Instructions for AI & optimization: labs 8 & 9

Isep Paris, 18th May 2022

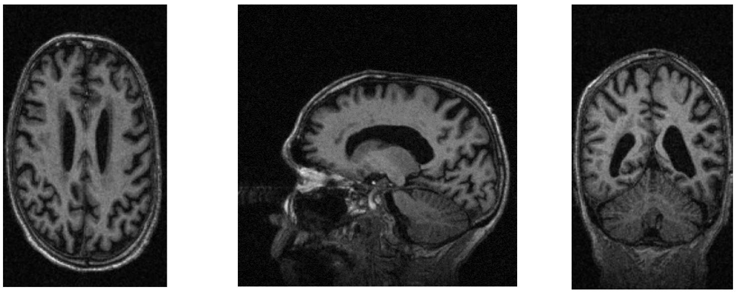
Automatic detection of planes in MR brain images

Deadline: Wednesday, May 25th 12 pm

Work alone, or in teams of 2 or 3 students

I. Context

Magnetic resonance imaging (MRI) is a medical imaging technique providing anatomical information about human body structures. MR images provide a particularly good contrast in the brain, allowing a more accurate diagnosis or follow-up. One other key advantage of MRI is the possibility to visualize the anatomy in three different views: the axial, sagittal and coronal planes. As shown in figure 1., the brain is symmetric, with respect to the inter-hemispheric plane. This property is visible on both axial and coronal planes.



(a) (b) (c)

Figure 1: visualization of T1w-images in the axial (a), sagittal (b) and coronal (c) planes

MR images are 3D volumes, saved in a specific format such as Nifti. However, they are usually analysed slice by slice when a visual expert assessment is required. As shown in Figure 2., the size or nature of the visualized structures vary from one slice to another.

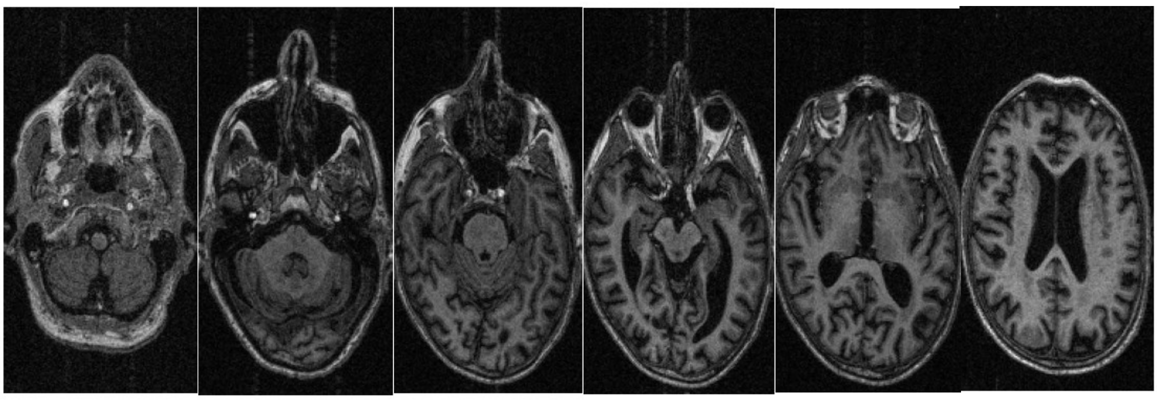


Figure 2: T1w axial MR images of the same subject

In this context, the goal of this lab is to develop a Machine learning approach to automatically detect the plane_axial, coronal or sagittal_ associated with a MR slice.

II. Material

The available dataset is made of T1w-MR images of brain slices, converted into .tiff. They are preprocessed so that the pixel size is 1x1 mm². Thus, if there is a distance of d pixels between two pixels, then the real life distance is d mm. Similarly, if some structure is made of p pixels, its real life area is p mm². The image is also masked to the brain: any pixel out of the brain has the same intensity value as the background. Finally, all the images are cropped to the 2D bounding box of the brain.

III. Training phase

The last lab is divided into two steps: training and test. In fact, it can be seen as a mini-challenge! The training phase takes place during the lab sessions 8 and 9.
The training set is made of 300 MR images of brain slices. They are available in the « training_data » folder. The plane associated to each image_axial, coronal or sagittal_ is given in the « training_labels.csv » file.
Your work during the training phase may be divided into up to four parts:

a) Data exploration (not mandatory but recommanded)
   Get familiar with the dataset. You can for example create a « quality control » folder, where the images are split into 3 subfolders according to their associated planes. It helps you visualizing the difficulty of the task, and choosing relevant features. You can also choose to discard some images that look irrelevant for training… but be careful, similar examples could appear in the future test set!
   You can also answer this useful question during this part: what is the label distribution among the training dataset ? It will be roughly the same among the test set.

b) Feature extractions (**mandatory**)
   Extract **homemade** features from each training image. Thus, **you have to implement your own features!** Of course, you can choose features implemented in the previous lab (symmetry, circle fitting, etc.), and use very basic libraries such as numpy.
   When all the features are collected for all the available images, you are strongly recommanded to save them in a csv file.

c) Training of a machine learning algorithm (**mandatory**)
   Train machine learning algorithm(s) on the selected features to classify the associated images into axial, coronal or sagittal. **The selected algorithm(s) should have been studied during the Optimization & AI or Data Analysis course.**
   More especially, you have to be careful about the following points:
   - feature selection, transformation or normalization (if needed)
   - hyperparameter tuning
   - algorithm selection
   - optimization algorithm
   - cross-validation

d) Test on the future dataset (not mandatory but recommanded)
Prepare the code to apply the trained model on the test dataset. It has to automatically save the results in a « test_labels.csv » file (same structure as « training_labels.csv »).

IV. Test phase

The test phase takes place during the last 30 minutes of the lab session 9.
A folder containing test images («test_data ») will be made available on Moodle. Apply your trained model on these images, and save the predicted labels in a « test_labels.csv ».

V. Expected outputs

At the end of the lab session 9, one single member of your team has to depose two files on Moodle.

a) **The ipynb code file**
A .**ipynb** file containing the code for the **training phase** (**feature extraction** and **ML training**), and **test phase**. This code has to be well structured and commented. More precisely, add in markdown cells any comment useful to justify your choices.
You have to clearly mention the names of all the members of your team in the code file.
As usually, assess that the following requirements are fulfilled:

- Run your Jupyter notebook before sending it back (kernel → Restart & Run all).
- Check that there is no error message.
- Do not write any space or special character in the name of your file
- Ensure that each line of code does is not (roughly) composed of more than 80 characters

The csv prediction file
A csv file containing two columns, the first one for the test image identifiers, and the second one for the predicted labels. The structure is the same as in the training_labels.csv file.

VI. Grading

The grading will be divided into two parts:

- 10 points: justification of the choices made for feature and machine learning frameworks, compliance with the instructions, quality and clarity of the ipynb file
- 10 points: **relative accuracy obtained for your group i on the test set** compared with the maximum score obtained among all the groups

$$10 * Acc_i / \max_{j \in \text{nb groups}} (Acc_j)$$