

The **next generation** GBCA  
from Guerbet is here

Explore new possibilities >

Guerbet | 

© Guerbet 2024 GUOB220151-A

# AJNR

## The "Peeking" Effect in Supervised Feature Selection on Diffusion Tensor Imaging Data

S. Diciotti, S. Ciulli, M. Mascalchi, M. Giannelli and N. Toschi

*AJNR Am J Neuroradiol* 2013, 34 (9) E107

doi: <https://doi.org/10.3174/ajnr.A3685>

<http://www.ajnr.org/content/34/9/E107>

This information is current as  
of September 18, 2024.

## The “Peeking” Effect in Supervised Feature Selection on Diffusion Tensor Imaging Data

We read with great interest the article by Haller et al<sup>1</sup> in the February 2013 issue of the *American Journal of Neuroradiology*. The authors used whole-brain diffusion tensor imaging-derived fractional anisotropy (FA) data, skeletonized through use of the standard tract-based spatial statistics (TBSS) pipeline, to achieve the following: 1) report significant group differences in FA among mild cognitive impairment (MCI) subtypes, and 2) perform individual classification of MCI subtypes by using a supervised feature selection procedure combined with a support vector machine (SVM) classifier. The study reports extremely high classification performances (100% sensitivity and 94%–100% specificity), which the authors describe as perhaps “too optimistic” and partially ascribe to “some degree of overfitting,” possibly also due to the use of feature selection.

The above-mentioned study presents a questionable use of supervised feature selection, which was performed on the entire dataset (ie, on both training and test data) instead of only on the training set of each partition generated during the cross-validation procedure. It is well-known that using test set labels to perform inference on a feature subset during the learning process can cause an overestimation of the generalization capabilities of the classifier (sometimes called the “peeking” effect) and that this effect is particularly severe when a large number of features are removed (like in this whole-brain DTI study, in which approximately 150,000 features were reduced to 1000).<sup>2,3</sup> In other words, training the classifier with the same instances (ie, data “points”) used for feature selection corresponds to providing it with “hints” about the solution of the classification problem, and Haller et al<sup>1</sup> recognized this circumstance as a “limitation” of their study. However, this methodologic mistake<sup>3</sup> (which unfortunately appears in several recent studies in the MR imaging literature) does not constitute a mere theoretic concern but rather can have important consequences on the final results.<sup>3</sup>

To better clarify and exemplify our point, we have analyzed DTI data in a patient cohort presented in a previous MCI-Alzheimer disease (AD) classification study.<sup>4</sup> Specifically, we attempted to discriminate between 30 patients with amnesic MCI and 21 with mild AD by using the processing pipeline (a Relief-F feature selection of the top 1000 features followed by an SVM classifier and 10 repetitions of a 10-fold cross-validation) and the same type of data (skeletonized

whole-brain FA data) used by Haller et al.<sup>1</sup> We repeated the analysis by using either incorrect cross-validation (ie, feature selection on the entire dataset followed by classification in cross-validation, as carried out by Haller et al<sup>1</sup>) or correct cross-validation (feature selection within each training set of the cross-validation).

In the former analysis, patients with mild AD were classified with 80.0% sensitivity and 96.7% specificity, while in the latter analysis, results dropped to 45.3% sensitivity and 67.3% specificity. These data demonstrate the remarkable amount of possible overestimation of the generalization capabilities due to the “peeking” effect in a cross-validation study which uses whole-brain TBSS data, and we speculate that the sensitivity/specificity values reported by Haller et al<sup>1</sup> would be substantially lowered if an orthodox feature-selection procedure was applied to their data.

In conclusion, given the relevance and potential of MCI subtype discrimination through MR imaging feature extraction and selection, full consideration of the methodologic pitfalls of combining supervised feature selection procedures with SVM in whole-brain imaging data analysis is highly recommended.

### REFERENCES

- Haller S, Missonnier P, Herrmann FR, et al. **Individual classification of mild cognitive impairment subtypes by support vector machine analysis of white matter DTI.** *AJNR Am J Neuroradiol* 2013;34:283–91
- Pereira F, Mitchell T, Botvinick M. **Machine learning classifiers and fMRI: a tutorial overview.** *Neuroimage* 2009;45(1 suppl):S199–209
- Smialowski P, Frishman D, Kramer S. **Pitfalls of supervised feature selection.** *Bioinformatics* 2010;26:440–43
- Diciotti S, Ginestroni A, Bessi V, et al. **Identification of mild Alzheimer’s disease through automated classification of structural MRI features.** *Conf Proc IEEE Eng Med Biol Soc* 2012;2012:428–31

S. Diciotti

S. Ciulli

M. Mascalchi

Department of Clinical and Experimental Biomedical Sciences

University of Florence

Florence, Italy

M. Giannelli

Unit of Medical Physics

Pisa University Hospital

Azienda Ospedaliero-Universitaria Pisana

Pisa, Italy

N. Toschi

Medical Physics Section, Department of Biomedicine and Prevention

Faculty of Medicine University of Rome Tor Vergata

Rome, Italy