Evaluation of a CAD (Computer Aided Detection) enhanced 2D synthetic mammogram: Comparison with standard synthetic 2D mammograms and conventional 2D digital mammography

JJ James [1], E Giannotti [1] & Y Chan [2]

[1] Nottingham Breast Institute, [2] Loughborough University

Introduction

Digital breast tomosynthesis (DBT) when combined with standard 2D digital mammography has been shown to improve the performance of breast cancer screening by increasing cancer detection rates [1-5]. The 2D component remains an important part of the examination and is used to facilitate assessment of symmetry between the breasts, aid comparison with prior mammograms and identify the presence of breast microcalcifications where the evidence for detection with DBT is less robust [1].

The mean glandular dose per view of a DBT image is around 2.3 mGy, which is between 1 - 1.5x more than the dose of standard 2D digital mammography [6]. Acquiring both a DBT and standard 2D digital mammogram on each woman leads to at least a doubling of the radiation dose, which may not be considered acceptable in an asymptomatic screening population. Consequently there has been much interest in the generation of synthetic 2D mammograms from the DBT data set eliminating the additional radiation burden of a separate 2D digital mammogram.

There is evidence from prospective and retrospective studies to support the use of synthetic 2D mammograms [5,7-9]. Several retrospective multi-reader studies, including the UK TOMMY trial, have demonstrated comparable performance between synthetic and conventional 2D mammography [7,8]. The Oslo and Storm-2 prospective studies of DBT in breast cancer screening found equivalent cancer detection rates regardless of whether the conventional 2D or the synthetic mammograms were read, concluding that synthetic mammograms were an acceptable replacement for directly acquired conventional 2D mammograms [5,9].

Another approach to improve performance is to combine the synthesised image with a Computer Aided Detection (CAD) algorithm. CAD has been used over the years to assist with the interpretation of 2D mammography. CAD software places marks or prompts on the images to draw the reader's attention to potential areas of concern, reducing observational oversights. A CAD algorithm has been developed with machine learning technology (iCAD Inc., Nashua , NH, USA and GE Healthcare, Buc, France) to assist in the detection of breast cancer on DBT images. Unlike a conventional CAD system which places marks on the image, areas of concern are automatically identified on each tomosynthesis slice and then blended onto a 2D synthetic image to provide a single CAD enhanced 2D synthetic image for each mammographic projection.

The aim of this study was to evaluate the diagnostic performance of the CAD enhanced synthetic mammogram in comparison with standard 2D synthetic mammograms generated from the DBT data set and standard 2D digital mammography.

Method:

The cases were collected as part of a performance study of DBT in the assessment of soft tissue screen detected abnormalities in the NHS Breast Screening Programme (NHSBSP). The study had local National Health Service (NHS) Research Ethics committee approval, with all participants giving written consent for the use of their images. 68 anonymised cases where raw data was available were retrospectively reprocessed by the vendor (GE Healthcare) to produce the CAD enhanced 2D synthetic image (Enhanced Volume Preview®). DBT images were only acquired on the side being investigated in the screening recall assessment clinic and so only a single oblique and cranio-caudal projection was available for each case. The original 2D screening mammogram and the standard synthetic 2D mammogram generated from the DBT data set (Volume Preview®) were also available for the affected side. DBT imaging was acquired on a commercially available GE Healthcare DBT system (SenoClaire®). The original 2D screening mammograms were performed on GE Healthcare full-field digital mammography machines (Senographe Essential®).

Retrospective reading was performed by two fellowship trained, breast radiologists with 3 and 18 years breast imaging experience respectively. Both readers fulfilled NHSBSP Quality Assurance Criteria for screening mammography. Each pair of oblique and cc images (standard 2D, standard synthetic 2D and enhanced synthetic 2D) were read separately and independently by each reader. Consequently there were 204 pairs of images and the reading order was separately randomised for each reader. The readers were blinded to the final pathology or outcome of assessment for each case and the type of image being read. The readers had no access to the tomosynthesis images used to generate the synthetic 2D images. Each reader assigned an imaging score to each pair of images using the UK Royal College of Radiologist 1-5 score [10]. Histopathology provided the ground truth outcome for those subjected to biopsy.

After the scores were assigned, the readers were unblinded and for the cancer cases determined whether the correct area had been enhanced by the CAD and the total number of areas enhanced on each image, in order to produce a measure of the sensitivity and specificity of the system. The sensitivity was determined by dividing the number of cancers that were correctly marked on the oblique or cranio-caudal projections by the total number of cancers. Specificity in CAD studies is traditionally measured by determining the number of false positive marks per image.

Differences in the performance were assessed using receiver operator characteristic (ROC) analysis. The area under the curves (AUC) was compared using Hanley two-tailed test. $P<0.05$ was regarded as statistically significant.

Results

The study population consisted of 68 cases. There were 34 malignant cases and 34 normal/benign cases. Of the malignant cases, 31 were invasive carcinoma (91%), two were ductal carcinoma in situ (DCIS)(6%) and one a case of aggressive fibromatosis. Four of the malignant cases had multifocal disease and so there were 38 malignant lesions in 34 cases.

ROC analysis was performed based on the imaging score assigned by each reader to each case (Figure 1). This showed that diagnostic accuracy was significantly improved with the CAD enhanced synthetic mammogram compared to the standard synthetic mammogram. Reader averaged AUC were 0.846 and 0.683 respectively (p=0.004). Diagnostic accuracy was also significantly improved with the CAD enhanced synthetic mammogram compared to the conventional 2D mammogram. Reader averaged AUC were 0.846 and 0.724 respectively (p=0.027). There was no significant difference in the diagnostic accuracy between the conventional 2D mammogram and standard synthetic mammogram. Reader averaged AUC were 0.723 and 0.683 respectively (p=0.52).

For the malignant lesions the CAD enhanced synthetic mammogram demonstrated correct enhancement of 37 or the 38 lesions - a sensitivity of 97.4% for malignant lesions in this case set. The only case that was not enhanced was a 5mm histological grade 1 tubular mixed carcinoma which manifested as a 5mm cluster of microcalcifications. For all 68 cases, there was an average of 3.2 areas enhanced per image. For the 34 cancer cases there were 2.1 false areas enhanced per image (and 1.1 true areas enhanced per image).

Discussion

The CAD enhanced synthetic mammogram is a novel approach to improving the performance of the 2D mammograms synthesised from the DBT data set. These results show that CAD enhancement can improve the diagnostic performance of the synthetic mammogram, potentially achieving a performance which is superior to the standard 2D mammogram. The performance of any synthetic mammogram is very dependent on the algorithm used to produce it. Software which produces synthesised 2D mammograms will often undergo multiple re-writes to optimise the image. In the Oslo study DBT screening study an earlier software version yielded a synthesised 2D image where the performance was inferior to the conventional 2D mammogram. Subsequent software improvements led to synthetic images which were equivalent to the conventional 2D mammogram [9]. CAD enhancement is an alternative software upgrade which also has the potential to improve performance of the synthesised mammogram.

In this study, the performance of the CAD enhanced synthetic mammogram exceeded that of the conventional 2D mammogram. The Oslo DBT study found that some cancers were better demonstrated on synthetic mammograms while others were more visible on conventional 2D mammography, although they were unable to discern any differences based on breast density, grade, size or radiological feature [9]. Zuley et al. performed a retrospective reading study and demonstrated that the majority of readers showed a trend for improved performance when reading synthetic 2D mammograms alone compared to standard 2D mammograms read alone, although the difference was not statistically significant [7]. CAD enhancement uses information from the DBT

data set to blend onto the synthetic image, so lesions better demonstrated on DBT may be better seen with CAD enhancement, which may explain the improved performance compared to the standard synthetic image and conventional 2D mammogram. Phantom studies have suggested that synthetic mammograms may be superior for demonstrating high contrast objects, whereas small lower contrast lesions such as microcalcificatons can get lost in the background noise of the inherently lower resolution synthetic image [11].

The performance of the CAD enhanced synthetic mammogram seems to be associated with the ability of the algorithm to correctly enhance cancers manifesting as soft tissue abnormalities. In this study all but one of the malignant lesions (97.4%) were correctly enhanced. Images were acquired in women undergoing assessment of soft tissue abnormalities so cases where microcalcification was the dominant abnormality were not included. Interestingly the only cancer in this study that was not enhanced by the CAD algorithm manifested as a small area of microcalcifications in a patient recalled for screening assessment for an asymmetric density that was benign on biopsy. The CAD enhancement algorithm used here is specifically designed to detect soft tissue lesions rather than microcalcifications. This is in marked contrast with traditional prompt based CAD systems used in the interpretation of conventional 2D screening mammography which have a high sensitivity for detecting microcalcifications, often approaching 100% in some studies [14, 15].

The CAD enhancement algorithm produces a large number of falsely enhanced areas on the mammogram (2.1 per image). Studies of traditional CAD used with conventional 2D mammography have suggested that the number of false prompts can adversely affect system performance, potentially increasing false positive interpretations [12,13]. In addition, a large number of falsely placed marks has the potential to distract the reader so that on occasions correctly place marks may be overlooked, leading to false negative interpretations [14]. The data presented here in this limited retrospective reading study shows no evidence of this, but the case set used was heavily enriched with cancers, so prospective evaluation of the algorithm in a representative screening population is required.

Our study has other limitations. It had a simple design to provide a basic comparison between the three different conditions (conventional 2D mammography, standard 2D synthetic mammogram and CAD enhanced synthetic mammogram), with the reading order separately randomised for each reader. In reality, synthetic images would always be read in conjunction with the DBT images and not in isolation as was undertaken here. Manufacturers design the synthetic studies to be used as guide during the interpretation of the DBT image set, rather than a standalone modality. Readers also had no access to images of the contralateral (normal) breast. Sometimes it is the lack of symmetry between sides which can draw attention to an abnormality. There is also potential bias towards the conventional 2D mammogram because all the abnormalities were originally detected and recalled as a result of the conventional 2D screening study. In addition, the performance of synthetic mammography and probably CAD enhancement is highly dependent on the computer software and algorithms used to generate the images and so it is difficult to extrapolate the findings to different manufacturers' solutions and software versions.

Further work is required to determine how best to integrate the CAD enhanced synthetic image into the DBT screening workflow. Although the performance of the CAD enhanced synthetic mammogram was superior to the standard synthetic image, it is unclear as to whether the CAD

enhanced synthetic image is best used as a replacement for the standard synthetic 2D image or should be used in a similar fashion to a conventional CAD system, with the areas of CAD enhancement being displayed as an overlay to the standard synthetic image at the readers discretion. The latter approach results in the areas of enhancement behaving like traditional CAD prompts drawing the reader's attention to potential areas of concern, with the aim of reducing observational oversights.

In conclusion, CAD enhancement offers an additional approach to further improve the performance of synthetic 2D mammography. Blending information from the DBT data set onto a synthetic image has the potential to produce a 2D image which is superior to conventional 2D mammogram for the detection of soft tissue breast lesions. This study adds to the body of evidence that the latest software developments should allow synthetic images to replace conventional 2D mammography for DBT screening, eliminating the additional radiation burden of a separate conventional 2D mammogram.

References

1 Gilbert FJ, Tucker L, Young KC. Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. Clinical Radiology 2016 71:141-150

2 Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol 2013 14: 583-589

3 Skaane P, Bandos AI, Guillien R, Eben EB, Ekseth U, Haakenaasen U, et al. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. Eur Radiol 2013 23:2061-2071

4 Lang K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality; results from the Malmo Breast Tomosynthesis screening trial, a population based study. Eur Radiol 2016 26:184-190

5 Bernardi D, Macaskill P, Pellegrini M, Valentini M, Fanto C, Ostillo L, et al. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. Lancet Oncol 2016 17:1105-1113

6 Bouwman RW, Van Engen RE, Young KC, den Heeten GJ, Broeders MJM, Schopphoven S, Jeukens CRLPN, et al. Average glandular dose in digital mammography and digital breast tomosynthesis: comparison of phantom and patient data. Phys Med Biol 2015 60:7893-7907

7 Zuley ML, Guo B, Catullo VJ, Chough DM, Kelly AE, Lu AH, Rathfon GY, Spangler ML, et al. Comparsion of two-dimensional synthesized mammograms versus original digital mammograms alone and in combination with tomonsynthesis images. Radiology 2014 271:664-671

8 Gilbert FJ, Tucker L, Gillan M, etal. Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a UK retrospective reading study (TOMMY Trial). Radiology 2015

9 Skaane P, Bandos AI, Eben EB, Jebsen IN, Krager M, Haakenaasen U, Ekseth U, Izadi M, Hofvind S, Gullien R. Two-view digital breast tomosynthesis screening with synthetically reconstructed projections images: Comparison with digital breast tomosynthesis with full-field digital mammographic images. Radiology 2014 271:655-663

10 Maxwell AJ, Ridley NT, Rubin G, Wallis MG, Gilbert FJ, Michell MJ; Royal College of Radiologists Breast Group. The Royal College of Radiologists Breast Group breast imaging classification. Clinical Radiology 2009 64:624-627

11 Nelson JS, Wells JR. How does c-view image quality compare with conventional 2D FFDM? Medical Physics 2016 43:2538-2547

12 Gilbert FJ, Astley SM, Gillan MGC, Agbaje OF, Wallis MG, James J et al. Single reading with computer-aided detection for screening mammography. NEJM 2008 359:1675-1684

13 Fenton FJ, Taplin SH, Carney PA, Abraham L, Sickles EA et al. Influence of computer-aided detection on performance of screening mammography. NEJM 2007 356:1399-1409

14 James JJ, Gilbert FJ, Wallis MG, Gillan MGC, Astley AM, Boggis CRM et al. Mammographic features of breast cancers at single reading with computer-aided detection and at double reading in a large multicentre prospective trial of computer-aided detection: CADET II. Radiology 2010 256:379-386

15 Brem RF, Hoffmeister JW, Zisman G, DeSimio MP, Rogers SK. A computer aided detection system for the evaluation of breast cancer by mammographic appearance and lesion size. Am J Roentgenol 2005 184:893-896