Feature Selection from CNN based features of classifying MCI patients as converter or non-converter to AD

Utkarsh Bairolia

Prof. Nikhil R. Pal

Table of Contents

Feature Selection from CNN based features of classifying MCI patients as converter or non-converter to AD

Utkarsh Bairolia

Indian Institute of Technology (ISM), Dhanbad, Jharkhand 826004

Guided By:

Prof. Nikhil R. Pal

Indian Statistical Institute, Kolkata, West Bengal 700108

Abstract

Alzheimer Disease is one of the most happening dementia. Mild cognitive impairment is pre stage of Alzheimer. Identifying MCI patients with high risk of AD is crucial factor to reduce disease effect. A CNN approach was taken to create identifiable features from patches of hippocampal region to classify convertible and non-convertible MCI patients. To reduce number of features generated for each image we use Group Feature Selection Multilayer perceptron. Then this features was used in Extreme Learning Machine to classify the subjects.

Keywords: Alzheimer, MCI, CNN, GFSMLP, EL

1 Introduction

Alzheimer's counts 60% of Dementia case. In this patient mainly have significant loss of memory, language disorder and disorientation. Still Cause of dementia is not yet known but early diagnosis of AD can slow down the effect. At Early stage of Alzheimer, patients develop Mild Cognitive Impairment (MCI) which is related to slight memory and intellect deterioration. At clinical level there is diverse transition stage between normal aging and dementia: i) revert to normal cognition, ii) stay stable as MCI (sMCI), iii) convert to AD (cMCI), or iv) develop other kind of dementia. There is no sharp boundary between these transitions which needs the machine learning approach to divide them accurately.

In this internship a CNN based architecture was built to extract high level features of registered and age corrected images and then a feature selection model was used to select relatively more importance features from extracted CNN based features. Then a special Extreme learning Machine was used for classification.

2 Material and Methods

The dataset was based on ADNI1 Dataset. The ADNI is an ongoing, longitudinal study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. The ADNI study began in 2004 and its first 6-year study is called ADNI1. Dataset include 188 AD, 229 NC, and 401 MCI subjects. These MCI subjects were grouped as: (1) MCI converters who were diagnosed as MCI at first visit, but converted to AD during the longitudinal visits within 3 years (n = 169); (2) MCI non-converters who did not convert to AD within 3 years (n = 139). The subjects who were diagnosed as MCI at least twice, but reverse to NC at last, are also considered as MCI non-converters; (3) Unknown MCI subjects who missed some diagnosis which made the last state of these subjects was unknown (n = 93). The demographic information of the dataset are presented below.

	AD	NC	MCIc	MCInc	MCIun
Subjects' number	188	229	169	139	93
Age range	55–91	60–90	55–88	55–88	55–89
Males/Females	99/89	119/110	102/67	96/43	60/33

MCIc means MCI converters. MCInc means MCI non-converters, MCIun means MCI unknown.

Fig 1 Demographic details of ADNI1 Dataset

2.1 Image Preprocessing

All the images were first skull stripped and then registered to MNI 152 template using Freeform Deformation Symmetric Diffeomorphic registration processes. The resulted images were normalize to MNI 152 template histogram using Nyul normalization technique. Finally all the images were in same template space and intensity range.

2.2 Age Correction

To reduce the effect of age related atrophy, age correction is necessary to remove age related effects. The effect was estimated by fitting pixel regression model to the subject's ages. If there are N healthy subjects and M voxels in each pre-processed MRI image then we say that Y_m is the vector intensity at m_{th} voxel and alpha as the vector of ages of N healthy subjects. Age related effect is calculated by regression model: $Y_m = W_m$*alpha + b_m at m_{th} voxel. For n_{th} subject, the new intensity of m_{th} voxel can be calculated as $Y'_{mn} = W_m(C - a_n)+Y_{mn}$, where y_{mn} is original intensity, a_n is age of nth subject. In this study, C is 75, which is the mean age of all subjects.

2.3 CNN-Based Features

CNN network was trained with AD and NC patients. From each MRI image 32*32 2.5D patch were generated from transverse, sagittal and coronal plane centered at a same point. Then these 3 patches were stacked together to create 2D RGB patch. From each MRI image 151 patches were generated on the basis of following points:

1. The patches must be originated in either left or right hippocampus region which have high correlation with AD.

2. There must be at least two voxels distance between each location.

3. All locations were random chosen.

The trained CNN is used to distinguish the converter and non-converter MCI as the converters are similar to AD patients and non-converters to NC.

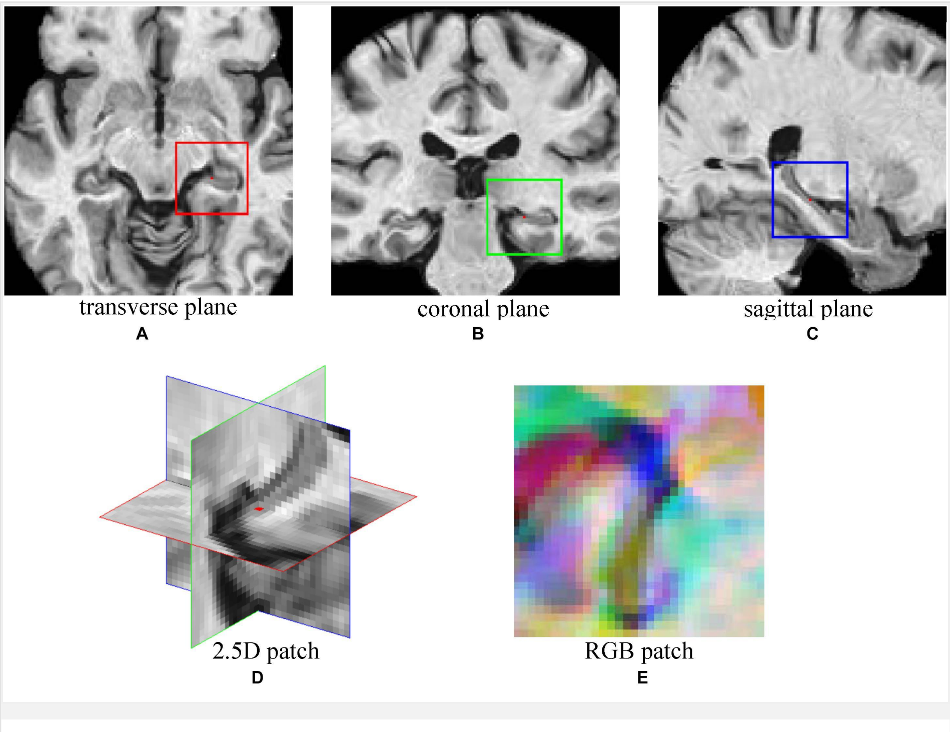

Fig 2 2.5D patch to 2D RGB patch

The architecture of the CNN is summarized in Figure below. The network has an input of 32 X 32 RGB patch. There are three convolutional layers and three pooling layers. The kernel size of convolutional layer is 5 X 5 with 2 pixels padding, and the kernel size and stride of pooling layers is 3 X 3 and 2. The input patch has a size of 32 X 32 and 3 RBG channels. The first convolutional layer generates 32 feature maps with a size of 32 X 32. After max pooling, these 32 feature maps were down-sampled into 16 X 16. The next two convolutional layers and average pooling layers finally generate 64 features maps with a size of 4 X 4. These features are concatenated as a feature vector, and then fed to full connection layer and softmax layer for classification. There are also rectified linear units layers and local response normalization layers in CNN, but are not shown for simplicity.

The CNN was trained with 417*151 images which are randomly split into 417 mini batches.Mini-batch stochastic gradient descent was used to update the coefficients of CNN.The momentum m, learning rate Etta and weight decay lambda are set as 0.9, 0.001, and 0.0001, respectively.The CNN was trained with 30 epochs.The 1024 features output by the last pooling layer were taken as CNN-based features. Thus, CNN generates 154624 (1024 X 151) features for each image.

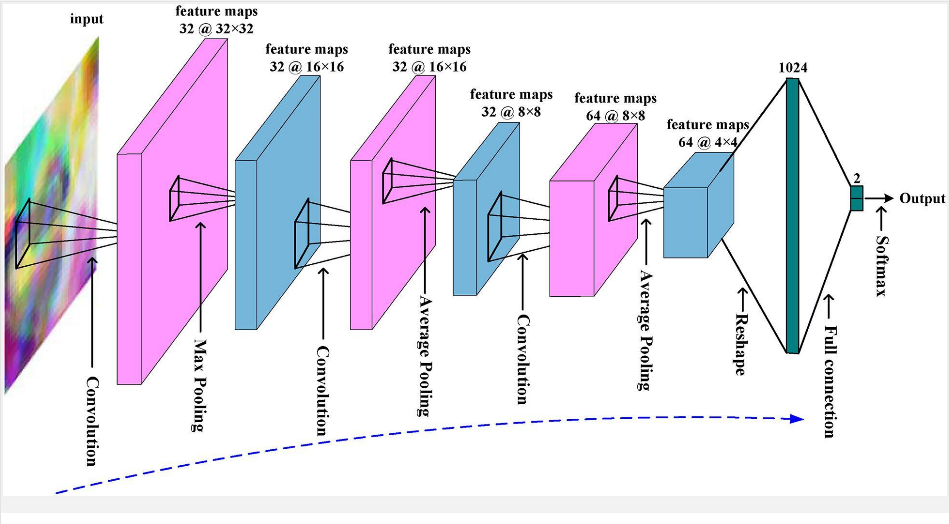

Fig 3 Architecture of CNN

2.4 Feature Selection

A Group Feature Selection MLP network was used to differentiate between features based on patches generated above. Important patches were selected from them. Each patch was a group of 1024 feature in this network therefore 151 groups were formed to find the better one from them. In this network e^{-beta^2} attenuation function was used. Only one hidden layer was created with 20 nodes in it. The network was iterated for 250 loops in which first 100 loops network weights were updated on Adam optimizer and for next 150 beta and weights both were updated on Adam. The learning rate for weights were 0.02 and for beta was 0.4. Initial value of beta was set to 3. Network was applied on the basis of 10 fold classification. Around 14-16 groups were selected as after the final iteration, value of attenuation function was greater than 0.1 for all of these groups.

2.5 Extreme Learning Machine

A special Extreme learning machine was used which avoided random generation of input weight matrix to classify converters/non-converters on chosen groups from GFSMLP network. Extreme Learning Machine is based on Weiming Lin paper.(http://dx.doi.org/10.3389/fnins.2018.00777)

3 Result

When all the 154626 features were used in ELM in 10 fold classification 63.34 % accuracy was obtained and similarly when only 16*1024 features which were selected during feature selection step, 63.32 % accuracy was obtained. So we can say that feature selected from the GFSMLP network showed better result on the basis of dimension reduction as number of features were significantly less from original number.

4 Referrences

1. Convolutional Neural Networks-Based MRI Image Analysis for the Alzheimer's disease Prediction from Mild Cognitive Impairment. By Weiming Lin, Tong Tong, Qinquan Gao, Di Guo, Xiaofeng Du, Yonggui Yang, Gang Guo, Min Xiao, Min Du1,Xiaobo Qu and The Alzheimer's Disease Neuroimaging Initiative†

2. Selecting Useful Groups of Features in a Connectionist Framework. By Debrup Chakraborty and Nikhil R. Pal, Fellow, IEEE.

3. Wyman, B. T., Harvey, D. J., Crawford, K., Bernstein, M. A., Carmichael, O.,Cole, P. E., et al. (2013). Standardization of analysis sets for reporting results from ADNI MRI data. Alzheimers Dement. 9, 332–337.

4. Van de Pol, L. A., Hensel, A., van der Flier, W. M., Visser, P. J., Pijnenburg, Y. A., Barkhof, F., et al. (2006). Hippocampal atrophy on MRI in frontotemporal lobar degeneration and Alzheimer's disease. J. Neurol. Neurosurg. Psychiatry 77,439–442.

5. Huang, G. B., Zhou, H., Ding, X., and Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. 42,513–529