

Editorial

Tryggo: Old nurse for truth

The real truth about ground truth: New insights into the challenges of generating ground truth maps for WSI CAD algorithm evaluation

Jason D. Hipp, Steven C. Smith, Jeffrey Sica, David Lucas, Jennifer A. Hipp, Lakshmi P. Kunju, Ulysses J. Balis

Department of Pathology, University of Michigan Health System, M4233A Medical Science I, 1301 Catherine St. Ann Arbor, Michigan 48109-0602

E-mail: *Ulysses J. Balis - ulysses@med.umich.edu

*Corresponding author

Received: 29 October 11

Accepted: 25 January 12

Published: 16 March 12

This article may be cited as:

Hipp JD, Smith SC, Sica J, Lucas D, Hipp JA, Kunju LP, Balis UJ. Tryggo: Old nurse for truth - The real truth about ground truth: New insights into the challenges of generating ground truth maps for WSI CAD algorithm evaluation. *J Pathol Inform* 2012;3:8.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2012/3/1/8/93890>

Copyright: © 2012 Hipp JD. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

BACKGROUND ON GROUND TRUTH MAPS

Tryggo: Old Nurse for Truth

Ground truth mapping has its origins in the computer science pattern recognition and machine vision fields and is a major activity associated with remote sensing (i.e. satellite images). Additionally, ground truth maps not infrequently find use in the following plurality of settings: motion video road detection and tracking, motion tasks, geoscience applications, sensor data and document analysis. Ground truth maps are needed to perform objective analysis and comparative evaluation of image analysis algorithms^[1] and are “defined as a representation of the agreed correct result of the ideal layout analysis method (i.e., the result of the method that, if it existed, would put an end to the research problem).”^[2] Operationally, such ground truth maps may be regarded as the “gold standard” by which results of other algorithms are compared.

Prior studies have discussed the issue of algorithm performance and can be grouped accordingly:^[3]

Comparative

Here an algorithm may be compared with others that attempt to address the same image processing task or its performance may be compared to ‘ground truth,’ or perhaps to human performance.

Analytical

The theory behind the algorithm is examined to try to

determine the limits to its operation. The computational complexity may be derived, or theoretical optimality may be determined under certain constraints. Frequently, the approach makes use of simplified input data to make the analysis feasible.

Performance

The manner in which the algorithm actually performs on test data is measured and execution times with different parameters may be reported.

Appropriateness to Task

The algorithm is shown in the context of a particular application, and the constraints of the task are used to justify the selection of the particular algorithm. The performance of the task as a whole is taken as the evaluation of the algorithm.^[3]

The most agreed-upon performance evaluation is longevity and public acceptance.^[3]

Access this article online

Quick Response Code:



Website:

www.jpathinformatics.org

DOI: 10.4103/2153-3539.93890

Within the collective medical specialties, ground truth maps first enjoyed utility in radiology, with them being employed as a performance metric in computer aided diagnosis (CAD) algorithm development applications.^[2,4-9] In radiology, ground truth maps are typically made by annotating digital images (MRI, CT, CR, etc.) to accurately represent the likelihood that pixels at each point correspond to a genuine abnormality or feature of interest. Additionally, ground truth maps can assist in generating difference estimates for boundary regions and in providing a central, spatially-based referential diagnostic map by which multiple annotations from different radiologists can be compared. By use of such comparisons, it becomes possible to realize a mechanistic approach that identifies interviewer variability and the associated possible subjectivity intrinsic to such annotated data.^[10]

Recent advances in high-resolution optical scanning technology have made rapid digital reproduction of whole slide imaging (WSI) data sets from glass slides a reality, not only for archival purposes but also at speeds compatible with diagnostic workflow in pathology. These technologies not only open new avenues for consultation, quality control, and telepathology, but also beg for research and development of computer aided diagnostic (CAD) technologies^[11-17] as adjunctive tests for surgical pathologists. Use of CAD has the potential to improve the practice of pathology in various ways^[14,18] by allowing the pathologist to exploit the particular advantages of artificial intelligence, including standardization and quantitation, in a fashion that is complementary to expert human diagnostic ability.

Operationally, quantitative assessment and comparisons of CAD algorithm performance involves comparison of distributions of an algorithm's predictions (in the form of scores) between ground truth positive (i.e., diagnosis present) areas of a WSI and ground truth negative areas (i.e., diagnosis not present). A nonparametric test statistic capable of quantifying the overall ability of such an algorithm to discriminate between ground truth positive and negative is the receiver operating characteristic (ROC), wherein performance is summarized by the area-under-the-curve (AUC).^[16,19,20] For a more detailed explanation regarding the creation and application of ROC curves and AUCs, the reader is directed to the provided references.^[21-23] These types of analyses require the use of ground truth maps. As described at the FDA Radiological Devices Panel meeting, March 2008 briefing package, "Ground truth determination includes whether or not disease is present within a patient as well as the precise location and local extent of disease" (www.fda.gov). More specifically, there are different types of ground truth constructs such as markup tags containing meta-data on a per-image basis such as "high lymphocytic infiltration," for diagnoses and grading such as prostatic adenocarcinoma Gleason Grade 3, and pixel-wise

classification (which requires the "painting" or "circling" of the lesional feature, which is often used for pathology images and for prostate MRI images). An extensive body of references that demonstrates the application of such ground truths can be found at Dr. Madabhushi's website (<http://lcib.rutgers.edu/lcib/publications>).

In our previous editorial, we highlighted the differences between radiological images and pathology images.^[24] Here, we will discuss the importance and challenges in defining ground truth for developing and assessing pathology CAD algorithms compared to "circling" as performed by a radiologist. In addition, a ground truth map rendered from a region of a pathology image must correspond to the intended specific surface area of interest, as relevant to the intended role of the algorithm under consideration.

Pathology Ground Truth Use Cases

Radiological images differ from those of pathology subject matter, owing to the former's limited spatial resolution for most diagnostic modalities, compared to WSI scans of digital slides. In the setting of such diminished resolution (which can be considered as being from the "meso-scale," and not the micro-scale of histology), many disease processes appear similar or even indistinguishable.^[24] Thus, as is standard practice in composition of radiologic reports, clinicians are provided with a context-dependent expert *impression* rather than a specific diagnosis.^[24] Because of this low to intermediate image resolution (certainly relative to histology) and relative homogeneity throughout the lesion, ground truths are often defined by annotating (circling) the entire lesion.

WSI scanners result in digital slides that are of extremely high resolution and possessing of color information (with both attributes enabling subnuclear feature detection). This allows for high fidelity digital representation of the complex constellation of cellular and architectural features, as interpreted by the pathologist. Large or small lesions of clinical interest are reduced to the cellular and subcellular level, where differences not discernible on the macro- and meso- scales (physical exam or radiography), respectively, are used to generate differential and final diagnoses. Entities may range from homogenous, discrete masses to ill-defined, rare, or infiltrative cells that can be difficult to identify and may admix with non-lesional tissue. In addition, they can consist of a multiple subjective features, that are not only difficult to identify but are highly variable and subjective, even among the world's experts.^[25] Currently, most pathology ground truth maps are made with digital slide viewers, which have an annotation "pen" that enables the circling of features of interest, akin to the methodology of radiologists.

One may intuit that one of the principal problems in constructing ground truth maps is that they are exceptionally tedious and time consuming. For example,

we recently made ground truth maps of three radical prostatectomy tissue sections, each containing prostate cancer, and manually painted only the malignant glands of the three digital slides (with this process taking nearly 8 hours).^[26] The possibility of incorporating pathologist-supervised automated “painting” of ground truth maps might enable increased speed and efficiency in the process, but as is evident, such a strategy easily engenders the concern that CAD algorithm performance becomes a self-fulfilling prophecy (it works optimally on the ground truth map it created!). We do note, however, that other fields, such as document analysis, have automated ground truth detection with the aid of algorithms, such as Aletheia.^[27]

A second basic issue concerns inclusion criteria and the extent of a lesion to include in a ground truth map (i.e., an atypical focus which autonomously would have been insufficient to diagnose cancer, but may be recognized as such in context with intermixed unequivocal cancer) should be included as “positive” regions in a ground truth map? Or, because their features alone are insufficient for diagnosis of cancer, should they be included as “negative?” Should such regions of lesion be excluded from performance testing of a CAD algorithm? This conundrum becomes all the more salient as numerous pathologic entities are amenable to ancillary testing for pathognomonic molecular characteristics. What if the same (as above) a typical but non-diagnostic foci within a malignancy show a pattern of immunohistochemical (IHC) staining that strongly supports or refutes a diagnosis of cancer? Is it then appropriate for such ancillary testing, perforce invisible to the H and E-based algorithm, to be used in the construction of ground truth maps?

In the end, evaluation of the performance of CAD requires algorithms comparison of predictions made by algorithms to a gold standard; the ground truth as defined by pathologists. The above concerns illustrate central tradeoffs apparent in the cost of construction of ground truth maps and the interplay between sensitivity and specificity in the diagnostic threshold for including or excluding part of an image as ground truth positive or negative. Perhaps being most important, however, is the extent to which the specific characteristics of pathologic entities complicate the concept of ground truth and highlight the varying histologic contexts where maps of ground truth must be constructed.^[2] Through the use of four use cases, we will highlight specific challenges of defining ground truth maps and show how ground truth varies, based on the length scale (magnification) and pathology of the disease.

USE CASES SHOWING A GROUND TRUTH MAP CONUNDRUM

Neoplasms Showing Mixed Morphologies

Figure 1 from Cheng and Hipp *et al.*^[28] shows an H and E

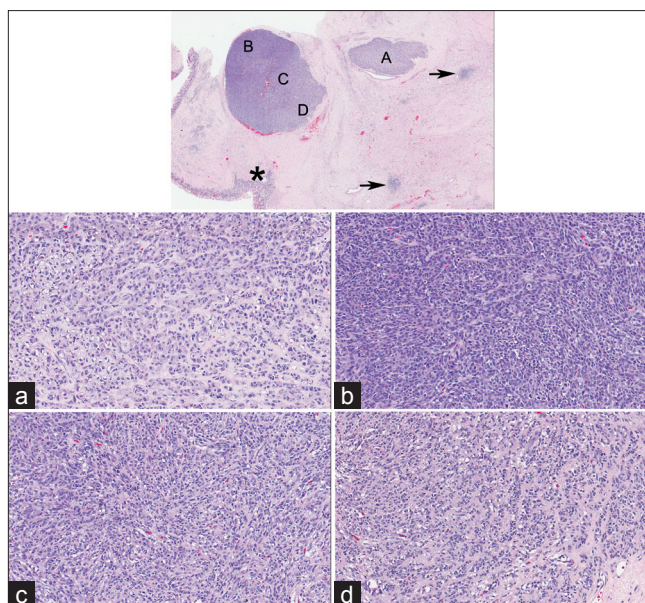


Figure 1: An H and E stained tissue section of a gastrointestinal stromal tumor (GIST) of the stomach was scanned into a digital slide as previously described by Cheng and Hipp *et al.*^[28] A low power view of the tissue section is shown above with corresponding representative high power fields of view shown in a-d. Reproduced with permission from Medknow Publications.

stained tissue section taken from a gastrointestinal stromal tumor (GIST) of the stomach as described by Cheng and Hipp *et al.* There we described, “two nodules of viable tumor set within a sparsely cellular myxohyaline matrix. Non-neoplastic gastric epithelium is present on the lower left side of the micrograph (see area near “*”) and there are two benign lymphoid aggregates (arrows).^[28] Within the tumor nodules there is histological heterogeneity. For example, the smaller nodule (a) comprises epithelioid cells arranged in cord-like arrays and separated by myxohyaline matrix. The larger nodule comprises three rather distinctive histological patterns. The upper left part of the nodule (b) consists of closely-spaced spindle cells, with oval nuclei and scant cytoplasm, with very little intervening matrix; while the lower right part (c) has a population of epithelioid and spindle cells, many exhibiting a vacuolated cytoplasm. Along the lower right edge of the large nodule (d), the histology is similar to that seen in the smaller nodule.”^[28]

If one had a low power field of view containing just the smaller nodule to the right [Figure 1a], a pathologist could “circle” the entire nodule and define this as a ground truth tumor region. However, at higher magnifications Figure 1a, one can see how the tumor cells are generously spaced amongst each other with intermixed myxohyaline matrix. From this length scale, the ground truth would be defined as the individual tumor cells themselves. Thus, ground truth determination is dependent on the corresponding length scale (magnification).

At higher power, one can appreciate the morphologic

heterogeneity of the larger nodule to the left [Figures 1b-1d]. Thus, if an image analysis algorithm were designed to generally recognize generally all these variants, a single ground truth map would be created by “painting” all the tumor cells; in contrast, if the CAD algorithm were very specific, targeting each of the morphologic variants, a series of ground truths maps would need to be created for each variant, as we described in Cheng and Hipp *et al.*^[28]

Malignancy with Admixed Precancer, Reactive Stroma, and Inflammation

Colonic adenocarcinomas often contain large proliferations of tumor cells, with them commonly admixed amongst stroma and inflammatory cells (and possibly nerves, vessels, muscle). Figure 2 of Cheng and Hipp *et al.*^[28] shows an H and E stained tissue section of a moderately-differentiated colonic adenocarcinoma.^[28] In this field of view, “the malignant glands are seen infiltrating through the benign stroma and into the muscularis propria, with focal areas of acute inflammation.”^[28] Thus, if one circles the area containing the tumor, this region would contain these benign cells and the CAD algorithm would be penalized for not identifying these as containing malignant cells. In such a use case, the ideal ground truth map would “paint” only the malignant epithelium. This would be time consuming to perform due to the ill-defined, infiltrative behavior of the tumor. Adding to the complexity of both algorithm development and performance testing, colorectal adenocarcinomas arise through an adenoma (pre-malignant, dysplasia)-adenocarcinoma sequence. As appropriate to its intended task, a ground truth map might or might not include dysplastic epithelium as ground truth positive. Finally, dysplastic morphologic elements, themselves, can be further subdivided into grades of dysplasia, with this reality further compounding the issue of setting an appropriate threshold.

Prostate cancer, a disease commonly characterized by small malignant acini infiltrating between larger benign prostatic glands, illustrates these concerns but to an even greater extent. The cancerous glands can often proliferate to form a tumor nodule that is seen to extensively infiltrate stroma, benign glands, inflammatory cells, and potentially adjacent precursor lesions. Such precursor lesions, such as high grade prostatic intraepithelial neoplasia (PIN), can show nuclear features of cancer and are often admixed within and nearby the periphery of the tumor nodule. Also, the glands of prostate cancer show luminal “white spaces” in proportion to tumor grade. Thus, circling the entire tumor nodule as ground truth “positive” would define everything within that nodule as cancer, resulting in the CAD algorithm being assessed for its ability to identify tumor regions in general, but not the specific tumor cells. The algorithm would be penalized for

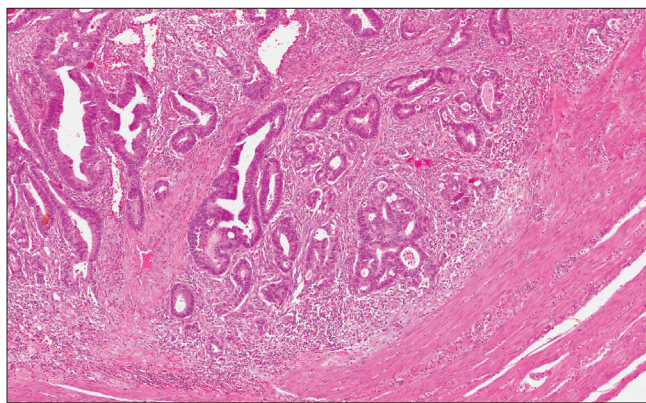


Figure 2: An H and E stained tissue section of a moderately differentiated colonic adenocarcinoma was scanned into a digital slide as previously described by Cheng and Hipp, *et al.*^[28] Reproduced with permission from Medknow Publications

not identifying the benign glands, stroma, inflammatory cells, and the luminal white spaces as cancer. In addition, as appropriate to its intended task, a ground truth map might or might not include high grade PIN (precursor lesions) as ground truth positive. In such a use case, the ideal ground truth map would include only malignant epithelium, to the exclusion of not only admixed non-neoplastic components, but also white spaces.

We recently described in work by Hipp and Monaco *et al.* that circling such a tumor nodule would be appropriate for an algorithm such as Probabilistic Pairwise Markov Modeling (PPMM)^[16] which analyzes luminal architecture and then scores the general area around these lumens as cancer.^[26] However, if one is evaluating or using a CAD algorithm that assesses and scores each pixel rather than an algorithm that scores the general area around the tumor, circling the tumor area would penalize the prior algorithm owing to the fact that the circled area would include the pixels associated with the benignity (as described above). Thus, the ground truth map must correspond to the CAD algorithm-specific selectivity under consideration.

Micropapillary urothelial carcinoma

Urothelial carcinoma (UC), the most common form of bladder cancer in Western countries, exhibits a peculiar capacity for “divergent” or “mixed histology” differentiation, comprising several unusual histologic variants.^[29] Recognition and documentation of these variants in the pathology report is critical as such findings have potential diagnostic, therapeutic and prognostic implications.^[30]

The micropapillary variant (MPUC), a variant of UC with aggressive clinical behavior, demonstrates a wide spectrum of architectural and cytologic features presenting a complex diagnostic problem for expert observers.^[31] A recent study sought to address the issue of interobserver reproducibility among 14 uropathologists, including the evaluation of

13 different, complex morphologic features commonly used for the diagnosis of MPUC and additionally, the utility of these individual features in challenging cases.^[25] This study showed significant interobserver diagnostic variability and a need to more precisely define MPUC.^[25] Recent reports suggest that quantitation of the relative amount of aggressive micropapillary morphology versus conventional UC may be of prognostic utility.^[32] Given these concerns, MPUC constitutes an entity uniquely in need of CAD, for which it is currently under investigation by our group.^[33]

MPUC presents a unique challenge in regards to ground truth: no single cell or nest of invading cells can be definitively identified as “ground truth positive.” Rather, a field of tumor cells that multifocally demonstrates a constellation of architectural features (e.g., retraction artifact, multiple separate nests within the same space) that themselves are of varying sensitivity and specificity for MPUC may be diagnosed as such based on the pathologist’s diagnostic integration of the entire morphology. In such a use case, a ground truth map might require an overlay of several, feature-specific ground truth maps, against which a CAD targeting a number of features might be compared. For that matter, in such a case where even experts may not be able to delineate strict boundaries around subareas showing variant morphology (i.e., a ground truth map), AUCs based on ground truth maps may not even be an appropriate performance metric.

CONCLUSION AND FUTURE DIRECTIONS

The construction and uses of ground truth maps for performance evaluation of CAD tools for pathology are many and reflect the complexity of diagnostic histopathology. Certainly, with this focused effort, we do no claim to have arrived upon a generalized solution that could be equally leveraged in any CAD algorithm validation setting. Rather, our use of targeted use cases serves to underscore two fundamental underlying realities intrinsic to ground truth maps: the foreground subject matter inclusion criteria is necessarily algorithm dependent and additionally, in the setting of proper matching of foreground criteria to CAD algorithm-specific selectivity, then (and only then) it becomes possible to realize sufficiently high ROC performance that is free from the artifact of over- or under-sampled “innocent-bystander” surface area (which in turn, can be seen as a cause of needlessly diminishing overall performance). Attaining the “holy-grail” of entire ground truth libraries will undoubtedly take decades of continued effort, recognizing the variability of use cases intrinsic to varying organs, disease entity, disease grade and algorithm foreground specificity (which presumably will each will require a specifically-tailored map).

Despite this challenge, the obtained results in this report demonstrates a pathway by which a workable solution for a plurality of prototypic use cases can be found. As a logical extrapolation, this model will very likely scale to ground truth map generation for the remainder of organs, diseases, grades and algorithm classes.

Recognizing that the science and art of ground truth map generation long precedes its use in digital histology, the digital pathology community can learn from the prior collective experiences of the computer science machine vision field (with specific examples being autonomous vehicle operation and active object tracking). In Hong *et al.*, a large image data repository (nearly a terabyte; or 1×10^{12} bytes) served as a target upon which system users could validate candidate ground truth classification algorithms for video-based road detection and navigational tracking.^[9] They also developed an application that “allows the ground truth to be extracted and stored in a database for later use.”^[9]

In part analogous to such efforts from the machine vision fields, we have developed a digital slide repository at the University of Michigan^[34,35] (www.WSIrepository.org) and have started to collect and deposit cohorts of ground truth imagery data on the repository site. Content submitted thus far includes original images and the associated manually generated ground truth maps from a number of our recent publications and submissions. For these and anticipated ground truth images, serving as reference datasets, we plan to include multiple ground truth annotations (circled areas, false-colored areas, etc.) and moreover, provide annotation tools, such that pathologists will be able to spatially document the context and implications of identified features, making such information immediately available to nonsubject-matter experts, such as for computer programmers. In addition, we hope the availability of a growing cohort of ground truth libraries will encourage others to generate additional annotation datasets that may be complimentary or incrementally accurate over that which is currently available today.

Finally, public availability of such curated ground truth map data sets would further enable a recent phenomenon known as crowd sourcing. We can envision that such repositories would contain ground truth data sets that encompass numerous disease examples, different diseases, length scales, special stains, and even examples showing biological and technical processing variability.

REFERENCES

1. Bridson D, Antonacopoulos A: A geometric approach for accurate and efficient performance evaluation of layout analysis methods. Manchester: Pattern Recognition & Image Anal. (PRIMA) Res. Lab, University of Salford; 2008. p. 1-4
2. Antonacopoulos A, Karatzas D, Bridson D: Ground truth for layout analysis

- performance evaluation document analysis systems VII. In: Bunke H, Spitz A, editors. Heidelberg: Springer Berlin; 2006. p. 302-11.
3. Takeuchi A, Shneier M, Hong TH, Chang T, Scrapper C, Cheok GS. Ground truth and benchmarks for performance evaluation. Bellingham, Washington USA: SPIE; 2003. p. 408-13.
 4. Shin MC, Goldgof D, Bowyer KW. An objective comparison methodology of edge detection algorithms using a structure from motion task. Tampa, FL: Department of Comput Sci and Eng, University of South Florida; 1998. p. 190-5.
 5. Baraldi A, Bruzzone L, Blonda P. Quality assessment of classification and cluster maps without ground truth knowledge. *IEEE Trans Geosci Remote Sens* 2005;43:857-73.
 6. Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med* 2008;359:1675-84.
 7. Fenton JJ, Taplin SH, Carney PA, Abraham L, Sickles EA, D'Orsi C, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 2007;356:1399-409.
 8. Chan HP, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM. Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography. *Med Phys* 1987;14:538-48.
 9. Hong TH, Takeuchi A, Foedisch M, Shneier MO. Performance evaluation of road detection and following systems. Bellingham, Washington USA: SPIE; 2004. p. 109-15.
 10. Astley SM. Evaluation of computer-aided detection (CAD) prompting techniques for mammography. *Br J Radiol* 2005;78:S20-5.
 11. Madabhushi A. Digital pathology image analysis: Opportunities and challenges. *Imaging Med* 2009;1:4.
 12. Doyle S, Feldman M, Tomaszewski J, Madabhushi A. A boosted bayesian multi-resolution classifier for prostate cancer detection from digitized needle biopsies. *IEEE Trans Biomed Eng* 2010. [In Press]
 13. Fatakawala H, Xu J, Basavanahally A, Bhanot G, Ganesan S, Feldman M, et al. Expectation-maximization-driven geodesic active contour with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology. *IEEE Trans Biomed Eng* 2009;57:1676-89.
 14. Basavanahally AN, Ganesan S, Agner S, Monaco JP, Feldman MD, Tomaszewski JE, et al. Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology. *IEEE Trans Biomed Eng* 2010;57:642-53.
 15. Lexe G, Monaco J, Doyle S, Basavanahally A, Reddy A, Seiler M, et al. Towards improved cancer diagnosis and prognosis using analysis of gene expression data and computer aided imaging. *Exp Biol Med* (Maywood) 2009;234:860-79.
 16. Monaco JP, Tomaszewski JE, Feldman MD, Hagemann I, Moradi M, Mousavi P, et al. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. *Med Image Anal* 2010;14:617-29.
 17. Gurcan MN, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B. Histopathological image analysis: A Review. *IEEE Rev Biomed Eng* 2009;2:147-71.
 18. Hipp J, Flotte T, Monaco J, Cheng J, Madabhushi A, Yagi Y, et al. Computer aided diagnostic tools aim to empower rather than replace pathologists: Lessons learned from computational chess. *J Pathol Inform* 2011;2:25.
 19. Hipp J, Cheng J, Daignault S, Sica J, Dugan MC, Lucas D, et al. Automated area calculation of histopathologic features using SIVQ. *Anal Cell Pathol (Amst)* 2011. [In Press]
 20. Doyle S, Feldman MD, Tomaszewski JE, Shih N, Madabhushi A. Cascaded multi-class pairwise classifier (CascaMPa) for normal, cancerous, and cancer confounder classes in prostate histology. *IEEE Int Symp Biomed Imaging* 2011.
 21. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol* 2004;5:11-8.
 22. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. *Crit Care* 2004;8:508-12.
 23. Grunkemeier GL, Jin R. Receiver operating characteristic curve analysis of clinical risk models. *Ann Thorac Surg* 2001;72:323-6.
 24. Hipp JD, Fernandez A, Compton CC, Balis UJ. Why a pathology image should not be considered as a radiology image. *J Pathol Inform* 2011;2:26.
 25. Sangoi AR, Beck AH, Amin MB, Cheng L, Epstein JI, Hansel DE, et al. Interobserver reproducibility in the diagnosis of invasive micropapillary carcinoma of the urinary tract among urologic pathologists. *Am J Surg Pathol* 2010;34:1367-76.
 26. Hipp J, Monaco J, Kunju LP, Cheng J, Yagi Y, Rodriguez-Canales J, et al. Integration of architectural and cytologic driven image algorithms for prostate adenocarcinoma identification, in submission. *Analytical Cellular Pathology*. [In Press]
 27. Antonacopoulos A, Bridson D, Papadopoulos C, Pletschacher S. A Realistic dataset for performance evaluation of document layout analysis. *Proc. ICDAR, Barcelona, Spain; 2009*. p. 296-300.
 28. Cheng J, Hipp J, Monaco J, Lucas DR, Madabhushi A, Balis UJ. Automated vector selection of SIVQ and parallel computing integration MATLAB: Innovations supporting large-scale and high-throughput image analysis studies. *J Pathol Inform* 2011;2:37.
 29. The-International-Agency-for-Research-on-Cancer. WHO Classification of Tumours: Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs. In: Elbe JN, Sauter G, Epstein JI, Sesterheim IA, editors. Oxford: Oxford University Press; 2004.
 30. Black PC, Brown GA, Dinney CP. The impact of variant histology on the outcome of bladder cancer treated with curative intent. *Urol Oncol* 2009;27:3-7.
 31. Sangoi AR, Beck AH, Amin MB, Cheng L, Epstein JI, Hansel DE, et al. Interobserver reproducibility in the diagnosis of invasive micropapillary carcinoma of the urinary tract among urologic pathologists. *Am J Surg Pathol* 2010;34:1367-76.
 32. Compérat E, Conort P, Rouprêt M, Camparo P, Mazerolles C. Pathologic diagnosis and management of flat lesions of urothelium detected with aminolevulinic acid (Hexvix®). *Prog Urol* 2011;21:157-65.
 33. Hipp J, Smith SC, Cheng J, Tomlins SA, Monaco J, Madabhushi A, et al. Optimization of complex cancer morphology detection using the SIVQ pattern recognition algorithm. *Anal Cell Pathol (Amst)* 2012;35:41-50.
 34. Hipp JD, Lucas DR, Emmert-Buck MR, Compton CC, Balis UJ. *Digital slide repositories* for publications: lessons learned from the microarray community. *Am J Surg Pathol* 2011;35:783-6.
 35. Hipp JD, Sica J, McKenna B, Monaco J, Madabhushi A, Cheng J, et al. The need for the pathology community to sponsor a whole slide imaging repository with technical guidance from the pathology informatics community. *J Pathol Inform* 2011;2:31.