Brain Tumour Segmentation using Convolution Auto-Encoder & U-NET

Hou Cheng Lam
houcheng.lam@studenti.unipd.it

Ava Finnegan
avalouise.finnegan@studenti.unipd.it

Abstract

Tumour Segmentation with Machine Learning has gained a lot of traction in recent years. In this report, we develop a 2 model pipeline with a Convolutional Autoencoder and a UNet model as a solution to the problem for users without capable hardware for typical machine learning task such as dedicated GPUs. Our model uses slices of 3D MRI scans so it is not reliant to a complete 3D scan. We show that using solely the FLAIR modality is enough to provide a high accuracy prediction.

1. Introduction

Tumour segmentation plays a pivotal role in the medical field, particularly in the diagnosis and treatment of diseases such as cancer. Accurate identification of the tumour area and the surrounding affected tissue is extremely important for successful treatment. In this paper, our focus is on glioblastoma, the most common type of brain tumour accounting for 47.7% of all cases. [2]

Our project is based on image segmentation, specifically semantic segmentation where the goal is to identify which pixels an object is contained in. The use of image segmentation in the medical field is becoming increasingly popular due to the requirement of high accuracy. Analysing MRI scans to identify the exact location in the brain for surgical removal is the most common method for treatment [5]. We will be investigating tumour segmentation on brain tumours using MRI images. The dataset that we are using is the BraTS dataset from the Brain Tumour Segmentation(BraTS) Challenge [1]. MRI scans are non-invasive for the patient but output 4 scans for each patient. However identifying accurate tumour volumes from 3D brain scans using the multiple modalities is a difficult task for doctors. This manual process of brain tumour segmentation can lead to mistakes and is time and labour intensive. These problems can be resolved by utilising automatic and robust methods to identify tumour areas with high accuracy.

The BraTS dataset is publicly available on Kaggle [3]. The dataset is composed of 4 modalities of the MRI scan which are T1, T1 contrast, T2 and T2 Fluid Attenuated Inversion Recovery (FLAIR). These were acquired from 1251 patients from various scanners and multiple institutions. The focus is to determine brain glioblastoma sub-regions in mpMRI scans. To assist with this task, an image segmentation file which is the ground truth of the tumour segmentation annotated by neuro-radiologists. Each segmentation map contains 4 classes which identifies the distinct tumour cells and surrounding areas. The classes are as follows;

- Label 0: No tumour
- Label 1: The necrotic tumour core
- Label 2: The invaded tissue
- Label 4: The GD-enhancing tumour

The model implemented in our paper will output a prediction mask which preforms a 4-class segmentation for each pixel of the patients brain. The data underwent a rigorous pre-processing phase to reduce it's size complexity which enabled us to run the models using our laptops. We implemented our model using only one of the four MRI modalities. Our focus of this project is to run a comparison on which of the 4 modalities achieve the best results for tumour segmentation.

2. Related Works

There are numerous scientific papers which use deep learning techniques to address tumour segmentation for successful treatment. The BraTS challenge for tumour segmentation was first introduced in 2012 and datasets have been released every year motivating a plethora of new techniques.

In 2015, Ronneberger et al [7] introduced the U-net model for medical image segmentation. It proved to be successful in tumour segmentation and has gained popularity due to it's faster implementation than traditional CNNs.

In 2022, a lightweight version of U-Net was proposed by Walsh et al. [9]. This lightweight version would allow for real-time segmentation of MRI scans without the requirement of heavy architecture and large amounts of data. The input data is broken into 3 planes namely axial, coronal and sagital rather than the original three-dimensional volumetric images, for simplified brain tumor segmentation.

Due to the complexity of the brain scans using MRI in various modalities, various pre-processing methods have been examined. In 2010 Tustison et al proposed an algorithm for field bias correction. [8] This N41TK algorithm has been utilised to correct the low-frequency uniformity of the magnetic field in the MRI images. Some papers such as Ranjbarzadeh et al. [6] heavily pre-process the image using thresholding and a morphological operator to remove irrelevant areas. In this paper once the data has been pre-processed, a cascading CNN model is implemented on the smaller preprocessed patch which speeds up computational time. In our method we have followed pre-processing steps which are common to almost all of the above papers. We focused on standardising the images and cropping the images to remove unnecessary background information and speed up computation.

In 2021, brain tumour segmentation using transformers were introduced by Wang et al. [10]. They applied Transformer on 3D CNN and proposed a novel network named TransBTS which is based on an encoder-decoder structure. This method utilities both local and global features to aid with segmentation.

In our approach we were inspired by the method proposed by Li et al.[4]. Their method consists of two steps. A 2D network with U-Net architecture is first used to localise the tumour Region of Interest (ROI) before tumour segmentation is preformed using a 3D U-Net model. We modified this by using the FLAIR modality to outline a region of interest (ROI) and used a Convolution Auto-encoder instead of the U-NET model due it's lighter architecture. Then we propose to use this smaller region containing the tumour to feed into our second model which preforms the segmentation on a 2D slice of the tumour rather than a 3D segmentation as seen from Li. This step significantly reduced the input data for the segmentation task which allowed us to preform segmentation on our laptops. Another benefits of the identification of a ROI is that segmentation can be done to a higher accuracy as majority of the brain does not contain the tumour.

3. Dataset

The dataset that we are using is the BraTS dataset from the 2021 Brain Tumour Segmentation (BraTS) Challenge [1]. We spent a large portion of our time on pre-processing. The dataset is massive with 1251 patients and 4 scans each, this amounts to over 5000 3D images which each have dimensions $240 \times 240 \times 155$. We decided to focus on one modality at a time to reduce computational cost. The modalities namely T1, T1Contrast, T2 and FLAIR all have unique characteristics that highlight different features of the tumour. The 3 tumour classes were one-hot encoded into a vector of size 4, including the class of No Tumour. The FLAIR modality is pictured in Figure 1 along-with the 3 tumour classes in which we aim to identify.

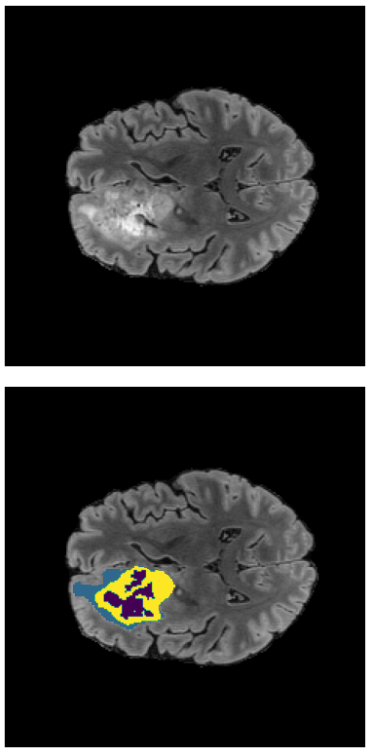


Figure 1: The segmentation classes of the tumour superimposed onto the FLAIR modality

The first step of our pre-processing was to reduce the volume of our data. The MRI scans of the data have additional voxels which reside outside the brain. This was unnecessary information and would have a significant computational impact on our training. With limited resources, it was essential for us to reduce this. Each individuals brain volume was cropped to exclude the background and then this was re-scaled so that the size of all scans were the same.

The image intensities can vary depending on which institution they were collected due to different MRI machines. To unify the data, the intensities were then standardised to have a mean 0 and standard deviation of 1.

$$I_{\text{standardised}} = \frac{I - \mu(I)}{\sigma(I)} \tag{1}$$

To summarise, after pre-processing the dataset consisted of 1251 patients with 1 scan modality per patient. Each scan was cropped, standardised and re-scaled to a dimension of size $64 \times 64 \times 155$.

4. Method

Our method can be divided into two stages and we worked with 2D slices of the brain rather than the full 3D

volume image. This was motivated by GPU constraints as training with 3D volumes caused memory issues. The first stage is the identification of the tumour ares using a convolutional auto-encoder model. The resulting ROI was used as a reference to the input of the U-Net model for tumour segmentation. The data was split into training, testing and validation sets. Both models were trained using Binary Cross-Entropy Loss and Adam optimization with L_2 Regularisation. The model architectures can be seen in Figure 2 and Figure 3.

4.1. ROI Identification

To reduce the size of the input to our U-Net model, we first identify the tumour region which we will call the region of interest (ROI). To find the ROI, we chose to use a Convolutional Auto-Encoder (CAE). The CAE consists of an encoder and a decoder. The model learns to encode input data into a lower-dimensional space representation and the decoder is used to reconstruct the original data or a target image such as mask images.

Padding was used between each intermediate convolutional layers which preserved the size of the image so that the output image was the same dimension as the input image before pooling. This model was shallow in comparison to the U-Net model as the ROI identification was a simpler task.

4.2. Tumour Segmentation

After we identified the patch where the tumour resides, this was resized to 64×64 for the segmentation task. This was to enlarge the region of interest for a more accurate segmentation. The model we implemented was a basic U-Net for semantic segmentation. We followed the U-Net architecture proposed by Ronneberger [7]. The model outputs an image of the same size $64 \times 64 \times 4$ where the last channel contains the probability of the pixel belonging to each of the 4 classes. The U-Net model consists of an contracting path and an expansive path. The contracting path follows the typical architecture of a convolutional neural network. There are two sequential convolution layers with 3×3 filters and a padding of 1. Each pair of convolution layers was followed by ReLU and a 2×2 max pooling operation with a stride of 2 for down-sampling. At each down-sampling step, the number of filters are doubled. Then the model follows an expansive path following the same pattern with a pair of convolution layers with 3×3 filters and same padding following by ReLU and a 2×2 convolution of stride 2 which halves the number of features.

The up-sampling produces an output which is larger than the input. There is a concatenation which links the contracting path to the expanding path. This concatenation is cropped due to the loss of border pixels at each up convolution. This concatenation improves the reconstruction of the image by making use of previous layers of the network. At the final layer a single convolution layer with kernel size 1×1 is used to map each pixel to one of the 4 classes.

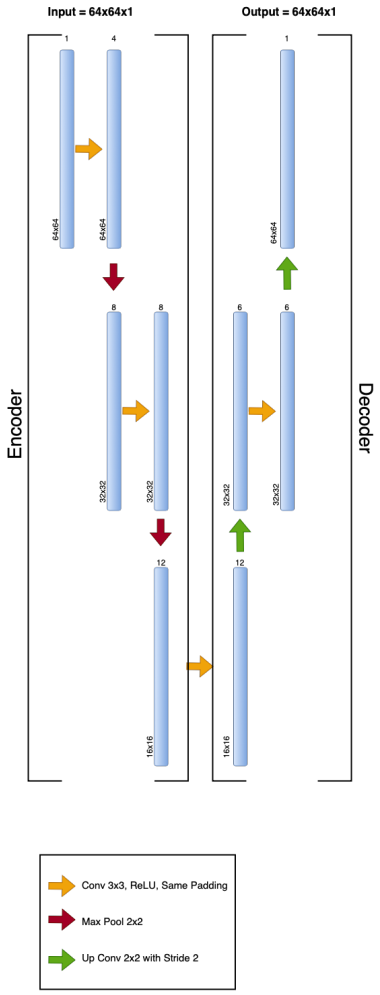


Figure 2: The structure of our CAE Model

4.3. Model Evaluation

To evaluate the performance of our model, the Dice similarity coefficient (DSC) is used. It is commonly used in image segmentation. It computes the similarity percentage between the ground truth and the output of the model. Let P denote the predictions of our model and T denotes the ground truth where $P \in \{0, 1\}$ and $T \in \{0, 1\}$. The Dice coefficient outputs the similarity between P and T where 1 indicates 100% accuracy and 0 represents zero accuracy.

$$\text{Dice coefficient} = 2 \cdot \frac{|P \cap T|}{|P| + |T|} \tag{2}$$

We use 2D slices to analyse the accuracy of the tumour segmentation. To obtain a full 3D segmentation of the tumours, the 155 2D slices of each patient can be combined.

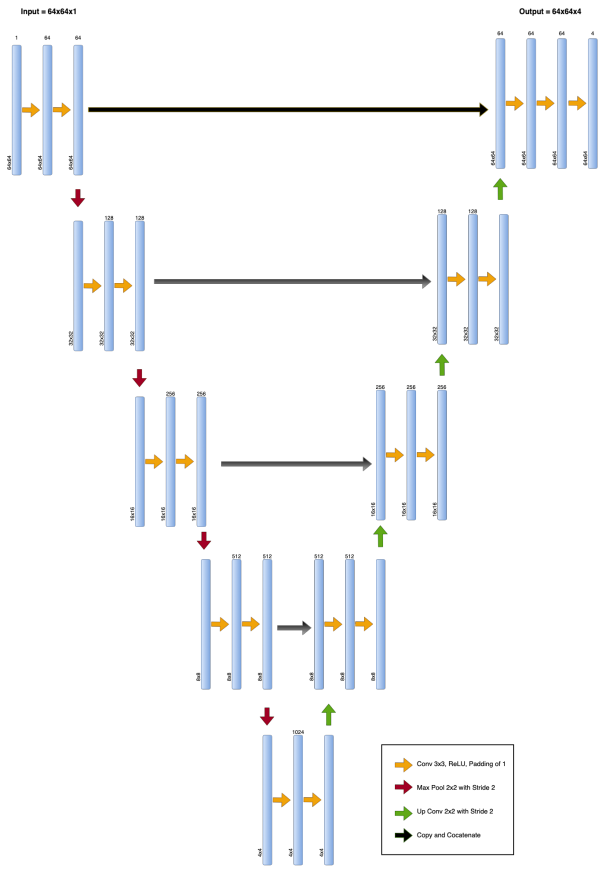


Figure 3: The structure of our U-Net Model

5. Experiment

The data was divided into a training, testing and validation set using a 70 : 10 : 20 split. This was done in a way to ensure that the 155 2D slices of the same patient would fall into the same category. This was important so that no bias was introduced into the validation sets which could influence the results. The training set contained 877 images, the testing set 126 images and the validation set contained 248. The BraTS dataset contains segmentations for the necrotic core tumour, the invaded tissue, the GD-enhancing core tumour, and the no tumour regions. Our aim is to obtain the highest possible score segmentation for all four regions using only one of the scan modalities. We ran our experiments on our CAE model to identify the region of interest. Any pixels that were not identified in the region of interest were given a label of 0, non tumour. We then ran experiments using the U-Net model for the final segmentation.

5.1. CAE model

The first model we analysed was the CAE model to determine the region that the tumour was located in. We ran this model with 1,531 training parameters using Binary Cross-Entropy Loss (BCE) as the loss function and Adam Optimisation with L^2-regularization at 0.01. With various learning rates, we had experimented with a selection of batch size and epochs. We observed that the learning rate for Adam Optimisation was a key factor in terms of the model fit in the validation data set. Setting it as 0.1, the model converged too fast and 0.001 was too slow. Having a learning rate too large might result in the algorithm failing to find the minima and the learning rate too small may trap the parameters in a local minima. Therefore, our final model was trained with 40 epochs, a batch size of 64 and a learning rate for Adam Optimisation of 0.01. We believe that our choice for the learning rate is a trade off between the training convergence speed and the risk of not locating the global minima point according to our loss function. The training time of our model was 2.1 hours.

Figure's 4 and 5 show the validation loss and accuracy curves across 40 epochs. The Flair scan preforms the best across all 4 scan modalities with a validation accuracy of 98.77%. T2 had an accuracy of 98.22% followed by similar accuracy's of 97.93% and 97.89% for T1 Contrast and T1 respectively.

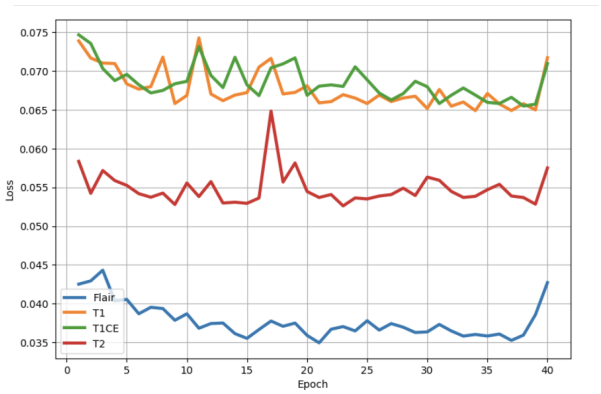


Figure 4: Validation Set Loss of CAE Model

5.2. UNet model

Now we move on to analyse the image segmentation. The result from our CAE model was used to indicate the area of the tumour scan which serves as an input for our U-Net model. The input is a cropped image of the brain scan which is significantly smaller than the original image of the full brain. During the training stage of this model, we cropped the training images according to the real location of the tumours. Hence, any scans without a tumour were discarded in the training phase. The architecture of this model is considerably more complex than our last model, as this model has 31,030,788 training parameters. This model again uses Binary Cross-Entropy Loss (BCE) as the

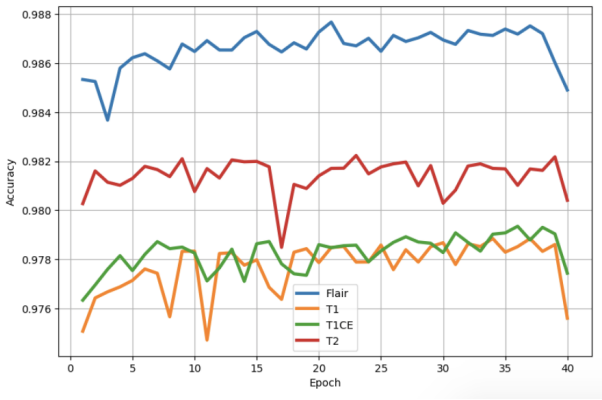


Figure 5: Validation Set Accuracy of CAE Model

loss function and Adam Optimisation with L^2 regularization. The same approach as the first model was employed to determine the best-suited model for the task, where we tested the model with various learning rates, epochs and different number of batch size. A similar effect to the model was observed by adjusting training rate for the Adam Optimiser. Thus, a learning rate of 0.0001 was chosen with a L^2-regularisation of 0.001 and we ran this for 50 epochs. The training time of the U-Net model was 12.2 hours.

In figures 6 and 7, the validation loss and accuracy are shown. Again Flair comes out as the scan with the best accuracy. The segmentation accuracy is 90.56%. The fact that Flair is the best scan modality for both the identification of the tumour and the image segmentation is the best case scenario. This leaves no doubt that the Flair scan is the overall best choice. The T1 Contrast had a similar image segmentation accuracy with 90.19% but it ranked third for the identification of the tumour area. The T2 modality had an accuracy of 88.71% and the T1 modality had an accuracy of 86.60%.

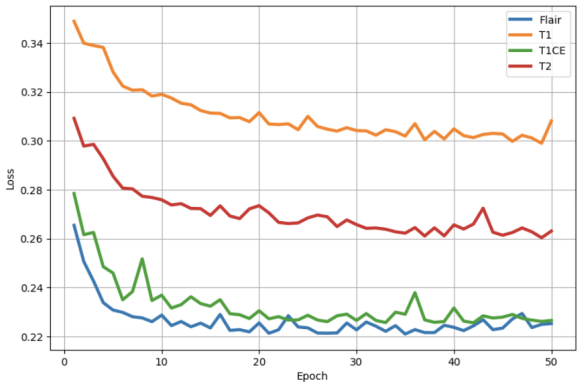


Figure 6: Validation Set Loss of U-Net Model

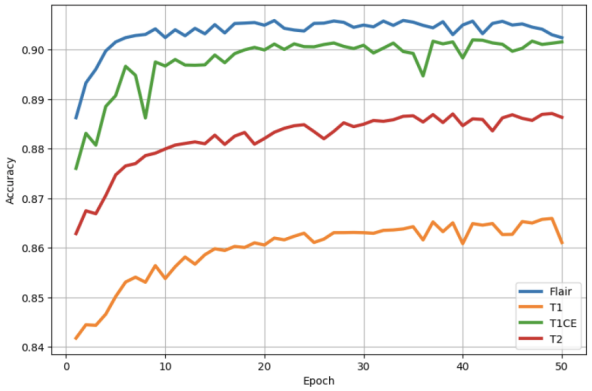


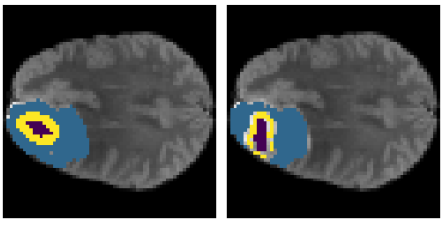
Figure 7: Validation Set Accuracy of U-Net Model

	Predicted Positive	Predicted Negative
Actual Positive	7599	824
True Negative	2194	8913

Table 1: Confusion Matrix

5.3. Dice Score

To predict for any unseen data, the 3D MRI scans are first converted into separate 2D slices. Each slice can be fed into the CAE model to identify any possible tumour region. Had a tumour been detected, the original slice of the MRI scan is cropped according to the ROI prediction. It is then fed into the UNet model to identify tumour segmentation within the region. Using the test dataset we had set aside, the Dice coefficient gave us a result of 97.84% accuracy. One important metric to note here is the false negative rate, where a tumour is found in the ground truth but was not detected in our ROI model. This amounts to about 4.2% of the data. Our results along with the ground truth can be visualised below. The ground truth tumour segmentations are on the left hand side and our tumour segmentations are on the right.



5.4. Future Work

Future work would be adopting further post-processing techniques in which we could use tumour characteristics to fill in pixels in which tumour pixels are located all around. Extensive research and knowledge surrounding tumours

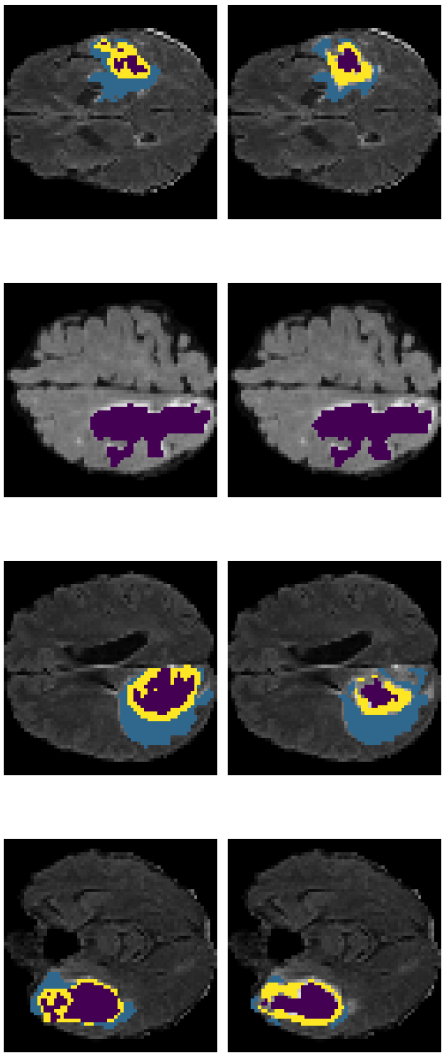


Figure 8: Ground Truth Segmentation vs Our Segmentation

would help with this task. With additional computational power, the model could be adapted to include all four scan modalities and this could be compared against the result of only using the Flair modality. Then a holistic evaluation could be made considering both accuracy and computation time.

Another model which we considered was the use of Detectron 2 built by Facebook AI Research in 2018. It is built on Pytorch and supports semantic segmentation. It uses maskrcnn as its benchmark and contains pre-trained models. These pre-trained models could be used as a starting point for brain tumour segmentation. One of the main limitations of our task is the complexity of the data and employing pre-trained models could help this issue. Fine-tuning could be used to reduce training times and obtain accurate results. This is an avenue to consider for future work.

6. Conclusion

In the project, the main hurdle we faced was the extremely large dataset. We largely reduced the data using pre-processing techniques and a two step procedure so we could run our image segmentation on our laptop. Every MRI scan comes with 4 scan modalities for each patient and our focus was to determine which one of these modalities would best assist accurate bran tumour segmentation. Our analysis determined that Flair is the best for the overall tumour segmentation. It returned the highest accuracy for both of our models and a Dice score of 97.84%.

References

[1] Brats. http://www.braintumorsegmentation.org. Accessed: 2024-04-02.

[2] Glioblastoma multiforme. https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Glioblastoma-Multiforme. Accessed: 2024-04-02.

[3] Kaggle. https://www.kaggle.com/datasets/dschettler8845/brats-2021-task1. Accessed: 2024-04-02.

[4] Shidong Li, Jianwei Liu, and Zhanjie Song. Brain tumor segmentation based on region of interest-aided localization and segmentation U-Net. *International Journal of Machine Learning and Cybernetics*, 13(9):2435–2445, Sept. 2022.

[5] Matia Martucci, Rosellina Russo, Francesco Schimperna, Gabriella D'Apolito, Marco Panfili, Alessandro Grimaldi, Alessandro Perna, Andrea Maurizio Ferranti, Giuseppe Varcasia, Carolina Giordano, and Simona Gaudino. Magnetic Resonance Imaging of Primary Adult Brain Tumors: State of the Art and Future Perspectives. *Biomedicines*, 11(2):364, Jan. 2023.

[6] Ramin Ranjbarzadeh, Abbas Bagherian Kasgari, Saeid Jafarzadeh Ghoushchi, Shokofeh Anari, Maryam Naseri, and Malika Bendechache. Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. *Scientific Reports*, 11(1):10930, May 2021.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, volume 9351, pages 234–241. Springer International Publishing, Cham, 2015. Series Title: Lecture Notes in Computer Science.

[8] Nicholas J Tustison, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee. N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, June 2010.

[9] Jason Walsh, Alice Othmani, Mayank Jain, and Soumyabrata Dev. Using U-Net network for efficient brain tumor segmentation in MRI images. *Healthcare Analytics*, 2:100098, Nov. 2022.

[10] Wenxuan Wang, Chen Chen, Meng Ding, Jiangyun Li, Hong Yu, and Sen Zha. TransBTS: Multimodal Brain Tumor Segmentation Using Transformer, June 2021. arXiv:2103.04430 [cs].