Accuracy and interpretability, tree-based machine learning approaches.

Submission Number:

2152

Submission Type:

Abstract Submission

Authors:

Marie Wehenkel[1], Pierre Geurts[1], Christophe Phillips[1]

Institutions:

[1]University of Liège, Liège, Belgium

Introduction:

For several years, neuroscientists have increasingly shown interests in pattern recognition techniques for the analysis of neuroimaging data [1,2]. The main aims are the development of accurate diagnosis systems but also the identification of brain regions related to the disease.
In particular, kernel methods such as Support Vector Machine (SVM)[3,4] are commonly used. With linear kernels, these approaches combine good accuracy, despite their simplicity, and good interpretability through the interpretation of feature weight maps [10]. At the cost of interpretability, non-linear kernels can potentially improve performance by capturing non-linear dependencies in the data.
So far, tree approaches have not been really popular in neuroimaging. Yet, with minimal tuning they exhibit several very interesting characteristics: they provide non-linear models, state-of-the-art accuracy on many problems and interpretable results through variable importance scores. In this work, we evaluate several standard tree methods and show that the best of them is competitive with SVM both in terms of accuracy and interpretability.

Methods:

We consider the single regression tree (ST) method along with 3 tree ensemble methods: random forests (RF) [7], extremely randomized trees (ET) [8], and LogitBoost (LB) [9]. RF and ET build parallel ensembles of randomized trees, while LB builds an additive ensemble of models in an iterative way. We employ two LB variants: either with a single optimized decision stump (i.e. a tree reducing to only one split) at each iteration (LB^1) or with an ensemble of 50 stumps (grown with the randomization of [8]) at each iteration (LB^2).

We compare the tree-based approaches, using our own implementation in Matlab, to the standard linear SVM available in PRoNTo [6].

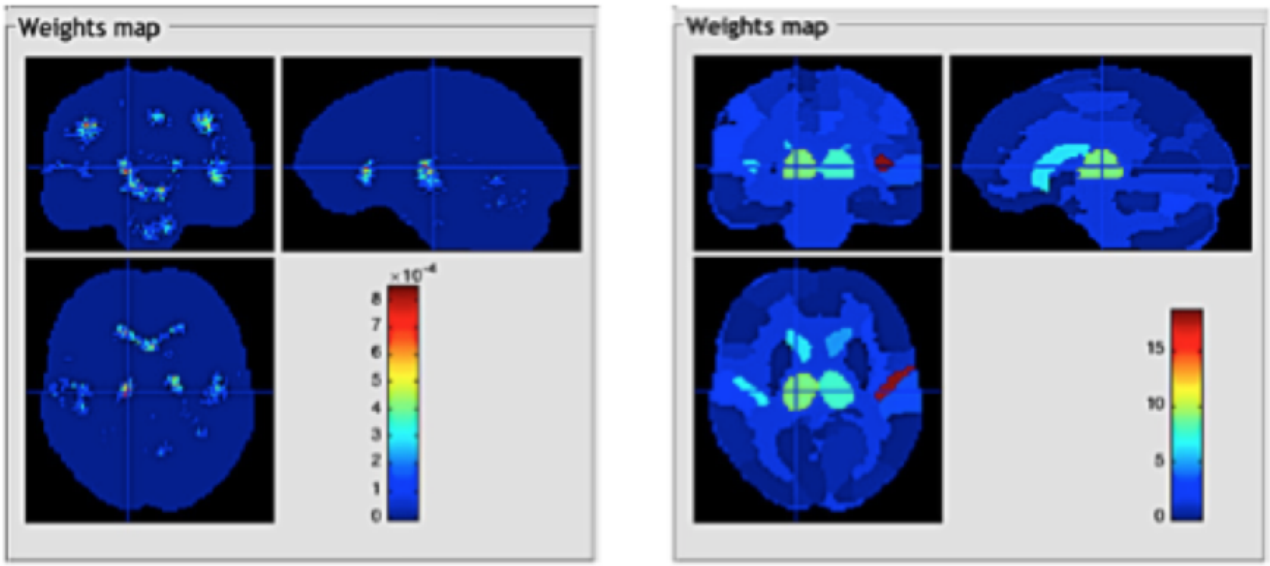
Results:

Methods are directly fitted from voxel-based feature vectors and are tested on the IXI dataset, composed of structural MRI from aged and young individuals [5]. We work in particular with scalar momentums as it was suggested to perform well in [6] to discriminate young and old people. A 5-fold cross validation (CV) procedure is used to assess all methods.
SVM, with or without parameter optimization (nested 5-fold CV with $C = 10^{[-3:1:3]}$), reaches a global accuracy of 98.51%. ST, RF and ET were evaluated with default parameters (M = 100 fully grown trees and $K = \sqrt{N}$ randomly selected attributes over N). For LB, the learning rate β was optimized with an additional nested 5-fold CV with $\beta = 5\times10^{[-4:1:-1]}$. We fixed the number of iterations at a sufficiently high value but we stopped the learning phase once the absolute difference between two iterations is below 10^{-6} to avoid over fitting.

Results are summarized in Table 1. ST is clearly inferior to ensembles as expected. The accuracies of randomized ensembles methods, RF and ET, do not exceed 96%. LB obtains a better accuracy, with LB^2 overtaking SVM performance with an error rate of 0.7435 %. LB also provides sparse models easily interpretable. Indeed, each LB stump involves a selection of the best voxel to split the node. The contribution of each voxel to the classification can then be measured by the number of times each voxel has been chosen over the iterations. Mimicking the procedure in [10], we used these counts to reconstruct a weight map and a weight map by region (as defined by the AAL atlas) for comparison with the same maps constructed from SVM (see Fig. 1&2). The maps are visually similar, despite the LB solution being much sparser than that of SVM, as confirmed by Table 2.

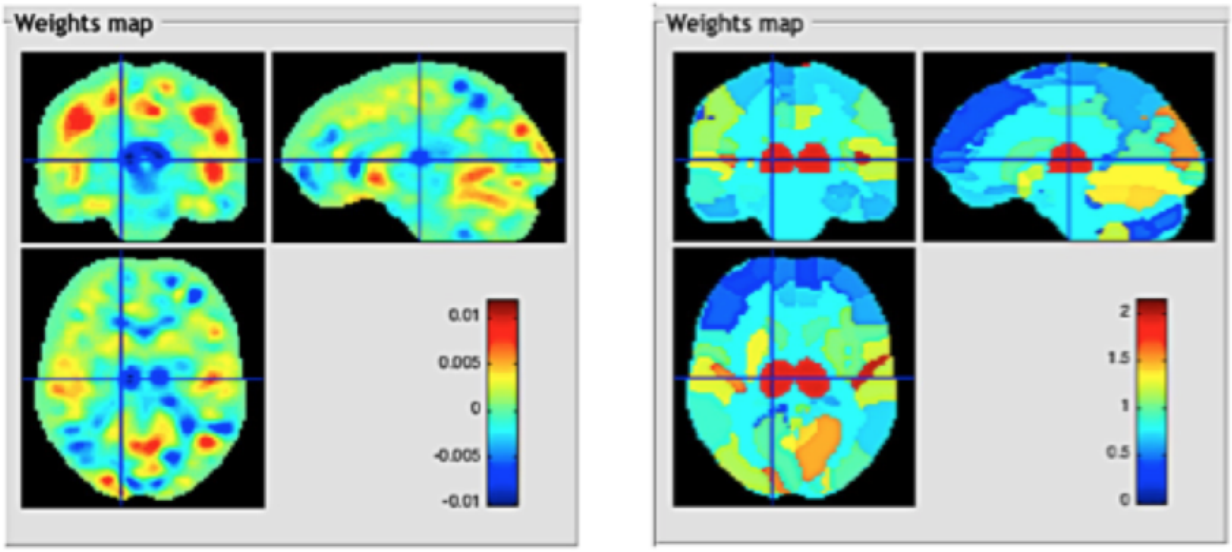
Method	Error rate (%)	Erroneous predictions in 269 samples
ST	15.24	41
RF	4.04 ± 0.17	≈ 11
ET	4.12 ± 0.12	≈ 11
LB with stumps, LB^1	2.23	6
LB with ensemble of trees, LB^2	0.74	2
SVM	1.49	4

Table 1 - Error rates for the 6 methods, obtained by 5-fold CV. For methods including randomization in the tree building procedure, we computed ten times each experience to obtain mean and standard deviation. Despite randomization in forest building procedure, the error rate is stable across folds for LB^2.



(a) Weights map. (b) Weights map by region.

Figure 1 - Weight map of voxel contributions (a) and summarized by region (b) providing information about the most predictive brain areas, for LB with forest of trees (LB^2) with a learning rate of 0.0005. A 5 fold CV for this method leads to 99.26% of accuracy.



(a) Weights map. (b) Weights map by region.

Figure 2 - Weight map of voxel contributions (a) and summarized by region (b) providing information about the most predictive brain areas, for SVM. A 5 fold CV for this method leads to 98.51% of accuracy.

LB^2	SVM
'Heschl_R'	'Vermis_6'
'Thalamus_L'	'Heschl_R'
'Thalamus_R'	'Thalamus_L'
'Caudate_L'	'Vermis_7'
'Heschl_L'	'Thalamus_R'
'Caudate_R'	'Paracentral_Lobule_R'
'Temporal_Sup_R'	'Vermis_8'
'Postcentral_L'	'Heschl_L'
'Cingulum_Mid_R'	'Occipital_Sup_L'
'Cerebelum_3_R'	'Calcarine_R'

Table 2 - Ranking of the first ten most contributing regions of AAL brain atlas selected by LB^2 and SVM respectively. Four regions are common in both top 10 lists.

Conclusions:

We show that tree methods can achieve competitive accuracy and provide interpretable models for the analysis of neuroimaging data. Although these results need to be confirmed on other datasets, we believe that tree methods are a promising alternative to linear SVM in this area. We also illustrate the added value of using randomized tree ensembles instead of ST within LB, which reveals to be even better than SVM.

Imaging Methods:

Anatomical MRI

Lifespan Development:

Aging

Modeling and Analysis Methods:

Classification and Predictive Modeling [1]
Methods Development [2]
Multivariate modeling

Keywords:

Aging
Machine Learning
Multivariate
STRUCTURAL MRI

[1][2]Indicates the priority used for review

Would you accept an oral presentation if your abstract is selected for an oral session?

Yes

I would be willing to discuss my abstract with members of the press should my abstract be marked newsworthy:

Yes

Please indicate below if your study was a "resting state" or "task-activation" study.

Other

Healthy subjects only or patients (note that patient studies may also involve healthy subjects):

Healthy subjects

Internal Review Board (IRB) or Animal Use and Care Committee (AUCC) Approval. Please indicate approval below. Please note: Failure to have IRB or AUCC approval, if applicable will lead to automatic rejection of abstract.

Yes, I have IRB or AUCC approval

Please indicate which methods were used in your research:

Structural MRI

For human MRI, what field strength scanner do you use?

3.0T

Which processing packages did you use for your study?

SPM
Other, Please list - PRoNTo

Provide references in author date format

[1] Pereira, F., et al. (2009). Machine learning classifiers and fMRI: a tutorial overview. Neuroimage, 45(1), S199-S209.
[2] Lemm, S., et al. (2011). Introduction to machine learning for brain imaging. Neuroimage, 56(2), 387-399.
[3] Mourão-Miranda, J., et al. (2005). Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. NeuroImage, 28(4), 980-995.
[4] Klöppel, S., et al. (2008). Automatic classification of MR scans in Alzheimer's disease. Brain, 131(3), 681-689.
[5] IXI - Information eXtraction from Images, funded by EPSRC GR/S21533/02, http://www.brain-development.org/
[6] Schrouff, J., et al. (2013). PRoNTo: pattern recognition for neuroimaging toolbox. Neuroinformatics, 11(3), 319-337.
[7] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
[8] Geurts, P., et al. (2006). Extremely randomized trees. Machine learning, 63(1), 3-42.
[9] Friedman, J., et al. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). The annals of statistics, 28(2), 337-407.
[10] Schrouff, J., et al. (2013, June). Localizing and comparing weight maps generated from linear kernel machine learning models. In Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on (pp. 124-127). IEEE.