Brain tumor classification on the patient level using attention-based AI methods and multi-sequences MRI

Vladimir Groza[1,2] VLADIMIR.GROZA@MEDIANTECHNOLOGIES.COM
Bair Tuchinov[1] BAIRT@NSU.RU
Evgeniya Amelina[1] AMELINA.EVGENIA@GMAIL.COM
Evgeniy Pavlovskiy[1] PAVLOVSKIY@POST.NSU.RU
Nikolay Tolstokulakov[1] N.TOLSTOKULAKOV@G.NSU.RU
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Median Technologies, Valbonne, France*

Mikhail Amelin[1,3] AMELIN81@GMAIL.COM
Sergey Golushko[1] S.K.GOLUSHKO@GMAIL.COM
[3] *FSBI "Federal Neurosurgical Center", Novosibirsk, Russia*

Andrey Letyagin[4] LETYAGINAY@BIONET.NSC.RU
[4] *Research Institute of Clinical and Experimental Lymphology, Branch of IC&G SB RAS, Novosibirsk, Russia*

Editors: Under Review for MIDL 2022

Abstract

Investigation of brain tumor structure and its type-dependent variations are among the list of most important research directions where the medical imaging methods are used. Structural and statistical analysis of these lesions originates various associated problems and projects such as detection of the tumors, shape and specific sub-regions segmentation (i.e. necrotic part, (non-)enhanced part, edema), classification of the tumor presence and treatment follow up prognosis. Almost all of these problems are usually solved numerically, specifically with the tendency to use the Artificial Intelligence (AI) related methods often including Deep Learning (DL) networks.

One of the most complicated, weakly explored and challenging tasks in this domain is the classification of the tumor types. This difficulty is explained by several reason where the most principle one is the strong limitation of the existing open-sourced datasets that include clinically confirmed tumor type labels based on the radiological examination protocols.

In this work we present current results of the brain tumor classification problem, where we consider and operate with four different lesion types such as meningioma, neurinoma, glioblastoma and astrocytoma. All the conducted research and presented results are obtained on the newly introduced dataset including 255 labeled volume MRI scans describing wide variety of the tumors and its clinically associated ground truth (GT) information.

Obtained in this work results demonstrate not only inspiring and strong Accuracy performance of 0.925 on patient level (and accordingly 0.894 slice-wise) but also very high potential and perspective for the future research in this field.

Keywords: Brain tumor, Deep learning, Medical imaging, Classification, MRI, Clinical analysis

1. Introduction

The field of medical imaging continues to grow rapidly, constantly raising new clinical problems. Such growth affects specific domains and tasks where a lot of new discoveries and achievements can be done. The high interest results in the promising opportunities to improve the clinical workflow, the screening process and potentially the treatment plan.

Among all possible regions of interests (ROI) in the human-body, there are several specific ones representing the highest focus for the AI-based research such as Lungs, Liver and Brain (Yan et al., 2020; Jiang et al., 2020). Depending on the study specification various CT/MRI/X-ray medical imaging modalities are used for solving problems like segmentation, detection, classification etc (Groza et al., 2020b; Dercle et al., 2021; Groza and Kuzin, 2020; Letyagin et al., 2020).

Unlike the well-known segmentation problems, one needs to consider the whole 3D volume as a single object when approaching the per-patient classification task. This crucial difference dramatically reduces the amount of the available data in terms of the units, so it significantly affects on the model performance and robustness.

Usually brain tumors are classified into benign (meningioma and neurinoma) and malignant (glioblastoma and astrocytoma) categories. Benign tumors are usually almost non-progressive (non-cancerous) and assumed to be less aggressive. Contrary the malignant tumors are cancerous and grow very rapidly with no constraints on the shape and size.

Recently several works about the tumors classification from MRI scans were published demonstrating interesting and promising results. The work of (M.Badza and Barjaktarovic, 2020) was done using only 3 types of the tumors such as meningioma, glioma and pituitary tumor which are quite evident for the distinguishing due to strong visual and localization differences between these classes. Also this work is based on the open sourced dataset of only T1-weighted scans that was only split into 10 train/validation folds. This could lead to some questionable results regarding the method robustness and data leakage due to the lack of proper validation and test set. The paper of (Khan et al., 2020) is focused on the classification of the acquisition sequences and not the tumor classes. This work is done using the well-known BraTS datasets (Reyes et al., 2018) which is unfortunately not acceptable for the tumor classification tasks. While being very famous and rich in terms of the given amount of the brain scans this dataset is created for the multi-class segmentation tasks first and is very limited with the represented tumor classes - only gliomas (i.e. glioblastomas and astrocytomas) are given. In the (Pei et al., 2020) paper authors consider the classification as a subtask which also includes only glioma vs astrocytoma differentiation.

There were published number of works with the approaches of the optimal usage of different MRI scans, improving the solutions stability, compensation the lack of missing data and increasing the models performance (Ge et al., 2019; Havaei et al., 2016; Varsavsky et al., 2018; Dorent et al., 2019).

In this work we propose the method trained and tested on the given dataset of 255 brain MRIs to classify them into 4 classes. Being still limited with the dataset size we obtained a generalized solution due to the balanced amount proportion of tumor types and its shape and localization variability.

Additionally, we demonstrated how small augmentation of the dataset can lead to the significant changes in the models performance. Such results behavior truly explains the

importance of the proper and rich dataset usage and secondly shed the light on the promising perspective in this research direction.

2. Dataset description

This work is based on the private Siberian Brain tumor (SBT) dataset and its 2 version namely v.0.3 and v.0.4, consisting relatively of 202 and 255 brain MRI volume scans. This was done in order to analyze the impact of the class imbalance and of the insufficient class representation. Dataset v.0.3 consists of 64 meningiomas, 58 neurinomas, 43 glioblastomas and 37 astrocytomas, while v.0.4 includes respectively 76, 77, 56 and 46 samples.

Additionally to this we created the independent hold-out test set of 40 MRI scans including 10 scans per each tumor category respectively. This hold-out test set was used as an additional verification of the obtained solution in order to confirm the robustness of the solution.

The dataset comprises four different kinds of brain tumors named meningioma, neurinoma, glioblastoma and astrocytoma. All volume patient scans are co-registered to the same anatomical template and are interpolated to the same pixel spacing and resolution (1 mm^3).

In order to make this dataset useful for the various experimental and numerical tasks, multiple associated manual annotation were created including the tumor type and multi class segments. All given patients have been segmented manually by two experienced board-certified neuro-radiologist raters, following the same annotation protocol and any label misclassification was manually corrected by an independent expert. All available GT labels were confirmed in accordance with the histological and immunohistochemical approved clinical patients history.

3. Methods

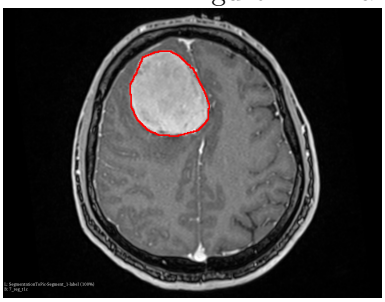
3.1. Image preprocessing and data augmentation

The MRI scans were processed to form the dataset of 2D PNG images that were normalized to [0;1] range and padded from the original size of 240x240 pixels to the common size of 256x256 pixels. Since all available volumes were co-registered to the identical anatomic structure, we removed first 6 and last 20 slices from each volume to exclude the specific registration artefacts. Next having the segmentation masks available we excluded all slices with no masks on it. Such manipulations are resulted in 8786 and 11094 2D slices respectively for dataset version v.0.3 and v.0.4. In both cases we selected the test set as 10% of the dataset while performing the split on the patient level. The rest 90% of the data was split into 5 folds using the GroupKFold method taking into account the patients ID numbers with the stratification by tumor type to prevent the data leakage. All slices were labeled with values 1, 2, 3 or 4 depending on the tumor type of the corresponding patient.

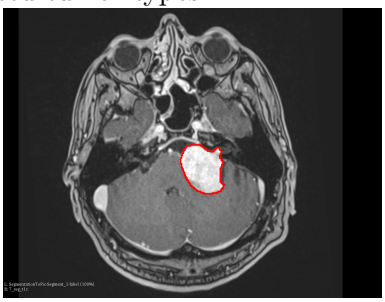
The examples of different tumors types are shown in Figure 1. The tumors are marked with a red outline.

This work was done using only T1C, T1 and T2-FLAIR sequences as the model inputs in order to create the pseudo RGB-like image(Groza et al., 2020a).

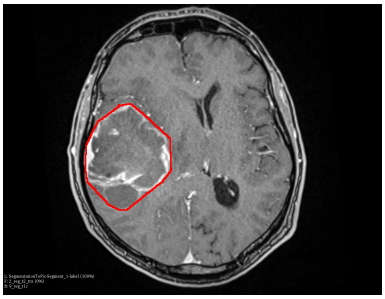
Figure 1: Example of presented tumor types



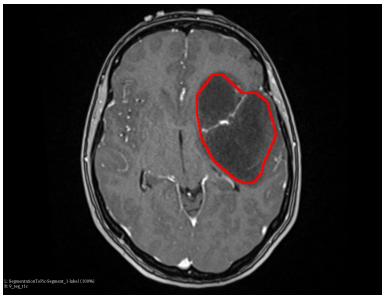
(a) Meningioma - T1c



(b) Neurinoma - T1c



(c) Glioblastoma - T2F



(d) Astrocytoma - T1c

In order to increase the dataset and to enhance the models generalization we used rich augmentations *on the fly* during the training process with the use of albumentations library (Buslaev et al., 2018). The list of randomly applied augmentations is the following: random 90-degree rotations, horizontal / vertical flips, ShiftScaleRotate, transpose transformation, random brightness, contrast and gamma. All augmentations were applied only to the training subset without modifying the validation one.

3.2. Neural Network Implementation Details

Tumor classification was performed using two different CNN developed in the Pytorch framework that are variants of the *se-resnext50-32x4d*, *se-resnext101-32x4d* networks (Hu et al., 2018).

Both of these networks consist of 4 blocks followed by a pooling layer, DropOut layer and Linear Fully Connected layer. Each of the blocks consists of the consequence of the sub-blocks of 2D Convolution layer, Batch Normalization and Activation, also including the channel attention and Squeeze-and-Excitation blocks. The main difference between these two architectures is the number of such sub-block, so for the *se-resnext50-32x4d* we have 3, 4, 6 and 3 respectively, while the *se-resnext101-32x4d* is constructed of 3, 4, 23 and 3.

We used additional attention Squeeze-and-Excitation blocks between the *basic* encoders block in order to increase the impact of tumor-specific features. We also included the AdaptiveAvgPool2d pooling layer before the classifier that is explained by the given tumor type variations.

3.3. Network Training and Performance Measurements

The proposed solution includes a 5-fold cross-validation scheme, so all networks were trained on 5 different folds as described above and ensembled for the final prediction. Both pipelines were trained for 200 epochs using the Focal Loss, *Adam* optimizer (Kingma and Ba, 2014) with initial learning rate (LR) of 1e-4 and the *CosineAnnealing* scheduling.

During the inference, we proposed two options for the slice-wise prediction: 1) based on all tumor-inclusive slices and 2) more adjusted approach excluding the slices with the only tumor edema presence. This approach can be explained by the very simple assumption that such slices are non-informative w.r.t. the tumor type.

In order to improve the performance of our model we applied the test time augmentations (TTA) in the form of 3 additional flip transformations (i.e. horizontal, vertical and its combination). Four different variants of the each sample were processed and output prediction probabilities were averaged before assigning the class.

We provide the models accuracy scores obtained both on the slice level and on the patient level. Tumor class prediction for each test MRI volume was assigned with the use of voting method among all tumor-inclusive slices.

4. Results

4.1. Initial results depending on the Dataset version

Quantitative results of the developed tumor classification method are shown in Table 1. These results stand for the multi-class classification task on the 2D slice level and were obtained on the independent test set of 40 patients using the models trained on the v.0.3 and v.0.4 of our dataset. Numbers in the brackets stand for the patient level accuracy obtained by the voting principle within the single patient.

Table 1: Initial brain tumor classification model performance accuracy

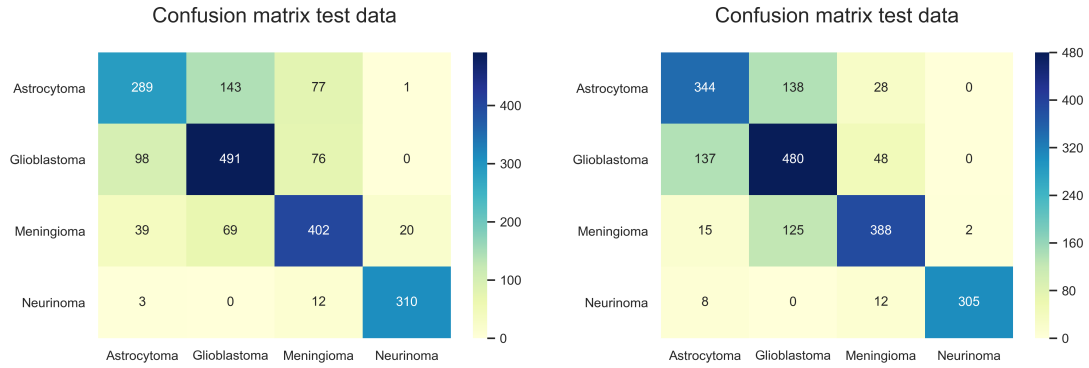
Dataset	Se-ResneXt50	Se-ResneXt101
Dataset v.0.3	**0.7845** (0.85)	0.7398 (0.75)
Dataset v.0.4	0.7616 (0.85)	**0.7809** (0.85)

The overall confusion matrices related to the training on extended dataset (v.0.4) are shown in Figure 2 for both networks respectively. Based on these figures, it is possible to note that the proposed approach is capable to efficiently grade the brain tumors even with the very limited dataset for this task.

4.2. Post-processing and final results

Additionally to the standard inference step we assumed that slices including only the edema part of the lesion are not informative and should be used excluded from the test set. This was possible to implement since the corresponding tumor segmentation masks were available. The test time augmentations (TTA) was used during the inference. These changes were done for the predictions obtained using the most performing model originated from Se-ResneXt101.

Figure 2: Confusion matrices of the initial results on Dataset v.0.4



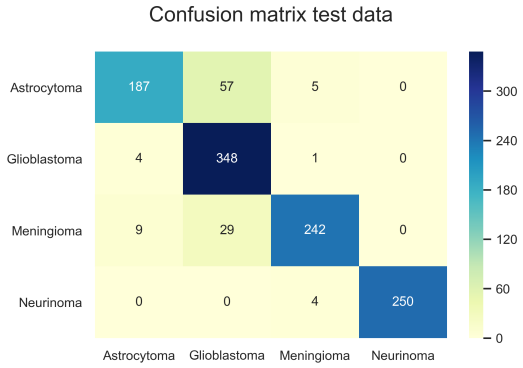
(a) Se-ResNeXt50 (b) Se-ResNeXt101

Obtained accuracy, average sensitivity and specificity results of this classification method are given in Table 2 compared to the baseline ones. Related confusion matrix is presented in Figure 3.

Table 2: Test accuracy on Dataset v.0.4 vs post-processing of edema cut and TTA

Name	ACC	Av. Sensitivity	Av. Specificity
Sx101	0.7809 (0.85)	0.773	0.908
Sx101 + post-proc	**0.8939 (0.925)**	**0.896**	**0.965**

Figure 3: Confusion matrix after post-processing and TTA on Dataset v.0.4



(a) Se-ResNeXt101 + Edema cut off + TTA

Returning to the final evaluation analysis, there are only 3 misclassified patients among given 40 in the test set. It is possible to investigate these cases individually as well as some other difficult cases that were tricky to resolve properly.

4.3. Overview of complex medical cases

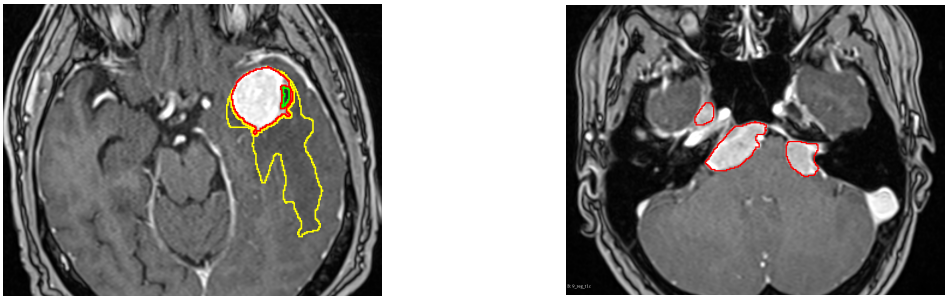
This section discusses the observed classification errors. The following color notations are used for the figures of this section: green - necrosis, red - enhanced tumor, blue - non-enhanced tumor, yellow - edema.

Case 1 (meningioma → astrocytoma, glioblastoma, meningioma). *Meningioma was classified with as nearly equal probability as any other type (Figure 4(a)).*

This tumor is characterized by a large vasogenic edema that is similar in nature, shape and volume to the typical edema of astrocytomas and glioblastomas. In addition, there is a slightly atypical location for a meningioma. They are more frequent in the parasagittal zone and have physical contact with the dura mater. It is highly likely that the combination of these factors was the basis for misclassication.

Case 2 (meningioma → neurinoma). *Firstly the tumor was classified as neurinoma while training on the Dataset v.0.3. However augmenting the training set with new cases of neurinomas and meningiomas (Dataset v.0.4) led to a correct final prediction (Figure 4(b)).*

This is due to the fact that a pair of meningiomas are located in the area of "symmetrical" cerebellar pontoon corners, where acoustic neuromas are most often located.



(a) Case 1. meningioma → any type

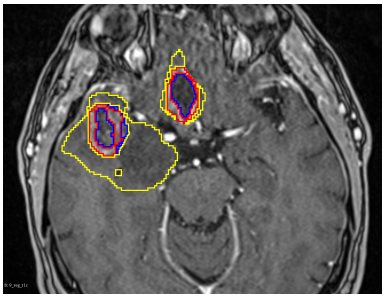
(b) Case 2. Meningioma → neurinoma

Figure 4: Case 1 and Case 2 examples

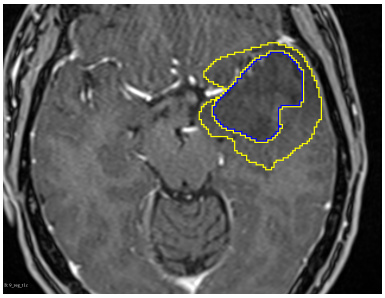
Case 3 (astrocytoma → glioblastoma). *Astrocytoma was misclassified as glioblastoma (Figure 5(a)).* The tumor type was assigned only on the basis of histological immunohistochemistry analysis, where a Grade 3 astrocytoma with a tendency to Grade 4 was found as a resulting conclusion. Thus, we can say that these tumors are similar by their MRI biomarkers, and on the other hand, we can assume that there is a transformation into glioblastoma that is quite frequently occurred in clinical practice.

Case 4 (astrocytoma → glioblastoma). *Low-contrast astrocytomas were classified in some cases as glioblastomas (Figure 5(b)).* Although usually the absence of the contrast-accumulating part or its very small volume is more typical to astrocytomas. The insufficient level of the differentiation is associated with a small representation of such cases in the training dataset.

In the conclusion of per-case investigation one could notice that the main mistakes are occurred in the classification of astrocytomas as glioblastomas. It should be noted that these are two closely related classes of glial tumors, where the astrocytomas tend to transform histologically into the glioblastomas with the time. Results of the classification



(a) Case 3. Grade 3 astrocytoma



(b) Case 4. Low-contrast astrocytoma

Figure 5: Case 3 and Case 4 examples

do not go beyond the class of gliomas but at the same time such mistakes are made towards overestimating the malignancy degree.

The second type of the classification errors is associated with the location of the tumor. Such errors are often observed with meningiomas located in a place typical for a neurinoma. In order to avoid this kind of errors, more attention should be paid not onlt to "lying on the surface" biomarkers of tumor localization but also to MRI biomarkers of the tumor structure.

5. Conclusions

This work presents the recent achievements and first numerical results for the multi-class brain tumor classification problem. We considered both approaches: previously investigated slice level and novel more complex and a priori less stable patient level.

The proposed method demonstrates the stable and strong performance in the multi-class brain tumor classification task in terms of the accuracy metric. It is important to note that results in this work were obtained on the very extensive dataset w.r.t. the tumor types. The stability and robustness of this method was confirmed with the various numerical experiments for four different lesion types as meningioma, neurinoma, glioblastoma and astrocytoma. Predictions on the patient level demonstrate promising results for the further research using more extensive dataset.

Additionally, obtained research outcomes demonstrate the efficiency of the Multi sequences MRI preprocessing approach in the various types of medical imaging problems.

Particular misclassification results are very important from the clinical point of view and can significantly impact on the GT labeling quality as well as on computer-aided systems where similar solutions are deployed.

Both of the classification error types presented in the "Overview of complex medical cases" could be possible to reduce by enlarging of the training dataset with the clinically relevant cases.

References

A. Buslaev, A. Parinov, E. Khvedchenya, V. Iglovikov, and A. Kalinin. *Albumentations: fast and flexible image augmentations.* arXiv:1809.06839, 2018.

L. Dercle, T. Henry, A. Carré, N. Paragios, E. Deutsch, and C. Robert. *Reinventing radiation therapy with machine learning and imaging bio-markers (radiomics).* State-of-the-art, challenges and perspectives, Methods, 2021. doi: 0.1016/j.ymeth.2020.07.003.

R. Dorent, S. Joutard, M. Modat, S. Ourselin, and T. Vercauteren. *Hetero-Modal Variational Encoder-Decoder for Joint Modality Completion and Segmentation.* arXiv:1907.11150, 2019.

C. Ge, I. Y. Gu, A. Store Jakola, and J. Yang. *Cross-modality augmentation of brain MR images using a novel pairwise generative adversarial network for enhanced glioma classification.* IEEE International Conference on Image Processing (ICIP), 2019.

V. Groza and A. Kuzin. *Pneumothorax Segmentation with Effective Conditioned Post-Processing in Chest X-Ray.* IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops), 2020. doi: 10.1109/ISBIWorkshops50223.2020.9153444.

V. Groza, B. Tuchinov, E. Pavlovskiy, E. Amelina, M. Amelin, S. Golushko, and A. Letyagin. *Data preprocessing via multi-sequencesmri mixture to improve brain tumor segmentation.* 8thInternational Work-Conference on Bioinformatics andBiomedical Engineering (IWBBIO), Springer Gabler, 2020a. doi: 10.1007/978-3-030-45385-5_62.

V. Groza, B. Tuchinov, E. Pavlovskiy, E. Amelina, M. Amelin, S. Golushko, and A. Letyagin. *Data Preprocessing via Multi-sequences MRI Mixture to Improve Brain Tumor Segmentation.* Bioinformatics and Biomedical Engineering, Springer, 2020b.

M. Havaei, N. Guizard, N. Chapados, and Y. Bengio. *Hemis: Hetero-modal image segmentation.* MICCAI, Springer International Publishing, 2016.

J. Hu, L. Shen, and G. Sun. *Squeeze-and-Excitation Networks.* IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. doi: 10.1109/CVPR.2018.00745.

Z. Jiang, C. Ding, M. Liu, and D. Tao. *Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task.* Lecture Notes in Computer Science, Springer, 2020. doi: 10.1007/978-3-030-46640-4_22.

M. Khan, I. Ashraf, M. Alhaisoni, R. Damasevicius, R. Scherer, A. Rehman, and S. Bukhari. *Multimodal Brain Tumor Classification Using Deep Learning and Robust Feature Selection: A Machine Learning Application for Radiologists.* Diagnostics, 2020. doi: 10.3390/diagnostics10080565.

D.P. Kingma and Jimmy Ba. *Adam: A method for stochastic optimization.* arXiv:1412.6980, 2014.

A. Letyagin, S. Golushko, M. Amelin, B. Tuchinov, E. Amelina, N. Tolstokulakov, E. Pavlovskiy, and V. Groza. *Multi-class Brain Tumor Segmentation via Multi-sequences MRI Mixture Data Preprocessing.* Cognitive Sciences, Genomics and Bioinformatics (CSGB), 2020. doi: 10.1109/CSGB51356.2020.9214645.

M.Badza and M. Barjaktarovic. *Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network.* Applied Sciences, 2020. doi: 10.3390/app10061999.

L. Pei, L. Vidyaratne, M. Rahman, and et al. *Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images.* Sci Rep 10, 2020. doi: 10.1038/s41598-020-74419-9.

M. Reyes, A. Jakab, S. Bauer, M. Rempfler, and A. Crimi et al. *Dentifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge.* arXiv preprint arXiv:1811.02629, 2018.

T. Varsavsky, Z. Eaton-Rosen, C. H. Sudre, P. Nachev, and M. J. Cardoso. *PIMMS: permutation invariant multi-modal segmentation.* arxiv.:1807.06537, 2018.

Q. Yan, B. Wang, D. Gong, C. Luo, W. Zhao, J. Shen, Q. Shi, S. Jin, L. Zhang, and Z. You. *COVID-19 Chest CT Image Segmentation - A Deep Convolutional Neural Network Solution.* arXiv:2004.10987, 2020.