

# FDA Validation Plan

Laima Marcinkeviciute

## Quantifying Hippocampus Volume for Alzheimer's Progression AI Assistant

### General Information

#### Intended Use Statement:

To assist radiologists in Quantifying Hippocampus Volume for Alzheimer's Progression on 3D T1-weighted (T1W) MRI.

#### Indications for Use:

The measurement of hippocampal volume has proven useful to diagnose and track progression in several brain disorders, most notably in Alzheimer's disease (AD). This Artificial Intelligence (AI) assistant intended to use for assisting radiologists in quantifying hippocampus volume on T1W MPRAGE MRI acquired in general population (males and females). The assistant can be used in the presence of many diseases which should be free of significant brain tissue deformations or edema/infarcts/tissue loss presence.

#### Device Limitations:

The system would require a proper integration with the existing PACs system and may require a high performance computing power. The assistant operating on a reduced FOV centered around the initial segmentation. The initial segmentation can be done using another AI assistant or manually. It requires properly labelled T1 weighted MRI DICOM images, preferably acquired using MPRAGE method. The assistant will process only images with the DICOM header 'Series Description' identified as 'HippoCrop'.

Assistant can't be used in population with significant medical or neurological illness, head injury, and derived values can be affected by active substance use or dependence.

#### Clinical Impact of Performance:

The following evaluation metrics were used to assess the assistance performance:

- Dice similarity coefficient (DSC)
- Jaccard similarity coefficient (JSC)
- Sensitivity
- Specificity

The Jaccard similarity coefficient is a statistic used for gauging the similarity and diversity of sample sets. The Jaccard coefficient measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets [https://en.wikipedia.org/wiki/Jaccard\\_index](https://en.wikipedia.org/wiki/Jaccard_index). The mean JSC for the test dataset is 0.825.

The Dice similarity (Dice, 1945) is an overlap metric commonly used to quantify segmentation accuracy. It is different from the Jaccard similarity coefficient which only counts true positives once in both the numerator and denominator. DSC is the quotient of similarity and ranges between 0 and 1 [https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice\\_coefficient](https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice_coefficient). The mean DSC for the test dataset is 0.904.

The high values of JSC and DSC means that the hippocampal volume calculated by our AI assistant very reliable and the derived values can be used for hippocampus volume quantification.

We also get pretty high mean sensitivity (proportion of accurately identified hippocampus, 0.921) and specificity (proportion of accurately identified non-hippocampus brain tissue, 0.997) values for the test dataset. These mean that the 92% of the hippocampus were properly segmented and 99.7% of the non-hippocampal tissue were identified properly, thus the derived hippocampus

volume values are reliable and could be used to assist radiologist in hippocampal volume quantification.

#### **Description of Dataset:**

The dataset was downloaded for the "Hippocampus" Medical Decathlon competition (MDC). The dataset consisted of MRI acquired in 90 healthy adults and 105 adults with a non-affective psychotic disorder (56 schizophrenia, 32 schizoaffective disorder, and 17 schizophreniform disorder) taken from the Psychiatric Genotype/Phenotype Project data repository at Vanderbilt University Medical Center (Nashville, TN, USA). Patients were recruited from the Vanderbilt Psychotic Disorders Program and controls were recruited from the surrounding community. The final clean dataset is consisted of 260 T1W MR images.

#### **Description of Training Dataset:**

The Dataset used in building this Assistant was extracted from the clean MDC dataset. The dataset was split into Training and Test datasets (80/20 - train/test).

#### **Description of Validation Dataset:**

The validation dataset was gotten by further split of the train dataset: 80/20 - train/validation.

#### **Algorithm Description**

The U-net convolutional network architecture was chosen for fast and precise segmentation of hippocampus because it has been shown as a successful scheme for several biomedical applications.

The algorithm was trained on a reduced FOV using patch approach centerer around the hippocampus.

### **FDA Validation Plan**

#### **Patient Population Description for FDA Validation Dataset:**

- General population both genders, males and females
- Restricted to population with significant medical or neurological illness and head injury
- Derived values can be affected by active substance use or dependence

#### **Ground Truth Acquisition Methodology:**

The structural images have to be preferably acquired with a 3D T1-weighted MPRAGE sequence (TI/TR/TE, 860/8.0/3.7 ms; 170 sagittal slices; voxel size, 1.0 mm<sup>3</sup>).

Manual tracing of the head, body, and tail of the hippocampus on images needed to be completed in order to create labels. This segmentation process is challenging since every slice of the 3D volume needs to be analyzed, and the shape of the structure needs to be traced.

The silver standard approach of using several radiologists would be the most optimal for AI assistant validation.

#### **Algorithm Performance Standard:**

The Dice similarity coefficient, the Jaccard coefficient, and the Hausdorff distance are good to measure similarity between to sets, and good for algorithm performance evaluation. We didn't calculate Hausdorff distance, which is distance metrics, and capable to evaluate the similarity in shape between the ground truth and segmentation.

We compared our algorithm performance metrics, such as DSC and JSC, with the state of the art algorithms presented in the reference [4] and DSC values from Medical Decathlon competition.

	<b>HippMapper</b>	<b>Medical Decathlon</b>	<b>Our AI Assistant</b>
Dice similarity coefficient	0.87	0.90	0.904
Jaccard similarity coefficient	0.77	-	0.826

Our “Quantifying Hippocampus Volume for Alzheimer’s Progression AI Assistant” performs about the same as the Rank#1 algorithm in Medical Decathlon competition and slightly outperform HippMapper. We didn't test our algorithm on HippMapper dataset, thus the most reliable comparison is the Medical Decathlon competition results.