Validation Plan for Quantifying Hippocampus Volume for Alzheimer's Progression

Your Name: Elliot Y.

Intended Use Statement

This algorithm is comprised of the U-Net architecture Convolutional Neural Network model to analyze MRI brain images to produce hippocampus volume measurements. It is intended to assist radiologists in the clinical setting by identifying individual MRI images that belong to one Series, analyzing every slice of the 3D volume, and outputting a report of hippocampal volume. This algorithm will provide consistency in its method from one Series to the next Series, reducing variances that are introduced by interobserver variability among radiologists.

It should be used in conjunction with a Radiologist's review of the MRI images and a review of a patient's medical history to diagnose and track progression of neurodegenerative brain disorders such as Alzheimer's Disease (AD).

Indications for Use

This algorithm may be applied to MRI brain images that are taken from the axial position.

Training Data

The training dataset was curated from the Medical Decathlon competition. This dataset consists of collection of 260 NIFTI files, with one file per MRI image volume and one file per corresponding segmentation mask. The original images are T2 MRI scans of the full brain. The training dataset images are cropped volumes of the original images, showing only the region around the hippocampus.

The segmentation masks had three label classes as follows:

Class Label	Corresponding Structure
0	Background
1	Hippocampal Anterior
2	Hippocampal Posterior

Clinical Validation Data Requirements

For clinical validation, another dataset is needed to show performance in the real world.

Partnering with a clinical facility and unlimited budget, the data set should have the following:

- Demographic information from the Medical Decathlon was not available, so it is necessary to collect images from all ages and genders.

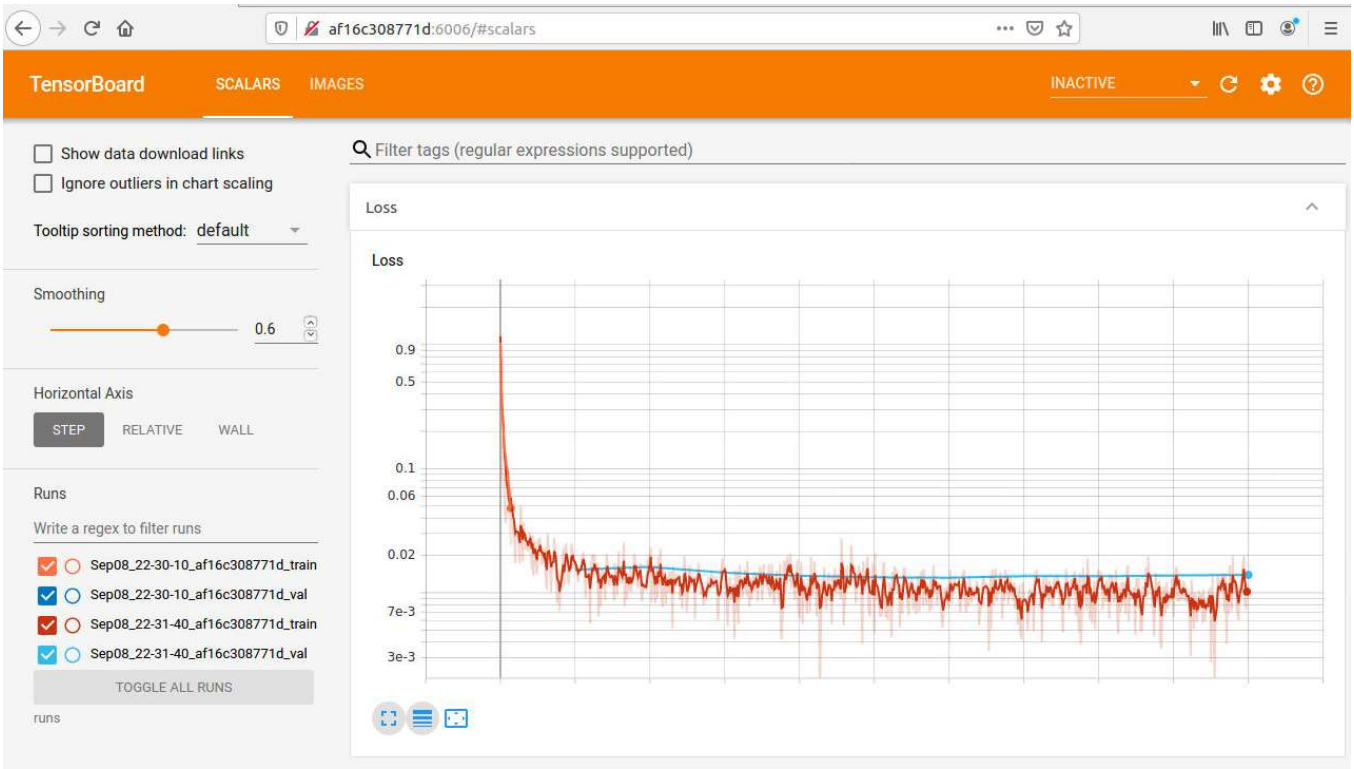
- MRI images do not have labels as the training dataset, so the labels need to be created. This algorithm is intended to assist radiologist create hippocampal volume measurements, so the ground truth will be based on the work of a hired team of radiologists who will each review the same data set of MRI scans and individually label each slice of the MRI scans using a tool such as 3DSlicer. To account for interobserver variability, a minimum of three radiologists will be needed. Their labels will be weighed according to years of experience.

Model Training Performance

This algorithm was trained with the U-Net architecture. The algorithm's output provides prediction of each pixel's probability of belonging to a certain class.

The optimizer was Adam, and it was set to minimize cross entropy loss, with a learning rate of 0.0002.

Loss for 10 epochs were tracked using Tensorboard:



Algorithm Performance

The performance of this model was measured by Dice and Jaccard similarity coefficients which represent similarity between model predictions and corresponding ground truth.

Utilizing the Test split of the parent dataset to understand the algorithm's performance, the Test split had the following results:

	Overall mean dice coefficient	**Overall mean Jaccard coefficient**
Test Split	0.9064	0.8298

For Clinical Validation, the Dice and Jaccard similarity coefficients will be utilized to compare model predictions and radiologist labeled data. Performance should be close to 0.90 Dice and 0.80 Jaccard.

Additionally, with demographic information available, right and left hippocampal volume by age distributions for each gender can be generated. This will allow comparison to work by other established groups such as UK Biobank's HippoFit. This algorithm's distributions should be similar to those by HippoFit.

Device Limitations

This algorithm performs well for T2 MRI brain scans that are cropped to the region around the hippocampus.

It will not perform well with uncropped MRI brain scans.

It may also not perform well with T1 MRI scans, which do not have contrast media.

This algorithm may have different performance with Flair MRI scans, which highlights only abnormal fluids.