Mammographic interpretation training: how useful is handheld technology?

Yan Chen, Alastair G. Gale, Hazel Scott
Applied Vision Research Centre, Loughborough University, Loughborough, UK, LE11 3TU

ABSTRACT

In the UK a national self-assessment scheme (PERFORMS) for mammographers is undertaken as part of the National Health Search Breast Screening Programme. Where appropriate, further training is suggested to improve performance. Ideally, such training would be on-demand; that is whenever and wherever an individual decides to undertake it. To use a portable device for such a purpose would be attractive on many levels. However, it is not known whether handheld technology can be used effectively for viewing mammographic images. Previous studies indicate the potential for viewing medical images with fairly low spatial resolution (e.g. CT, MRI) on PDAs. In this study, we set out to investigate factors that might affect the feasibility of using PDAs as a training technology for examining large, high resolution mammographic images. Two studies are reported: 20 mammographers examined a series of mammograms presented on a PDA, specifying the location of any abnormality. Secondly, a group of technologists examined a series of mammograms presented at different sizes and resolutions to mimic presentation on a PDA and their eye movements were recorded. The results indicate the potential for using PDAs to show such large, high resolution images if suitable Human-computer Interaction (HCI) techniques are employed.

Keywords: mammogram interpretation, Personal Digital Assistant (PDAs), training

1. INTRODUCTION

According to the World Health Organization (2006), cancer is responsible for 13% of all deaths globally and is the leading cause of death [1]. Breast cancer, the most common type of cancer amongst women, causes around 502,000 deaths each year. One in eight women in the US will develop invasive breast cancer over the course of their lifetime with a 1 in 33 chance of breast cancer causing their death [2]. The situation is similar in the UK; for instance approximately 36,900 new cases of breast cancer were diagnosed in 2005, with 10,300 deaths. The mortality rate amongst women, for this disease is 28 per 100,000 women [3]. It is reported that death rates for breast cancer have been declining in recent years in the US [4] and the UK [5], however breast cancer remains the most feared disease.

The National Health Service Breast Screening Programme (NHSBSP), which was initiated by the Department of Health in 1988 in the UK, was the first nationwide scheme of its kind in the world. With the purpose of facilitating the early detection of breast cancer and improving treatment, the scheme provides free breast screening every three years for all women in the UK aged 50 to 65 (and the upper age limit was recently extended to 70 years). Around one-and-a-half million women are screened in the UK each year. With the introduction of the scheme, the death rates began to fall as a result both of early cancer detection along with improved treatment. Based on the predictions of pre-screening rates in various age groups, by 1998 the death rate was about 20% lower than it would have been without the screening scheme [3].

Breast screening is a method of detecting breast cancer at a very early stage. The key step in the process is the visual examination of mammograms for the presence of abnormalities that are indicative of cancer. This task is carried out by trained mammographic film-readers (in the UK these are specially trained radiographers or consultant radiologists). Detecting early signs of breast cancer is an especially difficult task due to the rarity of the disease within the screening population: an abnormality will be present in only approximately seven cases per 1,000 women [6]. Furthermore, the various subtle ways in which an abnormality can present increases the difficulty of correct identification.

The difficulty of mammographic interpretation emphasises the importance of efficient training both to improve, and maintain, mammographers' everyday performance. Furthermore, changes within the NHSBSP have increased the need for the availability of a range of training approaches. The upper age limit of women invited to take up screening in the

Medical Imaging 2008: Image Perception, Observer Performance, and Technology Assessment,
edited by Berkman Sahiner, David J. Manning, Proc. of SPIE Vol. 6917, 691712, (2008)
1605-7422/08/$18 · doi: 10.1117/12.764483

UK has been extended from 65 to 70 years, and is set to change further to encompass women aged 47-73 years by 2012. This increased age range will increase the number of cases annually examined in the UK's breast screening centres. Thus, the combination of increased demand and limited capacity are placing pressure on the programme [7]. Consequently there is a need to train an increasing number of radiographers and to quickly bring them to advanced practitioner level (where they are qualified to examine and report breast screening cases within the national screening programme) in order that they can contribute fully to the NHSBSP.

Different methods of training are needed to help film readers to improve their expertise in identifying mammographic abnormalities at an early stage. An important practice in aiding film readers to develop the necessary skills could be to expose them to the appearances of a wide range of mammographic abnormalities through having them read a high volume of cases. Also, gaining appropriate experience of attending to specific abnormal features is another key factor which has been shown to help improve an individual's ability to recognise abnormalities [8]. However, appropriate mammographic interpretation training opportunities is, of necessity, presently somewhat limited.

In the UK there is a national self-assessment scheme which provides an opportunity for film readers to examine a wide range of especially selected, difficult exemplars of previously categorised mammographic screening cases within a short period of time. As part of the quality assurance programme for the NHSBSP, the PERFORMS scheme (PERsonal perFORmance in Mammographic Screening) was established in 1991 in response to the fact that feedback to film-readers on their screening performance on live cases at that time was of necessity very slow [8] (i.e. feedback on whether a case was a true negative or a false negative report was not obtained until that woman was scanned again in the next screening round). Thus, feedback on live screening cases plays a somewhat limited role as a training tool, although in the UK cases are discussed at weekly multidisciplinary meetings within the individual screening centres which serves as a very important training resource as radiological, pathological and surgical information per case can be presented and discussed. Limitations on opportunities for training are further exacerbated by the fact that mammographic films are traditionally viewed on a multi-viewer (Figure 1) which is a sizeable piece of equipment, statically sited at breast screening centres. This then limits the locations where training can be undertaken.

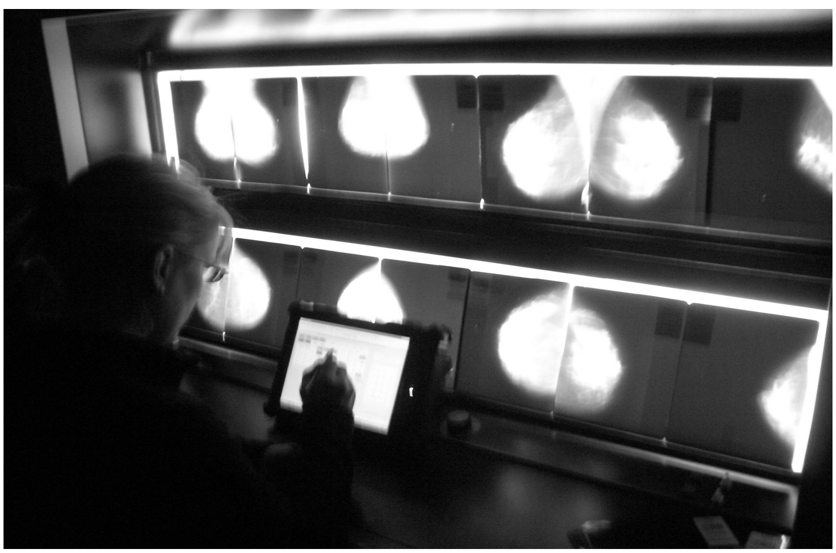

Figure 1. Mammogram multi-viewer

The increased use of digital mammography, and its forthcoming widespread adoption in the UK, offers the opportunity for extending the PERFORMS self-assessment scheme to provide increased dedicated and individualised training. Also, the success of implementing teleradiology in the domain of medical care, which allows the electronic transmission of radiological patient images from one location to another for the purposes of interpretation and/or consultation, offers the potential for more freedom in mammography interpretation training [9] -providing appropriate resolution images can be transmitted and displayed appropriately without loss of required resolution. There are some projects which already aim to employ advanced technologies to support mammographic interpretation training. For example, GIMI (Generic Infrastructure for Medical Informatics) is a collaborative project in the UK with the purpose of developing a prototype training tool for screening mammography which could offer radiologists a tailored educational experience based around the intelligent selection of training activities [10]. This is based on using grid technology to deliver training to individuals based at clinical workstations. Also, a related computer-based training (CBT) system has also been investigated to support both the improvement of the skills of experienced film-readers and the training of inexperienced ones using advances in high-quality computer displays at the mammogram viewing workstation and high-speed networking [11].

Both of these projects aim to implement training at the digital mammography workstation itself. However, technological advancements have enabled such images to be viewed on a laptop, or even on a PDA. This raises the possibility of having mammographic interpretation training delivered both whenever, and wherever, it suits the individual.

A few recent research projects have been carried out to develop mobile teleradiology systems on a PDA, which are independent of stationary and cable-bound computers. These projects are for better managing Digital Imaging and Communications in Medicine (DICOM) image data and to support group work. The wireless system allows efficient management of heavy loads of lossless DICOM image data and will be useful for collaborative work by radiologists in education, conferences, and research [12][13]. PDAs have been shown successfully to support the interpretation of CT images, whose small physical size and resolution is adequately handled by the PDA screen's resolution and size [14]. However, the potential for mammographic images to be viewed on PDAs has not yet been comprehensively researched, mainly due to an unanswered key question: can PDAs provide sufficient image quality? Particularly, the potential for using PDAs to deliver mammographic interpretation training has not been investigated. Therefore, an initial investigation into the possible factors which might affect the feasibility of using PDAs as a training technology for examining large high resolution mammographic images were investigated here.

The overall study employed digitised versions of mammographic cases which have previously been used in the PERFORMS scheme and for which both the known clinical outcome, as well as the opinions of all UK mammographers on their mammographic appearance, was known. Each participant's opinion was derived for the identification of the presence of a range of key mammographic features: ill-defined mass (IDM); spiculate mass (Spic); architectural distortion (AD); calcification (Calc); asymmetry (Asym); or no key mammographic features present (none).

The study comprised two parts. As a starting point, a pilot study was conducted at a major UK breast screening conference. This was to determine if a very small PDA screen can support mammographic interpretation amongst expert radiologists and also to collect feedback and comments from these individuals about the potential for a PDA in training. A more in-depth study was then undertaken to assess the influence of PDA screen size/pixels on performance in mammographic interpretation. In this, it was hypothesised that those viewing conditions which represent larger image size/higher pixel resolution would give rise to more correctly reported cases by the participants. Additionally it was hypothesised that the more gross mammographic features would be easier to identify on a PDA.

2. STUDY PART 1

The pilot study aimed to investigate the performance of experienced screening radiologists in making screening judgments using a PDA. This would provide an indication of the feasibility of showing mammograms on PDAs and suggest directions for further research.

2.1 Methods and materials

- Participants

Twenty experienced film-readers recruited from an opportunity sample which presented itself at the above mentioned conference.

- Materials

Visual stimuli: Seven pairs of mammographic images were selected from the PERFORMS archive of previously categorised cases. Five of the pairs featured a specific abnormality (i.e. IDM, Spic, AD, Calc and Asym) and two of the pairs featured no abnormality. Each image pair comprised the mediolateral oblique (MLO) view of both breasts. These were combined into one large image with a resolution of7200x4800 and then were transformed into 320x240 pixels which satisfy the resolution of the PDA that was used in the study.

PDA: An SPV M700 3G PDA phone was used. This PDA is commonly available in the UK and has a 2.8" screen with a resolution of 240 x 320 pixels, a 65536 colour display, and a MiniSD-expandable 64MB of memory (see Figure 2).

▪ Design

Each of the seven mammogram pairs was presented to each participant, using the PDA.

▪ Procedure

Each participant viewed each mammogram pair, in sequence, and made a decision if there was an abnormality present. Where an abnormality was identified, its location was indicated and noted by the experimenter. Also, his/her feedback from the examination of mammogram images on the PDA was collected after viewing all cases.

Figure 2. The PDA used to display mammograms in part 1 of the study

2.2 Results

The results show that the ability of detecting an abnormality, given an abnormality was present was greater than chance (Figure 3). However, performance on the normal image pairs in determining no abnormality was present (true negatives) fell below chance (Figure 3).

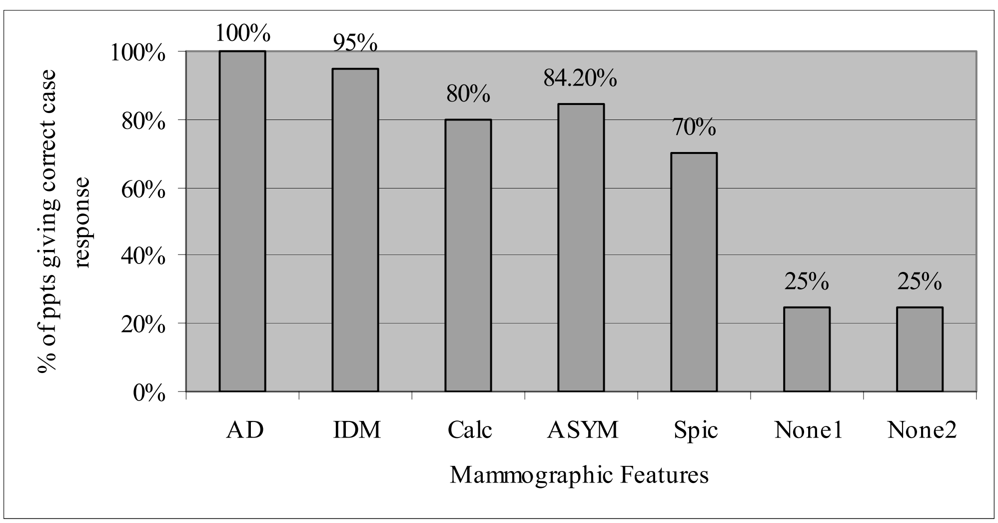

Figure 3: Participants: the percentage of participants giving correct case response per type of mammographic feature

For image pairs featuring abnormalities, the percentage of participants who specified the correct location is shown in Figure 4. In terms of performance across participants on identifying the correct location of an abnormality, performance varied between 30% for asymmetries and 80% for ill defined masses.

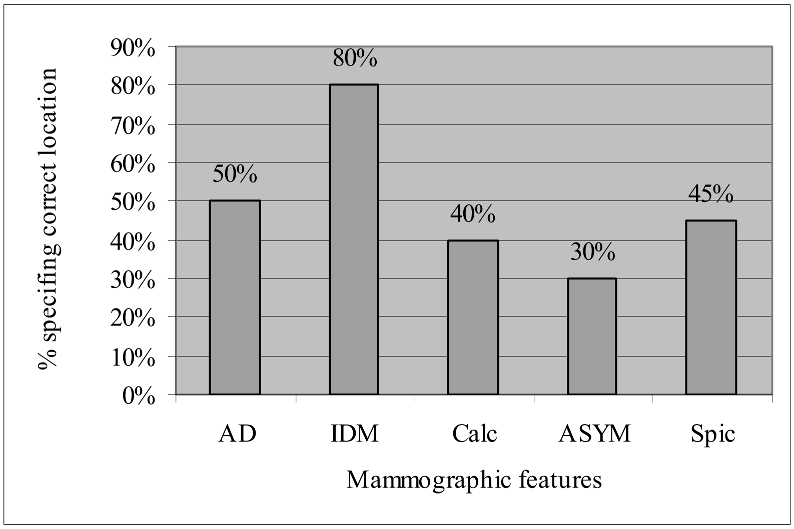

Figure 4. Participants: the percentage of participants specifying correct location by type of mammographic feature

2.3 Discussion

This pilot study provided a clear indication that it is possible to detect mammographic abnormalities (at least the specific exemplars used here) using a small PDA. However, the results indicate that there may be specific difficulties to overcome, particularly with respect to confidently determining that an abnormality is absent and also, although to a lesser degree, with respect to accurately locating abnormalities. There is an indication that these difficulties may be more pronounced for some kinds of abnormality than others. However, the present indicative data do not allow firm conclusions to drawn on this matter. Such difficulties may be overcome with a zoomable HCI interface to allow closer inspection of certain image areas. The second part of this study aimed to investigate the influence of display size to investigate further the issues associated with interpreting a mammographic image on a PDA display.

With respect to the comments from participants, although most were initially doubtful about the ability to make any decision on such a small image size, afterwards they were amenable to the idea of the potential use in training; particularly the idea of being able to zoom in to an area, or only presenting small areas of interest of mammograms, using such a device. Displaying mammograms on a PDA as a training tool is in many ways attractive but needs to be very carefully handled to maintain image quality. For viewing the images, zooming in/out with no effective pixel loss is suggested, which helps to see more details and dismiss suspicious areas. Even if zooming in/out is allowed, some mammographic features might still be expected to have difficulty to identify, e.g. calcification is expected to be very difficult to identify and masses are relatively easier.

3. STUDY PART 2

Part two of the study was conducted in order to compare the diagnostic efficacy associated with different sizes and resolution of the PDA display which, for experimental purposes, was simulated on an LCD computer monitor. The eye movements of the participants were recorded in this task. Of interest here was whether increasing the physical size of the displayed image or increasing the resolution affected the observers' behaviour.

3.1 Methods and materials

- Participants

Eight film-readers participated in the study: six second-year radiographer students with up to two years of radiography-related education and two radiographer tutors who have more than 10 years of general X-ray film reading experience. None specialised in mammography reading.

- Materials

Visual stimuli: Twenty-four pairs of mammographic images were selected from the PERFORMS archive. Each image pair comprised the mediolateral oblique (MLO) view of both breasts. Twelve of the pairs featured a specific abnormality which has been grouped into three (namely: Mass, AD/Asym and Calc) and twelve of the pairs featured no abnormality.

Hardware: The experiment was run on a PC with a 20" LCD monitor for displaying images; eye movements were recorded using a remote oculometer – a Tobii X50 eye-tracker. This records eye movements with no attachment to the participant and allows them some degree of free head movements (within 30 x 16 x 20 cm ; W x H x D) at 60 cm from the device. It has a reported accuracy of 0.5-0.7 degrees of visual angle.

- Design

Viewing conditions: Two different sizes and two different pixel resolutions were used. In terms of size, 3.5" (the most common screen size of PDA) and 6" (the biggest screen on the current general market in the UK). The two pixel groups were: 320x240 (the most common screen pixel for a PDA) and 800x600 (the highest pixel level on the current general market in the UK).

Performance in terms of proportion of correct responses was subject to a two-way, repeated-measure ANOVA: viewing condition (3.5", 6.0", 320x240, 800x600) x outcome (abnormality present, abnormality absent).

- Procedure

For each observer the eye tracker was first calibrated. Each observer was then given a short training on the appearance of the mammographic features used in the study before they completed the computer-based image examination task whilst their eye movements were discretely recorded (Figure 5). In this, the observers were asked to examine each case, identify if there was any abnormality present, and then rate their confidence in whether the image was normal or contained an abnormality. If the latter, they also had to specify the abnormality and its location. Also, they were asked to rate the quality of the images.

Each participant viewed three abnormal and three normal mammogram pairs in each of the four viewing conditions. The six mammogram pairs for each viewing condition were shown in blocks and the order of the viewing condition blocks was counterbalanced across participants using a Latin square design. Mammogram pairs were pseudo-randomly assigned to viewing condition with the condition that three normal cases and three abnormal cases must appear in viewing conditions and that the three abnormal cases must be made up of one architectural distortion (AD)/asymmetry (Asym); one calcification (Calc); and one mass.

Figure 5. Examples of viewing images in different conditions (Left: 3.5"; Right: 800x600 pixels) in part 2 of the study

3.2 Results

There was no main effect of viewing condition, F(3,21) = 1.272, p = .310.; a significant main effect of normality, F(1,7) = 10.162, p = .015; and no interaction between viewing condition and normality, F(3,21) = 0.132, p = .940, for more details see Table 1.

Table 1. Mean performance proportion of correct response by case known pathology and viewing condition

	3.5”	6”	320x240pixels	800x600pixels	Total
Abnormal	M=.750 SD=.295	M=.500 SD=.178	M=.646 SD=.187	M=.667 SD=.308	M=.640 SD=.254
Normal	M=.417 SD=.345	M=.250 SD=.295	M=.375 SD=.278	M=.417 SD=.295	M=.364 SD=.298

Figure 6 shows the percentage of correct response by viewing condition. Although there were no significant effects of viewing condition on participants' performance, it shows that there were slight differences amongst these conditions.

Although there is a main effect of outcome on participants' performance, the performance difference on the different feature groups is statistically insignificant. For image pairs featuring abnormalities, the percentage of participants who specify the correct abnormal area is shown in Figure 7. In terms of performance across participants on identifying the correct location of an abnormality, performance varied between 47.6% for calcification and 60% for masses.

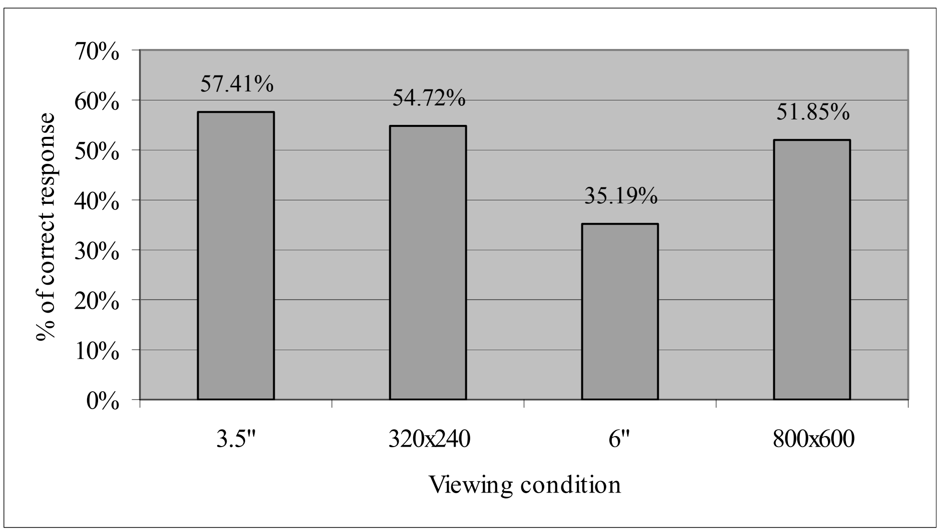

Figure 6. Participants: the percentage of correct response by viewing condition

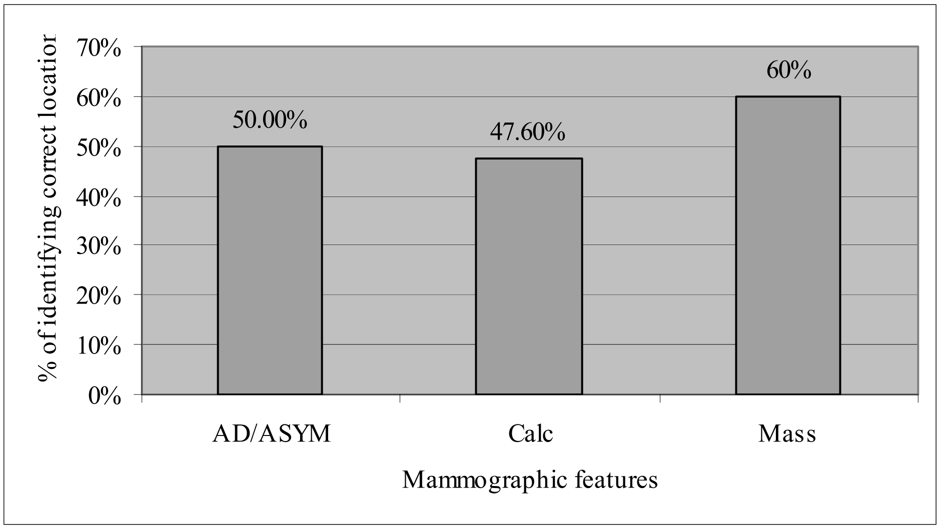

Figure 7. Participants: the percentage of giving correct location response by abnormality features

3.3 Discussion

It was hypothesised that those viewing conditions which represent a larger image size/higher pixel resolution would give rise to more correctly reported cases by the participants. However, according to the ANOVA results, viewing condition had no main effect on the performance of examining these images. Even though these trends did not reach statistical significance, the 3.5" screen, the smallest size of all the viewing conditions, appeared to support the highest level of performance. This suggests that, within the range of screen size and resolution reported here, simply increasing the PDA screen size or its pixel resolution might not be a straightforward solution for implementing PDAs for use as a mammographic interpretation training tool.

There is a significant difference between abnormal and normal cases ($p<.05$). The percentage of cases correctly classified by the participants ($n=8$) indicated that they made significantly less errors on abnormal (64% correct response) as compared to the normal cases (36.4% correct response). Similar results are also observed with in the PERFORMS scheme using normal sized mammographic film images and this may be exacerbated here by the small screen size [15]. Even so, this problem is not peculiar to PDAs and tends to result in false positive responses.

In terms of the hypothesis that some mammographic features would be easier to identify on PDAs, there was no significant difference between mammographic features on the percentage of correct responses and the percentage of correct location responses. Although the slight difference within features did not reach statistical significance, calcification appeared to be slightly more problematic, which is to be expected in accordance with the feedback from the pilot study and also from previous reported results from the PERFORMS scheme [15].

4. GENERAL DISCUSSION

This was an initial research investigation into aspects of the potential of viewing breast screening cases on a PDA for the purpose ultimately of providing some form of training in mammographic examination. Ostensibly a PDA is not a technology of choice as the typical PDA screen resolution can only accommodate about 1/10th a single mammogram – and in screening each woman must be represented for inspection by four such images. However the purpose here is to determine whether the general approach has merit and then to investigate HCI techniques fully to use such a small display to best effect.

The pilot study simply investigated whether experienced mammographers could identify a small range of abnormal features on a commonly available small PDA with a relatively low resolution screen. This study suggested the possibility of viewing mammograms on a PDA, although the performance on normal cases (with no key mammographic features present) was poor. This could be due to the nature of the study which caused observers to over read cases or that the small display actually affects their decision criterion. However, with a promising result, feedback and comments from experienced mammographers, the second study set out to collect more detailed information from a group of individuals who were familiar with radiographic appearances but not with mammography per se. Here, a significant main effect of normality was found with significantly less errors on the abnormal cases being made as compared to the normal cases, which is similar with the result of the pilot study and previous PERFORMS data [15][16].

The common difficulty in classifying normal cases which could be exacerbated by small screen size with low resolution suggests the possibility of delivering very specific training on PDAs, which takes the small size and pixel resolution into consideration. Within the different abnormal feature groups used here, the performance of detecting a mass was best, even though there is no significant difference within feature groups. Although there is no significant difference between the performances on different viewing conditions, the 3.5" screen was shown to be the best viewing condition – this may reflect a matching of the screen's spatial resolution to that of the human visual system and is a matter for further investigation. Also, the observers' eye movements were recorded in the second study (examples are shown in Figure 8), which allows examination of each individual's visual search behaviour and therefore enables the errors to be studied in more detail [17] and potentially gives additional opportunity of recommending tailored training to each individual. The analyses of the visual behaviour will be reported in a future paper.

It is suggested that further research needs to be carried out on how best to employ suitable HCI techniques to increase the feasibility of mammographic interpretation training on PDAs rather than overly, and simply, focusing on increasing the screen size and resolution. Overall, results from both studies indicate the, somewhat surprising, possibility of viewing mammograms, or parts of mammograms, on PDAs and to use these devices to deliver targeted training as, and when, required.

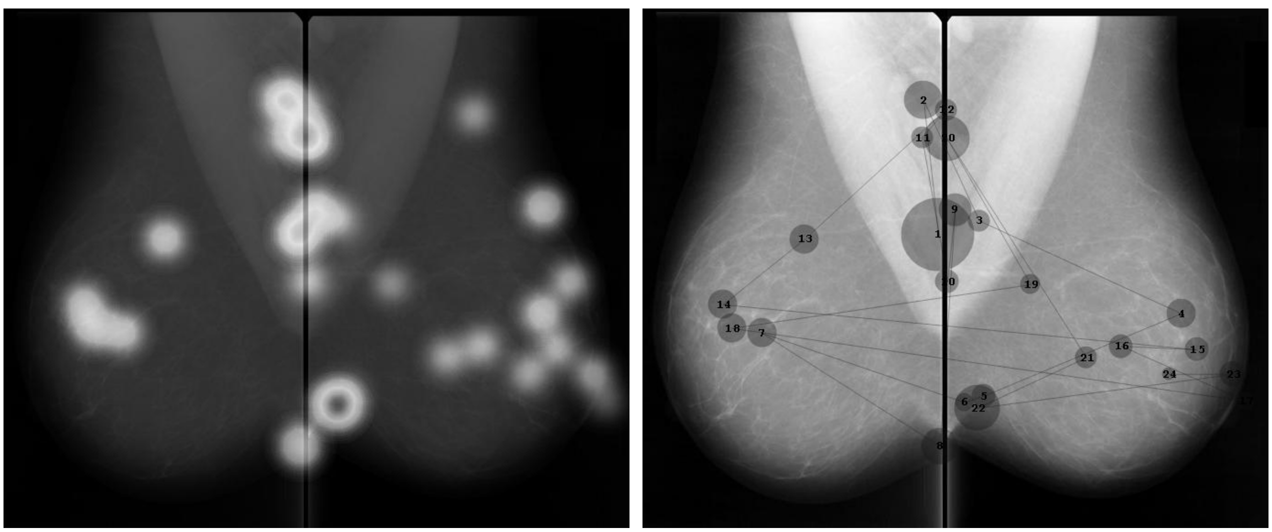

Figure 8. Examples of a mammogram used here with overlaid eye tracking data

5. CONCLUSION

The results indicate the technical potential of using PDAs as a training technology for examining mammograms. Whilst this is not an obvious choice for such high resolution images and would not be used as the sole training technology, it does demonstrate that PDAs can help deliver training to an individual, as and when they demand it. Further work is underway to improve the usability of a PDA system by facilitating the engagement of human-computer interaction (HCI) techniques to help such a mobile training scheme.

REFERENCES

[1] World Health Organization, "Cancer", Fact sheet 297, (2006). [Accessed: 14th Jan, 2008].

[2] American Cancer Society, "What Are the Key Statistics for Breast Cancer?", Cancer Reference Information, (2006). [Accessed: 16th Jan, 2008].

[3] National Statistics Online, "Breast Cancer: Incidence rises while deaths continue to fall", Office for National Statistics, (2007). [Accessed: 17th Jan, 2008].

[4] Espey, D. K., Wu, X. C., Swan, J., et al. "Annual report to the nation on the status of cancer, 1975-2004, featuring cancer in American Indians and Alaska Natives", Cancer 110 (10), 2119-52 (2007).

[5] Society for Women's Health Research, "Women's Fear of Heart Disease Has Almost Doubled in Three Years, But Breast Cancer Remains Most Feared Disease", Press Release, (2005).

[6] Patnick, J., editor. "One vision. NHS Breast Screening Programme Annual Review 2005", NHS Cancer Screening Programmes, (2005).

[7] Nickerson, C., Cush, S., "New ways of working in the NHS breast Screening Programme", Second report on Implementation, NHS Breast Screening Programme Publication, (2004).

[8] Gale, A.G., "PERFORMS-a self assessment scheme for radiologists in breast screening", Seminars in Breast Disease 6(3), 148-52 (2003).

[9] Ruggiero, C., "Teleradiology: a review", Journal of Telemedicine and Telecare, 4, Number 1, 1 March 1998, 25-35(11) (1998).

[10] GIMI, GIMI homepage http://www.gimi.ox.ac.uk/. [Accessed: 18th Jan, 2008].

[11] Soutter, J., Campos, J., Hartswood, M., Jirotka, M., Procter, R., Slack, R., et al., "Grid-based mammography training", Hospital Radiologist 5(6), (2003).

[12] Schweitzer, T., Engelmann, U., Schröter, A., Boraelv, E., Meinzer, H.P., "Teleradiology on a Personal Digital Assistant", 2nd Conference on Mobile Computing in Medicine 15, workshop of the Project Group MocoMed 2002, 73-80(2002).

[13] Nakata, N., Kandatsu, S., Suzuki, N., Fukuda, K., "Informatics in Radiology (infoRAD): Mobile wireless DICOM server system and PDA with high resolution display: feasibility of group work for radiologists", RadioGraphics 25, 273-83 (2005).

[14] Toomey, R.J., Ryan, J.T., McEntee, M.F., Evanoff, M., Ryan, P., Manning, D.J., Chakraborty, D.P., Curran, K., Brennan, P.C., "The Use of Personal Digital Assistants in the Reporting of Radiological Images", Proceeding of Medical Image Perception Society Meeting, (2007).

[15] Scott, H.J., Gale, A.G., "Breast screening: PERFORMS identifies key mammographic training needs", The British Journal of Radiology 79, 127-33 (2006).

[16] Scott, H.J., Gale, A.G., "Breast screening: when is a difficult case truly difficult and for whom?", Proceedings of SPIE 2005 5749, 557-65 (2005).

[17] Kundel, H.L., Nodine, C.F., Carmody, D., "Visual scanning, pattern recognition, and decision making in pulmonary nodule detection, Investigative Radiology, 12(1), 175-181 (1978).