



For reprint orders, please contact: reprints@futuremedicine.com

Digital pathology image analysis: opportunities and challenges

“We stand at the threshold of an era when predictive, preventive and personalized medicine will transform medicine by decreasing morbidity in cancer. We believe this transformation will be driven by the integration of multiscale heterogeneous data.”

The digital pathologist & computerized image analysis of histopathology

Over the last decade, the nature of diagnostic healthcare has changed rapidly owing to an explosion in the availability of patient data for disease diagnosis. Traditional methods of analysis of cancer samples were limited to a few variables, usually stage, grade and the measurement of a few clinical markers, such as estrogen receptor, progesterone receptor, HER2 for breast cancer and prostate-specific antigen for prostate cancer (CaP). The pathologist was trained to synthesize this information into a diagnosis that would help the clinician determine the best course of therapy. These data were also used to try to understand the molecular basis of cancer with the goal of improving therapy.

With the recent advent and cost-effectiveness of whole-slide digital scanners, tissue histopathology slides can now be digitized and stored in digital image form. With the availability and analysis of a much larger set of variables combined with sophisticated imaging and analysis techniques, the traditional paradigm of a pathologist and a microscopy could rapidly be replaced with a digital pathologist relying on a large flat screen panel to view and rapidly analyze digitized tissue sections.

Computer-aided diagnosis of histopathology

Over the past decade, dramatic increases in computational power and improvement in image analysis algorithms have allowed the development of powerful computer-assisted analytical approaches to biomedical data. Just as with digital radiology over two decades ago, digitized tissue histopathology has now become amenable to the application of computerized image analysis and machine-learning techniques for accurate

diagnosis. In the context of CaP, for example, of the approximately 1 million biopsies performed in the USA every year, only 20% are found to be positive for cancer. This implies that pathologists are spending a large fraction of their time looking at benign tissue, which in most cases is easily distinguishable from cancer [1,2]. This represents a huge waste of time and resources that might be better spent analyzing patients who actually have CaP, or to focus on the cases where the disease is difficult to identify/classify or presents with nonstandard features. Consequently, several researchers have begun to develop computer-aided diagnosis methods by applying image processing and computer vision techniques to try and identify spatial extent and location of diseases such as breast cancer [3–9], CaP [10–17], neuroblastomas and meningiomas [18–21] on digitized tissue sections.

One of the principal challenges in analysis of digital histopathology data is the enormous density of data that the algorithms have to contend with, compared with radiological and other imaging modalities. For instance, the largest radiological datasets obtained on a routine basis are high-resolution chest CT scans comprising approximately $512 \times 512 \times 512$ spatial elements or approximately 134 million voxels. A single core of prostate biopsy tissue digitized at $40\times$ resolution is approximately $15,000 \times 15,000$ elements or approximately 225 million pixels. To put this in context, a single prostate biopsy procedure can comprise anywhere between 12 and 20 biopsy samples or approximately 2.5–4 billion pixels of data generated per patient study. Thus, unlike computer-aided detection (CAD) algorithms previously proposed for radiology, histopathology CAD algorithms are typically constructed within a multiresolution framework [22] in order for them to be rapid, efficient and accurate.



Anant Madabhushi

Laboratory for Computational Imaging and Bioinformatics (LCIB), Department of Biomedical Engineering, Rutgers University, NJ, USA
Tel.: +1 732 445 4500 ext. 6213;
Fax: +1 732 445 3753;
anantm@rci.rutgers.edu

Computer-aided prognosis

A second important role of computerized image analysis of digital pathology is to identify prognostic markers and to predict disease outcome and survival. For instance, in both breast cancer [23–25] and CaP [26–28], cancer grade is known to be highly correlated to patient outcome and long-term survival. One of the issues with grade determination by a pathologist is the high degree of inter- and intra-observer variability [25,29–32]. Since pathologist grade is reflected in tissue architecture and nuclear arrangement, graph-based [4,5,7,33,34] algorithms have been proposed to quantitatively characterize spatial arrangement and distribution of histological structures such as cancer nuclei, lymphocytes and glands.

.....
“...of the approximately 1 million biopsies performed in the USA every year, only 20% are found to be positive for cancer. This implies that pathologists are spending a large fraction of their time looking at benign tissue, which in most cases is easily distinguishable from cancer...”

It is conceivable that these image-based predictors may in the future become powerful and accurate enough to be able to rival more expensive molecular prognostic assays in predicting disease outcome. For instance, for estrogen receptor-positive breast cancers, our group has been developing an image-based risk score predictor that on a small cohort of data appears to perform as well as a commercial molecular gene expression assay called Oncotype DX® [5,16] in predicting patient outcome.

Computer-aided theragnosis

It has always been accepted that cancer is a complex disease that we do not yet fully understand. In the clinic, the same treatment applied to two patients with diseases that look very similar have vastly different outcomes. A part of this difference is undoubtedly patient specific, but a part must also be a result of our limited understanding of the relationship between disease progression and clinical presentation. There is a consensus among clinicians and researchers that a more detailed approach, using computerized imaging techniques to better understand tumor morphology, combined with the classification of diseases into more meaningful molecular subtypes, will lead to better patient care and more effective therapeutics. The variables that can be used in such an analysis are the

molecular features of a tumor (as measured by gene-expression profiling or real-time PCR and FISH), results from the imaging of the tumor cellular architecture and microenvironment (as captured in histological imaging), the tumor 3D tissue architecture and vascularization (as measured by dynamic contrast-enhanced MRI) and its metabolic features (as seen by metabolic or functional imaging modalities e.g., magnetic resonance spectroscopy or PET) [35].

While digital pathology offers very interesting, highly dense data, one of the exciting challenges in the future will be in the area of multimodal data fusion for making therapy recommendations (theragnosis), especially as it pertains to personalized medicine. For instance, our group [36] has been exploring the correlation and integration of protein expression and histological image measurements to develop a combined classifier to predict which CaP patients will have disease recurrence following therapy.

Role of the pathologist in the digital age

While image analysis methods for digital pathology are rapidly finding application in the clinic, both imaging, computer scientists and pathologists alike need to appreciate that the primary purpose of these tools is to complement the role of the pathologist. They will not in the short or medium term be able to replace the vast domain of expertise that a pathologist brings to the table; a lesson that we can appreciate from radiology where the availability of commercial CAD systems over the last two decades has not in any way diminished the role of the radiologist.

.....
“We are living in an exciting time when disease diagnostics and treatment are becoming more accurate and patient specific. Computerized imaging methods are beginning to assist ... in making an accurate diagnosis of disease...”

The vast majority of histopathology image analysis researchers are computer vision researchers. As such, it is important to maintain a constant collaboration with clinical and research pathologists throughout the research process. There are unique challenges to analysis of histopathology imagery, particularly in the performances required for eventual use of the technique in a clinical setting. It is the pathologist who can best provide the feedback on the performance of the system, as well as suggesting new avenues of research that would provide

beneficial information to the pathologist community. Additionally, it is the pathologist that is best equipped to interpret the analysis results in light of underlying biological mechanisms which, in turn, may lead to new research ideas.

Looking to the future

We are living in an exciting time when disease diagnostics and treatment are becoming more accurate and patient specific. Computerized imaging methods are beginning to assist the pathologist and radiologist in making an accurate diagnosis of disease and identify morphological features correlated with prognosis. Molecular profiling of disease promises to help the clinician understand the underlying biology of the disease and suggest new and more effective therapeutics. We stand at the threshold of an era when predictive, preventive and personalized medicine will transform medicine by decreasing morbidity in cancer. We believe this transformation will be driven by the integration of multiscale heterogeneous data [9,36]. The goal of our research and the research of many other scientists is aimed at a future when disease diagnostics will involve the quantitative integration of multiple sources of diagnostic data, including genomic, imaging, proteomic and metabolic data acquired across multiple scales/resolutions that can distinguish between individuals or between subtle variations of the same disease to guide therapy. Quantitative

cross-modal data integration will also allow disease prognostics, enabling physicians to predict susceptibility to a specific disease as well as disease outcome and survival. Finally, the analysis will provide theragnostics; the ability to predict how an individual will react to various treatments. Such a theragnostic profile would be a synthesis of various biomarkers and imaging tests from different levels of the biological hierarchy. It would be used as the 'signature' of an individual patient, useful in predicting her/his response to drug treatment. A collection of these profiles, followed up over time, would provide insights into the disease process and be useful for improvements in developing future treatment options.

Financial & competing interests disclosure

This work was made possible via grants from the Wallace H Coulter Foundation, New Jersey Commission on Cancer Research, National Cancer Institute Grant Nos. R01CA136535-01, ARRA-NCI-3 R21CA127186-02SI, R21CA127186-01 and R03CA128081-01, the Society for Imaging Informatics in Medicine SIIM, The Cancer Institute of New Jersey and the Life Science Commercialization Award from Rutgers University. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Bibliography

- Allsbrook WC Jr, Mangold KA, Johnson MH, Lane RB, Lane CG, Epstein JI: Interobserver reproducibility of Gleason grading of prostatic carcinoma: general pathologist. *Hum. Pathol.* 32(1), 81–88 (2001).
- Epstein JI, Allsbrook WC Jr, Amin MB, Egevad LL: Update on the Gleason grading system for prostate cancer: results of an international consensus conference of urologic pathologists. *Adv. Anat. Pathol.* 13(1), 57–59 (2006).
- Anderson NH, Hamilton PW, Bartels PH, Thompson D, Montironi R, Sloan JM: Computerized scene segmentation for the discrimination of architectural features in ductal proliferative lesions of the breast. *J. Pathol.* 181, 374–380 (1997).
- Basavanahally A, Agner A, Alexe G, Ganesan S, Bhanot G, Madabhushi A: Manifold learning with graph-based features for identifying extent of lymphocytic infiltration from high grade breast cancer histology. Presented at: *MMBIA Workshop in Conjunction with MICCAI 2008*. New York, NY, USA, 10 September 2008.
- Basavanahally A, Xu J, Ganesan S, Madabhushi A: Computer-aided prognosis (CAP) of ER+ breast cancer histopathology and correlating survival outcome with oncotype Dx assay. Presented at: *IEEE International Symposium on Biomedical Imaging (ISBI)*. Boston, MA, USA, 28 June–1 July 2009.
- Dey P, Mohanty S: Fractal dimensions of breast lesions on cytology smears. *Advances Technol.* 29, 85–86 (2003).
- Doyle S, Agner S, Madabhushi A, Feldman M, Tomaszewski J: Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. Presented at: *IEEE International Symposium on Biomedical Imaging*. Paris, France, 14–17 May 2008.
- Petushi S, Garcia FU, Haber MM, Katsinis C, Tozeren A: Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC Med. Imaging* 6, 14 (2006).
- Chappelow J, Madabhushi A, Rosen M, Tomaszewski J, Feldman M: Multimodal image registration of *ex vivo* 4 Telsa prostate MRI with whole mount histology for cancer detection. *SPIE Med. Imaging* 6512(1), S1–S12 (2007).
- Brawer MK, Bigler SA, Deering RE: Quantitative morphometric analysis of the microcirculation in prostate carcinoma. *J. Cell. Biochem. Suppl.* 16H, 62–64 (1992).
- Monaco J, Tomaszewski J, Feldman M *et al.*: Detection of prostate cancer from whole-mount histology images using Markov random fields. Presented at: *Microscopic Image Analysis and Applications in Biology (MIAAB)*. New York, NY, USA, 5–6 September 2008.
- Naik S, Doyle S, Madabhushi A, Tomaszewski J, Feldman M: Automated gland segmentation and Gleason grading of prostate histology by integrating low-

- high-level and domain specific information. Presented at: *Workshop on Microscopic Image Analysis with Applications in Biology*. Piscataway, NJ, USA, 21 September 2007.
- 13 Doyle S, Madabhushi A, Feldman M, Tomaszewski J: A boosting cascade for automated detection of prostate cancer from digitized histology. In: *Medical Image Computing and Computer Assisted Intervention*. Springer Verlag, Copenhagen, Denmark, 504–511 (2006).
 - 14 Doyle S, Rodriguez C, Madabhushi A, Tomaszewski J, Feldman M: Detecting prostatic adenocarcinoma from digitized histology using a multi-scale, hierarchical classification approach. Presented at: *IEEE Engineering in Medicine and Biology Conference*. New York, NY, USA, 30 August–3 September 2006.
 - 15 Tabesh A, Teverovskiy M, Pang HY *et al.*: Multifeature prostate cancer diagnosis and Gleason grading of histological images. *IEEE Trans. Med. Imaging* 26(10), 1366–1378 (2007).
 - 16 Lexe G, Monaco J, Doyle S *et al.*: Towards improved cancer diagnosis and prognosis using analysis of gene expression data and computer aided imaging. *Exp. Biol. Med. (Maywood)* 234(8), 860–879 (2009).
 - 17 Jafari-Khouzani K, Soltanian-Zadeh H: Multiwavelet grading of pathological images of prostate. *IEEE Trans. Biomed. Eng.* 50(6), 697–704 (2003).
 - 18 Ruiz A, Sertel O, Ujaldon M, Catalyurek U, Saltz J, Gurcan M: Stroma classification for neuroblastoma on graphics processors. *Int. J. Data Min. Bioform.* 3(3), 280–298 (2009).
 - 19 Qureshi H, Sertel O, Rajpoot N, Wilson R, Gurcan M: Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. *Med. Image Comput. Comput. Assist. Interv. Int. Conf. Med. Image Comput. Comput. Assist. Interv.* 11(Pt 2), 196–204 (2008).
 - 20 Gurcan MN, Kong J, Sertel O, Cambazoglu BB, Saltz J, Catalyurek U: Computerized pathological image analysis for neuroblastoma prognosis. *AMIA Annu. Symp. Proc.* 11, 304–308 (2007).
 - 21 Kong J, Sertel O, Boyer KL, Saltz JH, Gurcan MN, Shimada H: Computer-assisted grading of neuroblastic differentiation. *Arch. Pathol. Lab. Med.* 132(6), 903–904; author reply 904 (2008).
 - 22 Boccignone G, Napolitano P, Caggiano V, Ferraro M: A multiresolution diffused expectation-maximization algorithm for medical image segmentation. *Comput. Biol. Med.* 37(1), 83–96 (2007).
 - 23 Anderson TJ, Sufi F, Ellis IO, Sloane JP, Moss S: Implications of pathologist concordance for breast cancer assessments in mammography screening from age 40 years. *Hum. Pathol.* 33(3), 365–371 (2002).
 - 24 Contesso G, Mouriesse H, Friedman S, Genin J, Sarrazin D, Rouesse J: The importance of histologic grade in long-term prognosis of breast cancer: a study of 1,010 patients, uniformly treated at the Institut Gustave-Roussy. *J. Clin. Oncol.* 5(9), 1378–1386 (1987).
 - 25 Dalton LW, Pinder SE, Elston CE *et al.*: Histologic grading of breast cancer: linkage of patient outcome with level of pathologist agreement. *Mod. Pathol.* 13(7), 730–735 (2000).
 - 26 Andrén O, Fall K, Franzén L, Andersson SO, Johansson JE, Rubin MA: How well does the Gleason score predict prostate cancer death? A 20-year followup of a population based cohort in Sweden. *J. Urol.* 175(4), 1337–1340 (2006).
 - 27 Bostwick DG: Grading prostate cancer. *Am. J. Clin. Pathol.* 102(4 Suppl. 1), S38–S56 (1994).
 - 28 Bostwick DG: Gleason grading of prostatic needle biopsies. Correlation with grade in 316 matched prostatectomies. *Am. J. Surg. Pathol.* 18(8), 796–803 (1994).
 - 29 Burchardt M, Engers R, Müller M *et al.*: Interobserver reproducibility of Gleason grading: evaluation using prostate cancer tissue microarrays. *J. Cancer Res. Clin. Oncol.* 134(10), 1071–1078 (2008).
 - 30 Engers R: Reproducibility and reliability of tumor grading in urological neoplasms. *World J. Urol.* 25(6), 595–605 (2007).
 - 31 Fukagai T, Namiki T, Namiki H, Carlile RG, Shimada M, Yoshida H: Discrepancies between Gleason scores of needle biopsy and radical prostatectomy specimens. *Pathol. Int.* 51(5), 364–370 (2001).
 - 32 Glaessgen A, Hamberg H, Pihl CG, Sundelin B, Nilsson B, Egevad L: Interobserver reproducibility of percent Gleason grade 4/5 in total prostatectomy specimens. *J. Urol.* 168(5), 2006–2010 (2002).
 - 33 Toussaint GT: The relative neighbourhood graph of a finite planar set. *Pattern Recognit.* 12, 261–268 (1980).
 - 34 Doyle S, Hwang M, Shah K, Madabhushi A, Tomaszewski J, Feldman M: Automated grading of prostate cancer using architectural and textural image features. Presented at: *International Symposium on Biomedical Imaging*. Washington, DC, USA, 12–15 April 2007.
 - 35 Lenkinski RE, Bloch BN, Liu F *et al.*: An illustration of the potential for mapping MRI/MRS parameters with genetic over-expression profiles in human prostate cancer. *MAGMA* 21(6), 411–421 (2008).
 - 36 Lee G, Monaco J, Doyle S *et al.*: A knowledge representation framework for integrating multi-modal, multi-scale imaging and non-imaging data: predicting prostate cancer recurrence by fusing mass spectrometry and histology. Presented at: *IEEE International Symposium on Biomedical Imaging (ISBI)*. Boston, MA, USA, 28 June–1 July 2009.