FDA Submission

Your Name: Slaouti-Jégou Yannis

Name of your Device: Hippocampus volume segmentation for brain MRIs

FDA Validation Plan

This algorithm was designed for assisting the radiologist in quantifying hippocampus volume for AD's progression over time in brain MRI scans.

To train this algorithm, we needed a good amount of data. To achieve this, brain MRI scans were collected by a radiology department that then ran a HippoCrop tool which cuts out a rectangular portion of a brain scan from every image series.

Radiologists have then collected and manually annotated a dataset of relevant volumes, and converted them to NIFTI format to obtain an exploitable ground truth. This method of manual labelling is widely considered as a golden standard by the researchers although it is very time-consuming. Indeed, this comparative study: A comparison of manual tracing and FreeSurfer for estimating hippocampal volume over the adult lifespan by Mike F Schmidt, Judd M Storrs Kevin B Freeman, Clifford R Jack Jr, Stephen T Turner, Michael E Griswold & Thomas H Mosley Jr shows that manual labelling of hippocampus volumes is pretty consistent dice-score-wise with a population of 262 males and 402 females aged 38 to 84, both when the same radiologist does the labelling (Intra-rater) and when the radiologist changes (Inter-rater).

Table 3

Comparisons of agreement between methods

		r		ICC		Dice	
		R	***L***	***R***	***L***	***R***	***L***
Trace comparisons							
Intra-rater	(*n* = 22)	0.69	0.90	0.84	0.95	0.88 (0.025)	0.86 (0.025)
Inter-rater	(*n* = 25)	0.68	0.69	0.59	0.59	0.82 (0.047)	0.82 (0.048)
Trace	6.0.0	0.73	0.72	0.34	0.32	0.76 (0.042)	0.75 (0.044)
Trace	sf+	0.78	0.77	0.72	0.70	0.78 (0.042)	0.78 (0.044)
Trace	sf–	0.78	0.77	0.51	0.48	0.75 (0.041)	0.76 (0.044)
Trace	5.3.0	0.71	0.70	0.29	0.27	0.76 (0.046)	0.75 (0.052)
Trace	5.2.0	0.72	0.70	0.31	0.27	0.76 (0.044)	0.75 (0.052)
FreeSurfer 6.0.0 comparisons							
6.0.0	sf+	0.92	0.89	0.52	0.51	0.84 (0.036)	0.83 (0.037)
6.0.0	sf–	0.91	0.89	0.19	0.18	0.77 (0.031)	0.77 (0.032)
6.0.0	5.3.0	0.87	0.89	0.89	0.84	0.87 (0.076)	0.86 (0.082)
6.0.0	5.2.0	0.90	0.89	0.92	0.86	0.88 (0.061)	0.87 (0.067)

Open in a separate window

Comparisons between multiple tracings are divided into intra-rater (the same human rater re-tracing the same hippocampus) and inter-rater (different humans tracing the same hippocampus). Manual tracings and FreeSurfer version 6.0.0 were compared to all other methods. Dice's coefficients are reported with standard deviations in parentheses. Older and alternate measurement comparisons are excluded. sf+ refers to hippocampal subfields' "whole_hippocampus" field; sf– refers to sf+ with tail subtracted. See methods for caveats regarding interpretation.

Figure 1 Dice Scores: state of the art methods

The performance of the algorithm is measured using multiple known metrics: The Dice score, the Jaccard score, sensitivity and specificity. To ensure that the algorithm performs competitively well, we can compare those metrics to the ones of the existing segmentation methods. As we can see in the Figure 1 above, state-of-the-art methods are able to achieve Dice scores ranging from 0.82 to 0.88 for the previously described population. Our own algorithm has to be able to perform as well or even better on such a population, and it can, as we have reached an average Dice score of 0.89 on our own population.

This algorithm is designed to work on a population of individuals aged over 40, as younger individuals are not prone contracting AD and were therefore not included in our population. For this reason, this algorithm might not perform well for individuals under 40 or individuals who have brain damage/brain lesions.