GRID COMPUTING FOR DIGITAL MAMMOGRAPHY

Mike Brady¹, David Gavaghan², Ralph Highnam³, Alan Knox⁴
Sharon Lloyd², Andrew Simpson², and David Watson⁴

¹Department of Engineering Science, University of Oxford
²Oxford University Computing Laboratory
³Mirada Solutions Ltd
⁴IBM UK

Abstract  A United Kingdom digital mammography national database would have major benefits for the United Kingdom. For example, such an archive could: provide a huge teaching and training resource; be able to aid in the evaluation of innovative software to compute the quality of each mammogram as it is sent to the archive; act as a significant resource for epidemiological studies; and be a tremendous step towards the support of centralised, automated computer-aided detection. This paper provides an overview of the eDiaMoND project, the principal aim of which is to develop a pilot system for such an archive.

Introduction

In this paper we provide an overview of the eDiaMoND project, the principal aim of which is to develop a pilot for a digital mammography national database to support the United Kingdom's breast imaging community.

We start by presenting the motivation for the project. We then consider the current situation with regards to the relationship between breast screening and information technology. Having considered this relationship, we then address the objectives of the eDiaMoND project and the constraints and considerations that the project team have to be aware of in pursuit of these objectives. Having addressed the objectives of the project, we discuss the anticipated contributions of the different members of the eDiaMoND consortium. We then consider the technical aspects of the eDiaMoND system as they currently stand: we present a functional view of the system, together with an overview of the architecture that is being employed. Next, we consider the proposed development of that architecture before providing an overview of some future directions for eDiaMoND.

Motivation

Breast cancer is a major public health issue in the western world, where it is the most common cancer among women. In the European Union, for example, breast cancer represents 19% of cancer deaths, and fully 24% of all cancer cases. Breast cancer is diagnosed in a total of 348,000 cases every year in the USA and EU, and kills almost 115,000 women annually. It is estimated that approximately one in eight women will develop breast cancer during the course of their lives; it is also estimated that one in 28 women will die of the disease. The earlier a tumour is detected, the better the prognosis for the woman concerned. A tumour that is detected when its size is just 0.5cm has a favourable prognosis in about 99% of all cases. Few women can detect a tumour by palpation (breast self-examination) when it is smaller than 1cm, by which time (on average) the tumour will have been in the breast for up to 6-8 years. The five-year survival rate for localised breast cancer is 97%; this drops to 77% if the cancer has spread by the time of diagnosis.

The World Health Organisation’s International Agency for Research on Cancer (IARC) has recently concluded that mass screening via mammography reduces mortality. The agency’s findings are based on the work of an IARC working group, comprising 24 experts from 11 countries. This working group evaluated all currently available international evidence on breast screening. The working group discovered a 35% reduction in mortality from breast cancer among women in the 50-69 age group who were screened, when compared to those who were not. This figure equates to one life being saved for every 500 women screened.
Although the precise figures vary from country to country, the above statistics present a very clear rationale for mass screening, which is currently - in the United Kingdom - based entirely on X-ray mammography. The NHS's Breast Screening Programme (BSP) began in 1987. Currently, the BSP invites women between the ages of 50 and 64 for breast screening every three years. This age criterion stems from the fact that the breasts of pre-menopausal women, particularly younger women, are composed primarily of milk-bearing tissue, which is calcium-rich; this milk-bearing tissue involutes to fat - which is transparent to X-rays - during the menopause.

If a mammogram displays any suspicious signs, the woman is invited back to an assessment clinic, in which other views and other imaging modalities are utilised. Currently, 1.3 million women are screened annually in the United Kingdom. There are 92 screening centres in the United Kingdom, employing a total of 230 radiologists. Each radiologist reads, on average, 5000 cases per year, with some reading up to 20,000.

The emergence of the NHS's Breast Screening Programme was as a result of the Government's acceptance of the report of the committee chaired by Sir Patrick Forrest. The report was rather bullish about the potential positive effects of a screening programme: "By the year 2000 the screening programme is expected to prevent about 25% of deaths from breast cancer in the population of women invited for screening ... On average each of the women in whom breast cancer is prevented will live about 20 years more. Thus by the year 2000 the screening programme is expected to result in about 25,000 extra years of life gained annually in the UK".

To date, the BSP has screened more than fourteen million women and has detected over 80,000 cancers. The programme is saving at least 300 lives per year. Furthermore, it is predicted that this number will rise to 1,250 by 2010.

Recent studies have suggested that the rate of interval cancers that appear between successive screening rounds is considerably larger than was predicted in the Forrest Report. Increasingly, there are calls for mammograms to be taken every two years and for both a cranio-caudal (head to toe direction) and mediolateral oblique (armpit to opposite hip) image to be taken of each breast. At present only mediolateral oblique images are taken.

The opportunity for information technology to assist healthcare professionals can be illustrated by consideration of statistics for screening in the USA. Currently, some 26 million women are screened in the USA annually (significantly, this is nearly half of the 55 million women who are screened each year worldwide). In the USA there are 10,000 mammography-accredited screening centres. Of these, 39% are community and/or public hospitals, 26% are private radiology practices, and 13% are private hospitals. Although there are 10,000 mammography centres, there are only 2,500 mammography specific radiologists.

Anecdotal evidence suggests that staff shortages in mammography in the USA seem to stem from the perception that it is “boring but risky”. This risk element is characterised by the fact that 12% of all malpractice lawsuits in the United States are taken against radiologists, with the failure to diagnose breast cancer becoming one of the leading reasons for malpractice litigation. In the USA, this shortage of radiologists is driving the development of specialist centres and technologies (computers) that aspire to replicate, or at least supplement, the skills of radiologists. It can be argued that screening environments are, in some ways, perfectly suited to the utilisation of information technology solutions, as they are repetitive and require objective measurements.

As we have noted, the development of mass screening programmes has already produced encouraging results. In addition, process changes can help detection rates and reduce recall rates. Indeed, it has been shown that recall rates drop by 15% when using two views of each breast. It has been demonstrated empirically that double reading (two radiologists examining each mammogram) greatly improves the results of screening. The number of cancers missed when mammograms are double read is half that of single reading. However, as one would expect, double reading is expensive and - in any case - there are too few screening radiologists. In addition, a study carried out at Yale of board-certified radiologists demonstrated that, when double reading was employed for a particular
study, the radiologists disagreed 25% of the time about whether a biopsy was warranted and 19% of the time in assigning patients to one of five diagnostic categories. Recently, it has been demonstrated that single screening plus the use of computer-aided diagnosis (CAD) tools – image analysis algorithms that aim to detect microcalcifications and small tumours – also greatly improves screening effectiveness, perhaps by as much as 20%.

Post-screening, the patient may be assessed by other modalities such as palpation, ultrasound and, increasingly, by MRI: 5-10% of those screened do have such an extended assessment. Following assessment, around 5% of patients have a biopsy taken. In light of the number of tumours that are missed at screening (which reflects the complexity of diagnosing the disease from a mammogram), it is not surprising that clinicians err on the side of caution and order a large number of biopsies. In the USA, for example, there are over one million biopsies performed each year, with a staggering 80% of these biopsies revealing a benign (non-cancerous) outcome.

**Breast screening and information technology**

Currently, in the United Kingdom, the NHS's Breast Screening Programme primarily involves the use of film, with some private hospitals and increasing numbers of symptomatic clinics exploring the potential for the use of full field digital machines. The use of film causes problems with storage for such large volumes of information; there is also, of course, the requirement to ensure that this information is kept secure, yet, at the same time, easily retrievable.

In common with other organisations, the NHS is moving to the use of digital information for health and patient records with the introduction of pilot projects for Electronic Patient Record Systems. However, due in part to previous problems of technological failures in the NHS, procurers of technology are keen to have a more integrated approach to workflow and systems.

Following the lead of the USA, where IHE (Integrating the Healthcare Enterprise) has been widely adopted for radiology, the IHE-UK is focusing on the interoperability of IT systems for radiology in the UK to enable an improvement in patient care through the better support of information systems.

An integral part of IHE is the deployment of ‘Picture Archiving and Communication Systems’, PACS, to support the storage and transfer of digital imaging information. These systems are typically used to store MRI, Ultrasound and CAT scan information within a hospital or a community of hospitals (termed Community PACS), to enable the easy access of information throughout the hospitals and, where possible, from a remote location to enable ‘tele-radiology’. Such implementations have been deployed by, for example, the Portsmouth and South West London Trusts.

Reporting Information Systems (RIS) and, increasingly in the UK, dedicated mammography reporting systems are also used widely to support the storage of digital information pertaining to a reading, and through the IHE initiative, it is envisaged that these systems will eventually interoperate seamlessly.

At the present time, trials and first installations of full field digital equipment are in progress, but it may be several years before the United Kingdom adopts this technology across the Breast Screening Programme. Secondary capture of images through digitising film would offer a means of capturing the current information for safe, efficient and secure digital storage.

Despite the encouraging moves in the United Kingdom towards an effective supporting infrastructure, the USA has advanced at a faster rate in its move to the digital world. In the USA, many radiology departments now use full field digital machines for mammography and often use computer aided detection tools to help detect cancers (acknowledging that in the USA there is no double reading of mammograms, unlike in the United Kingdom).

Of course, PACS systems have been developed over the years for non-mammography applications. Significantly, however, mammography poses somewhat unique problems due to the size of the images involved. For example, a mammogram in digital form yields 25-40MB of data at a pixel resolution of 50 microns. Furthermore, a typical
screening session for a patient involves approximately 120MB of data. In addition, if comparison to previous records is required, then a further 120MB of data is needed. These volumes of data are significant, especially when coupled with the timing constraints imposed by the working practices of radiologists.

In a typical screening setting, a radiologist spends around 30 seconds on each mammogram, and will read approximately 100-120 cases in an hour, so time is a vital factor in image retrieval.

Clearly, the development of an information technology infrastructure to support the United Kingdom's breast imaging community is not a trivial task, but it is recognised as an important step to enable system support for the workflow for mammography. In reviewing the needs for systems to support the Breast Screening Programme, it is apparent that the Programme operates as independent screening units, with some standardisation with respect to how they operate and with their own ways of working. This applies to the management of information, the roles and responsibilities matrices, and the methods for recording clinical diagnosis information. Any system to support such an entity would need to be sufficiently flexible and considered to offer any appropriate benefits.

Objectives and considerations

The main aim of the eDiaMoND project is to construct a large federated database of annotated mammograms at St George's and Guy’s Hospitals in London, the John Radcliffe Hospital in Oxford, and the Breast Screening Centres in Edinburgh and Glasgow. Applications to aid teaching, detection, diagnosis, and the facilitation of epidemiological studies are also being developed.

There are three main objectives to the initial phase of the project: the development of the Grid computing infrastructure to support federated databases of images (and related metadata and patient information) within a secure environment; the design and construction of a Grid-connected workstation and database of standardised images; and the development, testing and validation of the system on a set of important applications.

There are a number of key aspects that must be addressed in the development of a Grid infrastructure for a system such as eDiaMoND. One such aspect is, of course, security: ensuring secure file transfer and tackling the security issues involved in having patient records stored on-line, allowing access to authorised persons but also, potentially, the patients themselves at some time in the future, are all key issues.

The design and implementation of Grid-enabled federated databases for patient information, images and related metadata is another such aspect, with timely data transfer being the key design issue. Factors here revolve around (loss-less) data compression and very rapid and secure file transfer.

Data mining issues revolve around the need to run complicated queries across a very large federated database. This is also an issue for computer-aided diagnosis. Teaching tools that test radiologists via the production of either ‘random’ or pre-programmed cases have an obvious need to work at the speed of current breast screening, which means the next case will need to be displayed within seconds after the ‘next’ button is hit.

Ontologies are being developed for description of patient and demographic data, together with descriptions of the image parameters and of features within images.

Database design has been tailored to the needs of rapid search and retrieval of images, with the database being built within the IBM DB2 and Content Manager frameworks. The current and future needs of future epidemiological studies, and information related to high risk such as family history and known breast cancer causing mutations are being taken into consideration as the system is developed. By considering a sufficiently broad range of applications at the design phase, our aim is to develop standard interfaces to the system that will have the potential to support future research studies.

Some prototype data mining tools have already been developed within the context of a single database. Again, speed of access is crucial: the database architecture is being determined according to the frequency of requests for particular image types.
Within a project such as eDiaMoND, in which there is a need to consider the requirements for potentially implementing in a real clinical environment and using real patient data, the following are all key considerations.

- Ethical and legal considerations for the use of, and processing of, data for use by the project. This requires both clearance from the relevant ethics committees as well as compliance with The Data Protection Act (1998) and The Human Rights Act.
- Adopting appropriate security mechanisms and technologies: even though the data used within eDiaMoND is anonymised, the controls that are in place assume that it is not anonymised. Such controls include the use of Virtual Private Networks (VPN), encryption, certificates and keys, as well as fine-grained and flexible access control mechanisms and the provision of the appropriate education for all staff using the system.
- Designing a system with appropriate audit trail capabilities, recognising the need for traceable actions where any information is read or manipulated.
- Development of foolproof anonymisation techniques, as well as defined and managed processes for carrying out this function to prepare data for the project.

Furthermore, the project team has also had to consider the following additional issues.

- Analysis of NHS network constraints and understanding how to overcome these for an implementation of the pilot system. In addition, a second deliverable of the project - a blueprint document - will detail how the system might potentially be deployed throughout the NHS.
- Deployment of workflow methods enabling the project to demonstrate appropriate scenarios, but also offering flexibility for future exploitation.
- Understanding of the current and projected IT initiatives in healthcare and in the NHS in particular, e.g., IHE and its standards and also the work of the NHSIA.

The eDiaMoND consortium

It would be difficult for a project that is as ambitious as eDiaMoND to be undertaken by any single organisation. There is a clear requirement for the developed system to satisfy the needs of the end users if it is to be accepted: this project could not be undertaken without a detailed understanding of the procedures and needs of the radiologists, whom, after all, the system is being designed to support. In common with other complex e-Science projects, eDiaMoND is being undertaken by a project team, in which the contributing parties possess diverse and complementary skills. The partner organisations are described below.

IBM is providing the hardware for the project under a Shared University Research (SUR) grant, as well as the underlying software infrastructure on which the system is being built. In addition, a number of IBM employees have been seconded to the project and bring both significant managerial and technical experience.

The University of Oxford, which has academic and research staff based in the Engineering Science Department, the Software Engineering Centre (based in the Computing Laboratory), and the Oxford e-Science Centre associated with and working on the project, is the main academic partner.

Mirada Solutions is a University of Oxford spin-out company, whose main business is in the field of image processing technology. User applications for training and a range of leading edge technologies for processing mammograms are being developed by Mirada.

The project also involves a number of clinical partners. In each case, the clinical partners are breast screening centres that have interests that are closely related to the goals of the eDiaMoND project.

- The first clinical partner is King’s College, London and Guy’s and St. Thomas’ NHS Trust Hospitals, London, where strong expertise exists in medical imaging, including image-guided intervention, tissue modelling, and the measurement of change using medical images.
The John Radcliffe Hospital, Oxford contributes a strong history of research and practice in medical imaging;

St George's Hospital, London (with University College, London) has a large breast screening centre located at the hospital, and is providing significant expert opinion on user requirements for screening and training.

The final clinical partner is the Edinburgh Breast Care centre (together with the University of Edinburgh), where there is an extensive track record in evaluating the use of IT systems to support the work of breast-screening radiologists.

eDiaMoND: a functional description

Having described the motivation for, the goals of, and the organisations involved in, the eDiaMoND project, we are now in a position to discuss the development of the eDiaMoND system to date.

The core eDiaMoND system consists of middleware and a virtualised medical image store to support the eDiaMoND Data Grid concept. The virtualised medical image store comprises physical databases, with each being owned by a Breast Care Unit (BCU) participating in the eDiaMoND Grid. The assumption is that each of these BCUs will own and manage the hardware and software needed for its own data. The eDiaMoND Grid is formed by participating BCUs coming together as a virtual organisation, and uniting their individual databases as a single logical resource.

The key functions of eDiaMoND can be stated thus.

- Image acquisition: this is the process of inputting X-ray mammograms into the system. A radiologist at the Image Capture Workstation takes scanned X-ray films and adds patient information. The result is an image file, which is passed into the eDiaMoND Grid.

- Query: an administrator may query data in the system to set up a reading session as part of the screening process; a radiologist may make ad hoc queries in screening, or in constructing sets of images suitable as training cases; an epidemiologist may construct complex long-running queries that run across the entire archive.

- Image retrieval: this is retrieval from the Grid of specified image files and reports. Image files can be retrieved individually, or as a collection of all the files associated with a particular series or study or patient.

- Diagnosis reports: the system will capture and manage reports made by radiologists during the screening process.

- Image processing: this aspect covers processes that categorise or manipulate the image data in support of data mining and Computer Aided Detection (CADe) services in the Grid.

Each of the above functions must be implemented in a Grid that allows BCUs to collaborate with each other, while maintaining individual policies on how data for which they are responsible is distributed and shared. The system must allow BCUs to form a virtual organisation for breast screening without requiring any central authority or centralised IT resources. Furthermore, the system is required to support a workflow for breast screening. Finally, the system must allow the same or different BCUs to form virtual organisations for other applications for the mammography resource, with radiologist education and training and epidemiological studies being the examples used to demonstrate this in the project.

Architecture overview

Mirada Solutions are providing a GUI application to allow capture of images and patient data. Mirada Solutions are also providing a GUI application for Breast Screening. A browser-based client is being developed that allows a BCU administrator to drive the breast screening process. This will drive a database query component that will also be capable of use from the breast workstation.

Client interaction with the Grid is through a set of services that are managed by a registry. This set of services can - broadly - be divided into data services, which allow each hospital to see all of the data owned by the participating BCUs, and compute services, which have the ability to perform potentially complex and long-running calculations on the image data.
The set of services in a registry effectively defines the function of the Grid. Clients can interact with the registry to discover services. Flexibility is introduced by manipulating the service available in a registry. The registry may maintain several services with the same service interface offering different levels of quality of service.

Mammograms are very large binary objects. It is possible to store these in a relational database, but this is not as efficient as storing the image in a filesystem and holding a reference in the database. Mammograms are being stored as files in eDiaMoND. The intention is that each BCU maintains its own Image Store consisting of a relational database of patient data and image metadata with records linked to images in the filesystem. The eDiaMoND Grid will provide mechanisms whereby the individual Image Stores can be represented as a single large Virtual Image Store that are accessible to all BCUs.

Service-oriented architectures involve an approach to distributed computing that treats software and data resources as services available on a network. This approach is typified by the Open Grid Services Architecture (OGSA) definition of Grid Services. A Grid Service is essentially a stateful Web Service with a defined lifetime that conforms to a set of interfaces and behaviours with which a client may interact.

OGSA defines an architecture whereby service providers create a description of the service they offer and publish it in a registry. Service requestors “discover” service descriptions in the registry and “bind” to a service implementation offered by the provider.

The registry that a particular Grid client node interacts with to discover eDiaMoND services can be specified as a parameter. The contents of the registry effectively define the eDiaMoND Grid with respect to that particular client. A client may be easily re-configured to point to a different registry. All the clients in the Grid may share the same registry, but are not forced to, and Grid Services may be added to and removed from the registry dynamically. These features will be exploited to make the eDiaMoND Grid both flexible and extensible.

DICOM (Digital Imaging and Communications in Medicine) is a standard for communicating and managing digital medical data. DICOM defines communications protocols for manipulating medical objects in a distributed computing environment; it also defines file format for storing those objects.

The eDiaMoND system will have a DICOM interface at the acquisition workstation to capture images from DICOM compliant scanners (and later, Digital X-ray machines). The DICOM file format is used within the eDiaMoND system, but OGSA services rather than DICOM protocols are used internally for communication.

SMF is a technology being developed by Mirada Solutions that models the complete image creation process for X-ray mammography to determine the height of non-fatty tissue in the breast for each pixel in the image. In essence, SMF provides a normalization of X-ray mammograms that facilitates comparison of images and supports the development of data mining and computer-aided detection. The SMF algorithm takes a DICOM image file as input, and generates three outputs: a segmentation map, a breast tissue density map and some simple metrics.

Mirada Solutions are implementing the SMF algorithm on the acquisition and breast workstations. A newly captured DICOM image file will be processed on the acquisition workstation to create an additional DICOM image file containing the SMF breast density map in place of the “raw” image.

Mirada Solutions will further develop SMF through the lifetime of the project. This means that SMF images must be tagged with a version number and the raw DICOM image file must be kept to allow for reprocessing using new versions of the SMF algorithm. The acquisition workstation will send both the raw and SMF versions of the image to the Grid.

Changes in SMF are likely to be concerned with improvements in the way the tissue density map is derived or the way it is used to construct the image presented to a Radiologist. We do not expect changes to the format of the SMF data stored in the Grid.
The breast workstation will also be capable of generating SMF versions of raw DICOM image files. The primary reason for this is so that the breast workstation can be used with data resources other than eDiaMoND, where SMF is not directly available. The secondary effect of this is to allow the breast workstation to “recalculate” an SMF image if the algorithm has moved on to a new version and the particular SMF stored in the Grid is no longer useful. The eDiaMoND Grid will be capable of accepting SMF images from the breast workstation as well as the acquisition workstation. For any given raw image, only the most recent (according to version number) equivalent SMF image will be kept.

If time and resources permit, the SMF algorithm will also be implemented as a Grid Service to allow generation of SMF within the eDiaMoND Grid.

The eDiaMoND project will make use of OGSA-DAI services to represent the non-image and image data resources in the Grid.

Phased development

Development of the eDiaMoND pilot system is being divided into three phases. Two of these phases constitute versions of the system that will be deployed to the clinical evaluation sites. The Phase 1 deliverable implements a Grid that allows the eDiaMoND clinical team to amass data and test the system. The Phase 2 deliverable is the eDiaMoND pilot system.

These deliverable phases are preceded by a Phase 0 that is internal to the eDiaMoND development team. Phase 0 essentially allows for the integration and unit test of key system components on a single machine at OUCL. It serves as a test of the interfaces between the components and allows the methodologies for working across the development organisations to be developed.

The 2003 All Hands conference coincides with the end of Phase 0.

Future directions

The eDiaMoND project will build a prototype infrastructure that is capable of support digital breast imaging within the United Kingdom. By carefully considering user requirements, both for screening and for research, we aim to develop a generic, scalable and flexible Grid infrastructure. Our aim, once we have established proof-of-concept in this initial phase, is to develop a production system to support the NHS Breast Screening Programme and the work of the breast cancer research community in the United Kingdom.

In a wider context, by taking a generic approach to middleware development, we hope that the eDiaMoND system will be seen as a blueprint for similar breast screening systems in other countries, as well as for similar technologies that aim to support other modalities or diseases.

Ultimately, it is hoped that eDiaMoND will prove to be the first step on the road to providing a Grid-based infrastructure for 21st century patient-centred healthcare systems.