Team members: Bo Li and Esther Bron

Contact: Bo (b.li@erasmusmc.nl), Department of Radiology & Nuclear Medicine, Erasmus MC, Rotterdam/NL.

The details of the proposed method for Task 3 (lacune detection), and its initial results are described as follows:

- Preprocessing

The proposed preprocessing procedure includes steps of: resampling the input images to be isotropic $1m^3$ and in standard orientation ('RAS') using Nibabel, and applying bias field correction on FLAIR images using N4ITK (Tustison et al., TMI 2010; leave out this step in the submitted docker due to implementation problem).

In addition, we propose to use as additional inputs brain white matter (WM) map and the difference map between T1w and FLAIR image. As a result, **the input** to the network is the stack of T1w image, FLAIR image, WM probabilistic map, and the difference map. The white matter map is obtained by propagating the ICBM452 probabilistic atlas (Mazziotta et al., NeuroImage 1995). For the difference map, we first normalize the intensity of T1w and FLAIR images by their CSF-intensity, and subsequently subtract the normalized FLAIR image from the normalized T1w image. Elastix (version 5.0; Klein et al., TMI 2010) is used for the image registration, for which we applied affine registration during training and non-linear registration during the inference stage (6-9 min/pair). The choice is up to the balance between registration efficiency and accuracy.

- Method and CNN

The objective of the proposed algorithm is to detect all potential lacunes regions with a low false positive (FP) rate, and thereby assisting human observers. To achieve this, we divided the method into two steps to first ensure perfect instance-wise sensitivity, and subsequently reduce as many FPs as possible. Specifically, we implemented the candidate proposal part with a U-Net-like (Ronneberger et al., MICCAI 2015) fully convolution network (FCN) and a false negative (FN)-weighted binary cross entropy (BCE) loss function. **The weight (w)** is defined as follows: for true positives voxels (i.e., the predicted probability of being a lacune is larger than or equal to 0.5), w=1; while for false negatives voxels (i.e., p<0.5), w=(2-p)^2. We observed three types of FPs in the predicted candidates: sulcus, the boundary of brain ventricle, and lesion-like abnormalities. In the present algorithm, we omit lesion-like abnormalities, and propose to reduce the other two types of FPs using prior knowledge, i.e., by applying our additional inputs of WM probability map and a lacune ROI map.

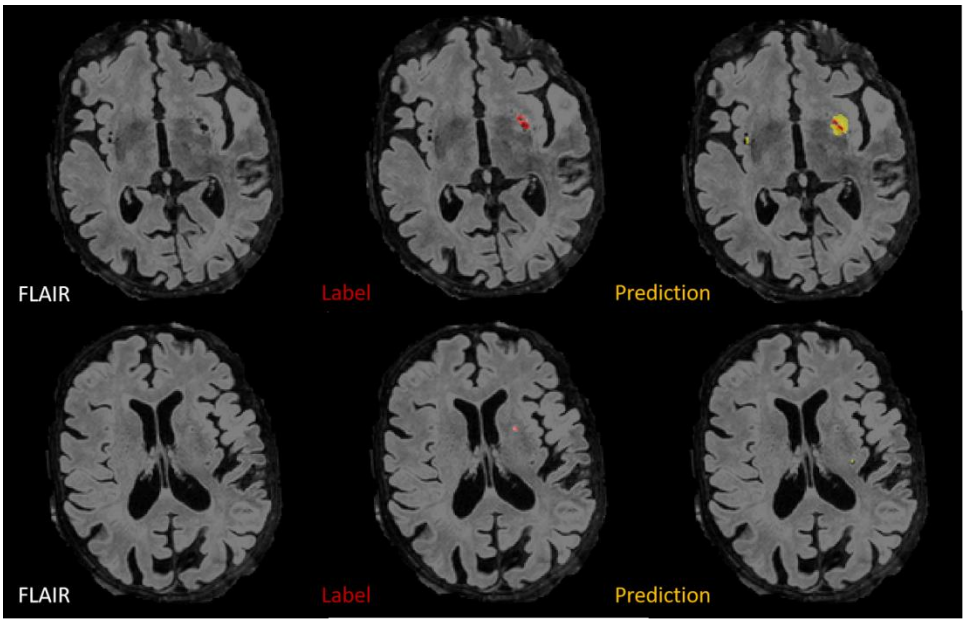
- Implementation

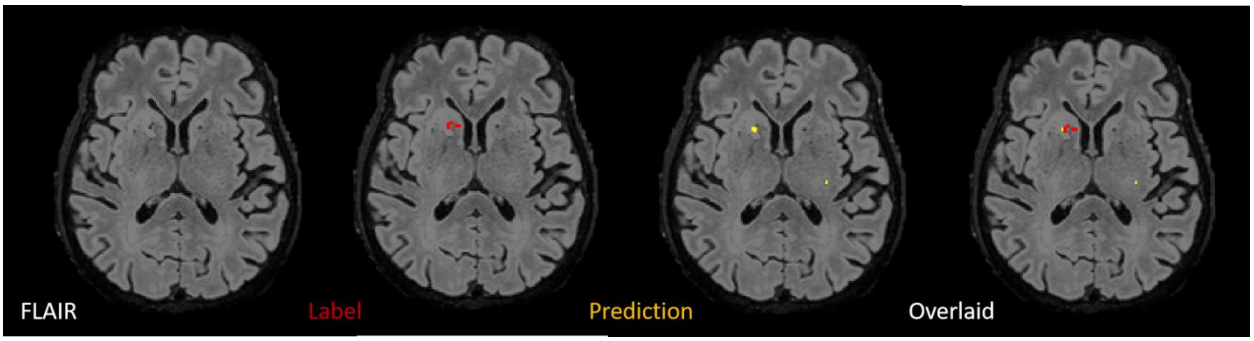
We split the data into 10 folds for model training, and consider as the first-stage results the ensemble of the ten predictions. The union of the two annotations was used as the training label. The learning rate was 5e-4; rotation (range: -15 to 15 degrees) and flipping (along x, y axis) were each applied with probability =0.5.

The prediction of the highest certainty was used for the Valdo challenge considering the evaluation metrics of the Dice coefficient and volume difference. For this, we (1) selected the model weights of the highest validation Dice coefficient, (2) multiplied the ensemble with the WM probability map and a lacune ROI map, and (3) classified the predicted instances into three levels using **multi-Otsu adaptive** thresholds (as available in scikit-image), in which the instances of the highest probability (i.e., the most certain predictions) were considered as the segmentation results. **The lacune ROI map** is generated using manual annotations of the training data, which is subsequently dilated to connect all components in the heat map. In the validation dataset, these strategies lead to a reduction of 0.3 in element-wise sensitivity (i.e., roughly dropped from 0.96 to 0.7). A better trade-off between sensitivity and specificity is expected with a finished second-stage pipeline for false positive reduction.

- Initial results

The Dice coefficient of the n=22 lesion-available Valdo datasets is around 0.26 ± 0.21 (range: 0 - 0.55). Instance sensitivity is around 0.5 ± 0.38 (range: 0 - 1), f1-score is around 0.52 ± 0.35 (range: 0 - 1). The absolute volume difference of all n=40 Valdo datasets is around 208 ± 280 mm^3 (range: 0 - 1080). Note that different from the Valdo definition, the Dice coefficient was not weighted by the rater-agreement, in the computation of the f1-score, distance and disconnected component were not taken into account, e.g., a lesion annotation with disconnected components was considered as multiple lesion labels. Results of sub-102 (upper figure; Dice=0.31, instance sensitivity=0.67, f1-score=0.44) and sub-105 (bottom figure; Dice=0.08, instance sensitivity=0.25, f1-score=0.33) are presented in below as example of Ture positives, FPs, and FNs.




FLAIR
Label
Prediction
Overlaid