

# Validation Plan

## Intended use of the product

- The algorithm aims to aid radiologists in accurately measuring hippocampal volume from MRI scans. This measurement plays a crucial role in diagnosing and monitoring the advancement of various brain disorders like Alzheimer’s disease.

## Algorithm description

The proposed algorithm is designed to seamlessly integrate into the clinical workflow, enhancing the diagnostic process for various brain disorders. Here’s a detailed outline of its integration:

1. **MRI Scan Processing:** Initially, a patient’s MRI brain scan undergoes preprocessing using an existing algorithm known as HippoCrop. This algorithm efficiently crops the scan to isolate the relevant hippocampus area, producing a rectangular volume.
2. **Segmentation Prediction:** Subsequently, the cropped volume scan is transmitted to the proposed algorithm for segmentation prediction. Employing advanced machine learning techniques, the algorithm generates a segmentation mask that accurately labels all pixels corresponding to the hippocampus across each slice of the scan.
3. **Volume Calculation:** Following segmentation, the volume of the hippocampus is computed by multiplying the total count of labeled pixels in the segmentation mask across all slices by the respective physical dimensions of the voxels. This precise volumetric measurement provides valuable quantitative data for diagnostic purposes.
4. **Clinical Validation:** The measurement results, along with the original MRI and cropped scans, are then transmitted to the clinician as supplementary information for expert validation. This additional data serves as a supportive tool for clinicians, aiding in their comprehensive assessment of the patient’s condition.
5. **Informed Decision-Making:** Equipped with this automated aid, clinicians can make informed decisions regarding diagnosis and treatment strategies. By leveraging accurate volumetric measurements provided by the algorithm, clinicians can enhance their diagnostic accuracy and optimize patient care pathways.

## Training data collection

- **Data Source:** The algorithm is trained on the “Hippocampus” dataset from the Medical Decathlon competition.

- **Data Format:** The dataset comprises NIFTI files, with one file for each volume scan and its corresponding segmentation mask.
- **Preprocessing:** Original T2 MRI scans of the full brain are cropped to isolate the hippocampus region, simplifying the machine learning task and reducing training times.
- **Dataset Size:** The dataset consists of 263 training and 131 testing images.
- **Data Labelling:** The dataset has undergone meticulous labeling and verification by a certified human expert rater, striving to emulate the precision necessary for clinical applications.

## Training the model

Using the recursive U-Net model which is a convolutional neural network architecture renowned for its efficacy in medical image segmentation tasks, was meticulously trained and optimized for the specific task of hippocampal volume quantification. This architecture comprises a series of down-sampling and up-sampling layers, facilitating the extraction of intricate spatial features while preserving contextual information crucial for accurate segmentation.

During training, the dataset was partitioned randomly into three subsets: training, validation, and test sets, adhering to best practices in machine learning experimentation. The training process involved iterative epochs, where the network learned to delineate hippocampal structures from MRI scans through minimizing a defined loss function.

Within each epoch, the validation set played a pivotal role in assessing the algorithm's performance and monitoring for potential pitfalls, such as overfitting or underfitting. This continual evaluation ensured the model's generalizability and robustness across diverse datasets.

Upon completion of training, the test set was utilized to rigorously evaluate the algorithm's performance. Metrics including the Dice similarity coefficient, Jaccard distance, sensitivity, and specificity were computed, providing quantitative measures of segmentation accuracy and model efficacy.

By employing the recursive U-Net architecture and adhering to rigorous training and evaluation protocols, the algorithm demonstrates a sophisticated approach to hippocampal volume quantification, holding promise for enhancing diagnostic capabilities in neuroimaging research and clinical practice.

## 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?**

Based on evaluation metrics results: - Mean Dice coefficient: 0.8966 (with 1 being the highest possible value) - Mean Jaccard index: 0.8139 (indicating an intersection of 81.39% between model output and ground truth) - Mean sensitivity: 0.9385 (suggesting that 93.85% of segmented areas identified as hippocampus truly belong) - Mean specificity: 0.9964 (highlighting that 99.64% of areas not identified as hippocampus are correctly segmented)

The algorithm showcases robust performance when applied to cropped human brain images, offering clear views of the hippocampus in MRI scans. It excels in accurately delineating non-hippocampal areas, boasting a remarkable specificity of 99.64%, while also demonstrating commendable sensitivity in identifying hippocampal regions, with a score of 93.85%. Nonetheless, its efficacy may be constrained when confronted with alternative scan types or images exhibiting partial hippocampal visibility. Moreover, given the limited patient information available in the dataset, including demographics and pre-existing conditions, the algorithm's performance may vary across different patient cohorts.