

Computational TMA Analysis and Cell Nucleus Classification of Renal Cell Carcinoma

Peter J. Schüffler, Thomas J. Fuchs, Cheng Soon Ong, Volker Roth,
Joachim M. Buhmann

Department of Computer Science, ETH Zurich, Switzerland

Abstract. We consider an automated processing pipeline for tissue micro array analysis (TMA) of renal cell carcinoma. It consists of several consecutive tasks, which can be mapped to machine learning challenges. We investigate three of these tasks, namely nuclei segmentation, nuclei classification and staining estimation. We argue for a holistic view of the processing pipeline, as it is not obvious whether performance improvements at individual steps improve overall accuracy. The experimental results show that classification accuracy, which is comparable to trained human experts, can be achieved by using support vector machines (SVM) with appropriate kernels. Furthermore, we provide evidence that the shape of cell nuclei increases the classification performance. Most importantly, these improvements in classification accuracy result in corresponding improvements for the medically relevant estimation of immunohistochemical staining.

1 Introduction

Cancer tissue analysis consists of several consecutive estimation and classification steps which are currently highly labour intensive. The tissue microarray (TMA) technology promises to significantly accelerate studies seeking for associations between molecular changes and clinical endpoints [10]. In this technology, $0.6mm$ tissue cylinders are extracted from primary tumor blocks of hundreds of different patients and these cylinders are subsequently embedded into a recipient tissue block. Sections from such array blocks can then be used for simultaneous in situ analysis of hundreds or thousands of primary tumors on DNA, RNA, and protein level. Although the production of tissue microarrays is an almost routine task for most laboratories, the evaluation of stained tissue microarray slides remains tedious, time consuming and prone to error. The high speed of arraying, the lack of a significant damage to donor blocks, and the regular arrangement of arrayed specimens substantially facilitates automated analysis.

This paper investigates an automated system to model such a workflow for renal cell carcinoma (RCC). Current image analysis software requires extensive user interaction to properly identify cell populations, to select regions of interest for scoring, to optimize analysis parameters and to organize the resulting raw data. Due to these drawbacks in current software, pathologists typically collect tissue microarray data by manually assigning a composite staining score for

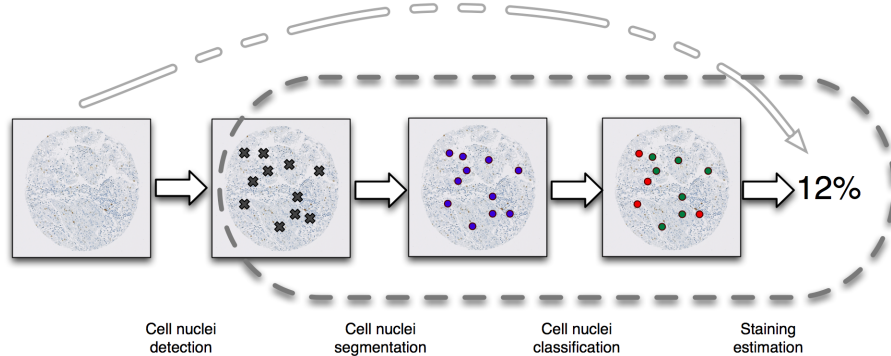


Fig. 1. TMA processing pipeline, consisting of the following stages: (1) Detection of cell nuclei within a high resolution TMA image, (2) nuclei segmentation, (3) nuclei classification into malignant and benign, and (4) calculation of the percentage of tumor cells and protein expressing tumor cells. The pathologists directly assign a staining score (pale dashed arrow). We consider the subset of tasks indicated by the boxed region.

each spot - often during multiple microscopy sessions over a period of days. Such manual scoring can result in serious inconsistencies between data collected during different microscopy sessions. Manual scoring also introduces a significant bottleneck that limits the use of tissue microarrays in high-throughput analysis.

The manual rating and assessment of TMAs under the microscope by pathologists is quite error prone due to the subjective perception of humans. In fact, one pathologist might differ in the detection and classification of the same cell nuclei within the same image in a second annotation round [7]. Reasons for such a discrepancy may relate to the partial destruction of a cell nucleus during the preparation of a TMA; the slow and often gradual transformation of a healthy cell to a tumor cell, and the subjective experiences of the pathologist. Therefore, decisions for grading and/or cancer therapy might be inconsistent among pathologists.

In this paper, we follow the workflow in [7] which consists of cell nuclei detection, segmentation and classification, followed by estimation of the proportion of nuclei stained by the antibody under investigation (Figure 1). In particular, we investigate the effect of different image features and kernels on the accuracy of classifying whether a cell nucleus is cancerous or not (Section 3.1). By considering the different types of features, we show in Section 3.2 the importance of using shape features. The effect of classification accuracy on staining estimation is shown in Section 3.3. Our contribution in this paper includes a comprehensive investigation of various image features and associated kernels on the performance of a support vector machine classifier for cancerous cells. Since the classifier is only a part of a larger pipeline, we show that the observed improvements in

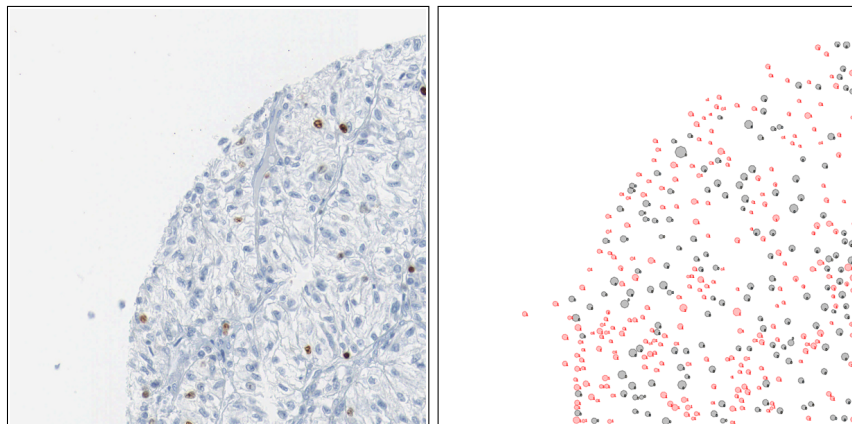


Fig. 2. Left: Top left quadrant of a TMA spot from a RCC patient. **Right:** A pathologist labeled all cell nuclei and classified them into malignant (black) and benign (red).

classification results in improvements in staining estimation. The estimation of staining of various antibodies forms the basis for localized biomarker assessment.

1.1 Tissue Micro Arrays

Tissue micro arrays (TMA) are an important preparation technique for the quantitative analysis of various types of cancer, like clear cell renal carcinoma (ccRCC). The TMA glass plates carry small round tissue spots of prospective cancerous tissue samples with a thickness of one cell layer for each spot. Our TMAs are immunohistochemically stained for the proliferation indicator protein MIB-1. Positive cell nuclei indicate cell division and appear brown in the image. Additionally, eosin counter-staining is used for discovering the morphological structure of the tissue and the nuclei on the TMA. Hence, MIB-1 negative nuclei are visualized as blue spots in the image.

In this study, we used the top left quarter of eight tissue spots from eight patients. Therefore, each image shows a quarter of the whole spot, i.e. 100-200 cells per image (see Figure 2). The TMA slides were immunohistochemically stained with the MIB-1 (Ki-67) antigen and scanned on a Nanozoomer C9600 virtual slide light microscope scanner from HAMAMATSU Photonics K.K.. The magnification of 40x resulted in a per pixel resolution of $0.23\mu m$. Finally the spots of single patients were extracted as separate three channel color images of 3000 x 3000 pixels size.

1.2 Support Vector Machines and Kernels

Support vector machines (SVM, e.g. [11, 2]) are in widespread and highly successful use for bioinformatics tasks. SVMs exhibit very competitive classification

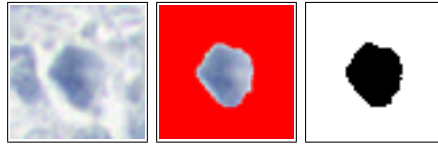


Fig. 3. Example of nucleus segmentation. **Left:** The original 80x80 pixel patch. **Middle:** The segmentation via graphcut. **Right:** The resulting shape of the nucleus.

performance, and they can conveniently be adapted to the problem at hand. This adaptation is achieved by designing appropriate kernel functions, which can be seen as problem-specific similarity functions between examples. The kernel function implicitly maps examples from their input space \mathcal{X} to a space \mathcal{H} of real-valued features (e.g. $\mathcal{H} = \mathbb{R}^d, d \in \mathbb{N} \cup \{\infty\}$) via an associated function $\Phi : \mathcal{X} \rightarrow \mathcal{H}$. The kernel function k provides an efficient method for implicitly computing dot products in the feature space \mathcal{H} via $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$.

The resulting optimization problem is convex and the global optimum can be found efficiently. Furthermore, there are many freely available software packages which can be used. One remaining question is the choice of features to use for a particular image x and how to compare them via the kernel $k(x_i, x_j)$. Our approach is described in detail in Section 2.4.

2 Methods

2.1 Nuclei extraction from labeled TMA images

TMA images of renal cell carcinoma were independently labeled by two pathologists [7]. From eight exhaustively labeled TMA images, we extracted 1633 patches of size 80x80 pixels centered at labeled cell nuclei (see Figure 3). For 1272 (78 %), the two pathologists agree on the label: 890 (70 %) benign and 382 (30 %) malignant nuclei.

We further labeled the nuclei according to their staining status. For each patch, a color histogram was calculated for a 30x30 pixel size center square. A higher mean of the red histogram than of the blue histogram indicated a stained nucleus.

2.2 Segmentation

Segmentation of cell nuclei was performed with graphcut [1, 4, 5, 9]. The gray intensities were used as unary potentials. A circular shape prior was used to prefer roundish objects (see Figure 3). To this end, the binary potentials were weighted based on their distance to the center. The contour of the segmented object was used to calculate several shape features as described in the following section.

2.3 Feature extraction

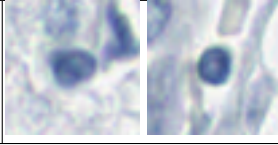
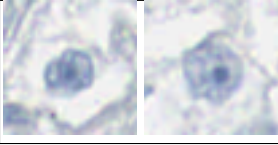
	Benign nucleus	Malignant nucleus
Shape	roundish	irregular
Membrane	regular	thick/thin irregular
Size	smaller	bigger
Nucleolus	none	dark spot in the nucleus
Texture	smooth	irregular
		

Table 1. Guidelines used by pathologists for renal nuclei classification. In addition to the depicted benign and malignant RCC nuclei the tissue consist of a large number of benign non-renal cells like lymphocytes and endothelial cells which were all labeled as non-cancerous.

One computationally beneficial aspect of RCC is the fact that the classification of cancerous cells can be performed locally (i.e. patch-wise). This advantage is absent for example in prostate cancer, where the morphology of whole glands is crucial for the classification. Pathologists use several intuitive guidelines to classify nuclei as described in Table 1. One aim of this study was to design features, which are able to capture these guidelines. The following list summarizes the extracted shape and histogram features:

- Histogram of **foreground intensity, FG** (nucleus, 32 bins)
- Histogram of **background intensity, BG** (surrounding tissue, 32 bins)
- **Freeman Chain Code, FCC:** FCC describes the nucleus' boundary as a string of numbers from 1 to 8, representing the direction of the boundary line at that point ([8]). The boundary is discretized by subsampling with grid size 8. For rotational invariance, the first difference of the FCC with minimum magnitude is used. The FCC is represented in a 8-bin histogram.
- **1D-signature, SIG:** Lines are considered from the object center to each boundary pixel. The angles between these lines form the signature of the shape ([8]). As feature, a 16-bin histogram of the signature is generated.
- **Pyramid histograms of oriented gradients, PHOG:** PHOGs are calculated over a level 3 pyramid on the gray-scaled patches ([3]).
- Shape descriptors derived from MATLAB's **regionprops** function (**PROP**) Area, BoundingBox(3:4), MajorAxisLength, MinorAxisLength, ConvexArea, Eccentricity, EquivDiameter, Solidity, Extent, Perimeter, MeanIntensity, MinIntensity, MaxIntensity;

Kernel	Definition
Linear	$u' * v$
Polynomial (degree $d=(3,5,7,10)$)	$((1/p)(u' * v))^d$
Gaussian	$e^{-(1/p) \sum_i (u_i - v_i)^2}$
Hellinger	$\sum_i \sqrt{u_i * v_i}$
Jensen Shanon	$\frac{-1}{\log 2} \sum_i u_i \log \frac{u_i}{u_i + v_i} + v_i \log \frac{v_i}{u_i + v_i}$
Total variation	$\sum_i \min u_i, v_i$
χ^2	$\sum_i \frac{u_i * v_i}{u_i + v_i}$
Distance	Definition
Euclidean	$\sqrt{\sum_i (u_i - v_i)^2}$
Intersection	$\min(\sum_i u_i, \sum_i v_i) * (1 - \frac{\sum_i \min(u_i, v_i)}{\min(\sum u, \sum v)})$
Bhattacharya	$-\log \sum_i \sqrt{u_i * v_i}$
χ^2	$\sum_i \frac{(u_i - v_i)^2}{u_i + v_i}$
Kullback Leibler	$\sum_i u_i * \log \frac{u_i}{v_i} + \sum_i v_i * \log \frac{v_i}{u_i}$
Earth Mover	$\sum_{i=1}^p \sum_{j=1}^i u_j - v_j $
ℓ_1	$\sum_i u_i - v_i $

Table 2. Commonly used kernels and distances for two scalar feature vectors u and v of the length p . For the histogram features all kernels and distances were employed, while for the PROP feature only the top most three kernels were used.

2.4 Kernel calculation

The feature vectors extracted from the patches were used to calculate a set of kernel matrices. All histograms are normalized. For the histogram features, ten kernel functions and eight distance measures for histograms have been investigated, whereas for the PROP features, only the linear, polynomial and Gaussian kernels were calculated (see Table 2).

Resulting dissimilarity matrices D were centered to have zero mean and checked for being positive semidefinite, to serve as kernel matrices K . Where needed, negative Eigenvalues were mirrored:

$$D_{centered} = -0.5 * Q * D * Q \text{ where } Q = \begin{pmatrix} 1 - \frac{1}{n} & & -\frac{1}{n} \\ & \ddots & \\ -\frac{1}{n} & & 1 - \frac{1}{n} \end{pmatrix}$$

$$K = V * |\Lambda| * V' \text{ where } V = \text{Eigenvectormatrix}(K)$$

2.5 SVM training

Training of the SVMs was preformed using the libSVM package for Matlab ([6]). To generate a more reliable gold standard, we selected the patches with consistent labels from both pathologists for training and testing. To combine different kernels, a normalization by their traces was performed. Normalization is especially crucial for the addition of kernels from different scales. Parameter optimization for capacity C was performed with exhaustive search, combining different parameter values with the SVM classifiers.

3 Results

We focus on three main questions in our experiments: (i) Can we at all classify the cell nuclei into a cancerous and a benign group? (ii) Which features are important for such a classification? Is shape important? (iii) Does superior classification rates lead to improved staining estimation by the complete analysis pipeline?

3.1 Cell nuclei classification

We investigated the performance of the classification task using 10-fold cross-validation over all patches. The results clearly demonstrate that the data support to automatically classify cell nuclei into benign and malignant at a comparable performance level of pathologists (see Figure 4). The best performing kernels utilize all features: foreground and background histograms, shape descriptors and PHOG. The median misclassification error is 17%. The best kernels showed a capacity of $C = 1000$. To confirm that we did not overfit the data, we chose the best kernel using a further cross validation level on the training data. The found best kernel was then tested on a separate test subset of samples that was not used for training. This classifier achieved a median misclassification error of 18%. In 6 out of 10 of the splits, the diffusion distance (with all histogram features) combined with a linear kernel for the PROP features was identified as best performing SVM kernel.

3.2 Importance of different image features

The features that we consider can be grouped into intensity features (foreground and background), shape features (FCC, SIG and PROP) and PHOG, which combines intensity gradients with a region of interest, i.e. the nucleus shape. To see how the different classes of features affect the performance of classifiers, we again performed a double CV over all kernels, separating the kernels into these three groups. Two conclusions can be drawn from Figure 5: (i) shape information improves classification performance, and (ii) the above mentioned feature classes measure different properties of the data; combining them improves the classifiers.

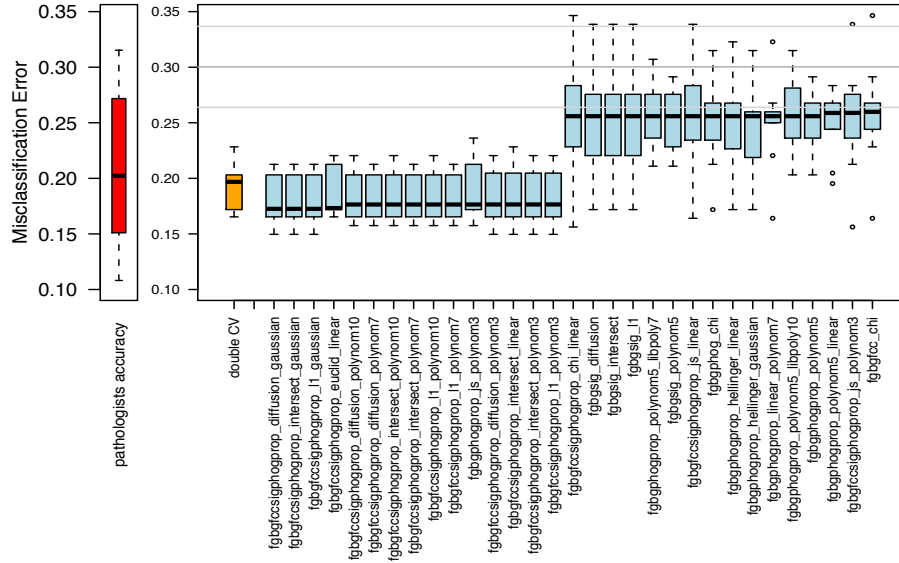


Fig. 4. Left: The “performance” of the pathologist is computed from the confusion matrix between the labels of the two pathologists. **Right:** Performance of kernels in nucleus classification. 15 best performing and 15 worse performing kernels (blue) are shown. Performance measure is the misclassification error in a 10-fold CV. The kernels’ names consists of the used features (see Section 2.3) and the kernel function (for histogram and non-histogram features, if needed). The orange bar represents the double CV result, indicating non-overfitting and the ability to classify new samples (see text). The horizontal line shows the mean (and standard deviation) of 100 permutation tests, indicating chance level of prediction.

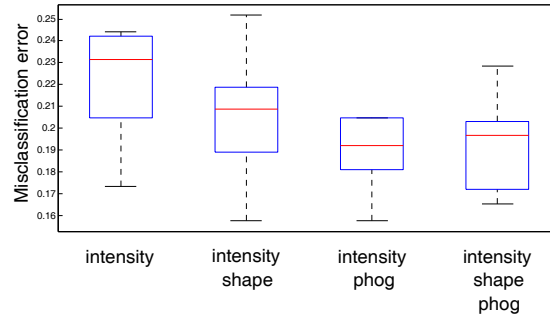


Fig. 5. Misclassification error of best kernels within a certain feature class (intensity: kernels using FG, BG; shape: kernels using FCC, SIG, PROP; phog: kernels using PHOG). Each bar shows the performance of the best kernel using a validation set and a double CV: in the inner CV, the best kernel in a feature class is chosen based on 90% the samples. In the outer CV, this kernel is tested on the remaining 10%. The plot shows that each image feature class carries information for classification.

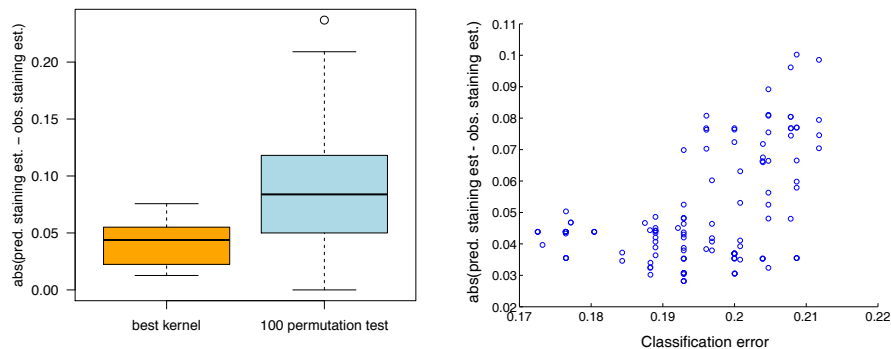


Fig. 6. Effect of nucleus classification performance on staining estimation. **Left:** Comparison between the best classifier and a random classifier with 100 permutation tests on the staining estimation task. In a 10-fold CV, the classifier was trained and used to predict the fraction of stained vs. all cancerous nuclei in the test set. The absolute difference of the predicted fractions to the fractions calculated from the labeled nuclei is shown in the plot. **Right:** Relation between the classifiers’ classification performance and staining estimation error (shown for the best 100 kernels). The staining error (absolute difference) of a classifier is calculated in the same way as in the left plot. The better the classification of a kernel is (more left), the better its staining estimation is (more down). The correlation coefficient 0.48.

3.3 Effect of classifier performance on staining estimation

Recall from Figure 1 that we are ultimately interested in estimating the fraction of cancerous cell nuclei which are stained. In Figure 6, we show the absolute difference in error between the predicted fraction of staining and the fraction of staining indicated by the pathologists’ labels. Note that we used a simple histogram method to estimate the colour. First, we compared the best classifier in Figure 4 to a random classifier. Our results show that a good classifier is able to estimate the staining of the cancerous nuclei with higher accuracy than a random classifier (see Figure 6 (left)). Since the fraction of stained cancerous nuclei is roughly 7% in the data, a classifier that results in an estimate of no staining will have an error of 7% on this plot. In Figure 6 (right) we show the relationship between different classification error rates with the staining estimation error.

4 Conclusion and Outlook

With this study, we contributed in several aspects: (i) We demonstrated that graph cuts can be employed in conjunction with a circular shape prior to segment cell nuclei in a robust fashion. (ii) The classification of nuclei into malignant

and benign is not only feasible but additional shape features boost the classification performance. (iii) We investigated and validated a large number of kernels, distance function and combination thereof (Figure 4). The lesson learned is that all extracted features are necessary for optimal performance and the best kernels perform significantly better than chance and comparable to human domain experts. (iv) Finally we demonstrated the influence of the classification task on the subsequent staining estimation problem.

Future goals of our research are going to be the incorporation of the proposed system into a general purpose TMA analysis framework. Part of that is building a robust nuclei detector and analyzing the resulting estimates in the scope of survival statistics. We are convinced that an automated and objective analysis pipeline will significantly further biomarker detection and cancer research.

Acknowledgements This work was supported by the FET-Open Programme within the 7th Framework Programme for Research of the European Commission, under project SIMBAD grant no. 213250.

References

1. Bagon, S.: Matlab wrapper for graph cut (December 2006), <http://www.wisdom.weizmann.ac.il/~bagon>
2. Ben-Hur, A., Ong, C.S., Sonnenburg, S., Schölkopf, B., Rätsch, G.: Support vector machines and kernels for computational biology. *PLoS Computational Biology* 4(10), e1000173 (2008)
3. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*. pp. 401–408. ACM, New York, NY, USA (2007)
4. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE transactions on Pattern Analysis and Machine Intelligence* 26(9), 1124–1137 (September 2004)
5. Boykov, Y., Veksler, O., Zabih, R.: Efficient approximate energy minimization via graph cuts. *IEEE transactions on Pattern Analysis and Machine Intelligence* 20(12), 1222–1239 (November 2001)
6. Fan, R.E., Chen, P.H., Lin, C.J.: Working set selection using second order information for training support vector machines. *J. Mach. Learn. Res.* 6, 1889–1918 (2005)
7. Fuchs, T.J., Wild, P.J., Moch, H., Buhmann, J.M.: Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients. *MICCAI* (2008)
8. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital image processing using matlab (2003), 993475
9. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? *IEEE transactions on Pattern Analysis and Machine Intelligence* 26(2), 147–159 (February 2004)
10. Kononen J, Bubendorf L, e.a.: Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med.* Jul;4(7), 844–7 (1998)
11. Schölkopf, B., Smola, A.J.: *Learning with Kernels*. MIT Press (2002)