Mini-Project Report (May 24th 2019)

Jiaxin Xiao

In this report, gender classification models were constructed by two approaches. One was volume-based classification and the other was image-based classification. For volume-based approach, best set of volume features was found. What’s more, best model selections of both approaches were explored as well. In the end, two approaches were compared and it was suggested that image-based classification was a better method.

# 

# Introduction

Magnetic resonance imaging (MRI) plays an important role in brain study and related clinical research. MRI-based measurement of brain size, segmented region volume, and other structure details provides evidence for brain disease detection, aging research, as well as evaluation of drug therapy. Thus automatic segmentation which supports for massive structural analysis of brain MR images was proposed to substitute pure manual segmentation. Rapid development of automatic segmentation was largely dependent on machine learning (Gryska, Schneiderman, & Heckemann, 2019). It is known that for different genders they own different volumes of brain structures(Yücel et al., 2001). However, overall brain volume varies across individuals, which makes it probably not predictive if taking the absolute volumes of tissue. Instead, relative volumes being computed as the ratios between each tissue and the whole  brain volume may be proper features for prediction. Gaussian Mixture Model is a commonly used way to implement segmentation of brain tissues. Aside from taking relative volume as feature to predict gender, pixels of spatially normalised greymatter maps by regeistering to common references could aslo be seen as features. However, to be practical, they need dimentional reduction methods to extract principal features. One common way used in reducing dimentions is principal component analysis (PCA).

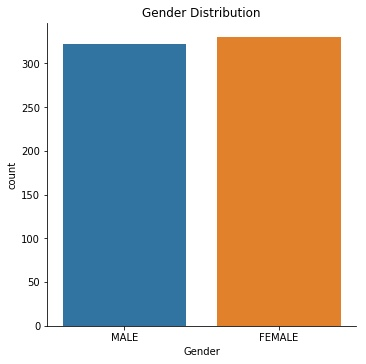
In model fitting process, features are taken as training data and corresponding predicting targets are taken as training targets. Different classifiers were tested and verified by cross-validation and their average accuracy were judeged by f1 score, precision, recall, area under receiver operating characteristics (ROC) curve (AUC) and support. Recall represents the ability of a model to find all the relevant cases within a dataset. Precision represents the ability of a model to identify only relevant data points. F1 score is to find an optimal blend of precision and recall. The larger F1 score is, the better the model performs. AUC-ROC curve is a performance measurement for classification problem at various threshold settings. ROC is a probability curve and AUC represents degree or measure of separability. An excellent mdoel has AUC near to the 1 which means it has good separability. A poor mdoel has AUC near to the 0 which means it has worst measure of separability. And when AUC is approximately 0.5, model has no discrimination capacity to distinguish between positive class and negative class.

# Material and Methods

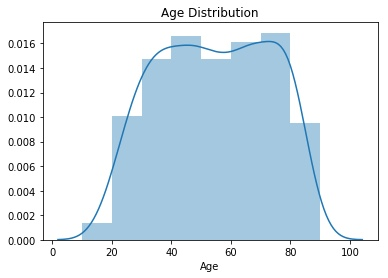
Data from 652 subjects were provided, including MRI images, brain masks, and grey matter maps which have already been extracted from a set of MRI scans and aligned to a common reference space to obtain spatially normalised maps. Meta data containing information about the subjects’ IDs, their age and gender was provided as well.

## Population Stastics Plot

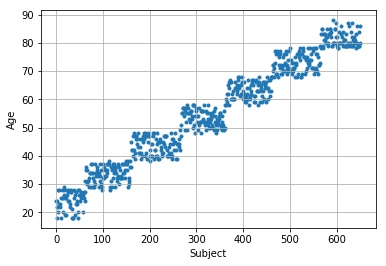
Meta data was firstly loaded to provide an overview and visualization of the statistics of the population of 652 subjects. Gender distribution (as shown in figure 1), age normal distribution (as shown in figure 2) and age scattered distribution (as shown in figure 3) were shown in plots.



**Gender distribution of 652 subjects.**



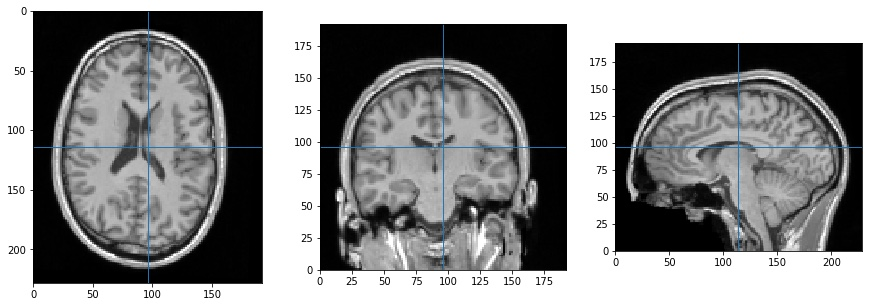
**Age normal distribution of 652 subjects.**



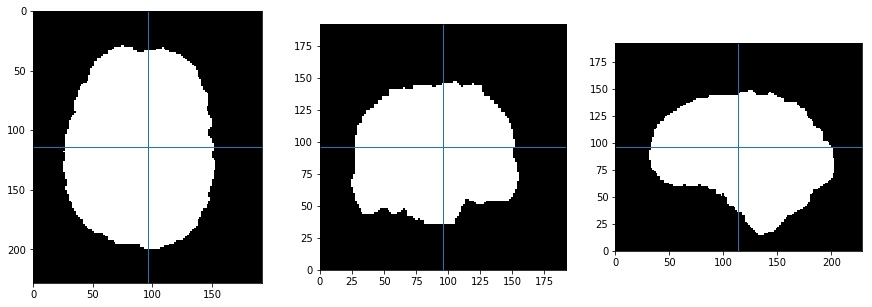
**Age scattered distribution of 652 subjects.**

## Brain Structure Segmentation

MRI images (as shown in figure 4) were preprocessed before segmentation. Non-brain volume of MRI images was masked out using brain masks given (as shown in figure 5).

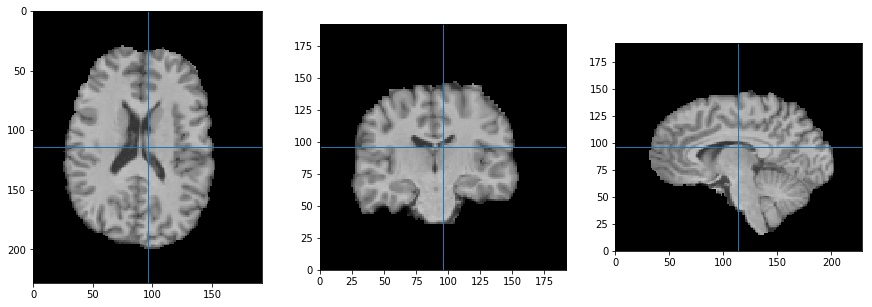


**Example of MR images of one subject used in volume-based classification.**

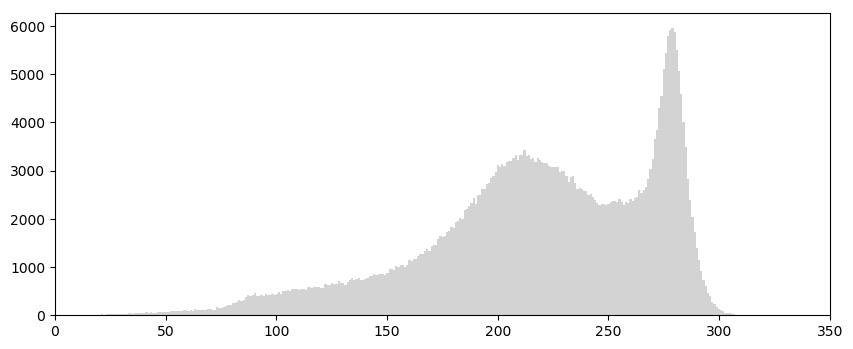


**Example of brain masks of one subject used in volume-based classification.**

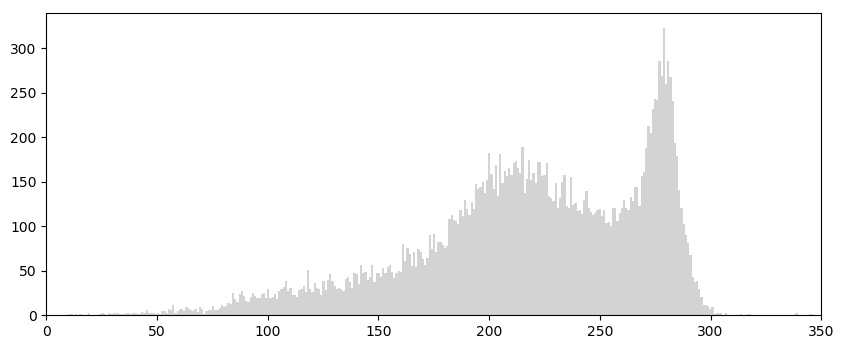
Pixel points of each image were firstly placed in one array. Masked images (as shown in figure 6) were obtained by setting MRI image pixels corresponding to zero intensity pixels in the brain mask as zero. Then, histogram of masked images was plotted (as in figure 7) and undersampled (as shown in figure 8).



**Example of masked images of one subject.**

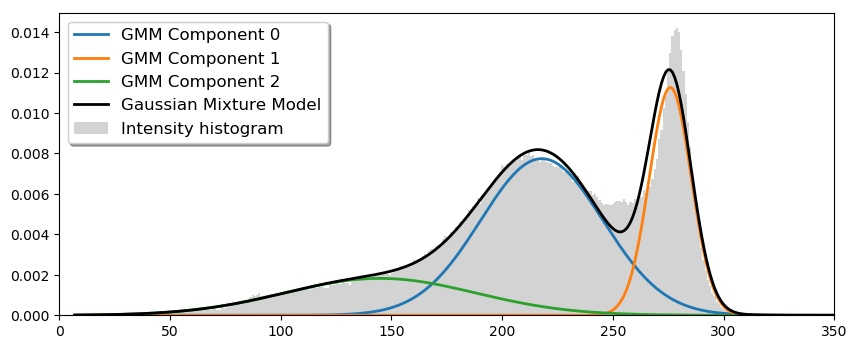


**Example of a histogram of full voxel points of masked iamges of one subject.**



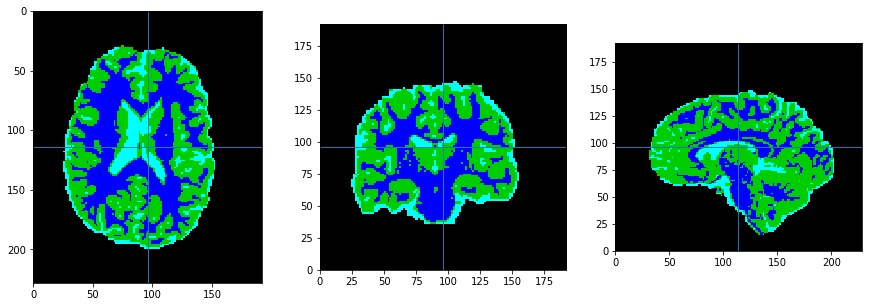
**Example of an undersampled histogram of masked iamges of one subject.**

Here in this report, Gaussian Mixture Model (GMM) was applied to doing the segmentation. Pixel points in the undersampled histogram of masked images were disttributed into three clusters by unsupervised learning method of GMM (as shown in figure 9).



**Three lusters obtained by GMM method.**

Segmented Images (as shown in figure 10) were then reconstructed from clustered image array. Three tissues, including cerebral blood fuid (CSF), greymatter (GM) and whitematter (WM), were labeled by distinct colors in the segmented images.



**Example of segmented images of one subject.**

Relative volumes of three tissues were calculated and stored in RV matrix. The formula to calculate relative volumes is shown as below:

For convenience of following model building all relative data was stored in the RV matrix, such as subject IDs, gender codes, age, and three relative volumes.

## Volume-based Gender Classification and Cross-validation

Dataset of 652 subjects was splited into two equally sized sets (X1, y1) and (X2, y2), which were used in training and testing in an alternating way, thus each set was used as (Xtrain, ytrain) and (Xtest, ytest) exactly once. Here in this report, a 2-fold RepeatedStratifiedKFold cross-validation method was used to split the dataset. Through this way, two classification models were fitted by training dataset and compared to validate each other. Cross-validation was also used in optimizing parameters when fitting supervector classifier (SVC). Method which computes area under ROC curve and Classification Report method which computes precision, recall, f1-score and support were performed to evaluate the average accuracy of different methods. To sum, four classifiers were trained and tested in the experiment, including Linear SVC, SVC(kernal=Linear/RBF/Polynomial), Stochastic Gradient Descent (SGD) and Decision Tree.

## Image-based classification using Greymatter Maps

Greymatter maps were smoothed with gaussian filter and then downsampled to reduce dimensionality before PCA. In this report, Discrete Gaussian method was used to smooth greymatter maps. Then the smoothed maps were downsampled by a factore of two. The preprocessed maps were stored in a big matrix with features of each sample in one row. Dataset of 652 subjects was splited into two equally sized sets according to cross-validation method same as above described. Then PCA was performed to reduce dimensionality and extract principal feature components. Cross-validation was also used in optimizing parameters when fitting supervector classifier (SVC). Method which computes area under ROC curve and Classification Report method which computes precision, recall, f1-score and support were performed to evaluate the average accuracy of different methods. To sum, four classifiers were trained and tested in the experiment, including Linear SVC, SVC(kernal=Linear/RBF/Polynomial), Stochastic Gradient Descent (SGD) and Decision Tree.

# Results

## Best Paramters for SVC

Grid Search Cross Validation method was performed to search for best parameters of SVC models trained by both relative volume features and principal greymatter map features.  Best parameters set of {’C’:1, ‘gamma’: 0.001, ‘kernel’: ‘rbf’} was found when training SVC by CSF relative volume, GM relative volume, and WM relative volume feature, which was also validated by cross-validation. Best parameters set of {’C’:10, ‘gamma’: 0.001, ‘kernel’: ‘rbf’} was found when training SVC by CSF relative volume, GM relative volume, and WM relative volume feature, which was also validated by cross-validation. Details could be checked by source code file attached to the report.

## Best Set of Relative Volume Features

Relative volume features of CSF, GM, WM were fitted to all four classes of classifiers and were evaluated by f1-score and auc, precision and recall. When comparing them with each other, the prior two of f1-score and auc were taken into account. The aucs of GM for all models were around 0.50, and the micro-averages of f1-score for all models  were 0.51. The aucs of WM for all models were around 0.52, and the micro-averages of f1-score for all models  were 0.51. The aucs of CSF for all models were around 0.55, and the micro-averages of f1-score for all models  were 0.51. Thus, the best set of relative volume feature should be CSF relative volume. Details could be checked by source code file attached to the report.

## Model Selection for Relative Volume Features

When comparing the models for relative volume features, best set of relative volume feature-CSF was considered. As the best set parameter for SVC was {’C’:1, ‘gamma’: 0.001, ‘kernel’: ‘rbf’} , parameters of SVC(kernel=RBF) were set as that for evaluation for models trained by relative volume features. For Linear SVC, the auc were around 0.55 but with relative large variance, and the f1-scores were around 0.51. For SGD model, the aucs were around 0.55 but with relative large variance when performing repeats by cross-validation method, and the f1-scores were around 0.51. For DecisionTree mdoel, the aucs were around 0.50, and the f1-scores were around 0.50. For SVC(kernel=Polynomial), the aucs were around 0.55 but also with relative small  variance, and the f1-scores were around 0.50. For SVC(kernel=RBF), the aucs were around 0.55 but with relative small variance, and the f1-scores were around 0.50. Therefore, DecisionTree model was suggested to be the most inappropriate one when taking relative volume features as training data.

## Model Selection for Principal Greymatter Map Features

As the best set parameter for SVC was {’C’:10, ‘gamma’: 0.001, ‘kernel’: ‘rbf’} , parameters of SVC(kernel=RBF) were set as that for evaluation of models trained by principal greymatter map features. For Linear SVC, the auc were around 0.98, and the f1-scores were around 0.92. For SGD model, the aucs were around 0.97, and the f1-scores were around 0.91. For DecisionTree mdoel, the aucs were around 0.69, and the f1-scores were around 0.69. For SVC(kernel=Polynomial), the aucs were around 0.94, and the f1-scores were around 0.49. For SVC(kernel=RBF), the aucs were around 0.98, and the f1-scores were around 0.64. Therefore, DecisionTree model should be the most inappropriate one when taking relative volume features as training data. Therefore, the relative good models for principal greymatter map features was suggested to be Linear SVC and SGD model, which are all linear separation models. Decision Tree model was suggested to be relativly poorly performed.

## Volume-based Classification Approach V.S. Image-based Classification Approach

When comparing these two approaches, best set of features as well as models of each one was selected. For volume-based classification approach, relative CSF volumes were chosen to be features to fit SVC or Linear SVC or SGD model. The aucs were around 0.55 and the f1-scores were around 0.51. For image-based classification approach, Linear SVC model was chosen to be trained. The aucs were around 0.98 and the f1-scores were around 0.92. Thus compared to volume-based classification approach image-based classification approach was suggested to be a relative good approach.

# Conclusions

In this report, two approaches for relative volume feature and principal feature extracted from greymatter maps on the basis of MRI brain scans were conducted to train classification models which can predict genders with obtained new images.

Different classifiers were tested and verified by cross-validation and their average accuracy were principally evaluated by f1 score and aucs. Grid searches were performed to find optimized parameters for SVC. Relative volume features were compared to each other and relative CSF volume was found to be the best feature set in the first approach. Using CSF relative volume feature to explore classifers for the volume-based approach, it was suggested that Decision Tree model was relative not appropriate. Principal greymatter map features were used to explore classifiers for the image-based approach, it was suggested that Linear SVC and SGD model were relative good models while Decision Tree classifer was as well relative inappropriate model. Overall, the comparison between two approaches found that Image-based classification method performed better than volume-based classification method.

# References

Gryska, E. A., Schneiderman, J., & Heckemann, R. A. (2019). Automatic brain lesion segmentation on standard MRIs of the human head: a scoping review protocol.. *BMJ Open*, *9*, e024824.

Yücel, M., Stuart, G. W., Maruff, P., Velakoulis, D., Crowe, S. F., Savage, G., & Pantelis, C. (2001). Hemispheric and gender-related differences in the gross morphology of the anterior cingulate/paracingulate cortex in normal volunteers: an MRI morphometric study.. *Cereb Cortex*, *11*, 17–25.