Validation Plan

What is the intended use of the product?

This software is capable of quantifying the hippocampal volume of patients in order to keep track of the progression of their Alzheimer disease.

This software is not intended to replace doctors. It is rather intended to help doctors keep track of the progressions of their patients with Alzheimer disease. So that, this software will alleviate the work overload of doctors. Moreover, patients and their relatives can also see the reports of this software and be more aware of the progression of the Alzheimer disease.

How was the training data collected?

The training data is made of MRI brain scans of multiple patients of both genders (male and female) and with different ages.

Once the MRI brain scans are taken, doctors cropped the volumes in which the hippocampuses are located in order to focus of those specific regions. These volumes are called "HippoCrop".

How did you label your training data?

The silver standard was used: A group of doctors sat down and annotated the voxels they think compose the hippocampus in every "HippoCrop" volume.

The gold standard is not used here and consists in taking the MRI brain scans with a special chemical that acts as a constrast and marks the hippocampus in a special way.

How was the training performance of the algorithm measured and how is the real-world performance going to be estimated?

The training performance was measured by using the Dice coefficient and the Jaccard metric. Here are the results:

mean_dice: 0.8834084001422208
mean_jaccard: 0.792622572904947

Real-world performance is going to be estimated in a similar way that training performance was estimated. If doctors want to estimate the real-world performance of this software, they need to annotate the voxels with the hippocampal structures of many "HippoCrop" volumes in a testing dataset. Then the Dice coefficient and the Jaccard metric can be computed. In that way, real-world performance can be measured.

What data will the algorithm perform well in the real world and what data it might not perform well on?

Obviously this software is intended to work with relatively older people. This software is not designed for children. The software works well with both genders: Male and female.

This software only works with "HippoCrop" volumes, in other words, parts of MRI brain scans which only contain the hippocampus. This software will not work well with entire MRI brain scans in which the volume contains all the parts of the brain.