

A peek into the future of radiology using big data applications

Amit T. Kharat, Shubham Singhal

Department of Radiology, Dr. D.Y. Patil Medical College, Hospital and Research Centre, Dr. D.Y. Patil Vidyapeeth, Pune, Maharashtra, India

Correspondence: Dr. Amit T. Kharat, Flat No. 6, Ashok Chakra One, Lane No. 7, Koregaon Park, Pune - 411 001, Maharashtra, India.
E-mail: kharatamit75@gmail.com

Abstract

Big data is extremely large amount of data which is available in the radiology department. Big data is identified by four Vs – Volume, Velocity, Variety, and Veracity. By applying different algorithmic tools and converting raw data to transformed data in such large datasets, there is a possibility of understanding and using radiology data for gaining new knowledge and insights. Big data analytics consists of 6Cs – Connection, Cloud, Cyber, Content, Community, and Customization. The global technological prowess and per-capita capacity to save digital information has roughly doubled every 40 months since the 1980's. By using big data, the planning and implementation of radiological procedures in radiology departments can be given a great boost. Potential applications of big data in the future are scheduling of scans, creating patient-specific personalized scanning protocols, radiologist decision support, emergency reporting, virtual quality assurance for the radiologist, etc. Targeted use of big data applications can be done for images by supporting the analytic process. Screening software tools designed on big data can be used to highlight a region of interest, such as subtle changes in parenchymal density, solitary pulmonary nodule, or focal hepatic lesions, by plotting its multidimensional anatomy. Following this, we can run more complex applications such as three-dimensional multi planar reconstructions (MPR), volumetric rendering (VR), and curved planar reconstruction, which consume higher system resources on targeted data subsets rather than querying the complete cross-sectional imaging dataset. This pre-emptive selection of dataset can substantially reduce the system requirements such as system memory, server load and provide prompt results. However, a word of caution, "big data should not become "dump data" due to inadequate and poor analysis and non-structured improperly stored data. In the near future, big data can ring in the era of personalized and individualized healthcare.

Key words: Automation; Big Data; Hospital Information Systems; informatics; Computer Aided Diagnosis; Picture Archiving and Communication Systems; Radiology Information Systems; virtual radiologist; Watson

Introduction

Big data is identified by four Vs^[1,2] –Volume, Velocity, Variety, and Veracity [Figure 1]. Let us see how radiology Picture Archiving and Communication Systems (PACS) data can meet these criteria.

Volume

Data available in radiology departments is high in volume as we can note by image size of computed tomography (CT),

angiography, radiography, and mammography images. Datasets are particularly large for cardiac CT studies and angiographic studies.

Velocity

This is the speed at which data is generated. Radiology data gets churned at high speed. CT and Magnetic Resonance Imaging (MRI) scanners are already pushing freshly acquired data in PACS, subsequently images are stored in short and long-term storage using Vendor

Access this article online

Quick Response Code:



Website:
www.ijri.org

DOI:
10.4103/ijri.IJRI_493_16

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

Cite this article as: Kharat AT, Singhal S. A peek into the future of radiology using big data applications. Indian J Radiol Imaging 2017;27:241-8.

Neutral Archives (VNA). This data gets generated in real time.

Variety

Radiology data is rich in variety, as the images come from various sources, these can be from Digital Radiography, Computed Radiography, Conventional Radiography, Interventional Radiology, MRI, Ultrasound, Colour Doppler, and Positron Emission Tomography–Computed Tomography (PET-CT) studies.

Veracity

This relates to the authenticity and credibility of the dataset. This must actually be the aim of an ideal big data project; however, there is also a need to do a proper and systematic analysis of data to get accurate results. Therefore, suboptimal scan, scans with motion artefacts, and incomplete studies can be deleted from the study group by quality checks. This will help in maintaining the uniformity of datasets.

To study such vast data and review these for clinically relevant and useful information using standard algorithms is unthinkable considering the time required to search for useful information. However, by applying different algorithmic tools and converting raw data to transformed data from such large datasets, there is a possibility of understanding and using such data for gaining new knowledge and insights.

Radiology departments are best suited to migrate to big data analysis of radiology images because of readily available information technology (IT) infrastructure.^[3] Busy radiology departments always have a PACS which is connected to the Hospital Information Systems (HIS) and Radiology Information Systems (RIS) to streamline the work flow. Connecting this huge library of images and their additional nonimage data such as patient demographics, history and laboratory records is not a difficult task compared to the digital environment in the rest of the departments within

the hospital. In addition, radiology departments are rich in information with a huge variety of data available on a platter where the data can be analysed and strengthened with algorithms.

Radiology data can be structured or unstructured [Figure 2] and in the form of image and nonimage data [Figure 3]. The accuracy of big data depends on the quality of the data. If the data available in the radiology departments is properly structured, then there is a high probability of finding desired information from such datasets. This can help to derive maximum valuable information, which can potentially modify decision support in patients.

The aim of using big data is to convert unstructured data to wisdom after extracting knowledge and information from given data[Figure 4].^[4] Here, wisdom means meaningful use and clinical relevance of the data.

Big data analytics and the process of analysis of complex data

Big data analytics consists of 6Cs^[5] [Figure 5]

Connection

This consists of networks between the various CT, MRI, Ultrasound, and Radiography equipment’s and patient order entry systems, which are the main source of patient image data, and the HIS, RIS, PACS, which are the main source of nonimage data.

Cloud

It means storing radiological data on off-shore (remote) servers hosted by proprietary internet networks and connecting them to the local hospital computers for fast access, processing and distribution of large data.

Cyber

It describes the computer processing power and memory that will be required to process the query to obtain specific answers. Complex query requires large main frame

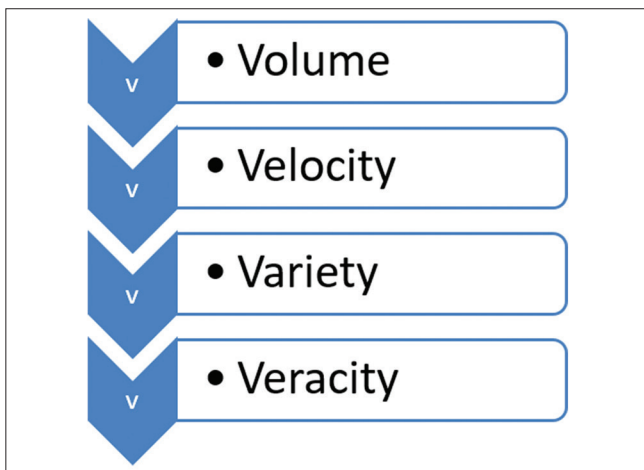


Figure 1: Four V's of Big Data

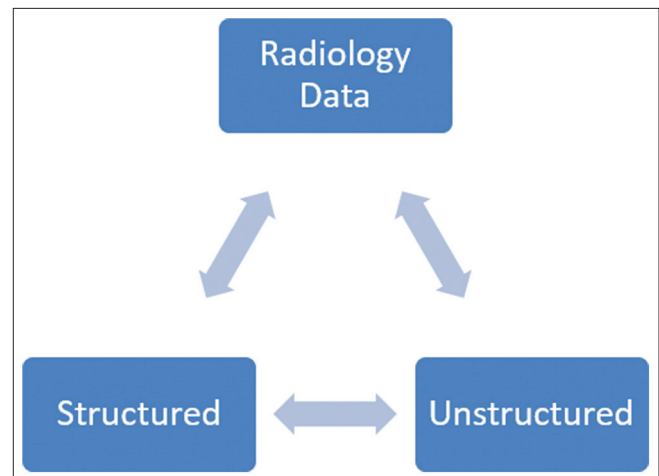


Figure 2: Types of radiology data

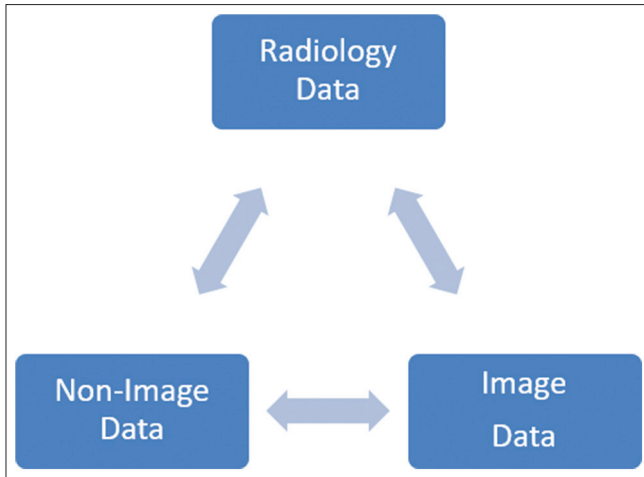


Figure 3: Forms of radiology data



Figure 4: Aim of Big Data

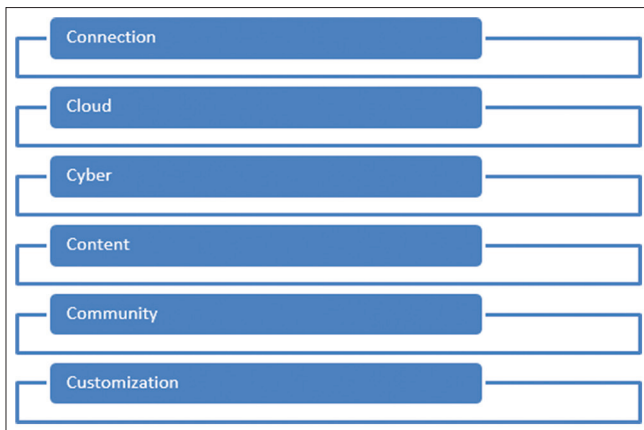


Figure 5: 6C's of Big Data Analytics

computers and higher processing power to obtain desired results in real time or near real time.

Content/context

This refers to the Digital Imaging and Communications in Medicine (DICOM) datasets which can be searched and information derived to get relevant, meaningful results and correlation, which can help alter patient line of management promptly.

Community

This implies to sharing data of a particular and sensitive nature with similar available data in other health care institutes to determine consistency and to get required information. This collaboration and sharing of radiology data can be of special interest in infectious diseases such as chest radiographic findings in H1N1 influenza. Big data

can be of potential help in early recognition of such imaging findings.

Customization

Radiology structured and non-structured data query should be customized and algorithms should be available on demand to answer specific queries to make this technology of clinical relevance and drive home the future of personalized healthcare. In this manner, big data can add value to radiology.

Process of big data analytics and implementation

There are three important steps in the analysis and implementation of big data:

Big data infrastructure: This refers to the connections, cloud, and computer processing powers. This also includes the data warehouses where the radiology data is stored.

Big data science: The art of extracting information by developing simple and complex algorithms. This is the engine of big data.

Big data execution: The algorithms used to extract and avail clinical useful information.

These tools should be made readily available to data scientists and knowledge workers such as radiologists and other specialty doctors handling radiology images and reports.^[6]

To facilitate large scale use of big data applications in radiology, there is an urgent need to incorporate structured reporting templates and standard radiology lexicon in the reporting process. This has two-fold advantages. First, this will help in decreasing confusion among the radiologists and referring doctors in the interpretation of the lesions described in the radiology reports. Secondly this will also help in easier comparison in follow-up imaging reports. Thirdly use of such lexicons can facilitate big data applications in creating specific radiology reporting templates by using disease specific keywords.

Intricacies of big data

The global technological prowess and per-capita capacity to save digital information has roughly doubled every 40 months since the 1980s.^[7] Since 2012, every single day 2.5 exabytes (2.5×10^{18}) of new data has been created, and as of 2014, each day 2.3 zettabytes (2.3×10^{21}) of data has been generated.^[8] Current research estimates approximately 2.5 quintillion bytes of data being created each day.^[9]

Processing large data using the current relational database management systems, standard desktop statistics and visualization dashboards, is a near impossible task. Instead,

the need of the hour is using “massive parallel software deployed on large number of servers.”^[10]

Possible use of big data in radiology in the near future

By using big data, the planning and implementation of radiological procedures in radiology departments can be giving a great boost. It can be used in the context of nonimage data such as that already present in the HIS, RIS, and image data already present in PACS and Vendor Neutral Archive (VNA).^[11]

When can big data be used?

Right from the time the patient walks in for an appointment to the dispatch of reports. The application can start when patient comes to schedule an appointment and continues when the technologists performs the scan and during the process when the radiologists reviews the images and generates reports. It can potentially support at each step in this complex process.

Just a few of potential applications of big data in near future

Though these are majorly potential big data application, boundaries get blurred as it is simultaneously working in tandem with other applications such as Cloud Computing, Data Mining, Deep Learning, and Computer Aided Diagnosis (CAD) [Figure 6].

In intelligent planning, scheduling of studies, and delivering patient-specific instructions

Examples

Prior to imaging

- Patients who have undergone cardiac stenting can be prompted to produce chest radiographs before scheduling MRI appointments
- Patients with right upper quadrant pain will be automatically triaged in the morning for ultrasound
- Patient’s scheduled for hysterosalpingography appointments after entering data regarding last menstrual period (LMP) may be provided auto prompts for the best dates for conducting the examination
- Female patients in reproductive age group undergoing

radiation-related examination can be prompted to give the date of the LMP

- CT studies for abdominal pain and bowel disorder may prompt the use of bowel-specific contrast agents under authorization by the radiologist

Applications numbered 1 to 5 given above are examples of big data analytics coupled with intelligent appointment scheduling systems

- Patient with liver lesion for MRI liver may be given prompts for liver-specific contrast on the dashboard during imaging or in RIS. This can then be authorized by a radiologist to decide on the type of liver MRI contrast.

Big data analytics can thus make the standard RIS into a “smart and intelligent RIS”.

During imaging

- CT studies for renal stones may produce direct prompts to technologists to do a plain CT
- Cardiac CT and Coronary Calcium scores may prompt radiologist regarding potential artefacts, and at the same time after analysis of the body habitus, weight, and heart rate, the scanner can give the best scan parameters required to image the patient. Same can also be applied for CT chest or body CT examinations, thus paving the way for “individualized and personalized imaging parameters”.

During image analysis

Assisting radiologists in decision support:

- After performing body CT or MRI, big data analytic tools can be used to perform texture analysis and correlate these findings with HIS data and give prognostication in a dashboard as well as tag the image for future correlative studies
- Spectroscopic analytic results can be comprehensively evaluated by using big data processes where similar data explored over a period of time can be evaluated to find the results and likely diagnosis in current scenarios
- SUV value assessment can be done intelligently using PET-CT, areas of highest metabolic activity can be auto highlighted. Big data can help across modalities for co-registration across images of the same region. Here, areas of high PET values can be cross-checked with similar areas of high cellularity (areas with diffusion restriction) on the MRI images or solid enhancing lesions on contrast CT and MRI studies by using co-registration protocols
- Intelligently comparing tagged areas on scans to assess change in volume of lesions can be a possibility in the near future
- Evolving infarcts on CT can be missed, especially if they are less than 6–12 hours old (hyper acute strokes). Big data can help in assessing the texture based on given history to find such cases and warn of a potential missed infarct in real time. This algorithm can be used in all patients of sudden onset weakness

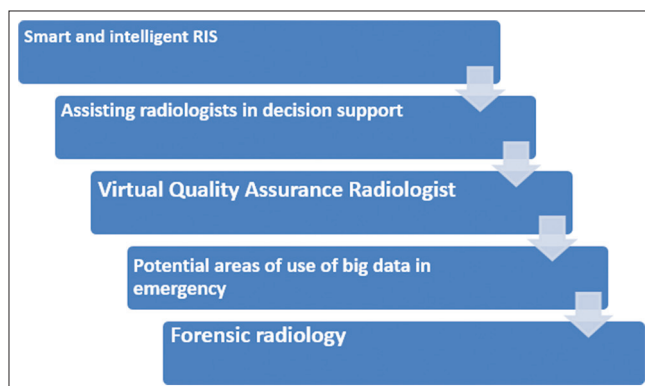


Figure 6: Potential applications of Big Data

- Big data can help in preparing direct reports of cases such as bleed, and mention size, shape, and midline shift; this is one place where the process can be automated. This algorithm will involve a complex interplay between big data analytics, computer aided diagnosis and deep learning tools
- By using dynamic contrast enhancement texture analysis (DCE-TA) tools,^[12] contrast enhancement patterns of liver can be used to diagnose subtle lesions and identify disease free liver parenchyma. Big data analytics of such data may lead to interesting results and need to be explored
- Serial accurate size and volume analysis of lung tumours or solitary pulmonary nodules is possible using big data metrics to decide temporal evolution of lesion. Big data can add these metrics which are difficult to assess using the naked eye. Thus, big data can be the “quintessential third eye” of a radiologist
- By using Hounsfield Units (HU) density differences on CT examinations, big data can decide possible contents of a suspected collection—serous, exudative, calcification, fat or air. This can be improvised by using big data applications than what has been done conventionally till date. This information can be available by a big data study of similar density lesions studied on scan by sifting through the HIS. This HU value however will need to be standardized prior to using these process and a range of values need to be identified prior to its implementation
- Radiologists perform pattern analysis for disorders, particularly in the lung. Big data based pattern analysis modules can detect areas of ground glass opacities, honey combing, reticular densities, fibrosis and thereby give a list of possible differentials in these situations using tools of computer aided diagnosis (CAD)
- In MRI studies of the knee, shoulder and spine, big data analytics can study bone bruise and contusion patterns and can recreate the sequence of events and method of injury likely in such scenarios opening a completely new dimension in trauma imaging.

Big data as a virtual quality assurance tool for radiologist

Big data based analytic tools can be the best quality control assessment tool in the hand of radiologists. After the initial diagnosis, the radiologists can run specific focused algorithms.

Example

In renal tumours, after performing contrast study, the radiologist can run an algorithm to check contrast enhancement characteristics, and such data metric can be compared to prior results and other data from the pathology department to give a narrowed specific differential list. In this manner, the big data based applications can help in quality control and assist a radiologist rather than replace them. Similar big data based tools can be applied for brain tumours or breast mass imaging. Big data based applications

may point to areas of abnormal density or intensity and alert the radiologist which may have been missed or overlooked otherwise. Here the big data tools will work hand in hand with deep learning algorithms.

These algorithms can be implemented by training the system to utilize the massive power of cloud computing.

Big data thereby paves the way to the concept of a “virtual radiologist” of the future, helping in monitoring scans in real time situations, analysis of parameters, and prompting suitable real time actions.

Potential areas of use of big data in emergency

- It has the potential to diagnose pneumothorax
- In CT and MR brain with intraparenchymal or extra-axial bleed it can give volume of bleed and midline shift
- It can be used to detect extraluminal air in the abdomen and renal stones
- Foreign body detection in radiographs and CT studies.

In the near future, it is in this area that automatic report generation can be attempted, above examples again show a complex interplay with deep learning and computer aided diagnosis.

Big data as a forensic radiologist

Dental mapping and radiography for bone age estimation in medicolegal cases to reduce inter-operator variability and to create new tools for “radiological bone fingerprinting” by trabecular pattern assessment in bones.

Big data should not be “Bug Data”

The crucial aspect in big data is designing algorithms to be run on structured datasets for analysis. This can save time and give real time results. Failure to implement this synergy will render “Big data to Bug data.” This highlights the importance of having good quality data. This will also save valuable time, computing power, and memory to find unknown variables.

Extracting information from the images

This is a challenging and crucial part in big data. It's using big data to read radiology images. There is a lot of practical difficulty in doing these studies, and it requires a lot of computing power and IT infrastructure. Moreover, the reports also need to be standardized to aid in “customized patient centric decision support system”. Few projects are already on, such as the IBM's Watson's^[13] supercomputer to provide decision support to radiologists in reading radiology images and initial results and literature is being made available. One such example is software Avicenna tested on the IBM supercomputer Watson for cardiology and breast imaging.^[14]

However, there are other ways and means to find important and relevant data in images such as “image tagging.”^[13]

Images can be tagged for specific lesions such as mass and stones by performing the image search. Big data in conjunction with data mining, and deep learning algorithms is the key to access this metadata and other wealth of information hidden in the radiology images.

Targeted use of Big Data applications can be done for images by supporting the analytic process by focusing on a region of interest, plotting its anatomy, and then running big data application on that specific body part, which can substantially reduce the system requirements and provide prompt results. Programs such as computer-aided lung informatics for pathology evaluation and rating (CALIPER) developed by biomedical imaging resource lab at the Mayo Clinic^[15] can be one such step. Algorithms and tools such as these in the future will be available on a platter within the department. Before running such application, the system should inform in advance the potential time required to search the data and expected time for results to be achieved.

Big data Practices (processing algorithms and security)

As is true with other IT solutions, such as cloud computing and data mining, here again there is a concern for safety and threat to data security as data is hosted on different cloud-based platforms and is accessed by various vendors and sourced from various platforms such as CT, MRI, digital X-rays, and mammography. This is sensitive and private information, and therefore, systems need to be HIPAA compliant. Layers of security can be built in with authorizations. Patient data need to be anonymized and the vendors need to be sensitized for using the best encryption protocol to protect patient health information (PHI). The overall progress and use of big data in radiology is held back due to issues such as privacy, security, and proprietary nature of data. Various health regulations limit the use of radiology and health data for analysis or research due to fear of breach of HIPAA or similar health regulations, and such concerns cannot be undermined.

Similarly, “Big data should not become Dump data” due to inadequate and poor analysis and non-structured improperly stored data. To avoid such issues there is a need for creating large data centres where images are stacked and properly categorized for rapid access. The data needs to be standardized and segregated in various data sets for example: CT chest non-contrast and CT contrast enhanced examination, CT angiography study. This makes it easy to search for data and run specific query. Such data needs to be anonymized so that they can be safely kept in research data centres, processed, and thoroughly analysed. The data centres need to be chain linked to promote these as “global virtual radiology research platforms”.

The radiology data centres thus created need to have 4 essentials [Figure 7]: rugged and protected infrastructure with data warehouses, well planned system with security

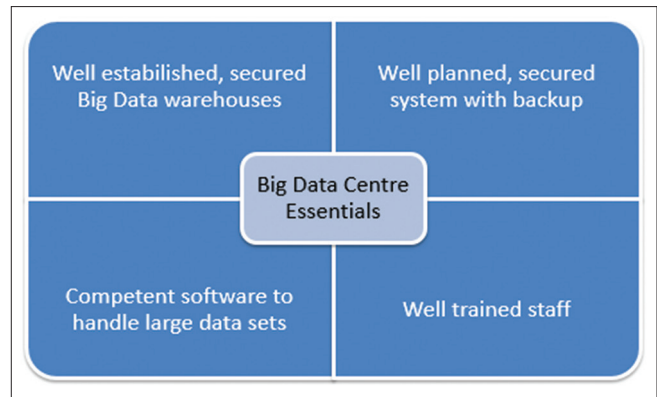


Figure 7: Big Data Centre Essentials

and back up, software that can handle all the raw data, process and transform it, and finally well trained IT staff.^[16]

Big data certification

Centres and radiology departments that are storing data in a standardized format and using methodical practices can be certified to be big data compliant. This can make the process of running and performing big data execution streamlined and cost effective to obtain useful and relevant results. This big data should be insured against possible theft, loss, or physical damage.

What we can do now to assist big data application

As big data applications are coming in vogue and algorithms are being tested. This is the right time for proper systematic data collection, systematic image tagging, and systematic plotting of data along with nonimage (text) data such as information in RIS and HIS to have clear patient demographics; this will help in future big data projects. Radiology departments which will be data rich, are uniquely placed to derive maximum benefits when proper algorithms are used to process the data. Therefore, now “it pays to have curated data”. Curated data here refers to organized, functionally fragmented data which provides important inferences after running analytical algorithms. The more the data in the radiology departments the more mature search results can be obtained and executed. Data which was once considered a liability to maintain is now can be considered a valuable asset. Radiology departments and administrators may see benefits in keeping data for a longer period of time, considering the information and wisdom they may acquire from these datasets in the coming future.

An ideal big data utilized patient management radiology work flow [Figure 8A and B]

In the near future when big data become inbuilt into the PACS, RIS, and HIS work flow [Figure 9], it can be a game changer as shown.

The below example again explains the use of tandem applications such as cloud computing, data mining, deep

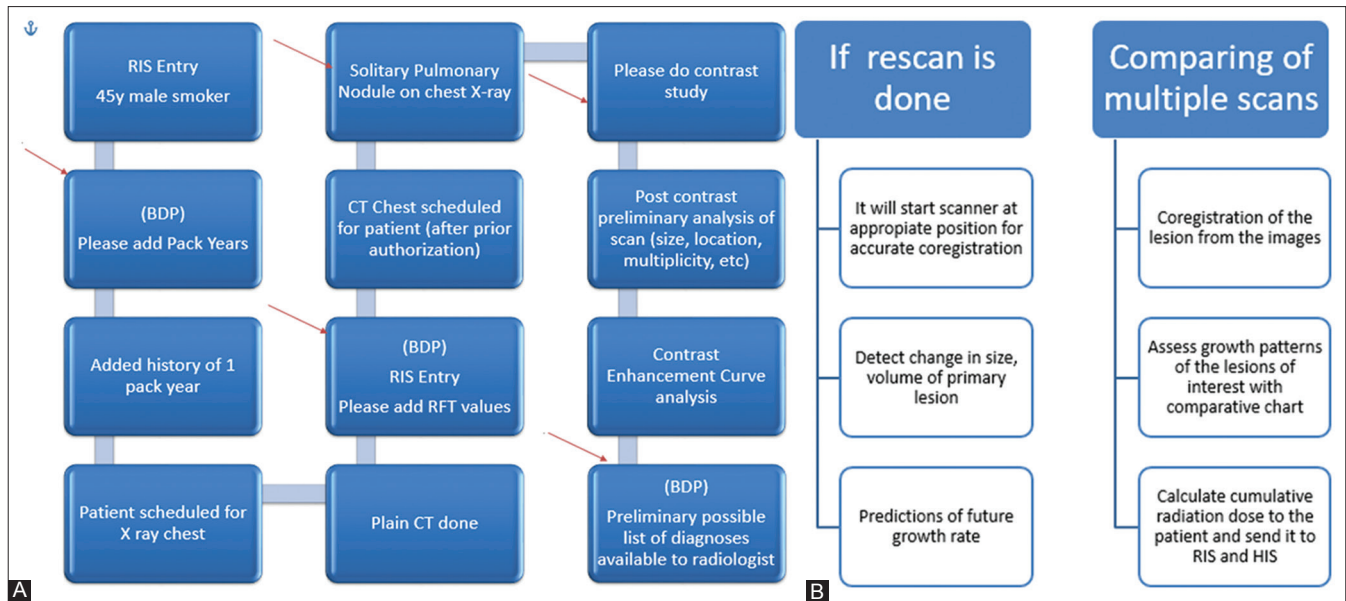


Figure 8 (A and B): (A, B) Ideal Big Data work flow. BDP – Big Data Prompt. Red arrows in this scenario indicates Big data system prompts

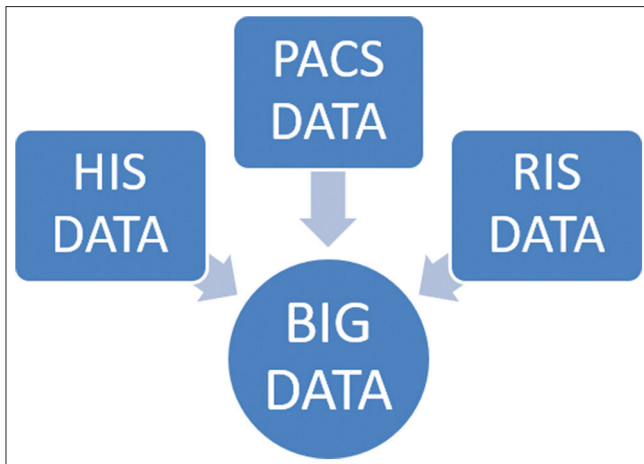


Figure 9: HIS, RIS, PACS workflow

learning, intelligent scheduling, patient management systems and CAD chain linked through big data analytics.

For a 45-year-old male patient with a history of cough and smoking, big data will prompt to ask for entering “pack years.” If X-ray chest reveals a suspicious nodule. The software prompts for a “CT chest.” Auto schedule for CT chest will be created. After authorization by the physician through the HIS and after confirming that no prior CT chest is available and the patient has consented to the exam, the further process begins.

During examination, the scanner will prompt for “CT thorax.” Plain CT scan will be done; if the scanner detects nodule, analytics tools will colour code and highlight it after a pattern analysis to give size and volume and will give a real-time recommendation to perform “contrast.”

After appropriate authorizations, technologist will go ahead with contrast study. The scanner will give auto prompt and calculate exact “correct contrast dose.” It will begin study at the proper time after detecting the contrast density and perform a contrast study; it may take delayed section and cover liver and adrenals. It will also auto prompt to take “contrast enhanced curves” of the solid area and suggests for possible sites for biopsy. Similarly, it will shift through the available data for the age, sex of the patient, and size of the nodule to predict its latitude of malignancy, this can be the crucial part of the big data analytics tool. This will be an ideal big data work flow chart. When the patient comes for follow-up, the scanner will automate the entire process, calculate the exact site of prior scans, and start scans at the same positions for co-registration, giving comparative analysis of the nodule size and volumes. It will inform about the interval change to understand its doubling, and therefore, predict future growth and changes with current medications because of its integration in HIS and its understanding of prior characteristics based on already fed information from similar prior experiences. It will also calculate the radiation dose received in the prior exam and current exam, giving a cumulative dose standard and graphical display. This information is available in the image itself or the metadata in the system.

Conclusion

Big data has the potential to usher in the era of personalized and individualized healthcare.

It starts with a systematic collection of data and ends with proper processing to obtain accurate and timely results. Big data is the logical next step in the evolution of radiology



Figure 10: Roadmap of the paper

departments. It can transform busy radiology departments and help in efficient management and providing intelligent and smart patient care options. It can improve the quality of performed scans, assist radiologists in decision support and can act as a virtual quality control tool for radiologist. Over a period of time, it can self-learn to find hidden information within the reports and images which are rather difficult to interconnect, or find a relationship using the standard routine or conventional protocols, and here lies the real advantage of this technology. In the near future, big data will work to assist radiologists by providing intelligent and targeted decision support rather than replacing radiologists. It is a technology whose time has come. Figure 10 shows roadmap of this paper.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

- IHTT. Transforming Health Care through Big Data Strategies for leveraging big data in the health care industry, 2013. [Last accessed on 27 July 2016].
- IBM Big Data Platform-Bringing Big Data to The Enterprise. Available from: www-01.ibm.com. N.P., 2016. Web. [Last accessed on 27 July 2016].
- Kansagra AP, Yu JP, Chatterjee AR, Lenchik L, Chow DS, Prater AB, *et al*. Big Data and the Future of Radiology Informatics. *Acad Radiol* 2016;23:30-42.
- Rowley J. The Wisdom Hierarchy: Representations of The DIKW Hierarchy. *J Info Sci* 2007;33:163-80.
- Big Data Infographics Images. Tech Talk. Available from: <https://tech-talk.org/2015/01/23/big-data-infographics-images/Web>. [Last accessed on 28 July 2016].
- Health Imaging. (2016). Diving Into the Business of Big Data in Radiology. Available from: <http://www.healthimaging.com/topics/imaging-informatics/diving-business-big-data-radiology?nopaging=1> [Last accessed on 26 Oct 2016].
- Hilbert M, López P. The World's Technological Capacity to Store, Communicate, and Compute Information. *Science* 2011;332:60-5.
- IBM What is big data? – Bringing big data to the enterprise. Available from: www.ibm.com. [Last accessed on 28 July 2016].
- Vcloudnews.com. (2017). Every Day Big Data Statistics – 2.5 Quintillion Bytes of Data Created Daily. [online] Available at: <http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/>[Last accessed on 27 Mar 2017].
- Rui P. Why Is Big Data So Big?. Available from: Blog.wedotechnologies.com. N.p., 2016. Web. [Last accessed on 28 July 2016].
- Agarwal TK, Sanjeev. Vendor Neutral Archive in PACS. *Indian J Radiol Imaging* 2016;22:242.
- Ganeshan B, Burnand K, Young R, Chatwin C, Miles K. Dynamic Contrast-Enhanced Texture Analysis of the Liver. *Investigative Radiology* 2011;46:160-8.
- Big Data in Radiology Will Drive Personalized Patient Care. Available from: AuntMinnie.com. N.p., 2016. Web. [Last accessed on 28 July 2016].
- Simonite T. IBM'S Automated Radiologist Can Read Images And Medical Records. *MIT Technology Review*. N.p., 2016. Web. [Last accessed on 30 July 2016].
- The Essentials: Information Technology for The Practice. *Radiology Business*. N.p., 2016. Web. [Last accessed on 28 July 2016].
- Maldonado F, Moua T, Rajagopalan S, Karwoski RA, Raghunath S, Decker PA, *et al*. Automated Quantification of Radiological Patterns Predicts Survival in Idiopathic Pulmonary Fibrosis. *Eur Resp J* 2016;43:204-12.