Validation plan

Intended use: Intended use is to assist radiologists in determining the hippocampus volume.

Indications for use: Indicated for use for determining hippocampus volume in the T2 MRI brain scans.

Patient population description: The patient population consist of adults, both males and females in the age group 53-71 years of age.

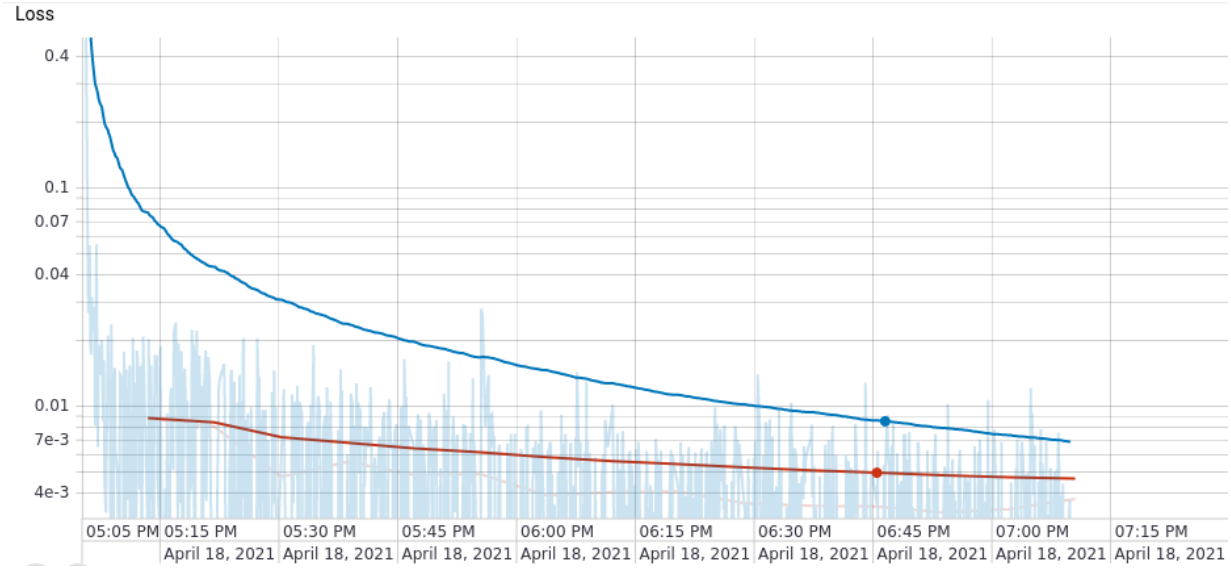
Training data collection: The dataset used is from the "Hippocampus" dataset from the Medical Decathlon competition. This dataset is stored as a collection of NIFTI files, with one file per volume, and one file per corresponding segmentation mask. The original images are T2 MRI scans of the full brain.

Evaluation of the training data set - There are 3 files that seem to be outliers – one of them because of a mismatch in hippocampus volume and other 2 possibly corrupt files. These have been removed for training the algorithm.

Ground truth acquisition methodology: There are 263 files for training and 131 for testing. The ground truths have been obtained from radiologists who have collected, annotated the hippocampus volumes and converted them into NIFTI file format.

Algorithm training performance: The UNet algorithm has been trained on the dataset, after removing the outliers. Image augmentation and drop out layers have been used to avoid overfitting.

The graph for the training loss and the validation loss is as below:



Blue Curve – Training Loss and Red Curve – Validation Loss

The Dice & Jaccard coefficients and Sensitivity & Specificity have been calculated on the test images and results are below:

1. Mean Dice Coefficient: 0.8448
2. Mean Jaccard Coefficient: 0.7323
3. Mean Sensitivity: 0.8762
4. Mean Specificity: 0.8363

Real World performance standard: The real-world segmentation done by the radiologists who are expert at their tasks is the standard the algorithm's performance will be compared against.

Algorithm limitations: The algorithm will perform well on T2 MRI scans of adults in the age group 53-71 years, both male and female.

However, the algorithm may not perform well on the following data:

1. MRI scans of patients outside the age range of 53-71 yrs.
2. MRI scans of brains with other brain ailments like tumours
3. T1 MRI scans of the brain