This paper discusses the design of a computer based training (CBT) system to support both the improvement of the skills of experienced readers and the training of new staff. This work is timely because recent advances in technology – particularly advances in high-quality computer displays and high-speed networking – hold the promise of delivering significant benefits for mammography. Through the use of ethnographic data we show how training packages and systems might be configured to take account of real world practice and how current practices would be impacted by a training system.

Introduction
The aim of breast screening is to diagnose cancers early to improve prognosis. This diagnosis process relies heavily on the skill of the human readers who examine screening mammograms. This paper discusses the design of a computer based training (CBT) system to support both the improvement of the skills of experienced readers and the training of new staff. As we will discuss, statistical analysis of readers’ performance provides a driver for this work because it demonstrates that reading skill is related to breadth and quality of experience. However, the real world practice of reading is more complex than the statistics reveal. The contribution of this paper is to examine the constraints on a design of a training system that are imposed by the real world nature of screening and the changes required in current practice to support its use.

This paper focuses on breast screening as practiced by the NHS screening programmes in England and Scotland. An X-ray study consisting of one or more views (mammograms) of each breast is conducted for each patient. These new images are married with previous mammograms and with a brief patient history to form a case. Ideally, cases are reviewed by at least two readers, including at least one consultant radiologist. For each case, the readers attempt to identify the several types of mammographic features (lesions) that are possible indicators of breast cancer. On the basis of this screening, 5% of the patients will be offered further assessment and of these patients 12.5% (0.6% overall) will be diagnosed with cancer [1]. Readers are aware of their responsibility both to spot the cancer cases and to avoid recalling healthy patients unnecessarily.

The work presented here has been undertaken as part of the requirements analysis phase for the eDiamond project, a large initiative concerned with the use of grid technology to support training, diagnosis and epidemiological applications for breast screening. To elicit requirements for a grid-enabled CBT tool, we have used video to capture the work of ‘mentoring’ (part of the process of training novice readers within NHS breast screening programmes) and have conducted a number of interviews with radiologists involved with training in three screening centres.

Our data collection and analysis are informed by ethnographic methods – it is
our contention that it is only through an understanding of how trainee readers acquire the practical skills associated with reading mammograms that we can appreciate the work of training.

Underlying this paper is a very simple, and hence possibly generic, model of CBT for mammography. The aim of CBT is to provide readers with additional reading experience on a broad range of cases accompanied by immediate, appropriate and accurate feedback. Due to the difficulties involved in managing physical film, training will be provided using high-resolution digital images and a soft copy workstation\(^1\). A large number of cases are required to provide practice in this way. We envisage the use of a grid infrastructure, e.g., high-speed computer networks and shared computing resources, to allow training cases to be shared between centres and to manage the collaborative work needed to select and annotate cases for use in training.

Statistical Drivers for CBT
This section explores previous statistical studies in the light of our ethnographic and interview data to uncover some of the drivers of grid enabled CBT.

Several studies have shown a correlation between the volume of cases read and the sensitivity and specificity of readers [2] [3]. One study concludes that a reader’s performance is related to the logarithm of the number of mammograms previously read [4]. This finding relates to readers with experience of between 10 and 12,000 cases, and is consistent with a power law relationship between practice and learning that is found in other domains [5]. At least one study suggests that readers are less good at detecting unusual cancers [6]. In the study, ill-defined masses represented only 3.0% of screening detected cancers (less than 1:5000 cases [1]), but 18% of false negative interval cancers, even though these lesions have a positive predictive value of 50%.

A number of studies suggest that the quality of feedback is important in teaching [7] [8]. We have observed two methods of providing trainee readers with feedback during screening practice. In side-by-side mentoring, the trainee examines mammograms under the direct supervision of a qualified reader. The qualified reader guides the trainee through the screening process. In third reading, the trainee reader works as an additional (third) reader in the normal screening process – comparing their analysis against the assessment written on the screening form by the other readers. One drawback of this approach is that the trainee learns with reference to the screening decision, where the ground truth (established by tests applied subsequently at assessment clinics) is not yet known. There is anecdotal evidence that trainee readers will develop their skills in a way that mirrors the strengths (and hence their weaknesses) of their mentor.

“I think you become what your mentor is in a way because I’m good on distortions, [my mentor] is good on distortions. She was taught by Dr [name] and Dr [name] is excellent on distortions… We laugh now because I write comments and [my mentor] will come along and she will just look down and say oh, you have already written it and she was just about to make the same comment.”

In summary, CBT could be used to improve screening accuracy by providing

---

\(^1\) We acknowledge that this raises issues about the relation between soft copy reading skills and physical film reading skills.
trainee readers with an additional depth and breadth of film reading experience in conjunction with reliable feedback.

**Situated Practices**

In our experience, every breast screening centre has its own local practices. Partly this is because every centre faces a unique set of local needs. This section will examine some of the implication of these situated practices on CBT.

We wish to draw a distinction between training readers to be better discriminators of benign and malignant presentations and training readers to discriminate between recallable and non-recallable cases. The premise is that some malignant lesions are sufficiently difficult to distinguish from benign presentations, and are sufficiently less often malignant than benign in the distribution seen in screening, that they are not recallable.

“It is like the little non-descript blobs. Every woman has got little non-descript blobs and you could say 1 in probably 10,000 is going to be a cancer – is the beginnings of cancer. When you have got hindsight you see that the one that had it but you can’t go out and call them.”

The need to provide reliable feedback in training suggests the use of cases where a ground truth is known. For example, a positive result from a needle biopsy adds significant weight to a cancer diagnosis. Similarly, some weight can be added to a normal (cancer free) diagnosis if there is a subsequent normal screening round. However, this material in itself does indicate whether the case is recallable. In practice, we find that the recall decision is a locally situated one and contingent on local circumstances.

An important activity in the production of a grid-enabled and shared teaching repository is the selection and work up of cases for inclusion in the teaching archive. One concern, given that the cases will originate from more than one centre, is that uniform criteria are applied. As a senior radiologist remarked:

“The last thing you want, if lots of people are putting data onto the system, is to have inaccurate intake… You need individuals on site that have allocated time. Their role in this is to find films, pull out the data and to say, “This is what we want to learn from this film.” That has to be done by somebody who is very experienced with the knowledge and skill to do [the job].”

Given the heterogeneous nature of screening practice there are issues of variability, including, but not limited to, whether mammograms might uniformly be judged as being of an appropriate technical standard, the sorts of tests usually conducted at assessment (for instance whether there is an aggressive policy of microcalcification biopsy, or a greater dependence on imaging), differences in reading practice (is the case double read, single read, is there third reader involvement?), patient history taking (is the women asked whether she is on HRT?), differences in X-Ray sets, film processing units and film manufacturer. Given differences of this nature, it may be difficult to ensure capture of uniform information about potential training cases.

One example of this concerns the annotation of lesions as one aspect of working up a case for training purposes. While most radiologists we have spoken to are aware of BI-RADS [9] and have suggested it as an annotation schema for training cases, all the clinics that we have observed use their own locally evolved notation for marking up lesions on the
screening form [10]. There is sometimes a lack of overlap between BI-RADS and this local notation. For example, readers might write “ISQ” meaning In Status Quo for a feature that had a similar presentation on a previous form and was therefore likely to be normal. ISQ has no BI-RADS equivalent.

This is indicative of the difficulties involved in managing the workup and sharing of training cases.

Discussion

This paper has attempted to provide a more detailed picture of the possible roles of computer based training in screening mammography than has been presented in previous work.

We have looked at the medical evidence for CBT and also the intricacies involved in achieving the goals implied by that evidence. In the process, we have shown that training is more complex than the statistics imply. Some parts of screening are inherently situated and resistant to formalisation. Some knowledge used in screening is embodied in the process of reading and is difficult to codify. New practices must be developed to allow training material to be shared.

One issue then is how to mesh a generic CBT system with local needs and with other modes of training. A related issue is how much a CBT system should aim to support local needs and practices. For example, a design might use cases taken from local screening to teach the recall decision. The same system might also use “interesting” cases, shared between centres, to teach the distinction between benign and malignant cases. In this way, the system could provide both a breadth of experience and some representation of local practice.

In future work, we aim to develop prototype systems in order to explore different strategies for meeting the design goals and constraints discussed in this paper.

Acknowledgements

We would like to thank the members of the NHS breast screening programmes in England and Scotland who participated in this study for their help. This work is funded by the Chief Scientist Office, Scottish Office and the UK e-Science research programme.

References