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

Modelo *multi-task* para classificação e segmentação de tumores cerebrais

Multi-task model for classification and segmentation of brain tumors

Modelo *multi-task* para la clasificación y segmentación de tumores cerebrales

Guilherme Müller Ferreira[1], Viviane Rodrigues Botelho[2] , Áttila Leães Rodrigues[3] , Carla Diniz Lopes Becker[2] e Thatiane Alves Pianoschi Alva[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.
3 Ph.D., Federal University of Rio Grande do Sul - UFRGS, DEMIN, Porto Alegre (RS), Brazil.

Autora correspondente: Prof. Dra. Thatiane Alves Pianoschi Alva
E-mail: thatiane@ufcspa.edu.br

Resumo

Objetivo: Validar se um modelo *multi-task (MTL)* para classificação e segmentação de tumores cerebrais é superior a um *single-task (ST)*. **Método:** a arquitetura do modelo é constituída de um encoder, que se bifurca em uma *fully connected* (classificação) e um *decoder* (segmentação). Para o *ST*, apenas a ramificação de classificação foi considerada. Ambos foram treinados utilizando a abordagem de 5-fold *cross validation* com os datasets SARTAJ e Figshare. **Resultados:** O *MTL* alcançou acurácia de 95.99% ± 0.70% em comparação com 95.87% ± 1.01% do *ST*. **Conclusão:** Ambos os modelos apresentaram desempenhos semelhantes, entretanto o *MTL* revelou algumas vantagens, como uma maior estabilidade de métricas, resultado do desvio padrão menor em todas as métricas. Em relação à literatura, o *MTL* obteve uma acurácia de apenas 3% abaixo do melhor modelo entre os analisados, e também apresentou um número significativamente menor de parâmetros, com até 187 vezes.

Descritores: Aprendizado Profundo; Neoplasias Encefálicas; Inteligência Artificial.

Abstract

Objective: To validate whether a multi-task model (MTL) for brain tumor classification and segmentation outperforms a single-task (ST) approach. **Method:** The model architecture consists of an encoder, branching into a fully connected (classification) and a decoder (segmentation). For the ST, only the classification branch was considered. Both were trained using the 5-fold cross-validation approach with the SARTAJ and Figshare datasets. **Results:** The MTL achieved an accuracy of 95.99% ± 0.70% compared to 95.87% ± 1.01% for the ST. **Conclusion:** Both models presented similar performances, however the MTL revealed some advantages, such as greater stability of metrics, resulting from the lower standard deviation in all metrics. Compared to the literature, the MTL achieved an accuracy only 3% below the best model analyzed and also had a significantly lower number of parameters, up to 187 times fewer.

Keywords: Deep Learning; Brain Neoplasms; Supervised Machine Learning; Artificial Intelligence.

Resumen

Objetivo: Validar si un modelo *multi-task* (*MTL*) para la clasificación y segmentación de tumores cerebrales es superior a un enfoque de *single-task* (ST). **Método:** La arquitectura del modelo consta de un *encoder*, que se bifurca en una *fully connected* (clasificación) y un *decoder* (segmentación). Para el *ST*, solo se consideró la rama de clasificación. Ambos fueron entrenados utilizando el enfoque de validación cruzada de 5 pliegues con los conjuntos de datos SARTAJ y Figshare. **Resultados:** El *MTL* logró una precisión del 95.99% ± 0.70% en comparación con el 95.87% ± 1.01% del *ST*. **Conclusión:** Ambos modelos presentaron desempeños similares, sin embargo el *MTL* reveló algunas ventajas, como una mayor estabilidad de las métricas, resultante de la menor desviación estándar en todas las métricas. En comparación con la literatura, el *MTL* logró una precisión solo un 3% por debajo del mejor modelo analizado y también tuvo un número significativamente menor de parámetros, hasta 187 veces menos.

Descriptores: Aprendizaje Profundo; Neoplasias Encefálicas; Inteligencia Artificial.

Introduction

Malignant brain tumors are a type of cancer that exhibits a disproportionate mortality rate in relation to their incidence. In other words, although they are uncommon, they have a significant impact on mortality[1]. In the early 21st century, there was an increase in the survival rate of patients with brain cancer, which can be associated, among other factors, with early diagnosis[2]. Given the above, the need for research directed towards the development of more effective and precise diagnostic tools is evident[3].

In the context of healthcare, the use of Machine Learning techniques, particularly Deep Learning, has emerged as a promising trend to assist in the diagnosis of medical images[3]. For example, Deep Learning is capable of assisting in the diagnosis of COVID-19[4]. Such models demonstrate the ability to learn patterns from existing data, enabling their application to new datasets[5]. In the field of computer vision, this is achieved through convolutional layers that perform feature extraction[3]. As an example of this, such layers can detect patterns at different levels of abstraction, ranging from identifying the presence or absence of edges in specific orientations to understanding more complex combinations[5].

Multi-task Learning (MTL) represents a promising approach in the field of Deep Learning, involving the training of a single model with multiple tasks (outputs) simultaneously[6]. Studies demonstrate that this approach offers the advantage of accelerating the training process and improving efficiency in data utilization[7]. The essence of this approach lies in exploring the relationships existing between tasks, which can lead to more efficient performance in each of them[6]. This is possible because the model shares representations and features among tasks, thereby enhancing its ability to generalize better[8].

In healthcare studies, there is already evidence that the MTL approach can improve the performance of Deep Learning models. Tardy and Mateus (2022)[9] achieved a 10% increase in the Area Under the Curve (AUC) for breast cancer classification by adding sub-tasks of classification and reconstruction. Oliveira et al. (2023)[10] developed a classification model with the segmentation sub-task to assess the severity of chronic

venous disorders and observed greater robustness and stability in training compared to other approaches.

In head magnetic resonance imaging (MRI) images, a study developed a segmentation model to identify small tumors with the reconstruction sub-task[11]. The results revealed that this approach proved to be superior in tumor segmentation with enhancement. Moreover, the metrics of the proposed neural network were similar to the Single-Task NVDLMED model. Regarding Single-Task models for brain tumor classification, various studies have reported significant results. Gómez-Guzmán et al. (2023)[12] achieved 97.12% accuracy with an InceptionV3-based architecture with 23.9M parameters. Ullah et al. (2022)[13] obtained 98.91% accuracy; however, their architecture was based on InceptionResNetV2, which has even more parameters (55.9M). Rasheed et al. (2023)[14] devised a strategy to achieve a precise yet lighter and faster model, constructing an architecture with 1.7M parameters and an accuracy of 97.84%.

In this context, this article presents a MTL model for the classification of head magnetic resonance images while simultaneously performing the tumor region segmentation task. This work will conduct two main comparisons. The first analysis will evaluate the performance of the MTL model against a ST model to validate whether the same benefits found in other works of better metrics[10] and greater stability[10] are found for the problem studied. The second comparison will involve our model and those from the literature, with an emphasis on performance and number of parameters.

Methods

Dataset

The data for this study originate from two datasets: SARTAJ[15] and Figshare[16]. From both datasets, images and corresponding classes were extracted. However, due to the lack of information regarding segmentation masks in SARTAJ[15], the masks identifying the lesion region in the images were exclusively obtained from Figshare[16]. Additionally, the "glioma" class from the SARTAJ[15] dataset was excluded due to identified data quality issues, as reported by Filatov (2022)[17]. In total, 5402 head magnetic resonance images

were used, with 937 meningioma tumors, 901 pituitary tumors, and 500 without tumors from SARTAJ[15], and 708 meningioma tumors, 930 pituitary tumors, and 1426 glioma tumors referring to Figshare[16]. Of these, 80% were set aside for training, and 20% for testing. Thus, 56.72% of our data has annotated masks.

ST and MTL architectures

Both ST and MTL models take as input a grayscale image of size (128, 128). The first part is an encoder, which consists of five blocks formed by: a convolutional layer with filters of size (3, 3); a Batch Normalization layer; a MaxPooling2D layer with size (2, 2). The number of filters in the convolutional layers started at 8 and doubled until reaching 128, ensuring that more complex features were represented by a greater number of feature maps. After the encoder, two branches were implemented: one for classification and another for segmentation.

The classification branch consists of a Flatten layer, followed by a Dense layer with 128 neurons, a Dropout layer with a rate of 0.5, and a Dense layer with 4 neurons representing the classes (no tumor, meningioma, glioma, and pituitary). The segmentation branch is a decoder consisting of 5 blocks similar to those in the encoder, with the modification of replacing the MaxPooling2D layer with an UpSampling2D layer with size (2, 2). Additionally, the number of filters started at 128 and was halved across the blocks. At the end of this branch, another convolutional layer with only 1 filter of size (1, 1) was added to result in a mask of size 128x128. The final model had only 608,645 parameters. When considering only the classification task (ST case), the model has 362,020 parameters.

Preprocessing

The images were normalized to the range [0, 1], and the labels were converted to the One-hot encoding format. For the SARTAJ[15] dataset, a mask with all elements equal to -1 was created for each image to identify the source.

Loss function

To enable the model to learn both classification and segmentation tasks simultaneously, a loss function based on two components was necessary. Equation 1 shows our total loss function associating real values (y) and predicted values (ŷ).

The first component is the loss function associated with classification (Equation 2), which is expressed by the Categorical Focal Loss, where default values from the package were chosen for alpha (0.25) and gamma (2). These parameters assist our model in balancing the importance for each class and adjusting the focus of our model on harder to predict examples.

The second component is the loss function associated with segmentation (Equation 3), which is expressed by the sum of two loss functions: Binary Focal Loss (Equation 4) and Dice Loss (Equation 5). For Equation 4, the same constants as Equation 2 were adopted, as they are standard package defaults.

$$L_{\text{Total}}(y, \hat{y}) = L_{\text{Classification}}(y, \hat{y}) + L_{\text{Segmentation}}(y, \hat{y}) \tag{1}$$

$$L_{\text{Classification}}(y, \hat{y}) = L_{\text{CategoricalFocal}}(y, \hat{y}) = -y \cdot 0.25 \cdot (1 - \hat{y})^2 \cdot \log(\hat{y}) \tag{2}$$

$$L_{\text{Segmentation}}(y, \hat{y}) = L_{\text{BinaryFocal}}(y, \hat{y}) + L_{\text{Dice}}(y, \hat{y}) \tag{3}$$

$$L_{\text{BinaryFocal}}(y, \hat{y}) = -y \cdot 0.25(1 - \hat{y})^2 \cdot \log(\hat{y}) - (1 - y) \cdot 0.25 \cdot \hat{y}^2 \cdot \log(1 - \hat{y}) \tag{4}$$

$$L_{\text{Dice}}(y, \hat{y}) = 1 - \frac{2 \cdot \sum(y \cdot \hat{y}) + 1}{\sum y + \sum \hat{y} + 1} \tag{5}$$

In this work, given the presence of missing labels, specifically the absence of segmentation masks for the SARTAJ[15] dataset, it was necessary to adapt the loss function. Inspired by the approach of Tardy and Mateus (2022)[9], the following condition was implemented during training: if the image belongs to SARTAJ[15], its prediction will contribute only to the classification component. In other words, in Equation 1, only the first component is considered for these samples.

Fit

The model was trained for a maximum of 50 epochs using the default RMSProp optimizer and a batch size of 32. The following callbacks were employed: EarlyStopping monitoring the validation loss with a patience of 10; ModelCheckpoint saving the weights that yield the lowest validation loss; and ReduceLROnPlateau reducing the learning rate by half every 3 epochs without a decrease in the validation loss. The 5-fold cross-validation approach was utilized.

Two scenarios were tested. In the first scenario, the model with classification and segmentation branch (MTL), and in the second scenario only with the classification branch (ST). Both were trained using a GPU P100.

Results and Discussion

ST and MTL models

The accuracy, precision and recall metrics obtained in the 5-fold cross-validation for both ST and MTL models are presented in Table 1 and Table 2 for the training, validation, and test sets. For the MTL model, as it presents a segmentation branch, the IoU metric was also calculated. Additionally, the loss function results for the training and validation sets are depicted in the form of graphs of the total loss function per epoch for ST (Graph 1), the classification component of the loss function per epoch for MTL (Graph 2) and the total loss function per epoch for MTL (Graph 3). A sample of the masks predicted by the MTL model can be visualized in Figure 1. Regarding processing time, ST presented an average time of 49s to complete an epoch, in contrast to MTL which presented 50s.

Table 1 – ST model evaluation results

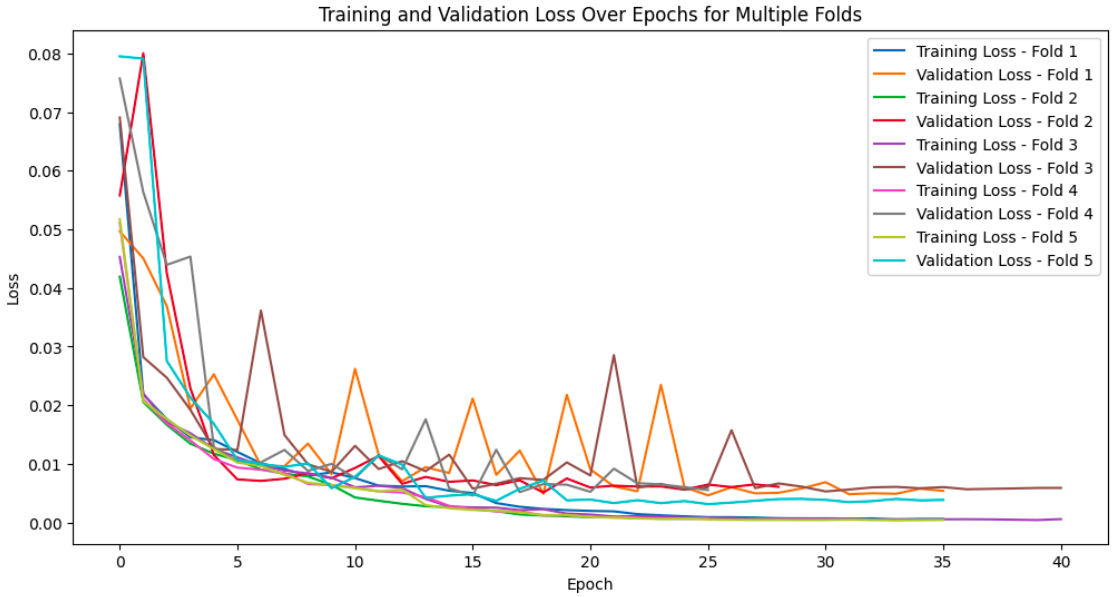
Set	Accuracy	Precision	Recall
Training	99.38% ± 0.56%	99.39% ± 0.55%	99.37% ± 0.57%
Validation	95.72% ± 0.84%	95.80% ± 0.87%	95.60% ± 0.81%
Test	95.87% ± 1.01%	95.92% ± 0.95%	95.80% ± 1.06%

Table 2 – MTL model evaluation results

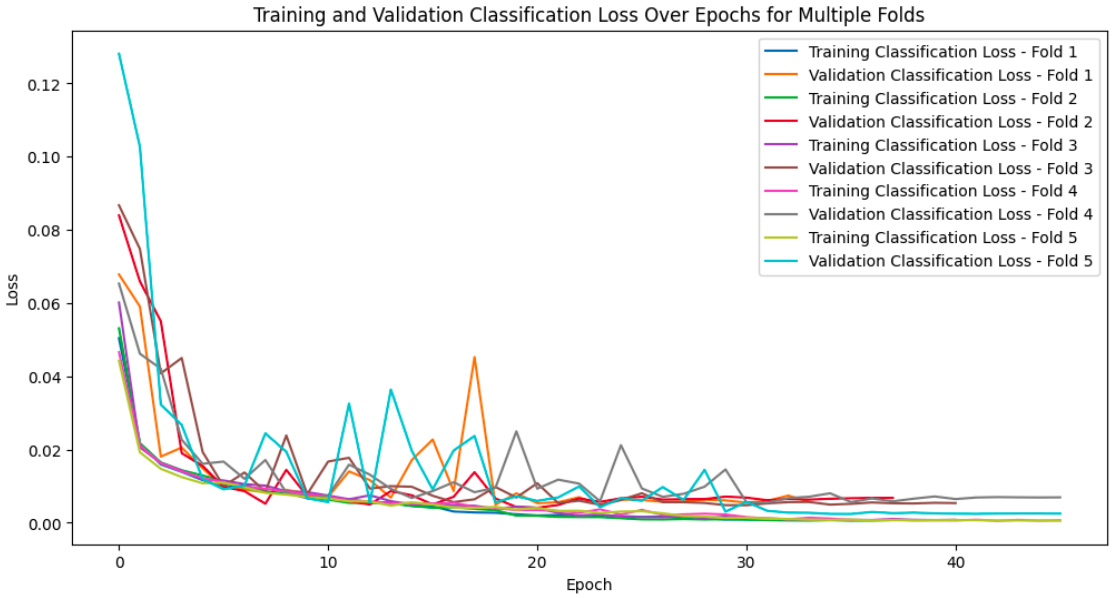
Set	Accuracy	Precision	Recall	IoU
Training	99.54% ± 0.34%	99.54% ± 0.34%	99.51% ± 0.34%	75.41% ± 2.85%
Validation	95.49% ± 0.76%	95.64% ± 0.82%	95.32% ± 0.64%	48.81% ± 1.16%

Test	95.99% ± 0.69%	96.14% ± 0.75%	95.87% ± 0.71%	46.09% ± 1.03%

Graph 1 – Loss function per epoch for the training and validation sets for each fold of the ST model



Graph 2 – Classification component of the loss function per epoch for the training and validation sets for each fold of the MTL model



Graph 3 – Loss function per epoch for the training and validation sets for each fold of the MTL

model

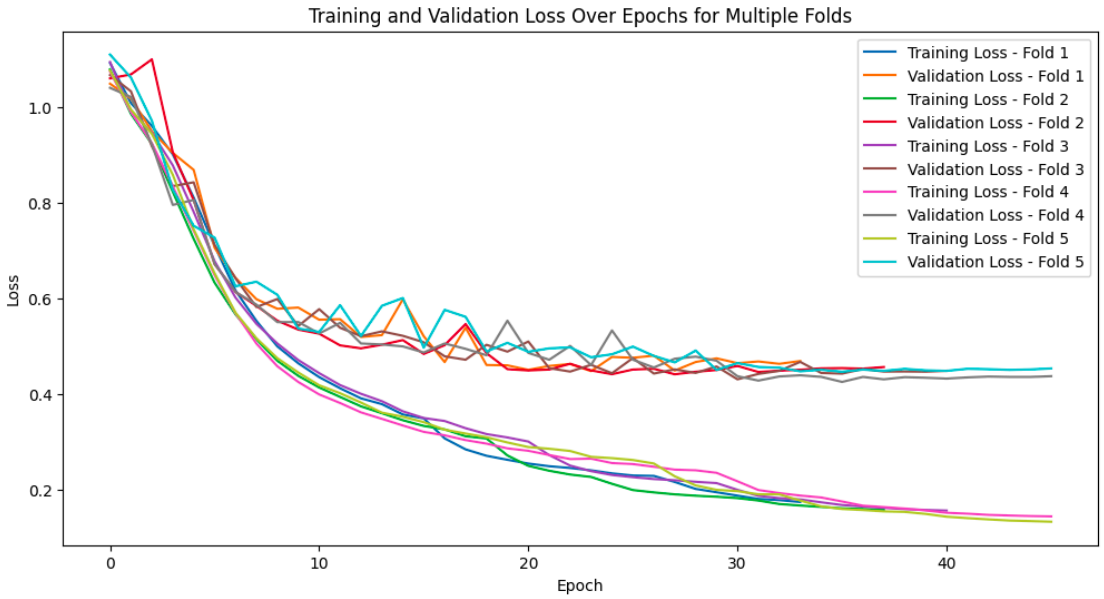
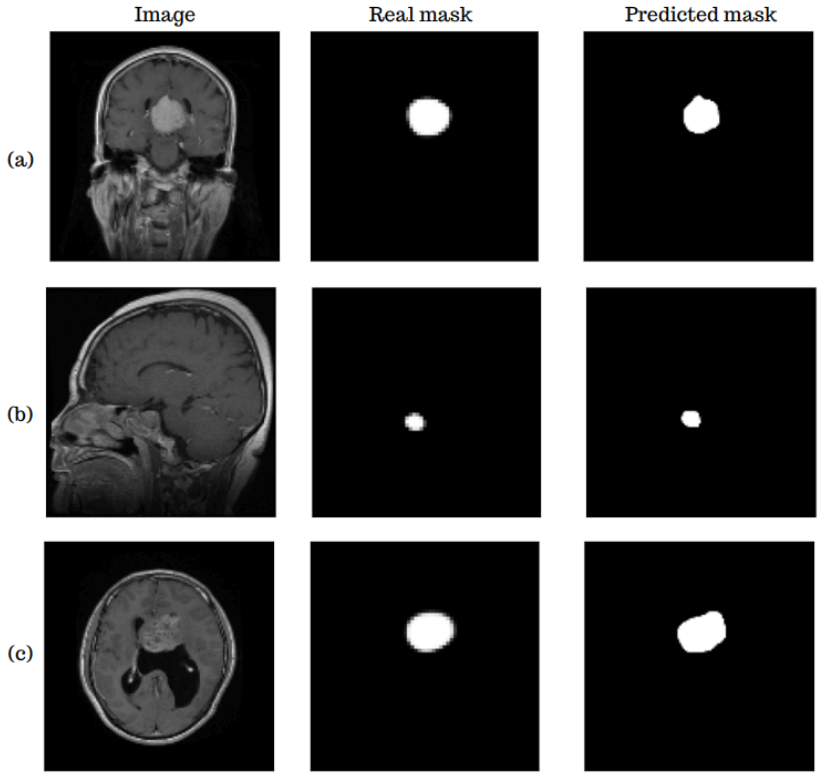


Figure 1 – Comparison of predicted masks on the test set with the real mask for: a) Meningioma (Figshare[16]), b) Pituitary (Figshare[16]), and c) Glioma (Figshare[16])



Regarding performance, both models presented similar metrics, with higher values in the training and testing set for the MTL, and in the validation set, for the ST. However, it is essential to highlight that this evaluation was based on a single iteration of the 5-fold cross-validation. A significant advantage observed in MTL was its greater stability, as can be seen in Tables 1 and 2, where smaller standard deviations are noticeable across all metrics in the all sets.

In Graph 3, it is possible to observe a growing difference over epochs between the total loss functions of the training and validation sets, a phenomenon not visible in the classification component (Graph 2). The explanation for this is evident in Table 2. The segmentation metric (IoU) shows a significantly larger difference between sets than the classification metrics. Thus, there are indications of overfitting for this task. However, as it is a secondary task, with the sole purpose of maximizing the primary task (classification), this is not a cause for concern. Additionally, by analyzing only the classification component (Graph 2), it is noticeable that there was no overfitting for this task.

Comparison with state of the art

Table 3 compares our MTL model with those found in the literature based on the number of classes, accuracy, precision, recall, IoU, and the number of parameters. For the work of Ullah et al. (2022)[13], the number of parameters related to the used backbone was considered. For our work, the parameter count was considered only for classification, with the segmentation; the value is 0.6M. It is important to highlight that Gómez-Guzmán et al.[12] and Rasheed et al. (2023)[14] used an additional dataset which was not used in this work, while Ullah et al. (2022)[13] used only one dataset (SARTAJ[15]).

Table 3 – Comparative table with models from the literature. Notes: * Parameter value refers to the backbone used. ** Parameter quantity is only for classification; with segmentation, it reaches 0.6M.

Articles	Number of classes	Accuracy	Precision	Recall	IoU	Number of parameters
Gómez-Guzmán et al.[12]	4	97.12	97.97	96.59	-	23.9M
Ullah et al.[13]	3	98.91	98.28	99.75	-	55.9M*
Rasheed et al.[14]	4	97.84	97.85	97.85	-	1.7M

Proposed MTL model	4	95.99	96.14	95.87	46.09	0.3M**

As shown in Table 3, the work presented similar results to those in the literature, with only 3% difference in accuracy for the best model among the analyzed studies. Additionally, the model has a significantly lower number of parameters, being approximately 6 times lighter than the model in the work of Rasheed et al. (2023)[14], which is the smallest model found in the literature. At the other extreme, the present model is approximately 187 times lighter than that of Ullah et al. (2022)[13]. This is an advantage of our work, since with fewer parameters, the model becomes simpler and has a lower capacity to memorize the data (overfitting), and tends to perform better with unseen data.

Another advantage of the model, in contrast to those found in the literature, is the production of segmentation masks that show the lesion region. This output not only provides additional information about the tumor's location but also corroborates the efficiency of the decoder, responsible for feature extraction, as illustrated in Figure 1. In this sense, it is possible to confirm that the classification performed by the network is based on important features of the images.

Conclusion

This paper suggests a MTL model designed for tumor classification and segmentation in skull magnetic resonance images and compares it with the ST approach. The same approach could be used in other imaging domains. Both models ST and MTL presented similar performances, however our results indicate an increase in accuracy, precision, and specificity metrics using the MTL approach in test and training sets. Greater stability was also noticed in the metrics of the MTL approach, which is a significant positive point. Furthermore, our results are comparable to findings in the literature, with only 0.6M parameters. This is important since, in addition to the guarantee of a more reliable model, requires less computational resources, without compromising accuracy, highlights the efficiency of our model.

References

1. Thierheimer M, Cioffi G, Waite KA, Kruchko C, Ostrom QT, Barnholtz-Sloan JS. Mortality trends in primary malignant brain and central nervous system tumors vary by histopathology, age, race, and sex. J Neurooncol. 2023;162:167-177.

2. Deorah S, Lynch CF, Sibenaller ZA, Ryken TC. Trends in Brain Cancer Incidence and Survival in the United States: Surveillance, Epidemiology, and End Results Program, 1973 to 2001. Neurosurg Focus. 2006;20(4):E1.

3. Yang S, Zhu F, Ling X, Liu Q, Zhao P. Intelligent Health Care: Applications of Deep Learning in Computational Medicine. Front Genet. 2021;12:607471.

4. Trombetta GBW, Fröhlich W da R, Rigo SJ, Rodrigues CA. Application of Deep Learning for Diagnosis of COVID-19-Induced Pneumonia from X-ray Images. J Health Inform [Internet]. March 15, 2021 [cited March 9, 2024];12. Available from: https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/828.

5. LeCun Y, Bengio Y, Hinton G. Deep Learning. Nature. 2015;521:436.

6. Zhang Y, Yang Q. An Overview of Multi-Task Learning. Natl Sci Rev. 2018;5(1):30-43.

7. Crawshaw M. Multi-Task Learning with Deep Neural Networks: A Survey. arXiv. 2020.

8. Ruder S. An Overview of Multi-Task Learning in Deep Neural Networks. arXiv. 2017.

9. Tardy M, Mateus D. Leveraging Multi-Task Learning to Cope With Poor and Missing Labels of Mammograms. Front Radiol. 2021;1.

10. Oliveira B, et al. A multi-task convolutional neural network for classification and segmentation of chronic venous disorders. Sci Rep. 2023;13:761.

11. Ngo DK, Tran MT, Kim SH, Yang HJ, Lee GS. Multi-Task Learning for Small Brain Tumor Segmentation from MRI. Appl Sci. 2020;10(21):7790.

12. Gómez-Guzmán MA, et al. Classifying Brain Tumors on Magnetic Resonance Imaging by Using Convolutional Neural Networks. Electronics. 2023;12:955.

13. Ullah N, et al. An Effective Approach to Detect and Identify Brain Tumors Using Transfer Learning. Appl Sci. 2022;12:5645.

14. Rasheed Z, et al. Brain Tumor Classification from MRI Using Image Enhancement and Convolutional Neural Network Techniques. Brain Sci. 2023;13(9):1320.

15. Bhuvaji S, Kadam A, Bhumkar P, Dedge S, Kanchan S. Brain Tumor Classification (MRI). Kaggle. 2020.

16. Cheng J. Brain Tumor Dataset. Figshare. 2017.

17. Filatov D, Ahmad GN, Yar H. Brain Tumor Diagnosis and Classification via Pre-Trained Convolutional Neural Networks. medRxiv. 2022.