PRISMA-CAD: Fully automathed method for Computer-Aided Diagnosis of Dementia based on structural MRI data

N. Amoroso[1,2], R. Errico[2,3], and R. Bellotti[1,2]

[1] Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy
[2] Dipartimento Interateneo di Fisica, Università degli Studi di Bari, Italy
[3] Istituto Nazionale di Fisica Nucleare, Sezione di Genova, Italy
for the Alzheimer's Disease Neuroimaging Initiative **

Abstract. Neurodegenerative association with structural changes of the brain has been widely investigated gaining profound knowledge on specific aspects of the healthy and diseased brain. Medial temporal lobe atrophy and, in particular, the hippocampal atrophy are important biomarkers for the Alzheimer's disease. In this paper we describe how MRI brain scans can be processed and analyzed, in a fully automated framework, to segment relevant anatomical structures, extract morphometric and statistical features and perform an accurate clinical classification on the basis of anatomical and statistical features. We trained an artificial neural on a population consisting of 288 subjects to discriminate normal control subjects (NC), from those affected by Alzheimer's disease (AD) and mild cognitive impairment (MCI) with a one versus one strategy. Performances were validated with k-fold procedure, NC-AD were discriminated with accuracy ACC (NC-AD) = 0.91 while the overall accuracy ACC (NC-MCI-AD) reached the 0.81 value.

1 Introduction

Neuroscience is generating exponentially growing volumes of data and knowledge on specific aspects of the healthy and diseased brain, in different species, at different ages. However, there is no effective strategy to experimentally map the brain across all its levels and functions, yet. A proof of interest in the field is the recent funding of worldwide initiatives, such as the Human Brain Project [4] and the Human Connectome Project [5]l.

** Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: *http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf*

[4] www.humanbrainproject.eu

[5] www.humanconnectomeproject.org

Medical image computing raises new challenges related to the scale and complexity of the required analyses. For example, magnetic resonance imaging (MRI) of the brain plays a fundamental role for detection of neurodegeneration. The manual segmentation on MRI has been so far considered the only available strategy to accurately access reliable structural biomarkers and therefore to achieve a sound quantitative clinical discrimination. Nevertheless, manual segmentation is a time-consuming task nor it can manage the intrinsic human intra-rater variability, this is why automated processing pipelines are needed as diagnosis support systems.

In the present paper a novel fully automated processing workflow is described. It consists of three main steps. Firstly the pre-processing, an automated rigid registration and histogram based equalization for spatial and intensity normalization. Then, a volume of interest (VOI) extraction is performed; this VOI individuates a gross region containing the left and right hippocampi, from this region important features as the hippocampal volume or its thickness are calculated. Finally, the classification (NC - MCI - AD) is obtained with an automated artificial neural network.

2 Materials

The goal of this work is to provide a fully automated and reliable diagnosis support system to discriminate NC - MCI - AD. The CADDementia challenge aims to compare several methods and protocols to unveil, on the basis of a common test set whether significant differences exist among the various algorithms. According to this a standardized evaluation framework is set up, consisting of 384 multi-center scans. The participating centers are: Erasmus MC (EMC), Rotterdam, the Netherlands; VU University Medical Center (VUmc), Amsterdam, the Netherlands; University of Porto / Hospital de São João (UP), Porto, Portugal This data set contains structural MRI (T1w) scans of subjects with the diagnosis of probable Alzheimer's disease (AD), mild cognitive impairment (MCI) and participants without a dementia syndrome (controls). In addition to the MR scans, demographic information (age, gender) and information on which data are from the same institute is included.

To reach this goal 30 MRI brain scan are provided by the MICCAI CADDementia challenge (http://caddementia.grand-challenge.org) for training. Nevertheless, an increased basis of knowledge should help classification to build more generalized model and this is why a second dataset consisting of 258 MRI brain scans shared by the Alzheimer's Disease NeuroImaging Initiative (ADNI) was used. The two training databases used are described with demographics given in table 1.

3 Methods

In this study a fully automated pattern recognition system for accurate and reproducible segmentation of the hippocampus and the peri-hippocampal region

Data	Size	Age	M/F	Subjects
ADNI	258	60 - 96	144/114	96 NC - 96 MCI - 66 AD
MICCAI	30	54 - 80	17/13	12 NC - 9 MCI - 9 AD

Table 1. Data demographics. Group size, range age (years) and resolution of the two clinical datasets, containing normal control (NC) subjects, Alzheimer's Disease (AD) and mild cognitive impairment (MCI) patients.

in structural Magnetic Resonance Imaging (MRI) was used. This procedure, described in detail in our previous works in [1,2,3,4], is schematically shown in figure 1.

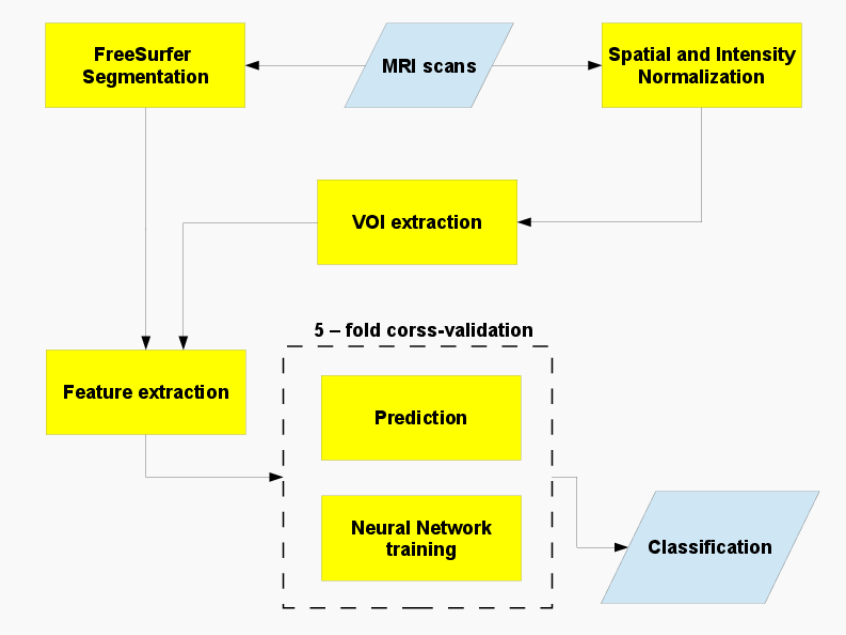

Fig. 1. The figure represents the overall processing pipeline. The MRI scans are processed independetly by FreeSurfer and an automated segmentation pipeline hippocampus-focused. The calculated features are used to train an artificial neural network in a 5-fold cross validation frameowrk, then finally classification is performed.

The system consisted of three processing levels: (a) MRI brain scans were linearly registered to the standard MNI152 template and an automated point distribution shape analysis method was used to define the peri-hippocampal region. (b) Feature extraction: the peri-hippocampal VOI was statistically analyzed; gray level distribution features such as means, standard deviations, kurtosis and skewness were calculated. Moreover, other morphometric hippocampal

based features such as the whole volume, the thickness, or local geometric features were calculated. In addition other features were calculated with a publicly available brain segmentation package FreeSurfer v.5.1 [6] [5] for an overall amount of 248 features. (c) Subject classification: a back propagation neural network was used to classify examples as NC, MCI and AD. An unsupervised filter was used to explore the feature space and determine correlations and linear dependences, in this way a subsample of about 150 features was determined. Features were normalized with the intra-cranial volume and finally normalized then, a one versus all strategy was adopted for training the network.

The network architecture was kept as simple as possible to avoid over-training issues, just one hidden layer with 10 neurons was used and a regularized cost function was adopted. To improve the generalization of the trained model a 5-fold strategy was used, in addition a random sampling of 50 features for every cross-validation round was performed. We repeated this procedure for a hundred times, thus obtaining 1500 trained networks. For every cross-validation round a 288×3 score matrix is obtained. Training performances are obtained by averaging the score matrix for every cross-validation and then averaging the class probabilities obtained from the different classifiers, in fact according to the one versus we trained a NC vs AD classifier, a NC vs MCI classifier and an AD vs MCI classifier, thus for example for NC subjects two distinct probabilities were given for each cross-validation step.

3.1 Computational infrastructure

The analyses presented in this paper were developed in MATLAB framework and required substantial computational resources. The previously described automated processing pipeline required an overall processing time of about 13 hours for subject, this processing time was almost entirely due to FreeSurfer. In fact, the processing time required to extract the hippocampal and the statistical features did not exceed one hour per subject. Therefore the use of dedicated workflow manager such as the LONI pipeline processing environment [6][7] was used: a user-friendly and efficient software for complex data analyses, available at http://pipeline.loni.ucla.edu and an adequate distributed infrastructure was of fundamental importance.

The analyses were carried out using the local computer farm BC2S [7]: a distributed computing infrastructure consisting of about 5000 CPU and allowing up to 1,8 PB storage. A further study for grid deployment was also performed, within the aim of creating a pipeline tool suitable for large clinical trials. It was carried out on the European Grid Infrastructure (EGI) which consists of about 300 geographically distributed sites around the world. The run-time reduction with the grid implementation allowed to produce results in a reasonable time with respect to the application execution as a sequential process on limited resources. The advantages of the grid execution were evident since we obtained the 90% of the analysis of 642 images after less than 16 hours.

[6] freesurfer.nmr.mgh.harvard.edu
[7] http://www.recas-pon.ba.infn.it

4 Results

The overall training results showed a significant discrimination among the three populations. Performances were measured in terms of accuracy (ACC) and area under the receiver operating characteristic (AUC). Figure **??** shows the overall training results.

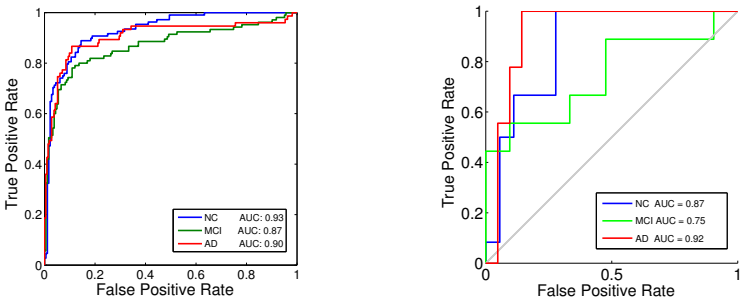

Fig. 2. The figure represents the receiver operating characteristics of the three classes: NC (blue), MCI (green) and AD (red) for both the overall training set (on the left) and the 30 MICCAI images (on the right). AUC is also reported for all of them, it can be seen how NC and AD are recognized slightly better than MCI.

Performances on the reduced training set (the MICCAI data) resulted significantly lower, with an average accuracy $\bar{ACC} = 0.67 \pm 0.3$. However the data size does not allow to draw statistically significant conclusions 3.

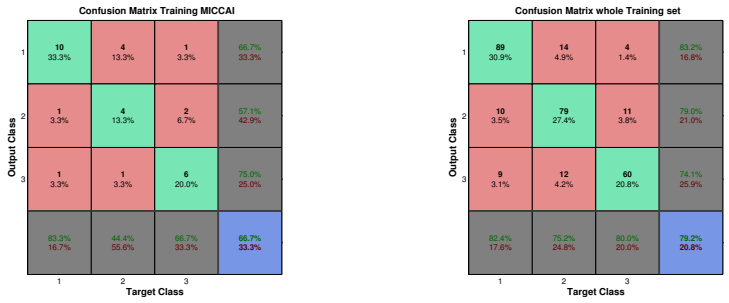

Fig. 3. The figure shows the confusion matrix relative to the overall training (on the left) and the 30 MICCAI training images (on the right). Classes 1-2-3 are respectively the NC, the MCI and the AD classes.

The networked already trained were then used to obtain the test predictions. To improve generalization, we randomly sampled 300 classifiers from the 1500

already trained (100 for each type: NC-AD, NC-MCI and AD-MCI) and obtained for test example a prediction to be NC, MCI and AD. As previously explained for training, the average class probability was reported as the final classification score.

5 Discussion and Conclusion

Accuracy and area under the curve results suggest the method is reliable, besides being fully automated it can be adopted for large studies without suffering of intra-rater variability nor requesting time-intensive manual work from experts.

Classification performances compare well state-of-the-art performances thus suggesting the overall analysis workflow is reliable. The reduced number of features used for classification also suggest the possibility to significantly improve performances with the individuation of new features. The most important features for classification resulted to be those correlated to temporal lobe and hippocampal atrophy, on one hand this demonstrates that the volumes obtained by the proposed workflow are reliable, on the other it should also suggest to explore and investigate new features to improve the clinical discrimination.

Unsupervised approaches, such as deep learning networks, could be naturally included in the proposed framework and will be investigated in future works. Besides, those issues deriving from the computational burden yielded by FreeSurfer should also be addressed, the most promising strategy, according to our results, could be the individuation of statistical features which could substitute those obtained by FreeSurfer.

Acknowledgments

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant

U01 AG024904) and DOD ADNI (Department of Defense award number W81XW H -12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

This research was supported by Istituto Nazionale di Fisica Nucleare (INFN), Italy. This research was also supported by grants from Università degli Studi di Bari, Italy.

All authors disclose any actual or potential conflicts of interest, including any financial, personal, or other relationships with other people or organizations that could inappropriately influence their work. All experiments were performed with the informed consent of each participant or caregiver in line with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Local institutional ethics committees approved the study.

References

1. Amoroso, N., Bellotti, R., Bruno, S., Chincarini, A., Logroscino, G., Tangaro, S., Tateo, A.: Automated Shape Analysis landmarks detection: Model and Developments. Proceedings of the International Symposium, CompIMAGE (2012) 139–142
2. Tangaro, S., Amoroso, N., Bruno, S., Chincarini, A., Frisoni, G.B., Maglietta, R., Tateo, A., Bellotti, R.: Active Learning Machines for Automatic Segmentation of Hippocampus in MRI . Industrial Conference on Data Mining (ICDM 2013). LECTURE NOTES IN COMPUTER SCIENCE (2013)
3. Maglietta, R., Amoroso, N., Bruno, S., Chincarini, A., Frisoni, G.B., Inglese, P., Tangaro, S., Tateo, A., Bellotti, R.: Random forest classification for hippocampal segmentation in 3D MR images. IMLA 2013 conference inproceedings (2013)
4. Tangaro, S., Amoroso, N., Boccardi, P., Bruno, S., Chincarini, A., Ferraro, G., Frisoni, G.B., Maglietta, R., Redolfi, A., Rei, L., Tateo, A., Bellotti, R.: Automated voxel-by-voxel tissue classification for hippocampal segmentation: Methods and Validation. in press (2014)
5. Fischl, B.: FreeSurfer. Neuroimage **62** (2012) 774–781

6. Dinov, I.D., Lozev, K., Petrosyan, P., Liu, Z., Eggert, P., Pierce, J., Zamanyan, A., Chakrapani, S., Van Horn, J.D., Parker, D.S., Magsipoc, R., Leung, K., Gutman, B., Woods, R.P., Toga, A.W.: Neuroimaging Study Designs, Computational Analyses and Data Provenance Using the LONI Pipeline. PLoS ONE 5 **9**
7. Rex, D.E., Ma, J.Q., Toga, A.W.: The LONI Pipeline Processing Enviroment. Neuroimage **19** (2003) 1033–1048