Accuracy and interpretability, tree-based machine learning approaches.

3791

M. Wehenkel[1,2], P. Geurts[1], C. Phillips[1,2,3]

[1]Department of Electrical Engineering and Computer Science, [2]GIGA *In Silico* Medicine, [3]GIGA CRC *In Vivo* imaging, University of Liège, Belgium

Université de Liège

Introduction

Main aims of **pattern recognition techniques** for neuroimaging:

- development of **accurate diagnosis** systems;
- **identification of brain regions** related to the disease.

Kernel methods (e.g. SVM, MKL) [1,2] commonly used:

- Good accuracy with linear kernels;
- Good interpretability through feature weight maps [3].

Tree approaches not really popular in neuroimaging **but**:

- State-of-the-art accuracy on many problems with minimal tuning;
- Results interpretable through variable importance scores.

<u>Aim:</u> to study **tree methods** and show their **good behavior** against those of traditional methods such as SVM and MKL.

Methods

<u>Data :</u>

Methods are tested on **two datasets** :

- IXI [4]:
  - ➢ Structural MRI;
  - ➢ 170 aged vs. 99 young individuals;
  - ➢ We work in particular with scalar momentums obtained with SPM8, like in [5].
- OASIS [6]:
  - ➢ Structural MRI;
  - ➢ 50 demented vs. 50 non-demented old subjects;
  - ➢ Age and gender matched;
  - ➢ Preprocessing with SPM8.

<u>Machine learning methods :</u>

- Linear support vector machines (SVM);
- Multiple kernel learning (MKL);
- Single regression tree (ST);
- Random forests [7] (RF);
- Extremely randomized trees [8] (ET);
- Logitboost [9] with ST (LB_1) and with ET (LB_2) .

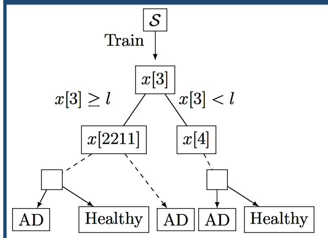
<u>Assessment :</u>

- Cross-validation (CV);
- 5 folds for IXI & 10 folds for OASIS;
- Nested CV for parameter optimization of SVM, MKL & LB.

<u>Weight map and weight map per region (MpR) built from:</u>

- Weight vector for SVM;
- Feature importance scores for ST, RF & ET;
- Number of times a voxel is choosen to split a node for LB;
- Aggregation of weights with AAL atlas for MpR.

Decision tree & Ensemble methods



Example of a decision tree classifying healthy vs. AD subjects from the voxel values of MRI images.

<u>Tree Ensemble Methods :</u>

- Combine the prediction of several trees;
- Trees grown either independently (as RF or ET) or sequentially (Boosting);
- Improve the bias-variance trade-off of single trees.

Results

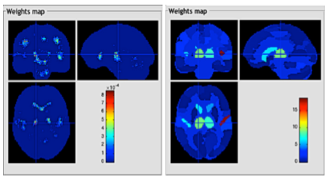
<u>Competitive accuracy :</u>

Method	IXI error rate (%)	OASIS error rate (%)
SVM	1.49	33.00
MKL	3.72	45.00
Single tree	15.24	44.00
Random forests	1.71 ± 0.26	35.50 ± 0.97
Extremely randomized trees	1.86 ± 0.30	33.50 ± 1.51
Logitboost LB_1	2.23	37.00
Logitboost LB_2	0.78 ± 0.12	33.60 ± 0.52

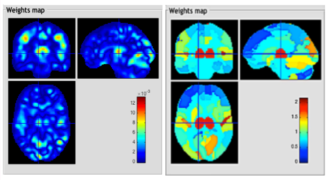
Table 1 : Summary of method performance for both datasets.

<u>Good interpretability:</u>

- Similar important regions;
- Sparser models with tree-based methods.

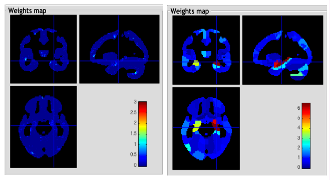


(a) Weight map for LB_2. (b) Weight map per region for LB_2.

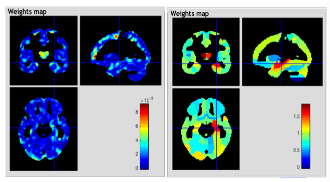


(c) Weight map for SVM. (d) Weight map per region for SVM.

Figure 1 : IXI dataset.



(a) Weight map for LB_2. (b) Weight map per region for LB_2.



(c) Weight map for SVM. (d) Weight map per region for SVM.

Figure 2 : OASIS dataset.

Rk	Method			
	SVM	MKL	RF	LB_2
1	Vermis6	TemporalMidL	**HeschlR**	HeschlR
2	**HeschlR**	**PostcentralL**	**CaudateL**	ThalamusL
3	**ThalamusL**	LingualL	**CaudateR**	ThalamusR
4	Vermis7	OccipitalMidL	**ThalamusL**	CaudateL
5	**ThalamusR**	**TemporalSupR**	**HeschlL**	HeschlL
6	ParacentralLobuleR	ThalamusL	**ThalamusR**	CaudateR
7	Vermis8	FrontalMidR	**PostcentralL**	TemporalSupR
8	**HeschlL**	PostcentralR	**TemporalSupR**	PostcentralL
9	OccipitalSupL	Cerrebelum6L	InsulaR	CingulumMidR
10	CalcarineR	TemporalMidR	**Cerebelum3R**	Cerebelum3R

Table 2 : Ranking of the first ten most contributing regions of AAL brain atlas selected by SVM, MKL, RF and LB_2 respectively for IXI dataset. We highlighted in bold regions in common with those of LB_2.

Conclusion

We show that **tree based methods** can achieve **competitive accuracy** and provide **interpretable models** for the analysis of neuroimaging data and thus we believe that tree methods are a **promising alternative** to traditional methods in this area.

<u>REFERENCES</u>

[1] Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K. R. (2011). Introduction to machine learning for brain imaging. Neuroimage, 56(2), 387-399. **[2]** Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., & Frackowiak, R. S. (2008). Automatic classification of MR scans in Alzheimer's disease. Brain, 131(3), 681-689. **[3]** Schrouff, J., Cremers, J., Garraux, G., Baldassarre, L., Mourão-Miranda, J., & Phillips, C. (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. **[4]** IXI - Information eXtraction from Images, funded by EPSRC GR/S21533/02. **[5]** Schrouff, J., Rosa, M. J., Rondina, J. M., Marquand, A. F., Chu, C., Ashburner, J., ... & Mourão-Miranda, J. (2013). PRoNTo: pattern recognition for neuroimaging toolbox. Neuroinformatics, 11(3), 319-337. **[6]** Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience, 19(9), 1498-1507. **[7]** Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32. **[8]** Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63(1), 3-42. **[9]** Friedman, J., Hastie, T., & Tibshirani, R. (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.

ACKNOWLEDGEMENTS & SPONSORS: Belgian National Fund for Scientific Research (F.R.S-FNRS) – University of Liège (ULg) – Cyclotron Research Centre (CRC)

MONTEFIORE INSTITUTE | http://www.montefiore.ulg.ac.be | Marie Wehenkel | m.wehenkel@ulg.ac.be