

Validation Plan

This model is meant to be used on medical images of the brain in a dicom format which contain the hippocampus, and it is intended to be used to assist clinicians in determining the volume of the hippocampus. This will be in tandem with a clinician's validation as an assistive tool to ease the process, and as such is not meant to replace the clinician's determination but to reduce the overall time that such a process takes. The model determines the overall volume of the hippocampus by observing its volume in both the anterior and posterior dimensions, then using these values to calculate the full volume of this area of the brain.

The dataset used for training this model was acquired as a part of the Medical Segmentation Decathlon. The original medical images were distributed through Vanderbilt University Medical Center for the competition. For training, testing, and validation, we used a series of brain images which have been cropped in order to better determine the volume of the hippocampus. Though cropping medical images is fairly common, it does mean that our model is likely to perform better on a cropped image than on an uncropped image. Our training data was labeled using masks of individual slices, which were used to calculate hippocampal volumes.

For our testing of this model, we calculated both dice and jaccard scores to determine the accuracy of its image segmentation procedures. Our dice score over the testing data was near 0.897 and our jaccard over the same was at 0.81. In order to validate this model's accuracy in a real world setting, we would need to compare the predicted hippocampus volume of the patients whose images are passed through our predictor to a significantly large enough population of other patients whose images have been seen by multiple clinicians. A simple way to achieve this could be to use a tool similar to HippoFit's "Automated Brain volume - by - age percentile calculator"¹, which would be able to compare the total volume of the hippocampus to a distribution made up of a large population of hippocampal volumes, noting the likelihood that such a volume would be possible. We could use this designation as a testament to the accuracy of the metric, as the majority of our predictions should be relatively close to volumes in the real-world population.

¹ <http://www.smanohar.com/biobank/calculator.html>