Predicting Conversion of Mild Cognitive Impairments to Alzheimer's Disease and Exploring Impact of Neuroimaging

Yaroslav Shmulev[1,2(✉)], Mikhail Belyaev[1,2],
and the Alzheimer's Disease Neuroimaging Initiative

[1] Kharkevich Institute for Information Transmission Problems, Moscow, Russia
yaroslav.shmulev@skolkovotech.ru
[2] Skolkovo Institute of Science and Technology, Moscow, Russia

Abstract. Nowadays, a lot of scientific efforts are concentrated on the diagnosis of Alzheimers Disease (AD) applying deep learning methods to neuroimaging data. Even for 2017, there were published more than hundred papers dedicated to AD diagnosis, whereas only a few works considered a problem of mild cognitive impairments (MCI) conversion to AD. However, the conversion prediction is an important problem since approximately 15% of patients with MCI converges to AD every year. In the current work, we are focusing on the conversion prediction using brain Magnetic Resonance Imaging and clinical data, such as demographics, cognitive assessments, genetic, and biochemical markers. First of all, we applied state-of-the-art deep learning algorithms on the neuroimaging data and compared these results with two machine learning algorithms that we fit on the clinical data. As a result, the models trained on the clinical data outperform the deep learning algorithms applied to the MR images. To explore the impact of neuroimaging further, we trained a deep feed-forward embedding using similarity learning with Histogram loss on all available MRIs and obtained 64-dimensional vector representation of neuroimaging data. The use of learned representation from the deep embedding allowed to increase the quality of prediction based on the neuroimaging. Finally, the current results on this dataset show that the neuroimaging does have an effect on conversion prediction, however cannot noticeably increase the quality of the prediction. The best results of predicting MCI-to-AD conversion are provided by XGBoost algorithm trained on the clinical and embedding data. The resulting accuracy is ACC $= 0.76 \pm 0.01$ and the area under the ROC curve – ROC AUC $= 0.86 \pm 0.01$.

1   Introduction

Alzheimer's Disease is irreversible progressive brain disorder mostly occurring in the middle or late life. At the same time, there is a transitional phase between the normal aging and dementia symptoms called mild cognitive impairment (MCI). People with MCI are at increased risk of AD development – approximately 15% of them converge to dementia every year. That's why, the early diagnosis of Alzheimer's Disease would allow patients to take preventive measures to temporarily slow the disease progression [10].

Neuroimaging is a variety of methods and technologies that reveal the structure and functions of brain. It includes Computer Tomography (CT), structural and functional Magnetic Resonance Imaging (sMRI and fMRI respectively) and etc. With a growth of deep learning applications in data analysis, neuroimaging is extensively used in many medical tasks such as image segmentation [1], diagnosis classification [11] and prediction of disease progression [5].

In the recent years, there were published a vast number of papers dedicated to classification of healthy controls from AD using deep learning approach applied to neuroimaging. However, only a few works considered predicting conversion of MCI to AD [5,6,8], which is a more complicated and clinically relevant problem. To classify stable and converged MCI the authors of [6] used different clinical biomarkers and complex feature maps extracted from neuroimaging. This method inherit the main drawbacks from manual feature extraction procedure. Cheng et al. in [5] consider the joint multi-domain learning for early diagnosis of AD to boost the learning performance. In this work, we are focusing on the conversion prediction using clinical and neuroimaging data. In addition, we want to explore the individual impact of different data types on the prediction performance for different prediction intervals. Finally, we obtain low-dimensional representation of high-dimensional MR brain images from a deep feed-forward embedding that is trained on the whole ADNI cohort.

2   Data

In this work, we use data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [2]. We choose patients that are diagnosed as normal controls (NC), mild cognitive impairments (MCI), and Alzheimer's Disease (AD). For each patient, we take visits for which both MR images and clinical data are available. The total number of available data samples is 8809.

Clinical Data. ADNI provides clinical information about each subject including recruitment, demographics, physical examinations, and cognitive assessment data. We add genetic and biospecimen data (cerebrospinal fluid concentration, blood, and urine) to the clinical dataset. The full list of attributes is available on the official ADNI website [2].

Neuroimaging Data. For the analysis, we take structural T1-weighted Magnetic Resonance Imaging (MRI), since they are available for all patients and for the most their visits. We fetch preprocessed images from ADNI database with the following preprocessing pipeline descriptions: **"MPR; GradWarp; B1 Correction; N3; Scaled"** and **"MT1; GradWarp; N3m"**. These images have different shapes and orientations and contain skull and other organs that might spoil a predictive performance. Thus, we apply the following preprocessing pipeline for the collected neuroimaging dataset. For the Brain extraction [3] and N4 bias correction [13] steps, we run ANTs Cortical Thickness Pipeline [14] for all available MR images. Then, we apply an affine transform, so that all brain images have the same orientation - **RAS** (Left - Right, Posterior - Anterior, Superior - Inferior). After the brain extraction step, the MRIs contain a lot of black voxels around the brain. We crop all images to the maximal extracted brain size, which is computed beforehand. Ultimately, all MR images in the dataset have a size of $(150, 208, 173)$. To increase a batch size that can be fitted to the Graphics Processing Unit (GPU), we downsample the dataset with the factor of 2 for each dimension, so that the resulting shape is $(75, 104, 87)$.

Conversion Dataset. To predict the MCI-to-AD conversion, we need to remove patients that are normal controls (NC) or have Alzheimer's Disease from the screening visit.

For the stable MCI, we consider participants that have not converged to AD for the known time-period. We also drop several last visits that are inside the prediction horizon, since the future for that patients is not known and they may converge to AD in the next visits.

For the converged MCI, we select participants that were diagnosed as MCI in earlier sessions and as AD later. We take visits that are within five year prediction interval. The total number of stable and converged patients are 532 and 327 correspondingly. The number of samples for two classes are 1764 and 1016.

3   Method

3.1   Clinical Data

For the classification based on clinical data, we use two machine learning algorithms: Logistic Regression and XGBoost. The first one is a linear method which is widely used in many practical applications because of its good interpretability and relative simplicity. The second method is an efficient implementation of gradient boosting on decision trees, which is a powerful machine learning algorithm that can catch nonlinear patterns in data [4].

3.2   Neuroimaging Data

Convolutional Neural Networks (CNN) have recently made a great breakthrough in the image classification and recognition tasks. Deep CNNs automatically

extract and combine from low- to high-level features from images and estimate target values in the end-to-end fashion. In this work, we use two deep architectures: VGG [12] and ResNet [7], that showed state-of-the-art performance in ImageNet classification challenge in 2014 and 2015 correspondingly. We generalize these architectures to the three-dimensional input size of MR images in the same way as was proposed in [11].

VoxCNN. The VGG-like network consists of ten 3D convolutional blocks, each of which consists of three 3D-convolutional layers with $3 \times 3 \times 3$ filter sizes, batch normalization and ReLU nonlinearity. Then, we use max pooling layer with $2 \times 2 \times 2$ kernel size to reduce the size of data propagated through the network. At the end of the net, there are three fully-connected layers with batch normalization and dropout layers in-between. For the experiments, we used the probability $p = 0.7$ of a neuron to be turned off. After the last fully-connected layer, there is softmax activation function to compute probabilities for each class.

ResNet3D. For the ResNet-like architecture, we use 6 residual blocks, each of which represents a sum of identity mapping and a stack of two convolutions with $3 \times 3x3$ filter size and 64 or 128 filters, batch normalization and ReLU. The convolutional layers from the standard ResNet are replaced with 3D ConvBlocks in the same way as we did for VoxCNN. We reduce the spacial size using three convolutions with strides $2 \times 2 \times 2$ before the residual blocks and one maximum pooling layer with $5 \times 5 \times 5$ kernel size before the fully-connected layer. Dropout with $p = 0.7$ and batch normalization are also used after the first fully-connected layer. At the end of the network, there is a second fully-connected layer with softmax activation to produce output probabilities.

4   Experiments

4.1   Setup

For the experiments with conversion prediction based on the neuroimaging data, we minimize a weighted binary cross-entropy loss function with balanced class weights. We use Nesterov momentum optimizer with initial learning rate 10^{-3} and scheduling learning rate policy: we decrease the learning rate ten-fold after 30 and 50 epochs. The batch sizes for ResNet3D and VoxCNN are 128 and 512 correspondingly. These numbers are chosen so that the full batch can be fitted to the GPU. The total number of epochs is 70.

4.2   Validation

To assess the classification performance more accurately, we run 5-fold group cross-validation with five different folds. As a group label, we use participant's ID to prevent the cases when different scans of one patient are simultaneously in train and test sets.

For neuroimaging data, on each step of cross-validation procedure, we train a separate neural network on a train set, use a validation set for early stopping and changing learning rate, and test the network model on a hold-out subset.

For hyperparameter tuning of Logistic Regression and XGBoost methods, another nested group cross-validation procedure is used.

We report the following metrics: accuracy, the area under the receiver operating characteristic curve (ROC AUC), sensitivity, specificity, and average precision.

4.3   Conversion Prediction

There are several ways how the disease progression problem can be formulated. In this work, we use binary classification setting to predict the fact of conversion within a five year interval: **class 0** - stable MCI, **class 1** - converged MCI. In other words, given a participant's visit, we would like to answer the question, whether an individual will converge to AD within the considered interval or not.

4.4   Embedding Learning

For conversion prediction, we used only 25% of all available MR images. To make use of all available data, we learn a deep feed-forward embedding on the whole neuroimaging dataset and, then, use it as a fixed feature extractor. We exploit the extracted features for conversion prediction task, as shown in Fig. 1.
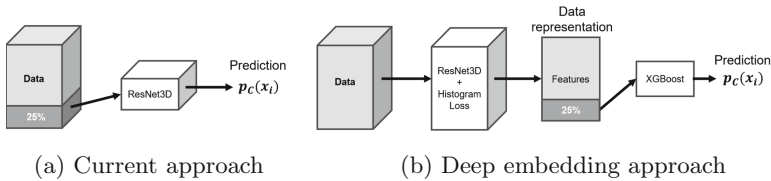

Fig. 1. Two approaches for conversion prediction task. (a) In the current approach, only 25% of all available MR images are used for the conversion prediction. (b) The embedding is trained on the whole MRI dataset and, then, used for feature extraction. We use the extracted features for the conversion prediction task.

Deep embedding is an approach, when complex high-dimensional input data are mapped into a smaller size semantic subspace preserving the most relevant information about the data. Generally, the mapping is learned from a large amount of supervised data. During the training process, semantically related samples are getting closer than semantically unrelated ones in the semantic subspace. To learn deep feed-forward embedding we use ResNet3D architecture up to the last fully connected layer. We add a fully connected layer with 64 output units and L2-normalization layer to the network. We use Histogram loss proposed in [15] as a training criteria, which is parameter-free batch loss function that firstly

estimates two distributions of distances between matching and non-matching pairs and, secondly, computes the overlap between these two distributions. Once the deep embedding is trained, we use it to extract the embedded features: all images from the conversion dataset are propagated through the network and 64-dimensional vector representations are obtained.

5   Results

The results of the conversion prediction are shown in Table 1. For the considered interval, the quality of prediction based on clinical data is significantly higher than one on the neuroimaging. On the neuroimaging, two network architectures provide comparative results for all experiments, although ResNet3D slightly outperforms VoxCNN. For the clinical data, the performance of XGBoost is slightly better than one of Logistic Regression model.

The results also show that the use of learned deep embedding helps increase the quality of prediction based on MR images, although it is still worse than one on the clinical data.

To investigate whether the neuroimaging data can add some new relevant information to the clinical data and, thereby, improve the prediction, we include extracted features from the embedding to the clinical ones. As can be seen from the results, the quality of prediction using clinical and embedding data is slightly higher than for clinical data, although still it is the same within the standard deviation.

Table 1. Conversion prediction results

Data/Method	ACC	ROC AUC	AV PREC	SENS	SPEC
Clinical data/Log Reg	.76 ± .01	.85 ± .01	.73 ± .05	.80 ± .03	.74 ± .02
Clinical data/XGBoost	.76 ± .01	.85 ± .01	.73 ± .03	.76 ± .02	.77 ± .01
Neuroimaging/VoxCNN	.61 ± .02	.70 ± .03	.52 ± .05	.70 ± .04	.56 ± .02
Neuroimaging/ResNet3D	.62 ± .01	.70 ± .02	.53 ± .02	.75 ± .03	.54 ± .01
Embedding/Log Reg	.69 ± .01	.71 ± .01	.54 ± .03	.60 ± .01	.75 ± .03
Embedding/XGBoost	.67 ± .02	.73 ± .01	.57 ± .02	.70 ± .02	.65 ± .05
Clinic. + Embed./Log Reg	.76 ± .02	.86 ± .02	.73 ± .03	.84 ± .02	.72 ± .03
Clinic. + Embed./XGBoost	.76 ± .01	.86 ± .01	.73 ± .02	.88 ± .03	.70 ± .03

Figure 2 shows the resulting representation from the learned deep embedding on a hold-out set. We applied T-SNE algorithm proposed in [9] to map our 64-dimensional feature vectors into 2-dimensional ones. In Fig. 2a, there are three clusters, each of which corresponds to one of the diagnoses: NC, MCI, or AD. From Fig. 2b and d can be seen that the separation between NC and AD is better than separation between MCI and AD.
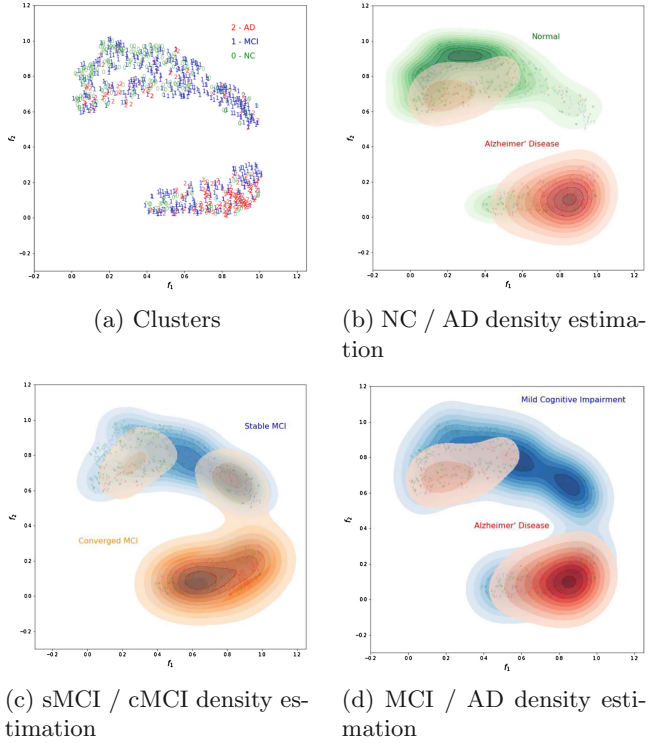
(a) Clusters

(b) NC / AD density estimation

(c) sMCI / cMCI density estimation

(d) MCI / AD density estimation

Fig. 2. Embedding visualization: (a) clusters in the embedded space, (b) Kernel Density Estimation (KDE) of NC and AD distributions, (c) KDE of stable and converged MCI distributions, (d) KDE of MCI and AD distributions.

The next observation from Fig. 2a is that MCI cluster is spread between normal controls (NC) and Alzheimer's Disease (AD). Since we know which MCI patient will converge to AD (cMCI) and which will not (sMCI), we plot the densities of stable and converged MCI. Figure 2c shows that these two groups of MCIs are quite good separated in the embedded space. The main mass of converged MCIs is closer to the AD cluster, whereas the stable MCIs are closer to the normal controls.

6 Conclusion

In this work, a problem of conversion prediction from mild cognitive impairment (MCI) to Alzheimer's Disease (AD) was considered. We collected, preprocessed and analyzed the clinical and neuroimaging data. We applied the state-of-the-art methods for image classification on the neuroimaging data and compared the quality of classification with the several machine learning methods trained on the clinical data. The results of the experiments showed that the clinical data allow

to obtain a better prediction quality than the neuroimaging and these models can be used for conversion prediction task.

We enhanced the performance on the neuroimaging data by training a deep feed-forward embedding. The embedding increased the quality of forecast, however, it is still worse than the clinical data yield. We further investigate the question whether the neuroimaging is able to add some new information for conversion prediction or not. According to the results on the current dataset, neuroimaging does have an effect on the conversion prediction, however it cannot noticeably increase the quality of the prediction when clinical data are used.

References

1. Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D.L., Erickson, B.J.: Deep learning for brain MRI segmentation: state of the art and future directions. J. Digit. Imaging **30**(4), 449–459 (2017)
2. Alzheimer's Disease Neuroimaging Initiative (2003). http://adni.loni.usc.edu/. Accessed 22 May 2018
3. Avants, B., et al.: Evaluation of an open-access, automated brain extraction method on multi-site multi-disorder data. In: 16th Annual Meeting for the Organization of Human Brain Mapping (2010)
4. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. ACM (2016)
5. Cheng, B., Liu, M., Shen, D., Li, Z., Zhang, D., Alzheimer's Disease Neuroimaging Initiative: Multi-domain transfer learning for early diagnosis of Alzheimer's disease. Neuroinformatics **15**(2), 115–132 (2017)
6. Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q.: Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol. Aging **32**(12), 2322–e19 (2011)
7. He, K., et al.: Deep residual learning for image recognition. In: Computer Vision and Pattern Recognition, December 2015
8. Hu, K., Wang, Y., Chen, K., Hou, L., Zhang, X.: Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis. Neurocomputing **175**, 132–145 (2016)
9. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(Nov), 2579–2605 (2008)
10. Roberson, E.D., Mucke, L.: 100 years and counting: prospects for defeating Alzheimer's disease. Science **314**(5800), 781–784 (2006)
11. Sarraf, S., Tofighi, G.: Classification of Alzheimer's disease structural MRI data by deep learning convolutional neural networks. arXiv preprint arXiv:1607.06583 (2016)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
13. Tustison, N.J., et al.: N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging **29**, 1310–1320 (2010)

14. Tustison, N.J., et al.: The ANTs longitudinal cortical thickness pipeline. In: Proceedings of SPIE (2013)
15. Ustinova, E., Lempitsky, V.: Learning deep embeddings with histogram loss. In: Advances in Neural Information Processing Systems, pp. 4170–4178 (2016)