Computerized Medical Imaging and Graphics 35 (2011) 515–530

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Review

Computational pathology: Challenges and promises for tissue analysis

Thomas J. Fuchs[a,b,*], Joachim M. Buhmann[a,b]

[a] Department of Computer Science, ETH Zurich, Universitaetstrasse 6, CH-8092 Zurich, Switzerland
[b] Competence Center for Systems Physiology and Metabolic Diseases, ETH Zurich, Schafmattstr. 18, CH-8093 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 29 June 2010
Received in revised form 21 January 2011
Accepted 23 February 2011

Keywords:
Computational pathology
Machine learning
Medical imaging
Survival statistics
Cancer research
Whole slide imaging

ABSTRACT

The histological assessment of human tissue has emerged as the key challenge for detection and treatment of cancer. A plethora of different data sources ranging from tissue microarray data to gene expression, proteomics or metabolomics data provide a detailed overview of the health status of a patient. Medical doctors need to assess these information sources and they rely on data driven automatic analysis tools. Methods for classification, grouping and segmentation of heterogeneous data sources as well as regression of noisy dependencies and estimation of survival probabilities enter the processing workflow of a pathology diagnosis system at various stages. This paper reports on state-of-the-art of the design and effectiveness of computational pathology workflows and it discusses future research directions in this emergent field of medical informatics and diagnostic machine learning.

Contents

* Corresponding author at: Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, USA. Tel.: +1 626 395 4866;fax: +1 626 795 8649.
E-mail address: fuchs@caltech.edu (T.J. Fuchs).

0895-6111/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2011.02.006

1. Computational pathology: the systems view

Modern pathology studies of biopsy tissue encompass multiple stainings of histological material, genomics and proteomics analyses as well as comparative statistical analyses of patient data. Pathology lays not only a scientific foundation for clinical medicine but also serves as a bridge between the fundamental sciences in natural science to medicine and patient care. Therefore, it can be viewed as one of the key hubs for translational research in the health and life sciences, subsequently facilitating translational medicine. In particular, the abundance of heterogeneous data sources with a substantial amount of randomness and noise poses challenging problems for statistics and machine learning. Automatic processing of this wealth of data promises a standardized and hopefully more objective diagnosis of the disease state of a patient than manual inspection can provide today. An automatic computational pathology framework also enables the medical user to quantitatively benchmark the processing pipeline and to identify error sensitive processing steps which can substantially degrade the final predictions, e.g. of survival times.

1.1. Definition

Computational pathology as well as the medical discipline pathology is a wide and diverse field which encompass scientific research as well as day-to-day work in medical clinics. The following definition is an attempt for a concise and practical description of this novel field:

> Computational Pathology investigates a complete probabilistic treatment of scientific and clinical workflows in general pathology, i.e. it combines experimental design, statistical pattern recognition and survival analysis within a unified framework to answer scientific and clinical questions in pathology.

Fig. 1 depicts a schematic overview of the field and three major parts it comprises: data generation, image analysis and medical statistics, which are described in detail in Sections 2–4.

2. Data: tissue and ground truth

2.1. Clear cell renal cell carcinoma

Throughout this review we use renal cell carcinoma (RCC) as a disease case to design and optimize a computational pathology framework. We argue that computational pathology frameworks for other diseases require a conceptually and structurally similar approach as for RCC.

Renal cell carcinoma figures as one of the 10 most frequent malignancies in the mortality statistics of Western societies [1]. The prognosis of renal cancer is poor since many patients suffer already from metastases at the time of first diagnosis. The identification of biomarkers for prediction of prognosis (prognostic marker) or response to therapy (predictive marker) is therefore of utmost importance to improve patient prognosis [2]. Various prognostic markers have been suggested in the past [3,4], but estimates of conventional morphological parameters still provide most valuable information for therapeutical decisions.

Clear cell RCC (ccRCC) emerged as the most common subtype of renal cancer and it is composed of cells with clear cytoplasm and typical vessel architecture. ccRCC exhibits an architecturally diverse histological structure, with solid, alveolar and acinar patterns. The carcinomas typically contain a regular network of small thin-walled blood vessels, a diagnostically helpful characteristic of this tumor. Most ccRCC specimen show areas with hemorrhage or necrosis (Fig. 3d), whereas an inflammatory response is infrequently observed. Nuclei tend to be round and uniform with finely granular and evenly distributed chromatin. Depending upon the grade of malignancy, nucleoli may be inconspicuous and small, or large and prominent, with possibly very large nuclei or bizarre nuclei occurring [1].

The prognosis for patients with RCC depends mainly on the pathological stage and the grade of the tumor at the time of surgery. Other prognostic parameters include proliferation rate of tumor cells and different gene expression patterns. Tannapfel et al. [2] have shown that cellular proliferation potentially serves as another measure for predicting biological aggressiveness and, therefore, for estimating the prognosis. Immunohistochemical assessment of the MIB-1 (Ki-67) antigen indicates that MIB-1 immunostaining (Fig. 3d) is an additional prognostic parameter for patient outcome. Tissue microarrays (TMAs, cf. Section 2.2) were highly representative of proliferation index and histological grade using bladder cancer tissue [5].

The TNM staging system specifies the local extension of the primary tumor (T), the involvement of regional lymph nodes (N), and the presence of distant metastases (M) as indicators of the disease state. Wild et al. [6] focus on reassessing the current TNM staging system for RCC and conclude that outcome prediction for RCC remains controversial. Although many parameters have been tested for prognostic significance, only a few have achieved general acceptance in clinical practice. An especially interesting observation of Wild et al. [6] is that multivariate Cox proportional hazards regression models including multiple clinical and pathologic covariates were more accurate in predicting patient outcome than the TNM staging system. On one hand this finding demonstrates the substantial difficulty of the task and on the other hand it is a motivation for research in computational pathology to develop robust machine learning frameworks for reliable and objective prediction of disease progression.

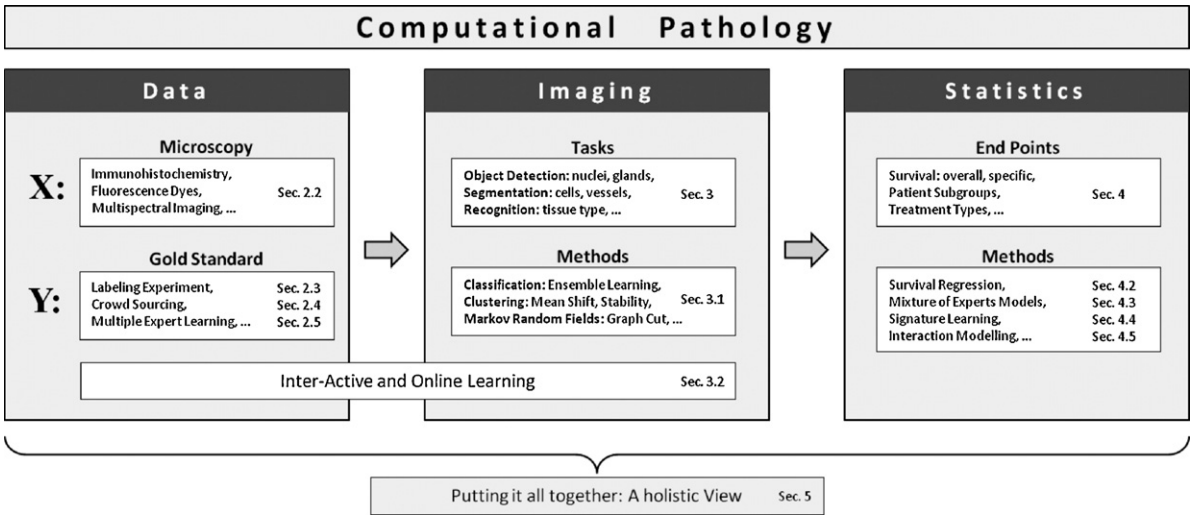

Fig. 1. Schematic overview of a workflow in computational pathology comprising three major parts: (i) the covariate data X is acquired via microscopy and the target data Y is generated in labeling experiments; Y provides training and testing information for supervised problems and it enables validation in unsupervised settings; (ii) image analysis in terms of nuclei detection, cell segmentation or texture classification yields detailed information about the tissue; (iii) medical statistics, i.e. survival regression and mixture of expert models are employed to investigate the clinical end point of interest, using data from the previous two stages. The aim is to build a complete probabilistic workflow comprising all three parts.

2.2. Tissue microarrays

The tissue microarray (TMA) technology significantly accelerated studies seeking for associations between molecular changes and clinical endpoints [7]. In this technology, tissue cylinders of a 0.6 mm diameter are extracted from primary tumor material 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 (cf. Fig. 3). These results can then be integrated with expression profile data which is expected to enhance the diagnosis and prognosis of ccRCC [8,3,9]. 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.

Although the production of tissue microarrays is an almost routine task for most laboratories, the evaluation of stained tissue microarray slides remains tedious human annotation work, which is time consuming and prone to error. Furthermore, the significant intratumoral heterogeneity of RCC results in high inter-observer variability. The variable architecture of RCC also results in a difficult assessment of prognostic parameters. State of the art commercial 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. Because of these drawbacks in current software, pathologists typically collect tissue microarray data by manually assigning a composite staining score for 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 hinders the use of tissue microarrays in high-throughput analysis.

2.3. Analyzing pathologists

To assess the inter- and intra-observer variability of pathologists we designed three different labeling experiments for the major tasks involved in TMA analysis. To facilitate the labeling process for trained pathologists we developed a software suite which allows the user to view single TMA spots and which provides zooming and scrolling capabilities. The expert can annotate the image with vectorial data in SVG (support vector graphics) format and he/she can mark cell nuclei, vessels and other biological structures. In addition each structure can be labeled with a class which is encoded by its color. To increase usability and the adoption in hospitals the software has been specifically designed for tablet PC so that a pathologist can perform all operations with a pen alone in a simple and efficient manner. All experiments were conducted with the same tablet PC employing the same conditions. Fig. 2 depicts the graphical user interfaces of the three applications.

2.3.1. Nuclei detection

The most tedious labeling task is the detection of cell nuclei. In this experiment two experts on renal cell carcinoma exhaustively labeled a quarter of each of the 9 spots from the previous experiment. Overall each expert independently marked the center, the approximate radius and the class of more than 2000 nuclei. Again a tablet PC was used so it was possible to split up the work into several sessions and the experts could use the machine at their convenience. The user detects a nucleus by marking its location and by drawing a circular semi-transparent polygon to cover it. The final step consists of choosing a class for the nucleus. In this setting it was either black for atypical nuclei or red for normal ones. This annotation work has to be repeated for each nucleus on each spot. Fig. 4 depicts a quarter of one of the RCC TMA spots together with the annotation and the disagreement between experts.

The average precision $(tp/(tp+fp))$ of one pathologist compared to the other is 0.92 and the average recall $(tp/(tp+fn))$ amounts to 0.91. These performance numbers show that even detecting nuclei on an histological slide is by far not an easy or undisputed task.

2.3.2. Nuclei classification

The second experiment was designed to evaluate the inter and intra pathologist variability for nuclei classification, i.e. determining if a nucleus is normal or atypical. This step crucially influences the final outcome due to the fact that the percentage of staining is only estimated on the subset of atypical nuclei. In the experiment, 180 randomly selected nuclei are sequentially presented in three different views of varying magnification. The query nucleus is indicated in each view with a red cross and the area which comprises the next magnification is marked with a red bounding box

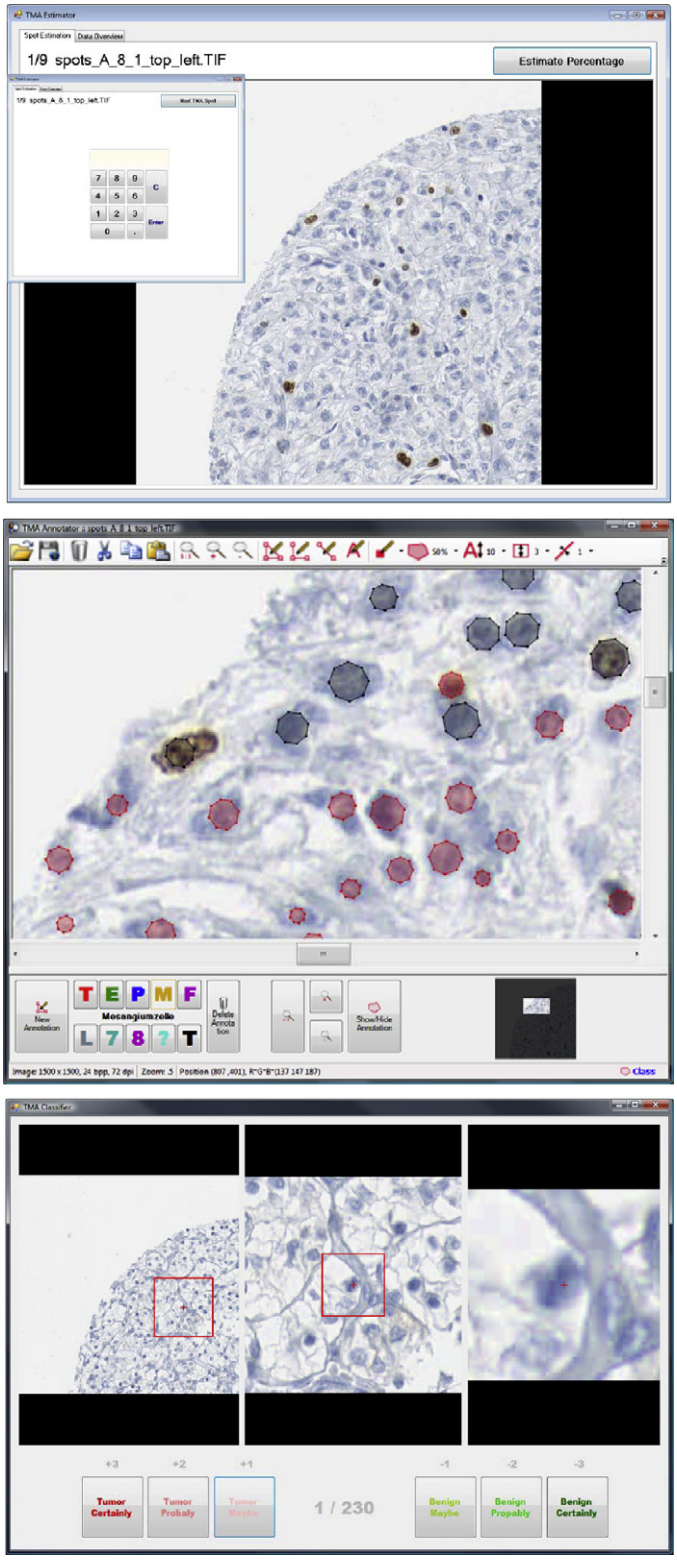

Fig. 2. Tablet PC labeling applications for (i) global staining estimation; (ii) nuclei detection and (iii) nuclei classification (from top to bottom).

(cf. Fig. 2). During the setup phase the user can adjust these views to simulate his usual workflow as good as possible. During the experiment the expert has to select a class for each nucleus and rate his confidence. Thus, he has the choice between six buttons: atypical certainly, atypical probably, atypical maybe, normal certainly, normal probably and normal maybe. After classifying all nuclei, which have been classified as tumor, are displayed again and the pathologist has to estimate if the nucleus is stained or not. Again he has to rate his confidence in his own decision on a scale of three levels. To test the intra pathologist's variability a subset of nuclei was queried twice but the images were flipped and rotated by 90° at the second display to hamper recognition.

The results for inter-pathologist variability for the binary classification task are plotted in Fig. 5a. Out of 180 nuclei all five experts agreed on 24 nuclei to be normal and 81 nuclei to be atypical, respectively cancerous. For the other 75 nuclei (42%) the pathologists disagreed.

The analysis of the intra-pathologist error is shown in Fig. 5b. The overall intra classification error is 21.2%. This means that every fifth nucleus was classified by an expert first as atypical and the second time as normal or vice versa. The self-assessment of confidence allows us also to analyze single pathologists. For example Fig. 5c shows the results of a very self-confident pathologist who is always very certain of his decisions but ends up with an error of 30% in the replication experiment. Fig. 5d on the other hand is the result of a very cautious expert who is rather unsure of his decision, but with a misclassification error of 18% he performs significantly better than the previous one. The important lesson learned is, that self-assessment is not a reliable information to learn from. The intuitive notion, to use only training samples which were classified with high confidence by domain experts is not valid.

In defense of human pathologists it has to be mentioned that these experiments represent the most general way to conduct a TMA analysis and analogous studies in radiology report similar results [10,11]. In practice, domain experts focus only on regions of TMA spots which are very well processed, which have no staining artifacts or which are not blurred. The nuclei analyzed in this experiment were randomly sampled from the whole set of detected nuclei to mimic the same precondition which an algorithm would encounter in routine work. Reducing the analysis to perfectly processed regions would most probably decrease the intra-pathologist error.

2.3.3. Staining estimation

The most common task in manual TMA analysis requires to estimate the staining. To this end a domain expert briefly (e.g. several seconds) views the spot of a patients and estimates the number of stained atypical cells without resorting to actual nuclei counting. This procedure is iterated for each spot on a TMA-slide to get an estimate for each patient in the study. It is important to note that, due to the lack of competitive algorithms, the results of nearly all TMA studies are based on this kind of subjective estimations. To investigate estimation consistency we presented 9 randomly selected TMA spots to 14 trained pathologists of the University Hospital Zurich.

The estimations of the experts varied by up to 20% as shown in Fig. 6a. As depicted in Fig. 6b the standard deviation between the experts grows linearly with the average estimated amount of staining. The high variability demonstrates the subjectivity of the estimation process. It is interesting to note that the ranking of TMA spots according to their staining degree is much more consistent than the direct estimation of the continuous percentage value (cf. Fig. 7).

This uncertainty is especially critical for types of cancer for which the clinician chooses the therapy based on the estimated staining percentage. This result not only motivates but emphasizes the need for more objective estimation procedures than current practice. Research in this field should be stimulated by the hope, that computational pathology approaches do not only automate such estimation processes but also produce better reproducible and more objective results than human judgment.

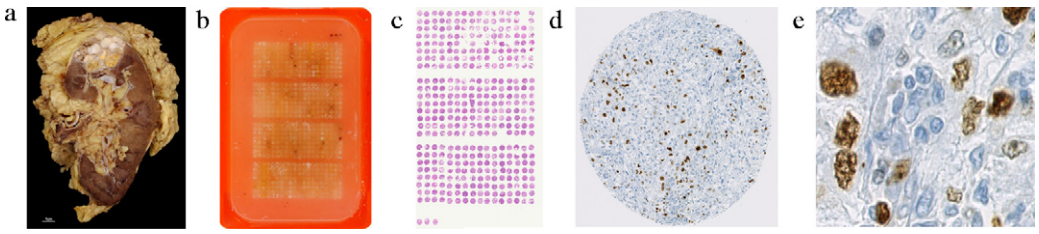

Fig. 3. Tissue microarray analysis (TMA): primary tissue samples are taken from a cancerous kidney (a). Then tissue cylinders of a 0.6 mm diameter are extracted from the primary tumor material of different patients and arrayed in a recipient paraffin block (b). Slices of 0.6 μm are cut off the paraffin block and are immunohistochemically stained (c). These slices are scanned as whole slide images and tiled into single images representing different patients. Image (d) depicts a TMA spot of clear cell renal cell carcinoma stained with MIB-1 (Ki-67) antigen. (e) shows details of the same spot containing stained and non-stained nuclei of normal as well as atypical cells.

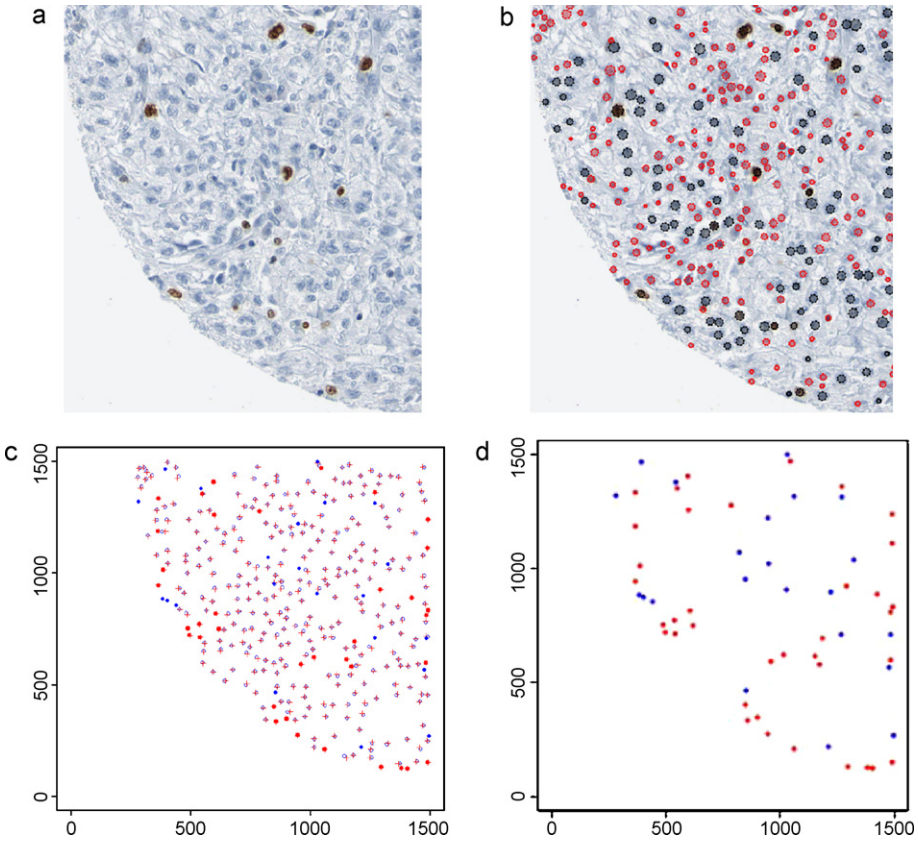

Fig. 4. (a) A quarter of an RCC TMA spot used for the nuclei detection experiment. (b) Annotations of one expert, indicating atypical nuclei in black and normal ones in red. (c) Overlay of detected nuclei from expert one (blue circles) and expert two (red crosses). (d) Disagreement between the two domain experts regarding the detection task. Nuclei which were labeled only by pathologist one are shown in blue and the nuclei found only by expert two are depicted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

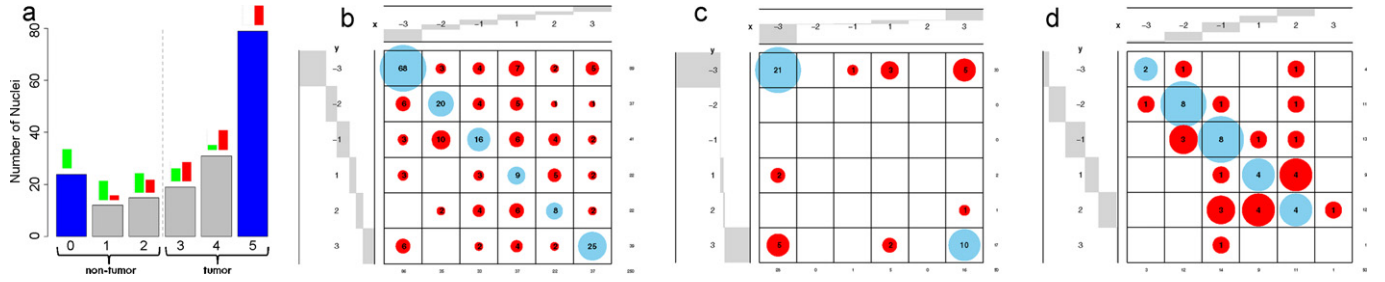

Fig. 5. (a) Inter-pathologist classification variability based on 180 nuclei labeled by five domain experts. The experts agree on 105 out of 180 nuclei (blue bars: 24 normal, 81 atypical). (b–d) Confusion matrices including reader confidence for intra-observer variability in nuclei classification: (b) The combined result of all five experts yields an intra pathologist classification error of 21.2%. (c) Example of an extremely self-confident pathologist with 30% error. (d) A very cautions pathologist with a misclassification error of 18%. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

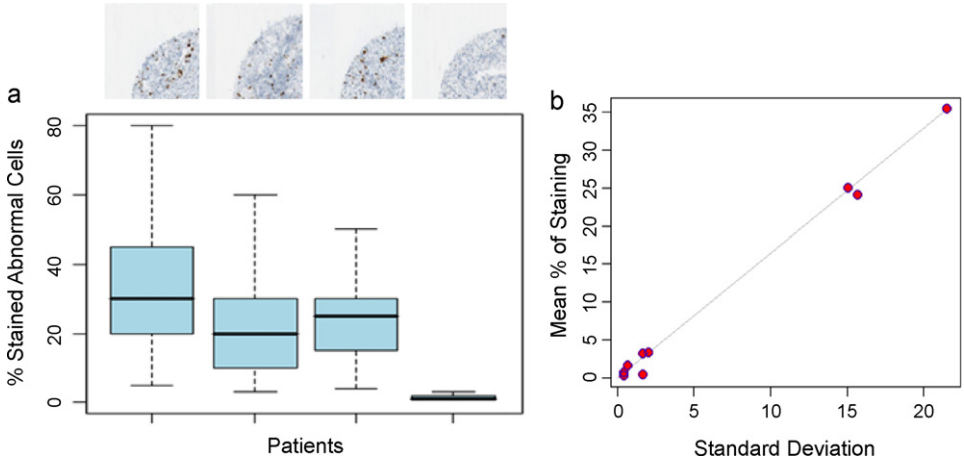

Fig. 6. (a) Results for 4 TMA spots from the labeling experiment conducted to investigate the inter pathologist variability for estimating nuclear staining. 14 trained pathologists estimated MIB-1 staining on 9 TMA spots. The boxplots show a large disagreement between pathologist on spots with an averages staining of more than 10%. The absolute estimated percentage is plotted on the y-axis. Spot 1 for example, yields a standard deviation of more than 20%. (b) The standard deviation grows linearly with the average estimated staining.

2.4. Expert variability in fluorescence microscopy

Complementary to immunohistochemical TMA analysis, fluorescence microscopy is applied often for high-throughput screening of molecular phenotypes. A comprehensive study evaluating the performance of domain experts regarding the detection of lymphocytes is presented by Nattkemper et al. [12]. In a best case, a medium-skilled expert needs on average one hour for analyzing a fluorescence micrograph. Each micrograph contains between 100 and 400 cells and is of size 658×517 pixel. Four exemplary micrographs were blindly evaluated by five experts. To evaluate the inter-observer variability Nattkemper et al. [12] define a gold standard comprising all cell positions in a micrograph that were detected by at least two experts.

Averaged over of CD3, CD4, CD7 and CD8 the sensitivity of the four biomedical experts is varying between 67.5% and 91.2% and the positive predictive value (PPV) between 75% and 100%. Thus the average detection error over all biomedical experts and micrographs is approximately 17%. Although fluorescence images appear to be easier to analyze due to their homogeneous background, this high detection error indicates the difficulty of this analysis task. These results corroborates the findings in the ccRCC detection experiment described in Section 2.1.

2.5. Generating a gold standard

The main benefit of labeling experiments, like the ones described before, is not to point out the high variability between pathologists or even their inconsistencies in repeated annotations of identical data, but to generate a gold standard. In absence of an objective ground truth measurement process, a gold standard is crucial for the use of statistical learning, first for learning a classifier or regressor and second for validating the statistical model. Section 5 shows an example how the information gathered in the experiments of Section 2.3 can be used to train a computational pathology system.

Besides labeling application which are developed for specific scenarios as the one described in Section 2.3 several other possibilities exist to acquire data in pathology in a structured manner. Although software for tablet PCs is the most convenient approach

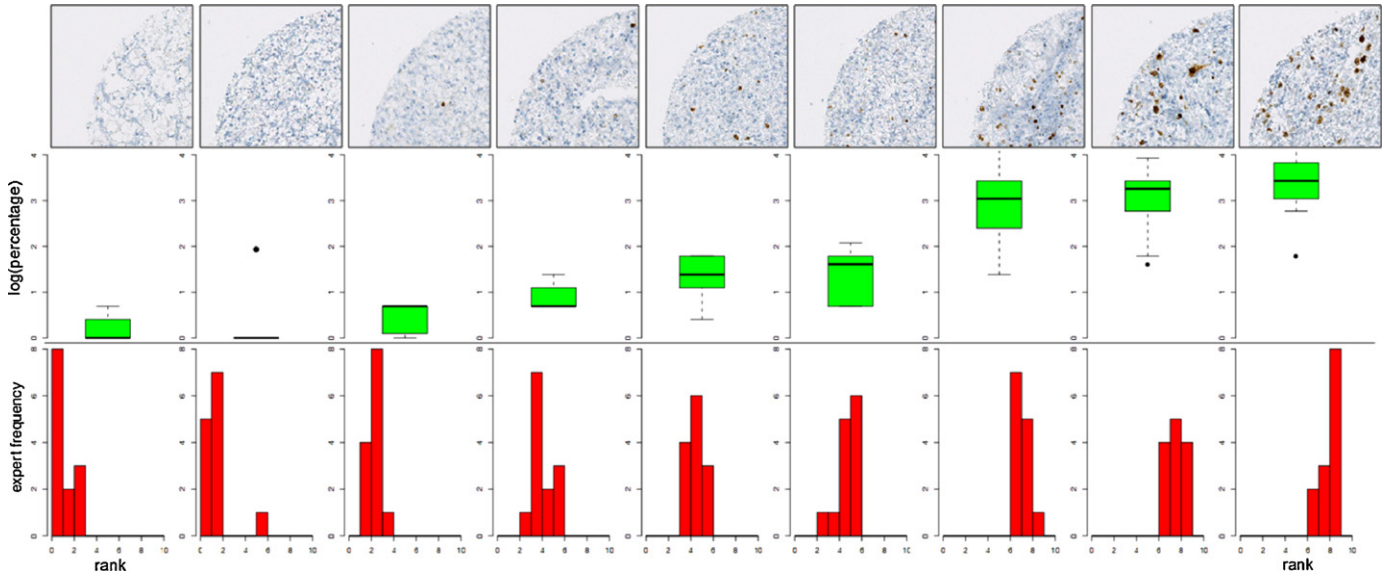

Fig. 7. Comparison between ranking and continuous staining estimation of nine renal cell carcinoma TMA spots with MIB-1 staining. The experiment was conducted by 14 trained pathologists and demonstrates the high consistency of the ranking data compared to the conventional direct estimation of the percentage of stained atypical nuclei.

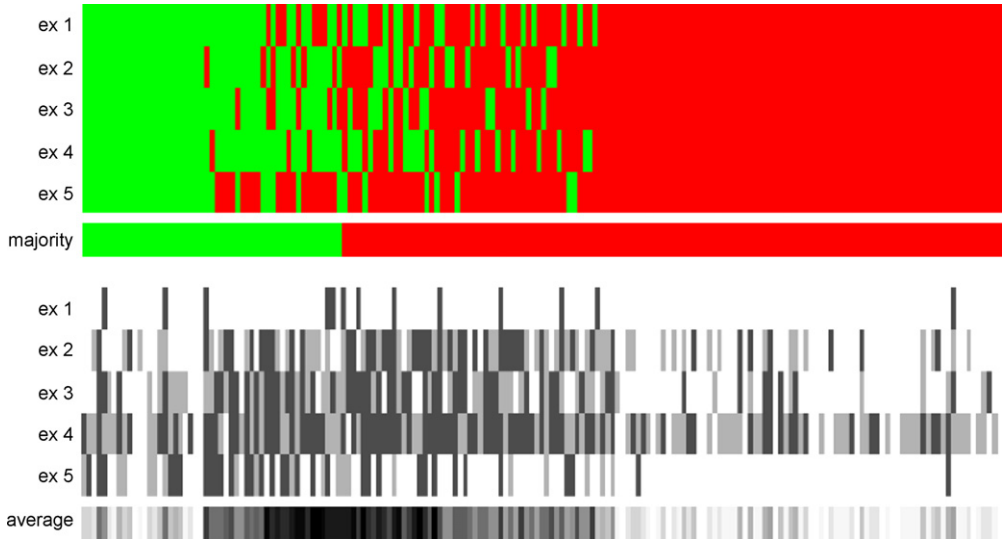

Fig. 8. Labeling matrix with majority vote (top) and confidence matrix with confidence average (bottom) of five domain experts classifying 180 ccRCC nuclei into atypical (red) and normal (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

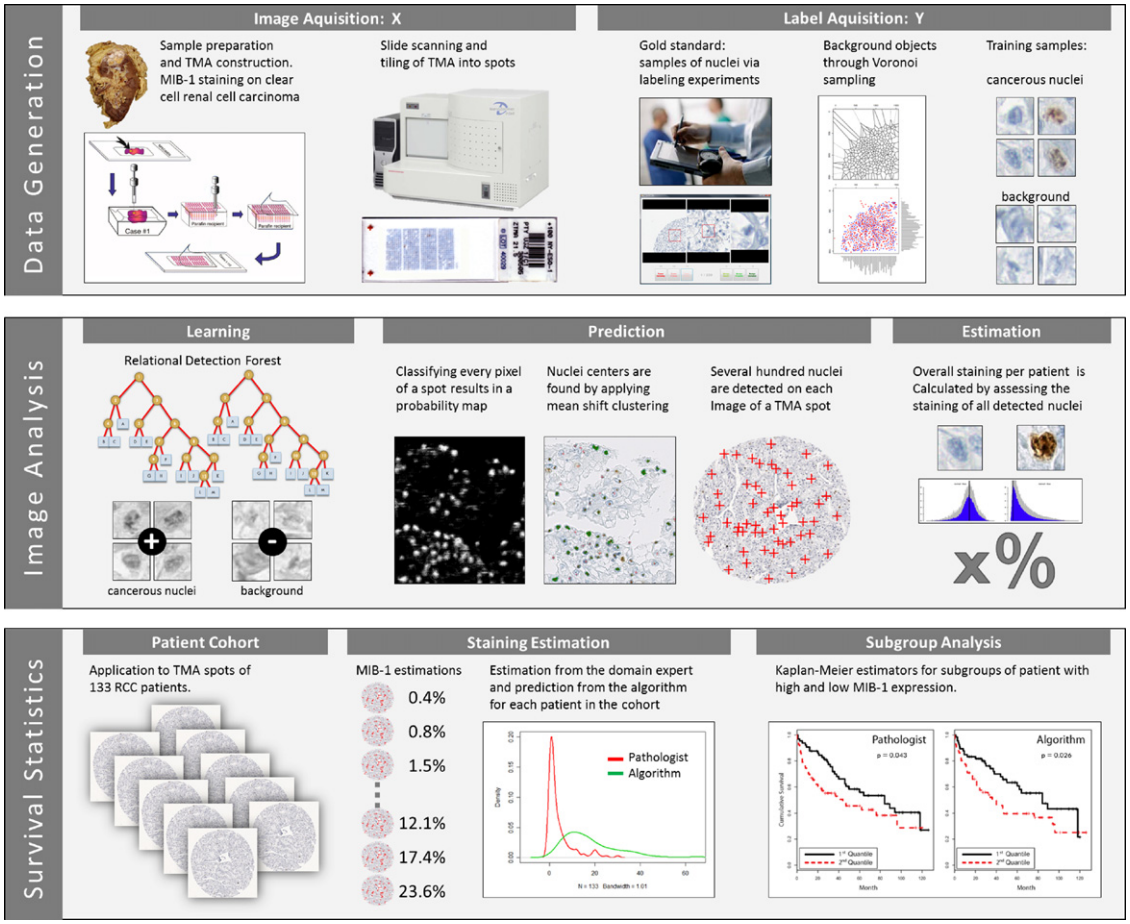

Fig. 9. A computational pathology framework for investigating the proliferation marker MIB-1 in clear cell renal cell carcinoma. Following the definition in Section 1.1 the framework consists of three parts: (i) the covariate data *X* existing of images of TMA spots was generated in a trial at the University Hospital Zürich. Extensive labeling experiments were conducted to generate a gold standard comprising atypical cell nuclei and background samples. (ii) Image analysis consisted of learning a relational detection forest (RDF) and conducting mean shift clustering for nuclei detection. Subsequently, the staining of detected nuclei was determined based on their color histograms. (iii) Using this system, TMA spots of 133 RCC patients were analyzed. Finally, the subgroup of patients with high expression of the proliferation marker was compared to the group with low expression using the Kaplan–Meier estimator.

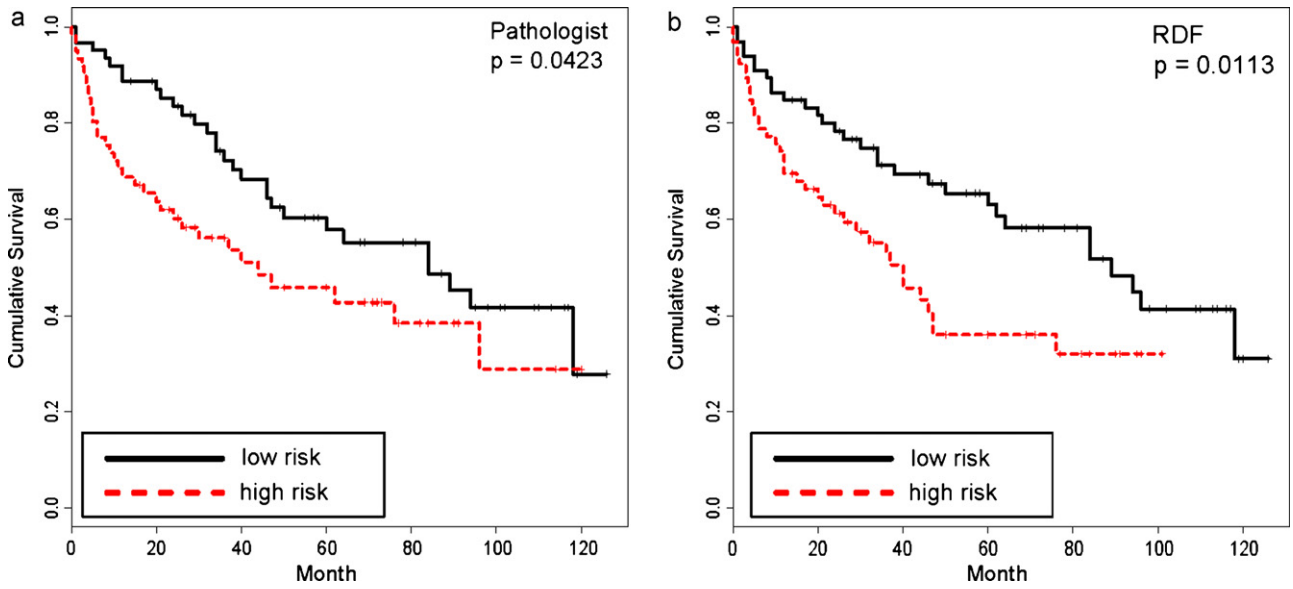

Fig. 10. Kaplan–Meier estimators show significantly different survival times for renal cell carcinoma patients with high and low proliferating tumors. Compared to the manual estimation from the pathologist (a) ($p = 0.04$), the fully automatic estimation from the algorithm and (b) compares favorable ($p = 0.01$) in terms of survival differences (log rank test) for the partitioning of patients into two groups of equal size [45].

to gather information directly in the hospital it is limited by the low number of test subjects which can complete an experiment. To overcome this limitation the number of labelers can be extended significantly by the use of web-based technologies.

Crowd-sourcing services like Amazon Mechanical Turk can be used to gather large numbers of labels at a low cost. Applications in pathology suffer from the main problem, that the labelers are all non-experts. While crowd-sourcing works well for task based on natural images [13], it poses considerable problems in pathology where for example the decision if a nucleus is normal or atypical is based on complicated criteria [14] which require medical training and knowledge. Likewise the recognition of some supercellular morphological structures requires years of training and supervision. Nevertheless crowd-sourcing could be useful in simple detection tasks like finding nuclei in histological slides. For the task of image segmentation, Warfield et al. [15] present an expectation maximization algorithm to estimate a gold standard based on the annotations of multiple experts and demonstrate its application in radiology.

2.6. Multiple expert learning

In classical supervised learning, a set of training data $\{(x_i, y_i)\}_{i=1,\ldots,n}$ is available which consists of objects x_i and their corresponding labels y_i. The task is to predict the label y for a new test object x. This approach is valid as long as the target variable $Y = \{y_1, \ldots, y_n\}$ denotes the ground truth of the application. If this condition is met, Y and $X = \{x_1, \ldots, x_n\}$ can be used for classifier learning and evaluation.

Unfortunately, for a large number of real word application ground truth is either not available or very expensive to acquire. In practice, as a last resort, one would ask several domain experts for their opinion about each object x_i in question to generate a gold standard as described in Section 2.5. Depending on the difficulty of the task and the experience of the experts this questioning often results in an ambiguous labeling due to disagreement between experts. In pathology, very challenging scenarios, like assessing the malignancy of cells, induce not only high inter-expert variability, but also the intra expert disagreement is quite large (cf. Section 2.3 and Fig. 8). Moreover, restricting the dataset to the subset of consistently labeled samples results in loss of the majority of data in these scenarios. Consequently, such a data acquisition procedure poses a fundamental problem for supervised learning. Especially in computational pathology there is clearly a need for novel algorithms to address the labeling problem and to provide methods to validate models under such circumstances.

More formally, each y_i is replaced by a D dimensional vector $\bar{y}_i = \{y_i^1, \ldots, y_i^D\}$, where y_i^d represents the ith label of domain expert d. To this end one is interested in learning a classifier $\Phi(X, \bar{Y})$ from the design matrix X and the labeling matrix $\bar{Y}$. To date it is an open research question, how such classifier $\Phi(X, \bar{Y})$ should be formulated.

Recently, Smyth et al. and Raykar et al. [16,17] presented promising results based on expectation maximization where the hidden ground truth is estimated in turn with the confidence in the experts. Also along this lines Whitehill et al. [18] introduced a probabilistic model to simultaneously infer the label of images, the expertise of each labeler, and the difficulty of each image. An application for diabetes especially for detecting hard exudates in eye fundus images was published by Kauppi et al. [19]. Although a number of theoretical results exist [20–22], empirical evidence is still lacking to establish that these approaches are able to improve over simple majority voting [23,24] in real world applications.

A further, promising line of research investigates the question if such a classifier $\Phi(X, \bar{Y})$ can be learned in an on-line fashion, especially when the new labels come from a different domain expert. An affirmative answer would show a high impact in domains where specific models can be trained for years by a large number of experts, e.g. medical decision support.

In summary, extending supervised learning to handle domain expert variability is an exciting challenge and promises direct impact on applications not only in pathology but in a variety of computer vision tasks where generating a gold standard poses a highly non trivial challenge.

2.7. Public datasets with labeling information

The availability of public datasets with labeling information is crucial for the advance of an empirical science. Although a comprehensive archive like the UCI machine learning repository [25]

does not exist for computational pathology, there are a number of datasets and initiatives which disseminate various kinds of data.

2.7.1. Immunohistochemistry

The most comprehensive database for antibodies and human tissue is by far the Human Protein Atlas [26,27]. Comprising spots from tissue micro arrays of 45 normal human tissue types, it contains anywhere from 0–6 images for each protein in each tissue type. The images are roughly 3000×3000 pixels in size, with each pixel approximately representing a 0.5×0.5 μm region on the microscopy slide.

A segmentation benchmark for various tissue types in bioimaging was compiled by Manjunath et al. [28], including 58 histopathological H&E stained images of breast cancer. The dataset provides labels from a single expert for the tasks of segmentation and cell counting. Finally, Maree et al. [29] presented an approach based on shared randomized vocabularies to search in such tissue image databases.

2.7.2. Cytology

Automation in cytology is the oldest and most advanced branch in the field of image processing in pathology. Reason therefore are that digital imaging is rather straightforward and that single cells on a homogeneous background are more easily detected and segmented than in tissue. As a result, commercial solutions are available since decades. Nevertheless especially the classification of detected and segmented nuclei still poses large difficulties for computational approaches. Lézoray and Cardot [30] published 10 color microscopic images from serous cytology with hand segmentation labels. For bronchial cytology Meurie et al. [31] provide eight color microscopic images. Ground truth information for three classes (nucleus, cytoplasm, and background pixels) is also available for each image. Pixels have a label specifying their classes (2: nucleus, 1: cytoplasm, 0: background). A dataset of 3900 cells has been extracted from microscopical image (serous cytology) by Lézoray et al. [32]. This database has been classified into 18 cellular categories by experts.

2.7.3. Fluorescence microscopy

A hand-segmented set of 97 fluorescence microscopy images with a total of 4009 cells has been published by Coelho et al. [33]; Maree et al. [34] presented 93 pairs of images corresponding to N- and C-terminal green fluorescent protein fusions of cDNAs. For fluorescence microscopy, the simulation of cell population images is an interesting addition to validation with manual labels of domain experts. Nattkemper et al. [12] and Lehmussola et al. [12] present simulation frameworks for synthetically generated cell population images. The advantage of these techniques is the possibility to control parameters like cell density, illumination and the probability of cells clustering together. Lehmussola et al. [35] supports also the simulation of various cell textures and different error sources. The obvious disadvantage are (i) that the model can only simulate what it encodes and therefore can not represent the whole variability of biological cell images and (ii) that these methods can only simulate cell cultures without morphological structure. The later disadvantage also prevents their use in tissue analysis. Although the thought of simulated tissue images in light microscopy is appealing, currently no methods exist which could even remotely achieve that goal.

3. Imaging: from classical image processing to statistical pattern recognition

In recent years, a shift from rule based expert system towards learned statistical models could be observed in medical information systems. The substantial influence that machine learning had on the computer vision community is also reflecting more and more on medical imaging in general and histopathology in particular. Classifiers for object detection and texture description in conjunction with various types of Markov random fields are continuously replacing traditional watershed based segmentation approaches and handcrafted rule-sets. Just recently, Monaco et al. [36] successfully demonstrated the use of pairwise Markov models for high-throughput detection of prostate cancer in histological sections. An excellent review of state-of-the-art histopathological image analysis methodology was compiled by Gurcan et al. [37].

As with most cutting edge technologies, commercial imaging solutions lag behind in development but the same trend is evident. Rojo et al. [38] review commercial solutions for quantitative immunohistochemistry in the pathology daily practice.

Despite the general trend towards probabilistic models, very classical approaches like mathematical morphology [39] are still used with great success. Recently, Lézoray and Charrier [40] presented a framework for segmentation based on morphological clustering of bivariate color histograms and we [41] devised an iterative morphological algorithm for nuclei segmentation. Besides common computer vision tasks like object detection, segmentation and recognition, histopathological imaging poses domain specific problems such as estimating staining of nuclei conglomerates [42] and differentiation nuclei by their shape [43].

3.1. Preprocessing vs. algorithmic invariance

Brightfield microscopic imaging often produces large differences of illumination within single slides or TMA spots. These variations are caused by the varying thickness of the slide or by imperfect staining. Such problems can be overcome either by preprocessing the image data or by designing and integrating invariance into the algorithmic processing to compensate for these variations.

Inconsistencies in the preparation of histology slides render it difficult to perform a quantitative analysis on their results. A normalization approach based on optical density and SVD projection was proposed by Macenko et al. [44] for overcoming some of the known inconsistencies in the staining process. Slides which were processed or stored under very different conditions are projected into a common, normalized space to enable improved quantitative analysis.

Preprocessing and normalization methods usually not only reduce noise induced differences between samples but often also eliminate the biological signal of interest. As an alternative to such an irreparable information loss during data acquisition, algorithms with illumination invariance or with compensation of staining artifacts are designed which are robust to these uncontrollable experimental variations.

Relational detection forests [45] provide one possibility to overcome this problem of information loss. Especially designed for detection of cell nuclei in histological slides, they are based on the concept of randomized trees [46]. The features, which are selected for this framework center around the idea that relation between features are more robust than thresholds on single features. A similar idea was applied by Geman et al. [47] to gene chip analysis where similar problems arise, due to the background noise of different labs. Contrary to absolute values, relations between DNA expressions are rather robust to preparation artifacts.

Object detection is commonly solved by training a classifier on patches centered at the objects of interest [48], e.g., the cell nuclei in medical image processing of histological slides. When considering only the relation between rectangles within these patches, such data preprocessing results in illumination invariant features which produce the same response for high and low contrast patches as long as the shape of the object is preserved. It has to be noted,

that such illumination invariant features fail for inverted images due to the directionality of the relation. In general, illumination invariance speeds up the whole analysis process because neither image normalization nor histogram equalization are required.

The feature base proposed in [45] is defined as follows: The coordinates of two rectangles R_1 and R_2 are sampled uniformly within a predefined window size w:

$$R_i = \{c_{x1}, c_{y1}, c_{x2}, c_{y2}\}, \quad c_i \sim U(x|0, w) \tag{1}$$

For each rectangle the intensities of the underlying gray scale image are summed up and normalized by the area of the rectangle. The feature $f(s, R_1, R_2)$ evaluates to a boolean value by comparing these quantities:

$$f(s, R_1, R_2) = \begin{cases} 1 & \text{if } \sum_{i|x_i \in R_1} \frac{x_i}{n_1} < \sum_{i|x_i \in R_2} \frac{x_i}{n_2} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where x_i is the gray value intensity of pixel i of sample $s = \{x_1, x_2, \ldots, x_n\}$ and n_1, n_2 denote the number of samples in R_1, R_2, respectively. From a general point of view this definition is similar to generalized Haar features but there exist two main differences: (i) the quantity of interest is the boolean relation between the rectangles and not the continuous difference between them and hence (ii) it is superfluous to learn a threshold on the difference to binarize the feature.

For example, a window size of 65×65 pixels was chosen in the validation experiments. Sampling two corners (x_1, y_1, x_2, y_2) of each of the two rectangles and taking into account their flipping invariance, results in $((64^4)/4)^2 \approx 2 \times 10^{13}$ possible features. Putting this number into perspective, the restriction of the detector to windows of size 24×24 leads to $\sim 6.9 \times 10^9$ features which significantly exceeds the 45,396 Haar features from classical object detection approaches [48].

For such huge feature spaces it is currently not possible to exhaustively evaluate all features while training a classifier. Approaches like AdaBoost [49] which yield very good results for up to hundreds of thousands of features are not applicable any more. These problems can be overcome by employing randomized algorithms [45,50] where features are sampled randomly for learning classifiers on these random projections.

3.2. Inter-active and online learning for clinical application

Day-to-day clinical application of computational pathology algorithms requires adaptivity to a large variety of scenarios. Not only are staining protocols and slide scanners constantly updated and changed but common algorithms like the quantification of proliferation factors have to work robustly on various tissue types. The detection of multiple objects like nuclei in noisy images without an explicit model still amounts to one of the most challenging tasks in computer vision. Methods which can be applied in an plug-and-play manner are still not available to date.

Fuchs and Buhmann [51] present an inter-active ensemble learning algorithm based on randomized trees, which can be employed to infer an object detector in an inter-active fashion. In addition this learning method can cope with high dimensional feature spaces in an efficient manner and in contrast to classical approaches, subspaces are not split based on thresholds but by learning relations between features.

Incorporating the knowledge of domain experts into the process of learning statistical models poses one of the main challenges in machine learning [52] and computer vision. Data analysis applications in pathology share properties of online and active learning which can be termed inter-active learning. The domain expert interferes with the learning process by correcting falsely classified samples. Algorithm 1 sketches an overview of the inter-active learning process.

In recent years online learning has been of major interest to a large variety of scientific fields. From the viewpoint of machine learning, Blum [53] summarizes a comprehensive overview of existing methods and open challenges. In computer vision online boosting has been successfully applied to car detection [54], video surveillance [55] and visual tracking [56]. One of the first inter-active frameworks was developed by Raman et al. [57] and applied to pedestrian detection.

Ensemble methods like boosting [49] and random forests [58,46] celebrated success in a large variety of tasks in statistical learning but in most cases they are only applied offline. Lately, online ensemble learning for boosting and bagging was investigated by Oza [59] and Fern and Givan [60]. The online random forest as proposed by Elgawi [61] incrementally adopts new features. Updating decision trees with new samples was described by Utgoff [62,63] and extended by Kalles and Morris [64] and Pfahringer et al. [65]. Update schemes for pools of experts like the WINNOW and Weighted Majority Algorithm were introduced by Littlestone [66,67] and successfully employed since then.

In many, not only medical domains, accurate and robust object detection specifies a crucial step in data analysis pipelines. In pathology for example, the detection of cell nuclei on histological slides serves as the basis for a larger number of tasks such as immunohistochemical staining estimation and morphological grading. Results of medical interest such as survival prediction are sensitively influenced by the accuracy of the object detection algorithm. The diagnosis of the pathologist in turn leads to different treatment strategies and hence directly affects the patient. For most of these medical procedures the ground truth is not known (see Section 2) and for most problems biomedical science lacks orthogonal methods which could verify a considered hypothesis. Therefore, the subjective opinion of a medical doctor is the only gold standard available for training such decision support systems.

Algorithm 1. Schematic workflow of an inter-active ensemble learning framework. The domain expert interacts with the algorithm to produce a classifier (object detector) which satisfies the conditions based on the experts domain knowledge.

Data: Unlabeled InstancesU = $\{u_1, \ldots, u_n\}$
%(e.g. image) **Input**: Domain Expert E
Output: Ensemble Classifier C

while *(expert is unsatisfied with current result)* **do**
  classify all samples u_i;
  while *(expert corrects falsely predicted sample u_i with $label_i$)* **do**
    update weights of the base classifiers
    learn new base classifiers
  end
end
return C

In such scenarios the subjective influence of a single human can be mitigated by combining the opinions of a larger number of experts. In practice consolidating expert judgments is a cumbersome and expensive process and often additional experts are not available at a given time. To overcome these problems online learning algorithms are capable of incorporating additional knowledge, so-called side-information, when it becomes available.

In an ideal clinical setting, a specialized algorithm for cell nuclei detection should be available for each subtype of cancer. By using and correcting the algorithm several domain experts as its users

continuously train and update the method. Thereby, the combined knowledge of a large number of experts and repeated training over a longer period of time yields more accurate and more robust classifiers than batch learning techniques.

The described setting differs from the conventional views of online learning and active learning insofar that new samples are neither chosen at random nor proposed for labeling by the algorithm itself. In addition, the adversary is not considered malicious but it is also not completely trustworthy. The domain expert reacts to the classification of unlabeled data and corrects wrongly classified instances. These preconditions lead to the success or failure of different combination rules.

It has to be noted, that these kind of machine learning approaches are in sharp contrast to classical rule based expert systems [68] which are still used by a number of commercial medical imaging companies. For these applications the user has to be an image processing expert who chooses dozens of features and thresholds by hand to create a rule set adapted to the data. Contrary to that strategy, in an inter-active learning framework the user has to be a domain expert, in our case a trained pathologists. Feature extraction and learning of statistical models is performed by the algorithms so that the expert can concentrate on the biomedical problem at hand. Inter-active learning frameworks like [54,51] show promising results, but further research especially on long term learning and robustness is mandatory to estimate the reliability of these methods prior to an application in clinical practice.

3.3. Multispectral imaging and source separation

Multispectral imaging (MSI) [69,70] for immunohistochemically stained tissue and brightfield microscopy seems to be a promising technology although a number of limitations have to be kept in mind.

To date, double- or triple-staining of tissue samples on a single slide in brightfield (non-fluorescence) microscopy poses still a major experimental challenge. Traditionally, double staining relied on chromogens, which have been selected to provide maximum color contrast for observation with the unaided eye. For visually good color combinations, however, technically feasible choices always include at least one diffuse chromogen, due to the lack of appropriate chromogen colors. Additional problems arise from spatial overlapping and from unclear mixing of colors. Currently, these problem are addressed by cutting serial sections and by staining each one with a different antibody and a single colored label. Unfortunately, localized information on a cell-by-cell basis is almost surely lost with this approach. In the absence of larger structures like glands, registration of sequential slices proved to be highly unreliable and often not feasible at all. Multispectral imaging yields single-cell-level multiplexed imaging of standard immunohistochemistry in the same cellular compartment. This technique even works in the presence of a counter stain and each label can be unmixed into separate channels without bleed-through. Image processing in pathology would profit from multispectral imaging even in experiments with a single foreground stain, due to the possibility to accurately separate the specific signal from the background.

Practical suggestions for immunoenzyme double staining procedures for frequently encountered antibody combinations like rabbit/mouse, goat/mouse, mouse/mouse, and rabbit/rabbit are discussed in [70]. The suggested protocols are all suitable for a classical red-brown color combination plus blue nuclear counter stain. Although the red and brown chromogens do not contrast very well visually, they both show a crisp localization and can be unmixed by spectral imaging.

Detection and segmentation of nuclei, glands or other structures constitute a crucial processing steps in various computational pathology frameworks. With the use of supervised machine learning techniques these tasks are often performed by trained classifiers which assign labels to single pixels or small image patches. Naturally one can ask if MSI could improve this classification process and if the additional spectral bands contain additional information? A study conducted by Boucheron et al. [71] set out to answer this question in the scope of routine clinical histopathology imagery. They compared MSI stacks with RGB imagery with the use of several classifier ranging from linear discriminant analysis (LDA) to support vector machines (SVM). For H&E slide the results indicate performance differences of less than 1% using multispectral imagery as opposed to preprocessed RGB imagery. Using only single image bands for classification showed that the single best multispectral band (in the red portion of the spectrum) resulted in a performance increase of 0.57%, compared to the performance of the single best RGB band (red). Principal components analysis (PCA) of the multispectral imagery indicated only two significant image bands, which is not surprising given the presence of two stains. These results [71] indicate that MSI provides minimal additional spectral information than would standard RGB imagery for routine H&E stained histopathology.

Although the results of this study are convincing it has to be noted that only slides with two channels were analyzed. For triple and quadruple staining as described in [70] MSI could still encode additional information which should lead to a higher classification performance. Similar conclusions are drawn in [72], stating that MSI has significant potential to improve segmentation and classification accuracy either by incorporation of features computed across multiple wavelengths or by the addition of spectral unmixing algorithms.

Complementary to supervised learning as described before, Rabinovich et al. [73] proposed unsupervised blind source separation for extracting the contributions of various histological stains to the overall spectral composition throughout a tissue sample. As a preprocessing step all images of the multispectral stack were registered to each other considering affine transformations. Subsequently it was shown that non-negative matrix factorization (NMF) [74] and independent component analysis (ICA) [75] compare favorable to color deconvolution [76]. Along the same lines Begelman et al. [77] advocate principal component analysis (PCA) and blind source separation (BSS) to decompose hyperspectral images into spectrally homogeneous compounds.

In the domain of fluorescence imaging Zimmermann [78] assembled an overview of several source separation methods. The main difficulty stems from the significant overlap of the emission spectra even with the use of fluorescent dyes. To this end, Newberg et al. [79] have conducted a study on more than 3500 images from the Human Protein Atlas [26,27]. They conclude that subcellular locations can be determined with an accuracy of 87.5% by the use of support vector machines and random forests [58,46]. Due to the spread of Type-2 diabetes there is growing interest in pancreatic islet segmentation and cell counting of α and β-cells [80]. An approach which is based on the strategies described in Sections 3.1 and 3.2 is described in [81].

It is an appealing idea to apply source separation techniques not only to multispectral imaging but also to standard RGB images. This approach could be useful for a global staining estimation of the separate channels or as a preprocessing step for training a classifier. Unfortunately, antigen–antibody reactions are not stoichiometric. Hence the intensity/darkness of a stain does not necessarily correlate with the amount of histochemical reaction products. With the exception of Feulgen staining also most histological stains are not stoichiometric. van der Loos [70] also state that the brown DAB reaction product is not a true absorber of light, but a scatterer of light, and has a very broad, featureless spectrum. This optical behavior implies that DAB does not follow the Beer–Lambert law, which describes the linear relationship between the concentration

of a compound and its absorbance, or optical density. As a consequence, darkly stained DAB has a different spectral shape than lightly stained DAB. Therefore attempting to quantify DAB intensity using source separation techniques is not advisable. Contrary to this observation, employing a non-linear convolution algorithm as preprocessing for a linear classifier, e.g. for segmentation could be of benefit.

3.4. Software engineering aspects

One of the earliest approaches for high performance computing in pathology used image matching algorithms based on decision trees to retrieve images from a database [82]. The approach was applied to Gleason grading in prostate cancer. Web-based data management frameworks for TMAs like [83] facilitate not only storage of image data but also storage of experimental and production parameters throughout the TMA workflow.

A crucial demand on software engineering is the ability to scale automated analysis to multiple spots on a TMA slide and even multiple whole microscopy slides. Besides cloud computing one possibility to achieve that goal is grid computing. Foran et al. [84] have demonstrated the feasibility of such a system by using the caGrid infrastructure [85] for grid-enabled deployment of an automated cancer tissue segmentation algorithm for TMAs.

A comprehensive list of open source and public domain software for image analysis in pathology is available at www.computational-pathology.org.

4. Statistics: survival analysis and machine learning in medical statistics

The main thrust of research in computational pathology is to build fully probabilistic models of the complete processing pipelines for histological and medical data. In medical research this nearly always also includes time to event data, where the event is either overall survival, specific survival, event free survival or recurrence free survival of patients. Statistics and machine learning within this scope is defined as *survival analysis*.

4.1. Censoring and descriptive statistics

Most difficulties in survival statistics arise from the fact, that nearly all clinical datasets contain patients with censored survival times. The most common form of censoring is right censored data which means that the death of the patient is not observed during the runtime of the study or that the patient withdrew from the study, e.g. because he moved to another location.

The nonparametric Kaplan–Meier estimator [86] is frequently used to infer the survival function from right censored data. This procedure requires first the survival times to be ordered from the smallest to the largest such that $t_1 \leq t_2 \leq t_3 \leq \ldots \leq t_n$, where t_j is the j th largest unique survival time. The Kaplan–Meier estimate of the survival function is then obtained as

$$\hat{S}(t) = \prod_{j:t_{(j)} \leq t} \left(1 - \frac{d_j}{r_j}\right) \tag{3}$$

where r_j is the number of individuals at risk just before t_j, and d_j is the number of individuals who die at time t_j.

To measure the goodness of separation between two or more groups, the log-rank test (Mantel–Haenszel test) [87] is employed to assesses the null hypothesis that there is no difference in the survival experience of the individuals in the different groups. The test statistic of the log-rank test (LRT) is χ^2 distributed:

$$\hat{\chi}^2 = \frac{\left(\sum_{i=1}^{m}(d_{1i} - \hat{e}_{1i})\right)^2}{\sum_{i=1}^{m}\hat{v}_{1i}} \tag{4}$$

where d_{1i} is the number of deaths in the first group at t_i and $e_{1i} = r_{1j}(d_i/r_i)$ where d_i is the total number of deaths at time $t_{(i)}$, r_j is the total number of individuals at risk at this time, and r_{1i} the number of individuals at risk in the first group. Fig. 10 depicts Kaplan–Meier plots for two subgroups each and the LRT p-values. The associated data is described in detail in Section 5.

4.2. Survival analysis

Survival analysis as a branch of statistics is not restricted to medicine but analyzes time to failure or event data and is also applicable to biology, engineering, economics, etc. Particularly in the context of medical statistics, it is a powerful tool for understanding the effect of patient features on survival patterns within specific groups [88]. A parametric approach to such an analysis involves the estimation of parameters of a probability density function which models time.

In general the distribution of a random variable T (representing time) is defined over the interval $[0, \infty)$. Furthermore, a standard survival function

$$S(t) = 1 - p(T \leq t_0) = 1 - \int_0^{t_0} p(t)\,dt \tag{5}$$

is specified based on the cumulative distribution over T. $S(t)$ models the probability of an individual surviving up to time t_0. The hazard function $h(t)$, the instantaneous rate of failure at time t, is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{p(T = t)}{S(t)}. \tag{6}$$

The model is further extended by considering the effect of covariates X on time via a regression component. In medical statistics, such effects are modeled by Cox's most popular proportionality hazards model [89]:

$$h(t|\boldsymbol{x}) = h_0(t)\,\exp(\boldsymbol{x}^T\boldsymbol{\beta}). \tag{7}$$

$h_0(t)$ is the baseline hazard function, i.e., the chance of instant death given survival till time t, $\boldsymbol{x}$ is the vector of covariates and $\boldsymbol{\beta}$ are the regression coefficients.

4.3. A Bayesian view of survival regression

Bayesian methods are gaining more and more popularity in machine learning in general and in medical statistics in special. A big advantage in survival analysis is the possibility to investigate the posterior distribution of a model. Especially in regularized survival regression models [57] it is possible to derive a posterior distribution also on zero coefficients, i.e. for biomarkers which hence were not included in the model.

A common choice of distribution for modeling time is the Weibull distribution

$$p(t|\alpha_w, \lambda_w) = \alpha_w \frac{1}{\lambda_w} t^{\alpha_w - 1}\, \exp\left(-\frac{1}{\lambda_w} t^{\alpha_w}\right), \tag{8}$$

where α_w and λ_w are the shape and scale parameters, respectively. The Weibull distribution models a variety of survival functions

and hazard rates in a flexible way and it is also the only distribution which captures both the accelerated time model and the proportionality hazards model [90]. Based on the definition (8) and assuming right-censored data [88], the likelihood assumes the form

$$p(\{t_i\}_{i=0}^N|\alpha_w,\lambda_w)=\prod_{i=1}^N\left(\frac{\alpha_w}{\lambda_w}t_i^{\alpha_w-1}\right)^{\delta_i}\exp\left(-\frac{1}{\lambda_w}t_i^{\alpha_w}\right), \quad (9)$$

where $\delta_i = 0$ when the i th observation is censored and 1 otherwise. Furthermore, to model the effect of covariates $\boldsymbol{x}$ on the distribution over time, Cox's proportional hazards model can be applied with covariates having a multiplicative effect on the hazard function.

4.4. Higher order interactions

A reoccurring question in biomedical research projects and especially in TMA analysis studies interactions of markers and their influence on the target. Two modern approaches within the scope of computational pathology try to solve this question from a frequentist [91] and a Bayesian [57] point of view.

The most frequent approach for modeling higher order interactions (like pairs or triplets of features, etc.) instead of modeling just the main effects (individual features) are polynomial expansions of features. For example the vector $\boldsymbol{x}=\{x_1,x_2,x_3\}$ can be expanded up to order 2 as $\boldsymbol{x}'=\{x_1,x_2,x_3,x_1:x_2,x_1:x_3,x_2:x_3,x_1:x_2:x_3\}$. $x_i:x_j$ denotes the concatenation of the feature vectors x_i, x_j. Additional flexibility is built into this model by including a random effect in η in the following manner:

$$\eta=\boldsymbol{x}^t\boldsymbol{\beta}+\epsilon, \quad \text{where } \epsilon\sim N(0,\sigma^2). \quad (10)$$

To include the covariate effect the likelihood of Eq. (9) is modified as follows:

$$p(\{t_i\}_{i=0}^N|\boldsymbol{x}_i,\alpha_w,\lambda_w)=\prod_{i=1}^N\left[\frac{\alpha_w}{\lambda_w}t_i^{\alpha_w-1}\exp(\eta_i)\right]^{\delta_i}\cdot\exp\left(-\frac{1}{\lambda_w}t_i^{\alpha_w}\exp(\eta_i)\right)$$

These kind of models can be seen as enhancement of generalized linear models [92] and are called random-intercept models. For a fully Bayesian treatment of the model, suitable priors have to be defined for the parameters of the model, namely α_w, λ_w, σ and $\boldsymbol{\beta}$. Useful priors for this model are described in [57].

4.5. Mixtures of survival experts

Frequently, sub-groups of patients specified by characteristic survival times have to be identified together with the effects of covariates within each sub-group. Such information might hint at the disease mechanisms. Statistically this grouping is represented by a mixture model or specifically by a mixture of survival experts.

To this end, Rosen and Tanner [93] define a *finite* mixture-of-experts model by maximizing the partial likelihood for the regression coefficients and by using some heuristics to resolve the number of experts in the model. More recently, Ando et al. [94] suggest a maximum likelihood approach to infer the parameters of the model and they use Akaike's information criterion (AIC) to determine the number of mixture components.

A Bayesian version of the mixture model [95] analyzes the model with respect to time but does not capture the effect of covariates. On the other hand the work by Ibrahim et al. [96] performs variable selection based on the covariates but ignores the clustering aspect of the modeling. Similarly, Paserman [97] defines an infinite mixture model but does not include a mixture of experts, hence implicitly assuming that all the covariates are generated by the same distribution with a common shape parameter for the Weibull distribution.

Raman et al. [57] unify the various important elements of this analysis into a Bayesian mixture-of-experts (MOE) framework to model survival time, while capturing the effect of covariates and also dealing with an unknown number of mixing components. To infer the number of experts a Dirichlet process prior on the mixing proportions is applied, which solves the issue of determining the number of mixture components beforehand [98]. Due to the lack of fixed-length sufficient statistics, the Weibull distribution is not part of the exponential family of distributions and hence the regression component, introduced via the proportionality hazards model, is non-standard. Furthermore, the framework of Raman et al. [57] includes sparsity constraints to the regression coefficients in order to determine the key explanatory factors (biomarkers) for each mixture component. Sparseness is achieved by utilizing a Bayesian version of the Group-Lasso [99,100] which is a sparse constraint for grouped coefficients [101].

5. The computational pathology pipeline: a holistic view

This chapter describes a genuine computational pathology project, which has been designed following the principles described in the previous sections. It is an ongoing project in kidney cancer research conducted at the University Hospital Zürich and ETH Zürich. Parts of it were published in [102] and [45], where also algorithmic details of the computational approach can be found.

Fig. 9 depicts a schematic overview of the project subdivided into the three main parts which are discussed in the following.

5.1. Data generation

The data generation process consists of acquiring images of the TMA spots and of annotating these images by pathologists. The TMA spots represent the covariates X in the statistical model and the detection and classification labels for nuclei denote the target variable Y.

The tissue microarray block was generated in a trial study at the University Hospital Zürich. 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. The magnification of 40 × resulted in a per pixel resolution of 0.23 μm. The tissue microarray was tiled into single spots of size 3000 × 3000 pixel, representing one patient each.

Various strategies can be devised to estimate the progression status of cancerous tissue: (i) we could first detect cell nuclei and then classify the detected nuclei as atypical or normal [41]; (ii) the nucleus detection phase could be merged with the normal/atypical classification to simultaneously train a sliding window detector for atypical nuclei only. To this end samples of atypical nuclei were collected using the labeling experiments described in Section 2.3. Voronoi sampling [45] was used to generate a set of negative background patches which are spatially well distributed in the training images. Hence a Voronoi tessellation is created based on the locations of the positive samples and background patches are sampled at the vertices of the Voronoi diagram. In contrast to uniform rejection sampling, using a tessellation has the advantage that the negative samples are concentrated on the area of tissue close to the nuclei and few samples are spent on the homogeneous background. (The algorithm should not be confused with Lloyd's algorithm [103] which is also known as Voronoi iteration.) The result of the data generation process is a labeled set of image patches of size 65 × 65 pixel.

5.2. Image analysis

The image analysis part of the pipeline consists of learning a relational detection forest [45] based on the samples extracted in

the previous step. The feature basis described in Section 3.1 is used to guarantee illumination invariance.

The strong class imbalance in the training set is accounted for by randomly subsampling the background class for each tree of the ensemble. The model parameters are adjusted by optimizing the out of bag (OOB) error [46] and they consist of the number of trees, the maximum tree depth and the number of features sampled at each node in a tree.

For prediction each pixel of a TMA spot is classified by the relation detection forest. This analysis step results in a probability map where the gray value at each position encodes the probability of being the location of a atypical nucleus. Finally, weighted mean shift clustering is conducted with a circular box kernel based on the average radius r of the nuclei in the training set. This process yields the final coordinates of the detected atypical nuclei.

A simple color model is learned to differentiate a stained nucleus from a non-stained nucleus. Located on the labeled nuclei, color histograms are calculated for both classes using all pixels within a radius r of a considered location. A test nucleus is then classified based on the distance to the centroid histograms of both classes. The final staining estimation per patient is achieved by calculating the percentage of stained atypical nuclei.

5.3. Survival statistics

The only objective endpoint in the majority of TMA studies is the prediction of patient survival, i.e., of the number of months a patient survived after disease diagnosis. The experiments described in Section 2.3 document the large disagreement between pathologists for the estimation of staining. Hence, fitting an algorithm to the estimates of a single pathologist or to a consensus voting of a committee of pathologist is not desirable.

To this end the proposed computational pathology framework is validated against the right censored clinical survival data of the 133 ccRCC patients. In addition these results were compared to the estimations of an expert pathologist specialized on renal cell carcinoma. He analyzed all spots in an exceptional thorough manner which required him more than two hours. This time consuming annotation exceeds the standard clinical practice significantly by a factor of 10–20 and, therefore, the results can be viewed as an excellent human estimate for this dataset.

Fig. 10 shows Kaplan–Meier plots of the estimated cumulative survival for the pathologist and the computational pathology framework. The farther the survival estimates of the two groups are separated the better is the group estimation. Quantifying this difference with a log-rank test shows that the proposed framework performs favorable ($p = 0.0113$) to the trained pathologist ($p = 0.0423$) and it can differentiate between the survival expectancy of the two groups of patients.

5.4. Framework properties

The presented computational pathology framework can be characterized by the following properties: (i) *Simplicity:* It can be used in a plug-and-play fashion to train object detectors in near real time for large variety of tasks. (ii) *Novel feature basis:* The introduced relational features are able to capture shape information, they are illumination invariant and extremely fast to evaluate. (iii) *Randomization:* The randomized tree induction algorithm is able to exploit the richness of the intractable large feature space and to take advantage of it by increasing diversity of the ensemble. (iv) *Real world applicability:* The proposed algorithms perform well not only on renal cancer tissue but also in fluorescent imaging of pancreatic islets [81] and in quantifying staining in murine samples [104].

6. Future directions

6.1. Histopathological imaging

One promising research direction in medical image analysis points to online learning and interactive learning of computer vision models. Histopathology covers not only a broad and heterogeneous field but also new biomarkers, antibodies and stainings are developed on a daily basis. To this end, real world applications have to quickly adapt to changing tissue types and staining modalities. Domain experts should be able to train these models in an interactive fashion to accustom novel data. For example, a classifier for object detection can be trained by clicking on novel objects or correcting for false detections.

A necessary prerequisite for research in computational pathology proved to be the scanning of whole slides and TMAs. Huisman et al. [105] describe a fully digital pathology slide archive which has been assembled by high-volume tissue slide scanning. The Peta bytes of histological data which will be available in the near future pose also a number of software engineering challenges, including distributed processing of whole slides and TMAs on clusters or in the cloud, multiprocessor and multicore implementations of analysis algorithms and facilitating real time image processing on GPUs.

6.2. Clinical application and decision support

In today's patient care we observe the trend to integrate pathological diagnoses in web based patient files. Avatar based visualization proved to be helpful not only for medical experts but also for a new generation of patients who are better informed and demand online and appropriately visualized information about their own disease state and treatment procedures.

Furthermore this approach can be extended for decision support by statistical models which are able to utilize this unified view of patients incorporating data from a large variety of clinical sources, e.g. pathology, cytology, radiology, etc.

6.3. Pathology @ home

Real-time, in vivo cancer detection on cellular level appears as a futuristic dream in patient care but could be a reality in a few years. Shin et al. [106] constructed a fiber-optic fluorescence microscope using a consumer-grade camera for in vivo cellular imaging. The fiber-optic fluorescence microscope includes a LED light, an objective lens, a fiber-optic bundle, and a consumer-grade DSLR. The system was used to image an oral cancer cell line, a human tissue specimen and the oral mucosa of a healthy human subject in vivo. The fiber-optic microscope resolved individual nuclei in all specimens and tissues imaged. This capability allowed qualitative and quantitative differences between normal and precancerous or cancerous tissues to be identified. In combination with a computational pathology framework, this technique would enable real time classification of cancerous cells in epithelial tissues. Such a portable and inexpensive system is especially interesting for patient care in low-resource settings like the developing world.

Constructing a microscope for mobile phones defines the future of patient care in remote sites with centralized analysis support. Breslauer et al. [107] built a mobile phone-mounted light microscope and demonstrated its potential for clinical use by imaging sickle and P. falciparum-infected red blood cells in brightfield and M. tuberculosis-infected sputum samples in fluorescence with LED excitation. In all cases the resolution exceeded the critical level that is necessary to detect blood cell and microorganism morphology. This concept could provide an interesting tool for disease diagnosis

and screening, especially in rural areas where laboratory facilities are scarce but mobile phone infrastructure is available.

6.4. Standards and exchange formats

One of the major obstacles for wide spread use of computational pathology is the absence of generally agreed upon standards and exchange formats. This deficit not only handicaps slide processing management and whole slide digital imaging [108], but it also extends to statistical models and analysis software. Standardized exchange formats would support project specific combinations of object detectors, staining estimation algorithms and medical statistics. It would be very much desirable if at least the research community would agree on a few simple interfaces for data and model exchange.

6.5. Further reading

All links and references presented in this review together with software, statistical models and a blog about the topic are available from www.computational-pathology.org.

Acknowledgments

The authors wish to thank Holger Moch and Peter Schraml for their help in conducting the RCC TMA project, Peter Wild and Peter Bode for annotating the medical data and Monika Bieri and Norbert Wey for scanning and tiling the TMA slides. Special thanks also to Volker Roth and Sudhir Raman for valuable discussions. We also acknowledge financial support from the FET program within the EU FP7, under the SIMBAD project (Contract 213250).

References

[1] Grignon D, Eble J, Bonsib S, Moch H. Clear cell renal cell carcinoma, World Health Organization Classification of Tumours. Pathology and genetics of tumours of the urinary system and male genital organs. IARC Press.
[2] Tannapfel A, Hahn H, Katalinic A, Fietkau R, Kühn R, Wittekind C. Prognostic value of ploidy and proliferation markers in renal cell carcinoma. Cancer 1996;77(January (1)):164–71.
[3] e.a. Moch H, Schraml P. High-throughput tissue microarray analysis to evaluate genes uncovered by cdna microarray screening in renal cell carcinoma. Am J Pathol 1999;154(April (4)):981–6.
[4] Sudarshan LW. Genetic basis of cancer of the kidney. Semin Oncol 2006;33(October (5)):544–51.
[5] e.a. Nocito A, Bubendorf L. Microarrays of bladder cancer tissue are highly representative of proliferation index and histological grade. J Pathol 2001;194(July (3)):349–57.
[6] Wild PJ, Fuchs TJ, Stoehr R, Zimmermann D, Frigerio S, Padberg B. Detection of urothelial bladder cancer cells in voided urine can be improved by a combination of cytology and standardized microsatellite analysis. Cancer Epidemiol Biomark Prevent 2009;18(6):1798–806.
[7] Kononen J, Bubendorf L, Kallionimeni A, Barlund M, Schraml P, Leighton S. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med 1998;4(7):844–7.
[8] e.a. Takahashi M, Rhodes DR. Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. Proc Natl Acad Sci USA 2001;98(August (17)):9754–9.
[9] e.a. Young AN, Amin MB. Expression profiling of renal epithelial neoplasms: a method for tumor classification and discovery of diagnostic molecular markers. Am J Pathol May 2001;158(5):1639–51.
[10] Saur SC, Alkadhi H, Desbiolles L, Fuchs TJ, Székely G, Cattin PC. Guided review by frequent itemset mining: additional evidence for plaque detection. Int J Comput Assist Radiol Surg 2009;4(3):263–71.
[11] Saur SC, Alkadhi H, Stolzmann P, Baumüller S, Leschka S, Scheffel H. Effect of reader experience on variability, evaluation time and accuracy of coronary plaque detection with computed tomography coronary angiography. Eur Radiol 2010;20(7):1599–606.
[12] Nattkemper TW, Twellmann T, Ritter H, Schubert W. Human vs. machine: evaluation of fluorescence micrographs. Comput Biol Med 2003;33(1):31–43.
[13] Welinder PPP. Online crowdsourcing: rating annotators and obtaining cost-effective labels. In: CVPR Workshop on Advancing Computer Vision with Humans in the Loop (ACVHL), 2010. Workshop on Advancing Computer Vision with Humans in the Loop (ACVHL).
[14] Eble JN, Sauter G, Epstein JI, Sesterhenn IA. Pathology and genetics of tumours of the urinary system and male genital organs, vol. 7 of classification of tumours. World Health Organization; 2004.
[15] Warfield S, Zou K, Wells III W. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. IEEE Trans Med Imag 2004;23(7):903–21.
[16] Smyth P, Fayyad UM, Burl MC, Perona P, Baldi P. Inferring ground truth from subjective labelling of venus images. In: NIPS. 1994. p. 1085–92.
[17] Raykar VC, Yu S, Zhao LH, Jerebko A, Florin C, Valadez GH. Supervised learning from multiple experts: whom to trust when everyone lies a bit. In: ICML'09: proceedings of the 26th annual international conference on machine learning. 2009. p. 889–96.
[18] Whitehill J, Ruvolo P, fan Wu T, Bergsma J, Movellan J. Whose vote should count more: optimal integration of labels from labelers of unknown expertise. In: Bengio Y, Schuurmans D, Lafferty J, Williams CKI, Culotta A, editors. Advances in neural information processing systems, vol. 22. 2009. p. 2035–43.
[19] Kauppi T, Kamarainen J-K, Lensu L, Kalesnykiene V, Sorri I, Kälviäinen H. Fusion of multiple expert annotations and overall score selection for medical image diagnosis. In: SCIA'09: proceedings of the 16th Scandinavian conference on image analysis. 2009. p. 760–9.
[20] Lugosi G. Learning with an unreliable teacher. Pattern Recogn 1992;25(1):79–87.
[21] Smyth P. Bounds on the mean classification error rate of multiple experts. Pattern Recogn Lett 1996;17(12):1253–7.
[22] Dekel O, Shamir O. Vox populi: collecting high-quality labels from a crowd. In: Proceedings of the 22nd annual conference on learning theory (COLT 2009). 2009. p. 377–86.
[23] Tullock G. Problems of majority voting. J Polit Econ 1959;67(6):571–9.
[24] Downs A. Problems of majority voting: in defense of majority voting. J Polit Econ 1961;69(2):192–9.
[25] Frank A, Asuncion A. UCI machine learning repository; 2010. http://archive.ics.uci.edu/ml.
[26] Berglund L, Björling E, Oksvold P, Fagerberg L, Asplund A, Al-Khalili Szigyarto C. A genecentric Human Protein Atlas for expression profiles based on antibodies. Mol Cell Proteom 2008;7(10):2019–27.
[27] Pontén F, Jirström K, Uhlen M. The Human Protein Atlas—a tool for pathology. J Pathol 2008;216(4):387–93.
[28] Drelie Gelasca E, Obara B, Fedorov D, Kvilekval K, Manjunath B. A biosegmentation benchmark for evaluation of bioimage analysis methods. BMC Bioinform 2009;10(1):368.
[29] Marée R, Denis P, Wehenkel L, Geurts P. Incremental indexing and distributed image search using shared randomized vocabularies. In: Proceedings of the international conference on multimedia information retrieval, MIR'10. 2010. p. 91–100.
[30] Lézoray O, Cardot H. Cooperation of color pixel classification schemes and color watershed: a study for microscopical images. IEEE Trans Image Process 2002;11(7):783–9.
[31] Meurie C, Lézoray O, Charrier C, Elmoataz A. Combination of multiple pixel classifiers for microscopic image segmentation. IJRA (IASTED Int J Robot Automat) 2005;20(2):63–9, special issue on Colour Image Processing and Analysis for Machine Vision, ISSN 0826-8185.
[32] Lézoray O, Elmoataz A, Cardot H. A color object recognition scheme: application to cellular sorting. Mach Vis Appl 2003;14:166–71.
[33] Coelho LP, Shariff A, Murphy RF. Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms. In: ISBI'09: proceedings of the sixth IEEE international conference on symposium on biomedical imaging. Piscataway, NJ, USA: IEEE Press; 2009. p. 518–21.
[34] Maree R, Geurts P, Wehenkel L. Random subwindows and extremely randomized trees for image classification in cell biology. BMC Cell Biol 2007;8(Suppl. 1):S2.
[35] Lehmussola A, Ruusuvuori P, Selinummi J, Huttunen H, Yli-Harja O. Computational framework for simulating fluorescence microscope images with cell populations. IEEE Trans Med Imag 2007;26(7):1010–6.
[36] Monaco JP, Tomaszewski JE, Feldman MD, Hagemann I, Moradi M, Mousavi P. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. Med Image Anal 2010;14(4):617–29.
[37] Gurcan M, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B. Histopathological image analysis: a review. IEEE Rev Biomed Eng 2009;2:147–71.
[38] Rojo M, Bueno G, Slodkowska J. Review of imaging solutions for integrated quantitative immunohistochemistry in the pathology daily practice. Folia Histochem Cytobiol 2009;47(3):349–54.
[39] Soille P. Morphological image analysis: principles and applications. Secaucus, NJ, USA: Springer-Verlag New York, Inc.; 2003.
[40] Lézoray O, Charrier C. Color image segmentation using morphological clustering and fusion with automatic scale selection. Pattern Recogn Lett 2009;30(4):397–406.
[41] Fuchs TJ, Lange T, Wild PJ, Moch H, Buhmann JM. Weakly supervised cell nuclei detection and segmentation on tissue microarrays of renal cell carcinoma. In: Pattern recognition. DAGM 2008, vol. 5096 of lecture notes in computer science. Berlin/Heidelberg: Springer; 2008. p. 173–82.
[42] Halama N, Zoernig I, Spille A, Westphal K, Schirmacher P, Jaeger D. Estimation of immune cell densities in immune cell conglomerates: an approach for high-throughput quantification. PLoS ONE 2009;4(11):e7847.
[43] Arif M, Rajpoot N. Classification of potential nuclei in histology images using shape manifold learning. In: Proceedings of the international conference on machine vision (ICMV'2007). 2007. p. 113–8.

[44] Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X. A method for normalizing histology slides for quantitative analysis. In: ISBI'09: proceedings of the sixth IEEE international conference on symposium on biomedical imaging. Piscataway, NJ, USA: IEEE Press; 2009. p. 1107–10.
[45] Fuchs TJ, Haybaeck J, Wild PJ, Heikenwalder M, Moch H, Aguzzi A. Randomized tree ensembles for object detection in computational pathology. In: ISVC (1), vol. 5875 of lecture notes in computer science. Springer; 2009. p. 367–78.
[46] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
[47] Geman D, d'Avignon C, Naiman DQ, Winslow RL. Classifying gene expression profiles from pairwise mrna comparisons. Stat Appl Genet Mol Biol 2004;3(1):19.
[48] Viola P, Jones M, I-511–I-518 Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol. 1. 2001.
[49] Freund Y, Schapire R. Experiments with a new boosting algorithm. In: Machine learning: proceedings of the thirteenth international conference. 1996. p. 148–56.
[50] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn 2006;63(1):3–42.
[51] Fuchs TJ, Buhmann JM. Inter-active learning of randomized tree ensembles for object detection. In: ICCV workshop on on-line learning for computer vision, 2009. IEEE; 2009.
[52] Vapnik V. Statistical Learning Theory. New York, NY, USA: Wiley-Interscience; 1998.
[53] Blum A. On-line algorithms in machine learning. In: Online Algorithms; 1996. p. 306–25.
[54] Nguyen TT, Grabner H, Bischof H, Gruber B. On-line boosting for car detection from aerial images. In: RIVF. IEEE; 2007. p. 87–95.
[55] Celik H, Hanjalic A, Hendriks E, Boughorbel S. Online training of object detectors from unlabeled surveillance video. In: IEEE computer society conference on computer vision and pattern recognition workshops, 2008. CVPRW'08. 2008. p. 1–7.
[56] Grabner H, Leistner C, Bischof H. Semi-supervised on-line boosting for robust tracking. In: ECCV 2008;1:234–47.
[57] Raman S, Fuchs T, Wild P, Dahl E, Buhmann J, Roth V. Infinite mixture-of-experts model for sparse survival regression with application to breast cancer. BMC Bioinform 2010;11(Suppl 8):8, doi:10.1186/1471-2105-11-S8-S8.http://www.biomedcentral.com/1471-2105/11/S8/S8.
[58] Amit Y, Geman D. Shape quantization and recognition with randomized trees. Neural Comput 1997;9:1545–88.
[59] Oza NC. Online ensemble learning. Ph.D. Thesis. The University of California, Berkeley, CA; September 2001.
[60] Fern A, Givan R. Online ensemble learning: an empirical study. Mach Learn 2003;53(1–2):71–109.
[61] Elgawi OH. Online random forests based on corrfs and corrbe. In: IEEE computer society conference on computer vision and pattern recognition workshops, 2008. CVPRW'08. 2008. p. 1–7.
[62] Utgoff PE. Incremental induction of decision trees. Mach Learn 1989;4:161–86.
[63] Utgoff PE. An improved algorithm for incremental induction of decision trees. In: ICML. 1994. p. 318–25.
[64] Kalles D, Morris T. Efficient incremental induction of decision trees. Mach Learn 1996;24(3):231–42.
[65] Pfahringer B, Holmes G, Kirkby R. New options for hoeffding trees. In: Orgun MA, Thornton J, editors. Australian conference on artificial intelligence, vol. 4830 of lecture notes in computer science. Springer; 2007. p. 90–9.
[66] Littlestone N, p. 285—318 Learning quickly when irrelevant attributes abound: a new linear-threshold algorithm. In: Machine learning; 1988.
[67] Littlestone N, Warmuth MK. The weighted majority algorithm. In: FOCS. 1989. p. 256–61.
[68] Hayes-Roth F, Waterman DA, Lenat DB. Building expert systems. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.; 1983.
[69] Levenson RM, Mansfield JR. Multispectral imaging in biology and medicine: slices of life. Cytometry Part A 2006;69A(8):748–58.
[70] van der Loos CM. Multiple immunoenzyme staining: methods and visualizations for the observation with spectral imaging. J Histochem Cytochem 2008;56(4):313–28.
[71] Boucheron L, Bi Z, Harvey N, Manjunath B, Rimm D. Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery. BMC Cell Biol 2007;8(Suppl 1):8.
[72] Cukierski WJ, Qi X, Foran DJ. Moving beyond color: the case for multispectral imaging in brightfield pathology. In: ISBI'09: proceedings of the sixth IEEE international conference on symposium on biomedical imaging. Piscataway, NJ, USA: IEEE Press; 2009. p. 1111–4.
[73] Rabinovich A, Agarwal S, Laris CA, Price JH, Belongie S. Unsupervised color decomposition of histologically stained tissue samples. In: Advances in neural information processing systems. MIT Press; 2003. p. 667–73.
[74] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999;401(6755):788–91.
[75] Hyvarinen A. Independent component analysis. Neural Computing Surveys 2.
[76] Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. Analyt Quant Cytol Histol 2001;23:291–9.
[77] Begelman G, Zibulevsky M, Rivlin E, Kolatt T. Blind decomposition of transmission light microscopic hyperspectral cube using sparse representation. IEEE Trans Med Imag 2009;28(8):1317–24.
[78] Zimmermann T. Spectral imaging and linear unmixing in light microscopy. In: Rietdorf J, editor. Microscopy techniques, vol. 95 of Advances in Biochemical Engineering/Biotechnology. Berlin/Heidelberg: Springer; 2005. p. 245–65.
[79] Newberg JY, Li J, Rao A, Pontén F, Uhlén M, Lundberg E. Automated analysis of Human Protein Atlas immunofluorescence images. In: ISBI'09: proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging. Piscataway, NJ, USA: IEEE Press; 2009. p. 1023–6.
[80] Herold J, Zhou L, Abouna S, Pelengaris S, Bernard Alper Epstein D, Khan M, et al. Integrating semantic annotation and information visualization for the analysis of multichannel fluorescence micrographs from pancreatic tissue. Comp Med Imag Graph 2010;34(6):446–52.
[81] Floros XE, Fuchs TJ, Rechsteiner MP, Spinas G, Moch H, Buhmann JM. Graph-based pancreatic islet segmentation for early type 2 diabetes mellitus on histopathological tissue. In: MICCAI 2009;1:633–40.
[82] Wetzel A. Computational aspects of pathology image classification and retrieval. J Supercomput 1997;11(3):279–93.
[83] Thallinger G, Baumgartner K, Pirklbauer M, Uray M, Pauritsch E, Mehes G. Tamee: data management and analysis for tissue microarrays. BMC Bioinform 2007;8(1):81.
[84] Foran DJ, Yang L, Tuzel O, Chen W, Hu J, Kurc TM. A cagrid-enabled, learning based image segmentation method for histopathology specimens. In: ISBI'09: proceedings of the sixth IEEE international conference on symposium on biomedical imaging. Piscataway, NJ, USA: IEEE Press; 2009. p. 1306–9.
[85] Oster S, Langella S, Hastings S, Ervin D, Madduri R, Phillips J. caGrid 1.0: an enterprise grid infrastructure for biomedical research. J Am Med Inform Assoc 2008;15(2):138–49.
[86] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53(282):457–81.
[87] Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;4(22):719–48.
[88] Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. New York: Springer-Verlag Inc.; 1997.
[89] Cox DR. Regression models and life-tables. J R Stat Soc Ser B (Methodol) 1972;34(2):187–220.
[90] Joseph DS, Ibrahim G, Chen M-H. Bayesian survival analysis. New York: Springer-Verlag Inc.; 2001.
[91] Dahinden C, Ingold B, Wild P, Boysen G, Luu V-D, Montani M. Mining tissue microarray data to uncover combinations of biomarker expression patterns that improve intermediate staging and grading of clear cell renal cell cancer. Clin Cancer Res 2010;16(1):88–98.
[92] McCullaghand P, Nelder J. Generalized linear models. Chapman & Hall; 1983.
[93] Rosen O, Tanner M. Mixtures of proportional hazards regression models. Stat Med 1999;18:1119–31.
[94] Ando T, Imoto S, Miyano S. Kernel mixture survival models for identifying cancer subtypes, predicting patient's cancer types and survival probabilities; 2004.
[95] Kottas A. Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. J Stat Plann Inference 2006;136(3):578–96.
[96] Ibrahim JG, Hui Chen M, Maceachern SN. Bayesian variable selection for proportional hazards models; 1996.
[97] Paserman MD. Bayesian inference for duration data with unobserved and unknown heterogeneity: Monte Carlo evidence and an application. Tech. rep. Institute for the Study of Labor (IZA); January 2004.
[98] Rasmussen CE, Ghahramani Z. Infinite mixtures of Gaussian process experts. In: Advances in neural information processing systems, vol. 14. MIT Press; 2001. p. 881–8.
[99] Raman S, Fuchs TJ, Wild PJ, Dahl E, Roth V. The Bayesian group-lasso for analyzing contingency tables. In: ICML'09: proceedings of the 26th annual international conference on machine learning. 2009. p. 881–8.
[100] Raman S, Roth V. Sparse Bayesian regression for grouped variables in generalized linear models. In: Proceedings of the 31st DAGM symposium on pattern recognition. 2009. p. 242–51.
[101] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc B 2006:49–67.
[102] Fuchs TJ, Wild PJ, Moch H, Buhmann JM. Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients. In: Medical image computing and computer-assisted intervention. MICCAI 2008, vol. 5242 of lecture notes in computer science. Berlin/Heidelberg: Springer; 2008. p. 1–8.
[103] Lloyd SP. Least squares quantization in pcm. IEEE Trans Inform Theory 1982;28(2):129–37.
[104] Bettermann K, Vucur M, Haybaeck J, Koppe C, Janssen J, Heymann F. Tak1 suppresses a nemo-dependent but nf-[kappa]b-independent pathway to liver cancer. Cancer Cell 2010;17(5):481–96.
[105] Huisman A, Looijen A, van den Brink SM, van Diest PJ. Creation of a fully digital pathology slide archive by high-volume tissue slide scanning. Human Pathol 2010;41(5):751–7.
[106] Shin D, Pierce MC, Gillenwater AM, Williams MD, Richards-Kortum RR. A fiber-optic fluorescence microscope using a consumer-grade digital camera for *in vivo* cellular imaging. PLoS ONE 2010;5(6):e11218.
[107] Breslauer DN, Maamari RN, Switz NA, Lam WA, Fletcher DA. Mobile phone based clinical microscopy for global health applications. PLoS ONE 2009;4(7):e6320.
[108] Daniel C, Garca Rojo M, Bourquard K, Henin D, Schrader T, Mea VD. Standards to support information systems integration in anatomic pathology. Arch Pathol Lab Med 2009;133(11):1841–9.