

## Challenge of Alzheimer's disease classification based on multicenter DTI data

### Project overview:

Diffusion tensor imaging (DTI) has been widely used to show structural integrity and delineate white matter degeneration in AD through diffusion properties. It is confirmed that WM integrity measures are effective in classifying AD using machine learning.

This project aims to evaluate and develop an analysis framework for optimal performance of Alzheimer's disease classification (AD vs normal controls, NC) using diffusion measurements (e.g. FA/MD, etc.) along 18 major white matter tracts. The goal is to achieve both overall high prediction accuracy in cross-validated samples and high consistency in different sites. Secondly, an exploratory goal of this project is to investigate that to what extent can these measurements contribute to the prediction of mild cognitive impairment(MCI), that is, the classification of MCI.

### Project description:

The data we provided is processed by AFQ (Automated Fiber Quantification) pipeline (1, 2) which can quantify eight kinds of diffusion measurements (including FA/MD/RD/AD/CL/curvature/torsion/volume) at multiple locations (100 points for each fiber) along the trajectory of 18 major white matter tracts. Our dataset contains 825 subjects across seven different sites in which 700 subjects are provided for the users to train the models, the left 125 (no label) will be released around Nov. 15. Simultaneously, we provided a clinical diagnosis (classification label) and population features (age and gender) as well as the site label of each subject.

Simply, the task is to classify AD patients and healthy people with these features. You can freely handle these data (such as feature extraction, dimension reduction, smoothing, etc.) and freely divide the training set and validation set for training and testing to achieve higher classification accuracy.

### Project outcomes:

Up to 3 researchers can form a team & work during the competition to design classification models. Each team can submit up to 5 different models for evaluation. The submitted models will later be reviewed using our retained 125 subjects. Teams submitting the selected out-performing models will be contacted to get involved for potential collaborations.

The models will be evaluated based on classification accuracy. A post hoc analyses will evaluate the consistency of estimated predictors in each model.

### DATA:

Here we provide all raw data processed by AFQ in [MCAD\\_AFQ\\_competition.mat](#). The order of subjects in matrixes (label/center/popu/MCAD\_AFQ\_data) is matched.

#### ➤ Details:

- Load data
  - Matlab  
`Load('MCAD_AFQ_competition.mat');`
  - Python  
`from scipy.io import loadmat`

```
data = loadmat('MCAD_AFQ_competition.mat', mat_dtype=True)
```

■ R

```
library(R.matlab)
```

```
nms<-readMat ('MCAD_AFQ_competition.mat')
```

If you load the data successfully, the following matrices will show up:

- **Train\_diagnose:**  
This matrix contains diagnosis information for each subject. Numeral 1 means normal control (NC). Similarly, Numeral 2 represents mild cognitive impairment (MCI), and numeral 3 represents Alzheimer's disease patients (AD).
- **Train\_sites:**  
In this matrix, different numbers represent different sites that subjects belong to.
- **Train\_population:**  
Here we provide age and gender information for each subject. The 1st column is gender in which 0 represents male and 1 represents female. The 2nd column is age.
- **Train\_set:**  
This cell provides all data processed by AFQ. Each row is a fiber tract whose name can be found in a variable called "fgnames" (matched order). Each column is a tract property. We already put all subjects from different sites together.

*Note:* The features provided here are not normalized.

### Expected outputs:

Teams are expected to submit a model as well as a brief description in a compressed zip file. It is better to send us your code. We will review and test your model using the testing data and return you a test report including accuracy, precision, etc. Please make sure to include all the required information from the submission checklist. Teams can submit up to 5 different predictions for further evaluation.

### Evaluation Criteria:

The ACC, AUC, and F-Score for (AD vs. NC, and AD vs. MCI vs. NC) will be computed for the 125 testing subjects to assess the performance of the models.

### Where to start:

Step 1: Clone this repository

First and foremost, clone this repository to have access to the provided data. More information is provided in this issue.

Step 2: Join our mattermost channel

It is strongly recommended to join our channel to chat about any questions regarding the project at [yliu@nlpr.ia.ac.cn](mailto:yliu@nlpr.ia.ac.cn)

Step 3: Register your team

Please make sure to register your team to Yida Qu [quyida2019@ia.ac.cn](mailto:quyida2019@ia.ac.cn) by submitting the necessary information.

Step 4: Start coding :

After registering your team and loading the data successfully through the way we mentioned above,

you can start to build your machine learning model to achieve this classification task. Good luck!

Step 5: Submit your model and necessary instructions

You can submit your model and necessary instructions by filling this submission template and submitting it as a confidential issue.

#### **Advisory committee**

Ying Han, Xuanwu Hospital of Capital Medical University  
Tianzi Jiang, Institute of Automation, Chinese Academy of Sciences  
Bing Liu, Institute of Automation, Chinese Academy of Sciences  
Jie Lu, Xuanwu Hospital of Capital Medical University  
Xi Zhang, Chinese PLA General Hospital  
Qing Wang, Qilu Hospital of Shandong University  
Yuying Zhou, Tianjin Huanhu Hospital

#### **Organizing Committee**

Yong Liu, Institute of Automation, Chinese Academy of Sciences  
Dawei Wang, Qilu Hospital of Shandong University  
Pan Wang, Tianjin Huanhu Hospital  
Hongxiaong Yao, Chinese PLA General Hospital  
Bo Zhou, Chinese PLA General Hospital

#### **Licenses**

The data can only be used for this competition, and can not be used (it) for any commercial/business purpose. If you want to join us to use it for further research, please contact Yong Liu at [yliu@nlpr.ia.ac.cn](mailto:yliu@nlpr.ia.ac.cn)

#### **Contact:**

For any question about the project, classification models, and submission process, you can contact

Yong Liu, Institute of Automation, Chinese Academy of Sciences, [yliu@nlpr.ia.ac.cn](mailto:yliu@nlpr.ia.ac.cn)

Yida Qu, Institute of Automation, Chinese Academy of Sciences, [quyida2019@ia.ac.cn](mailto:quyida2019@ia.ac.cn)

#### **Reference:**

1. J. D. Yeatman, A. Richie-Halford, J. K. Smith, A. Keshavan, A. Rokem, A browser-based tool for visualization and analysis of diffusion MRI data. *Nat Commun* **9**, 940 (2018).
2. X. Dou *et al.*, Characterizing white matter connectivity in Alzheimer's disease and mild cognitive impairment: An automated fiber quantification analysis with two independent datasets. *Cortex* **129**, 390-405 (2020).