Automated Classification and Grading of Tumors in Mass Spectrometric Images using Postprocessed Random Forests

Michael Hanselmann[1], Ullrich Köthe[1], Marc Kirchner[1], Bernhard Y. Renard[1], Erika R. Amstalden[2], Kristine Glunde[3], Ron M. A. Heeren[2], Fred A. Hamprecht[1]

[1] Heidelberg Collaboratory for Image Processing (HCI), Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg, Heidelberg, Germany
[2] FOM AMOLF, FOM-Institute for Atomic and Molecular Physics, Amsterdam, The Netherlands
[3] Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, USA

Introduction

Imaging Mass Spectrometry (IMS) allows a detailed analysis of the spatial distribution of biomolecules in tissue and their relation to histological properties. In clinical applications, this "digital staining" constitutes a powerful complement to chemical staining since it is not limited to one tissue type at a time. Only automated approaches can handle the resulting data sets and offer enhanced interpretability.
If spatially resolved labels are available, supervised classifiers can be used for automated distinction of different tissue types. We combine the Random Forest classifier with Markov-Random-Fields and vector-valued median-filtering to achieve high robustness against spatial noise.

Data and Data Processing

Experimental data was acquired from orthotopic human breast cancer xenografts grown in mice from MCF-7, a weakly metastatic and estrogen-sensitive breast cancer cell line. For each tumor slice, a Hematoxylin-Eosin (H&E) stained parallel slice is available, which was employed for labeling training data. 7 MS images were used:

- 6 parallel tissue slices of the same tumor (S3, S4, S5, S7, S9, S11)
- 1 slice (T1) of a 2nd tumor (same cell line, grown in a genetically identical mouse)

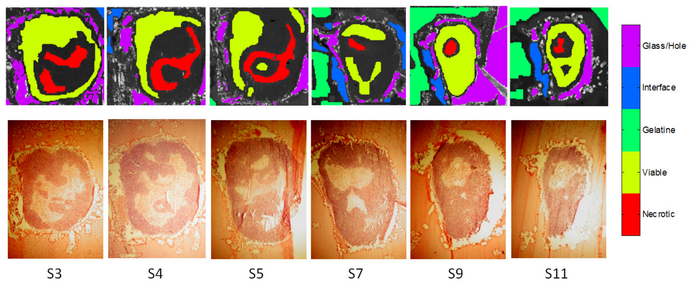


Figure 1: Label maps for the S-Slices with corresponding H&E stained parallel slices.

After calibration all spectra were baseline-corrected and normalized by their total ion count. Features were extracted by a threshold-based peak-picker.

References

[1] Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[2] Ulintz et al. Improved classification of mass spectrometry database search results using newer machine learning approaches. *Molecular & Cellular Proteomics*, 5:497–509, 2006.

[3] Geman and Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Patt. Anal. & Mach. Intell.*, 6(6):721–741, 1984.

[4] Welk et al. Median filtering of tensor-valued images. *Patt. Rec.*, 2781:1724, 2003.

Results

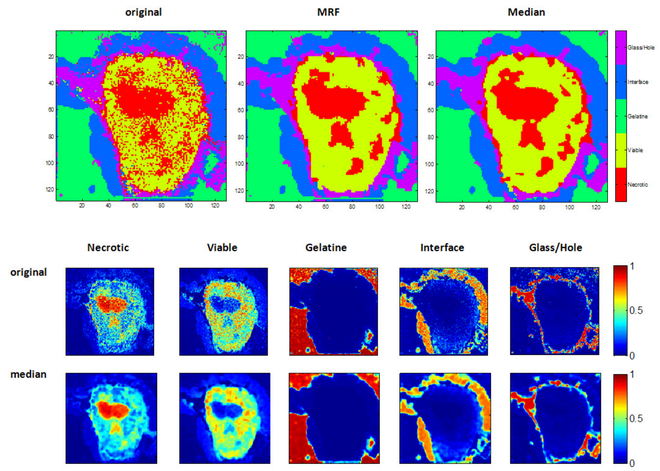


Figure 2: Results on slice S7 after training on S3, S4, S5, S9 and S11.

		tissue class					
	meas.	necrotic	viable	gelatine	interf.	glass	mean
S3	SE	92.8	82.7	-	89.7	89.5	**88.7**
	PPV	57.6	96.8	-	41.9	98.6	**73.7**
S4	SE	64.7	98.3	-	89.9	97.1	**87.4**
	PPV	95.6	85.7	-	98.3	86.3	**91.4**
S5	SE	94.7	91.3	86.7	-	98.1	**92.7**
	PPV	86.7	96.8	100	-	99.4	**95.8**
S7	SE	99.4	75.4	99.1	99.6	89.1	**92.5**
	PPV	30.4	99.7	99.5	94.8	99.5	**84.8**
S9	SE	81.0	96.2	84.0	96.4	97.2	**90.1**
	PPV	54.0	97.3	99.4	60.0	97.0	**81.5**
S11	SE	96.4	90.4	87.4	99.1	91.0	**92.8**
	PPV	40.0	98.7	99.4	68.3	99.7	**81.2**

Table 1: The results of the leave-one-out cross validation experiment.

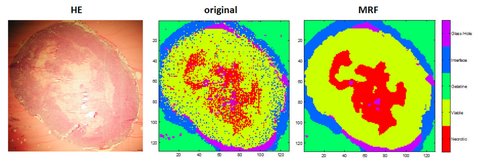


Figure 3: Results for T1 after training with samples from all S-slices.

		tissue class					
	meas.	necrotic	viable	gelatine	interf.	glass	mean
T1	SE	71.9	91.6	99.9	90.3	99.6	**90.4**
	PPV	94.4	84.0	97.6	93.4	93.4	**92.5**

Table 2: Results for T1 after training with samples from all S-slices.

A Random Forest was trained on 5 of 6 S-slices and the remaining slice was used for testing (leave-one-out cross-validation). Classifications with high sensitivity (SE) and positive predictive values (PPV) were obtained (table 1), where

$$SE = \frac{true\ pos.}{true\ pos.+false\ neg.}$$

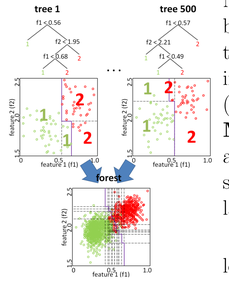
$$PPV = \frac{true\ pos.}{true\ pos.+false\ pos.}.$$

In a second experiment, we trained a Random Forest on the 6 S-slices of the first tumor and classified the T1-slice of the second tumor (table 2).
In both scenarios, post-hoc smoothing of the classification maps improved the SE and PPV by $\approx 3-5\%$ (regarding the gold standard labels).

Methods

Random Forests. The Random Forest classifier [1] is a supervised, decision-tree based ensemble machine learning method (see figure 4) with favorable properties:

- **High prediction accuracy** – Random Forests perform equally good as Support Vector Machines (SVMs) and other current classifiers [2]
- **Robustness to overfitting** even with large numbers of input variables
- **Robustness to parameter settings**
- **Class probabilities available** in addition to crisp classification results



Noisy input data leads to noisy predictions as indicated by significant label variation even in small spatial contexts. We remove these outliers via spatial regularization in a post-hoc smoothing step with Markov-Random-Fields (MRF) [3] or vector-valued median-filtering (VVM) [4].
MRF. The single site potential at pixel i encourages the agreement of each label z_i with the local classification result while the pair potentials require consistency of each label with the labels of surrounding pixels $\pi(i)$:

$$\log(p(Z|S)) = \underbrace{\sum_{i=1}^{N} \log(\varphi(z_i))}_{single\ site\ pot.} + \lambda \underbrace{\sum_{i}^{N} \sum_{j\in\pi(i)} \log(\vartheta(z_i, z_j))}_{pair\ potential} + \text{c.} \tag{1}$$

VVM. Given weights w_k, for each pixel the weighted vector-valued median μ of the set $\tilde{S} = \{\tilde{\mathbf{x}}_1, ..., \tilde{\mathbf{x}}_K\}$ of K neighboring M-dimensional vectors is calculated from

$$\mu(\tilde{S}) = \underset{a\in\mathbb{R}^M}{\operatorname{argmin}} \left(\sum_{k=1}^{K} w_k \left\| x_k - a \right\|_2 \right). \tag{2}$$

Conclusions

- Digital staining constitutes a **powerful complement** to chemical staining
- Random Forests are well suited for **automated classification** of IMS data
- Necrotic, viable, gelatine, interface and glass areas were classified with **high sensitivity** ($\approx 90\%$) **and positive predictive values** ($\approx 85\%$)
- Post-hoc smoothing (if appropriate) makes classification **more robust**
- Future work: additional experiments on data featuring genetic variation

Acknowledgements

We gratefully acknowledge financial support by the DFG (grant no. HA4364/2-1) (M.H, B.Y.R., F.A.H), the Robert Bosch GmbH (F.A.H.), the Netherlands BSIK program 'Virtual Laboratory for e-science' (R.M.A.H., E.R.A.), and the National Institutes of Health of the USA (grant no. NIH R01 CA134695) (K.G., R.M.A.H.). The experimental data were acquired as part of the research program of the "Stichting voor Fundamenteel Onderzoek der Materie", which is supported by the "Nederlandse organisatie voor Wetenschappelijk Onderzoek".