Strategies for Medical Data Extraction and Presentation Part 1: Current Limitations and Deficiencies

Bruce Reiner

Published online: 10 February 2015
© Society for Imaging Informatics in Medicine 2015

Abstract Data overload is a burgeoning challenge for the medical imaging community; with resulting technical, clinical, and economic ramifications. A primary concern for radiologists is the timely, efficient, and accurate extraction of imaging and clinical data, which collectively are essential in determining accurate diagnosis. In current practice, imaging data retrieval is limited by the fact that imaging and report data are decoupled from one another, along with the non-standardized and often ambiguous free text data contained within narrative radiology reports. Clinical data retrieval is equally challenging and flawed by the lack of information system integration, paucity of clinical order entry data, and diminished role of the technologist in providing clinical data. These combined factors have the potential to adversely affect radiologist performance and clinical outcomes by diminishing workflow, report accuracy, and diagnostic confidence. New and innovative strategies are required to improve and automate data extraction and presentation, in a context- and user-specific fashion.

Keywords Data extraction · Decision support · Imaging informatics

Introduction

One of the greatest challenges for all practicing medical professionals is data overload, which encompasses both data volume and complexity [1]. In current practice, the lack of integration between medical information system technologies exacerbates this data conundrum by creating a working environment of ubiquitous data without easy access [2]. While these challenges affect

all medical disciplines and stakeholders, radiologists are particularly sensitive due to existing time and workload constraints and de-coupling of text-based report and pixel-based imaging data. The net result is that many radiologists are forced with deciding between optimizing workflow at the expense of limited data or optimizing data at the expense of limited workflow.

Information system providers (e.g., PACS, RIS, and EMR) have attempted to address this data shortfall by adapting intelligent computerized data mining techniques to automate data extraction and directly integrate into existing workflow strategies. One commonly employed technique is to use natural language processing (NLP) to extract high priority data from historical imaging and clinical reports [3]. The lack of standardized data in existing text reports renders this process somewhat problematic and subject to inferencing and potential error [4].

Another challenge for data extraction strategies is the reality that data extraction and presentation are both context and user specific. Data requirements may fundamentally change based upon the specific task being performed (e.g., modality, anatomy, and clinical indication), individual patient (e.g., medical, imaging, and surgical histories), and radiologist (e.g., specialty training, clinical experience, and personal preferences). The net result is that technical solutions must address multi-disciplinary data requirements, lack of data standardization and integration, and contextual- and user-specific variability. If not challenging enough, one must realize that any proposed innovation will not be widely accepted and adopted into everyday practice unless it is workflow enhancing (or at least workflow neutral) [5].

Current Practice of Data Extraction

Historical Imaging Data

Before one can begin to develop a comprehensive innovation strategy which takes these multiple (and often competing)

B. Reiner (✉)
Department of Radiology, Maryland VA Healthcare System,
10 North Greene Street, Baltimore, MD 21201, USA
e-mail: breiner1@comcast.net

variables into account, existing practice and technology must be understood and analyzed for deficiencies. While each end-user (i.e., radiologist) and technology (i.e., PACS) have its own unique attributes, several common variables exist which can be used for illustrative purposes. Table 1 outlines a series of steps routinely used by radiologists in the performance of imaging exam interpretation and reporting, with some steps being routinely performed and others performed on an “as needed” basis. The number of steps employed is often variable and dependent upon the exam type, clinical indication, current exam findings, patient history, individual and institutional provider attributes, and workload.

One commonly performed step of high variability is the review of historical imaging data, which includes both prior imaging data and reports. Radiologists commonly have individual preferences and data requirements regarding historical imaging data retrieval, which is often context specific. The time radiologists devote to historical imaging data retrieval often correlates with the complexity of the imaging exam and patient medical history. Generally speaking, the larger and more complex the imaging dataset, the greater amount of time is typically allocated to historical imaging data retrieval. Needless to say, these assumptions are only valid when historical imaging data is readily available and accessible.

Since historical imaging data consists of both prior reports and images which are distinct and separate from one another (i.e., de-coupled), radiologists routinely decide whether one or both types of data are required for accurate interpretation of the current imaging study. In many cases, radiologists will rely on the textual report data alone, obviating direct visualization of the historical imaging dataset. This makes the assumption

that the prior imaging exam was accurately interpreted and the text-based report findings are easily understood and can be directly applied to interpretation of the current imaging dataset. If this assumption is not true, the radiologist interpreting the current imaging exam runs the risk of rendering an inaccurate (or non-definitive) diagnosis based upon inaccurate and/or incomplete data contained in the historical report.

The obvious way to counteract this potential liability in inaccurate and/or incomplete report data is to directly review the historical imaging dataset, in conjunction with the corresponding report. The downside of this approach is that additional time and energy are expended by the radiologist (or other operator), which can be quite significant in the case of complex and large historical imaging datasets, which may contain in excess of 1000 individual images. In the absence of annotated “key images”, the radiologist must navigate through the comprehensive dataset in search of specific images which correspond to individual report findings. In some cases, this review process may be enhanced when the radiologist authoring the prior report will incorporate text into the report specifying the specific series and image numbers of interest.

In historical report retrieval and review, there are three distinct strategies radiologists commonly employ which include comprehensive report review, isolated review of the report impression, and a hybrid approach which skims through the body of the report in conjunction with review of the impression. The strategy employed can often vary for each individual radiologist (i.e., intra-radiologist variability), based upon a number of factors including individual exam type, report content, clinical indication, and workload. It is not unusual for radiologists to modify their data retrieval patterns throughout the course of a given workday; given factors such as cumulative fatigue, changing workload demands, and time constraints. The potential for intra-radiologist variability becomes magnified in current practice by the fact that existing historical imaging data retrieval is in large part a manual process. With the exception of fairly rudimentary hanging protocols (which retrieve comparable historical imaging exams), radiologists are in large part individually responsible for the retrieval of historical imaging reports.

The net result of these combined factors is that imaging data retrieval in its current form is largely manual, highly variable, and often incomplete. Innovation is required to enhance the quality, consistency, clarity, and workflow associated with data extraction if fundamental change is to occur.

Clinical Data

While the retrieval of historical imaging data is primarily at the control of the individual radiologist, the retrieval of clinical data is largely outside of the radiologist control. The lack

Table 1 Representative steps in radiology interpretation process

1. Current exam order opened on PACS.
2. Order entry data reviewed (i.e., clinical indication and technologist notes).
3. Historical imaging folder reviewed
 - a. Prior imaging exams
 - b. Prior reports
4. Current imaging data reviewed.
5. Current and selected historical imaging exam data reviewed (and compared) in tandem.
6. Depending upon findings observed, additional historical imaging data may be reviewed.
7. Interpretation of current imaging study is performed which may incorporate added computerized tools (e.g., advanced visualization, image processing, and decision support).
8. If required, supplemental clinical data may be accessed to assist in interpretation (e.g., laboratory/pathology data, consultation report, and procedural note).
9. Report generated following completed data analysis.
10. Consultation may be performed with referring clinician or staff to communicate and/or clarify report findings.

of existing data integration between information system technologies effectively prohibits radiologists from manually retrieving clinical data. As a result, radiologists become passive participants in clinical data retrieval and must rely upon three primary sources of clinical data; order entry data provided by the ordering clinician, clinical data contained within historical medical imaging reports, and data obtained by the technologist (which most commonly is derived from direct communication with the patient and/or clinical staff).

This latter source of clinical data has undergone substantive change since the transition from analog to digital imaging practice. In analog practice, it was commonplace for technologists to obtain detailed clinical data at the time of exam preparation and record this data on a paper-based clinical datasheet, which was submitted to the radiologist along with the hard-copy images for interpretation. Once the exam interpretation was completed, this clinical datasheet would either be stored in the patient imaging folder (i.e., film jacket) for future use or simply discarded. In situations where patients routinely returned for follow-up imaging exams (e.g., screening mammography), these paper-based clinical datasheets would be stored and modified over time to reflect updates to the patient's clinical history. The ability to update and modify these clinical datasheets provided an easy to use and workflow enhancing tool for storing clinical data and taking advantage of prior data input to minimize current time and resource requirements. Since data attribution was frequently identified by the initials of the technologist who recorded the data (i.e., data source), inaccurate or incomplete data entries could easily be rectified and corrected, along with feedback provided to the data source. This provided an easy to use method for longitudinal clinical data entry and retrieval, along with internal quality assurance.

With the advent of PACS and digital imaging, many of the vestiges of analog practice were eliminated, including the analog clinical datasheet. With the exception of breast imaging, many imaging departments do not routinely use clinical datasheets in everyday practice. Instead, technologists are given the option of manually entering clinical data into the RIS or PACS in the form of "notes" which are frequently inserted alongside the order entry data. In the absence of rigorous standards and quality assurance, a great deal of previously available clinical data has become lost in digital practice.

The net effect is that the quality and quantity of clinical data provided to the radiologist has diminished over time, which has the potential to adversely affect the clinical outcomes. All radiologists have experienced cases where the clinical order entry data consists of a single word (e.g., pain, weakness, and fever). On the most superficial level, additional clinical data related to duration, location, and physical exam findings would provide valuable insight to the interpreting radiologist in focusing attention to specific anatomic locations and associated disease processes. On an intermediate level, clinical

data relating the patient's medical history, prior surgeries, and clinical data (e.g., laboratory and pathology) would provide enhanced knowledge as to differential diagnosis and organ system dysfunction. On a more granular level, clinical data relating pharmacology, genetics, and family/occupational histories provide a far greater understanding of disease probabilities and predilections. The reality is that clinical data empowers the diagnostician, assists in determining statistical probabilities, and increases diagnostic confidence (which is a frequently reported deficiency of existing radiology reports). Without clinical data and resulting knowledge, the radiologist is largely reduced to an intelligent and well trained "picture reader". Clinicians who lament the indecisiveness and ambiguity of radiology reporting must also take responsibility for providing accurate and thorough clinical data to enhance accuracy and reliability of medical imaging interpretation. In the absence of comprehensive clinical order entry data, alternative strategies must be developed to overcome these existing data deficiencies.

Review of the Literature

The ability to enhance radiologic diagnosis by providing clinical information at the time of interpretation was first proposed by Schreiber in 1963 [6]. This influence of clinical data on imaging diagnosis occurs at two stages; perception and interpretation [7, 8]. Perception refers to the identification of abnormal findings while interpretation refers to the attribution of observed abnormalities to disease.

Numerous studies have established the dependence of accurate imaging diagnosis on clinical data [9–11]. In addition to the intuitive fact that accurate clinical data improves diagnostic accuracy, the corollary is also true, inaccurate clinical data adversely affects diagnostic accuracy [12]. This interdependence between clinical data and interpretation accuracy becomes greater as patient and exam complexity increase [13, 14]. Ironically, mammography is the one area in contemporary medical imaging practice where clinical data is routinely recorded and made accessible to the interpreting radiologist.

The impact of clinical data on imaging diagnosis is not exclusive to the radiologist community, but also affects non-radiologist imaging providers. Studies have shown that diagnostic performances of cardiologists and orthopedists interpreting imaging studies are also affected by clinical data, and can actually be of greater magnitude when compared with radiologists' diagnostic accuracy [15, 16].

The criticality and existing deficiencies of clinical data are well recognized in the practicing radiologist community [4, 17]. In a survey of academic radiologists, Boonn and Langlotz [18] reported 72 % of radiologist respondents reported insufficient clinical information at the time of interpretation. While 94 % of surveyed radiologists stated that they would use

additional clinical data sources if readily available, these were currently used less than 15 % of the time, predominantly due to existing time constraints. This underscores the fact that while radiologists routinely require additional clinical data, they are currently not accessing external data sources due to technical and workflow limitations. The practical solution lies in creating an automated (or pre-populated) method of clinical and imaging data extraction, which can be iteratively refined based upon context and user-specific use.

Conclusion

The interpretation of a medical imaging exam is a multi-step process which incorporates a variety of clinical, imaging, and historical data elements; each of which plays an important role in determining diagnostic accuracy. The ability to retrieve and comprehend historical imaging data is currently limited by the decoupling of imaging and report data and conventional report format, in the form of narrative free text reports. The existing workflow model calls for radiologists to manually retrieve this data, which can result in workflow inefficiencies, inconsistent data extraction, and/or overlooked important data.

The extraction of clinical data is equally if not more fundamentally flawed due to deficiencies in clinical data availability, information system technology integration, and clinical data standardization. Existing order entry requirements are minimal and allow for clinicians to order imaging exams without sufficient clinical data to facilitate accurate and confident radiologist diagnosis. At the same time, the transition from analog to digital medical imaging practice has resulted in a paradoxical decrease in clinical data, in part due to the reduction in technologist acquired clinical data. While clinical data repositories have expanded with medical digitization, the majority of these clinical data repositories are not directly integrated with PACS, prohibiting timely data access.

These clinical and imaging data impediments have the potential to adversely affect interpretation accuracy and diagnostic confidence; both of which can adversely affect clinical and economic outcomes. It is essential that new innovative strategies be developed to circumvent these existing data deficiencies; while taking into account the workflow and productivity challenges associated with data overload. While automated data mining would serve as a logical source for innovation, the existing non-standardized format of imaging and clinical data presents challenges to traditional data mining techniques. The ultimate answer may lie in a combination of human and computer data entry, extraction, and analysis, which provides

end-users with the ability to customize data extraction templates specific to their individual needs and preferences.

References

1. Reiner BI, Krupinski E: The insidious problem of fatigue in medical imaging practice. *J Digit Imaging* 25:3–6, 2012
2. Ash JS, Berg M, Coiera E: Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc* 11:104–112, 2004
3. Demner-Fushman D, Chapman WW, McDonald CJ: What can natural language processing do for clinical decision support? *J Biomed Inform* 42:760–772, 2009
4. Reiner B: Uncovering and improving upon the inherent deficiencies of radiology reporting through data mining. *J Digit Imaging* 109–118, 2010
5. Reiner BI, McKinley M: Innovation economics and medical imaging. *J Digit Imaging* 3:325–329, 2013
6. Schreiber MH: The clinical history as a factor in the roentgenogram interpretation. *JAMA* 185:399–401, 1963
7. Loy CT, Irwig L: Accuracy of diagnostic tests read with and without clinical information: a systematic review. *JAMA* 292:1602–1609, 2004
8. Berbaum KS, Franken Jr, EA, Dorfmann DD, et al: Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. *Invest Radiol* 21:532–539, 1986
9. Zalis ME, Barish MA, Choi JR, et al: CT colonography reporting and data system: a consensus proposal 1. *Radiology* 236:3–9, 2005
10. Miniati M, Prediletto R, Formichi B, et al: Accuracy of clinical assessment in the diagnosis of pulmonary embolism. *Am J Respir Crit Care Med* 159:864–871, 1999
11. Grenier P, Chevret S, Beigelman C, et al: Chronic diffuse infiltrative lung disease: determination of the diagnostic value of clinical data, chest radiography, and CT on Bayesian analysis. *Radiology* 191:383–390, 1994
12. Zhianpour M, Janghorbani M: Effect of clinical information on brain CT scan interpretation: a blinded double crossover study. *MJIRI* 17: 173–177, 2003
13. Leslie A, Jones AJ, Goddard PR: The influence of clinical information on the reporting of CT by radiologists. *Br J Radiol* 73:1052–1055, 2000
14. McNeil BJ, Hanley JA, Funkenstein HH, et al: Paired receiver operating characteristic curves and the effect of history on radiographic interpretation: CT of the head as a case study. *Radiology* 149:75–77, 1983
15. Simons M, Parker JA, Donohoe KJ, et al: The impact of clinical data on interpretation of thallium scintigrams. *J Nucl Cardiol* 1:365–371, 1994
16. Berbaum KS, Franken Jr, EA, El-Khoury GY: Impact of clinical history on radiographic detection of fractures: a comparison of radiologists and orthopedists. *AJR* 153:1221–1224, 1989
17. Reiner BI, Siegel EL, Knight N: Radiology reporting: past, present, and future: the radiologist perspective. *J Am Coll Radiol* 5:313–319, 2007
18. Boonn WW, Langlotz CP: Radiologist use of and perceived need for patient data access. *J Digit Imaging* 22:357–362, 2009