CrossMark

PathBot: A Radiology-Pathology Correlation Dashboard

Linda C. Kelahan[1] · Amit D. Kalaria[1] · Ross W. Filice[1,2]

Abstract Pathology is considered the "gold standard" of diagnostic medicine. The importance of radiology-pathology correlation is seen in interdepartmental patient conferences such as "tumor boards" and by the tradition of radiology resident immersion in a radiologic-pathology course at the American Institute of Radiologic Pathology. In practice, consistent pathology follow-up can be difficult due to time constraints and cumbersome electronic medical records. We present a radiology-pathology correlation dashboard that presents radiologists with pathology reports matched to their dictations, for both diagnostic imaging and image-guided procedures. In creating our dashboard, we utilized the RadLex ontology and National Center for Biomedical Ontology (NCBO) Annotator to identify anatomic concepts in pathology reports that could subsequently be mapped to relevant radiology reports, providing an automated method to match related radiology and pathology reports. Radiology-pathology matches are presented to the radiologist on a web-based dashboard. We found that our algorithm was highly specific in detecting matches. Our sensitivity was slightly lower than expected and could be attributed to missing anatomy concepts in the RadLex ontology, as well as limitations in our parent term hierarchical mapping and synonym recognition algorithms. By automating radiology-pathology correlation and presenting matches in a user-friendly dashboard format, we hope to encourage pathology follow-up in clinical radiology practice for purposes of self-education and to augment peer review. We also hope to provide a tool to facilitate the production of quality teaching files, lectures, and publications. Diagnostic images have a richer educational value when they are backed up by the gold standard of pathology.

Background

In the current climate of quality measures and standards, a seamless, consistent system (i.e., a dashboard) for presenting correlative pathology across the entire range of interventional and diagnostic radiology examinations provides several benefits. First, the imager receives pathology results on images where a novel finding (i.e., a new liver mass) prompts a differential diagnosis. Many times, the radiologist plans to "follow up" an interesting and complex imaging case, but workflow demands and cumbersome electronic medical record query can exhaust even the most dedicated and well-intentioned imager. Second, the procedure-oriented imager should have knowledge of the success rate of their image-guided biopsies and, ultimately, if the pathology results are concordant or discordant with imaging findings. Such consistent feedback achieves two main endpoints: (1) provides meaningful assessment of the value added of an image-guided biopsy and (2) identifies the need for further evaluation if discordance is discovered.

The importance of pathology correlation continues to be emphasized in radiology training. In a survey of radiology chief residents in 2015, 78% of radiology residency programs sent all of their residents to lectures at the American Institute for Radiologic Pathology, a 4-week course dedicated to

✉ Linda C. Kelahan
Lkelahan2@gmail.com

1 Department of Radiology, MedStar Georgetown University Hospital, 3800 Reservoir Road NW, Washington, DC 20007, USA

2 MedStar Medical Group Radiology, Washington, D.C, USA

teaching the imaging appearance of disease by juxtaposing associated pathology images and emphasizing underlying pathophysiology [1]. The value of pathologic correlation in radiology, especially in regard to patient care, is perhaps best illustrated in breast imaging. Breast imagers must correlate their biopsy results and determine if these are concordant with imaging findings, due to the requirements set forth by the Mammography Quality Standards Act (MQSA) in 1992 [2]. Other radiology subspecialties correlate pathology results less rigorously, particularly for diagnostic (non-interventional) examinations, though correlation in this setting can provide rich and meaningful feedback at all levels of experience. If a radiologist diagnoses or misses a case of appendicitis on a computed tomography (CT) scan, knowledge of the pathologic diagnosis of appendicitis serves as direct feedback to the radiologist. Potentially, pathologic correlation can serve as an objective metric to augment traditional peer review. In current practice, radiologists assess each other's diagnostic radiology report quality, which is essentially subjective with varying degrees of formality. Attempts to formalize peer review, such as the American College of Radiology's web-based numerical scoring system, RADPEER, which grades degree of interpretation discrepancy between two radiologists have limitations [3]. One limitation is hindsight bias, as the reviewing radiologist is evaluating the radiologist who interpreted the prior study. For example, a lesion that is much larger on the current study was present in hindsight but significantly smaller and therefore *missed* in retrospect. Second, there is a lack of anonymity between the reviewer and the radiologist being reviewed, potentially introducing some personal bias. In addition, these scores are not actually available to the individual radiologist, but are instead submitted to the head of the department, eliminating any real-time or specific feedback to the radiologist. Third, there have been reported rates of low interobserver correlation in the peer review process, again suggesting this is a highly subjective and imperfect standard [4, 5]. Feedback from referring clinicians can be infrequent, depending on the practice setting, and are generally not tracked.

When an image-guided biopsy is performed, the adequacy of a sample to render a pathologic diagnosis is a key quality measure. Further, if the resulting pathologic diagnosis does not appear consistent (concordant) with the imaging findings that prompted the biopsy, this should prompt further evaluation. Alternatively, if the pathologic diagnosis was never even initially considered in the radiologist's differential, or if the finding was missed outright and the patient was taken to surgery on the basis of clinical judgment (i.e., appendicitis), this serves as direct educational feedback for the imager.

In clinical practice, pathologic correlation is less emphasized for multiple reasons. Time can be a limiting factor for pathology correlation, as is high imaging volume. Accessing pathology reports through cumbersome electronic health records or by contacting the referring physician is inefficient.

Ideally, access to pathology results that are relevant to the radiologist should be more seamless, easy-to-access, easy-to-read, and current. We felt that a web-based radiology-pathology correlative dashboard could best achieve these goals. The fundamental goal of any effective dashboard is to unify multiple moving parts of a system and function as a real-time display of a quality metric, without overwhelming the user. In a panel discussion defining a quality dashboard, one of the specific quality metric candidates cited was pathologic comparison as a measure of radiologists' accuracy [6]. A radiology-pathology dashboard has previously been presented as an abstract at the Radiological Society of North America (RSNA) 2014 Scientific Assembly and Annual Meeting [7]. However, from the author's abstract, it appears that this dashboard was limited to only addressing correlation of interventional procedures (not diagnostic imaging) and relied on temporal relationships between procedures and pathology results. That is, any pathology results on a patient following a procedure for a given time period were included, without any attempt to capitalize on further analysis to make these pairings more relevant to the radiologist. While this non-specific method of radiology-pathology correlation may be effective in procedural feedback, it does not address pathology correlation for diagnostic (non-procedural) imaging. A searchable database centralizing radiologic and pathologic information (to eliminate manual query of multiple hospital systems) has been created elsewhere for facilitating research and teaching file creation by radiologists [8]. This is a valuable resource for radiologists searching for specific diagnoses, but does not provide direct feedback for the cases the radiologist has dictated. Direct feedback for the radiologist with the diagnostic "gold standard" of pathology is possible via a dashboard. We present a novel approach in utilizing the RadLex ontology in annotating radiology and pathology reports in our dashboard development, which allows us to expand and provide pathology correlation for both diagnostic imaging reports and image-guided procedures.

Methods

Our study met the criteria for exemption from approval by the institutional review board at our institution. We query all pathology results for patients who had a radiology exam or image-guided biopsy performed within 60 days prior on a daily basis from one of our clinical databases (Microsoft Amalga). The patient's medical record number (MRN) serves as the unifying frame of reference to link these reports. The pathology and radiology reports (both diagnostic and image-guided procedural dictations) are then both annotated by the National Center for Biomedical Ontology (NCBO) Annotator [9] hosted locally on a virtual machine utilizing a limited subset of anatomical concepts from the RadLex ontology.

The RadLex ontology was developed to index educational materials found online, especially for the RSNA Medical Imaging Resource Center (MIRC), as an improvement on the ACR Index numbering organizational scheme [10]. Our method is a novel application of this ontology. The result is a list of the most frequent anatomic concepts in the pathology report. These terms are further mapped to synonyms and parent terms in a heuristic fashion that worked for the unique features of reporting at our institution. If any of the two most frequent concepts found in the pathology report are found in the radiology report, this is categorized as a "match" and sent to a separate database. Matched reports are pulled from this database and displayed on a web-based dashboard, which has a secure login for all radiologists in our department. On this dashboard, users can select a date range, categorize their matches as "Concordant" or "Discordant" in a similar fashion as practiced by breast imagers, and provide feedback on incorrect matches ("Irrelevant") button (Fig. 1). Concordant means the pathology result is expected given the interventional or diagnostic findings. Discordant means the pathology result is unexpected and that further workup may be warranted. Irrelevant means the presented correlative pathology result is completely irrelevant (e.g., a lung biopsy result presented for a CT of a lower extremity). The first two categories are metrics for the user to refer to. The last category is intended to help us identify non-sensical matches and to then refine our matching algorithm to present more meaningful results.

The benefit of our algorithm is feedback for both diagnostic studies and for procedures. For example, a rib lesion diagnosed on a chest CT as a chondroid neoplasm is confirmed by biopsy (Fig. 2). The radiologist who made the imaging diagnosis receives direct feedback confirming their findings. The radiologist who performs the biopsy can also see the pathology results alongside their procedural dictation. The pathology results for a CT-guided bone marrow biopsy confirm both the adequacy of the sample provided, as well as the diagnosis of remission (Fig. 3).

Results

Initially, a small subset of potential radiology-pathology report match candidates are reviewed as to whether our matching algorithm would correctly identify them as a match (if relevant) or left them unmatched (if irrelevant). Specifically, 124

pathology-radiology match candidates, which are marked as matched or unmatched by our algorithm, are reviewed by a second-year radiology resident. The actual radiology images and pathology slides are not reviewed, only the reports. This "small batch" review is conducted to identify large gaps or errors in our approach. The sensitivity of our algorithm on this small batch review is 47%, specificity is 92%, and accuracy is 71%. Of the five false-positive reports, three are attributed to information in the "clinical history" section of the radiology report that is irrelevant. For example, a patient history of a "small bowel transplant" on dictations for a chest radiograph, a lower extremity Doppler, and a pelvic ultrasound report is erroneously linked with a pathology report of a small bowel biopsy. For this reason, the clinical history section of the radiology report is subsequently excluded from annotation. Additionally, given the number of irrelevant radiographs and lower extremity ultrasounds discovered on an initial review, these studies are considered low yield for purposes of pathology follow-up and are excluded from query in the next iteration of our algorithm.

With these exclusions (ultrasound and plain film reports and the clinical history subsection of radiology reports) to our algorithm, a second, larger manual review by a second-year radiology resident is again undertaken. This review of 576 match candidate reports yields an increased sensitivity of 60%, increased specificity of 93%, and increased accuracy of 77%.

Discussion

Dashboards in academic radiology are continuing to evolve. While a recent survey suggests their usage primarily centers on quantitative metrics of productivity such as study volume, revenue, and turn-around-time [11], there is a growing interest in the utilization of dashboards for radiology education [12–15].

Our user-friendly platform was built with the novel application of the RadLex ontology to effectively annotate radiology and pathology reports and display pathologic results for relevant imaging studies for direct feedback to the radiologist. Our initial algorithm had a sensitivity of only 47%. This was found to be at least in part to information contained in the subheading of clinical history that gave an erroneous presumption of the presence of that diagnosis in the report (i.e.,

Fig. 1 RadPath dashboard heading. An example of a typical user display with customizable date range (*left*). The matched pathology and radiology reports can be organized by the user's dictated diagnostic or procedural

("biopsy") reports. The user can provide feedback on their results as "concordant", "discordant", or "irrelevant"
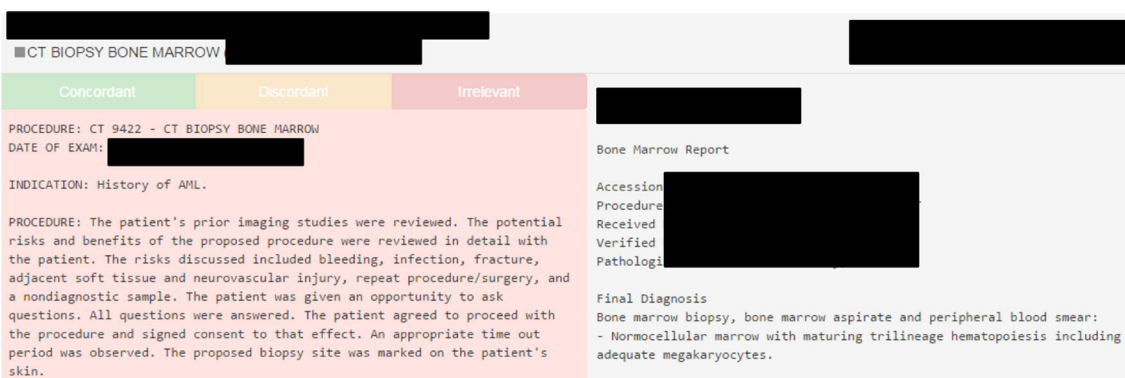
Fig. 2 Rib biopsy RadPath correlation. Radiology report (*left*) from a chest CT describing a rib chondroid neoplasm, which was confirmed on subsequent biopsy per the corresponding pathology report (*right*)

small bowel transplant as described earlier). In addition, the usefulness of including extremity ultrasound examinations in our algorithm is low as the evaluation for a deep vein thrombosis will neither prompt a biopsy of the thrombus nor typically relate in any way to a meaningful pathologic diagnosis. The yield of interesting pathology was limited enough in plain film imaging that these studies were excluded (with the exception of mammographic images). For example, many pathology reports contain descriptions of removed orthopedic hardware and were linked to pelvis radiographs. After excluding these imaging modalities and clinical history, we still had a sensitivity of 60%. This was related to many factors: terms or synonyms not found in RadLex; gaps in our algorithm; differences in reporting styles between radiologists and pathologists; and possibly others. For example, a radiology report of a sinus CT was not successfully matched with pathology results for a sinus biopsy because the term "sinus" is not in the RadLex ontology. Pleural effusion cytology reports and chest CT radiology reports were not matched because although RadLex does recognize "pleura", it does not include the term "pleural". The term "CSF" is not in the RadLex ontology, making lumbar puncture cytology results and brain and/or

spine MRIs difficult to reconcile. "Disc" and "disk" are not included as synonyms of "intervertebral disk," which could be useful. "L3-L4" could be included as a synonym of "L3/L4" to account for different ways of dictating intervertebral disc levels. "Renal" could be included as a synonym of "kidney" which would have been helpful when a radiology report refers to a "renal biopsy" or the diagnostic imager gives a differential for a renal mass but the pathology report describes the specimen as a "kidney".

There were also instances when differences in terminology between the radiologist and pathologist were not addressed by our synonym/parent term mapping algorithms. For example, while the radiology report describes a "left lower extremity" and "left tibia-fibula", the word "leg" is never explicitly used. The pathology report from the surgical debridement of this lower extremity uses the terms "debrided free flap left leg" and "left leg". The problem in rectifying this failure to match is that mapping to a too-general parent term increases the risk of false-positive matches. For example, mapping the different subsets of leg anatomy (the "tibia" and "toe") to a much broader term ("leg") could potentially result in linking unrelated pathology and radiology reports (toe lesion biopsy result

Fig. 3 Bone marrow biopsy RadPath correlation. CT-guided biopsy report (*left*) from a left iliac bone marrow aspirate with corresponding pathology report (*right*) confirming sample adequacy and remission of acute myeloid leukemia (AML)

to hip CT imaging report). Given these limitations, we were reasonably satisfied, as we prefer to keep our specificity high at the expense of sensitivity in order to reduce false-positive radiology-pathology report matches that may potentially contribute to user fatigue/disinterest in dashboard utilization, although further improvements could be made. Arguably, another limitation of our algorithm is the assumption that pathology is always the gold standard. There are situations where the pathology report is not necessarily the definitive answer, such as in cases of chondroid neoplasms, where imaging may actually provide a more specific diagnosis.

Other institutions have applied RadLex and natural language processing (NLP) to automate the creation of teaching files [16]. However, these methods have only been applied to the radiology report and have not included the pathologist's input. Our dashboard could potentially build upon these practices by automatically indexing paired radiology and pathology reports by diagnoses for an easy-to-search repository or generating resident quizzes of pathology-proven cases. We also plan to track resident utilization of our dashboard and see if there is any associated increase in the resident submission of educational exhibits and publications of case series or imaging atlases (for example, an atlas of pathology-proven prostate cancer and associated prostate MRIs). We also hope to track and identify any potential departmental trends in either the over diagnosis or under diagnosis of certain disease processes. For example, although our institution is a highly active transplant center, our accuracy rate in identifying cases of transplant rejection on imaging is currently unknown. We also hope to utilize PathBot as a replacement for the current paper-based method of tracking pathology results for our breast imaging department. Currently, a pathology addendum is made to a breast biopsy dictation only once the breast imager receives a paper copy of the pathology report (a process normally facilitated by the office manager). Arguably, the alternative of an automated email prompting the addendum could reduce the potential for human error inherent in the current paper-based system.

Potentially, our algorithm could be adjusted to alert the radiologist in cases where a pathology report was never generated despite a recommendation for a biopsy in any radiologic subspecialty. This could prompt the radiologist to contact the ordering clinician to ensure that the recommendation was received and that any barriers to the patient's biopsy may be adequately addressed or, alternatively, a follow-up examination can be scheduled. Future endeavors would also ideally incorporate input from our pathology department.

Conclusion

Presenting radiology-pathology correlation for diagnostic imaging and interventional procedures in a dashboard format is an important means for continuing educational feedback for radiologists. Application of RadLex to intelligently match radiology and pathology reports is a novel application in generating radiology-pathology correlation for both interventional and diagnostic radiologists. The main limitations of this study include a sensitivity of 60% in achieving successful pairing of relevant radiology and pathology reports, which was largely a consequence of either (1) absence of relevant terminology in the RadLex lexicon or (2) differences in terminology utilized by the radiologist and pathologist (i.e., renal versus kidney), where our synonym and parent term mapping algorithms did not address these differences. Our real-life application of the RadLex ontology uncovered potential additions to this valuable ontology. We hope our experience encourages further pursuit of regular radiology-pathology correlation in clinical practice, as well as expansion of the RadLex ontology.

References

1. Hammer MM, Shetty AS, Cizman Z, et al.: Results of the 2015 survey of the American alliance of academic chief residents in radiology. Acad Radiol. 22(10):1308–1316, 2015. doi:10.1016/j.acra.2015.07.007.
2. Health C for D and R. Regulations (MQSA)—Mammography Quality Standards Act (MQSA): https://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/Regulations/ucm110823.htm. Accessed March 13, 2017.
3. Jackson VP, Cushing T, Abujudeh HH, et al.: RADPEER scoring white paper. J Am Coll Radiol. 6(1):21–25, 2009. doi:10.1016/j.jacr.2008.06.011.
4. Verma N, Hippe DS, Robinson JD: JOURNAL CLUB: assessment of interobserver variability in the peer review process: should we agree to disagree? Am J Roentgenol. 207(6):1215–1222, 2016. doi:10.2214/AJR.16.16121.
5. Bender LC, Linnau KF, Meier EN, Anzai Y, Gunn ML: Interrater agreement in the evaluation of discrepant imaging findings with the Radpeer system. Am J Roentgenol. 199(6):1320–1327, 2012. doi:10.2214/AJR.12.8972.
6. Forman HP, Larson DB, Kazerooni EA, et al.: Masters of radiology panel discussion: defining a quality dashboard for radiology—what are the right metrics? Am J Roentgenol. 200(4):839–844, 2013. doi:10.2214/AJR.12.10469.
7. Kohli MD, Kamer AP: Story of Stickr —design and usage of an automated biopsy follow up tool. In: *Story of Stickr—Design and Usage of an Automated Biopsy Follow Up Tool.*; 2014. http://archive.rsna.org/2014/14010707.html. Accessed March 12, 2017.
8. Rubin DL, Desser TS: A data warehouse for integrating radiologic and pathologic data. J Am Coll Radiol. 5(3):210–217, 2008. doi:10.1016/j.jacr.2007.09.004.
9. Jonquet C, Shah NH, Musen MA: The open biomedical annotator. Summit on Translat Bioinforma. 2009:56–60, 2009. http://www.ncbi.nlm.nih.gov/pubmed/21347171. Accessed March 13, 2017.

10. Langlotz CP: RadLex: a new method for indexing online educational materials. Radiographics. 26(6):1595–1597, 2006. doi:10.1148/rg.266065168.

11. Mansoori B, Novak RD, Sivit CJ, Ros PR: Utilization of dashboard technology in academic radiology departments: results of a national survey. J Am Coll Radiol 10(4):283–288.e3, 2013. doi:10.1016/j.jacr.2012.09.030.

12. Gorniak RJT, Flanders AE, Sharpe RE: Trainee report dashboard: tool for enhancing feedback to radiology trainees about their reports. Radiographics. 33(7):2105–2113, 2013. doi:10.1148/rg.337135705.

13. Sharpe RE, Surrey D, Gorniak RJT, Nazarian L, Rao VM, Flanders AE: Radiology report comparator: a novel method to augment resident education. J Digit Imaging. 25(3):330–336, 2012. doi:10.1007/s10278-011-9419-5.

14. Kelahan LC, Fong A, Ratwani RM, Filice RW: Call case dashboard: tracking R1 exposure to high-acuity cases using natural language processing. J Am Coll Radiol. 13(8):988–991, 2016. doi:10.1016/j.jacr.2016.03.012.

15. Kalaria AD, Filice RW: Comparison-bot: an automated preliminary-final report comparison system. J Digit Imaging. 2015. doi:10.1007/s10278-015-9840-2.

16. Do BH, Wu A, Biswal S, Kamaya A, Rubin DL: Informatics in radiology: RADTF: a semantic search–enabled, natural language processor–generated radiology teaching file. RadioGraphics. 30(7):2039–2048, 2010. doi:10.1148/rg.307105083.