

Combined DICOM and HL7 viewer in support of a bridge from content-based image retrieval to computer-aided diagnosis

Petra Welter, Fatih Topal, Sebastian Jansen, Thomas M. Deserno, Jörg Riesmeier, Christoph Grouls, Rolf W. Günther

Abstract Content-Based Image Retrieval (CBIR) is a field of rising interest for the application in Computer-Aided Diagnosis (CADx). Exploiting the visual information hidden in the images for retrieval, it facilitates the identification of similar past examinations, thereby providing a second opinion. Still, CBIR is not an integral part of a radiologist's daily work. A comprehensive representation of CBIR results in a standard format utilizing established clinical infrastructure together with all referenced comparable examinations will support a bridge from CBIR to CADx. In this paper, we introduce the general IRMACON viewer, a system that represents results from CBIR systems encoded as a DICOM Structured Reporting document in a layout adjusted for CADx. IRMACON also allows the insight into identified similar examinations for convenient comparison by accessing patient's information and diagnostic findings from the Hospital Information System (HIS) using HL7 messages. Our system is embedded into the clinical setting and the radiologist's workflow. We applied the IRMACON viewer to the Image Retrieval in Medical Applications (IRMA) framework.

Petra Welter, Fatih Topal, Sebastian Jansen, Thomas M. Deserno
Department of Medical Informatics, RWTH Aachen University, Aachen, Germany
e-mail: pwelter@mi.rwth-aachen.de

Jörg Riesmeier
ICSMED AG, Oldenburg, Germany

Christoph Grouls, Rolf W. Günther
Department of Diagnostic Radiology, RWTH Aachen University Hospital, Aachen, Germany

1 Introduction

Content-Based Image Retrieval (CBIR) [1] relates to the identification of images from a database using the visual information hidden in the images instead of attached textual annotations. Computer-Aided Diagnosis (CADx) can generally be defined as “a diagnosis made by a radiologist who uses the output of a computerized scheme for automated image analysis as a diagnostic aid” [2]. Physicians usually refer in their decision making of a diagnosis to already closed cases with verified diagnostic results. In the context of CADx, the typical application of CBIR supports differential diagnosis, i.e. the distinguishing between two or more diseases by systematical comparison. The aim of CBIR is to provide the radiologist with a diagnostic aid by automatically identifying relevant past cases along with their diagnosis and other suitable information. This provides the radiologist with a second opinion, thereby enhancing the diagnostic assessment.

CBIR has a high potential of improving the quality and efficiency of clinical care processes [3] and entails proved high benefits for CADx [4, 5, 6]. There are promising CBIR based CADx systems, the majority concentrating on a particular application area, e.g., lung cancer [7], mammography [8], computed tomographic images of the chest [9]. However, CBIR is not established yet in clinical routine. CAD schemes using CBIR approaches face a number of challenges and this technology needs more research work [10].

One essential factor is the seamless and robust integration into Picture Archiving and Communication Systems (PACS) and the radiologist’s workflow. This includes a comprehensive and integrative representation of the CBIR results along with all corresponding evidences adapted to the physicians needs. The appropriate and contextualized preparation of necessary data required for differential diagnosis delivers an indispensable brick in the bridge from CBIR to real CADx.

A substantial requirement is the use of open standards and the integration with a hospital-wide communication structure [4]. Digital Imaging and Communications in Medicine (DICOM) Structured Reporting (SR) [11, 12] is an approved standard format for exchanging CADx results in clinical environments, e.g., for mammography [13, 14], which can be stored to and retrieved from PACS. Because of the many advantages [15, 12], the Integrating the Healthcare Enterprise (IHE) employs DICOM SR in several integration profiles [16]. There are several proposals for the general integration of CAD or CBIR systems into clinical environments, also based on DICOM SR documents, e.g., [17, 18]. That motivated us to apply DICOM SR encoded CBIR results in our concept presented in this paper.

There are viewers capable of interpreting SR documents and producing a generic layout, but they are not adjusted to the special needs of CBIR for CADx. Diagnostic findings from different clinical departments, e.g. pathology, laboratory or endoscopy, are usually archived centrally by the Hospital Information System (HIS), utilizing the Health Level Seven (HL7) [19] protocol.

Clinical viewers for diagnostic findings are usually not capable of integrating CBIR results or even CADx results in general. We close this gap and present in this paper the novel IRMACON viewer, a combined viewer for CBIR SR documents together with referenced diagnostic findings from HIS.

2 Materials and methods

We first analyzed the general CBIR-based CAD scheme, in which we embedded our IRMACON viewer. Our proposed IRMACON viewer system makes use of DICOM SR and HL7. Relevant information on both is explained in the following. Thereafter, we analyzed requirements concerning a successful data capturing.

2.1 CBIR-based CADx scheme

CADx systems with CBIR approaches incorporate the following workflow in principle: The physician examines medical images to make a diagnosis. Either, the CADx request is done automatically, e.g., by configuring an internal mechanism inside of the PACS, or the physician manually demands the CADx request. Data from the current patient's context is then transferred to the CBIR system. The CBIR system processes the query input and generates a SR document from the CBIR results which is stored in PACS. The DICOM archive holds all SR documents and delivers requested CBIR results to the physician at his diagnostic workstation, who is awaiting the CBIR support for his diagnosis. The CBIR SR document lists identified images with a certain minimum similarity to the current examination image. Normally, the physician then requests diagnostic details of selected similar images from HIS manually to collect all relevant information for his decision making.

2.2 DICOM SR encoded CBIR results

DICOM provides by means of the SR format a standard encoding of structured data. Templates define valid contents and value types to restrain different encoding of the same content in order to simplify the processing of SR documents. They facilitate an automated processing and interpretation of SR documents [20]. DICOM standard templates do not cover the CBIR application for CADx. We proposed a template adapted to the special requirements of CBIR [21]. It comprises the query image, identified similar images along

with their similarity score, and a description of the applied CBIR system. Our template is applicable by any CBIR system and the resulting SR document may be stored to PACS.

2.3 Query of HL7 diagnostic findings

HIS generally archives diagnostic findings from radiology, laboratories and other affiliated departments in a clinical repository. Reports may be plain text or encoded as, e.g., Portable Document Format (PDF), Microsoft Word. Communication regarding documents exchange is accomplished by HL7 messages. Message type QRY^R02, “Query For Results Of Observation”, requests diagnostic findings, which are returned by message type ORF^R04. A HL7 query message is composed by a message header (“MSH”), followed by a query definition (“QRD”) and a query filter (“QRF”).

In clinical environments, HIS are usually configured to manage communication by a dedicated communication server. This implies that all HL7 messages are sent to this central point, specified by an IP address and a port. The header of the HL7 message contains the final destination, given in the field named “Receiving Application”. The communication server forwards the message to its addressee.

The following small fragment of an exemplary ORF^R04 message is taken from a response to a QRY^R02 message and contains requested diagnostic findings. Parts denoted by “...” have been omitted to simplify the example. HL7 messages are divided into mandatory and optional segments and fields, depending on the particular message type. Each message begins with a header segment, introduced by the keyword “MSH”. The header of this example names “EKG” as “Sending Application” and our viewer “IRMACON” as “Receiving Application” of the message. Date of this message is 2010-02-18 and message type is “ORF^R04”. The segment starting with “OBX” contains the actual diagnostic information and is of value type “TX”, which stands for text, followed by the “Observation Identifier” and the textual diagnostic information.

```
MSH|^~\&|EKG||IRMACON||201002180944||ORF^R04|...
...
OBX|1|TX|79000&ADT^EKG COMM|<diag info>|...
```

The actual diagnostic information has to be extracted from the message and to be displayed to the radiologist.

2.4 Mapping of DICOM and HL7 identifiers

The IRMACON viewer queries diagnostic findings by HL7 protocol that relate to patients and examinations whose medical images have been retrieved by the CBIR system. This requires a mapping of identifiers between the DICOM and the HL7 domain. Relevant keys are patient and order which are used in the HL7 query message to query particular diagnostic findings. There is no official standard that defines a mapping. But based on generally applied principles, e.g., by [16, 22], we decided to use the following mapping:

- PACS Order Number
 - DICOM: (0008,0050) Accession Number
 - HL7: OBR-3 Filler Order Number
- Patient Identifier
 - DICOM: (0010,0020) Patient ID
 - HL7: PID-3 Patient ID

2.5 Analysis of data capture process

We investigated the appropriate design of a viewer of similar medical cases and their corresponding diagnostic findings used in CADx. The following requirements have to be fulfilled in order for the radiologist to benefit from the viewer [23]:

- (R1) Presentation and arrangement of information must promote an easy and quick comprehension. This includes an appropriate structuring according to semantics and interrelations of data. Furthermore, the limitation on only relevant data and the provision of details only on demand helps to accelerate data capturing.
- (R2) Handling of the viewer must be intuitive and simple. This allows a quick initial training and instant functionality recognition by the radiologist.
- (R3) Presentation of data must be adapted to the particular application. For example, the identified similar images retrieved by the CBIR system are captured easily when presented pictorially. The corresponding DICOM identifiers are of little use in the context of CADx.
- (R4) Data retrieval must not slow down the radiologist's work. Otherwise, the physician will hardly be willing to use a CADx system.

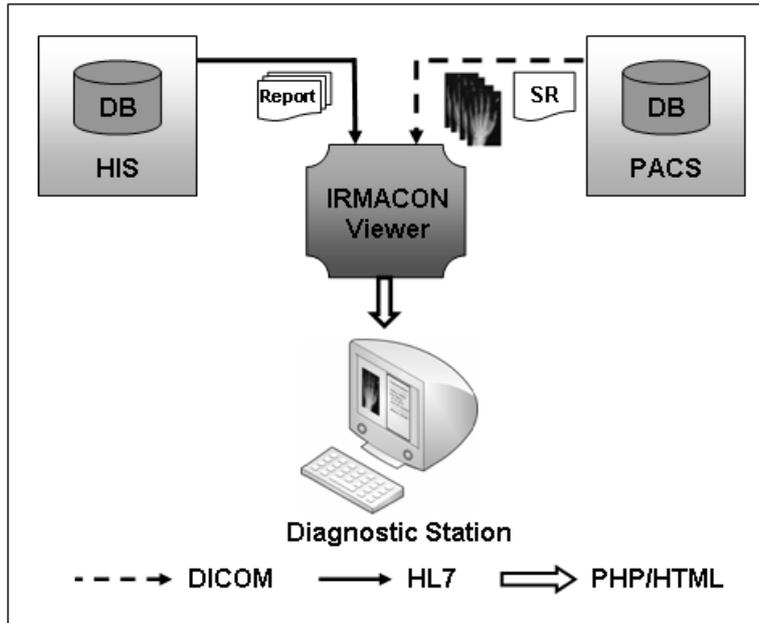


Fig. 1 Overview of the IRMACON system

3 Results

3.1 System design

Our proposed architecture is illustrated in Figure 1:

1. The radiologist at his diagnostic workstation starts the webpage of the IRMACON viewer, which is integrated into the PACS client as a plug-in by utilizing PACS' application programming interface (API). The current patient's context is passed to IRMACON.
2. The IRMACON viewer retrieves the SR document and all referenced images from PACS by DICOM C-MOVE messages. The images along with basic information are presented to the physician.
3. The physician may choose reports of diagnostic findings that are requested from the HIS by the IRMACON viewer.
4. The IRMACON viewer starts an appropriate application associated with the retrieved report to facilitate the display of the report's content.

Our system design is based on using the already established infrastructure by PACS and HIS. Data is transferred by DICOM and HL7 protocol. DICOM SR documents containing CBIR results are managed by PACS. Assuming

that PACS and HIS are balanced systems configured for high performance and reliability, our design meets the requirement (R4).

3.2 Implementation of IRMACON viewer

The IRMACON viewer is a web-based application implemented in Hypertext Preprocessor (PHP) that dynamically generates output in Hypertext Markup Language (HTML) format. OFFIS DCMTK [24] is used for realizing DICOM commands from within IRMACON. HL7 messages are implemented using HAPI (HL7 Application Programming Interface) [25]. SR documents are converted to Extensible Markup Language (XML) by DCMTK and then parsed by the XML extension of PHP. The retrieved DICOM images are converted to Portable Network Graphics (PNG) in order to have an image format suitable for the representation in web browser.

We created a testbed for our viewer limiting PACS and HIS to their relevant functionalities needed in our scenario. It includes a DICOM database accomplished by DCMTK configured with an DICOM Application Entity Title (AET) for accessing DICOM images and SR documents. Further, the HIS is simulated by a simple HL7 application. Our CBIR system of choice is the Image Retrieval in Medical Applications (IRMA) framework [26]. It has a Simple Object Access Protocol (SOAP) interface for transferring input data and results. CBIR output is created according to our SR template.

3.3 Layout of IRMACON viewer

The layout is illustrated in Figure 2:

1. At the top border, the query or examination image for diagnosis is shown. The identified similar images by the CBIR system are listed below in descending order of their similarity for easy comparison. This structures the information according to their semantics and interrelation, as postulated in (R1). Presenting on the main page images instead of textual data simplifies the capturing of the information (R3).
2. Selecting an image opens a new window giving details on this image, namely examined body part, date of birth and sex. Furthermore, a button for retrieval of reports from the HIS is supplied. Providing further details only on demand meets (R1).
3. The navigation through the viewer is enabled by arrows pointing to the left for switching to images more similar or to the right to images less similar. Details are shown on pressing the field named "Select" on the bottom of the thumbnail. This enables an intuitive and easy handling as postulated by (R2).

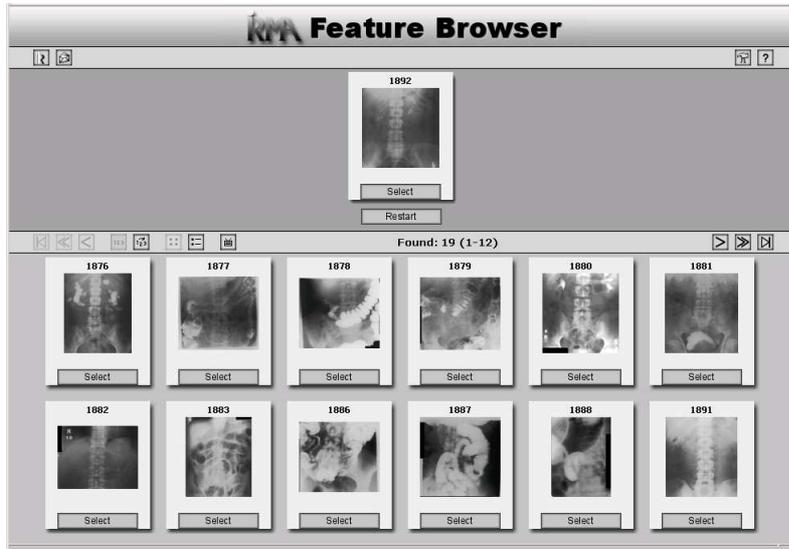


Fig. 2 Exemplary output of CBIR results in the IRMACON browser

4 Conclusions

The introduced IRMACON viewer provides a comprehensive and consistent representation of CBIR results integrated into the clinical setting of a radiologist's every day work. It assembles evidences sourced from CBIR, PACS and HIS necessary for diagnostic analyses. Hence, it follows the principal goal in health systems, "the right information, in the right place, at the right time" [27]. IRMACON establishes an essential link between CBIR systems and applied CADx. We hope that our proposal will contribute to the widespread use of CBIR for CADx in daily radiological routine. Our ongoing work comprises (i) the extension of our testbed and (ii) the processing of specified regions of interest for restricting the CBIR to a certain phenomenon.

Acknowledgements This research was partly funded by the German Research Foundation (DFG), Le 1108/9.

References

1. Smeulders AWM, Worring M, Santini S et al. (2000) Content-based image retrieval at the end of the early years. *IEEE Trans Pattern Anal Mach Intell* 22(12):1349–80.
2. Doi K (2004) Overview on research and development of computer-aided diagnostic schemes. *Semin Ultrasound CT MR* 25:404–10.

3. Müller H, Rosset A, Garcia A, et al. (2005) Informatics in radiology (Inforad): Benefits of content-based visual data access in radiology. *Radiographics* 25(3):849–58.
4. Müller H, Michoux N, Bandon D et al. (2004) A review of content-based image retrieval systems in medical applications – Clinical benefits and future directions. *Int J Med Inform* 73:1–23.
5. Deselaers T, Deserno TM, Müller H (2008) Automatic medical image annotation in ImageCLEF 2007: overview, results, discussion. *Pattern Recognit Lett* 29(15):1988–95.
6. Nakayama R, Abe H, Shiraishi J et al. (2009) Potential usefulness of similar images in the differential diagnosis of clustered microcalcifications on mammograms. *Radiology* 253:625–31.
7. Li Q, Li F, Shiraishi J et al. (2003) Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules. *Med Phys* 30(10):2584–93.
8. Muramatsu C, Li Q, Suzuki K et al. (2005) Investigation of psychophysical measure for evaluation of similar images for mammographic masses: Preliminary results. *Med Phys* 32(7):2295–304.
9. Aisen AM, Broderick LS, Winer-Muram H et al. (2003) Automated Storage and Retrieval of Thin-Section CT Images to Assist Diagnosis: System Description and Preliminary Assessment. *Radiology* 228:265–70.
10. Zheng B (2009) Computer-Aided Diagnosis in Mammography Using Content-Based Image Retrieval Approaches: Current Status and Future Perspectives. *Algorithms* 2:828–49.
11. National Electrical Manufacturers Association (2009) DICOM Part 16: Content Mapping Resource.
<ftp://medical.nema.org/medical/dicom/2009/09.16pu.pdf>. Cited 2010 Feb 23
12. Clunie DA (2000) DICOM Structured Reporting. PixelMed, Bangor PA.
13. Philips Medical Systems (2007) Integration of digital mammography on PACS. http://www.healthcare.philips.com/phpwc/main/shared/assets/documents/about/events/2009/ecr/pdf/hi/integration_of_digital_mammography_on_pacs.pdf. Cited 2010 Feb 23
14. Eastman Kodak Company (2006) Make mammography work flow. http://www.carestreamhealth.com/mammoPACS_brochure_N1-102.pdf. Cited 2010 Feb 23
15. Noumeir R (2006) Benefits of the DICOM Structured Report. *J Digit Imaging* 19:295–306.
16. Integrating the Healthcare Enterprise (2008) IHE Technical Framework, vol. I: Integration Profiles.
http://www.ihe.net/Technical.Framework/upload/ihe_tf_rev9-0ft_vol1_2008-06-27.pdf. Cited 2010 Feb 23
17. Le A, Mai L, Liu B et al. (2008) The workflow and procedures for automatic integration of a computer-aided diagnosis workstation with a clinical PACS with real world examples. *Procs SPIE* 6919.
18. Welter P, Deserno TM, Fischer B et al. (2009) Integration of CBIR in radiological routine in accordance with IHE. *Procs SPIE* 7264.
19. Health Level Seven, Inc. (HL7) (2010) Health Level 7 Standards. <http://www.hl7.org/implement/standards/index.cfm>. Cited 2010 Feb 23
20. Hussein R, Engelmann U, Schroeter A et al. (2004) DICOM Structured Reporting: part I. overview and characteristics. *RadioGraphics* 24(3):891–6.
21. Welter P, Deserno TM, Gülpers R et al. (2010) Exemplary Design of a DICOM Structured Report Template for CBIR Integration into Radiological Routine. *Procs SPIE Medical Imaging* 7628.
22. Kinsey TV, Horton MC, Lewis TE (2000) Interfacing the PACS and the HIS: Results of a 5-year Implementation. *RadioGraphics* 20:883–91.

23. Dayhoff R, Kirin G, Pollock S et al. (1994) Medical data capture and display: the importance of clinicians' workstation design. *Procs Annu Symp Comput Appl Med Care* 541–45.
24. OFFIS (2010) DICOM Toolkit DCMTK.
<http://dicom.offis.de>. Cited 2010 Feb 23
25. University Health Network (2010) HAPI (HL7 application programming interface).
<http://hl7api.sourceforge.net>. Cited 2010 Feb 23
26. Lehmann TM, Güld MO, Thies C et al. (2004) Content-based image retrieval in medical applications. *Methods Inf Med* 43(4):354–61.
27. Winter A, Lagemann A, Budig B et al. (1996) Integration of health professional workstation in hospital information systems. *Procs MIE* 685–89.