Intended use

The algorithm can assist the radiologist in the segmentation of hippocampus on T1-weighted brain MRI images.

Training data

The training data is 260 cropped 3D T1-weighted brain MRI images from the Medical Decathlon competition. The dataset was acquired from 90 healthy adults and 105 adults with a non-affective psychotic disorder (56 schizophrenia, 32 schizoaffective disorder, and 17 schizophreniform disorder). All subjects were free from significant medical or neurological illness, head injury, and active substance use or dependence. Hippocampus was traced manually.

Simpson, Amber L., et al. "A large annotated medical image dataset for the development and evaluation of segmentation algorithms." *arXiv preprint arXiv:1902.09063* (2019).

Algorithm Limitations

Significant medical or neurological illness, head injury, and active substance may affect the performance of the algorithm. The algorithm may not perform well on MRI images generated by weighting methods other than T1.

Validation plan

Patient Population Description for FDA Validation Dataset:

The patients should be healthy adults or adults with a non-affective psychotic disorder. They should be free from significant medical or neurological illness, head injury, and active substance use or dependence.

Ground Truth Acquisition:

The validation dataset should be the 3D T1-weighted brain MRI scan generated with similar parameters (TI/TR/TE, 860/8.0/3.7 ms; 170 sagittal slices; voxel size, 1.0 mm^3) as the training set.

We can have three radiologists manually label the MRI images and use their consensus areas as the ground truth. To make sure our ground truth is reliable. We can first check whether labels from the three radiologists agree with each other. We can then check whether the hippocampus volume of each person falls into the normal range according to age and sex. If there are some outliers, we can go back to the raw image and double check.

Algorithm Performance Standard:

Dice similarity and Hausdorff distance between the predictions and ground truth can be used to measure the algorithm performance. The state-of-art algorithms achieve dice

similarity of 0.85-0.9. We hope our algorithm can achieve the similar or better performance.

Shi, Yonggang, Kun Cheng, and Zhiwen Liu. "Hippocampal subfields segmentation in brain MR images using generative adversarial networks." *Biomedical engineering online* 18.1 (2019): 1-12.