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.