CSE 559A Final Project Report

Shuyu Xu
shuyu@wustl.edu

Abstract

Alzheimer's disease(AD) can be characterized by a decline in cognitive ability. When patients are sill in a mild cognitive impairment(MCI) stage, the prediction for the future progression of AD may be helpful in treatment for patients. To identify the MCI patients, the longitudinal MRI sequences will provide useful information. In this report, the method used the similarity between local image patches to encode local displacements due to the atrophy between a pair of longitudinal MRI scans. Two estimators, conventional logistic regression classifier and SVM classifier are used for learning, and both of their accuracy of prediction reach to over 80%.

1 Introduction

Alzheimer's disease (AD) is characterized by a cognitive abilities regression, usually occurs among the aged. The diagnosis for AD comes before a mild cognitive impairment (MCI) stage, which is a transition between normal cognitive ability regression due to age and AD[1]. The signs of neurodegeneration can be observed using magnetic resonance images (MRI)[4]. MRI-based fetures such as tissue probability maps, cortical thickness and hippocampal morphology, or a combination of these features can be used to characterized different stages of AD[12].

There were two stages of MCI that will discriminate between patients. One is that patients remain stable in MCI stage (s-MCI), and another is that it will progress into AD in the future (p-MCI)[7]. Moradi *et al.*[6] showed that discrimination of p-MCI and s-MCI takes up to 3 years for transition. Therefore, distinguishing p-MCI and s-MCI will help increase the possibility of preventive care treatment for the disease before it completely turns into AD.

From the physiological point of view, AD can be characterized with the lesions of cerebral cortex, shrinkage of volume in intrinsic cortex, amygdala and hippocampal. Eskildsen *et al.*[8] presented the prediction for AD using patterns of cortical thinning. Since one of the most obvious characteristic of AD is the regression of cognitive abilities, and hippocampal is considered to function the main part for memory and spatial positioning, detection on its abnormality is common in observing AD. A method using hippocampal grading by Coupe *et al.*[2] used hippocampal abnormality based on local similarities to discriminate the two stages, which indicated a high accuracy.

2 Background & Related Work

Sanroma *et al.*[7] proposed an idea to make use of the high level of detail to describe the atrophy patterns across a pair od MRI scans from the same subject ar different time points. This method used a descriptor to compute the local patch-wise similarities between baseline and follow-up images. Considering only the early prediction for AD, hippocampal region was extracted to be fit into the descriptors, and then using a conventional logisitic regression classifier to test the accuracy. Using the dexcriptors, this method helps to predict the MCI stage for patients up to 3 years earlier before it will converse into AD.

There are two hippocampals in human's brain, as shown in Figure 1, which are position symmetrical. They lie under the cerebral cortex, mainly functioning human's long-term memory, short-term-memory and spatial positioning ability. When AD is present, hippocampal is among the first atrophies for the disease. Thus, observing hippocampal's lesion is important for the prediction of AD.

There are many works related to the predictions for AD. Sinchai *et al.*[9] used morphometric hippocampal surface map for the possible future cognitive scores prediction of AD. Most of the research on prediction for AD are based on the shape and volume of hippocampal. Csernansky *et al.*[3] did precinical detection for AD using the shape and volume of hippocampal in the elderly. Suppa *et al.*[10] predicted the AD patients who are at dementia risk with hippocampal volumetry with FSL-FIRST.

Using the longitudinal data for prediction of AD, there are other learning-based methods for the early stage of the disease. Zhu *et al.*[13] used SVM which is particular for the longitudinal data. But for Sanroma *et al.*[7]'s method, the descriptors are longitudinally-specific, unlike the longitudinally-aware classifiers using the conventional MRI images in Zhu *et al.*[13]'s paper. Therefore, conventional classifiers can be used for Sanroma *et al.*[7]'s method.

Figure 1: Hippocampal ROI (red) overlaid onto template image. From left to right the images correspond to Axial plane, Coronal plane and Sagittal plane.

Figure 2: Sum of weight images matches exactly with the right and left masks. From left to right the images correspond to Axial plane and Sagittal plane.

3 Proposed Approach

The method used *i*-th patient's baseline images B_i and follow-up images F_i, and then by extracting the hippocampal as the region of interest, the subtle atrophy patterns which will discriminate s-MCI and p-MCI will be captured. The proposed method can be divided into four steps: (1) preprocessing the raw MRI images, (2) extract the region of interest (ROI), (3) compute the patch-based similarity maps, (4) buid the atrophy descriptor, (5) implement the conventional regression classifier for learning.

3.1 Imaging Preprocessing

The imaging preprocessing is presented as the following steps:

- Prepare the image in nii format.
- Normalize the values in each image.
- Use the inverse non-rigid transformations between template and baselines, denoted as $T^{-1}_{T\rightarrow B_i}$[7].

The raw MRI images obtained are dcm format, for the convenience of following processing steps, they are converted into nii format. Nomalization will also be performed for the images values. Gaussian normalization is used here for normalization:

$$I = \frac{\bar{I}}{\sigma(I)} \tag{1}$$

where I is denoted for the image, and $\sigma(I)$ is the standard devation of an image.

Since the left and right masks of hippocampals for all the images are the same, non-rigid registration is needed for correcting the shape and position of the raw images with a template. Figure 3 showed the image before and after registration. It can be seen that the image is almost well matches with the template.

3.2 Extract Region of interest

Since the hippocampal is considered as one of the first regions which are atrophied due to AD[11] and hippocampal can be chosen as the ROI for the early prediction for AD[2]. The ROI around the hippocampal is extracted using two masks, indicating the left and right hippocampals in human brain. Figure 1 shows the hippocampal ROI which are overlaid into the template B_i.

3.3 Patch-Based similarity Maps

After obtaining the hippocampal ROI for B_i and F_i, which are a 3-dimensional matrices, the patch-based similarity maps are built and have a dimension of 4, where the lsat dimension is a $3 \times 3 \times 3$ matrix. For each pixel in B_i ROI, its neighboring 26 pixels are needed for caluculation for the weight images. And each corresponding pixel in F_i ROI, each of its neighboring pixel's 26 neighbering pixels are needed for the calculation for weight images. The similarity maps can be calculated using the function[7]:

$$W_i^{(j)}(x) = exp(\frac{-||P_i^B(x) - P_i^F(N_x(j))||_2^2}{h^2}) \tag{2}$$

where, $N_x(j)$ is the *j*-th pixel in the cubic neighborhood of size s^3around the point x, and $j = 1, 2...s^3$. $P_i^B(x)$ and $P_i^F(x')$ are the patches centered at x and x' in the baseline and follow-up images correspondingly. Also, the patch

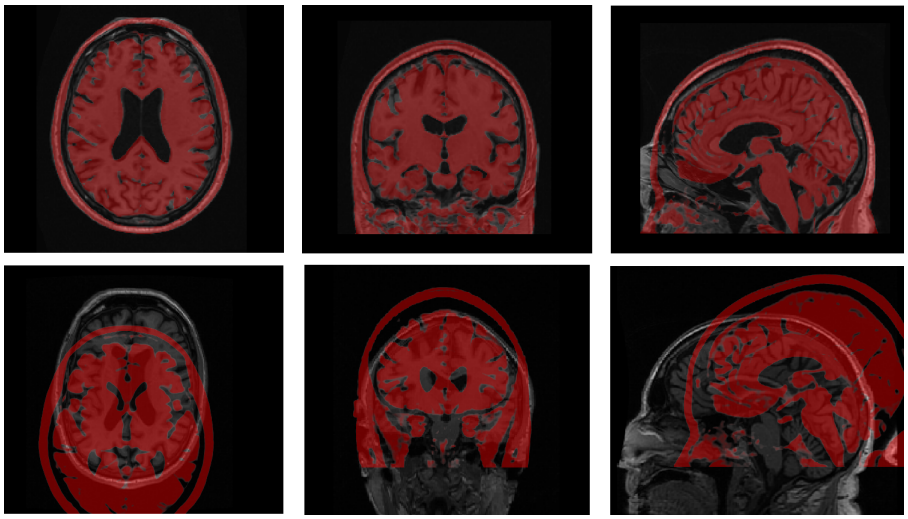

Figure 3: Images after and before non-rigid registration. From left to right are Axial plane, Coronal plane and Saggital plane. Template in red, registered image in grey. Number of resolution = 3; Iterations = 500

size here are of size $3\times3\times3$. For the h^2 in the function, h^2 is equal to the smallest $-||P_i^B(x) - P_i^F(N_x(j))||_2^2$ values for each centered pixel, that is $h^2 = min(-||P_i^B(x) - P_i^F(N_x(j))||_2^2)$. This part is different from the paper, since there are some problems with the function for h^2 in paper. When calculating the weight images, a small value ϵ will be used for stabling the result. Then, by normalizing the weights and the sum of the last dimesion in weight images is either zero or one. Figure 2 shows that the weight images matches exactly with the right and left masks.

Figure 4 is the computed weight images with 27 layers. Each layer matches with the size of mask. It is similar to Figure 2 but the fronter shows 27 layers.

3.4 Build the Descriptors

After obtaining the weight images $W_i^{(j)}$, that is the similarity maps, they will be encapsulated and turned into feature vectors. It can be roughly described into the following step:

- Smooth the warped weight images using Gaussian kernel, which has the size of $\sigma \times \sigma$.
- Tie up the similarities across ROI locations and neighbors, that is, $W_i^{(j)}(x)|x \in \Omega, j = 1, 2...s^3$.

The images obtained from patients may have ROI position variation, so the inverse non-rigid transformation is needed for corrections of the ROI positions.

For the small errors which may occur during the registration, a Gaussian kernel will be used to smooth the weight images, which hsa the width of $\sigma = 1.0$ here.

When building the longitudinal atrophy descriptors for i-th patient, denoted as z_i, in practice, the locations of ROI will be subsampled will a certain step size ρ along each dimension, since if all the samples are included the vectors's length will be pretty large and redundancy will increase.

For the descriptors, all the parameters used here are $\sigma = 1.0, s^3 = 27, \rho = 2$. So the final feature vectors will have a length about $\frac{|\Omega|s^3}{\rho^3}$.

3.5 Classification and Learning

For building the conventioanl logistic regression to predict the possible outcome for each patients, denoted as $y_i = 0, 1$ for s-MCI and p-MCI respectively. The training subjects are denoted as $z_i, i = 1, 2...n$, which have been given the descriptors using the method in Section 2.3[7].

Logistic regression is one of the most intuitive and simplest classification for data. The multiple logistic function for $LogisticRegression$ is

$$p(X) = \frac{e^{\beta_0+\beta_1X_1+...+\beta_pX_p}}{1 + e^{\beta_0+\beta_1X_1+...+\beta_pX_p}} \tag{3}$$

,where $X = (X_1, X_2...X_p)$ are p predictors.

Another estimator Support vector machines (SVM) is used. SVM is effective in high dimensinal spaces, even if the number of sample is small.

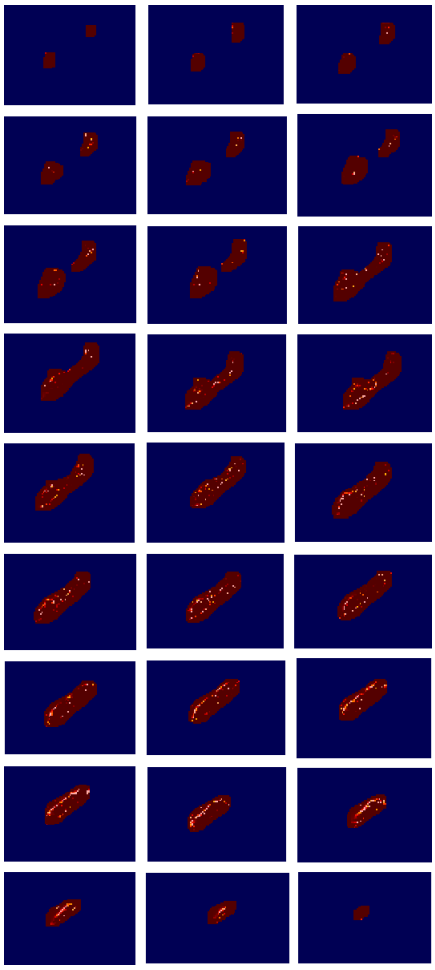

Figure 4: Similarity Maps with 27 layers (Sagittal) matches exactly with the mask. Each layer corresponds to a neighbor of the centered pixel, and the clearset image corresponds to the centered pixel.

4 Experiment

4.1 Preparation for the data

The MRI data used are obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database[1]. ADNI is an open database source for Alzheimer's disease, aiming at helping to reduce the cost and time for clinical trials on AD. There are over 2 thousand subjects recruited, most of them are over 50 years old, including cognitive normal and early MCI or late MCI people, as well as people with AD. For the standardized MRI data sets, there are typically two kinds, incluidng 1.5T and 3T. The data used here are the same as in [7][5]. The T1-weighted MP-RAGE sequence is typically at 1.5T. Image size are about $256 \times 256 \times 166$. Limited by the computer RAM, only 28 p-MCI and 28 s-MCI images are used for testing. The RID labels for p-MCI and s-MCI are randomly chosen from the supplementary data in [5].

Imaging proceesing is presented using pakages *nibabel* and *dicom2nifti*[2], which are used for processing for neuro images, including transfer the nii dataset into matrix-like data, and convert the dcm file into nii file for convience of non-rigid regestration. For the registration, the package *pyelastix*[3] is used.

numpy is used to process the pixel data, *cv2* imported from *OpenCV* is used for smoothing the images, *collections* for generating the *nii* file which is usually used for storing the MRI images data. During the processing of the data, *ITKsnap*[4] can be used for an intuitive understanding of the possible mistakes occurs, such as the shapes and positions.

[1] http://www.adni-info.org/

[2] https://github.com/icometrix/dicom2nifti

[3] https://github.com/almarklein/pyelastix

[4] http://www.itksnap.org/pmwiki/pmwiki.php

To perform the classification and learning, $sklearn$ is used.

Specifically, for the registration package $pyelastix$, it calls the $Elastix$[5] to do the registration. The package works by consistently detecting the possible positions of images to template, and has the image registered by moving to the template's corresponding place instead of using deformations. The package can process numbers of images at the same time by putting them into a numpy array.

4.2 Experiments Methods

The procedures are as described in Section 2.

- Convert the images files into nii format.
- Transform the *nii* dataset into matrix-like data using *nibabel*.
- Normalize the images using Gaussian normalization.
- Implement inverse non-rigid transformation to all images.
- Extract the ROI images using left and right masks on both the baseline and follow-up images.
- Obtain the similarity maps using the function as described in Section 2.2.
- Smooth the similarity maps using a Gaussian kernel.
- Build the descriptors to obtain feature vectors.
- Use the conventional regression classification for learning.

4.3 Results

One problem met during the classification is that the built descriptors have different length. For conveninece, the shortest length among all the s-MCI and p-MCI decriptors is found for cutting the descriptors to have the same length. Dtaset without scaling is usually randomly distributed and does not behave like the Gaussian distribution with zero mean and unit variance, which is a common requiremnet for many laerning estimators. So it is necessary to preprocess the dataset by removing the mean and then scaling to variance equalt to one.

Due to the limitation of number of samples, $cross-validation$ is used during learning. It divides the sample in n parts, denoted as n-fold, usually one for testing and the others for training, and repeats the process until all the parts have been used as testing set. This will help improving the correction of the model when there is a limited number of samples. For the randomness of choosing samples from s-MCI and p-MCI, $ShuffleSplit$ is used to generate a 4-fold cross-validation.

To lower the dimensional space, principle component analysis (PCA) is also performed. Here $PCA = 10, 20, 50, 100, 500$ are tested, the result is as Table 1. Also, selecting the important features in the descriptors, which affect the result most, can increase the accuracy of learning. $random\ forest$ package is used to select the most important features[7], where they have the $feature\ importances$ larger than half of the averge importance of the descriptors, denoted as 0.5σ, where σ is the average importance of the descriptors. Figure 5 compared the result with and without selecting the best features, though some accuracy result with $feature\ importances$ applied are far too high from expected, the total tendency with the best features considered showed better result than without them. One possible problem for the too ideal result may be due to the limited number of samples.

SVM is also used for training the dataset, the result is shown as in Figure 5. Before applying the best features, the accuracy is better than logistic regression without using best features. However, when $feature\ importance$ is applied, SVM seems to be less sensitive than logistic regression, but the total result is still good. Different PCA are also tested for SVM, the result is as in Table 1. The parameters used for SVM are searched using permutations between several possible values of different parameters, and the best parameters are used for the training, which are shown in Figure 5.

Moreover, when neither PCA or $feature\ importances$ are considered for learning in both logistic regression and SVM estimators, the accuracy of logistic regression are higher overall than using PCA, see Figure 5. The highest accuracy for this method is about 0.71. But for SVM, the method without using neither PCA or $feature\ importances$ is lower than when PCA is present, probably this is because the features are much greater than the number of samples, so the result is not good.

5 Conclusion

In conclusion, the patched-based similarity maps compared the difference between baseline images and follow up images, using hippocampals as the region of interest. During the building of descriptors for the similarity maps, inverse non-rigid registration is an important part to match the raw images with the template, since they all use the same hippocampal masks. For the learning result, with the application of $features\ importance$ collected from random forest, Logistic regression shows a better result than SVM. Thus selecting the best features for fitting does help improve the accuracy. But with PCA only, SVM showed a better result. This is not the same as the paper[7] said.

[5] http://simpleelastix.github.io/

Number of components	10	20	50	100	500
Logistic regression	0.500	0.571	0.589	0.536	0.536
SVM	0.589	0.600	0.625	0.480	0.500

Table 1: Best mean accuracy for cross-validation with different PCA. The best accuracy occurs at PCA = 50 for Logistic regression and SVM. The range of regularization is from 0.5 to 50, step size = 0.5;

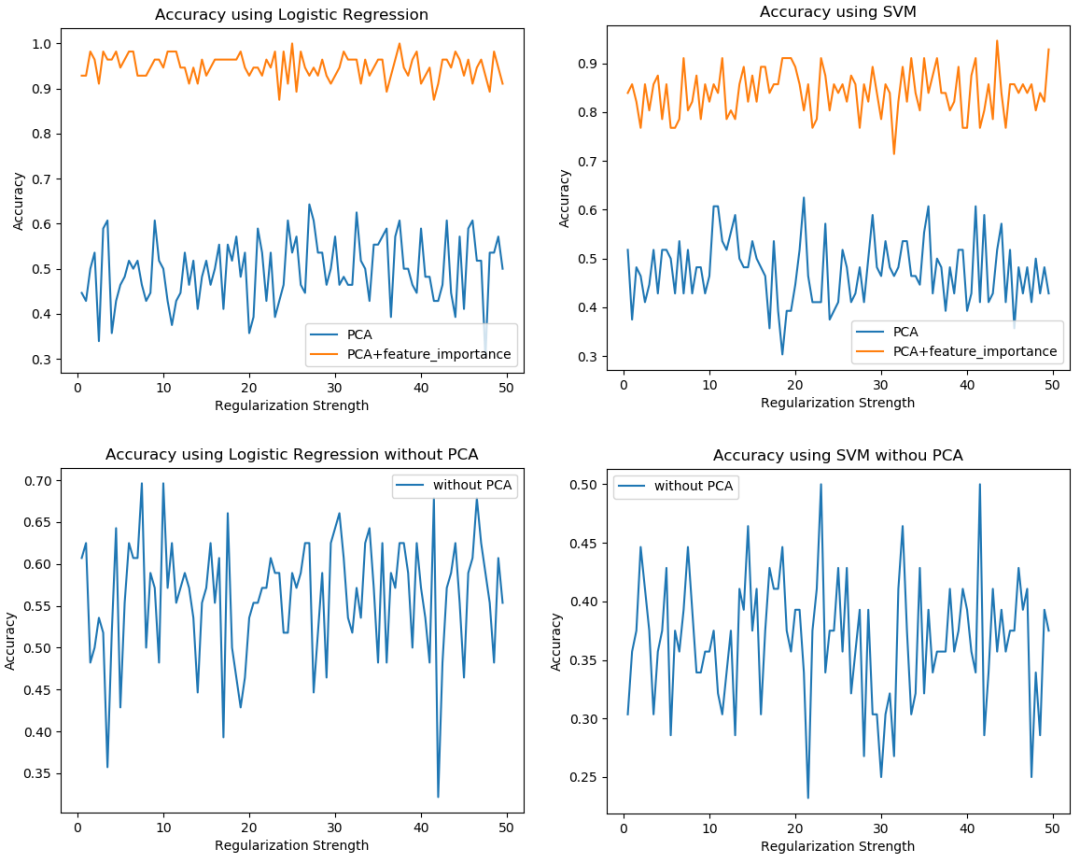

Figure 5: Accuracy changes with regularization strength of Logistic regression and SVM. Top left is the logistic regression using PCA; top right is SVM using PCA; bottom left is logistic regression without using PCA or important features; bottom right is SVM without using PCA or important features. Logistic regression: PCA = 50, choosing the best result from Table 1; penalty using for logistic regrassion is L2 regularization strength ranges from 0.5 to 50, with stepsize = 0.5; cross validation splits to 4 parts, test size = 0.25; Random forest classifier with 1000 trees; SVM: regularization strength ranges from 0.5 to 50, with step size = 0.5; cache size = 0.25; kernel = linear.

One possible reason for this is the limited number of samples used here, and in [7] 100 s=MCI and 164 p-MCI images are used. Another possible reason is the iteration times for registration. However, due to the limited time, resource of data, and computer RAM, running hundreds of images with large iteration number for registration has not been performed. If further research will be proposed, dimensionality reduction techniques to the extracted features may be applied for applying the method to a larger brain area. Because due to the high-dimensionality of the extracted descriptors, this method can only be reasonably applied to small areas such as the hippocampals, and applying the method to larger areas would allow to obtain more discriminative descriptors.

Acknowledgments

Some data including the template image and two hippocampal masks are obtained from Dr. Sanroma, the author of this paper[7].

References

[1] P. Birju B. *Mild cognitive impairment: Hope for stability, plan for progression.* CCJM, 2012.

[2] P. Coupe. *Scoring by nonlocal image patch estimator for early detection of Alzheimers disease.* NeuroImage, 2012.

[3] J. Csernansky. *Preclinical detection of Alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly.* Neuroimage, 2004.

[4] G. Frisoni. *The clinical use of structural MRI in Alzheimer disease.* Nat. Rev. Neurol., 2010.

[5] E. Moradi. *Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects.* Neuroimage, 2014.

[6] E. Moradi. *Machine learning framework for early MRI-based Alzheimers conversion prediction in MCI subjects.* NeuroImage, 2015.

[7] G. Sanroma. *Early Prediction of Alzheimers Disease with Non-local Patch-Based Longitudinal Descriptors.* Springer&MICCAI, 2017.

[8] E. SF. *Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning.* Neuroimage, 2013.

[9] T. Sinchai. *Feature selective temporal prediction of Alzheimer's disease progression using hippocampus surface morphometry.* Brain Behav, 2017.

[10] P. Suppa. *Performance of Hippocampus Volumetry with FSL-FIRST for Prediction of Alzheimer's Disease Dementia in at Risk Subjects with Amnestic Mild Cognitive Impairment.* JAD, 2016.

[11] P. Thompsom. *Dynamics of gray matter loss in Alzheimers disease.* J. Neurosci., 2003.

[12] R. Wolz. *Multi-method analysis of MRI images in early diagnostics of Alzheimers disease.* PLOS ONE, 2011.

[13] Y. Zhu. *Early diagnosis of Alzheimers disease by joint feature selection and classification on temporally structured support vector machine.* MICCAI, 2016.