Multimodal Segmentation of Brain Tumors in 3D MRI Scans and Survival Prediction

Chaman Singh

Northeastern University
360 Huntington Avenue
Boston, Massachusetts, USA, 02115

singh.ch@northeastern.edu

Abstract: Brain tumour segmentation poses a challenging task even in the eyes of a trained medical practitioner. Advances in diagnostic imaging, surgical techniques and radiotherapy are being incorporated into routine clinical practice. Deep learning methods can help solve the problem of detecting tumor with precision and even segment it. We used a 3D deep convolutional neural network Unet for segmentation purpose. We used autoencoders to convert the input images into a latent feature distribution space for the task of predicting days of survival. We tested our network against the MICCAI BraTS 2020 dataset that comprised of 235 magnetic resonance imaging (MRI) scans and yielded a validation accuracy of 89.1% for whole tumor segmentation and a mean absolute error of 173.75 days in predicting the number of days of survival.

1. Introduction

There are two types of brain tumors: primary and secondary. Primary brain cancers develop from brain cells, whereas secondary tumors spread from other organs to the brain. Cancer is defined as the abnormal division of cells that results in the formation of masses of tissues that disrupt the human body's functions. These masses, known as tumors, have the ability to spread to other parts of the body (metastasis), latch on to vital organs, and obstruct normal functions. This may result in organ failure and the patient's death. Some tumors are benign, meaning they are generally localized and do not constitute a life-threatening hazard to the patient. In a study published in 2016 [Dasgupta et al. (2016)], 2% of malignant tumours are formed in the brain, mostly in the glial cells(classified as gliomas). 59.5% of these gliomas are highgrade, meaning they are malignant. Even though some gliomas are benign, they can be devastating because they disrupt normal brain activities. Brain tumor segmentation from MRIs is crucial for disease diagnosis, effective monitoring, and therapy planning. Manual delineation procedures necessitate anatomical expertise, are costly, time consuming, and subject to human error.

In the recent years, deep convolutional neural networks have been very successful for a variety of visual recognition tasks [Girshick et al. (2014)]. Neural networks can learn a hierarchical representation of features from the data by itself but suffer from a performance loss in limited data size settings. Overcoming this issue, Unet [Ronneberger et al. (2015)] is a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. It has been shown that Unets can be trained end-to-end from

very few images and perform well.

In this paper, we implement a 3D Unet architecture for the segmentation task of each class. To reduce the number of features of the input image, we model the latent space using autoencoder and variational autoencoder and apply support vector regression (SVR) to predict the survival days. The paper is divided into various sections. Section 2 talks about the BraTS dataset, section 3 presents some related works on this kind of problems. Section 4 introduces the approach followed in this project and then we discuss the experiments and conclude and briefly explore some future scopes to improve upon in sections 5 and 6.

2. Dataset

We used MICCAI BraTS 2020 dataset for this project [Bakas et al. (2017)]. 235 Routine clinically-acquired pre-operative multimodal MRI scans of glioblastoma (GBM/HGG) and lower grade glioma (LGG), with pathologically confirmed diagnosis and overall survival are provided as the training and validation data. All BraTS multimodal scans are available as NIfTI files (.nii.gz) and describe a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (T2-FLAIR) volumes, and were acquired with different clinical protocols and various scanners from multiple institutions. All the imaging datasets have been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations comprise the GD-enhancing tumor (ET - label 4), the peritumoral edema (ED - label 2), and the necrotic and non-enhancing tumor core (NCR/NET - label 1), as described both in the BraTS 2012-2013 TMI paper [Menze et al. (2015)] and in the latest BraTS summarizing paper [Bakas et al. (2018)]. The provided data are distributed after their pre-processing, i.e., co-registered to the same anatomical template, interpolated to the same resolution $1mm^3$ and skull stripped. The overall survival (OS) data, defined in days, are included in a comma-separated value (.csv) file with correspondences to the pseudo-identifiers of the imaging data. The .csv file also includes the age of patients, as well as the resection status. BraTS 2020 dataset also contains 133 images in test set with available segmentation labels. But the test set does not have overall survival data, hence we did not use it in this project at all but will be a good candidate to enrich our dataset in future experiments.

The four types of MRI scans[1] bring a unique variation to the resulting image and are described below:

- T1-weighted: It is a native image and it uses short echo time (TE) and short relaxation time to measure spin-lattice relaxation (TR). It works well for reducing signal from high watery areas and increasing signal from fatty parts, improving the scan quality of tumors.
- T2-weighted: It is a native image and uses a long Relaxation Time and a long echo time to measure the spin-spin relaxation. This is the inverse of T1 and as a result, we see greater signal in watery areas and less signal in fatty ones.
- T2-FLAIR (Fluid Attenuated Inversion Recovery): It is a T2 -weighted native image and suppresses fluid imagery by setting an inversion time. This helps in suppressing cerebrospinal fluid and exposes hyper lesions.
- T1 weighted with Contrast Enhancements (T1Gd): T1 weighted MRI scan subjected to post contrast enhancement, enhancing the signals from tumours substantially.

It is reported that T1Gd and the T2-FLAIR volumes were most useful to produce the ground truth segmentations. The ground truth of the scans consists of 3 different regions labelled as

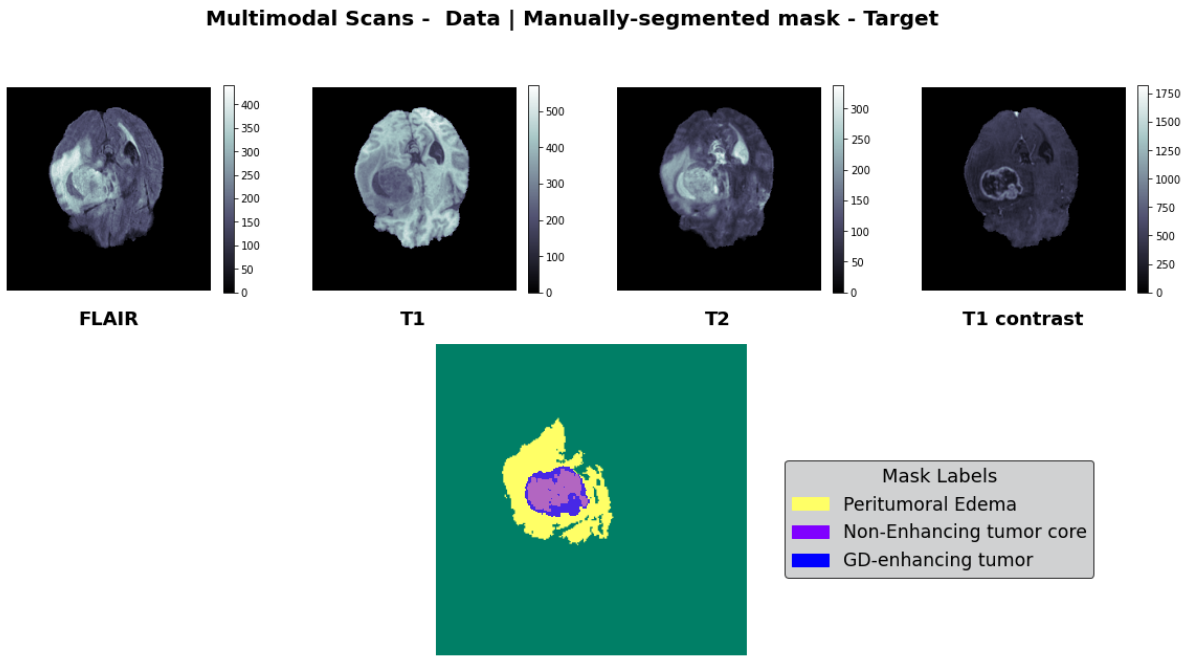

Fig. 1. 2D representation of the four modalities and the ground truth mask in a MRI scan

different modalities. The union of all three labels gives the whole tumor segmentation (WT).

- Tumor Core (TC): The non enhancing core of the tumors consisting of dead tissue.
- Enhancing Tumour (ET). The region where the tumor cells actively divide
- Edema: The region mostly consists of swollen tissue filled with fluids.

3. Related Works

Unets [Ronneberger et al. (2015)] and its variants have been successfully employed for biomedical image segmentation over different datasets. [Dutta et al. (2020)] achieved 58.3% accuracy using time distributed Unet (TD-Unet) on BraTS 2015 dataset. [Myronenko (2018)] implemented a variational autoencoder architecture for brain tumor segmentation and reported a validation Dice score of 91% for WT segmentation using an ensemble of 10 models. [Huang et al. (2021)] used NLSE-VNet model, which integrates the Non-Local module and the Squeeze-and-Excitation module into V-Net to segment three brain tumor sub-regions in multimodal MRI on the BraTS 2019 and BraTS 2020 datasets. They reported an average Dice of brain tumor segmentation tasks up to 79% and an average RMSE of the survival predictive task as low as 311.5 days. [Anand et al. (2021)] used a 3D fully convolutional neural networks with hard mining and achieved a Dice score of 0.85 for WT segmentation task and an accuracy of 0.448 for the survival prediction task on BraTS 2020 dataset. [Ben Ahmed et al. (2022)] yielded an accuracy of 74% on BraTS 2019 dataset for the survival prediction task using an ensemble of 3D deep convolutional networks.

4. Approach

Image segmentation involves identifying object(s) in an image and separating it from the rest of the image. Unet[2] is a very popular learning paradigm in the field of medical image segmentation

as it can work on less amount of data as well which is typically the case of annotated medical imaging datasets.

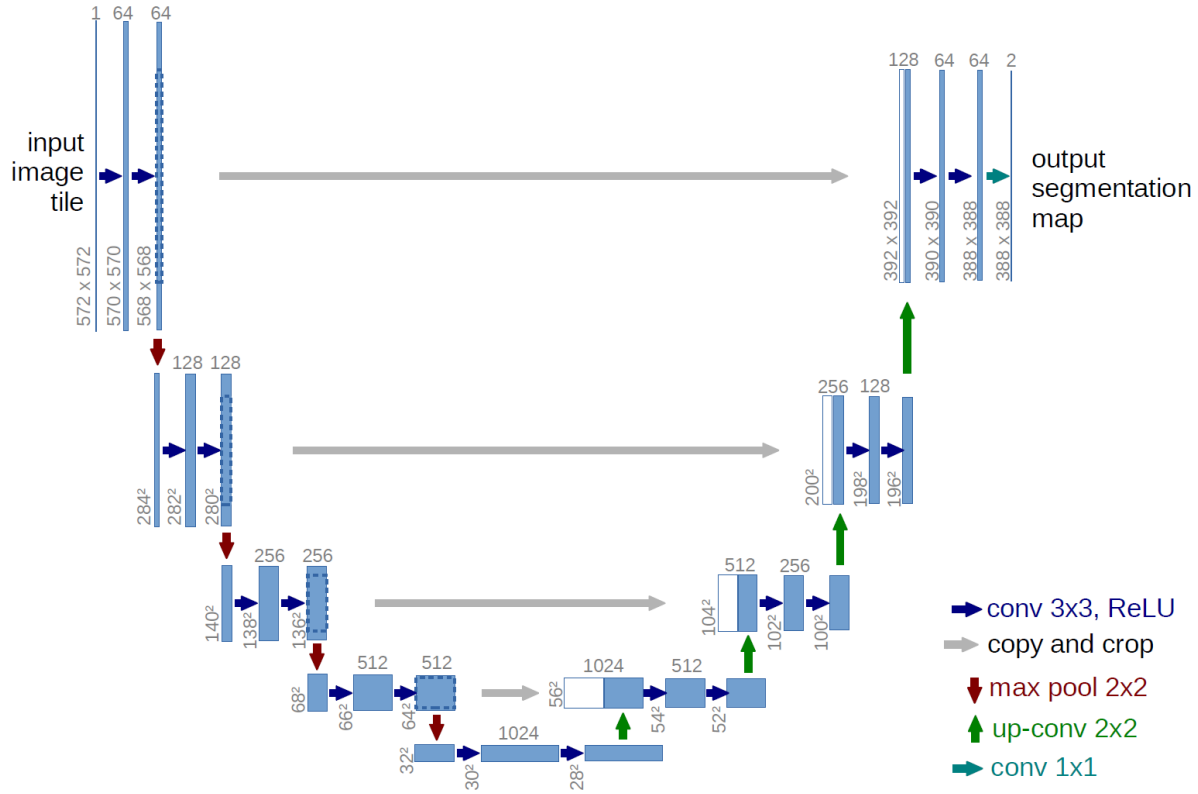

Fig. 2. Unet Architecture for 2D images from the original paper by Ronneberger et al. (2015). A similar architecture is used to report results in this paper for processing 3D images

4.1. Unet

Figure 2 illustrates the Unet architecture. It has a shrinking path on the left and an expansive path on the right (right side). The contracting path follows the standard convolutional network architecture. It comprises of two 3x3 convolutions (unpadded convolutions) that are applied repeatedly, each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling.We double the number of feature channels with each downsampling step.

Every step in the expansive path consists of an upsampling of the feature map followed by a 2x2 convolution (up-convolution) that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. Due to the loss of boundary pixels in every convolution, cropping is required. A 1x1 convolution is employed at the final layer to convert each feature vector to the desired number of classes. The 3D Unet used in this project consisted of 5,652,507 trainable parameters. The table of number of parameters in each layer is not provided here due to its large size. In our model, we are minimizing for the binary cross entropy (BCE) and dice error (1 - dice score).

4.2. *Autoencoder*

Autoencoder is an encoder-decoder based CNN architecture with an asymmetrically larger encoder to extract image features and a smaller decoder to reconstruct the segmentation mask [Chen et al. (2018)]. The encoder part consists of a repeat of convolution layers followed by max pooling layers. The decoder structure is similar to the encoder one, but with deconvolution layers followed by unpooling layers. The encoder and decoder parts are connected through a linear layer. The autoencoder model Figure 3 used in this project consisted of 390,827,364 trainable parameters in total. We optimize the model for minimum mean squared error loss.

```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
           Conv3d-1    [-1, 16, 153, 238, 238]           1,744
         MaxPool3d-2  [[-1, 16, 76, 119, 119], [-1, 16, 76, 119, 119]]               0
           Conv3d-3     [-1, 32, 74, 117, 117]          13,856
         MaxPool3d-4  [[-1, 32, 24, 39, 39], [-1, 32, 24, 39, 39]]               0
           Conv3d-5      [-1, 96, 23, 38, 38]          24,672
         MaxPool3d-6  [[-1, 96, 11, 19, 19], [-1, 96, 11, 19, 19]]               0
           Linear-7                  [-1, 512]     195,183,104
           Linear-8               [-1, 381216]     195,563,808
      MaxUnpool3d-9       [-1, 96, 23, 38, 38]               0
  ConvTranspose3d-10       [-1, 32, 24, 39, 39]          24,608
     MaxUnpool3d-11     [-1, 32, 74, 117, 117]               0
  ConvTranspose3d-12     [-1, 16, 76, 119, 119]          13,840
     MaxUnpool3d-13    [-1, 16, 153, 238, 238]               0
  ConvTranspose3d-14     [-1, 4, 155, 240, 240]           1,732
================================================================
Total params: 390,827,364
Trainable params: 390,827,364
Non-trainable params: 0
----------------------------------------------------------------
```

Fig. 3. The Autoencoder network used for modelling the latent space for survival days prediction

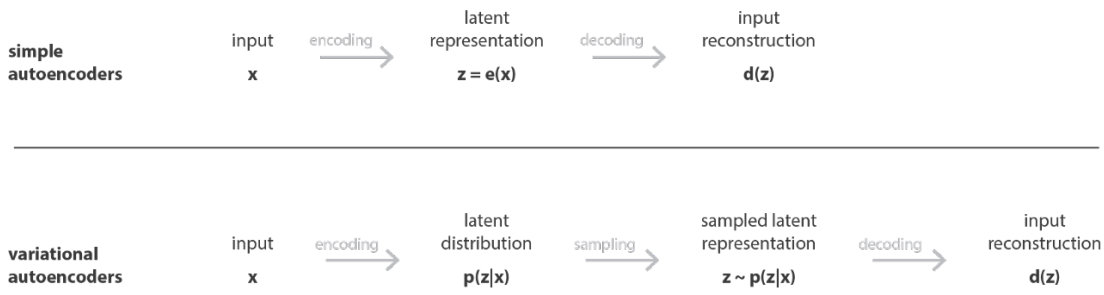

Fig. 4. Autoencoder and variational autoencoder modelling comparison

4.3. *Variational Autoencoder (VAE)*

A variational autoencoder [Kingma and Welling (2013)] can be thought as an autoencoder with regularized training to avoid overfitting and ensuring that the latent space has acceptable properties that enable generative process. In VAE, instead of encoding the input as a single

point, we encode it as a distribution over the latent space. The model training follows these steps repeatedly:

1. The input is encoded as a distribution over the latent space
2. A point from the latent space is sampled from that distribution
3. The sampled point is decoded to compute the reconstruction error
4. The reconstruction error is backpropagated through the network

Starting from the encoder endpoint output, we first reduce the input to a low dimensional space of 512. Then, a sample is drawn from the Gaussian distribution with the given mean and standard deviation, and reconstructed into the input image dimensions following the same architecture as the decoder. L_{KL} is standard VAE penalty term [Doersch (2016)], a KL divergence between the estimated normal distribution $\mathcal{N}(\mu, \sigma^2)$ and a prior distribution $\mathcal{N}(0, 1)$, which has a closed form representation where N is total number of image voxels

$$L_{KL} = \frac{1}{N}\sum \mu^2 + \sigma^2 - \log \sigma^2 - 1$$

$$Loss = L_{reconstruction} + L_{KL}$$

Where $L_{reconstruction}$ is modeled as mean squared error between the ground truth and the reconstruction and we train the network to minimize *Loss*.

4.4. Evaluation Metrics

We use Sorensen–Dice coefficient to evaluate tumor segmentation. Dice coefficient is used to gauge the similarity of two samples (predicted mask and ground truth) and is defined as:

$$\mathrm{DSC} = \frac{2|\mathrm{X} \cap \mathrm{Y}|}{|\mathrm{X}| + |\mathrm{Y}|}$$

We also report the Jaccard similarity coefficient between the predicted mask and the ground truth. Jaccard index is defined as:

$$\mathrm{J} = \frac{|\mathrm{X} \cap \mathrm{Y}|}{|\mathrm{X} \cup \mathrm{Y}|}$$

To evaluate the performance of the number of days of survival task, we are going to use the mean absolute error (MAE) between the predicted days of survival and the provided ground truth.

5. Experiments

We use the batch size of 1 and the input images are of dimensions 155x240x240. We use Adam optimizer to adjust the learning rate during the training process with a starting learning rate of 5e-4 for all the three models reported in this paper. We train the models until the training loss converges. For reporting purpose, we also measure the validation loss after every epoch but no feedback was given to the model from validation data as the validation was used to report the final results here. Figure 5 shows the training progress of the Unet model used for the tumor segmentation task while figure 6 illustrates variations in the Dice score during the training phase.

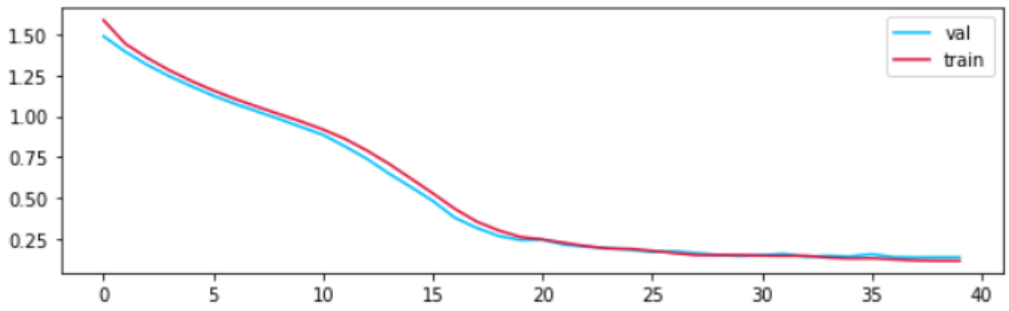

Fig. 5. Unet model training loss for train and validation data per epoch for the tumor segmentation task

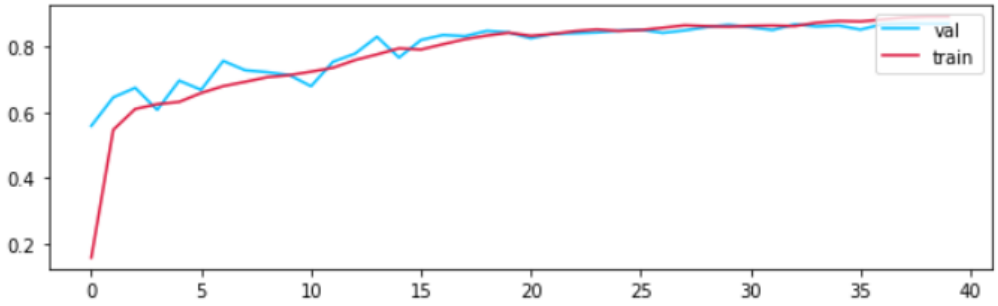

Fig. 6. Unet model DICE score for train and validation data per epoch for the tumor segmentation task

Figure 7 shows the ground truth comparison with the prediction by the Unet model for a randomly selected sample for the whole tumor (WT) segmentation task. We can see that the prediction is quite accurate. Figure 8 shows the variational Autoencoder reconstruction using the latent distribution features for a randomly selected sample and seems to be a good approximation of the ground truth. Similar results were observed with autoencoder but are not reported here.

For the days of survival prediction task, we fed the latent features from the autoencoder models to a support vector regression model using sigmoid kernel function. We also tried using multi layer perceptron regressor to model the relationship among the input features and the target label. Both, SVR and MLP regressor models gave similar results and clearly suffered from the lack of data samples available for training compared to the input feature space. So it may be good experiment to explore the latent feature space of size 256 or less.

Table 1. Survival day prediction over validation set

Algorithm	Mean Absolute Error (Days)
Autoencoder	**173.75**
Variational Autoencoder	173.80

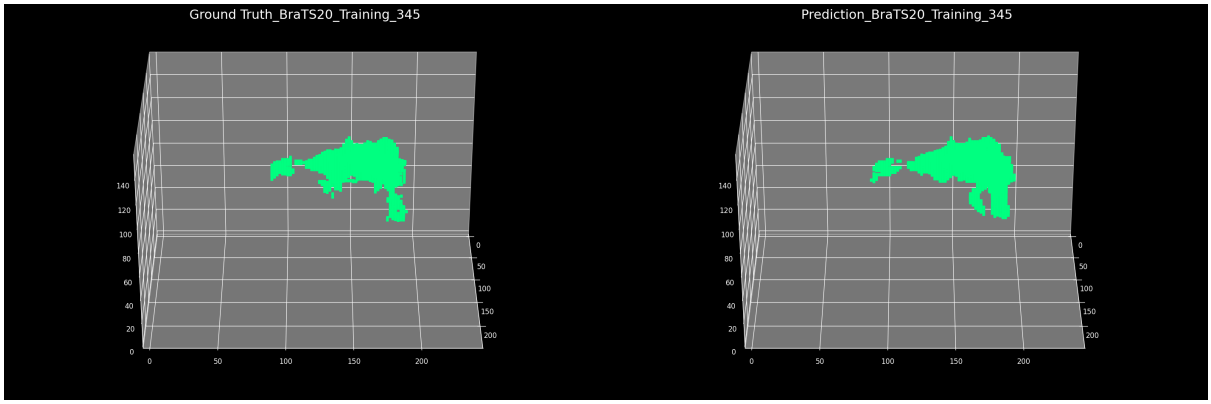

Fig. 7. Ground truth vs prediction of a sample for whole tumor (WT) segmentation

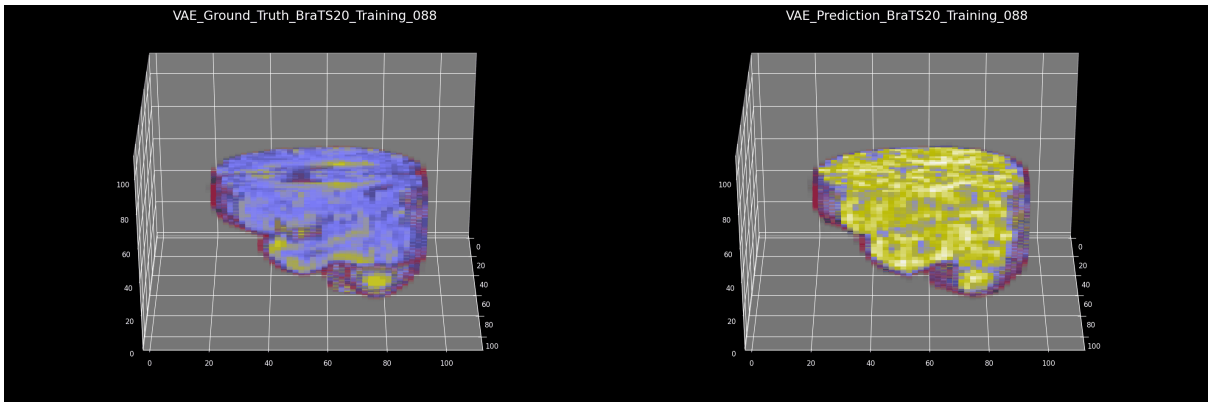

Fig. 8. Variational Autoencoder reconstruction (Ground truth vs prediction) for a sample using the latent distribution features

6. Conclusion

The effectiveness of Unet for medical image segmentation has been long proven. In this project, we achieved an accuracy of 89.1% using a single model without any parameter tuning. The accuracy is very close to the highest reported in the literature, 91%, using an ensemble of 10 variational autoencoder models. Applying autoencoder/ VAE models for the task of modelling reduced feature dimension latent space seems like a promising idea as it yielded a smaller mean absolute error (173.75 vs 311 days) than the ones reported in previous works for previous versions of dataset. But more experiments are required to substantiate this claim. Furthermore, we expected variational autoencoder model to perform better than the simple autoencoder model given its regularization property but that was not observed here.

6.1. Improvements

Due to the time and computing resources constraints, hyper parameter tuning was not performed to select the best set of parameters for the models used. The tuning to optimal parameters is expected to improve the performance. The strength of variational autoencoder is that it can be used to generate samples in a limited data environment but we did not use that option in this paper. It will be interesting to see the effect of generated samples using Variational autoencoder or generative adversarial networks (GANs) on the performance of both the segmentation and survival prediction models. In this project, only the basic versions of Unet and VAE were

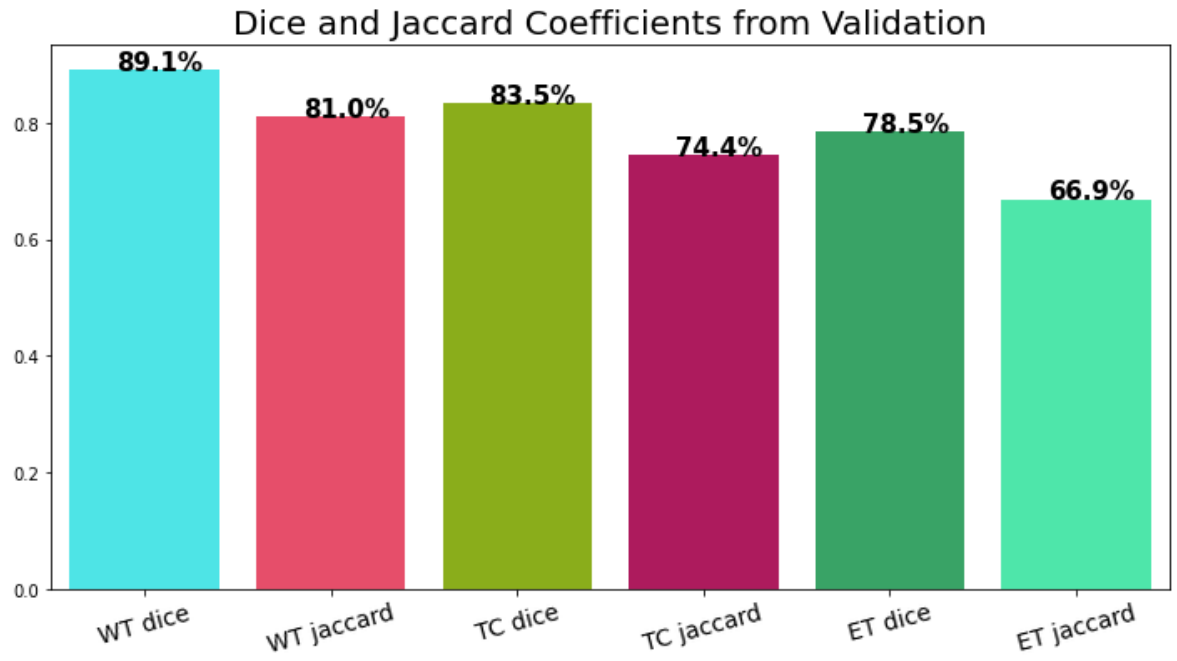

Fig. 9. Mean Dice and Jaccard score for each class of the tumor segmentation task

explored and with the availability of more advanced models of Unet, VAE, the results may be further improved. The paper describing the human annotation process states that the T2-FLAIR and T1 with contrast images were most helpful in the annotation. So it will be interesting to study the variation in the performance of the models using only one or a combination of the modalities. Given the small size of the training data, we can also explore the effect of simple data augmentation such as image flip and rotation. Moreover, we did not utilize the test set in this project and utilizing it in a supervised or semi supervised way can potentially improve the model performances.

References

Brats2020 dataset (training + validation), Jul 2020. URL https://www.kaggle.com/datasets/awsaf49/brats20-dataset-training-validation.

Vikas Kumar Anand, Sanjeev Grampurohit, Pranav Aurangabadkar, Avinash Kori, Mahendra Khened, Raghavendra S Bhat, and Ganapathy Krishnamurthi. Brain tumor segmentation and survival prediction using automatic hard mining in 3d cnn architecture, 2021. URL https://arxiv.org/abs/2101.01546.

Eugenia Anello. Variational autoencoder with pytorch, Mar 2022. URL https://medium.com/dataseries/variational-autoencoder-with-pytorch-2d359cbf027b.

Awsaf. Brats2020 dataset (training + validation), Jul 2020. URL https://www.kaggle.com/datasets/awsaf49/brats20-dataset-training-validation.

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data*, 4:170117, September 2017.

Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, and Alessandro Crimi et. al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge, 2018. URL https://arxiv.org/abs/1811.02629.

Kaoutar Ben Ahmed, Lawrence O Hall, Dmitry B Goldgof, and Robert Gatenby. Ensembles of convolutional neural networks for survival time estimation of High-Grade glioma patients from multimodal MRI. *Diagnostics (Basel)*, 12(2), January 2022.

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation, 2018. URL https://arxiv.org/abs/1802.02611.

Archya Dasgupta, Tejpal Gupta, and Rakesh Jalali. Indian data on central nervous tumors: A summary of published work. *South Asian J Cancer*, 5(3):147–153, July 2016.

Carl Doersch. Tutorial on variational autoencoders, 2016. URL https://arxiv.org/abs/1606.05908.

Jeet Dutta, Debajyoti Chakraborty, and Debanjan Mondal. Multimodal segmentation of brain tumours in volumetric mri scans of the brain using time-distributed u-net. In Asit Kumar Das, Janmenjoy Nayak, Bighnaraj Naik, Soumen Kumar Pati, and Danilo Pelusi, editors, *Computational Intelligence in Pattern Recognition*, pages 715–725, Singapore, 2020. Springer Singapore. ISBN 978-981-13-9042-5.

Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

He Huang, Wenbo Zhang, Ying Fang, Jialing Hong, Shuaixi Su, and Xiaobo Lai. Overall survival prediction for gliomas using a novel compound approach. *Frontiers in Oncology*, 11, 2021. ISSN 2234-943X. doi: 10.3389/fonc.2021.724191. URL https://www.frontiersin.org/article/10.3389/fonc.2021.724191.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013. URL https://arxiv.org/abs/1312.6114.

Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, and Justin Kirby et. al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015. doi: 10.1109/TMI.2014.2377694.

Andriy Myronenko. 3d mri brain tumor segmentation using autoencoder regularization, 2018. URL https://arxiv.org/abs/1810.11654.

Polomarco Polomarco. $Brats20_3dunet_3dautoencoder, Nov2020. URL$.

Joseph Rocca. Understanding variational autoencoders (vaes), Mar 2021. URL https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73.

O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. URL http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a. (available on arXiv:1505.04597 [cs.CV]).