Improving automatic classification of brain tumors with deep learning techniques

Aprimorando a classificação automática de tumores cerebrais com técnicas de aprendizado profundo

Mejora de la clasificación automática de tumores cerebrales con técnicas de aprendizaje profundo

Willian de Vargas[1] , Dieine Estela Bernieri Schiavon[1] , Viviane Rodrigues Botelho[2] , Thatiane Alves Pianoski[2] , Carla Diniz Lopes Becker[2]

1 Master's 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: Willian de Vargas
E-mail: willianv@ufcspa.edu.br

Abstract

The accuracy in the automatic classification of brain tumors plays a crucial role in the method's reliability for healthcare applications. Classification errors can lead to inaccurate diagnoses, resulting in inappropriate and potentially harmful approaches. **Objective:** To propose an approach aimed at minimizing classification errors. **Method:** We developed a two-stage convolutional neural network model: first, four binary models for tumors with significant differentiation challenges; then, an Ensemble model for multiclass classification. Additionally, we employed a technique to interpret model predictions and identify regions of interest in medical images. **Results:** The proposed approach achieves an accuracy of 98%. **Conclusion:** This work contributes to applying deep learning in brain tumor classification, emphasizing the importance of transparent and robust approaches for precision and safety in predictions.

Keywords: Brain tumors; Automatic Classification; Deep Learning

Resumo

A precisão na classificação automática de tumores cerebrais desempenha um papel determinante para a confiabilidade do método para aplicações na saúde. Erros de

classificação podem resultar em diagnósticos imprecisos, levando a abordagens inadequadas e potencialmente prejudiciais. **Objetivo:** Propor uma abordagem visando minimizar erros de classificação. **Método:** Desenvolvemos um modelo de rede neural convolucional em duas etapas: primeiro, quatro modelos binários para tumores que apresentam maiores desafios de diferenciação; depois, um modelo *Ensemble* para classificação multiclasse. Adicionalmente, empregamos uma técnica para interpretar as previsões dos modelos e identificar as regiões de interesse nas imagens médicas. **Resultados:** Os resultados demonstram que a abordagem proposta alcança uma acurácia de 98%. **Conclusão:** Este trabalho trouxe contribuições para a aplicação de aprendizado profundo na classificação de tumores cerebrais, destacando a importância de abordagens transparentes e robustas para garantir precisão e segurança nas previsões.

Descritores: Tumores cerebrais; Classificação Automática; Aprendizado Profundo

Resumen

La precisión en la clasificación automática de tumores cerebrales desempeña un papel crucial en la confiabilidad del método para aplicaciones en la salud. Los errores de clasificación pueden dar lugar a diagnósticos imprecisos, resultando en enfoques inadecuados y potencialmente perjudiciales. **Objetivo:** Proponer un enfoque para minimizar errores de clasificación. **Método:** Desarrollamos un modelo de red neuronal convolucional en dos etapas: primero, cuatro modelos binarios para tumores con desafíos significativos de diferenciación; luego, un modelo Ensemble para clasificación multiclase. Además, empleamos una técnica para interpretar las predicciones del modelo e identificar las regiones de interés en imágenes médicas. **Resultados:** Los resultados demuestran que el enfoque propuesto logra una precisión del 98%. **Conclusión:** Este trabajo aporta a la aplicación del aprendizaje profundo en la clasificación de tumores cerebrales, resaltando la importancia de enfoques transparentes y robustos para garantizar precisión y seguridad en las predicciones.

Descriptores: Tumores cerebrales; Clasificación Automática; Aprendizaje profundo

Introduction

Tumors that affect the Central Nervous System result from the abnormal proliferation of cells in the tissues that make up this anatomical region.[1] In 2020,

approximately 308,102 new cases of brain tumors and other tumors of the central nervous system were diagnosed in worldwide, leading to an estimated 251,329 deaths, as reported in the study by Sung *et al*.[2]

A variety of intracranial lesions can radiologically mimic meningioma.[3] Tumor location and growth pattern can be useful for differential diagnostic considerations. Meningiomas in the cerebral hemispheres can be confused with dural metastases, especially from glial tumors that extend into the subarachnoid space.[4] Typical imaging changes along the optic nerve sheath or cavernous sinus, initially suggestive of meningioma, may indicate glioma or inflammatory diseases. Additionally, pituitary neoplasms such as adenomas or craniopharyngiomas can be confused with meningiomas.[3]

This study aims to investigate the causes underlying the frequent incorrect classification of meningioma tumors when using the following databases: Brain Tumor Classification MRI[5], the Brain Tumor Dataset[6], and the Br35H - Brain Tumor Detection[7], all available in public repositories for training deep learning models. We intend to test approaches that can reduce the number of misclassifications, particularly when differentiating between meningioma and no tumor. Furthermore, this study aims to incorporate the Local Interpretable Model-agnostic Explanations (LIME)[8] technique for the interpretability and understanding of model predictions.

Related Work

The article by Rasool *et al*.[9] proposes a hybrid architecture based on Convolutional Neural Network (CNN) to classify three types of tumors using magnetic resonance images. The method proposed by the authors combines CNN with Support Vector Machine (SVM), achieving 98.10% accuracy when using Google-Net as a feature extractor. Analyzing the confusion matrices presented in the study, we observed that the tumor with the most classification errors was glioma, with 23 errors classified as meningioma. Meningioma was also incorrectly classified as glioma (2 errors) and pituitary (8 errors).

In another study, Rasool *et al*.[10] proposed two strategies for automatically

classifying brain tumors. The first combines unsupervised classification by an SVM with feature extraction by a pre-trained CNN, known as SqueezeNet (SN-SVM). The second

approach incorporates the supervised softmax classifier with a SqueezeNet Fined Tuned (SN-FT). The results of the experiments showed that the adjusted SqueezeNet model achieved an accuracy of 96.50%. However, when using SqueezeNet as a feature extractor with an SVM classifier, the accuracy increased to 98.70%. The authors also presented confusion matrices and hit rates by class. In the SN-SVM approach, it was observed that the meningioma tumor was incorrectly classified as a glioma in two images and as absence of tumor in two others. In the SN-FT approach, the meningioma class was incorrectly classified as glioma in 12 images.

The authors Mahajan and Chavan[11] presented an analysis of three transfer learning processes with the VGG16, Inception V3, and EfficientNet B2 architectures applied to classify and detect brain tumors in magnetic resonance images. The proposed model using EfficientNet B2 achieved an overall accuracy of 97.50%. We noticed several classification errors involving the meningioma class when analyzing the results.

The paper by Gómez-Guzmán *et al.*[12] evaluates seven convolutional neural network models, highlighting InceptionV3 with 97.12% accuracy. The authors, however, emphasized the importance of identifying inaccuracies in the classification, which were revealed through confusion matrix analysis. According to the authors' interpretation, InceptionV3 demonstrated greater accuracy than alternatives due to its lower number of incorrectly classified cases. They also emphasize that, during the analysis, the classification of "Glioma" and "Pituitary Tumor" proved effective in all models evaluated. However, the "Meningioma" and "No tumor" classes presented learning challenges, revealing lower performance than the other classes.

In the study by Özkaraca *et al.*[13], the authors used VGG16Net and DeseNet to classify brain tumors. The main objective was to evaluate how transfer of learning impacts classification success rates. The results reached 96% in the sensitivity metric. It is noteworthy that, in the analysis of the confusion matrix, 23 images were classified incorrectly. Among these errors, we highlight the 5 cases of False Negatives in the "Meningioma" class, erroneously classified as "No Tumor". Additionally, it was observed that the "Meningioma" class was the most prone to classification errors, totaling 15

incorrect classifications among the 23 error cases.

The studies reviewed point to advances but also challenges and recurring errors in the classification of brain tumors, especially meningioma. Among the articles reviewed, only one study[12] addressed this issue, highlighting the need to investigate further the causes of this error, which may be related to the data, model architecture, image quality, or variables yet to be identified. With these challenges, it is important to seek computational and scientific solutions to mitigate this type of error and, thus, improve the reliability of brain tumor classification systems.

Methods

Dataset

The dataset used in this work is the Brain Tumor MRI Dataset[14], available in a public repository. It comprises 7,023 Magnetic Resonance Imaging (MRI) images of the human brain, divided into two subsets: Training (5,712 images) and Test (1,311 images). The images were classified into Glioma, Meningioma, No Tumor, and Pituitary. This dataset is composed of results from merging information from three distinct datasets.[5,6,7] The dataset[5] initially provided the brain MRI images for the final dataset. However, a critical flaw in the "Glioma" category was identified, with incorrectly categorized images, as demonstrated by previous research results.[12,15] To correct this inaccuracy, the "Glioma" images were replaced with the images contained in the dataset[6]. Furthermore, dataset[7] provided images from the "No Tumor" category for the final dataset, enabling the inclusion of images of healthy brains.

Preprocessing

Image preprocessing is a fundamental step in preparing data for analysis in computer vision applications. Each image must be unique, and duplicate images have been removed. It uses a function that compares images based on their hash values, ensuring that only one copy of each image is kept. Individual images are resized to 150x150 pixels and converted to grayscale.

When developing machine learning models, ensuring data integrity, and preventing leaks are crucial to obtaining reliable and meaningful results. Some

strategies were adopted for this purpose: The first was separating data into training and testing sets. Ensures that the model is evaluated on a dataset not used during training. After splitting the data, the images were normalized to ensure all input values were on the same scale. Normalizing the data after dividing the training, validation, and testing sets can prevent data leakage so that the model works efficiently and fairly.

Architecture and Configuration of the Models

The CNN architecture was implemented using the Keras library and adopted for both binary and multiclass classification. The structure includes convolutional and pooling layers, starting with 32 filters, a 5x5 kernel, and ReLU activation function, followed by a repeated pattern, increasing the filters to 64, 128, and then returning to 64, with max pooling for dimensionality reduction. Global Average Pooling was applied to reduce the image dimensionality to the global average, facilitating the training of the subsequent dense layer. A dropout layer with 0.2 rate was used for regularization to prevent overfitting. The final dense layer has 4 units for multiclass classification and 1 unit for binary classification, using softmax and sigmoid activation functions, respectively, assigning probabilities to the output classes. The adopted optimizer was Adam, with a learning rate of 0.001, and the loss function is sparse_categorical_crossentropy for multiclass and binary_crossentropy for binary. During training, accuracy was monitored as a metric for performance evaluation.

Training

The training process was conducted in two phases. The first phase involved developing four binary classification models to address the challenges of tumor classification, specifically meningioma and glioma. These models were designed to classify Glioma x Meningioma, Glioma x Pituitary, Meningioma x NoTumor, and Meningioma x Pituitary. In the second phase, an Ensemble model was created by combining the binary models for multiclass classification. To evaluate the models' performance across different subsets of data, we employed Repeated K-Fold Cross-validation. The learning curves in Figures 1 and 2 illustrate the training process for the Glioma, Meningioma, and No Tumor classes.

Figure 1 – Learning curves of the Glioma x Meningioma model

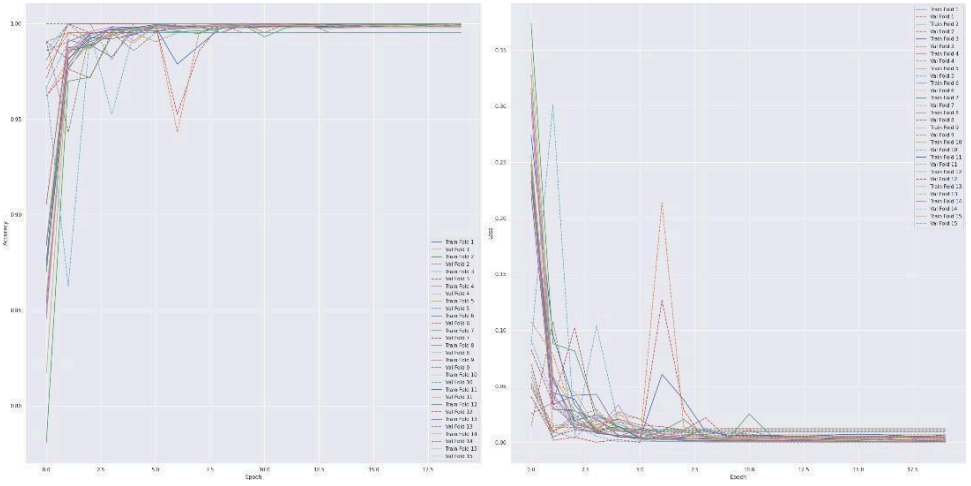
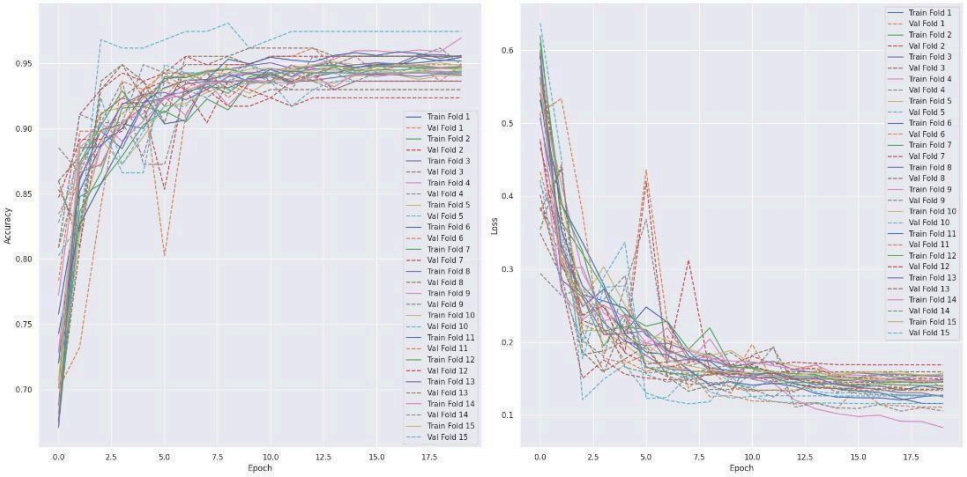


Figure 2 – Learning curves of the Meningioma x No Tumor model



After observing a low performance in the "Glioma x Meningioma" model, we adopted two distinct datasets, one for glioma[16] and another for meningioma[17], containing pre-processed images of the respective tumors. Following the training of the four binary models, we created an Ensemble model (Figure 3) to leverage the individual knowledge of each binary model in multiclass classification, considering the specialized information each model acquired during training. The model's performance during the 20 training epochs is highlighted in the learning curves (Figure 4).

Figure 3 – Ensemble Model Architecture

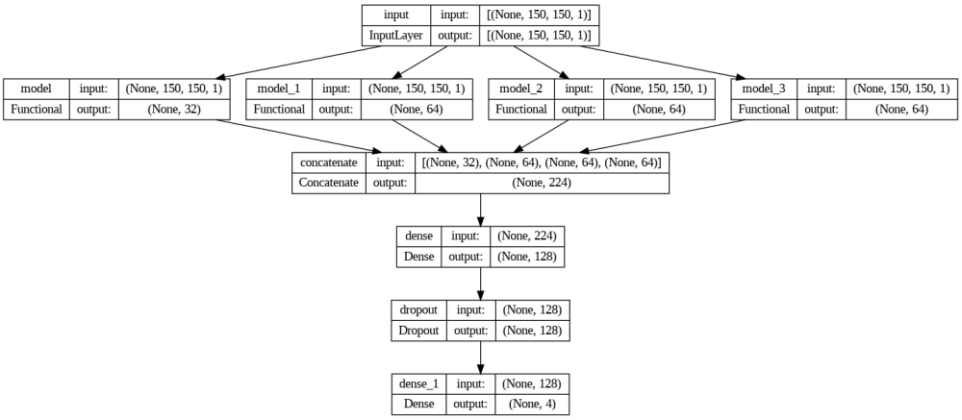
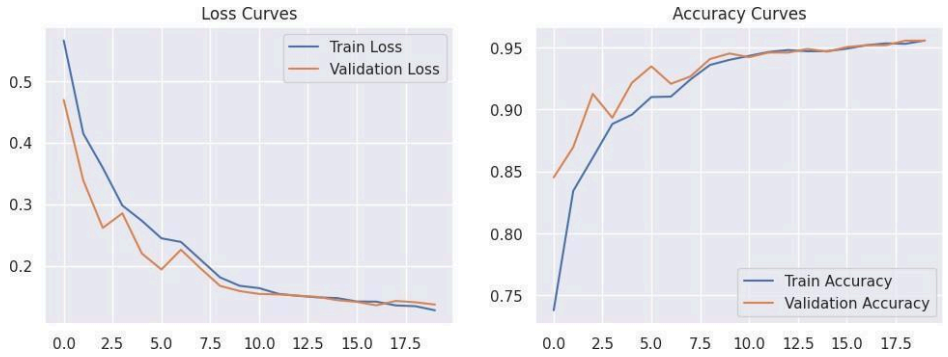


Figure 4 – Learning Curves of the Multiclass Ensemble Model



Final Model Testing

After completing the training and validation phases, the Ensemble model underwent an additional 30 epochs of training using all previously employed training and validation images. Subsequently, the model was assessed using a distinct test dataset to evaluate its generalization capability. The Brain-tumor-classification Dataset[18] was used, divided into four categories: glioma, meningioma, notumor, and pituitary, totaling 13,196 images. Of these images, 92% were allocated for training (12,093 images) and 8% for testing (1,103 images). During testing, the set of 1,103 images was used exclusively, applying basic preprocessing procedures, such as converting the images to grayscale and resizing to (150, 150, 1).

Model Explainability

Understanding the reasons behind the predictions of deep learning models is fundamental, especially when the planned use is in healthcare, which directly impacts human well-being. Currently, the demand for explainability of models previously considered black-blox is growing. The ability to interpret model predictions raises the level of confidence and promotes the responsible and ethical adoption of new technological advances in medicine and healthcare. In this context, we use the LIME[8] interpretability methodology for the explainability of the final model predictions. Using the technique allows the user to understand how the model makes predictions based on specific parts of the image according to the predicted class.

Results and discussion

This section discusses the results of the proposed Ensemble model for classifying brain tumors, comparing them with related studies presented in the corresponding section. Performance evaluation, crucial to validating the model's effectiveness, is measured by precision, sensitivity, F1-score, and accuracy.

The model achieved considerable accuracy, such as 100% for “No Tumor”, 95% for “Meningioma”, 99% for “Pituitary”, and 98% for “Glioma”. Sensitivity was consistent, standing out at 99% for "No Tumor" and "Pituitary", 97% for "Meningioma", and 96% for "Glioma". The F1-score, which combines precision and sensitivity, indicated consistent balance, approaching 0.97 for all classes. Furthermore, the accuracy of the final model was 98%, in line with recent literature on deep learning models on the same database. The accuracy surpasses previous studies[11,12,13] and is equal to recent studies.[9,10]

The confusion matrix (Figure 5), generated in the final tests, offers a detailed view of the predictions for each class. The results indicate solid model performance, with 98% accuracy and only 2% (24 images) incorrectly classified. We highlight that the false negative prediction for the meningioma class, a concern in the health area, occurred only once. Tumors with greater malignancy were not misclassified as non-tumors. The model had challenges differentiating between gliomas and

meningiomas, aligning with previous studies but with a lower error rate in these classes than previous studies.[9,10,11].

Figure 5 – Ensemble Model Test Confusion Matrix

As in medical practice, various other lesions, including glioma, can radiologically mimic meningioma.[3] Because we are working with real-world problems, deep learning may be susceptible, even to a lesser degree, to this pattern of difficulty, just like the radiologist. Other hypotheses for this classification problem may be the quality of images in available public databases, inappropriate adjustments of hyperparameters about the problem, and less robust methods for dividing and training data, such as simple holdout.

Regarding the explainability of the Ensemble model predictions, Figure 6 represents the meningioma class, while Figure 7 represents the glioma class. In Figure 6, we have an example of correct classification, while in Figure 7, although the actual class is glioma, it was classified as meningioma with 94.89% confidence. The areas highlighted in yellow in the central images indicate the most relevant superpixels for class classification, with these critical areas highlighted while the rest of the image is hidden. The third image on the right presents the same interpretation, but with a different representation style. The colors green and red play a crucial role in interpretation, indicating positive and negative contributions to correct classification,

respectively. In all images, the areas around the tumor are highlighted, identifying the points close to the tumor considered for positive classification. It is important to note that the images used for explainability were obtained from public and untrained sources, ensuring an unbiased evaluation of the Ensemble model. The choice of images from health websites aimed to test the generalization of the model to new images, simulating real-world situations and ensuring the robustness of the classification in different contexts outside the known environment during training.

Figure 6 – Explainability of the Meningioma class with LIME. The original image is on the left—source: Internet [19]

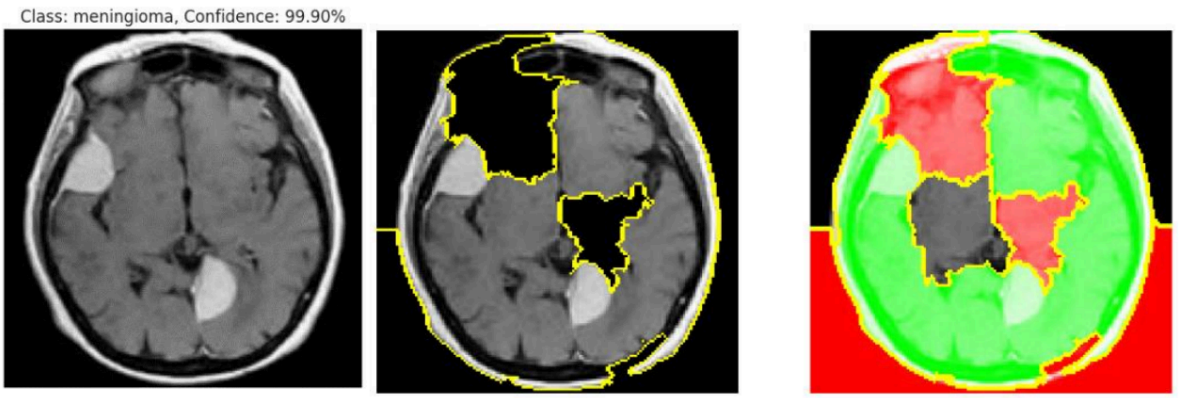
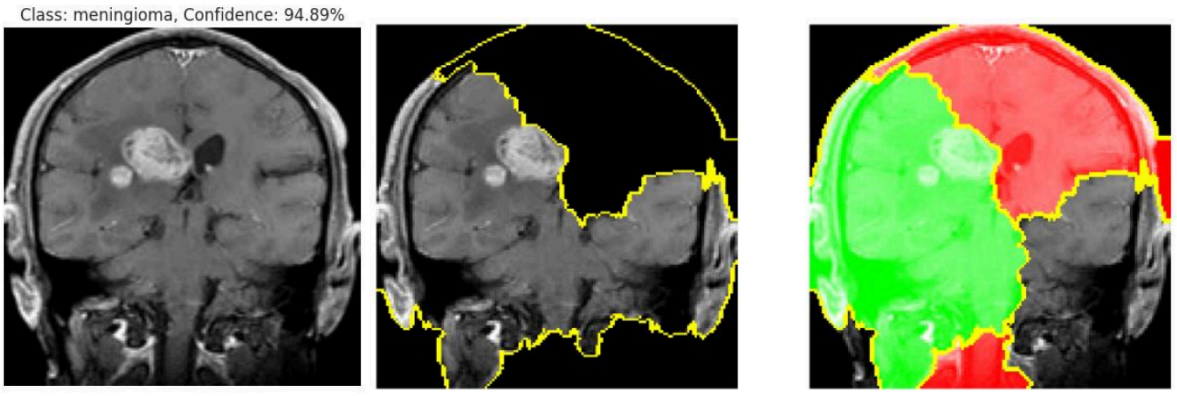


Figure 7 – The incorrect classification of the Glioma tumor by Meningioma with LIME The original image is on the left—source: Internet [20]

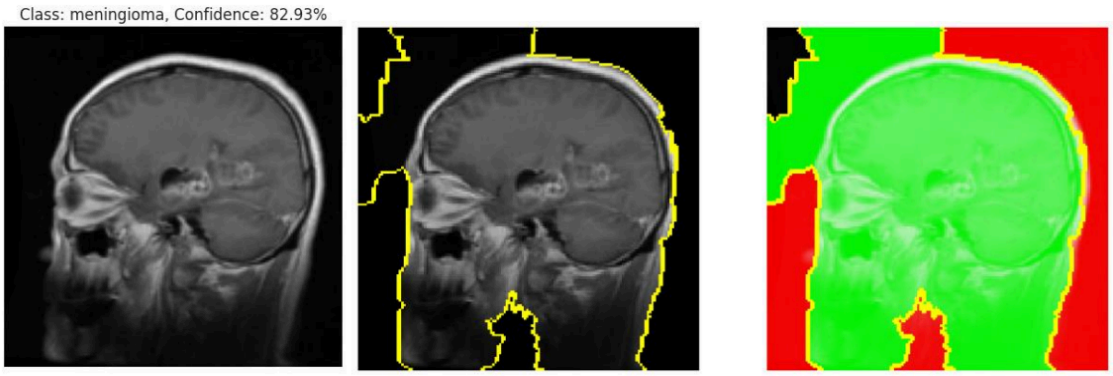


In the images in Figure 7, which show the classification error by the Ensemble model, we highlight the contour and the green area in the region identified as the tumor area. However, as discussed in the literature and mentioned in this study, the tumor's

location may be in a gray area common to both tumors. This similarity in location makes accurate classification difficult since possibly many examples of meningiomas used to train the model are at similar points.

In Figure 8, the glioma class was again incorrectly classified as meningioma, with a confidence of 82.93%, we observed the model's difficulty in identifying the central point for the classification. The exclusion of the black background led us to infer that the central image, which is our point of interest, is the focus of the model. However, perhaps due to the small size of the tumor, the low contrast, or the cropping of the image, it hindered correct classification. These observations highlight the classification challenges in cases of tumors with similar anatomical characteristics or in low-resolution or low-contrast imaging conditions.

Figure 8 – Explainability of the incorrect classification of the Glioma by Meningioma tumor



Our study presented promising results in classifying brain tumors using the Ensemble model. However, it also presented some limitations. The difficulty distinguishing between glioma and meningioma, possibly due to their similarities, represents a significant challenge. Furthermore, the quality and quantity of MRI images can affect the model's accuracy.

Conclusion

The presented paper aimed to test deep learning approaches for classifying brain tumors in medical images, emphasizing reducing classification errors and interpretability of model predictions. The results showed an accuracy of 98%, indicating

the high potential to aid diagnosis and improve precision.

The LIME technique played a crucial role in allowing the interpretation of model decisions in detail. LIME was used to identify regions of interest in medical images, facilitating the understanding of the specific characteristics taken into consideration by the model in the classification of brain tumors. Not only can this increase clinical confidence in decision-making, but it also assists in identifying areas of uncertainty, which

is particularly important in challenging situations such as distinguishing between glioma and meningioma.

Although the results were promising, the study highlighted essential limitations that highlight the need for improvements. Specifically, expanding the dataset, using preprocessing techniques to improve image quality, and improving interpretation should be further explored. The study presents contributions to applying deep learning for the automatic classification of brain tumors, emphasizing the importance of transparent and robust approaches to guarantee model predictions' accuracy, safety, and reliability.

References

1. Câncer do sistema nervoso central [Internet]. Instituto Nacional de Câncer - INCA. [cited 2024 Jan 26]. Available from: https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/sistema-nervoso-central

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a Cancer Journal for Clinicians. 2021 Feb 4;71(3):209–49.

3. Nowosielski M, Galldiks N, Iglseder S, Kickingereder P, von Deimling A, Bendszus M, et al. Diagnostic challenges in meningioma. Neuro-Oncology [Internet]. 2017 Nov 1 [cited 2024 Jan 26];19(12):1588–98. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5716093/

4. Wilms G, Lammens M, Marchal G, Demaerel P, Verplancke J, Van Calenbergh F, et al. Prominent dural enhancement adjacent to nonmeningiomatous malignant lesions on contrast-enhanced MR images. AJNR American journal of neuroradiology [Internet]. 1991 [cited 2024 Jan 26];12(4):761–4. Available from: https://pubmed.ncbi.nlm.nih.gov/1882761/

5. Bhuvaji S. SartajBhuvaji/Brain-Tumor-Classification-DataSet [Internet]. GitHub. 2023 [cited 2024 Jan 26]. Available from: https://github.com/SartajBhuvaji/Brain-Tumor-Classification- DataSet

6. Cheng, J. (2017). Brain tumor dataset [Internet]. [cited 2024 Jan 26]. Available from: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427. DOI: 10.6084/m9.figshare.1512427.v5

7. Ahmedhamada0. Br35H: Brain Tumor Detection 2020 [Internet]. [cited 2024 Jan 26]. www.kaggle.com. Available from: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection

8. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?" Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16

[Internet]. 2016; [cited 2024 Jan 26]. Available from: https://arxiv.org/pdf/1602.04938.pdf.

9. Rasool M, Ismail NA, Boulila W, Ammar A, Samma H, Yafooz WMS, et al. A Hybrid Deep Learning Model for Brain Tumour Classification. Entropy [Internet]. 2022 Jun 8;24(6):799. [cited 2024 Jan 26]. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9222774/

10. Rasool M, Ismail NA, Al-Dhaqm A, Yafooz WMS, Alsaeedi A. A Novel Approach for Classifying Brain Tumours Combining a SqueezeNet Model with SVM and Fine-Tuning. Electronics. 2022 Dec 29;12(1):149.

11. Mahajan, N., & Chavan, H. (2023). A Robust Approach for Brain Tumor Detection using Transfer Learning. Available from: https://ieeexplore.ieee.org/abstract/document/10220906?casa_token=ZmdSd_mFuJ0AAAAA:l5j9izaUrS-LA26hclujQgSDBscVkYwPc9MsmLEmXBrS3XRWZ2-dOfnBend_8uD_RDg4scQ7gg

12. Gómez-Guzmán MA, Jiménez-Beristaín L, García-Guerrero EE, López-Bonilla OR, Tamayo-Perez UJ, Esqueda-Elizondo JJ, et al. Classifying Brain Tumors on Magnetic Resonance Imaging by Using Convolutional Neural Networks. Electronics [Internet]. 2023 Jan 1;12(4):955. [cited 2024 Jan 26]. Available from: https://www.mdpi.com/2079-9292/12/4/955

13. Özkaraca O, Bağrıaçık Oİ, Gürüler H, Khan F, Hussain J, Khan J, et al. Multiple Brain Tumor Classification with Dense CNN Architecture Using Brain MRI Images. Life. 2023 Jan 28;13(2):349.

14. Nickparvar, M. (2023). Brain Tumor MRI Dataset [Internet]. [cited 2024 Jan 26]. www.kaggle.com. Available from: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset/

15. Filatov D, Yar GNAH. Brain Tumor Diagnosis and Classification via Pre-Trained Convolutional Neural Networks [Internet]. arXiv.org. 2022. [cited 2024 Jan 26]. Available from: https://arxiv.org/abs/2208.00768

16. Glioma Object Detection Dataset and Pre-Trained Model by test [Internet]. Roboflow. [cited 2024 Jan 26]. Available from: https://universe.roboflow.com/test-786lz/glioma-4mibx

17. Meningioma Detector Object Detection Dataset and Pre-Trained Model by Kelompok BRIN [Internet]. Roboflow. [cited 2024 Jan 26]. Available from: https://universe.roboflow.com/kelompok-brin-dxxif/meningioma-detector

18. brain-tumor-classification Classification Dataset by opendataacademia [Internet]. Roboflow. [cited 2024 Jan 26]. Available from: https://universe.roboflow.com/opendataacademia/brain-tumor-classification-qkeoa

19. Desai PK. Multiple meningiomas | Radiology Case | Radiopaedia.org [Internet]. Radiopaedia. [cited 2024 Jan 26]. Available from: https://radiopaedia.org/cases/multiple-meningiomas-1

20. Gomes L. Glioma de Baixo Grau - Tumores Cerebrais [Internet]. Dr. Gustavo Noleto. 2022 [cited 2024 Jan 26]. Available from: https://www.neurodrgustavonoleto.com/glioma-de-baixo-grau/