

---

# Project Proposal - ECE 176

---

Ishan Agrawal  
ECE  
A18869769

Hao Lin  
Mathematics  
A18932572

## Abstract

Brain tumors represent a serious threat to human health due to their ability to disrupt essential brain functions. In particular, Gliomas are the most prevalent collection of brain tumors which form in the Central Nervous System (CNS), impacting motor response Holland (2001). Early and accurate detection of brain tumors from magnetic resonance imaging (MRI) scans is a critical step in improving patient outcomes. However, manual reading of these scans is not only a subjective and time-consuming task, but also requires expert radiologists Bakas et al. (2017). Thus, with deep learning showing success in semantic segmentation, we propose automating this process using a 3D U-Net architecture to predict segmentation masks from input MRI scans. We believe that the encoder-decoder structure along with the skip connections the U-Net has will allow it to find both low-level details and high-level context within the data. Ronneberger et al. (2015). For our data, we will be using the BraTs2020 dataset which contains roughly 370 individual brain scans. Additionally, we intend to experiment with Attention based U-Nets which incorporate learned attention gates to help the model de-filter noise within the scans Oktay et al. (2018). To evaluate these models, we will be using the Dice Similarity Coefficient (DSC) which measures the overlap between the two segmentation masks by rewarding true positives and penalizing false positives and false negatives. Additionally, we will be using the Average Hausdorff Distance (AHD) metric which measures the average of the shortest distances between the ground truth and the segmentation mask Müller et al. (2022).

## 1 Problem Definition

Primary malignant brain tumors, also known as intracranial tumors, are abnormal masses that form within the brain and divide uncontrollably McFaline-Figueroa and Lee (2018). In the case of high-grade malignant Gliomas, the median survival length is between two and three years. Even the less aggressive variants such as astrocytomas and oligodendroglomas, have relatively short survival periods Ohgaki and Kleihues (2005). Therefore, due to the aggressive nature of the cancer, neuroimaging is used constantly to evaluate the progression of the condition as well as the efficacy of the current treatment. However, reading these scans is time-consuming and often inconsistent. Therefore, errors in diagnosis can cause the cancer to grow and metastasize, potentially spreading to other regions of the body if not discovered. Additionally, even if treatment shows success in eliminating the vast majority of cancerous cells, reemergence is still possible.

For this reason, the detection and monitoring stage represents a critical part in the fight against brain tumors. Successful detection in the early stages significantly improves patient prognosis and reduces the risk of severe neurological damage. We believe that the 3D U-Net architecture, with its encoder-decoder structure and its skip connections makes it well suited for semantic segmentation tasks, as not only can it capture high-level features but it can also retain information from previous layers. In addition to this, we have decided to compare the results to the Attention U-Net architecture to see if the attention mechanism can help the model focus on cancerous cells while de-emphasizing healthy cells.

## 2 Tentative Method

To begin with, we will be using the NiBabel Python package to visualize and read the .nii.gz files from the dataset. This package is widely used neuroimaging applications for viewing and manipulating data. For our case, it is useful because it provides functions for converting the scans into numpy arrays. Since we are given four different scans for each brain, I will be stacking each feature to make a four channel 3D input for the model. In our case, data is constrained as we only have 370 scans in the dataset in total. Therefore, since most of the scans are unlabelled as they contain a lot of empty space, we will be cropping the arrays. If this wasn't done then the model may develop a bias towards under-predicting the size of the segmentation mask, since the cancerous region would represent a very small portion of the scan. Additionally, this is also important because improves the rate at which model learns, as less parameters are needed for smaller input sizes. After this, we will normalize the data by using a MinMax scaler to map all the data points in the scan from zero to one.

Next, we will be constructing the U-Net model architecture in PyTorch. For the encoder, the architecture will include four layers of two 3D convolutional layers with a kernel size of three followed by the ReLU activation function. Each time we downsample the number of output features will double, and we will apply a max-pooling layer with kernel and stride of two, halving the spatial dimensions of the output features after each layer Ronneberger et al. (2015). When we reach the bottleneck, additional convolutional and ReLU layers will be added, further down-sampling the image while increasing the number of channels. The decoder layer will consist of four layers of transpose convolutions with a kernel size of two and a stride of two, doubling the spatial dimensions of the scans. To preserve fine-grained spatial information, skip connections are used to concatenate the feature maps from each encoder layer with the corresponding decoder input. This allows the model to combine high-resolution localization features with deep semantic information, improving segmentation accuracy, particularly around object boundaries. For the Attention U-Net, we will be adding attention gates which contain learned attention weights which are applied to the skip-connections from the encoder layers Oktay et al. (2018). Finally, in the last layer, we will apply a softmax activation function over the feature dimension for our prediction.

From there, we will begin to train the model. We will be experimenting with various hyper-parameters such as the learning rate, batch size, and optimizer used. We may additionally incorporate features such as LayerNorm which normalizes across the feature dimension of each sample independently so that the mean is zero and the standard deviation is one Ba et al. (2016). This is important because without LayerNorm the weights in the next layer is highly dependent on the weights in the current layer, which can cause a vanishing gradient problem as the network becomes too large. However, we may also decide not to include this because it may cause training times to increase a lot, since it is very computationally expensive to compute the mean and variance for each training example, especially as the hidden sizes become larger.

## 3 Experiments

The BraTs2020 dataset is a dataset from Kaggle that contains MRI scans in the form of .nii.gz files, which are 3D scans. The dataset is intended to solve four tasks: segmentation of gliomas, prediction of patient survival rate, clinical evaluation of progression status, and uncertainty estimation for predicted tumor subregion. However, for our proposal, we will be focusing on the segmentation of gliomas in the brain. The dataset contains four different channels which represent different MRI imaging modalities to highlight a wide range of different features. In particular, the dataset contains one native and one contrast-enhanced T1 image, which highlights the fatty tissues in the brain. The dataset also contains T2 and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes, which highlight brain tissue with water in them. Since the vast majority of the scan is actually un-labelled empty space, we will be cropping the scans before we train the model.

During training, we will be experimenting evaluating our model on DSC and AHD to see which loss function helps the network converge faster and produce more accurate segmentation masks through their intersection over the union. DSC rewards true positive results, while penalizing false negatives and false positives by taking the ratio between the true positives and the sum of the other components Müller et al. (2022). In contrast, AHD measures the similarity between two sets by taking the average of the distance from every point in the first set and its closest point in the other set. The direction is then reversed and the two quantities are averaged to get the final AHD Müller et al. (2022).

## References

- Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *arXiv preprint arXiv:1607.06450*.
- Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., and Davatzikos, C. (2017). Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13.
- Holland, E. C. (2001). Progenitor cells and glioma formation. *Current opinion in neurology*, 14(6):683–688.
- McFaline-Figueroa, J. R. and Lee, E. Q. (2018). Brain tumors. *The American journal of medicine*, 131(8):874–882.
- Müller, D., Soto-Rey, I., and Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1):210.
- Ohgaki, H. and Kleihues, P. (2005). Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *Journal of Neuropathology & Experimental Neurology*, 64(6):479–489.
- Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.