Fully automatic segmentation of microbleedings on MR images using mathematical morphology and Random Forest

Elodie Puybareau[2], Hélène Urien[1] and Guillaume Tochon[2]
[1] LISITE, ISEP, FRANCE
[2] LRDE, EPITA, France

Introduction

Microbleedings in brain MRI appear on T2S modality as small dark spots, while they are not clearly visible in the other modalities. Moreover, these lesions are quite small and not really present in the brain. For those reasons, we chose not to use Deep Learning for this problem, but a combination of Mathematical morphology and Random Forest.

Method

Candidate extraction

As said in the introduction, microbleedings are small dark spots on T2S images. For each slice of the T2S volume, we process an area-top-hat. We do first an area closing of size 30 (this removes all the dark spots smaller than 30 pixels in area) and subtract the original slice from it to recover only these small spots. We do exactly the same process on the T2S images multiplied by the T1 image. We keep candidates that are present in both results. We filter this result (we remove the lesion less than 3 pixels) and obtain all our 2D candidates. We reconstruct the volume in 3D and consider each 3D connected component (with an equivalent of the 4-connectivity) as a candidate.

Features selection

We extract 15 features : mean, max and min intensity of the connected component, the distance of the centroïd from the edges of the brain, the intensities of the centroïd in modality T2 and T1, and the normalized centroïd coordinates.

Random Forest training

We use the sklearn random forest module, using its default parameters.
We selected all the candidates that are microbleedings (total of 224 candidates, label=1), and 300 random candidates that are not microbleedings per patient (total of 21000 label=0). Note that the total number of candidate of all the patients is 2439221, so our « training database » represents about 10% of our base. We use the function of sklearn to do a train_test_split with a train size of 0.8.
We randomly set the seed and trained our random forest classifyier. We validate it using the entire database.

Results and analysis

The preliminary results show a F1-score of about 0.3, due to a large amount of false positive. We have very few false negative (21 over the 224). By studying the importance of features, we noticed that the minimum value in the T2S modality and the coordinates of the centroïd are the features with the higher weight. A refinement phase is necessary to decrease the number of false positive, but we did not have time to investigate.