Article published online: 2021-08-02

Data mining in radiology

Amit T Kharat, Amarjit Singh, Vilas M Kulkarni, Digish Shah
Dr. D Y Patil University, Pimpri, Pune, Maharashtra, India

Correspondence: Dr. Amit T Kharat, Dr. D Y Patil University, Sant Tukaram Nagar, Pimpri, Pune - 410 018, Maharashtra, India. E-mail: kharatamit75@gmail.com

Abstract

Data mining facilitates the study of radiology data in various dimensions. It converts large patient image and text datasets into useful information that helps in improving patient care and provides informative reports. Data mining technology analyzes data within the Radiology Information System and Hospital Information System using specialized software which assesses relationships and agreement in available information. By using similar data analysis tools, radiologists can make informed decisions and predict the future outcome of a particular imaging finding. Data, information and knowledge are the components of data mining. Classes, Clusters, Associations, Sequential patterns, Classification, Prediction and Decision tree are the various types of data mining. Data mining has the potential to make delivery of health care affordable and ensure that the best imaging practices are followed. It is a tool for academic research. Data mining is considered to be ethically neutral, however concerns regarding privacy and legality exists which need to be addressed to ensure success of data mining.

Key words: Data; data mining; hospital information system; information; knowledge; knowledge discovery; radiology; radiology data analysis; radiology information system

Introduction

Modern radiology departments have huge databases of images and text. These, like any corporate databases, are rich in data content, but poor in information content.[1] In simple terms, it means that the best use of the available radiology data does not translate into its directly benefiting patient care. An effective tool that can bridge this gap is by way of "data mining" or extracting useful information from the huge database of images and text.

Data mining technology utilizes the available data in Radiology Information System (RIS) and Hospital Information System (HIS). It instantly provides meaningful information which adds value to diagnosis, plans further patient management, saves time, and reduces costs for the entire healthcare industry. By using the data analysis tool, the radiologist can make informed decisions as well as predict the future outcome of a particular imaging finding.[2]

Access this article online	
Quick Response Code:	Website: www.ijri.org
	DOI: 10.4103/0971-3026.134367

Historical Background and Definition

Gregory Piatetsky-Shapiro created the term "Knowledge Discovery in Databases" in 1989. This term was readily accepted by the Artificial Intelligence (AI) and Machine Learning Community.[3] The term "Data Mining" was first introduced to the database community in the year 1990.[4]

It describes a process that facilitates the study of radiology data in various dimensions. It converts the large patient image and text datasets into useful information that helps in improving patient care and provides informative reports.

Process

The process of knowledge discovery occurs by analyzing data from varying perspectives and summating it into useful information.[5] It is done by making use of specialized software which assesses relationships and agreement in available information and instantly displays results in response to various radiological queries.

Using this process, the radiologist can locate and understand concealed patterns in data contained in the departmental and/or hospital database of clinical, radiological, and laboratory reports. This further helps

the radiologist in making a definitive or possible diagnosis and, therefore, directs the physician toward effective treatment.[6]

Data mining applications and software can be run on easy-to-use personal computers to a large mainframe system. The size of the application depends on the size of the database that needs to be mined and the complexity of the query.

To understand data mining, three terminologies need to be understood: Data, information, and knowledge [Figure 1]. Their concepts and representative examples are given in Table 1.

This information is usually obtained by electronic sifting through the pathology and surgical records, post-surgical findings, and subsequent follow-up of the patients. Such data was unavailable to radiologists in the past in an electronic database format.

Types of Data Mining

There are seven types of data mining [Figure 2].[7]

Classes

Here, archived data is used to retrieve data in preplanned fixed groups. Illustratively, the radiology department can

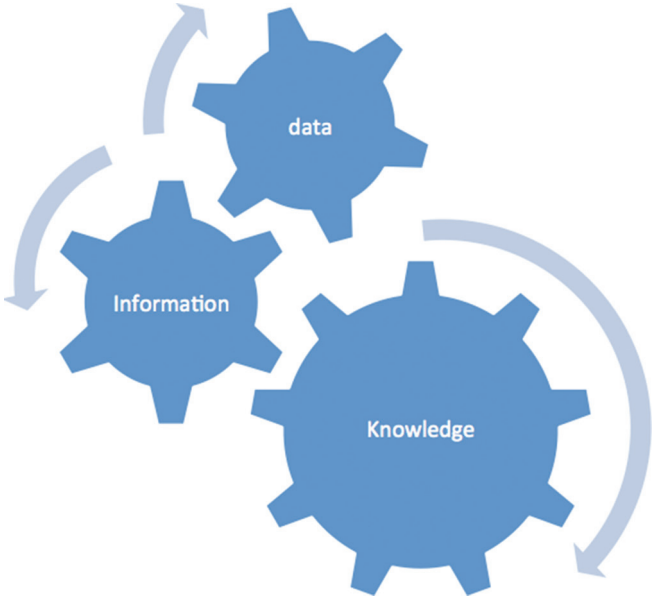

Figure 1: Parts of data mining process

Figure 2: Types of data mining

Table 1: Data, information, and knowledge: Concepts and representative examples

Examples	Data	Information	Knowledge
Term concept	Facts, numbers, or text processed by a computer is data	Patterns, associations, or relationships within the data	Information and facts skills converted into knowledge about historical patterns, future trends, and best practice
Examples	Contrast medium used and its amount	Type and pattern of enhancement	Centripetal enhancement and contrast puddling
	Diffusion and ADC*values	Diffusion and ADC* characteristics of various lesions	Diffusion restriction in fresh infarcts, epidermoid, and/or malignant lesions
	Mechanical and thermal index in USG examination	Artifacts seen in B scan orbit	Reverberation due to foreign bodies and asteroid bodies in B scan
	CA-125 levels in ovarian lesions	Imaging characteristics of ovarian mass	Eccentric mural nodules, solid components, and papillary excrescences
	Axial and longitudinal dimensions of appendix	Imaging appearance of inflamed appendix	Tubular aperistaltic blind-ended structure in right iliac fossa in acute appendicitis
	Serum amylase	Imaging appearance of pancreas	Modified CT severity index in pancreatitis
	SUV† in PET/CT	PET/CT imaging features of neoplastic masses	SUV† characteristics of masses and to assess for primary malignancies
	Cobb's angle on radiograph and CT	Imaging appearance of congenital kyphoscoliosis	Imaging characteristics of hemi vertebrae, vertebral fusion, and butterfly vertebra
	Hounsfield units in fat-containing lesions on CT	CT imaging appearances of fat-containing lesions	Negative CT density of fat-containing lesions such as lipoma, dermoids, angiomyolipomas

*ADC: Apparent diffusion coefficient, †SUV: Standardized uptake values, PET: Positron emission tomography, CT: Computed tomography

mine the data to find a particular cutoff weight which prevented a computed tomography (CT) or magnetic resonance imaging (MRI) examination. In this case, the group searched by body weight is "cross-sectional imaging by CT and MRI." Similarly, for purposes of quality assurance, data regarding exposure factors for CT, mammographic and radiographic examinations can be mined from a single group modality using ionizing radiation for diagnostic purposes."

Clusters

Data items are grouped according to coherent relationships or radiology preferences. Example: renal stone patients may need to undergo conventional radiography, ultrasonography (USG), intravenous urography (IVU), CT urography, and nuclear renal scans. Likewise, brachial plexus trauma patients need recommendation for MRI and CT myelographic studies. These are examples of clusters and one can predict the other investigations based on the clinical data.

Associations

Data can be mined to identify connections. For instance, a patient referred for MRI joint frequently undergoes conventional radiograph of the affected joint or associated joints. Patients referred for mammography studies may need to undergo sono-mammography. If this investigation is not done, the software automatically anticipates, recommends, and books an appointment.

Sequential patterns

Data is mined to foresee behavior patterns and inclination. Nonfunctioning kidneys or poorly excreting kidneys on IVU can be triaged to perform nuclear renal scans. Altered marrow signal in the vertebral bodies on MRI can be subjected to nuclear bone scans. Lesions which are multisystem can be predicted to undergo a positron emission tomography-computed tomography (PET/CT).

Classification

This is also referred to as machine learning technique. Example: pregnant ladies who undergo first trimester screening without nuchal translucency (NT) measurement will be suggested for a rescan at 11-13 weeks if the prior report does not state a comment on NT. Studies where NT scan is already performed can be rescheduled for anomaly scan at a specific date.

Prediction

A data mining technique that studies the relationship between autonomous variables and the relationship between non-autonomous and autonomous variables. Example: productivity of radiology department for the current year can be predicted based on current data and previous historical data. This can also be categorized as modality specific, specialist specific, and radiologist specific.

Decision tree

This is a frequently used data mining technique, as it is an easy model to understand. In decision tree technique, the foundation of the decision tree is a basic situation or position that can have multiple answers.[7] Example of a decision tree is highlighted in Figure 3.

Data mining consists of five major elements. The elements perform the actual process of data handling as outlined in Figure 4.

- All information in RIS and HIS is first extracted, loaded, and transformed in the main database server. This process requires all pertinent data such as radiology reports, technician notes, and history and laboratory findings to be stored in the main database
- The data is then stored in the database and managed in the multidimensional database system
- A mode allows easy and rapid access of this data to researchers, radiologists, and other specialists
- Analysis of data is then performed by using varying grades of complexity using application software
- Finally, data is presented in a useful format, such as a graph or table.

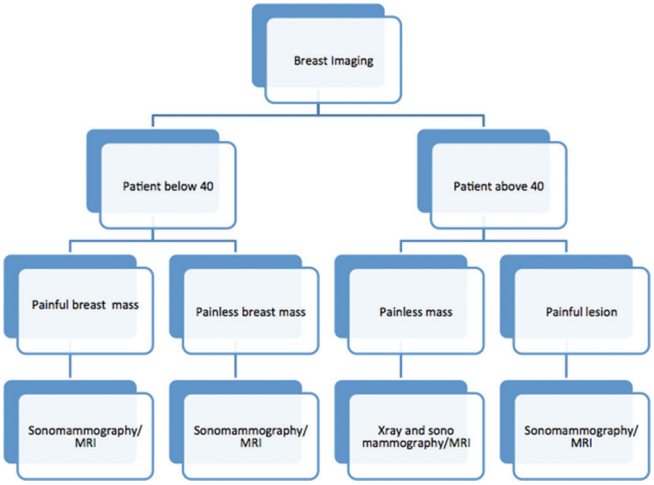

Figure 3: Example of decision tree data mining process

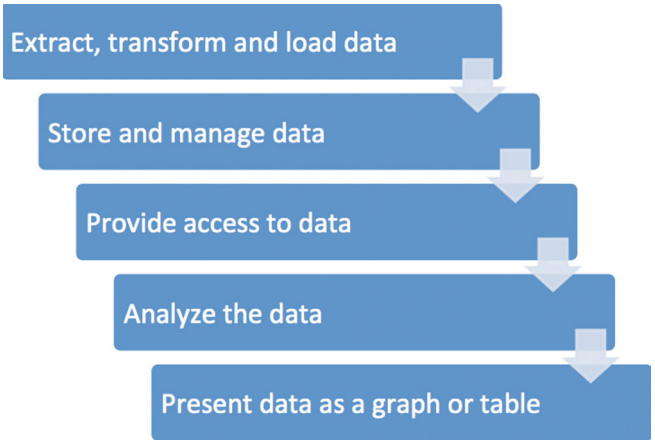

Figure 4: Elements of data mining

Different levels of analysis are available [Figure 5].

Advanced analytical programs can provide us instantaneous feedback tools such as business and clinical analytics.

Business analytic tools give us information on equipment utilization, personnel utilization, merging and blending of services.

Clinical analytic tool can give us information on scan follow-up, peer feedback, radiation dose limitation, and scheduling patients.[8]

Prior to establishing data mining protocols for the radiology database, the data needs to be pre-treated. Patterns can be discovered within data only if they are present in the selected sample. The sample of data that needs to be mined is cleaned to remove non-useful data and presented for data mining procedures.

The above process can be well exemplified by a study done in 2011 by Harvard Medical School, Boston, where the authors were able to extract the dose of radioactive drugs given to individual patients, from 204,561 unstructured nuclear medicine reports in their department in the last 25 years.[9] They created an open-source toolkit for the purpose of extraction of radiation exposure data from the reports. First, the text was extracted from each report and duplicate statements were deleted. Within each report, all units of radioactivity were then converted to a standardized format ("mCi" for millicurie and "uCi" for microcurie). Next, the unit of radioactivity was matched to user-defined array of radio-pharmaceuticals (14mCi of Tc-99 Sestamibi). The results were manually validated against randomly selected 2359 reports, which yielded a recall rate of 97% and a precision of 98.7%.

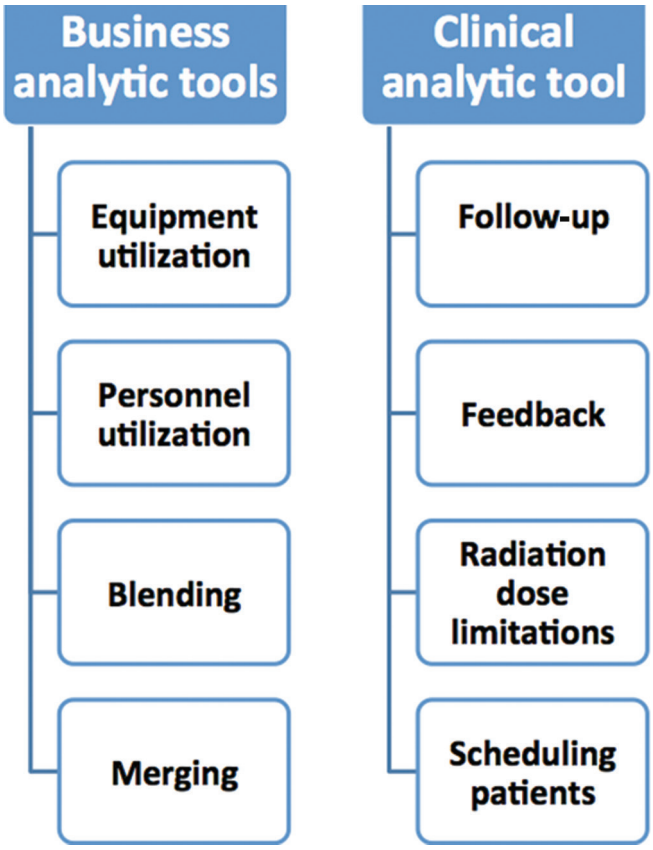

Figure 5: Levels of data mining

Before conducting a search on the radiology data for research by a radiologist or radiology administrator, it is imperative to understand how to conduct a search. Again, considering the example of bone metastasis, the researcher may also need to search for terms such as "skeletal metastasis" and "osseous metastasis"; otherwise, these can be omitted from the search. Current algorithms built in the Picture Archiving and Communication System (PACS) and RIS do not allow such synonym-based search. However, the system can be made to learn and use this information by installing complex algorithms. Currently, we lack a "standard lexicon" which can be a hurdle in achieving the complete merits of data mining process. However, steps are being taken in this direction, such as Radiological Society of North America's (RSNA) initiative RadLex - a valuable tool with more than 30,000 terms for systematic storage and archival and retrieval of radiology information.[10] Developers and researchers can use this archive to develop data mining tools.

The process of data mining can be made significantly easier by improving ways of designing Digital Information and Communications in Medicine (DICOM) and PACS reporting systems, where proper planning beforehand will make the data more searchable. A good way to enhance the discoverability of data is to make part of the report properly structured, highlighting the keywords prior to saving reports with future searches being restricted only to that part of the report rather than the whole text. This would significantly reduce the complexity of the software involved, saving time, curtailing costs, and improving the quality of recovered data.

The future of data mining also holds prospects for image-based data mining. Image-based data mining concept is based on the color, texture, and shape of the image.[11] Images with similar appearances can be mined from a database; however, the process is complex due to the large image size. Larger the data, more complex is the algorithm and computational costs to mine the image-based data. The complexity can further increase if the data is from cine loops and video files.

Advantages

The advantages of data mining are presented in Table 2.

Data mining can aid in making proper decision making in terms of better diagnosis, follow-up, and choosing the

Table 2: Summary of advantages and disadvantages of data mining

Advantage	Disadvantage
Instant	Privacy
Real time	Ethical issues
Adds knowledge to reports	Confidentiality
Evidence-based medicine	
Reduced costs	
Meaningful use of technology	

Table 3: Query complexity

Categories	Example	Query complexity
Intra-modality	Calculation of average IV contrast requirement for CT studies per day	Simple query: Single modality
Inter-modality	Number of liver masses diagnosed on imaging	Mild complexity: Inter-modality query
Inter-departmental	Hepatocellular carcinoma: Typical imaging features and their confirmation by HPE*	Moderate complexity: Inter-departmental query
Intra-hospital	Hepatocellular carcinoma, confirmation, follow-up, findings post chemotherapy and surgery	Complex query: RIS, PIS† and HIS query
Inter-hospital	Hepatocellular carcinoma stage IV follow-up	Complex query: Using multiple hospital study centers

*HPE: Histopathological evaluation, †PIS: Pathology information system, RIS: Radiology information system

right modality. It can identify the best imaging protocols and proper treatment planning. Overall, it can make delivery of health care affordable.

Queries can be simple intra-modality or complex inter-modality, intra-hospital, as highlighted in Table 3. Data mining can identify the effective treatments and best practices while patients can receive better care.

It is a tool for academic search. A search query on "bone metastasis" in the software will identify reports with related terms such as bone metastases, metastases in the bone, metastasis in the appendicular or axial skeleton, and skeletal metastases. This kind of information tool can help students and researchers in archiving of cases. Images of cases can be pulled out instantly from the system without losing valuable time.

Similarly, specific queries as in the case of a student who wants to study cases of "hypersensitivity interstitial pneumonitis" or "aortic aneurysm" can be searched in the RIS and HIS with instantaneous access. This can assist in thesis,research, paper writing and retrospective case studies on a multitude of topics covering various aspects of radiology, pathological and surgical follow-up.

Disadvantages

The disadvantages of data mining are presented in Table 2.

Data mining follows the accepted principles of right and wrong that govern the conduct of a profession, and is therefore considered to be ethically neutral.[12]

However, there are questions regarding privacy, legality, and ethics.[13] Data mining requires data preparation, which can uncover information or patterns that may compromise confidentiality and privacy obligations.

Conclusion

Data mining has the potential to change the face of radiology practice due to advances in software and hardware with better integration of HIS and RIS. Data for data mining is generated in radiology departments; this can act as a huge storehouse of knowledge, which can turnaround existing patterns of practice and improvise efficiency in workflow. Radiologists and radiology administrators can use data mining to efficiently manage the radiology practice and make radiology reports informative. Studies have shown that data mining can be an asset to drive workflow in a radiology department as compared to onsite workflow assessment, thereby saving time and effort.[14] Data mining can provide impetus to radiology research by saving time and be cost-effective in the long run allowing meaningful use of technology.

References

1. Doug A. Data Mining. Available from: http://www.laits.utexas.edu/~norman/BUS.FOR/course.mat/Alex/. [Last accessed on 2014 Jan 01].
2. Howell WL. Data Mining and Analytics in Radiology. PACS and Informatics, RIS. 2012. Available from: http://www.diagnosticimaging.com/pacs-and-informatics/data-mining-and-analytics-radiology. [Last accessed on 2014 Jan 1].
3. Fayyad U, Piatetsky-Shapiro G, Smyth P. From Data Mining to Knowledge Discovery in Databases. AI Magazine 1996;17:37-54.
4. Mena J. Machine Learning Forensics for Law Enforcement, Security and Intelligence. Boca Raton, FL: CRC Press (Taylor and Francis Group), Auerbach Publications; 2011.
5. Craig D. Cloud Computing History 101. 2010. Available from: Http://www.constructioncloudcomputing.com/2010/08/14/cloud-computing-history/. [Last accessed on 2014 Jan 01].
6. Kantardzic M. Data Mining: Concepts, Models, Methods, and Algorithms. John Wiley and Sons; 2003.
7. Data Mining Techniques by Zentut. Home/Data Mining/Data Mining Techniques. Available from: http://www.zentut.com/data-mining/data-mining-techniques. [Last accessed on 2014 Jan 01].
8. Trevino M. Radiology Tackles Broad Range of Applications with Data-mining by Diagnostic Imaging. 2005. Available from: http://www.diagnosticimaging.com/articles/radiology-tackles-broad-range-applications-data-mining. [Last accessed on 2014 Jan 01].
9. Ikuta I, Sodickson A, Wasser EJ, Warden GI, Gerbaudo VH, Khorasani R. Exposing exposure: Enhancing patient safety through automated data mining of nuclear medicine reports for quality assurance and organ dose monitoring. Radiology 2012;264:406-13.
10. What is RadLex? Available from: Http://www.rsna.org/RadLex.aspx. [Last accessed on 2014 Jan 01].

11. Sahu M, Shrivastava M, Rizvi MA. Image Mining: A New Approach for Data Mining Based on Texture. Computer and Communication Technology (ICCCT), 2012 Third International Conference; 2012. p. 7-9.
12. Seltzer W. The Promise and Pitfalls of Data Mining: Ethical Issues. Fordham University. Available from: Http://www.amstat.org/committees/ethics/linksdir/Jsm2005Seltzer.pdf. [Last accessed on 2014 Jan 01].
13. Pitts JW. The End of Illegal Domestic Spying? Don't Count on It. Washington Spectator; 2007. Available from: http://en.wikipedia.org/wiki/Data_mining. [Last accessed on 2014 Jan 01].
14. Lang M, Kirpekar N, Bürkle T, Laumann S, Prokosch HU. Results from data mining in a radiology department: The relevance of data quality. Stud Health Technol Inform 2007;129:576-80.

Cite this article as: Kharat AT, Singh A, Kulkarni VM, Shah D. Data mining in radiology. Indian J Radiol Imaging 2014;24:97-102.
Source of Support: Nil, **Conflict of Interest:** None declared.

ERRATUM

Indian Journal of Radiology and Imaging 2014; Vol 24; Issue 1

Title: Reading in Devanagari: Insights from functional neuroimaging
Page 44-50

Corrected version of this article available on the below given link:
http://www.ijri.org/article.asp?issn=0971-3026;year=2014;volume=24;issue=1;spage=44;epage=50;aulast=Singh

The error is regretted

- Editor, IJRI