Evaluation of an AI tool to measure mammographic density for use in a future FAST MRI trial

S Gomes, LM Warren, M Halling-Brown, KC Young, M Trumble, Lyn I Jones, K Klimczak, J Rose and T Timlin

From the Department of Scientific Computing (S.G., M.T., M.D.H-B., L.M.W) and National Co-ordinating Centre for the Physics of Mammography (K.C.Y), Royal Surrey NHS Foundation Trust, Egerton Road, Guildford GU2 7XX, England; North Bristol NHS Foundation Trust (L.I.J., T.T), Bristol, England; Centre for Vision, Speech and Signal Processing (M.D.H-B); Department of Physics (K.C.Y.), University of Surrey, Guildford, England; Independent Cancer Patients' Voice, London (J.R), North Bristol NHS Trust, Southmead Hospital, Southmead Road, Westbury on Trym, Bristol (KK,LJ)

Introduction

High breast density can increase a woman's risk of developing breast cancer [1] and can mask cancers on mammograms, making them harder to detect [2]. Breast composition on a mammogram can be categorised using the American College of Radiology BI-RADS 5th edition classification into one of four density categories - a (almost entirely fat), b (scattered areas of fibroglandular density), c (heterogeneously dense) and d (extremely dense).

First post-contrAst SubtracTed images Magnetic Resonance Imaging (FAST MRI) is an abbreviated breast MRI protocol that can detect cancers otherwise undetected by mammography for women with dense breasts (BI-RADS 5th ed. density categories c and d), but it is not known if it could similarly improve cancer detection for women with BI-RADS category b density. FAST MRI DYAMOND (Diagnostic Yield study for Average MammOgraphic screeNing Density) [3] "ISRCTN74193022" is a multi-centre study led by North Bristol NHS Trust designed to evaluate the ability of FAST MRI to find cancers missed by mammography within the UK's population-risk NHS Breast Screening Programme (NHSBSP) [4]. The study will invite women attending their first screening mammogram with BI-RADS b density category and a normal mammogram result for supplemental screening with a FAST MRI scan. To determine eligibility for supplemental imaging using FAST MRI, an efficient and reliable method for measuring breast density from mammograms is needed.

The purpose of this study is to evaluate the accuracy and reliability of a previously developed research AI breast density classification tool, for use in the FAST MRI DYAMOND study.

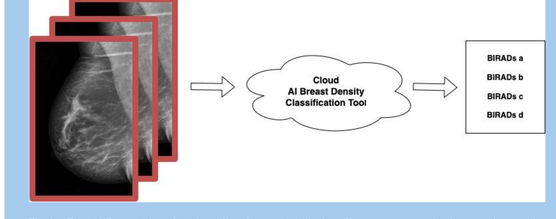


Fig.1 – The AI Breast density classification tool will classify mammography images into a density category associated with one of the 4 BI-RADS 5th ed. density categories.

Methods and materials

A stand-alone evaluation of a previously trained research AI tool was performed. The research AI tool predicts four BI-RADS 5th edition-associated categories of breast density. The four-class classification model within the AI tool was trained using a 5-layer convolutional neural network (CNN), depicted in Figure 2, and was developed using Pytorch (version 1.4.0).

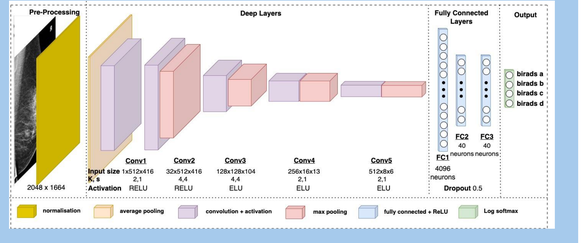


Fig.2 - Overview of network architecture and characteristics of each layer for the previously developed AI tool that is being tested in this paper. Layers' visual sizes are indicative only and not to scale, please refer to the input size indicated as n x h x w where n is the number of features or depth, h is height and w is the width of the input of each layer. K and s represent the kernel and stride of the convolution layer. The optimizer was Adam, learning rate = 10^{-4}, batch size 4. The weights were initialised by randomly drawing them from a gaussian distribution with standard deviation of sqrt(2/n) where n is the number of weights.

The inputs into the model were processed images, which were padded to the same aspect ratio and then resized to 2048 X 1664, to avoid distorting them. The mean and standard deviation of the pixel values in the training images were calculated. These were used during training and validation to normalise the images to zero mean and unit standard deviation as is common in such studies [9]. No other augmentation or pre-processing was performed.

Processed mammograms acquired on Hologic X-ray systems and their Volpara® (version 1.5.4) 5th Edition BI-RADS density classifications from unprocessed images were collected for 12,627 women from two UK screening sites. Inclusion criteria specified mammograms from women most likely to benefit from supplemental screening (age 50-55, first screen). BI-RADS classifications from the AI tool were compared, to those from Volpara® (ground truth). Volpara® is a commercial software tool widely regarded as the gold standard for breast density measurements. 95% confidence intervals, calculated using 1000 bootstrap samples were used to compare performance across age groups and ethnicity.

Results

The percentage of women in BI-RADs categories a, b, c and d respectively correctly categorised by the AI tool was 85% (95%CI:82-88), 87% (95%CI:86-88), 81% (95%CI:80-82) and 83% (95%CI:82-85).

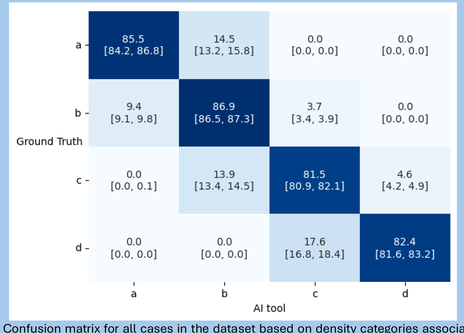


Fig.3 - Confusion matrix for all cases in the dataset based on density categories associated with the Breast Imaging-Reporting and Data System (BI-RADS) 5th edition classification. The 95 % confidence intervals are calculated using bootstrapping with 1000 samples.

Within each BIRADs category, the 95% confidence intervals across age and ethnicity overlapped, indicating the difference in performance was not statistically significant.

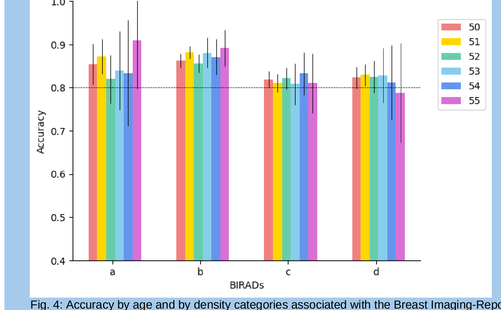


Fig. 4: Accuracy by age and by density categories associated with the Breast Imaging-Reporting and Data System (BI-RADS) 5th edition classification. The error bars represent the 95 % confidence intervals calculated using 1000 samples' bootstrap.

The distribution of cases was 6%, 42%, 34% and 18% for BIRADs a, b, c and d for Volpara. This compared with 9%, 42%, 32% and 17% for the AI tool.

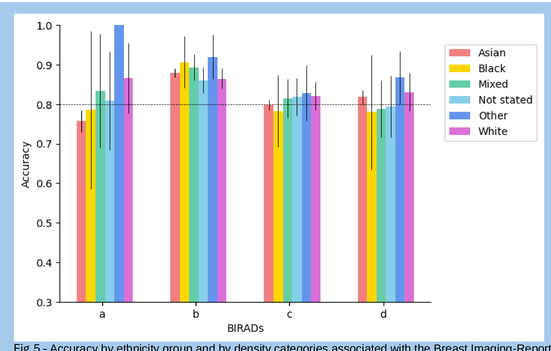


Fig.5 - Accuracy by ethnicity group and by density categories associated with the Breast Imaging-Reporting and Data System (BI-RADS) 5th edition classification. The error bars represent the 95 % confidence intervals calculated using 1000 samples' bootstrap (for density categoryBI-RADs a, ethnicity 'Other' there were only 2 samples and so it was not possible to calculate confidence intervals).

Acknowledgments

The creation of the OPTIMAM Image Database was funded by Cancer Research UK (C30682/A28396). The DYAMOND study was funded by Southmead Hospital Charity Research Fund and sponsored by North Bristol NHS Trust. The authors would like to thank Volpara for providing the use of their software for this study

Conclusion

In conclusion, an AI research tool that uses processed screening mammography images for predicting BI-RADS 5th edition-associated density categories has shown comparable performance to the reference density measurements calculated with Volpara® on paired unprocessed mammography images. This validates the use of the research AI tool in the DYAMOND study, to randomise women for supplemental imaging with Fast MRI.

References

1.McCormack VA, dos Santos Silva I. Breast Density and Parenchymal Patterns as Markers of Breast Cancer Risk: A Meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2006;15(6):1159-1169. doi:10.1158/1055-9965.EPI-06-0034

2.Chiu SYH, Duffy S, Yen AMF, Tabár L, Smith RA, Chen HH. Effect of Baseline Breast Density on Breast Cancer Incidence, Stage, Mortality, and Screening Parameters: 25-Year Follow-up of a Swedish Mammographic Screening. *Cancer Epidemiol Biomarkers Prev*. 2010;19(5):1219-1228. doi:10.1158/1055-9965.EPI-09-1028Sickles, EA, D'Orsi, CJ, Bassett, LW, et al. ACR BI-RADS® Mammography. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Published online 2013.

3.FAST MRI Research Programme | North Bristol NHS Trust. Accessed June 13, 2023. https://www.nbt.nhs.uk/FASTMRI

4.NHS England. Breast screening: quality assurance standards in radiology. Published online 2011. https://www.gov.uk/government/publications/breast-screening-quality-assurance-standards-in-radiology