INCREASING THE CREDIBILITY OF MR SPECTROSCOPY-BASED AUTOMATIC BRAIN TUMOR CLASSIFICATION SYSTEMS

Martin Berger[1,2], Klaus Sembritzki[1], Joachim Hornegger[1], Christina Bauer[3]

[1] *Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany*
[2] *Graduate School Heterogeneous Image Systems, Friedrich-Alexander University Erlangen-Nuremberg*
[3] *Siemens Healthcare, MR Division, Siemens AG, Erlangen, Germany*

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

SIEMENS

Introduction

Many approaches were introduced for automatic classification of brain tumors using MR spectroscopy and pattern recognition [1,2].

Little effort has been paid regarding the interpretability of such classification systems, which might be one reason that none of these approaches found its way into a clinical application.

In this work, we introduce a novel reliability measure for a classifier's decision and propose a visualization method for a linear decision boundary directly in the spectral domain.

Materials and Methods

Classifier comparison to confirm previous results [1]

- Four classifiers with and without feature extraction (PCA): LDA, ν-SVM, RBF-SVM, Random Forests (RF)
- Evaluation: 10x10-fold CV / Measure: Balanced error rate

Visualization of linear classifiers in the spectral domain

Decision Boundary: $C(\boldsymbol{x}) = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x} + \alpha_0$

- $\boldsymbol{\alpha}$ encodes regions that are important for the classifier and can be visualized directly in the spectral domain
- Also works with prior feature extraction or feature selection

New reliability measure for newly classified samples

Training

Cross Validation → Labeled dataset: $S = \{(\boldsymbol{x}_1, y_1, \cdots, (\boldsymbol{x}_N, y_N)\}$

→ Evaluated dataset: $S^e = \{(\boldsymbol{x}_1, y_1, \hat{y}_1, \hat{q}_1), \cdots, (\boldsymbol{x}_N, y_N, \hat{y}_N, \hat{q}_N)\}$

quality measure, e.g. posterior probability

Classification

→ Classification of new sample $\tilde{\boldsymbol{x}}$ yields a quality measure $\tilde{q}$

- Extract all training samples that had a lower quality measure

$$S^e_{low} = \{(\boldsymbol{x}, y, \hat{y}, \hat{q}) \in S^e \mid \hat{q} \leq \tilde{q}\}$$

- ***Correct*** ($\hat{y} = y$) and ***incorrect*** outcome binomial distributed
- Binomial confidence interval based on S^e_{low} yields a lower bound for the accuracy of this particular classification

INTERPRET database [3]:

Short-Echo Single-Voxel Spectra	215
Glioblastoma & metastases (AGG)	122
Meningioma (MEN)	60
Low-grade glioma (LGG)	33

Pre-processing steps (adopted from [1])

a) Klose algorithm
b) Water removal
c) Apodization
d) Baseline correction
e) L2-normalization
f) Crop [0.5, 4.3] ppm
g) Resize to 256

Results and Discussion

- Spectral representation clearly depicts metabolites that are important for a classifier's decision (cf. Fig.1)
- We propose a reliability measure that yields the lower bound of the accuracy for a particular classification
- Reliability measure is more stable than the posterior probability and in good agreement with the real accuracy (cf. Fig.3)
- Combination of reliability measure and spectral weight vector gives further insight into a classifiers decision (cf. Fig.2)

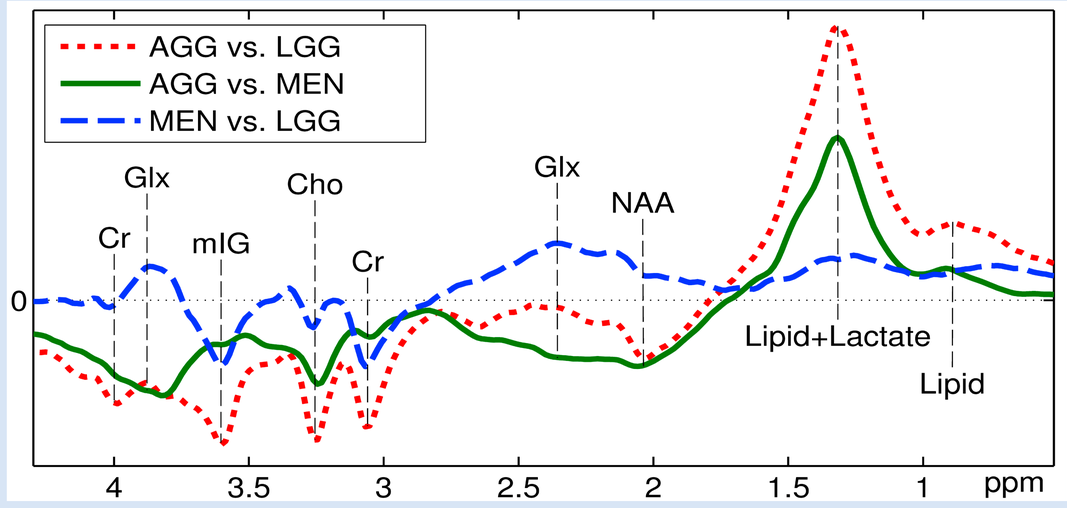

Figure 1: Spectral weights of the LDA for three different binary decision problems. The metabolites are depicted by: creatine (Cr), glutamate+glutamine (Glx), myo-inositol+glycine (mIG), choline (Cho), N-acetylaspertate (NAA), lactate and lipids.

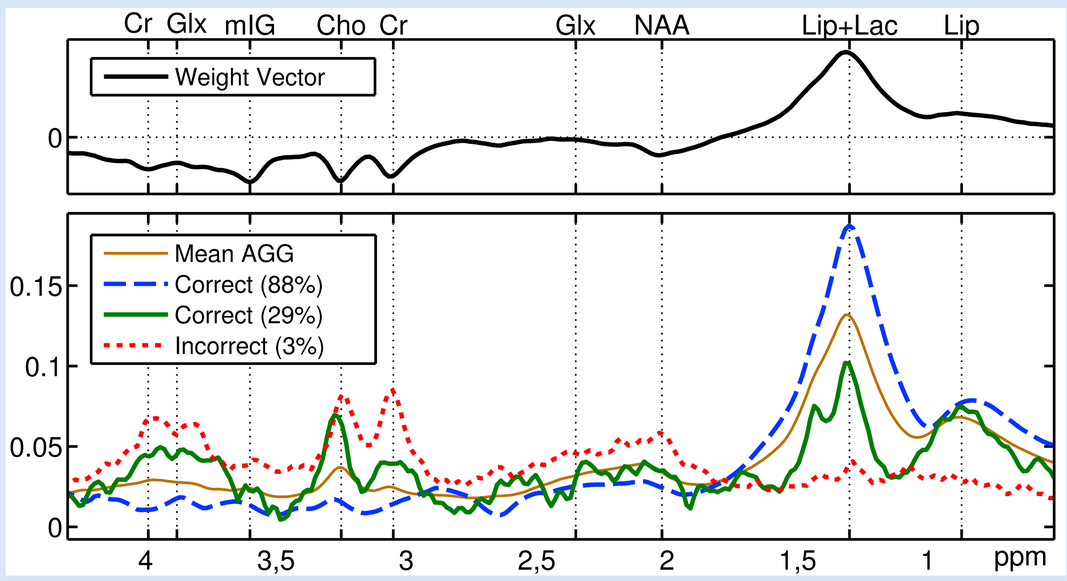

Figure 2: Top: Spectral weight vector of an LDA classifier discriminating between AGG and LGG. Bottom: Three exemplary classifications and the mean spectrum of all AGG cases. The reliability measure is given in the legend.

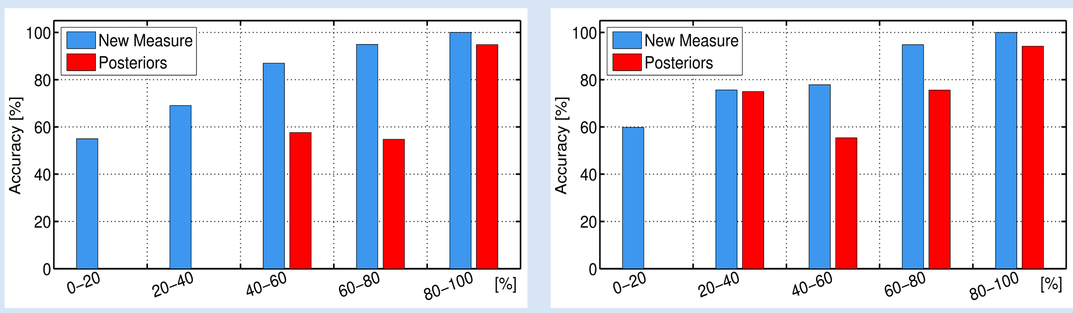

Figure 3: Accuracies after CV when binning the results according to the new reliability measure based on a 95% confidence interval (blue bars) and the posteriors directly (red bars). Results for LDA (left) and ν-SVM (right).

Conclusion

- We introduce a novel reliability measure and propose a classifier visualization directly in the spectral domain
- Combination of reliability measure and spectral weights could help to further improve clinical acceptance of automatic classification systems
- For future work we plan a more profound evaluation of the reliability measure

References

[1] Garcia-Gomez et al., "Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy.," Magma, vol. 22, no. 1, pp. 5–18, Feb. 2009.

[2] Menze et al., "Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors.," NMR in biomedicine, vol. 19, no. 5, pp.599–609, Aug. 2006.

[3] Julia-Sape et al., "A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients.," Magma, vol. 19, no. 1, pp. 22–33, Mar. 2006.

Contact
✉ martin.berger@cs.fau.de
http://www5.cs.fau.de/~berger

The authors gratefully acknowledge funding of the Research Training Group 1773 "Heterogeneous Image Systems", funded by the German Research Foundation (DFG).