Biomedical imaging informatics in the era of precision medicine: progress, challenges, and opportunities

William Hsu,[1] Mia K Markey,[2,3] May D Wang[4,5,6]

INTRODUCTION

Biomedical informatics is the interdisciplinary field that studies and pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, motivated by efforts to improve human health.[1 2] Not only do biomedical informaticians study and develop theories, methods, and processes for the generation, manipulation, and sharing of biomedical data, but they also investigate how to model and reason on these data in order to effect beneficial change in the healthcare enterprise. In addition, an important aspect associated with developments in this field is the consideration of social and behavioral sciences in the design and evaluation of technical solutions. As a subfield of biomedical informatics, biomedical imaging informatics (BMII) encompasses all of the aforementioned aspects from the perspective of imaging. BMII has emerged as one of the fastest growing research areas in recent years given the evolution of techniques in molecular imaging, anatomical imaging, and functional imaging and advancements in imaging biomarker generation. Developments have also been accelerated by efforts to realize precision medicine,[3] which necessitates a multiscale understanding of diseases that integrate insights in areas such as radiology, pathology, and genetics. This focus issue highlights the growing impact of BMII, demonstrating the increasing breadth of imaging modalities (eg, optical, molecular, in addition to traditional diagnostic modalities) and the diversity of specialties that depend on imaging information (eg, dermatology, pathology, surgery).

Early efforts in BMII can be traced to the 1980s when the rise in radiological imaging techniques such as CT and MRI necessitated a digital, filmless approach to acquiring and interpreting images. The ability to acquire and distribute images electronically using picture archiving and communication systems (PACS) spawned a variety of applications aimed at improving radiological practice, research, and education. Imaging informatics efforts resulted in the development of specialized standardized terminologies such as BI-RADS (Breast Imaging–Reporting and Data System),[4] adoption of data standards such as the Digital Imaging and Communication in Medicine (DICOM) standard,[5] and the use of telemedicine via teleradiology.[6] In addition, these efforts have provided a foundation for the current developments in BMII, including image acquisition, management, quality control, information extraction, modeling, computer-aided detection and diagnosis, content-based image retrieval (CBIR), visualization, and quantitative imaging.[7 8] For instance, diffusion-weighted MRI provides detailed structural and functional information that can be used to characterize complex disease phenomena in vivo (eg, estimate tumor cell proliferation rate), bridge evidence across different biological scales (eg, link tissue and cellular levels), and quantitatively assess the effectiveness of interventions for individual patients (eg, generate maps of treatment response).[9] With the ongoing healthcare reform, adoption of electronic health records, and the rise of big data (with imaging contributing a major part), new approaches to extracting, modeling, and acting on this information are needed to help scientists, providers, caregivers, patients, and public health professionals to understand the importance of image findings in context (eg, taking into account the full medical record) and, ultimately, to achieve efficient and effective care of individuals.

Many professional societies such as the American Medical Informatics Association (AMIA), the International Society for Optics and Photonics (SPIE), the Radiological Society of North America (RSNA), the Society for Imaging Informatics in Medicine (SIIM), and the Medical Image Computing and Computer Assisted Intervention Society (MICCAI) incorporate imaging informatics-related themes. Nevertheless, recent developments in the field underscore the pressing need to engage individuals from multiple disciplines to address shared challenges such as managing large datasets, developing common standards to achieve interoperability, and integrating multiscale evidence for personalized medicine. AMIA is uniquely poised to advance imaging informatics because it has a diverse membership that spans numerous organizational types (eg, academia, industry, healthcare) and disciplines (eg, medicine, engineering, social sciences). Expertise within the broader biomedical informatics community would be beneficial in: (1) developing technical and policy guidance on the implementation of a multimedia electronic health record and the role of imaging in the meaningful use criteria from the Office of the National Coordinator for Health Information Technology; and (2) incorporating the perspectives of other medical specialties (eg, primary care, cardiology), providers (eg, nurses, pharmacists), stakeholders (eg, payors, policymakers), and disciplines (eg, health services, cognitive sciences). We believe that AMIA would provide a fertile environment for fostering new collaborations among these individuals, complementing existing societies, to define how imaging-derived information should be integrated, represented, interpreted, and acted upon.

ARTICLES IN THIS FOCUS ISSUE

This focus issue provides a cross-section of topics in BMII, characterizing the breadth of current work through original research articles, a review on whole-slide imaging informatics, and a perspective on developing imaging-based observational databases for research. Going beyond traditional information extraction (ie, image analysis), these articles describe the promise and challenges of applying analytical techniques to actual clinical cases, developing clinical decision support tools, and incorporating a cognitive perspective. We broadly group articles into the five categories below.

Image characterization. New developments in image processing and analysis provide methods for handling noise, artifacts, and differences in acquisition techniques to generate accurate and reproducible

[1]Medical Imaging Informatics (MII) Group, Department of Radiological Sciences, UCLA David Geffen School of Medicine, Los Angeles, California, USA; [2]UT Biomedical Informatics Lab (UT BMIL), Department of Biomedical Engineering, Cockrell School of Engineering, The University of Texas at Austin, Austin, Texas, USA; [3]Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA; [4]The Wallace H Coulter Joint Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, USA; [5]The School of Electrical and Computer Engineering, Parker H Petit Institute for Bioengineering and Bioscience, and Institute for People and Technology, Georgia Institute of Technology, Atlanta, GA, USA; [6]Department of Hematology and Oncology, and Winship Cancer Institute, Emory University, Atlanta, GA, USA

Correspondence to Dr William Hsu, Medical Imaging Informatics (MII) Group, Department of Radiological Sciences, UCLA David Geffen School of Medicine, Los Angeles, CA 90024, USA; willhsu@mii.ucla.edu

information from images. Yang and Fei[10] describe an automatic method for segmenting the skull in magnetic resonance (MR) images. Their novel approach performs bilateral filtering in radon-transformed MR images to create a multiscale image series. Multiscale processing allows the results from a coarser scale to be used as a mask for the next fine scale. The method is robust to noise, as shown by results on phantom, simulated brain and real MRI data. The proposed method can be used during attenuation correction in brain MR/positron emission tomography (PET) applications. Neubert *et al*[11] present informatics tools for automated detection of abnormalities in lumbar intervertebral discs (IVD) from both high-resolution three dimensional (3D) and routine two-dimensional (2D) clinical T2-weighted MRI. Using the proposed informatics tools, they quantify signal intensity and IVD morphology in both 3D and 2D MRIs of patients with varying degrees of disc degeneration to detect abnormalities. Their results suggest that the 3D features contain important information for the detection of abnormalities, compared with the traditionally used 2D MRI features.

Image management. A longstanding challenge in the imaging informatics community has been how to process semantic information from images for tasks such as retrieval. While the area of CBIR has been widely reported,[12 13] recent efforts address the 'semantic gap', which refers to the divide that exists between low-level pixel data provided by feature extraction techniques and the high-level interpretations that are made by domain experts.[14] Crespo Azcárate *et al*[15] offer a solution to improve the effectiveness of query expansion for image retrieval systems based on hierarchical Medical Subject Headings (MeSHs). They propose two strategies: using the Unified Medical Language System (UMLS) to identify medical concepts; and dividing the query text into n-grams for sequences corresponding to the heading descriptors. Their results show a significant improvement over the results of the ImageCLEF 2011 competition. The proposed solution has a high level of clinical relevance, especially in light of widespread growth of electronic clinical imaging systems that could incorporate these query methods to improve the relevance of retrieved images. Chen *et al*[16] describe a methodology, based on support vector machines (SVMs), to improve and customize image retrieval systems for specific disease categories. The authors describe techniques to retrieve an image showing corresponding diseased body parts when a search is made for a specific disease. They reduce the number of annotations required by training the classifier on a set of body parts as opposed to specific diseases. Their algorithm uses MeSH terms in conjunction with body part identification to retrieve specific medical images. In addition, two articles describe web-based frameworks for the retrieval and analysis of massive imaging datasets. Bourgeat *et al*[17] describe a tool called 'MilxXplore', which displays multimodal data (eg, MRI, PET), permitting users to perform quantitative image analysis over the web. To permit such analysis, 3D volumes are converted into 2D representative slices with segmentation results overlaid. In addition, built-in statistical modules that use R are presented in tables and graphs showing the changes over time, when possible. An article by Gutman *et al*[18] details a platform for storing and visualizing whole-slide pathological images called the Cancer Digital Slide Archive (CDSA). Currently, CDSA hosts over 20 000 images from 22 cancer types provide by The Cancer Genome Atlas (TCGA). The authors present a case study on glioblastoma and discuss challenges and solutions while developing such a large-scale platform.

Modeling and decision support. The integration of semantic and quantitative features from biomedical images with other clinical and genomic data is an ongoing area of research, but has already started to yield new insights about diseases. A perspective by Bui *et al*[19] highlights ongoing imaging informatics challenges in establishing image-based observational databases for holistic disease modeling using imaging, genomic, and clinical data. The article focuses on organ imaging, but the framework discussed can easily be extended to other medical images. They categorize the challenges into five broad areas: (1) image processing, including normalization, standardization, and feature extraction; (2) natural language processing, including automatic concept identification, ontological mapping, and structuring; (3) context-driven image interpretation; (4) integration and construction multiscale models; and (5) development of advanced applications for computer-aided diagnosis and image retrieval. In addition, two articles present the results of combining multiscale features drawn from imaging and other data sources for classification and prediction. In the first article, Liu *et al*[20] describes an automatic architecture for glaucoma diagnosis by integrating imaging, genomic, and clinical data. They generate predictive models that incorporate genomic single-nucleotide polymorphism data, retinal imaging features, and personal patient data using a multiple kernel SVM learning framework. The article shows that combining these heterogeneous data sources provides increased accuracy compared with single sources, and also substantially improves performance compared with the current clinical standard of care. Second, Golden *et al*[21] discuss an exploratory study to predict response of early-stage triple-negative breast cancer to neoadjuvant chemotherapy. Logistic regression models were constructed from different combinations of clinical findings, quantitative image features extracted from gray-level co-occurrence matrix of lesion kinetic maps, BI-RADS scores, and patterns of response derived from dynamic contrast-enhanced MRI. Pre-chemotherapy imaging features are shown to generally correlate with pathological complete response, residual lymph node metastases, and residual tumor with lymph node metastases. Finally, an article by Kothari *et al*[22] presents a review of challenges, state-of-the-art methods, and future directions in histopathological whole-slide imaging informatics. They discuss key components of a clinical decision support system such as: quality control of whole-slide images to eliminate image artifacts and batch effects; comprehensive image description using different types of pixel-, object-, and semantic-level image features; prediction modeling for patient-level diagnosis using features such as grade and subtype; and data and image visualization for exploratory analysis and semantic interpretation. The review also highlights the importance of some informatics methods in a case study on clinical endpoint prediction for patients from TCGA with kidney renal clear cell carcinoma.

Applications for improving clinical practice. Informatics methods are also improving the efficiency of interpreting imaging information and the communication of evidence derived from images to stakeholders such as patients. An article by Tourassi *et al*[23] investigates the modeling of the relationship between image content, human perception, cognition, and error while diagnosing breast cancer using mammograms. Using gaze data collected from six radiologists tasked with making diagnostic decisions, the authors find that quantitative image features can predict perceptual behavior (tendency to dwell) of experts, and images and gaze features can predict cognitive behavior (ease of diagnosis). Such models can be used in a decision support system to predict diagnostic errors. The study establishes that behavior prediction models differ from expert to expert. Arnold *et al*[24] demonstrate an online portal that helps patients effectively view and understand their radiology images and reports during the phases of diagnosis and treatment. The portal features three important components: (1) generation of a timeline for disease progression based on results from an image processing module; (2) concept extraction from radiology reports using natural language processing; and (3) navigation of results via a consumer-oriented web-based user interface. A dataset of 2883 brain MR images from 277 patients with brain cancer were leveraged to demonstrate the usability of the portal. The results indicate that informatics tools can be supportive of patient education and information interpretation by providing simplified radiological images and clinical reports. As a majority of health information systems are designed for providers, this study presents the importance of clinical patient engagement by introducing educational components using imaging informatics solutions.

LOOKING FORWARD

An overarching challenge in the informatics community is matching our ability to generate and acquire data with a comparable ability to understand and act on this information. The ability to mine clinical data is essential for advancing biomedical research on a wide range of health conditions, for providing actionable evidence to caregivers based on analyses of images, and for studying population health. We highlight several emerging areas as reflected in discussions at recent workshops and articles appearing in this issue.

Making imaging and imaging-derived information accessible. A growing number of imaging datasets are being made publicly available through initiatives such as The Cancer Imaging Archive, Osteoarthritis Initiative, and Alzheimer's Disease Neuroimaging Initiative.[25] These datasets can be used to advance biomedical research by providing researchers with test datasets collected at different institutions using various pieces of equipment for validating image analysis and imaging-based prediction algorithms. However, the usefulness of these images in developing decision support systems is dependent on the availability of contextual information such as relevant image acquisition details (eg, to normalize measured intensity values) and clinical and outcome information for each individual. Standards for capturing the semantic content of imaging data and incorporating other contextual information (eg, acquisition parameters, clinical/outcome information) are integral to enable reuse of collected data for subsequent analyses. Annotation Image Markup (AIM) is an example of an emerging format for sharing image annotations, but the AIM standard does not address issues related to variability in how image features are interpreted by radiologists and in how quantitative image features are reported.

Managing increased use of imaging across clinical domains. Imaging is integral to many clinical departments, including cardiology, dermatology, orthopedics, neurology, ophthalmology, pulmonary medicine, radiation oncology, surgery, and urology. For example, in cardiology, imaging provides structural (eg, fiber tracking) and functional (eg, perfusion) information to detect ischemia, scarring, and stenosis. Images are currently maintained in separate information systems (eg, PACS) and only loosely integrated with electronic health records. Moving forward, we need to define an architecture and associated standard for making imaging an integral part of the multimedia electronic health record[26] that permits the meaningful use of imaging data in the context of diagnostic and treatment decisions.

Using imaging to link biological scales. Imaging biomarkers provide functional molecular information about underlying physiologic or cellular processes related to cell proliferation, vascularity, and metabolism. The combination of medical imaging and genomics presents a new opportunity to link observations made at the tissue or organ levels to cell behavioral phenotypes such as proliferation, apoptosis, and cell motility. Analysis related to the correlation of genomic information and quantitative imaging biomarkers is often referred to as radiogenomics or imaging genomics.[27–29] These studies result in the generation of association maps between imaging features and genetic mutations, linking anatomical regions with functional information at the (sub)cellular level. In addition, contextual information about the patient's history, course of treatment, and other diagnostic data gleaned from the electronic medical record can be integrated with imaging and genomic level information, establishing a unique picture of the patient and creating a roadmap for optimizing treatment. Such studies have led to insights into redefining diseases—for instance, a study that combined MRI features with information about molecular and genetic mutations into a Cox regression model resulted in a more accurate indication of outcome over existing classifications.[30]

Ensuring the reproducibility and utility of imaging-derived evidence. Methods are needed to evaluate the accuracy and reproducibility of information extracted from imaging data. For instance, techniques for image enhancement (eg, iterative image reconstruction techniques) and quantification (eg, image registration and segmentation algorithms) need to be evaluated on a dataset that is representative of potential variations in image quality and acquisition parameters. While mean squared error, peak signal-to-noise ratio, and qualitative assessments made by a domain expert are commonly reported, they each have limitations in their ability to objectively characterize the perceptual quality of each image. Several new image quality indexes have been proposed, such as structural similarity (SSIM)[31] and the gray relational coefficient.[32] Furthermore, quantitative biomarkers that are generated from imaging data should be robust against variations in acquisition hardware (eg, scanners from different manufacturers) and image processing approaches (eg, effect of using a rigid vs non-rigid registration method).

The landscape of biomedical imaging has changed significantly since the last *JAMIA* focus issue on imaging informatics[33]; the focus of research has shifted from acquiring and managing imaging data to understanding its significance to interpreting derived knowledge in the context of the entire health record. As evidenced by the articles in this issue, BMII is not limited to traditional radiological imaging; rather, the field encompasses a broad range of imaging modalities and serves a number of stakeholders. As we enter the era of precision medicine, future advancements in BMII will provide precise characterizations of pixel data and approaches to understanding imaging information in the context of other biomedical data (eg, clinical, genomic). We invite members from the imaging and informatics communities and organizations such as AMIA to collaborate and provide leadership towards a shared vision for imaging informatics in the broader context of biomedical informatics.

Contributors WH, MKM, and MDW contributed equally to the intellectual content and preparation of this editorial.

Competing interests None.

Provenance and peer review Commissioned; internally peer reviewed.

Editor's choice
Scan to access more
free content

To cite Hsu W, Markey MK, Wang MD. *J Am Med Inform Assoc* 2013;**20**:1010–1013.

Received 31 August 2013
Accepted 7 September 2013

J Am Med Inform Assoc 2013;**20**:1010–1013.
doi:10.1136/amiajnl-2013-002315

REFERENCES

1 Kulikowski CA, Shortliffe EH, Currie LM, *et al*. AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *J Am Med Inform Assoc* 2012;19:931–8.
2 Shortliffe EH, Cimino JJ. *Biomedical informatics: computer applications in health care and biomedicine*. Springer, 2006.
3 National Research Council (U.S.). *Committee on a framework for developing a new taxonomy of disease. Toward precision medicine building a knowledge network for biomedical research and a new taxonomy of disease*. Washington, DC: National Academies Press, 2011.
4 Bassett LW. Standardized reporting for mammography: BI-RADS™. *Breast J* 1997;3:207–10.
5 Bidgood WD, Horii SC, Prior FW, *et al*. Understanding and using DICOM, the data interchange standard for biomedical imaging. *J Am Med Inform Assoc* 1997;4:199–212.
6 Ruggiero C. Teleradiology: a review. *J Telemed Telecare* 1998;4:25–35.
7 Geis JR. Medical imaging informatics: how it improves radiology practice today. *J Digit Imaging* 2007;20:99–104.
8 Sinha U, Bui A, Taira R, *et al*. A review of medical imaging informatics. *Ann NY Acad Sci* 2002;980:168–97.
9 Bui AA, Taira RK. *Medical imaging informatics*. Springer, 2010.
10 Yang X, Fei B. Multiscale segmentation of the skull in MR images for MRI-based attenuation correction of combined MR/PET. *J Am Med Inform Assoc* 2013;20:1037–45.
11 Neubert A, Fripp J, Engstrom C, *et al*. Three-dimensional morphological and signal intensity features for detection of intervertebral disc degeneration from magnetic resonance images. *J Am Med Inform Assoc* 2013;20:1082–90.
12 Müller H, Michoux N, Bandon D, *et al*. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *Int J Med Inform* 2004;73:1–23.
13 Akgül CB, Rubin DL, Napel S, *et al*. Content-based image retrieval in radiology: current status and future directions. *J Digit Imaging* 2011;24:208–22.
14 Deserno TM, Antani S, Long R. Ontology of gaps in content-based image retrieval. *J Digit Imaging* 2009;22:202–15.
15 Crespo Azcárate M, Mata Vázquez J, *et al*. Improving image retrieval effectiveness via query expansion using MeSH hierarchical structure. *J Am Med Inform Assoc* 2012;20:1014–20.
16 Chen Y, Ren X, Zhang G-Q, *et al*. Ontology-guided organ detection to retrieve web images of disease manifestation: towards the construction of a consumer-based health image library. *J Am Med Inform Assoc* 2013;20:1076–81.
17 Bourgeat P, Dore V, Villemagne VL, *et al*. MilxXplore: a web-based system to explore large imaging datasets. *J Am Med Inform Assoc* 2013;20:1046–52.
18 Gutman DA, Cobb J, Somanna D, *et al*. Cancer digital slide archive: an informatics resource to support integrated in silico analysis of TCGA pathology data. *J Am Med Inform Assoc* 2013;20:1091–8.
19 Bui AAT, Hsu W, Arnold C, *et al*. Imaging-based observational databases for clinical problem solving: the role of informatics. *J Am Med Inform Assoc* 2013;20:1053–8.
20 Liu J, Zhang Z, Wong DWK, *et al*. Automatic glaucoma diagnosis through medical imaging informatics. *J Am Med Inform Assoc* 2013;20:1021–7.
21 Golden DI, Lipson JA, Telli ML, *et al*. Dynamic contrast-enhanced MRI-based biomarkers of therapeutic response in triple-negative breast cancer. *J Am Med Inform Assoc* 2013;20:1059–66.
22 Kothari S, Phan JH, Stokes TH, *et al*. Pathology imaging informatics for quantitative analysis of whole-slide images. *J Am Med Inform Assoc* 2013;20:1099–108.
23 Tourassi G, Voisin S, Paquit V, *et al*. Investigating the link between radiologists' gaze, diagnostic decision, and image content. *J Am Med Inform Assoc* 2013;20:1067–75.
24 Arnold CW, McNamara M, El-Saden S, *et al*. Imaging informatics for consumer health: towards a radiology patient portal. *J Am Med Inform Assoc* 2013;20:1028–36.
25 Toga AW. The clinical value of large Neuroimaging data sets in Alzheimer's disease. *Neuroimaging Clin N Am* 2012;22:107–18.
26 Seto B, Friedman C. Moving toward multimedia electronic health records: how do we get there? *J Am Med Inform Assoc* 2012;19:503–5.
27 Rutman AM, Kuo MD. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. *Eur J Radiol* 2009;70:232–41.
28 Lambin P, Rios-Velazquez E, Leijenaar R, *et al*. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441–6.
29 Thompson PM, Martin NG, Wright MJ. Imaging genomics. *Curr Opin Neurol* 2010;23:368–73.
30 Gutman DA, Cooper LA, Hwang SN, *et al*. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA Glioblastoma data set. *Radiology* 2013;267:560–9.
31 Wang Z, Bovik AC, Sheikh HR, *et al*. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 2004;13:600–12.
32 Lin SC, Lin YC, Feng WS, *et al*. A novel medical image quality index. *J Digit Imaging* 2011;24:874–82.
33 Kulikowski CA. Medical imaging informatics: challenges of definition and integration. *J Am Med Inform Assoc* 1997;4:252–3.