S.L.I.C.E. M.E.D.

Segmenting **L**esions for **I**mproved **C**are, **E**nhanced **M**anagement, and **E**xact **D**iagnosis

By Chase Guenther

UNIVERSIDAD DE CHILE

NDSU Computer Science

Introduction

Multiple Sclerosis (MS) is a debilitating neuroinflammatory disease characterized by demyelination, or damage to the protective sheath surrounding nerve fibers in the brain and spinal cord. Magnetic Resonance Imaging (MRI) plays a vital role in diagnosing and monitoring MS by visualizing lesions, areas of abnormal signal intensity, within the brain. Accurate segmentation of these lesions is crucial for several reasons. It can:

- Improve diagnostic accuracy and efficiency: Automated segmentation can reduce the time radiologists spend on manual segmentation, allowing them to focus on interpretation and diagnosis.
- Increase objectivity and consistency: Manual segmentation can be subjective and prone to variability between readers. Automated methods provide more consistent results.
- Quantify disease burden: Accurate segmentation allows for measurement of lesion volume, a valuable biomarker for tracking disease progression and treatment response.

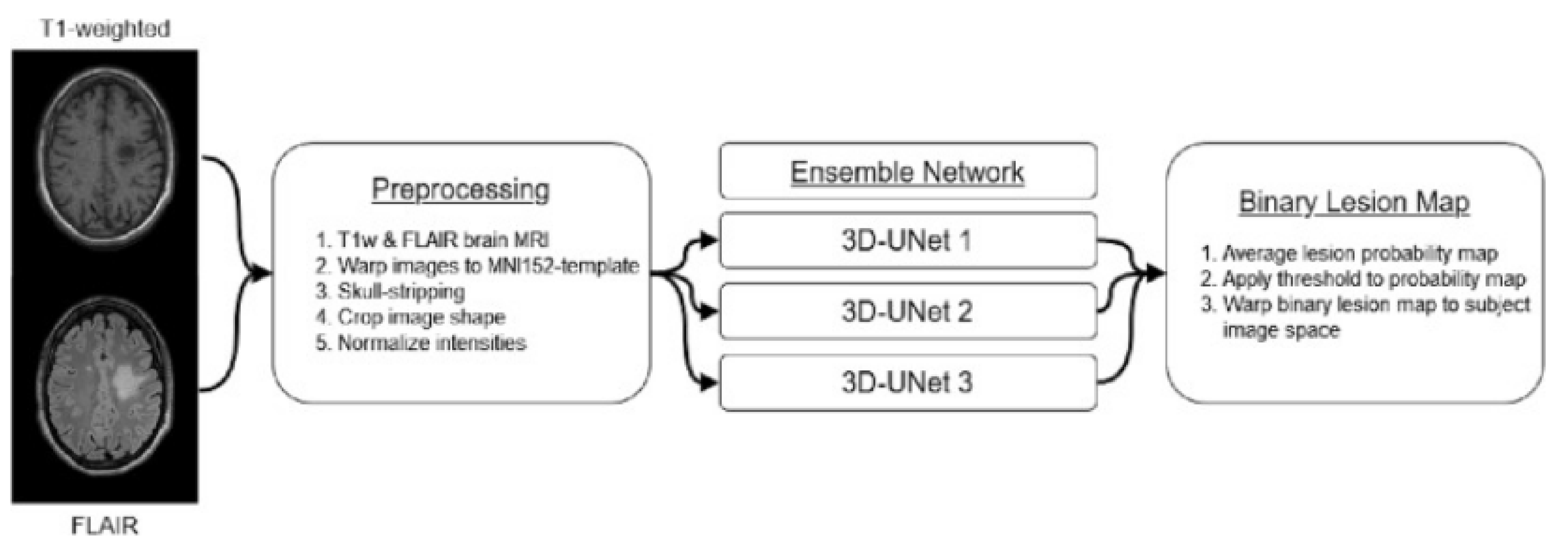

Figure 1. LST-AI program design concept.

What is the Problem?

Currently, although results are promising for image segmentation of brain lesions, they are still lacking. For most algorithms dice scores do not achieve numbers above .7, and even the best is only barely surpassing that on benchmark datasets. A more accurate and robust algorithm would allow for better diagnosis and tracking of MS.

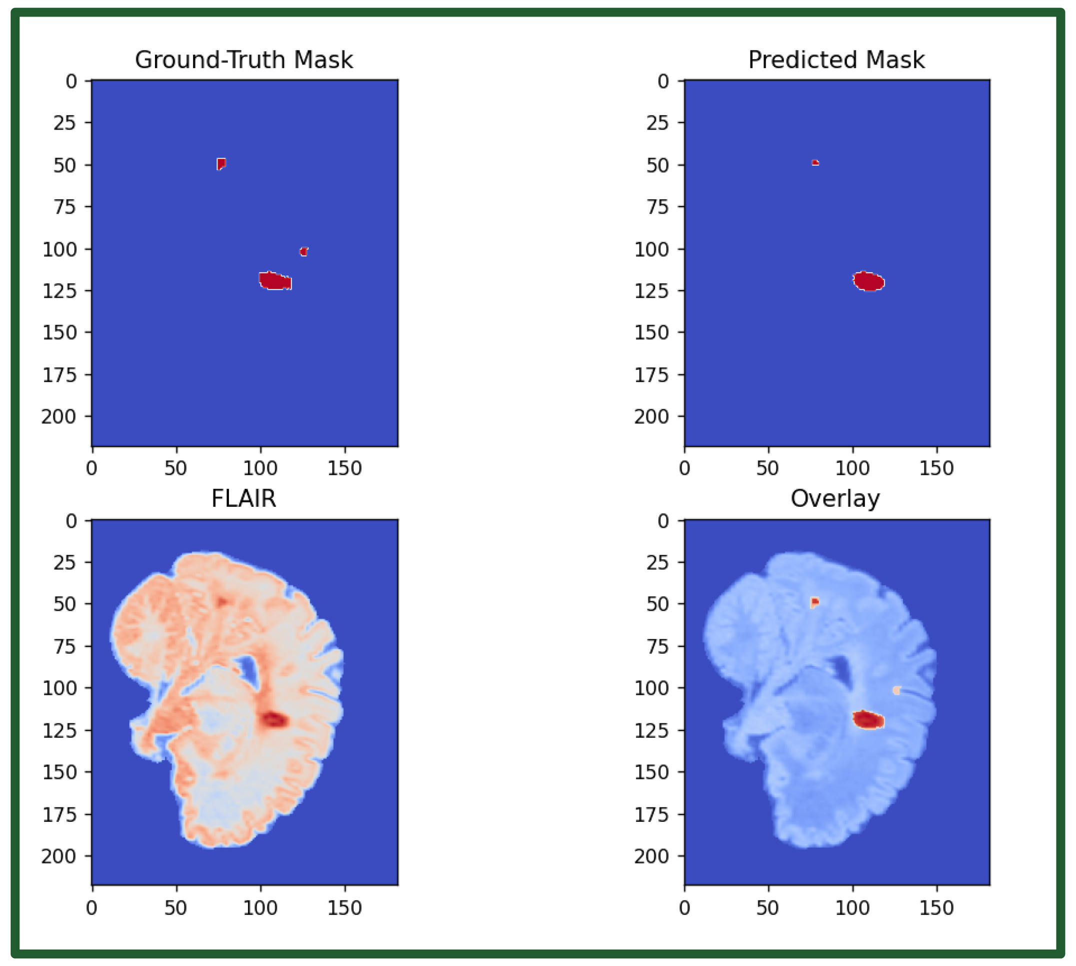

Figure 2. Example output using LST-AI base models on our dataset.

Our Methodology and Approach

Hardware. The Training and results of the tests will be obtained using North Dakota State University's supercomputer, "CCAST". To be able to run the program within a reasonable time, due to some of the more computationally intensive tasks of image segmentation techniques, we had to choose a powerful machine. WE are using 16 CPU cores and 1 a100 Nvidia GPU for my current testing efforts. It was found that even with these powerful numbers, more memory would be needed on the GPU to be able to contain the model if we wanted to make it any larger.

Dataset. The training data is made up of 56 Patients at a total of 93 time points. At each time point, there is a T1 weighted (T1w), T2 weighted (T2w), Fluid attenuated inversion recovery (FLAIR) MRI scan, and ground truth segmentation mask. Normally, MRI scans start out in DICOM format (Digital Imaging and Communications in Medicine), but these images have already been partially preprocessed. The extra DICOM information was discarded, and the images were converted to NIfTI (Neuroimaging Informatics Technology Initiative), a 3D image file format. Each image was also skull-stripped, removing the skull and leaving the brain tissue. Each 3D is in the shape 182 x 218 x 182 (Width x Height x Depth). To briefly explain the difference between the types of MRI scans, The T1w images darken lesions and both T2w and FLAIR images brighten lesions, however FLAIR will suppress cerebrospinal fluid (CSF), which appears bright in T2 images. The testing data, although similar to the training data, does not contain time points or masks. Instead, there is only one time point for each of the 22 patients. Each patient file still contains a T1w, T2w, and FLAIR image of the NifTI format.

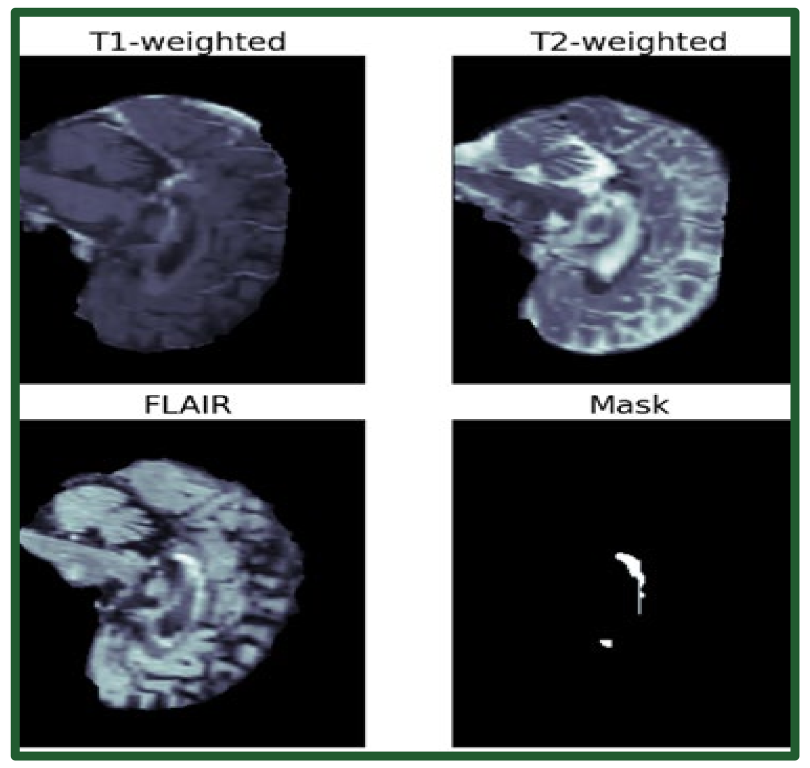

Figure 3. Example image data from our dataset.

Method. The foundation of our approach is an ensemble of 3D U-Net architectures. The U-Net is a convolutional neural network architecture specifically designed for medical image segmentation tasks. It consists of a contracting (encoder) path and an expanding (decoder) path.

- Contracting Path: These convolutional layers progressively reduce the spatial dimensions (width, height, and depth) of the feature maps while increasing the number of feature channels to learn complex representations of the input. It typically consists of several repeated blocks, each containing two or three 3D convolutional layers (Conv3D) followed by a non-linear activation function (e.g., Leaky ReLU) and batch normalization.
- Expanding Path: The expanding path upsamples the captured features and integrates them with high-resolution features from the contracting path to produce a segmentation map with high spatial resolution. It utilizes transposed convolution (Conv3DTranspose) layers to increase the spatial dimensions and concatenate feature maps from the corresponding contracting path block at each level.

In our implementation, we address the computational complexity of the U-Net architecture by replacing the standard Conv3D layers in both the contracting and expanding paths with depthwise separable convolution (DSC) layers. A DSC consists of two separate steps: a depthwise convolution and a pointwise convolution ([Chollet, 2017]).

- Depthwise Convolution: In the depthwise convolution, separate filters are applied to each input channel, extracting features independently. This significantly reduces the number of computations compared to a standard convolution, as it avoids learning redundant filters for capturing spatial correlations across channels
- Pointwise Convolution: The pointwise convolution then uses 1x1x1 filters to project the outputs from the depthwise convolution layer into a new feature space, introducing the non-linearity and combining features across channels. We use a 1x1x1 conv3D layer to fulfill this purpose.

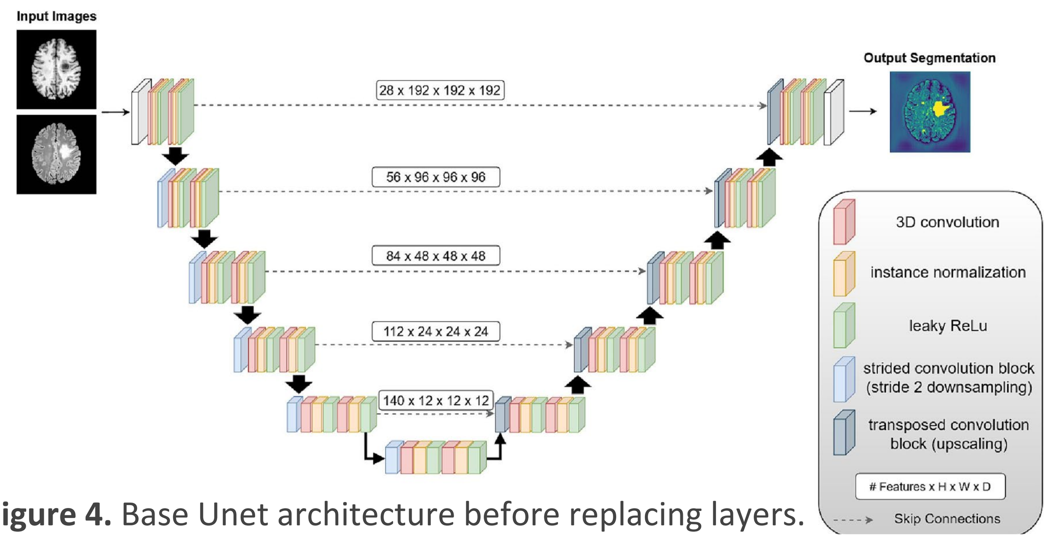

Figure 4. Base Unet architecture before replacing layers.

Experimental Results

Results. Extensive training attempts on our model and consistent troubleshooting has led to a failure to achieve valid results. Although the program is capable of training and only outputs a few warnings, it does not achieve proper learning. Statistics that are output by the validation dataset achieve results that at first look like nonsense but begin to make more sense as you view the predicted images and see they are solely predicting that there is never a lesion (black images).

After 62 epochs, this is the resulting statistical output of one model:
Overall loss (including deep supervision layers): 1.0036
Output loss: 1.0000
Output accuracy: 0.9999
Output precision: 0.0000e+00
Output recall: 0.0000e+00
Output dice coefficient: 9.4029e-20
Output intersection-over-union: 4.7015e-20
Output f1 score: 8.1508e-28

As can be seen, the losses are not improving at all during training. Accuracy shows as near perfect, but this is due to the minimal number of lesion pixels in the images, meaning that guessing the entire image is black is already near perfect accuracy, so it is a poor way of measuring the capability of the model. All other methods of measuring the model capabilities are close to 0, making this output near useless.

Conclusion

After working through our approach, the outcome turned out to be a failure for now. However, though this internship is over I plan to continue my effort, learning from the mistakes of this project and starting again. As I continue the research, I will work to not only fix and receive proper experimental results for the current algorithm, but also try more changes to the inner architecture of the 3D U-Nets, such as implementing residual connections to assist in solving the vanishing/exploding gradient problem and theoretically achieve superior results.