Computer aided diagnosis system based on random forests for the prognosis of Alzheimer's disease

M. Wehenkel[1,2], C. Bastin[2], P. Geurts[1], C. Phillips[2]
[1]Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgium,
[2]GIGAResearch, University of Liège, Liège, Belgium.

Introduction: Over the last decade, a large number of computer aided diagnosis (CAD) systems have been developed by researchers in neuroimaging to study neurodegenerative diseases or other kinds of brain disorders [1,2,3]. Briefly, machine-learning (ML) techniques help doctors to distinguish groups of people (e.g. healthy vs. diseased) by automatically identifying characteristics in the images that discriminate the groups. The challenge in the modelling of CAD systems is not only to perform well in terms of prediction but also to provide relevant information about the diagnosis, such as regions of interest in the brain that are affected by the disease.

In this abstract, we propose an original CAD system consisting in the combination of brain parcelling, ensemble of trees methods, and selection of (groups of) features using the importance scores embedded in tree-based methods. Indeed, on top of their ease of use and accuracy without ad hoc parameter tuning, tree ensemble methods such as random forests (RF) [4] or extremely randomized trees (ET) [5] provide interpretable results in the form of feature importance scores. We also compare the performance and interpretability of our proposed method to standard RF and ET approaches, without feature selection, and to multiple kernel learning (MKL) [6]. The latter was shown to be an efficient method notably capable of dealing with anatomically defined regions of the brain by the use of multiple kernels.

Methods: Our CAD system is designed to discriminate older adults with Mild Cognitive Impairments (MCI) in terms of their clinical outcome 4 years later, based on their current PET images. More precisely, 45 individuals presenting mild cognitive impairments (MCI) at the beginning of the study were followed during 4 years and their diagnostic updated based on neuropsychology tests (no further imaging was performed). Among those subjects, 22 patients were eventually diagnosed with Alzheimer's disease (AD) in the course of the study. These were labelled "MCI converters" (MCIc). The others showed no cognitive decline and are thus denominated stable MCI (sMCI). The aim of such a CAD system is thus to predict the likelihood of progression to dementia based on the images acquired before the onset of the disease.

The PET images were pre-processed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/). This included spatial normalization to the MNI reference space and intensity normalization by the cerebellar intensities. Then a feature vector for each individual was built by extracting the voxel values within the brain volume.

The first step of our diagnosis system consists in learning a tree ensemble model and attributing a score per AAL region [7] from the mean of the voxel importance scores in each region. In a second step, the k best regions according to these scores are selected and used to learn a new model (with k set to 10 in our experiments). This approach is thus a combination of group selection and ensemble methods. The procedure used for its assessment is summarized in Figure 1. We evaluate it with a "leave 10% of subjects per group out" cross validation (CV) procedure for RF and ET respectively with default parameter values (M = 500 trees and K= $\sqrt{N}$ where N is the total number of features). Standard RF and ET (without feature selection) and MKL, all with default parameter setting, are also assessed for comparison with the same CV procedure. As RF and ET involve randomization, experiments were repeated ten times, called runs here under, to obtain mean and standard deviation of performance metrics.

To interpret the results of the proposed method and to have insights about regions involved in the prognosis, we compute importance scores for each ensemble of trees. We then average the scores over the folds and the runs and we subsequently compute a score for each brain region. For MKL, we use the weights attributed to each brain area. Finally, as our proposed method embeds a selection process, we also analyze the frequency of selection of brain areas over the folds and the runs to have additional information about important regions.

Algorithm Protocol of model assessment.

Require: Divide the learning set (LS) into X folds.
  for $i = 1 : X$ **do**
    Remove the i^{th} fold from LS.
    Fit an ensemble of trees from the learning set $LS \backslash \{i^{th} fold\}$ to obtain importance scores.
    Compute a score W_R for each set of features.
    Rank the groups of features and choose the ten best groups.
    Build an ensemble model using the ten groups and the set $LS \backslash \{i^{th} fold\}$.
    Test the model on the i^{th} fold.
  end for

Figure 1: Protocol to assess the proposed CAD system.

Results: Table 1 summarizes the accuracy, sensitivity and specificity obtained with each method. MKL is less efficient in terms of accuracy than tree-based approaches. Moreover, we observe that extremely randomized trees, which include supplementary randomization, provide better accuracy than RF. Our proposed CAD system obtains also a better accuracy with ET than RF. The preliminary step of group selection slightly increases mean values of accuracies and sensitivities of ensemble methods and decreases the variance caused by randomization with a large number of features (more or less 200 000 voxels to consider).

Table 1: Summary of method performance and corresponding p-values (obtained using a permutation test with 100 repetitions). The asterisk indicates a $p - value < 0.05$. GS abbreviation is used for group selection.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
MKL	68.89 (0.02)*	59.09 (0.23)	78.26 (0.01)*
RF	77.11 ± 2.58 (0.01)*	71.82 ± 4.18 (0.02)*	82.17 ± 2.47 (0.03)*
ET	80.22 ± 3.22 (0.01)*	77.73 ± 5.85 (0.01)*	82.61 ± 4.10 (0.01)*
GS and RF	78.00 ± 1.26 (0.01)*	76.36 ± 1.92 (0.01)*	79.57 ± 2.93 (0.01)*
GS and ET	80.44 ± 1.75 (0.01)*	78.18 ± 3.59 (0.01)*	82.61 ± 0 (0.01)*

In terms of interpretability with weights for MKL and importance scores for the ensemble methods, we can observe in Table 2 the listing of the ten most contributing regions for each method for the discrimination between MCIc and sMCI. The areas TemporalMidR, AngularR and TemporalMidL are common to the five models. Moreover, ParietalInfR, Vermis7 and TemporalInfR are identified among the most important by each of the tree-based methods. Finally, we analyse the regions that have been selected the most frequently over the folds and the runs during the selection process of our procedure. For RF, in order of decreasing frequency, the ten most frequent are TemporalMidR, AngularR, ParietalInfR, TemporalMidL, TemporalInfR, CuneusL, Vermis7, TemporalInfL, Cerebelum6R and Vermis8 whereas ET identifies TemporalMidR, AngularR, TemporalMidL, ParietalInfR, TemporalInfR, Vermis7, TemporalInfL, Cerebelum6R, Vermis8, Vermis6 as the first ten. Nevertheless, the frequency of selection for the last three listed areas for both methods is at most half the time. Given this information, those regions should likely not be considered as informative to decide if an individual will convert to AD within four years following the start of cognitive impairments.

Table 2: Ranking of the first ten most contributing regions of AAL brain atlas.

Rk	Method				
	MKL	RF	ET	GS and RF	GS and ET
1	TemporalMidR	AngularR	TemporalMidR	AngularR	TemporalMidR
2	AngularR	TemporalMidR	AngularR	TemporalMidR	AngularR
3	Vermis6	ParietalInfR	TemporalMidL	ParietalInfR	ParietalInfR
4	ParietalSupR	TemporalMidL	ParietalInfR	TemporalMidL	TemporalMidL
5	TemporalMidL	Vermis7	Vermis7	CuneusL	Vermis7
6	FrontalSupMedialR	CuneusL	TemporalInfR	Cerebelum10L	TemporalInfR
7	Vermis8	TemporalInfR	TemporalInfL	Vermis7	Cerebelum6R
8	OlfactoryL	Vermis8	Vermis8	TemporalInfR	Cerebelum10L
9	Cerebelum10L	TemporalInfL	Vermis6	ThalamusL	TemporalInfL
10	ThalamusL	TemporalSupR	Cerebelum6R	Cerebelum6R	Vermis6

Discussion: We have shown that, at least for the data and problem considered here, tree-based ensemble methods are competitive methods and that they can outperform other advanced methods like MKL. They exhibit better accuracy, sensitivity and specificity and provide good interpretability through importance scores. Furthermore, group selection combined with ensemble of trees adds more insight about the regions that are relevant to diagnose a MCI patient who is likely to develop Alzheimer's disease within four years. Indeed, group selection enables us to study the frequency of selection of a brain area among the whole set. It should also be noted that the results regarding the most involved regions are coherent with studies showing that MCI patients who are about to develop Alzheimer's disease exhibit more hypometabolic temporoparietal areas than MCI patients who remain stable in the next few years [8]. Another advantage of feature selection is that it improves the sensitivity of the diagnosis, which is the quantity relative to true positive (i.e. MCI converters), and largely reduces the variance induced by the initial huge number of features and the randomization process. Finally, the ET approach, with or without group selection, gives rise to accuracy slightly higher than that of RF. Nevertheless, supplementary tests are needed to assess if the differences of accuracy between the distinct methods are statistically significant.

To conclude, we show that using group selection combined with ensemble of trees compose a good CAD system which can help making a correct early prognosis of people suffering of mild cognitive impairments.

References

[1] Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R. Jr., Ashburner, J., Frackowiak, R.S. 2008. Automatic classification of MR scans in Alzheimer's disease. Brain. 131, 681-689.

[2] Garraux, G., Phillips, C., Schrouff, J., Kreisler, A., Lemaire, C., Degueldre, C., Delcour, C., Hustinx, R., Luxen, A., Destée, A., Salmon, E. 2013. Multiclass classification of FDG PET scans for the distinction between Parkinson's Disease and Atypical Parkinsonian Syndromes. NeuroImage Clinical. 2, 883-893.

[3] Fu, C.H., Mourao-Miranda, J., Costafreda, S.G., Khanna, A., Marquand, A.F., Williams, S.C., Brammer, M.J. 2008. Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. Biol. Psychiatry 63, 656-662.

[4] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[5] Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63(1), 3-42.

[6] Bach, F., Lanckriet, G., Jordan, M. 2004. Multiple kernel learning, conic duality, and the SMO algorithm. Proceedings of the 21st International Conference on Machine Learning. 41-48.

[7] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M. 2002. Automated Anatomical Labeling of activations in SPM using a Macroscopic Anatomical Parcellation of the MNI MRI single-subject brain. NeuroImage 15, 273-289.

[8] Chételat, G., Desgranges, B., De La Sayette, V., Viader, F., Eustache, F., & Baron, J. C. (2003). Mild cognitive impairment: Can FDG-PET predict who is to rapidly convert to Alzheimer's disease? Neurology, 60, 1374-1377.