# A New Database for Medical Images and Information

Dave Tahmoush and Hanan Samet

University of Maryland, College Park

**Abstract**

We present a medical image and medical record database for the storage, research, transmission, and evaluation of medical images, as well as tele-medicine applications. Any medical image from a source that supports the DICOM standard can be stored and accessed, as well as associated analysis and annotations. Information and image retrieval can be done based on patient info, date, doctor's annotations, features in the images, or a spatial combination of features. Secure access and transmission is addressed for tele-medicine applications. This database application follows all HIPAA regulations.

## Introduction

Most retrievals in medical image database systems are based on the patient identification information or image modality [1]as it is defined in the DICOM standard [2], and it is hoped that inclusion of other features can improve the effectiveness of this type of system. Archimedes includes retrieval based on features, as well as on patient identification information and image modality.

The number of digital medical images is rapidly rising, prompting the need for improved storage and retrieval systems. Image archives and imaging systems are an important economic and clinical factor in the hospital environment [3]. The management and the indexing of these large image and information repositories is becoming increasingly complex. Archimedes is an effort at bringing the latest in information technology to the medical community.

Picture Archiving and Communication Systems (PACS) are the main software components used to store and access the large amount of visual data in medical departments. Often, several layer architectures exist for quick short-term access and slow long-term storage [4], and a web-based PACS architecture has been proposed [5]. Web interfaces have been described for medical image databases [6], and Archimedes also includes a web interface. A web interface simplifies the deployment of the system and enables tele-medicine, but requires tighter security measures. The Archimedes system was designed as a web-based system from the start, and provides a platform to evaluate the usefulness and effectiveness of incorporating those changes into PACS.

Several frameworks for distributed image management solutions have been developed such as I2Cnet [7, 8]. Image retrieval based on visual features is often proposed but unfortunately little is said about the visual features used or the performance obtained. One competing framework with at least a partial implementation is the IRMA (Image Retrieval in Medical Applications) framework [9, 10]. IRMA enables the classification of images into anatomical regions, modality, and orientation. Some frameworks support telemedicine as well as data and image management [11], which Archimedes supports as well.

Archimedes is an image analysis and patient records management tool intended for the use of the medical community. It allows doctors to search for common features in a database of images via innovative combinations of search techniques and algorithms. This system allows the rapid retrieval of images and patient records, and can also find patients with similar images, conditions, or annotations to compare treatment successes.

The support of the National Science Foundation under Grants EIA-00-91474 and CCF-0515241, Microsoft Research, and the University of Maryland Graduate Research Board is gratefully acknowledged.

Proceedings of SPIE Vol. 6516 Medical Imaging 2007: PACS and Imaging Informatics, Steven C. Horii; Katherine P. Andriole, Editors, 65160G - San Diego, USA, Feb 2007.

Searchable images types include common medical images such as x-rays, mammograms, CAT scans and MRIs. Doctors can quickly and easily retrieve patient records with the use of Archimedes' electronic image searching capabilities. They can also use it to perform electronic comparisons of images more simply than traditional hard-copy comparison. In addition to software improvements, the software archives the addition of markups and notations to images, and supports auditing and reviewing of doctors' decisions. This satisfies strict legal constraints of patient information for hospitals and other medical institutions.

The rest of this paper is organized as follows. Section 2 discusses the design of Archimedes for data and image collection and analysis, as well as telemedicine. Section 3 describes the capabilities of Archimedes including the upload, storage, and sharing of images and data. Section 4 draws conclusions and discussed future work.

## 2. Archimedes Design

The design of Archimedes had to take into account the sensitive nature of the data as well as the multitude of regulations coming to govern this field. The design focused on satisfying HIPAA regulations in the US while maintaining the ability to adapt to other regulations.

There are four main sections to the design of a distributed database system. The first layer is the client including the Graphical User Interface (GUI). The second and third layers are the server and network protocol. The final layer is the underlying database selection and design. The issues driving the design are the security of the system and the capabilities needed to operate effectively.

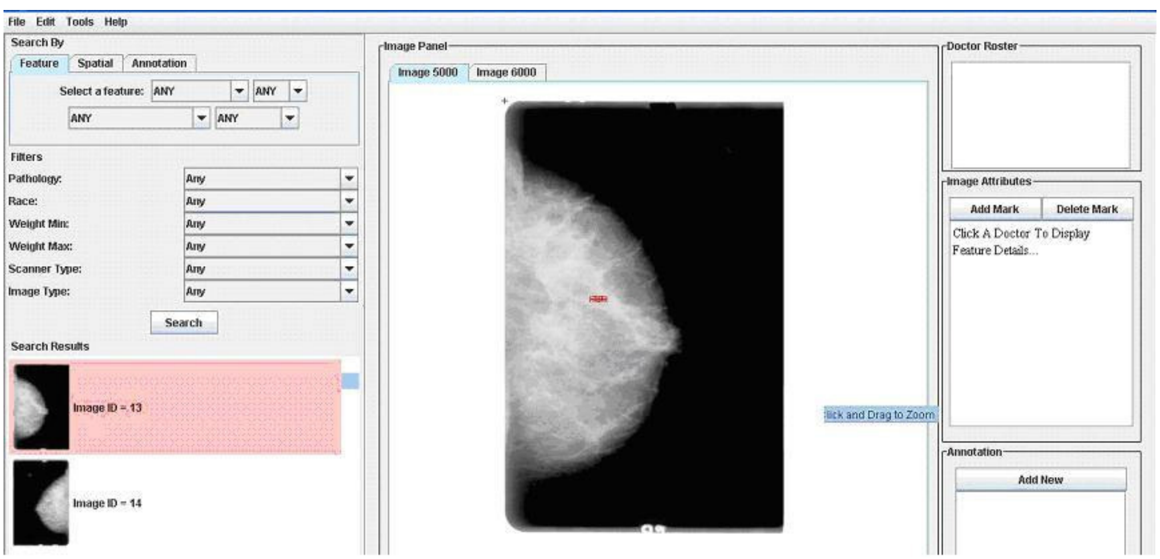


Figure 1. The Archimedes user interface. The search panel and results set are on the left, the zooming interface is in the middle, and adding features and annotations are on the right. This is the semi-private view where no patient information is viewable or searchable.

The client section of the design was difficult because of the need to integrate an image tool with a data and search tool. The GUI is shown in Figure 1. Instead of devoting most of the space to images as is a standard practice in image applications, Archimedes devoted a smaller space but augmented that space with the ability to zoom in on interesting parts of the image as well as grab the image and move it around within the available space. The images can also be downloaded into other viewing systems, and image sets are viewed through a tabbing system. Moving the images around appeared to be intuitive, and is similar to the technique used in GoogleMaps. However, the zooming was designed to be smooth and not stepped, with a mouse interface that was difficult to master. A planned upgrade will be to simplify the mouse interface for zooming in and out on images. The extra space was used to display auxiliary data like doctor's notes and the data searching interface and search results.

The client design for searching and managing patient records and images focuses on searching by patient information, image features, and text. For search by patient information in Archimedes, the medical professional is allowed to enter patient information (i.e. first name, last name, date of birth, etc.)into Archimedes. Once information is entered, they can use filters to further refine their search results. Searching for images with specific features or spatial combinations of features allows doctors to further specify the results. Archimedes also allows search over text comments doctors previously made about patients or images. This works like a primitive yahoo search over the text of the medical annotations. This free form of text search, combined with patient information, was surprisingly popular and simple to use, possibly because it mimicked a well-known searching application.

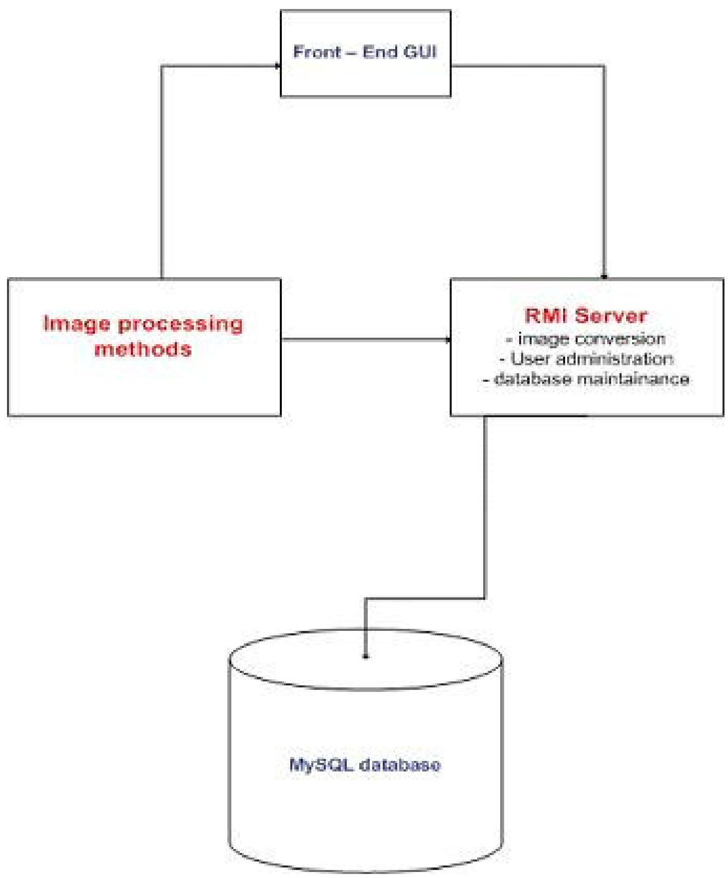


Figure 2. Archimedes High-Level Design. A web interface GUI is connected to a RMI server and a database. Note that the Image Processing methods are separate form the server.

The program must maintain a high level of security due to privacy issues associated with maintaining sensitive patient medical information. The application is web-based for simplified deployment and tele-medicine uses, but this makes security more of an issue. Information transmitted from the server to the front-end is encrypted via the AES encryption scheme. All modifications during program use are monitored and logged by the system, and the viewing of the logs is limited to administrators. Images are transferred from the server to the client, where the client allows manipulations and exploration of the images. The transmission of images over an encrypted connection caused significant problems for the usability of the system. One key to improving the performance on images is to strip out all HIPAA-regulated information from the images and transmit it separately over an encrypted connection. Note that DICOM is not used to communicate between the client and the server, but is used when images are entered or downloaded into other applications. Because the choice of server environment was flexible, Java was chosen as the development language for the server.

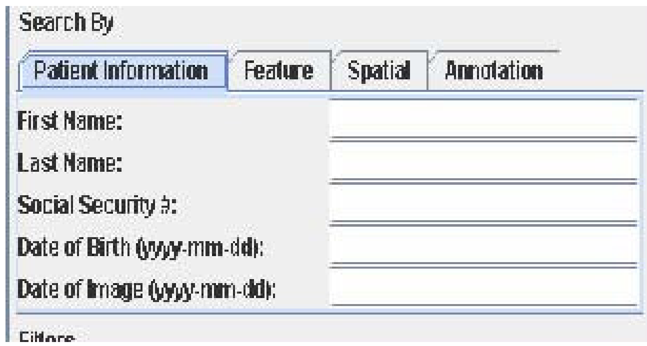


Figure 3. Archimedes Patient Information Search Panel. Searches over patient information can be done using the patient's name, date of birth, social security number, or the date an image was taken. Additional search tabs are available as well.

There are many possible databases that could have been selected for use in Archimedes, including products from Oracle and Microsoft. However, the spatial search requirements of Archimedes dictated the database choice. Our prototype used MySQL for simplicity as is shown in Figure 2, but the final design uses the PostgreSQL open source SQL compliant relational database. PostgreSQL runs on all major operating systems including Linux, UNIX (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, SunOS, Tru64), BeOS, and Windows, which makes it highly portable and therefore extensible in the scope of our project. PostgreSQL allows all of the features of an advanced database, including transactions, tablespaces, and foreign keys. Using PostGIS, an extension for the PostgreSQL database, adds support for geographic objects and spatially enables the database. PostGIS complies with the Simple Features Specification for SQL and is an Open Source project as well. Both programs have been extensively tested and are considered secure, stable products. Using our Database API we will also be able to support Oracle databases with the Spatial Extension. This will make the project more extensible for the future.

The design paid careful attention to the access to images because of privacy issues. Access to images can be either tightly controlled and private, public, or semi-private, while access to patient information is always tightly controlled and private. Administration is simplified through the use of groups, where semi-private images have groups of trusted medical professionals associated with them to help provide analysis. This is helpful for administration, where the hospital doctors or a subset are defined as the image default group setting.

The incorporation of DICOM 3 capability is necessary in any medical image system. However, XML support was also included mainly because it was simple. It became the preferred method for uploading data that was not already in DICOM format. The XML schema included tags such as <Patient>, <Image>, <Doctor>,<Feature>,

<XPos>, <YPos>, and <Annotation>.

Image processing packages can be incorporated into Archimedes through the ability to upload not only text and images, but also features in images, their positions, and associated annotations. The separation of image processing packages and the database application was chosen to maintain the flexibility of the system and the ability to incorporate multiple different packages and is shown in Figure 2. The feature characteristics need to be flexible, and can be defined at upload.

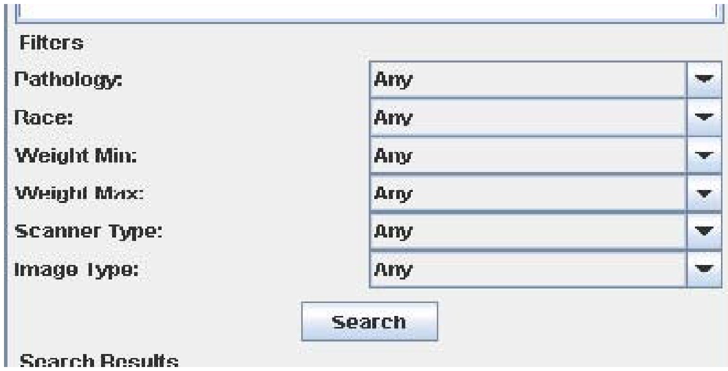


Figure 4. Archimedes Filter Panel. Results from searches can be filtered based on pathology, image and scanner type, weight range, and race. This was useful in analyzing performance data across multiple parameters.

Archimedes is a three-tiered application including backend server, server logic unit, and web front end user interface. The server can run on any machine using a Unix, Linux, or Windows operating system that can support Java.

## 3. Archimedes Capabilities

Archimedes was designed to have extensive search capabilities. Medical professionals can search the database based on information about the patient, as can be seen in Figure 3, and they can use filters to further refine their search results as shown in Figure 4. Searching by features, spatial combinations of features, and text allows the doctors to specify feature parameters they wish to see in the results.

Medical professionals can view and manipulate images on the Archimedes system and tab through sets of images. There is a zooming interface in order to focus in on interesting parts of the images. Point features can be inserted and described as overlays to the images. Text annotations can be entered. Multiple overlays can be captured for each image, allowing double reading of images. The overlays can be blinded for running studies on radiologist techniques. The radiologists' diagnoses are captured and the data can be accessed remotely, thereby allowing tele-medicine applications to be run on the Archimedes platform.

Images and image sets can be viewed directly on the system or downloaded to an alternate image viewing system. Image annotations can be entered directly into the system, and features can be marked as well, or uploaded from another program in an XML schema. Archimedes supports both the DICOM format and an XML format for the upload of data. Images can be resized, and image zooming is supported. Further details on the capabilities of Archimedes can be found in [12, 13].

## 4. Conclusions and Future Work

We have created a secure database for the storage, retrieval, manipulation, and annotation of medical images and medical records. We found that complying with HIPAA regulations was possible. A test site has been deployed and we are awaiting deployment at a larger hospital. Although Archimedes was not originally designed to be a complete medical information application, it did provide a wealth of information on important issues in the development of this type of application

Several difficulties arose during the creation of this project, including encryption of images and GUI issues. The transmission of images over an encrypted connection was not an optimal choice as it introduced a noticeable delay over unencrypted transmission. Perhaps a better choice of encryption algorithm could have reduced the effect, but the alternative of unencrypted anonymous images and encrypted data would perform significantly better. The GUI did show several areas that needed improvement. The zooming image interface needs to be simplified. Instead of a single monolithic user interface that includes everything on one page with tabbing inside sections, a redesign with tabbing between pages was found to be a more optimal use of space.

The GUI redesign is one of the issues that will be addressed in the future of this project. Additionally, the capability of having area and volume features will be included, with the required capabilities in performing spatial comparisons on these features.

The use of groups to which individual users belong greatly simplified the administration of the system. The incorporation of XML as well as DICOM capabilities improved the flexibility of the system and allows backwards compatibility between systems that are DICOM based and newer systems that might incorporate XML. XML became the preferred format for uploading data that was not already in DICOM format because it is so flexible and easy to code. The inclusion of XML was simple because of the use of Java which contains much of the required capability.

The incorporation of filtering of search results was useful in analyzing trends in the medical data. For example, a particular radiologist's diagnoses can be checked by searching for their diagnosis feature, then filtering with malignant to see how often the diagnosis was correct. Missed diagnoses can be checked by comparing all of a radiologist's patients without a diagnosis feature that were eventually malignant.

**Thanks**

This project was developed in cooperation with the Software Engineering At Maryland (SEAM) cooperative at the University of Maryland. This cooperative matches project with outstanding senior level computer science and computer engineering students at the University of Maryland. Some of the students who helped with this project include Matt Fowle, Dan Ilkovich, Nima Negabhan, James Wren, Guilherme Bandeira, Duane Gilbert, Matt Weinstein, Hassan Shaukat, Sureshmi Wijewardena, Bernard Ng, Paul Carlson, Ratandeep Singh Achreja, Zvi Alexander Band, Jay Ming-Chie Liu, Pratik Mathur, Htin Kyaw Nyo, Kristofer Patrick Quinn, Obaid Siddiqui, Michael Andrew Tantino. The founder of SEAM, Professor Jim Purtilo, deserves special thanks.

**Bibliography**

1 T. M. Lehmann, M. O. Guld, C. Thies, B. Fischer, M. Keysers, D. Kohnen, H. Schubert, B.B. Wein, Content-based image retrieval inmedical applications for picture archiving and communication systems, in: Medical Imaging, Vol. 5033 of SPIE Proceedings, San Diego, California, USA, 2003, 109-117.

2 B. Revet, DICOM Cook Book for Implementations in Modalities, Philips Medical Systems, Eindhoven, Netherlands, 1997.

3 M.W. Vannier, E.V. Staab, L.C. Clarke, Medical image archives -present and future, in: Proceedings of the International Conference on Computer-Assisted Radiology and Surgery, Paris, France, 2000, 565-570.

4 H.U. Lemke, PACS developments in Europe, Computerized Medical Imaging and

Graphics 27 (2002) 111-120.
5 J. Zhang, J. Sun, J. N. Stahl, PACS and web-based image distribution and display, Computerized Medical Imaging and Graphics 27 (2) (2003)197-206.
6 T. Frankewitsch, U. Prokosch, Navigation in medical internet image databases, Medical Informatics 26 (1) (2001) 1-15.
7 S.C. Orphanoudakis, C.E. Chronaki, S. Kostomanolakis, I2Cnet: A system for the indexing, storage, and retrieval of medical images by content, Medical Informatics 19 (2) (1994) 109-122.
8 S.C. Orphanoudakis, C.E. Chronaki, D. Vamvaka, I2Cnet: Content-basedsimilarity search in geographically distributed repositories of medicalimages, Computerized Medical Imaging and Graphics 20 (4) (1996) 193-207.
9. M. O. Guld, B. B.Wein, D. Keysers, C. Thies, M. Kohnen, H. Schubert, T. M. Lehmann, A distributed architecture for content-based image retrieval in medical applications, in: Proceedings of the International Conference on Enterprise Information Systems (ICEIS2001), Setubal, Portugal, 2001, pp. 299-314.
10 Keysers, J. Dahmen, H. Ney, B. B. Wein, T. M. Lehmann, A statistical framework for model-based image retrieval in medical applications, Journal of Electronic Imaging 12 (1) (2003) 59-68.
11 M. Tsiknakis, D. Katehakis, C. Orphanoudakis, Stelios, Intelligent image management in a distributed PACS and tele-medicine environment, IEEE Communications Magazine 34 (7) (1996) 36-45.
12 D. Tahmoush, H. Samet, A web collaboration system for content-based image retrieval of medical images, in Proceedings of SPIE –Medical Imaging 2007: Image Processing, San Diego, CA, February 2007.
13 D. Tahmoush, H. Samet, Archimedes, an archive of medical images, In American Medical Informatics Association 2006 Annual Symposium Proceedings Biomedical and Health Informatics: From Foundations to Applications to Policy, Washington, DC, D. W. Bates, J. H. Holmes, and G. Kuperman, editors, November 2006.