Neuroradiology imaging database: using picture archive and communication systems for brain tumour research

Yang G L, Tan Y F, Loh S C, Lim C C T

ABSTRACT

Introduction: Disease registries and databases form an important component of research in clinical medicine, and can be useful to support retrospective studies and prospective clinical trials. However, analysis of radiological imaging databases has not been explored: imaging and clinical data often exist as separate silos of information, even in modern digital-enabled hospitals in Singapore. We describe a computerised method for creating a radiological research database using data from the picture archive and communication system (PACS) and hospital information system (HIS).

Methods: Using a relational database and Java programming language, we created the neuroradiology imaging database (NRID). A web-interface for keyword searches were tested with the clinical data from PACS of a tertiary referral hospital for neurological diseases. Keyword and wildcard searches were conducted for various brain neoplasms and compared to HIS discharge diagnosis.

Results: The NRID was deployed successfully and keyword search could be completed in real time. Lists of patients with meningioma, oligodendroglial tumour, neurocytomas, cerebral abscess, and neurocysticercosis could be exported and compared with the HIS discharge diagnosis. Patients with neurological diseases could be obtained by manually combining lists.

Conclusion: An imaging database can be created using clinical PACS data, which can enable keyword search functions to support brain tumour research. Radiological databases can help support clinical research, but further work needs to be done in order to take full advantage of the potential of digital health information.

Keywords: biomedical research, database management systems, picture archive and communication system, radiology information systems, neuroradiology

Singapore Med J 2007; 48(4):342–346

INTRODUCTION

Databases form an important component of medical research. Disease registries and research databases are important for epidemiology and health service research, and can be used to support a variety of other clinical studies and potentially data mining in future.[1–3] Currently, various investigators and consortia are establishing initiatives in local disease registries in order to support retrospective and prospective clinical studies in Singapore.[4] Although clinical databases have been well established, image databases have been largely overlooked. The potential of using a large body of radiological imaging data to establish epidemiological statistics, normative values, and outcomes research has not been explored, either alone or in conjunction with larger clinical research databases.[5]

Recent trends and advances in the computerisation of radiological images, stored in a picture archive and communications system (PACS), have resulted in most local hospitals having access to image information in the digital format in the radiology department. However, although a knowledge repository would be useful to support clinical and radiological research, none of the PACS vendors currently provides a solution for the creation of research databases as a commercial option to support applications other than clinical throughput. Furthermore, the advantages of digital records are negated by the disparate systems within any single hospital. Hence, even though a wealth of useful clinical, research and teaching material is hidden in the PACS, a mechanism to identify, catalogue or analyse relevant cases does not exist.[6] Data mining and sharing of details between incompatible computer systems are

Biomedical Imaging Laboratory, Agency for Science Technology and Research Singapore, 30 Biopolis Street, #07-01 Matrix, Singapore 138671

Yang GL, MSc
Software Engineering Manager

Tan YF, B Eng
Intern

Casemix Office, Tan Tock Seng Hospital, 11 Jalan Tan Tock Seng, Singapore 308433

Loh SC, MA
Assistant Director,

Department of Neuroradiology, National Neuroscience Institute, 11 Jalan Tan Tock Seng, Singapore 308433

Lim CCT, MBBS, MMed, FRCR
Senior Consultant and Adjunct Associate Professor

Correspondence to: Dr Tchoyoson CC Lim
Tel: (65) 6357 7021
Fax: (65) 6358 1259
Email: tchoyoson_lim@nni.com.sg

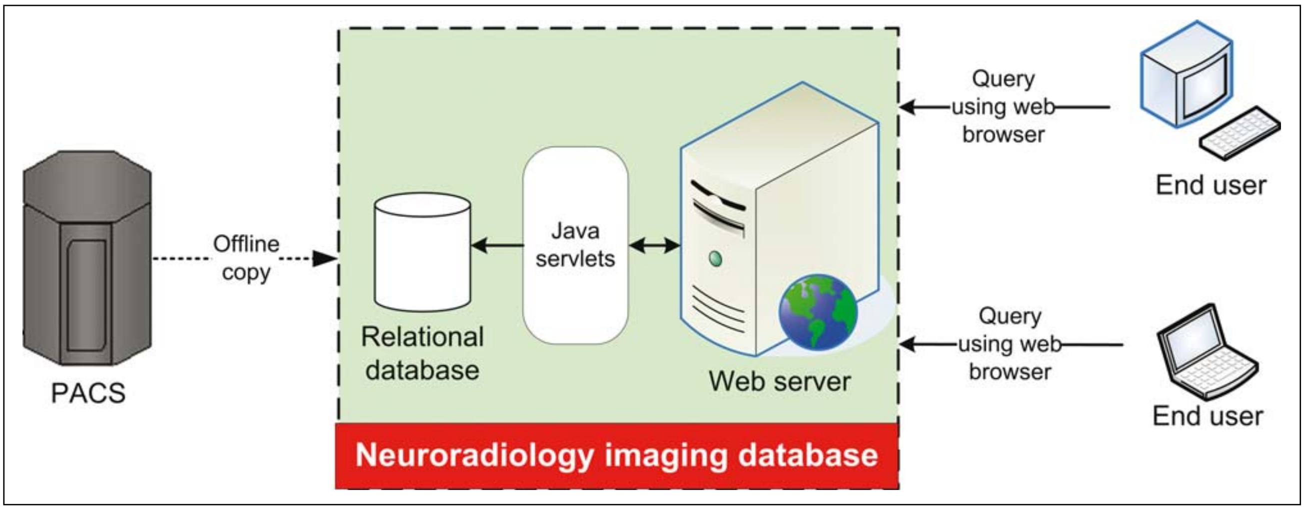

Fig. 1 System Diagram. There are three major components of the Neuroradiology Imaging Database (NRID): a relational database, Java servlets and a web server. The database PACS information is first copied into the database. End users can access the NRID through a web browser and submit queries. Web server receives the query, invokes the Java servlets to search in the database, and returns the query results to the end users in web pages.

difficult, and collaborative research between hospitals is greatly hampered. For example, clinical research in neuro-oncological imaging is hampered by lack of comprehensive lists of patients with brain tumours who have undergone neuroimaging. A method that can identify patients with a particular disease (such as brain tumour) and their imaging features would be desirable. We describe the neuroradiology imaging database (NRID), a radiology research database containing information from PACS, that incorporates a simple keyword search interface to generate patient lists that can be used for research.

METHODS

We designed a client-server system based on a relational database and used Java programming language to create a web-interface in order to search for and export relevant data in the form of patient lists for research. The NRID was populated with the caseload of the clinical PACS of the Neuroradiology Department, National Neuroscience Institute (NNI). Data from the search results was then compared and combined with other data sources such as hospital information system (HIS) discharge diagnosis to yield definitive lists of patients with neurological diseases such as brain tumours for research analysis.

The relational database forms the core of the NRID system (Fig. 1), with database fields including information that can be matched to Digital Imaging and Communication in Medicine (DICOM) standard formats exported from PACS. These fields include patient identifying parameters (such as patient name, date of study), image descriptors (such as modality, slice thickness, pulse sequence), and most importantly, the clinical report (a free text radiological report typically describing the study findings and diagnosis). We deployed Oracle 9i (Oracle, Redwood, Shores, CA, USA) on a networked personal computer running Windows 2000 (Microsoft, Redwood, WA, USA) operating system. We also developed a web interface module that interacts with the database so users can navigate and search the database using a web browser (such as Internet Explorer, Microsoft, Redwood, WA, USA). The web interface incorporates a search function that can query DICOM fields, such as dates, series descriptors and free text search within the examination radiological reports for text keywords (such as "meningioma" or "cysticercosis"). In addition to simple keyword search, search using wildcards (such as "oligo*") (Figs. 2 & 3) and Boolean ("MRI" AND "abscess") filters were enabled. Initial data from 2000 to 2004 was imported from clinical PACS and tested.

The HIS discharge diagnosis comprises a searchable spreadsheet (Microsoft Excel, Redmond, WA, USA) that contains the final discharge diagnosis coded into the inpatient clinical system of the Tan Tock Seng Hospital (TTSH). Only those patients who underwent neuroimaging at NNI and were represented in the clinical PACS in NNI were included in this database. After discharge from the inpatient service, the "International Classification of Diseases, 9th Revision, Clinical Modification" (ICD-9-CM) codes of primary and secondary diagnoses, and Morphology (M) code for neoplasms, if applicable, were recorded. This project was approved by the Institutional Review Boards and Medical Boards of both hospitals.

We tested the NRID and HIS discharge diagnosis for lists of patients with meningioma, oligodendroglial tumours, neurocytoma, medulloblastoma, cerebral

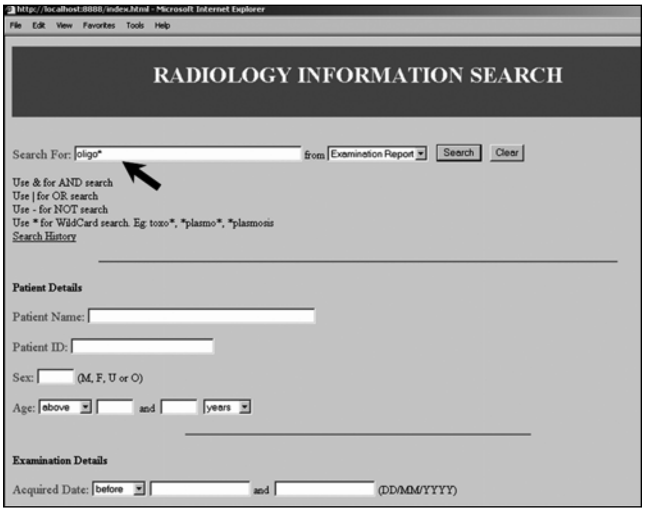

Fig. 2 Keyword search function. Screenshot of the NRID search page showing wildcard free text search (in this example – for "oligo*", arrow) within the Examination Report field. Note that Boolean search (AND, OR and NOT), and search within fields other than Examination Report (such as age range and acquisition dates) can also be performed.

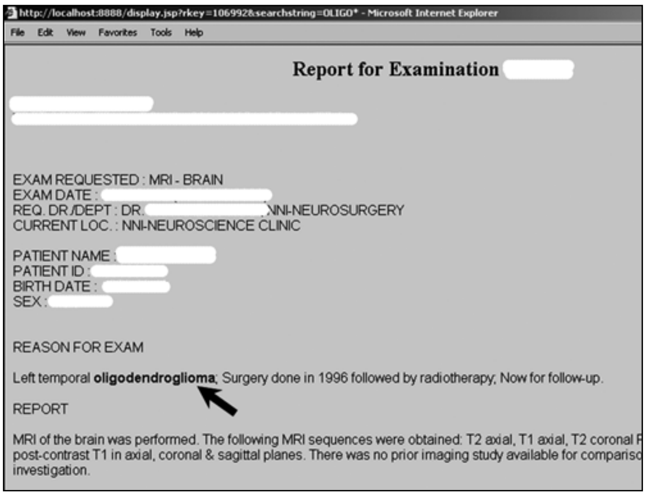

Fig. 3 Keyword "hit". Screenshot showing an example of a positive result ("hit") in the keyword search for "oligo*" in Fig. 2. The word "oligodendroglioma" appeared in the Reason for Examination (arrow) and Impression (not shown).

abscesses and neurocysticercosis. Using simple text search or wildcard search functions, lists of positive occurrences (or hits) in the NRID were exported (in Microsoft Excel, format). The HIS discharge diagnosis was also searched for the equivalent ICD diagnosis code or tumour M code and similarly exported. These lists were then manually compared and combined to form the final list of patients with the diagnosis.

RESULTS

An operational NRID client server computer system was installed and tested at the Neuroradiology Department, NNI, Singapore. This is a tertiary referral centre for neurological diseases and serves an 800-bed acute hospital, 29,000 outpatient visits and 25,000 radiological investigations annually, comprising computed tomography (CT), magnetic resonance (MR) imaging, diagnostic and interventional angiography. The PACS database from January 2000 to December 2004 was successfully exported to the NRID, and searched for keywords and wildcard keywords. Boolean search using "AND" and "OR" functions were also successful. Lists of keyword matches could be exported in spreadsheet format using the software. Keyword text and wildcard searches were completed in 3–10 minutes, depending on the complexity of the search.

The HIS discharge diagnosis database was also successfully created. A list of all unique patients could be extracted from the NRID and compared against the HIS records (which contains all TTSH inpatients including NNI as well as non-NNI patients), from January 2000 to December 2004. Positive matches were selected and discharge diagnosis was successfully recorded for all patients found in the NRID. This formed the HIS discharge database, which contained relevant ICD diagnosis codes. Using the "Find" function in Microsoft Excel software for ICD numeric codes, manual selection of patient records could be performed to identify patients with specific diagnosis. However, the process of extracting selected cases and for manual comparison and subsequently the time taken with NRID results took several days for each diagnosis.

The results of NRID keyword and discharge diagnosis search are shown in Table I. For all keyword searches, large numbers of "hits" were returned, with many patients having many "hits" in multiple different studies. For example, a patient with a tumour may have a "hit" in CT on diagnosis, another on MR post-surgical baseline study and a third MR or CT on long-term follow-up imaging. These duplicate studies from a single patient had to be manually reduced or concatenated to yield lists of individual patients. Patient lists from the HIS discharge diagnosis were similarly duplicated by patients with multiple admissions, each with a unique discharge summary, for instance two episodes of inpatient admission for surgical resection of original and recurrent tumour, possibly with malignant progression of tumour grade or complications such as wound infection.

The keyword "meningioma" was a common occurrence in the radiological reports, as it is one of the commonest intracranial neoplasms. These were found in the radiological reports of PACS records, usually describing the diagnosis (e.g. "likely to be from meningioma"). However, the keyword also appeared in negative reports in patients who did not have meningiomas

Table I. Results of neuroradiology imaging database and discharge summary keyword search in patients with tumour and tumour-like mimics.

Keyword/ICD-9 Code (M Code)	Meningioma/ (M953)	Oligo*/ (M945)	Neurocytoma/ (M9506-0)	Medulloblastoma/ (M947)	Abscess/ 324	Neurocyticerosis/ 123
Number of NRID occurrences	2447	225	23	109	961	29
Number of unique patients from NRID	601	38	15	38	154	12
Number of patients from discharge summary	282	28	3	4	132	14
Number of common patients	197	21	3	4	116	10

but other tumours (e.g. "unlikely to be meningioma") and these were also recorded as "hits". Comparing the two patient lists obtained using NRID and HIS discharge diagnosis in most neurological diseases, a patient would correctly appear on both lists. However, some patients appeared only on one list and not the other. Those patients that were in NRID but not HIS discharge diagnosis included patients who underwent only outpatient neuroimaging and therefore did not have an inpatient episode. Patients with HIS discharge diagnosis but were not found on the NRID list included patients in whom the radiological report did not specify a pathological diagnosis matching the ICD codes (e.g. "malignant primary tumour" in a patient with oligodendroglioma). Despite this discrepancy, lists of patients with a final discharge diagnosis of neurological disease and who underwent neuroimaging could be compiled by combining lists.

Using the wildcard search for "oligo*", various types of oligodendroglial neoplasms could be found in the NRID, including oligodendroglioma, oligodendroastrocytoma, anaplastic oligodendroglioma and anaplastic oligodendroastrocytoma. These "hits" occurred mainly in the "reason for request" portion of the text rather than diagnosis or comments, as these patients were on follow-up neuroimaging for brain tumour surveillance. In the search for central neurocytoma, similar results were found. Medulloblastoma search yielded very few inpatients discharge instances as this tumour primarily affects the paediatric population and the NRID "hits" were mostly from outpatient neuroimaging studies referred from the KK Women's and Children's Hospital. Finally, tumour-like space occupying lesions from cerebral abscess and neurocysticercosis were also successfully searched. In these cases, we found several instances of HIS discharge diagnosis miscoding of neurocysticercosis as abscesses or "seizures".

DISCUSSION

Using the data from clinical PACS, the NRID of neurological diseases could be created, and keyword searches could yield lists of patients with specific diagnoses. Although primitive in its current form, combining NRID and HIS discharge diagnosis can potentially be a useful tool for medical research. In brain tumour imaging, this may represent a small step towards a comprehensive research database, especially as it can access the rich but unorganised data held within clinical PACS. The web-based search interface could be intuitively and quickly manipulated. We had previously described a case-based electronic radiology teaching file interfaced with PACS that can be applied to medical education.(7-9) This educational resource and the current NRID initiative are examples of how PACS can be used for research and teaching beyond its clinical utility.(10-12)

We demonstrated lists of patients with tumour and tumour-like mimics, and were able to find common as well as rare neurological diseases, using both PACS imaging data as well as HIS discharge diagnosis. Due to the nature of the data collection in each database, there were unique patients (such as outpatients with neuroimaging not reflected in inpatient HIS) in each dataset. The data completeness of each diagnosis was also dependent on the individual epidemiological and presentation patterns, and paediatric tumours or rare and specific pathological diagnosis (such as central neurocytoma) may be under-represented in the database. However, an advantage of our study was that data from NRID and HIS discharge diagnosis could be combined. Previously, data from such individual "silos" were isolated even though combining them would have increased the yield of patient numbers and improved the quality of research.

The NRID and HIS discharge diagnosis in their current form do not represent a full-featured research database. The simple search function in NRID is limited to keywords only, and it is capable of exporting patient lists but there is no provision for cleaning up databases to decrease redundancy and improve completeness. It is also not linked to existing clinical trials research database, although this would be desirable to further enhance collaborative research. There are also many limitations to keyword text search compared to more powerful structured reports and natural language search functions.(13,14) For example, keyword "abscess" search will also identify the phrase "not abscess" or "abscess would be unlikely". However, though this currently represents an obstacle to creating a complete list of patients with the correct diagnosis, the data is also valuable for health services research and outcomes research to identify misdiagnosis and see how decision-making can be improved. Although the actual search function in NRID using software was quick and was usually accomplished within a few minutes, manual extraction and comparison with HIS discharge diagnosis were extremely laborious. A software method for automating the laborious manual identification (and elimination) of such duplicate cases would be desirable.

Radiological reports are also hampered by a plethora of words to convey the same meaning, and apart from mammography, which uses the BIRADS lexicon, there is no widely accepted standard neurological medical lexicology.[15] Different words such as "haemorrhage", "haematoma" or "haemorrhagic lesion" could be used to describe the same finding. Furthermore, the current situation of free text radiological report may be enhanced by DICOM structured reporting, which could organise reports into diagnostic fields and facilitate the search method.[16-18] Another area of research focus could be classification by imaging features, instead of merely by pathological diagnosis. A method of describing, classifying, and segmenting the images themselves, using digital image metadata could be useful. Finally, studying other more complex neurological diseases or syndromes such as stroke, epilepsy or trauma, which unlike neoplasms do not have a single diagnosis, may be a more challenging prospect. Future studies and software development may be helpful to eventually allow data mining and more sophisticated research.

In conclusion, we describe a computerised method for creating a radiological research database incorporating keyword search in PACS. This enabled us to combine imaging database with hospital discharge summary data to identify patients with neurological diseases. Although NRID and HIS discharge summary represent early efforts at imaging research, further work needs to be done in order to take full advantage of the potential of digital health information to support medical research.

ACKNOWLEDGEMENTS

The study was supported partly by the Biomedical Research Council, National Healthcare Group and Singapore Health Service grants PTD/03004 and PTD/01010. This paper was presented as part of a free paper at the Singapore Malaysia Congress 2005. The authors thank Dr Wieslaw Nowinski and Dr Francis Hui for their suggestions and support.

REFERENCES

1. Toga AW. Imaging databases and neuroscience. Neuroscientist 2002; 8:423-36.
2. Wong ST, Hoo KS Jr, Cao X, et al. A neuroinformatics database system for disease-oriented neuroimaging research. Acad Radiol 2004; 11:345-58.
3. Sasso G, Marsiglia HR, Pigatto F, et al. A visual query-by-example image database for chest CT images: potential role as a decision and educational support tool for radiologists. J Digit Imaging 2005; 18:78-84.
4. Tan KP, Hewitt RE. Design a secure web-based tissue repository database system: ethical considerations. Cell Preserv Tech 2005; 3:131.
5. Yousem DM, Bryan RN, Beauchamp NJ Jr, Arnold AM. A national neuroimaging database: a call to action. Acad Radiol 2004; 11:829-31.
6. Siegel E, Reiner B. Electronic teaching files: seven-year experience using a commercial picture archiving and communication system. J Digit Imaging 2001; 14 (2 suppl 1):125-7.
7. Lim CC, Yang GL, Nowinski WL, Hui F. Medical Image Resource Center – making electronic teaching files from PACS. J Digit Imaging 2003; 16:331-6.
8. Yang GL, Lim CC. Singapore National Medical Image Resource Center (SN.MIRC): a world wide web resource for radiology education. Ann Acad Med Singapore 2006; 35:558-63.
9. Yang GL, Lim CCT, Nowinski WL. A method and apparatus for creating radiological teaching files from clinical image archive. United States patent application no. 10/307190, 2002.
10. Yang GL, Aziz A, Narayanaswami B, et al. Informatics in radiology (infoRAD): multimedia extension of medical imaging resource center (MIRC) teaching files. Radiographics 2005; 25:1699-708.
11. Yang GL, Lim CCT, Banukumar N, Aziz A, Hui F. Design and deployment of a large brain image database for clinical and non-clinical research. In: Ratib OM, Huang HK, eds. Proceedings of the Medical Imaging 2004: PACS and Imaging Informatics; 2004 Feb 14-19; San Diego.
12. Yang GL, Lim CCT, Narayanaswami B, Hui F, Nowinski WL. A large brain image database for education and research. In: Lemke HU, Inamura K, Vannier MW, et al, eds. Proceedings of the 17th International Congress on Computer Assisted Radiology and Surgery; 2003 Jun 25-28; London.
13. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. J Am Med Inform Assoc 1994; 1:161-74.
14. Mendonca EA, Haas J, Shagina L, Larson E, Friedman C. Extracting information on pneumonia in infants using natural language processing of radiology reports. J Biomed Inform 2005; 38:314-21.
15. Ohno-Machado L, Boxwala AA, Ehresman J, Smith DN, Greenes RA. A virtual repository approach to clinical and utilization studies: application in mammography as alternative to a national database. Proc AMIA Annu Fall Symp 1997; 369-73.
16. Pendergrass HP, Greenes RA, Barnett GO, et al. An on-line computer facility for systematized input of radiology reports. Radiology 1969; 92:709-13.
17. Langlotz CP. Automatic structuring of radiology reports: harbinger of a second information revolution in radiology. Radiology 2002; 224:5-7. Comment on: Radiology 2002; 224:157-63.
18. Radlex: a lexicon for uniform indexing and retrieval of radiology information resources. Available at: www.rsna.org/RadLex/. Accessed June 23, 2006.