Validation Plan

Intended use of the product

- The algorithm's intended use is to assist radiologists in quantifying hippocampus volume from an MRI scan. Quantification of the hippocampus volume of a patient is beneficial for diagnosing and tracking progression of several brain disorders, most notably Alzheimer's disease.

Algorithm description

- The proposed algorithm should be integrated into the clinical life cycle:
  - First, a patient's MRI brain scan is sent to an existing algorithm called HippoCrop to crop the scan into a rectangular volume around the hippocampus area.
  - This cropped volume scan is then sent to the proposed algorithm to predict a segmentation mask that labels all pixels belonging to the hippocampus for each slice of the scan.
  - The volume of the hippocampus is retrieved by multiplying the counted labeled pixels in the segmentation mask for all slices with the related physical dimensions of the voxels.
  - The measurement results are then sent to the clinician alongside the MRI and cropped scans as a supplemental aid for expert validation.
- With this automated aid, the clinician can make an informed decision.

Training data collection

- The training data is collected from the "Hippocampus" dataset from the Medical Decathlon competition.
- This dataset is stored as a collection of NIFTI files, with one file per volume, and one file per corresponding segmentation mask.
- The original images are T2 MRI scans of the full brain, but we are using cropped volumes where only the region around the hippocampus has been cut out to simplify our machine learning problem and allow for reasonable training times.
- The dataset consists of 263 training and 131 testing images.

Labelling the training data

- The data has been labeled and verified by a human expert rater with best effort to mimic the accuracy required for clinical use.

Training the model

- The recursive U-Net was trained with a random split of the data into training, validation and test set. With each epoch, the validation set is used to measure performance and monitor for common pitfalls (e.g., over and underfitting) on the held-out set. The test set was used to compute the average performance metrics: Dice similarity coefficient, Jaccard distance, sensitivity and specificity.

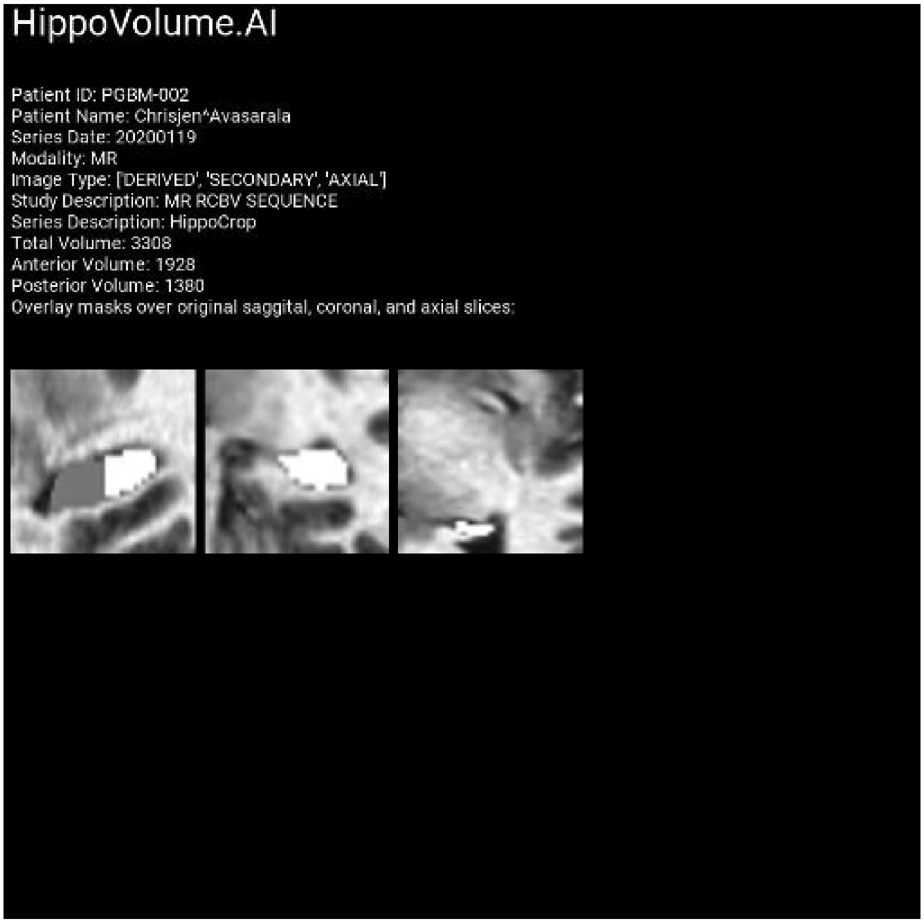
Training performance measurement and real-world performance estimation

- Training performance of the algorithm was measured with Jaccard distance, Dice similarity coefficient, Sensitivity and Specificity scores.
- Real-world performance can further be estimated by validating model performance with radiologists.

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

- The algorithm should perform well with cropped human brain images that give a clear view of the hippocampus from an MRI scan.
- The overall performance metrics achieved were:
  - mean dice: 0.8907 (the highest is 1)
  - mean jaccard: 0.8045 (the intersection of the model output with ground truth is 80.45%)
  - mean sensitivity: 0.8322 (of all areas segmented as belonging to hippocampus, 83.22% really do belong)
  - mean specificity: 0.9984 (of all the areas that are not belonging to hippocampus, 99.84% are correctly segmented)
- This means that our algorithm is very good at correctly segmenting out areas that do not belong to the hippocampus (99.84%), and is fairly good at segmenting areas that belong to the hippocampus (83.22%).
- The algorithm may not work well on other types of scans or images that do not or only partially show the hippocampus.
- The dataset did not include much patient information such as demographics, pre-conditions, etc. – the algorithm may perform better in some groups and worse in others.

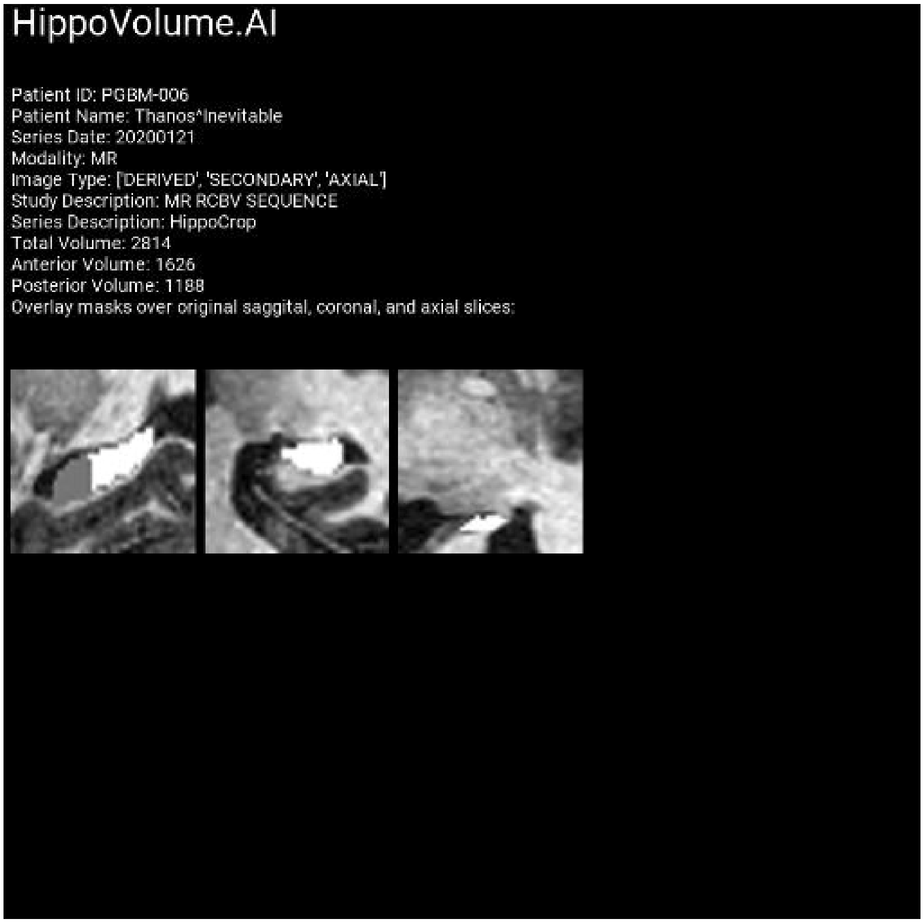
Sample OHIF Reports



HippoVolume.AI

Patient ID: PGBM-002
Patient Name: Chrisjen^Avasarala
Series Date: 20200119
Modality: MR
Image Type: ['DERIVED', 'SECONDARY', 'AXIAL']
Study Description: MR RCBV SEQUENCE
Series Description: HippoCrop
Total Volume: 3308
Anterior Volume: 1928
Posterior Volume: 1380
Overlay masks over original saggital, coronal, and axial slices:

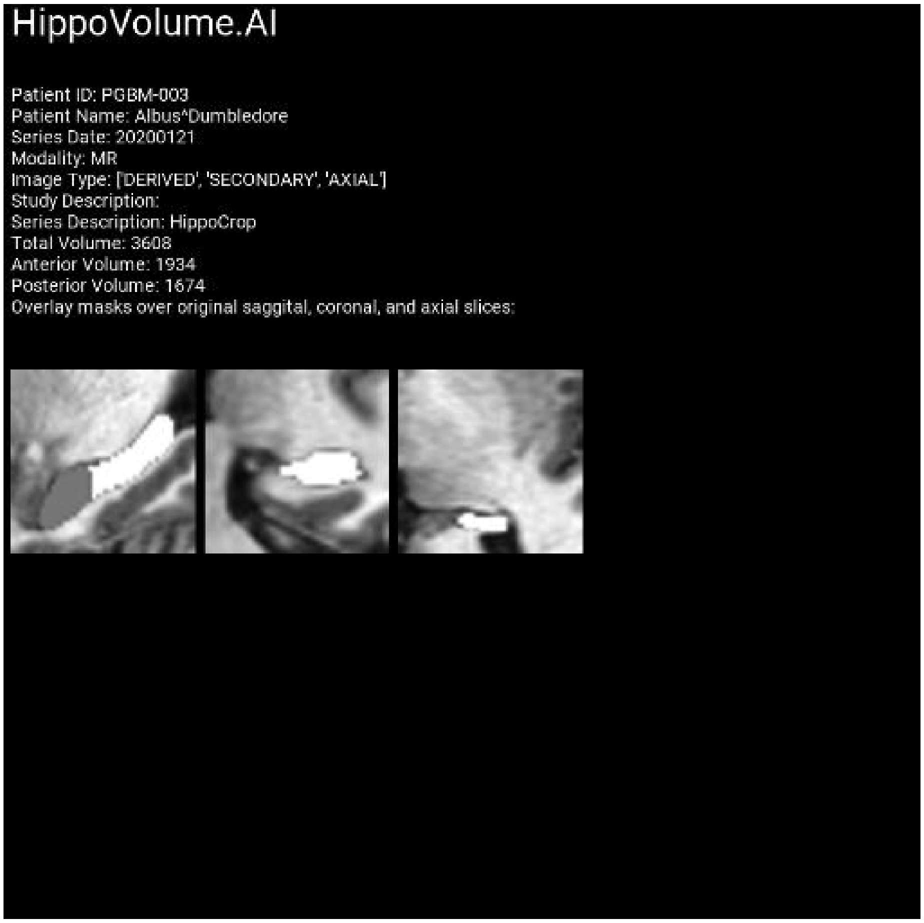
Figure 1: Study 1 OHIF Report



HippoVolume.AI

Patient ID: PGBM-006
Patient Name: Thanos^Inevitable
Series Date: 20200121
Modality: MR
Image Type: ['DERIVED', 'SECONDARY', 'AXIAL']
Study Description: MR RCBV SEQUENCE
Series Description: HippoCrop
Total Volume: 2814
Anterior Volume: 1626
Posterior Volume: 1188
Overlay masks over original saggital, coronal, and axial slices:

Figure 2: Study 2 OHIF Report



HippoVolume.AI

Patient ID: PGBM-003
Patient Name: Albus^Dumbledore
Series Date: 20200121
Modality: MR
Image Type: ['DERIVED', 'SECONDARY', 'AXIAL']
Study Description:
Series Description: HippoCrop
Total Volume: 3608
Anterior Volume: 1934
Posterior Volume: 1674
Overlay masks over original saggital, coronal, and axial slices:

Figure 3: Study 3 OHIF Report