Structured Multivariate Pattern Classification to Detect MRI Markers for an Early Diagnosis of Alzheimer's Disease

Cécilia Damon, Edouard Duchesnay, Marine Depecker

Structured multivariate pattern classification to detect MRI markers for an early diagnosis of Alzheimer's disease

Cécilia Damon
CEA, LIST, LIMA
91191 Gif-sur-Yvette, France
Email: cecilia.damon@cea.fr

Edouard Duchesnay
CEA, I2BM, Neurospin
91191 Gif-sur-Yvette, France
Email: edouard.duchesnay@cea.fr

Marine Depecker
CEA, LIST, LIMA
91191 Gif-sur-Yvette, France
Email: marine.depecker@cea.fr

***Abstract*—Multiple kernel learning (MKL) provides flexibility by considering multiple data views and by searching for the best data representation through a combination of kernels. Clinical applications of neuroimaging have seen recent upsurge of the use of multivariate machine learning methods to predict clinical status. However, they usually do not model structured information, such as cerebral spatial and functional networking, which could improve the predictive capacity of the model and which could be more meaningful for further neuroscientific interpretation. In this study, we applied a MKL-based approach to predict prodromal stage of Alzheimer disease (*i.e.* early phase of the illness) with prior structured knowledges about the brain spatial neighborhood structure and the brain functional circuits linked to cognitve decline of AD. Compared to a set of classifical multivariate linear classifiers, each one highlighting specific strategies, the smooth MKL-SVM method (*i.e.* L_p MKL-SVM) appeared to be the most powerful to distinguish both very mild and mild AD patients from healthy subjets.**

I. Introduction

Multivariate machine learning methods are increasingly used in neuroimaging to detect predictive patterns of a variable of interest [1]. Contrary to univariate methods, such as voxel-based morphometry [2], they provide label prediction based on the information carried by a set of features. Recently, they have also been exploited to identify, in cerebral anatomical MRI, spatial distributed structural abnormalities underlying prodromal stage of Alzheimer's Disease (AD) [3], [4]. Nowadays, only postmortem histopathological examinations can make AD diagnosis certain while accurate diagnosis of prodromal AD is crucial for early therapeutic care of dementia sufferers.

Classical multivariate learning methods analyze all the available information together in a blind way without considering well-known constraints inherent to the field, such as the functionning and structural networking of the brain or some additional contextual information, such as multiple sources of input data. Multiple Kernel Learning (MKL) alows to resolve the learning problem in a more realistic way by managing the introduction of contextual information under the form of a linear convex combination of kernels [5]. That is combining subsets of features or several data modelings (*i.e.* different parametrizations of kernel functions), consistent according to a prior knowledge. Besides attempting to improve the predictive capacity of the model [6], [4], an important objective of MKL is to make the interpretation of the results more meaningful [7].

Based on this idea, we proposed an MKL-based approach aiming at exploiting both spatial and functional information, by considering respectively sets of neighboring subcortical regions of interest (ROIs) and classical functional network of ROIs, more specifically those with injuries underlying cognitive decline in AD. We highlighted the benefits of taking into account structured information by comparing the performances of an SVM-based MKL method with those of classical multivariate linear methods for the prodromal AD (very mild and mild stages) prediction. Each considered method reflects a particular strategy in its solution modeling and learning. We dealt in particular with the impact on the predictive capacity of two characteristics present in the proposed strategies, through a set of models comparisons : features interactions modeling and the level of model complexity. Moreover, making use of linear methods and kernels yields to easy-to-interpret models, giving us good intel on the predictive MRI patterns for prodromal AD and enabling us to visualize these anatomical predictive patterns.

This paper is organised as follows. The general framework of this study is described in Section II. Then, our main contribution is detailed in Section III, giving the description of the conducted experiments. Finally, the performances of the learning methods are compared and analyzed in Section IV. A discussion related to these results and further perspectives are given in Section V.

II. Materials and Methods

Data: the data used for the experiments presented in this paper come from the cross-sectional dataset of the Open Access Series of Imaging Studies (OASIS) (www.oasis-brains.org), a neuroimaging database freely available to the scientific community. This dataset contains T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) scans for 416 subjects aged from 18 to 96 years, male and female, and all right-handed [8]. The Alzheimer dementia stage is characterized using the Clinical Dementia Rating (CDR) scale

that rates gradual onset and progression of decline in memory, cognitive and functional domains [9]. A CDR equal to 0.5, 1 and 2 corresponds to very-mild (or MCI), mild and moderate AD patients, respectively. A CDR equal to 0 corresponds to healthy subjects.

Neuroscientists being mainly interested in understanding AD at the region level, we decided to analyse the whole-brain subcortical anatomy instead of the classical whole brain voxel-based approach. Hence, we tested our methods on a dataset containing volumes of 43 subcortical regions automatically extracted with the Freesurfer software [10]. To take account for head size differences, subcortical ROIs volumes have been normalized respectively by the intra-cranial volume (ICV) [11]. Regarding age, we selected only subjects over 60 years as the observation of patients with dementia symptoms (*i.e.* CDR>0) begins from that age and we have excluded the two only subjects diagnosed as moderate AD (*i.e.* CDR=2) because of their weak representativeness. We thus obtain that the age range, mean and standard deviation of the different groups (*i.e.* control, very mild and mild AD) are relatively similar. Regarding gender, the three groups do not include the same proportions of males and females.We thus have rebalanced the proportions in such a way to have a similar number of males and females in each group. We also have removed critical subjects (MMSE<20) in order to restrict the dementia sufferers group to typical patients.

Learning methods: we first considered the Gaussian Naive Bayes (GNB), a parametric model assuming features independance conditionnally to each class. Then, we considered three methods that model features interactions : Regularized logistic regression (RLR), Linear Support Vector Machine (SVM) and Multiple kernel learning based on support vector machines (MKL-SVM) [12], [5].

Regularized logistic regression (RLR) is based on classical logistic regression that represents the class-conditional probabilities through a linear function of features. For sparse (resp. smooth) logistic regression, a L_1 (resp.L_2) regularization is introduced through the penalization of the weighted vector $\mathbf{w}$. The predictive model is then obtained by maximizing the L_1 (resp.L_2) penalized log-likelihood.

Linear Support vector machine (SVM) is a non-parametric classifier whose predictive function is based on a linear kernel function. For binary classification problems, it builds a separating hyperplane in the samples-space maximizing the margin between the two classes.

Multiple kernel learning based on Linear SVM (MKL-SVM) considers a convex linear combination of several basis kernels, one per subset of features (or distinct data origin). In our experiment, we have assigned a linear kernel per structured cluster of anatomical ROIs spatially and functionally neighboring. We build ten clusters and thus ten linear kernels, called subkernels, to produce a final single kernel matrix which is optimized to both find the best separation on the training set and limit the complexity model (*i.e.* $||\beta||_l^l$, where β_k is the subkernel weight of the k-th kernel). For the norm l of the subkernel weight vector, we tested three levels of model capacity constraint : sparse with the 1-norm, non-sparse with the p-norm ($l = p \in]1, 4]$) and the extreme case of non-sparsity with the ∞-norm ($l = 10^6$), incrementing all subkernel weights to one [13]. Kernels are normalized to avoid biased contributions due to magnitude differences present in the initial data. The MKL-SVM problem can be written as follows :

$$\begin{array}{ll} \arg\min_{\mathbf{w}_k,\epsilon,\beta_k,b} & \sum_k \frac{||\mathbf{w}_k||_2^2}{\beta_k} + C\sum_{i=1}^m \epsilon_i + ||\beta||_l^l, \\ \text{under the constraints:} & y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \epsilon_i, \\ & \epsilon_i \geq 0, i = 1, \ldots, m \end{array} \quad (1)$$

Validation algorithm: for purpose of testing and comparing several learning methods, the original dataset has been divided into two subsets : a training dataset containing $2/3$ of the original dataset, the remaining data being used for testing.

For penalized logistic regressions and linear SVM, we considered a range of powers of ten values spaced between 10^{-3} and 10^3 for the penalty parameter C, which has been optimized with an inner 10-folds cross-validation on the training sample. Data and kernels normalization for MKL-SVM have been systematically computed on each training fold and applied on the corresponding test fold. Moreover, initial imbalance of the group sizes have been preserved using stratified re-sampling.

Classification accuracy has been evaluated by averaging sensitivity and specificity scores to take into account imbalanced groups of subjects. Sensitivity gives the rate of AD sufferers correctly identified while specificity evaluates the proportion of healthy patients correctly identified. Statistical significance of the classifiers accuracy (*p-value*) has been assessed by computing the probability that a random classifier following a binomial distribution with probability 0.5 ($\mathcal{B}(0.5)$) could get a better score.

III. Contributions

Given the three groups, we aim at discriminating the following populations :

- Control subjects (CDR=0, 67 subjects) from very mild dementia sufferers (CDR=0.5, 51 patients). We called this comparison "CvsVM" for "Control *versus* Very Mild".
- Control subjects (CDR=0, 95 subjects) from mild dementia sufferers (CDR=1, 19 patients). We called this comparison "CvsM" for "Control *versus* Mild".

The main objective of this study was to assess the benefits of introducing structured priors in the learning process, *i.e.* some dependencies between features, to detect biomarkers that may enable an earlier identification of patients in the prodromal stage of AD. Based on *a priori* knowledges and using the MKL-SVM learning method, we have been able to incorporate contextual information about the spatial and functional neighboring of the brain subcortical ROIs. Therefore, we considered ten clusters of regions of interest that both corresponded to some parts of well-known anatomical systems (basal ganglia, limbic system) and networks of neighboring subcortical structures [14]. Each cluster has been then associated to a kernel

of the MKL-SVM method.

We first proposed two clusters gathering left- and right-hemispheric ROIs, called *left- and right-Brainstem clusters*, the *brainstem* representing the main intermediate for motor and sensory innervation between brain and body. Then, we formed two additional clusters, called *left- and right-Center clusters*, based on the hemispheric neighbors of the medial part of each lateral ventricle and related to the *corpus callosum* that connects the right and the left hemispheres and allows their communication. An other cluster has been considered to represent the ventricular system composed of three connected cavities containing cerebrospinal fluid, called *Ventricles cluster*. By extending the latter, we created a new cluster containing structures belonging to the diencephalon, called *Diencephalon cluster*. Finally, we introduced two functional sets of subcortical structures, one corresponding to the basal ganglia, underlying action selection and reinforcement learning, and the other related to a part of the limbic system associated with emotion, memory and motivation. For each we considered one cluster per hemisphere, thus called *left- and right-Basal ganglia* and *left- and right-limbic system*.

To deal with the benefit of modeling structured priors, we first compared MKL-SVM with Linear SVM, which considered all the available features inside a single kernel. Then, we compared MKL-SVM with RLR, which did not include structured priors either but has been a goog competitor. We only compared L_1 MKL-SVM with L_1 RLR and L_p MKL-SVM with L_2 RLR to confront models with similar type of regularization.

Given all the tested methods, several comparisons have been made, which emphasized the impact of different strategies on podromal AD prediction. In particular, we highlighted the impact of modeling features interactions in terms of classification accuracy by considering the comparison between a GNB classifier, that does not take into account interactions between features, with the L_2-RLR classifier, the latter being the best competitor among the set of classical multivariate methods modeling features interactions.Then, we studied the impact of model complexity constraints on the classifiers performances for the regularized logistic regression technique, *i.e.* sparse L_1 RLR *versus* smooth L_2 RLR, and for the MKL-SVM approach, *i.e.* sparse with L_1 norm *versus* non-sparse leading to a smooth constraint with L_p norm ($p > 1$) *versus* non-sparse leading to a uniform-weighted with L_∞ norm.

IV. Results

The purpose of this section is to present the empirical comparison of the learning methods previously described based on their performances for the two study cases (*CvsM* and *CvsVM*) given in table 1.

For both experiments, all the tested methods gave significant accuracies at least for a statistical threshold of 5%. L_p MKL-SVM method achieved the best accuracies and the highest sensitivity to discriminate both very mild and mild dementia sufferers from healthy subjects.

Experiments	Classifiers						
	GNB	SVM	L_2 RLR	L_1 RLR	L_1 MKL-SVM	L_p MKL-SVM	L_∞ MKL-SVM
CvsM	85,9*** (71,9 ; 100,0)	64,1* (78,1 ; 50,0) C=0.1	85,9*** (71,9 ; 100,0) C=10^{-3}	72,9** (62,5 ; 83,3) C=1	85,9*** (71,9 ; 100,0)	89,1*** (78,1 ; 100,0) p=4.0	80,7*** (78,1 ; 83,3)
CvsVM	71,5** (78,3 ; 64,7)	70,1** (69,6 ; 70,6) C=1	81,8*** (69,6 ; 94,1) C=0.1	71,6** (60,0 ; 82,4) C=1	75,2*** (73,9 ; 76,5)	82,5*** (82,6 ; 82,4) p=3.6	82,5*** (82,6 ; 82,4)

Fig. 1. For each experiment *CvsM* and *CvsVM*, the table shows the classification accuracy of each tested strategy with its statistical significance. Stars represent significance threshold (*:0.05, **: 0.01, ***: 0.001. Between brackets, two numbers represent respectively sensibility and specificity scores. and below is given the model parameter selected by cross-validation.

Modeling of structured a priori knowledges: the MKL-SVM method, that used multiple kernels to model structured a priori knowledges, improved the preformances obtained with the SVM and the RLR methods for both classification experiments.

Modeling features interactions: the comparison between the GNB classifier and the L_2 RLR, showed that both classifiers obtained identical results (overall accuracies, senstivity and specificity scores) for the experiment *CvsM*.. We noted that they achieved similar performances with really opposite modeling strategies since the L_2 RLR model involved the weakest penalization ($C = 10^{-3}$), the amount of information, including features interactions, being thus only slightly constrained. For the experiment *CvsVM*, the L_2 RLR classifier outperformed the GNB classifier. Modeling features interactions is probably relevant to detect very mild stage of AD because the volumic losses and enlargements are smaller and and less diffuse in the brain than in the mild stage of AD. Hence, features interactions provide additional information enabling to improve the learning process and leading to a better model.

Level of complexity: we first focused on the comparison of sparse (L_1 penalization) and smooth (L_2 or L_p penalization) models. For both classification experiments, L_2 RLR model (resp. L_p MKL-SVM model) gave higher and more significant accuracies than L_1 RLR model (resp. L_1 MKL-SVM model). We concluded that, with a sparse penalization, relevant information could be lost since some kernels and all the features inside (regarding the L_1 MKL-SVM) or some features (regarding the L_1 RLR) are set to zero.

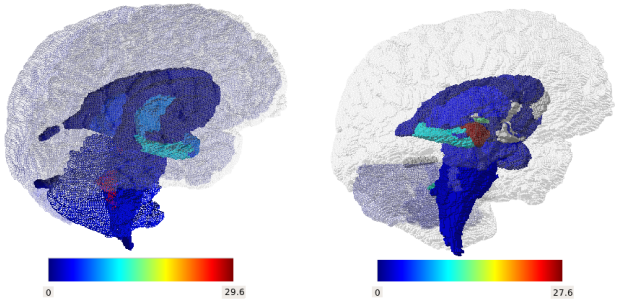
Then, by comparing the three levels of penalization for MKL-SVM (sparse, smooth and uniform), we noted that L_p MKL-SVM gave the most predictive performances whatever the classification experiments. A smooth constraint allowed to detect a more relevant predictive pattern of anatomical ROIs networks than the extreme cases of sparse and uniform constraints. Moreover, despite the fact that the L_p MKL-SVM method involved additional computational cost since we had to select the regularization norm p, it offered the advantage to learn the intrinsic complexity of the set of kernels.

Predictive anatomical patterns of L_p MKL-SVM models: by using only linear kernels, we have been able to visualize the predictive anatomical pattern since a weight has been assigned to each kernel and also to each feature. Hence, we computed the final weight of each subcortical ROI through all the kernels that highlighted its contribution to the classification decision.

Larger is this weight, larger is its relevance to discriminate both groups.

For the experiment *CvsM*, the highest weight of the discriminative pattern (see Figure 2(a)) is located on the fourth ventricle and average weights to its neighbouring structures, the cerebellum and brain stem ROIs. This is coherent with some neuroscientific researches having identified the brainstem area as being affected by the senile plaques in the early stage of AD [15].MKL-SVM model also gave an important contribution to the hippocampus, thalamus, amygdala and to the corpus callosum, which belong to the limbic system essential for long-term memory and well-known to be impacted in AD [16].

For the experiment *CvsVM*, the L_p MKL-SVM model assigned the highest weight to amygdala structure (in the right hemisphere) and high and average weights to close structures, all involved in the limbic system, such as the hippocampus, a part of corpus callosum, the lateral ventricle and a part of the thalamus (see Figure 2(b)). The discriminative pattern is also defined by an important and an average contribution from the volumes of the pallidum and the putamen structures, respectively. Both of these structures border and are in relation to the basal nucleus of meynert (part of the limbic system) suffereing a loss of cholinergic neurones in the early stage of AD [17]. Like for the experiment *CvsM*, the volume changes of the brainstem and fourth ventricle structures appeared to be discriminant.



(a) (b)

Fig. 2. 3D map of the discriminative pattern captured by the L_p MKL-SVM model for the experiments *CvsM* (a) and *CvsVM* (b). The weight of each subcortical ROI computed through all the kernels is represented by a color (cf palette) in the brain 3D volume that highlighted its contribution to the classification decision.

V. Conclusion and perspectives

The modeling of a priori knowledges with the multiple kernel learning improved the classification of individuals in a prodromal stage of Alzheimer's disease. Moreover, the use of an optimal smooth L_p norm brought more flexibility than fixed-norm counterparts by finding the intrinsic complexity of the given set of kernels. Then, L_p MKL-SVM strategy is an interesting competitor method to sparse strategies widely used on real data to improve interpretability. It is also an efficient technique to test some hypothesis such as the involvement of a specific anatomical network in AD.

Further works will come to confirm these results on the sizeable Alzheimer's disease Neuroimaging Initiative (ADNI) database and to study the introduction of multi-modality data and the identification of neuroimaging markers predictive of the progression of mild cognitive impairement (MCI) to AD which is an important medical challenge.

References

[1] P. Golland, B. Fischl, M. Spiridon, N. Kanwisher, R. Buckner, M. Shenton, R. Kikinis, A. Dale, and W. Grimson, "Discriminative analysis for image-based studies," *Medical Image Computing and Computer-Assisted Intervention MICCAI 2002*, pp. 508–515, 2002.

[2] J. Ashburner and K. Friston, "Voxel-based morphometry–the methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000.

[3] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M. O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *Neuroimage*, vol. 56, pp. 766–781, May 2011.

[4] S. Kloppel, C. Stonnington, C. Chu, B. Draganski, R. Scahill, J. Rohrer, N. Fox, C. Jack, J. Ashburner, and R. Frackowiak, "Automatic classification of mr scans in alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689, March 2008.

[5] G. Lanckriet, N. Cristianini, P. Bartlett, L. Ghaoui, and M. Jordan, "Learning the kernel matrix with semidefinite programming," *The Journal of Machine Learning Research*, vol. 5, pp. 27–72, 2004.

[6] C. Hinrichs, V. Singh, G. Xu, and S. Johnson, "Predictive markers for ad in a multi-modality framework: An analysis of mci progression in the adni population," *Neuroimage*, 2010.

[7] R. Cuingnet, M. Chupin, H. Benali, and O. Colliot, "Spatial and anatomical regularization of svm for brain image analysis," *Proceedings of the Neural Information Processing Systems conference*, 2010.

[8] D. Marcus, T. Wang, J. Parker, J. Csernansky, J. Morris, and R. Buckner, "Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.

[9] J. Morris, "Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the alzheimer type," *International Psychogeriatrics*, vol. 9, no. 1, pp. 173–176, 1997.

[10] B. Fischl, D. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness *et al.*, "Whole brain segmentation:: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, 2002.

[11] R. Buckner, D. Head, J. Parker, A. Fotenos, D. Marcus, J. Morris, and A. Snyder, "A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume," *Neuroimage*, vol. 23, no. 2, pp. 724–738, 2004.

[12] T. Hastie, R. Tibshirani, and et al. et al., Eds., *The Elements of Statistical Learning*. Springer Series in Statistics, 2001.

[13] M. Kloft, U. Brefeld, S. Sonnenburg, P. Laskov, K. Müller, and A. Zien, "Efficient and accurate lp-norm multiple kernel learning," *Advances in neural information processing systems*, vol. 22, no. 22, pp. 997–1005, 2009.

[14] N. Makris and al. (Center for Morphometric Analysis), *General Brain Segmentation-Method and utilization, Version 3*, 2004. [Online]. Available: http://www.cma.mgh.harvard.edu/manuals/segmentation/

[15] L. Grinberg, U. Rueb, and H. Heinsen, "Frontiers: Brainstem: neglected locus in neurodegenerative diseases," *Frontiers in Dementia*, vol. 2, 2011.

[16] M. De Leon, A. George, L. Stylopoulos, G. Smith, and D. Miller, "Early marker for alzheimer's disease: the atrophic hippocampus." *Lancet*, vol. 2, no. 8664, p. 672, 1989.

[17] I. Sassin, C. Schultz, D. Thal, U. Rüb, K. Arai, E. Braak, and H. Braak, "Evolution of alzheimer's disease-related cytoskeletal changes in the basal nucleus of meynert," *Acta neuropathologica*, vol. 100, no. 3, pp. 259–269, 2000.