pituitary cases. Errors included 5 meningioma cases as no tumor, 9 as pituitary, 2 no tumor cases as meningioma, and 2 pituitary cases as meningioma. The base model correctly classified 286 glioma, 287 meningioma, 404 no tumor, and 297 pituitary cases. Errors included 11 meningioma cases as no tumor, 8 as pituitary, 1 no tumor case as meningioma, 3 pituitary cases as meningioma, and 1 pituitary case as no tumor. The attention model shows fewer misclassifications for meningioma and pituitary cases, indicating a slight advantage in performance over the base model.

Figures illustrating validation loss and accuracy across 100 epochs (Figures 3 and 4) facilitate visual comparisons between the models. These visual representations offer insights into the models' learning processes and adaptability to new datasets.

Figure 3 – Loss and Accuracy graphs of the 5 cross validation folds in the base model.

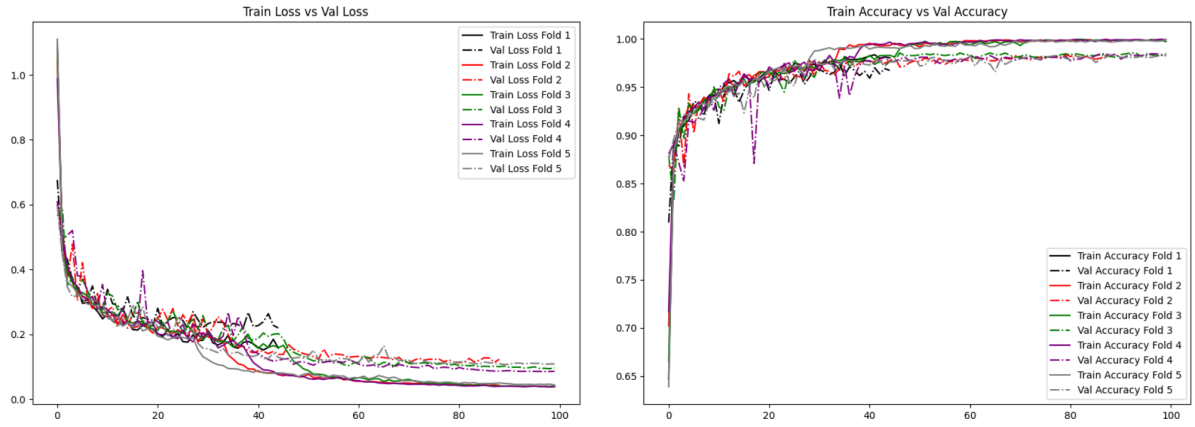
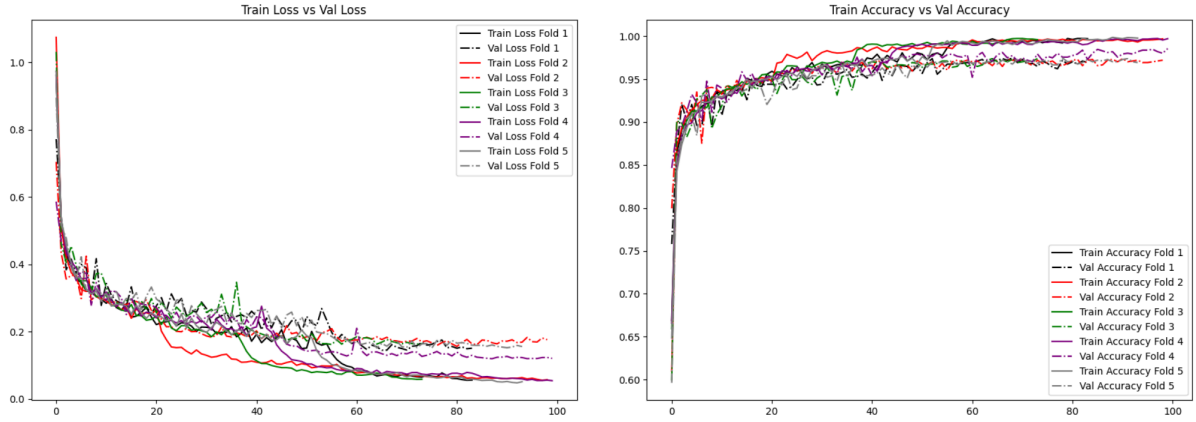


Figure 4 – Loss and Accuracy graphs of the 5 cross validation folds in the attention model.



The Receiver Operating Characteristic (ROC) curve and its Area Under the Curve (AUC-ROC), depicted in Figures 5 and 6, evaluate the models' binary classification performances. A higher AUC-ROC value indicates a superior classification capability. Precision-recall analyses provide further insights into the models' discrimination capabilities and their handling of false positives and negatives.

Figure 5 – ROC and Precision-Recall curve generated from the base model.

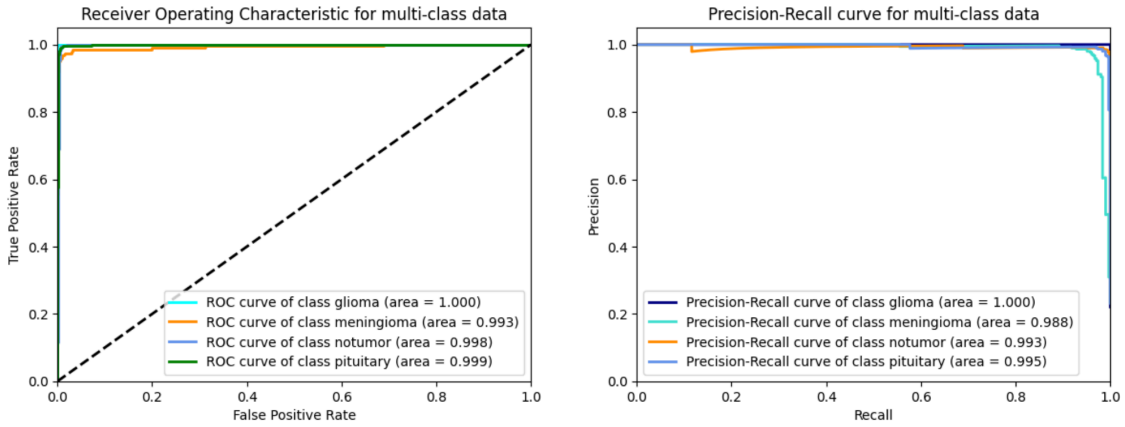
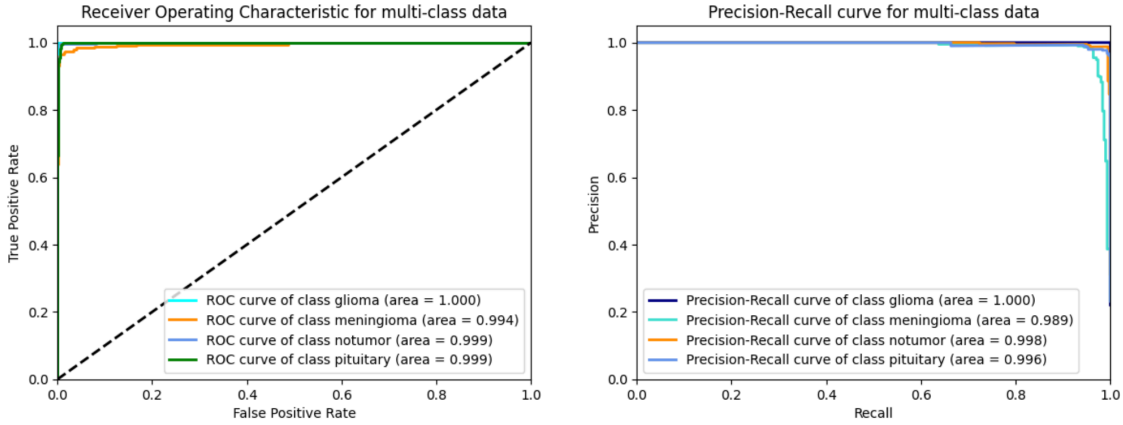
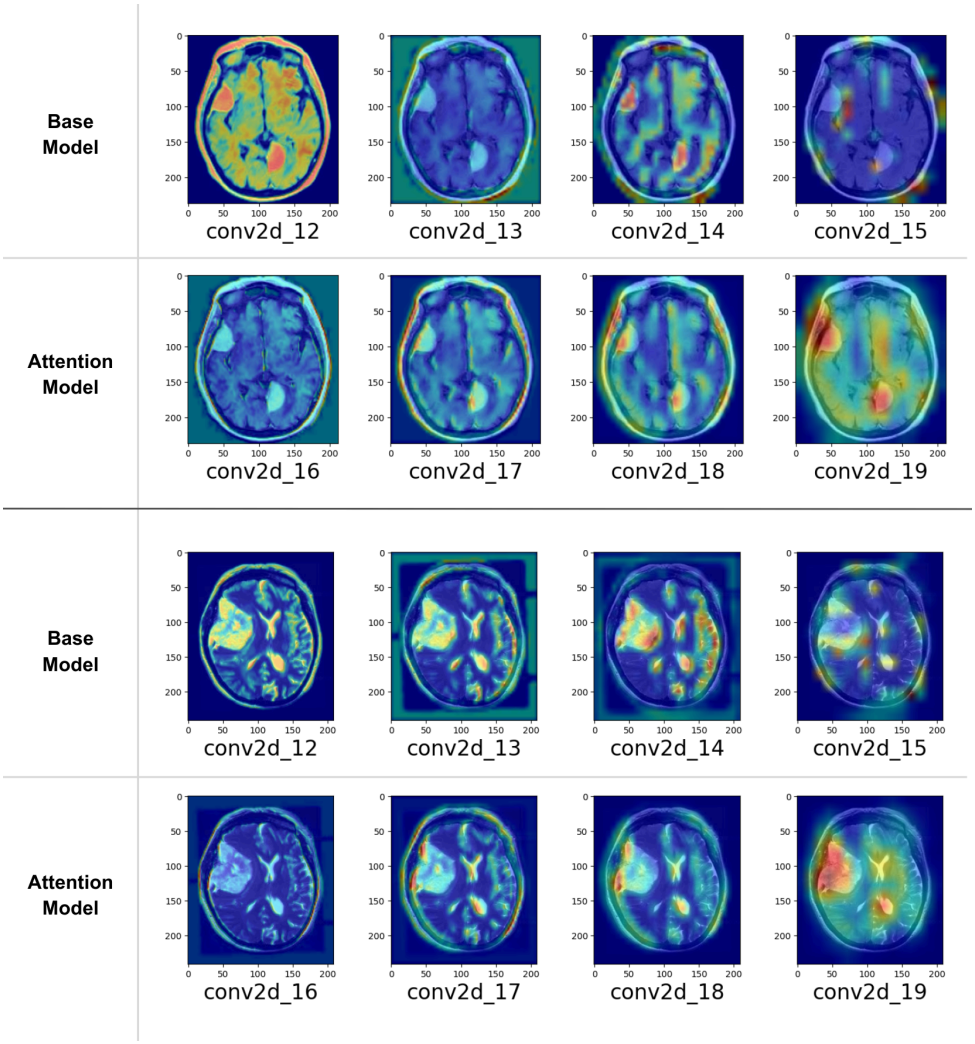


Figure 6 – ROC and Precision-Recall curve generated from the attention model.



Both approaches incorporated Grad-CAM to enhance model interpretability, visualizing influential areas within images at different convolutional layers (Figure 7). Analyzing the last convolutional layers of the attention model versus the base model provides substantive evidence that the attention mechanism focuses more on the actual tumor areas.

Figure 7 – Grad-CAM heatmaps comparing the base model and the attention model across different convolutional layers on two Meningioma MRIs from the ”Brain Tumor MRI Dataset” [7] test set. These images illustrate the differences in focus and highlight areas between the two models.

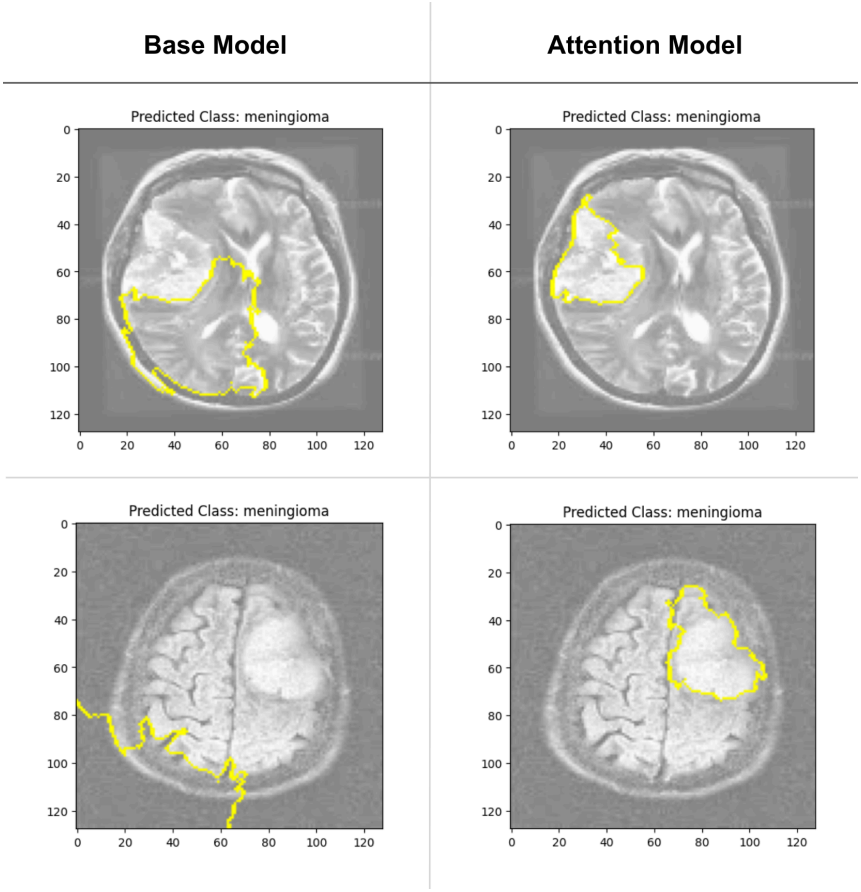


We used the LIME technique to assess the interpretability of both CNN models and set the parameters to focus on the top four classes, allowing us to examine the models' decision-making across multiple likely outcomes. Additionally, we focused our analysis on the most influential feature to provide a detailed view of the primary factor driving the models' classification accuracy.

Both models correctly classified the images, but the LIME analysis uncovered significant differences in their interpretative processes (Figure 8). Although the base model classified the images accurately, it focused on areas not clinically relevant to the tumor features, indicating a potential misalignment between its decision-making process and meaningful pathological markers. In contrast, the attention model not only classified

correctly but also focused on the appropriate tumor regions, aligning its interpretive areas with clinically relevant features. These findings highlight the importance of integrating attention mechanisms into CNN models to improve the reliability and relevance of their interpretative outputs in medical imaging.

Figure 8 – LIME images, generated from the Meningioma images in the "Brain Tumor MRI Dataset" [7] test set, mark the regions in yellow that each model used to classify the images.



Conclusion

In this comprehensive analysis, we juxtapose the performance of two CNN models: one that operates without attention mechanisms and another that leverages these to refine its analytical prowess. Our findings highlight a clear distinction in performance, with the attention-enhanced model demonstrating slightly superior accuracy metrics. Though

modest, this incremental improvement in accuracy is a testament to the nuanced yet impactful role that attention mechanisms play in deep learning.

The enhanced model's ability to selectively concentrate on more relevant features within the input data allows for a more focused and effective learning process. This strategic focus is critical in tasks where the margin for error is minimal, and even slight enhancements in accuracy can have significant implications. Through Grad-CAM and LIME, we observed that the model with attention mechanisms achieved better accuracy and displayed higher precision in identifying and emphasizing the features most relevant to the task at hand.

This comparative study sheds light on the inherent benefits of integrating attention mechanisms into CNNs, particularly in applications where the clarity and preciseness of model interpretations are paramount. The improvement in accuracy underscores the potential of attention mechanisms to fine-tune model performance, marking a step forward in the evolution of machine learning models toward greater efficacy and interpretability.

Referências

1. Felipe C, Alva T, Winck A, Becker C. An approach in brain tumor classification: The development of a new convolutional neural network model. In: Anais do XX Encontro Nacional de Inteligência Artificial e Computacional. Porto Alegre: SBC; 2023. p. 28-42. doi:10.5753/eniac.2023.233530.

2. An J, Joe I. Attention map-guided visual explanations for deep neural networks. Applied Sciences. 2022;12(8):3846. Available from: https://doi.org/10.3390/app12083846

3. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision; 2017 Oct 22-29; Venice, Italy. p. 618-26.

4. Mercaldo F, Brunese L, Martinelli F, Santone A, Cesarelli M. Explainable Convolutional Neural Networks for Brain Cancer Detection and Localisation. Sensors (Basel). 2023;23(17):7614. Published 2023 Sep 2. doi:10.3390/s23177614

5. Hussain T, Shouno H. Explainable Deep Learning Approach for Multi-Class Brain Magnetic Resonance Imaging Tumor Classification and Localization Using Gradient-Weighted Class Activation Mapping. Information. 2023; 14(12):642. https://doi.org/10.3390/info14120642

6. Alzahrani SM. ConvAttenMixer: Brain tumor detection and type classification using convolutional mixer with external and self-attention mechanisms. J King Saud Univ Comput Inf Sci. 2023;35(10):101810. doi: 10.1016/j.jksuci.2023.101810.

7. Nickparvar M. Brain tumor MRI dataset [Data set]. Kaggle. 2021. Available from: https://doi.org/10.34740/KAGGLE/DSV/2645886

8. Jun W, Liyuan Z. Brain Tumor Classification Based on Attention Guided Deep Learning Model. Int J Comput Intell Syst. 2022;15:35. https://doi.org/10.1007/s44196-022-00090-9

9. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018 Jun 18-23; Salt Lake City, UT, USA. p. 7132-41.

10. Figshare Brain Tumor Classification [Data set]. Kaggle. Available from: https://www.kaggle.com/datasets/rahimanshu/figshare-brain-tumor-classification. Last accessed 2023 May 17.

11. Sousa HS, Pereira Neto AA, Paula Júnior IC de, Melo CR de. Segmentação de infecções pulmonares de COVID-19 com a rede Mask R-CNN. J Health Inform [Internet]. 2023 Jul 20 [cited 2024 Mar 22];15(Especial). Available from: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/1100