Acta Radiologica
2018, Vol. 59(9) 1051–1059

DOI: 10.1177/0284185117748487
journals.sagepub.com/home/acr
SAGE

New reconstruction algorithm for digital breast tomosynthesis: better image quality for humans and computers

Alejandro Rodriguez-Ruiz[1], Jonas Teuwen[1], Suzan Vreemann[1], Ramona W Bouwman[2], Ruben E van Engen[2], Nico Karssemeijer[1], Ritse M Mann[1], Albert Gubern-Merida[1] and Ioannis Sechopoulos[1,2]

Abstract

Background: The image quality of digital breast tomosynthesis (DBT) volumes depends greatly on the reconstruction algorithm.

Purpose: To compare two DBT reconstruction algorithms used by the Siemens Mammomat Inspiration system, filtered back projection (FBP), and FBP with iterative optimizations (EMPIRE), using qualitative analysis by human readers and detection performance of machine learning algorithms.

Material and Methods: Visual grading analysis was performed by four readers specialized in breast imaging who scored 100 cases reconstructed with both algorithms (70 lesions). Scoring (5-point scale: 1 = poor to 5 = excellent quality) was performed on presence of noise and artifacts, visualization of skin-line and Cooper's ligaments, contrast, and image quality, and, when present, lesion visibility. In parallel, a three-dimensional deep-learning convolutional neural network (3D-CNN) was trained (n = 259 patients, 51 positives with BI-RADS 3, 4, or 5 calcifications) and tested (n = 46 patients, nine positives), separately with FBP and EMPIRE volumes, to discriminate between samples with and without calcifications. The partial area under the receiver operating characteristic curve (pAUC) of each 3D-CNN was used for comparison.

Results: EMPIRE reconstructions showed better contrast (3.23 vs. 3.10, $P = 0.010$), image quality (3.22 vs. 3.03, $P < 0.001$), visibility of calcifications (3.53 vs. 3.37, $P = 0.053$, significant for one reader), and fewer artifacts (3.26 vs. 2.97, $P < 0.001$). The 3D-CNN-EMPIRE had better performance than 3D-CNN-FBP (pAUC-EMPIRE = 0.880 vs. pAUC-FBP = 0.857; $P < 0.001$).

Conclusion: The new algorithm provides DBT volumes with better contrast and image quality, fewer artifacts, and improved visibility of calcifications for human observers, as well as improved detection performance with deep-learning algorithms.

Keywords
Digital breast tomosynthesis, visual grading analysis, deep learning, reconstruction algorithms

Date received: 7 July 2017; accepted: 23 November 2017

Introduction

Digital mammography (DM) is currently the most used technique for breast cancer detection, and population-based mammography screening programs have been proven to reduce mortality among women while being cost-effective (1,2). However, mammography projects a three-dimensional (3D) object, the breast, onto a two-

[1]Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, the Netherlands
[2]Dutch Expert Centre for Screening (LRCB), Nijmegen, the Netherlands

Corresponding author:
Ioannis Sechopoulos, Department of Radiology and Nuclear Medicine, Radboud University Medical Centre, Geert Grooteplein 10, 6525 GA, Post 766, Nijmegen, The Netherlands.
Email: Ioannis.sechopoulos@radboudumc.nl

dimensional (2D) image. As a consequence, there is an inherent loss of sensitivity and specificity due to anatomical noise arising from tissue superposition. Digital breast tomosynthesis (DBT) can overcome the limitations of DM by providing a pseudo-3D image of the breast (3), and many prospective trials and retrospective studies have demonstrated the clinical benefit of introducing DBT for breast cancer detection (4–9). Therefore, DBT might be considered a potential candidate to replace DM for population-based screening (10,11).

DBT consists of the acquisition of several low-dose planar X-ray projections of the compressed breast over a limited angular range, which are then reconstructed into a pseudo-3D volume. This acquisition strategy has inherent challenges that deteriorate image quality (3). The limited angle acquisition gives rise to out-of-plane artefacts and low vertical resolution (12–15), the low-dose per projection increases the impact of noise, and X-ray scatter decreases contrast (16). The reconstruction algorithm is one of the main aspects of image creation that could ameliorate these technical drawbacks, and therefore can greatly affect the final quality of DBT images.

Many different reconstruction approaches have been studied over time (17). Traditionally, the most widespread algorithm across DBT systems is filtered back projection (FBP), an analytical reconstruction method widely used in computed tomography (CT) and adapted for DBT (18,19). Fully iterative reconstruction algorithms are also in use (20–22). In order to make the most out of both approaches, FBP is recently being complemented with *a posteriori* iterative optimizations, in order to reduce artifacts and noise, and increase contrast of the DBT images (23,24), without lengthening the reconstruction time substantially (one of the main drawbacks of iterative reconstructions).

One manufacturer has followed this approach in their DBT system (Mammomat Inspiration, Siemens Healthineers, Forchheim, Germany), recently updating the clinical standard reconstruction algorithm on their system from FBP to FBP with *a posteriori* iterative optimizations (called EMPIRE), with preliminary results pointing to a decrease in artifacts and noise while enhancing image contrast of DBT volumes (23–25) (online only).

In this work, we compare this new DBT reconstruction algorithm to the previous one using clinical patient images with two methodologies. First, in order to assess the benefits of the new algorithm in terms of image quality and lesion depiction, we perform a visual grading analysis (VGA) study (26) with human readers. Second, we assess if the new DBT reconstruction algorithm provides images that also benefit automated computer detection systems. In particular, we trained and tested two equivalent deep-learning based 3D convolutional neural networks for the task of detecting calcifications in DBT, one using FBP images and the other with EMPIRE images. Deep learning is an artificial intelligence computer technique (27) that has achieved similar to superior performance to humans for many complex medical imaging tasks (28). In mammograms, a small calcification may indicate the presence of cancer, either in situ or invasive, thus detection is important (29). However, their small size (range = 0.050–3 mm) increases detection time, and certainly deep-learning based computer systems could aid humans in this task (30).

Material and Methods

Reconstruction algorithms

The two reconstruction algorithms compared in this work are both clinical standard algorithms used by the Siemens Mammomat Inspiration DBT system: the FBP algorithm; and the new Enhanced Multiple Parameter Iterative Reconstruction (EMPIRE), introduced in 2016.

The FBP algorithm for DBT is described in detail in the work by Mertelmeier et al. (19). It basically back projects the DBT projections after application of different filters to account for the limited sampling of DBT in the vertical direction throughout the breast. The EMPIRE algorithm is based on FBP, but it includes additional processes aiming to achieve better artifact suppression, higher resolution, and less noise (23–25).

Patient data

Out of a total of 2071 DBT patient studies acquired during clinical routine work-up as per standard practice at our institution between December 2014 and December 2015, 374 were consecutively collected, without any exclusion criteria, to obtain a case set with the proportions described in Table 1. All participants consented to participate in research studies within our institution and the need for specific written informed consent for this study was waived by the ethics committee.

All patients underwent an imaging protocol consisting of at least unilateral one-view DBT and digital mammography with a Siemens Mammomat Inspiration DBT system. All images were acquired in automatic exposure control mode. For a full DBT scan, the X-ray tube moves in an arc of 50° and acquires 25 projection images with an angular range of approximately 46°, during a total scan time of 20 s. The projection images were subsequently reconstructed by the DBT system into a pseudo-3D volume with focal planes

Table 1. DBT patient studies used in each experiment.

Total included patients (n = 374)			
	Normal (BI-RADS 1–2)	Biopsied benign	Biopsied malignant
VGA study (n = 100)	30	30 (soft tissue, n = 19; calcifications, n = 10; both types, n = 1)	40 (soft tissue, n = 22; calcifications, n = 11; both types, n = 7)
Automated computer detection study* (n = 305)	245	18 (calcifications n = 18)	42 (calcifications n = 42)

*No cases with soft tissue lesions were included in the automated computer detection study.

parallel to the detector 1 mm apart, using the standard FBP algorithm. These raw projection images were reconstructed using the EMPIRE reconstruction algorithm on an off-line workstation only for this study, so this process took place well after the acquisition of each case.

Visual grading analysis study population

For the VGA study, 100 patient unilateral mediolateral oblique (MLO) view DBT studies were consecutively selected out of the 374 described above to achieve the desired proportion of patient cases (Table 1): 40 biopsy proven malignant cases; 30 biopsy proven benign cases; and 30 normal cases. The latter were scored as BIRADS® 1 or 2 and had at least one year of negative follow-up. The ground truth location of the lesions was annotated under the supervision of an experienced radiologist (13 years of experience with mammography, three with DBT) with access to pathology and radiology reports.

Automated computer detection study population

For the computer detection study, out of our set of 374 cases, all abnormal cases due to calcifications scored as BI-RADS 3, 4, or 5 cases were selected. Cases with calcifications were used since visibility of this type of lesion has been proposed to be the main advantage of EMPIRE over FBP (24). No cases with soft tissue lesions were included in this study. This yielded 60 DBT patient studies (Table 1). From these, 114 DBT volumes (either MLO, cranio-caudal [CC], or both views) were available. Location of calcifications were annotated individually for each reconstructed volume (independently in EMPIRE and FBP), under the supervision of the same experienced radiologist with access to pathology and radiology reports. A sample of 245 normal patient studies (bilateral, BI-RADS 1 or 2) was also selected for training of the computer detection algorithms.

Visual grading analysis study

An absolute VGA observer study (26) was performed to assess several aspects of image quality in both reconstruction algorithms. It was carried out by four readers specializing in breast imaging (one radiologist, one clinical PhD student, and two physicists specializing in mammography), who had a median of 12 years of experience in breast imaging (range = 3–21 years).

Two reading sessions separated by at least two weeks were performed in order to avoid possible bias in the results due to a direct comparison between reconstruction algorithms of the same patient. Both reconstructions (FBP and EMPIRE) of each patient were alternatively and randomly split between the two reading sessions. In total, 50 FBP volumes and 50 EMPIRE volumes were scored during each session. Scoring was performed on a 5-point scale (1 = poor quality to 5 = excellent quality) on six aspects of normal anatomy (presence of noise and artifacts, visualization of skin line and Cooper's ligaments, contrast, and overall image quality) and, when present, visibility and sharpness of both types of lesions (calcifications and soft tissue). The location of the lesions was outlined for the readers. The reading was performed on an in-house developed workstation (CIRRUS Observer, Diagnostic Image Analysis Group, Nijmegen, the Netherlands) (Fig. 1), using high-resolution mammographic monitors of at least 5 MP.

To account for repeated measures and multiple independent reader variability, the average results were analyzed with generalized estimating equations (GEE) models, using as outcome the scores of each of the questions. The two-way GEE models were built using the reconstruction algorithm and reader as main effects as well as their interaction term. An exchangeable working correlation matrix structure was chosen. Wald 95% confidence intervals (CI) were computed. Differences in the scores between reconstruction algorithms for each reader were tested with the Mann–Whitney U (Wilcoxon) non-parametric test.

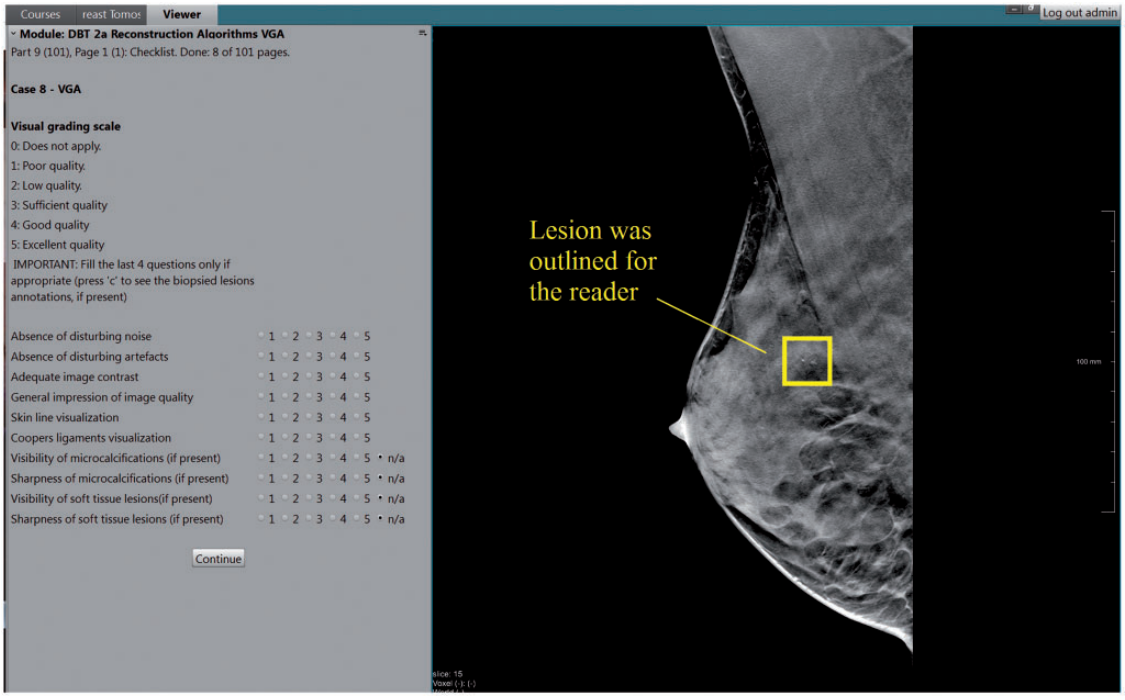

Fig. 1. In-house developed workstation for the scoring of the visual grading analysis reader study. The readers answered ten questions on a 5-point scale (1 = poor quality to 5 = excellent quality) and the lesions were outlined. The workstation automatically registered the results and provided a summary report per reader after each session.

A two-tailed P value < 0.05 was considered to indicate significant difference. All analyses were performed using SPSS (version 24, IBM Inc., Armonk, NY, USA).

Computer automated detection study

A 3D deep-learning based computer detection system based on convolutional neural networks (3D-CNN) was trained, validated, and tested for the task of detecting suspicious calcifications (scored as BI-RADS 3, 4, or 5) in DBT images, using data reconstructed both with EMPIRE and FBP. At the end, the performance of the network trained and evaluated with EMPIRE data was compared to the performance of the network trained and evaluated with FBP data. The 305 DBT patient studies were split into training, validation, and test in a case-level to avoid bias, with the proportions shown in Table 2.

The 3D-CNN used in this study is an extension of the 2D deep-learning approach to detect calcifications in mammography developed by Mordang et al. (31). It was trained to discriminate between 3D DBT patches (size = $29 \times 29 \times 9$ voxels) with and without suspicious calcifications. More details regarding the architecture and training strategy of the 3D-CNN can be found in Appendix 1.

The 3D-CNN with the parameters that achieved the best accuracy on the validation data during the training was then used to compute the receiver operating characteristic curve (ROC) on the test dataset. The partial area under the receiver operating characteristic (ROC) curve (pAUC) for a false-positive rate of 0–0.05 was computed. This range was empirically defined as the range where the largest difference of pAUC between EMPIRE and FBP was found. The pAUC was

Table 2. Number of DBT patient studies, DBT image volumes, and extracted patches used for the training, validation, and testing of the 3D-CNNs.

	Training	Validation	Test
Patients			
Positive	42	9	9
Negative	172	36	37
Volumes			
Positive	79	17	18
Negative	624	124	94
Patches*			
Positive	EMPIRE: 928 FBP: 725	EMPIRE: 201 FBP: 178	EMPIRE: 119 FBP: 86
Negative	EMPIRE: 928 FBP: 725	EMPIRE: 201 FBP: 178	EMPIRE: 47,000 FBP: 39,500

*Differences on a patch level between EMPIRE and FBP reconstruction algorithms are due to different individual calcification annotations between reconstructed volumes.

Table 3. Average scores (1 = poor quality to 5 = excellent quality) of each of the parameters of the visual grading analysis (VGA) for each reconstruction algorithm, obtained with a generalized estimating equations (GEE) model, which accounts for the variability of repeated measures by multiple independent readers.

	FBP	EMPIRE	*P* value*
General image quality			
Absence of disturbing noise	3.09 (3.02–3.15)	3.12 (3.06–3.19)	0.424
Absence of artifacts	2.97 (2.87–3.07)	3.26 (3.15–3.36)	<0.001
Adequate image contrast	3.10 (2.99–3.20)	3.23 (3.12–3.33)	0.010
Overall image quality	3.03 (2.94–3.13)	3.22 (3.12–3.31)	<0.001
Skin line visualization	3.10 (3.02–3.18)	3.11 (3.01–3.20)	0.855
Cooper's ligaments visualization	3.39 (3.32–3.47)	3.47 (3.40–3.54)	0.057
Lesions			
Visibility calcifications	3.37 (3.19–3.55)	3.53 (3.35–3.71)	0.053
Sharpness calcifications	3.02 (2.85–3.16)	3.03 (2.88–3.18)	0.875
Visibility soft tissue	3.77 (3.58–3.96)	3.84 (3.64–4.04)	0.365
Sharpness soft tissue	3.51 (3.33–3.69)	3.52 (3.34–3.70)	0.918

Within parentheses, 95% Wald CIs are shown.
*A two-tailed *P* value < 0.05 was considered to indicate significant difference between reconstruction algorithms.

compared between the 3D-CNN trained with FBP data (3D-CNN-FBP) and the 3D-CNN trained with EMPIRE data (3D-CNN-EMPIRE) after bootstrapping (n = 5000), via the Mann–Whitney U (Wilcoxon) non-parametric test. A two-tailed P value < 0.05 was considered to indicate significant difference. All statistical analyses were performed using MATLAB© R2017a (MathWorks, Natick, MA, USA).

Results

Visual grading analysis study

The averaged results from the GEE model (Table 3) yielded that EMPIRE reconstructions showed slightly better contrast (significant for one reader, the radiologist) and fewer artifacts (significant for all readers). In general, a better overall image quality (significant for three readers) was also assessed for the EMPIRE DBT volumes. No significant difference was found between reconstruction algorithms for the level of noise and the skin line visualization, while Cooper's ligaments were slightly better represented with EMPIRE (significant for one reader, the radiologist). Regarding the lesion representation of both algorithms, on average a better visibility of calcifications was found for EMPIRE. All readers scored EMPIRE higher than FBP for visibility of calcifications (significant for one reader, the clinical PhD student), while no difference was found for soft tissue lesions.

There was significant inter-reader variability in all the scores ($P < 0.001$). Cumulative percentage of the scores of all readers are shown in Fig. 2, which shows

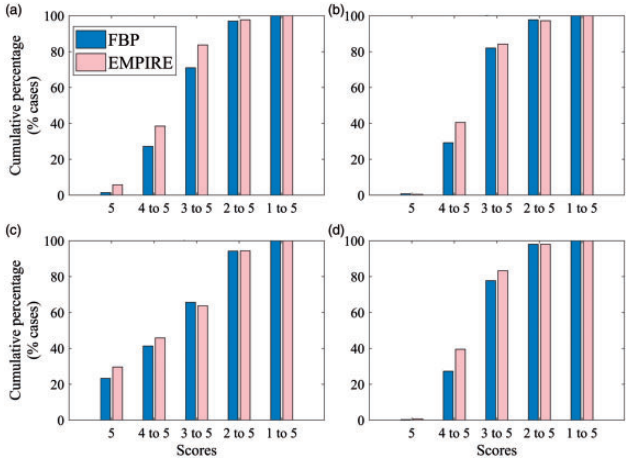

Fig. 2. Cumulative percentages of the scores (1 = poor quality, 5 = excellent quality) across readers for the four most relevant aspects that were found on average better for EMPIRE compared to FBP. (a) Absence of artifacts, (b) Image contrast, (c) Visibility calcifications, (d) Overall image quality.

that EMPIRE achieves higher scores for the four most significant aspects found on the GEE models: presence of artifacts; adequate image contrast; visibility of calcifications; and overall image quality. For these, the results for each reader are also shown in Fig. 3. Two examples of cases that were scored by most readers higher with EMPIRE than with FBP for the visualization of calcifications are shown in Fig. 4. Fig. 5 shows a case with a soft tissue lesion, equally well-visualized in EMPIRE as in FBP. Finally, an example of a case scored by all readers as

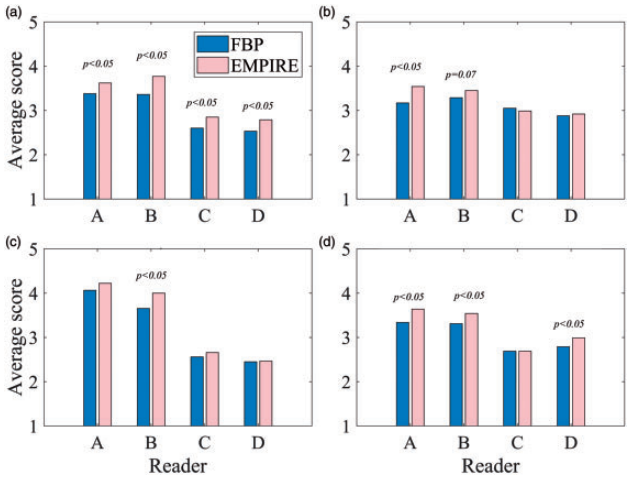

Fig. 3. Average scores per reader (1 = poor quality, 5 = excellent quality) for the four more relevant aspects that were found on average better for EMPIRE in comparison with FBP reconstruction. Differences between reconstruction algorithms for each reader were tested with the Mann–Whitney U (Wilcoxon) non-parametric test. (a) Absence of artefacts, (b) Image contrast, (c) Visibility calcifications, (d) Overall image quality.

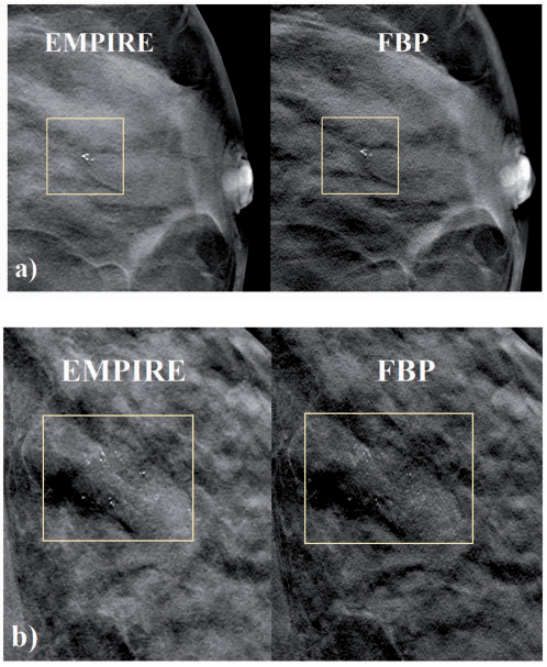

Fig. 4. Example ROIs of two DBT cases containing malignant calcifications (outlined) reconstructed with EMPIRE (left) and standard FBP (right). Three observers scored calcification visibility higher for EMPIRE in case (a), while all four of them scored EMPIRE higher in case (b). These images are displayed with the default window width and level set by the DBT system.

better in EMPIRE regarding artefacts is displayed in Fig. 6.

Computer automated detection study

The ROC curves of the 3D-CNN for FBP and EMPIRE are shown in Fig. 7a. The 3D-CNN-EMPIRE showed similar high performance as the one trained and tested with FBP (AUC-EMPIRE = 0.990 vs. AUC-FBP = 0.986). This is mainly influenced by the operating points at high false-positive rate (FPR, or 1 – Specificity), which have a sensitivity almost equal to 1. However, at low FPRs, we observed that 3D-CNN-EMPIRE performed better than 3D-CNN-FBP. For instance, at FPR = 0.01, 3D-CNN-EMPIRE achieved a sensitivity of 0.958 while 3D-CNN-FBP achieved a sensitivity of 0.845. The partial ROC curve delimited in the range with FPR of 0–0.05 is shown in Fig. 7b. After bootstrapping, the partial AUC (pAUC) of EMPIRE is 0.880 (95% CI = 0.846–0.897), significantly better ($P < 0.001$) than pAUC-FBP = 0.857 (95% CI = 0.815–0.881).

Discussion

The comparison of breast tomosynthesis reconstruction algorithms shows that the new EMPIRE reconstruction improves the image quality of the standard FBP reconstruction on the Siemens Mammomat Inspiration DBT system. The VGA results yielded in average better results for EMPIRE in some of the analyzed aspects of image quality. Also, the 3D-CNN using EMPIRE images achieved higher performance with a better ROC curve, specially at the range of high specificity, relevant for screening.

In general, performing additional iterative processes on the FBP reconstructed volumes appears useful in order to enhance the visualization of DBT images, heavily degraded due to the acquisition limitations of DBT. In particular, we have observed that image contrast can be enhanced and the presence of artifacts reduced. In addition, Cooper's ligaments are slightly better visualized with EMPIRE. Cooper's ligaments are fibrous connective tissue between the inner side of the skin and the pectoral muscles. Usually, changes in their structure yield a high predictive value for malignant mass lesions (32).

Furthermore, skin line visualization was similar among both algorithms. An excellent skin line visualization and sharpness is one of the main reported benefits of FBP in comparison to fully iterative algorithms (17). This remains unchanged with EMPIRE. Assessment of possible breast skin thickening anomalies is of importance since it may be associated with malignancy (33).

As pointed out in preliminary studies (24), it has been confirmed in our study that the new EMPIRE

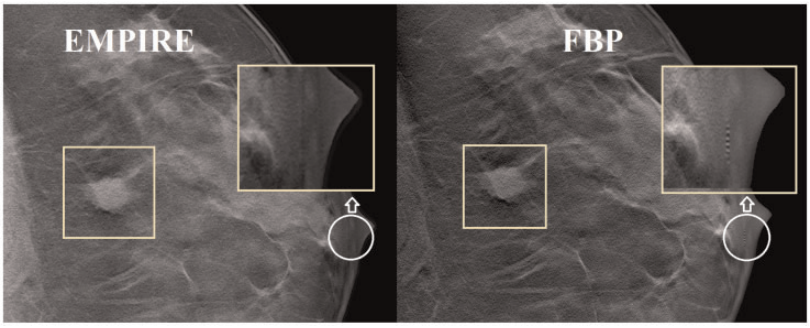

Fig. 5. Example ROIs of a DBT case containing a malignant soft tissue lesion (outlined) reconstructed with EMPIRE (left) and standard FBP (right). Three observers scored soft tissue visibility similar between EMPIRE and FBP (one reader scored EMPIRE higher than FBP). Also note how an artefact nearby the nipple (white circle), due to a calcification in another DBT plane, is visible in FBP but not in EMPIRE. These images are displayed with the default window width and level set by the DBT system.

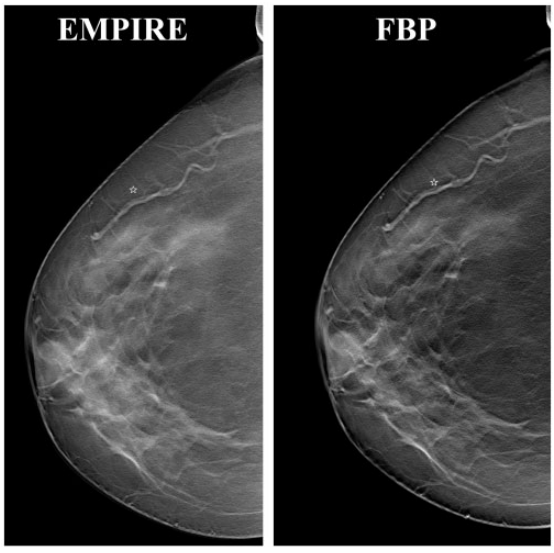

Fig. 6. Example of patient DBT slice reconstructed with EMPIRE (left) and standard FBP (right). All four observers scored the artefacts on the FBP volume worse than on EMPIRE. It can be seen that for tissue near the skin line, EMPIRE provides a better visualization compared with FBP. Also, the large vein on the lateral side of the breast (under the star mark) shows more overshooting artefact (shadow like artefact, 21) in FBP than in EMPIRE. These images are displayed with the default window width and level set by the DBT system.

algorithm significantly improves the visibility of calcifications in the DBT volumes for humans. In addition, we also showed a similar benefit for a deep-learning based computer detection system when it comes to classification of calcifications. The higher contrast of calcifications achieved by EMPIRE, combined with a similar visualization of soft tissue lesions, suggests that EMPIRE might improve the clinical performance of DBT for lesion detection in a clinical setting.

A topic of future work is to study the impact of the EMPIRE algorithm on tests designed for quality control of the reconstructed slices of breast tomosynthesis (13). Moreover, further expansion of the 3D-CNN for EMPIRE is also still required, since here we just used a basic network while, similar techniques can also be applied in order to detect/classify groups of calcifications, as well as other types of lesions.

A limitation of this study is the fact that an actual detection reader study was not performed to account for lesion visibility. In addition, some of the observers were not breast radiologists, but given the non-clinical task of evaluating image quality, we believe this is a minor limitation. Also, the medical physicist observers provided the least number of significantly different assessments between the two reconstruction algorithms in the VGA study. Therefore, any potential bias would be in favor of the FBP algorithm.

It should also be noted that, although images from both algorithms were objectively and independently annotated, not the same calcifications were included for evaluation of the 3D-CNN with EMPIRE and FBP. We observed that more calcifications were annotated in EMPIRE. This might support that calcification visibility for human observers is higher in EMPIRE. As a consequence, this might lead to a bias in favor of FBP, since likely many true-positive calcifications for EMPIRE were labeled as true negatives in FBP, while they could have been considered as false negatives.

In conclusion, the new EMPIRE reconstruction algorithm, in comparison with FBP, provides breast tomosynthesis volumes with better contrast and overall image quality, fewer artifacts, and improved visibility of calcifications according to the human observers, as well as improved detection capability in deep-learning systems. As a consequence, this new algorithm might enhance DBT clinical performance of radiologists and

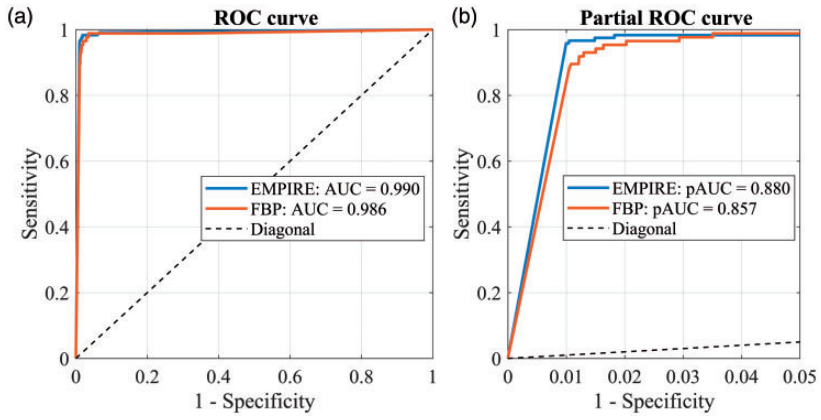

Fig. 7. Complete (a) and partial (b) ROC curves of the same 3D-CNN trained and validated with EMPIRE images and trained and validated with FBP images, for the task of detecting suspicious calcifications in DBT slices.

improve the accuracy of deep-learning based computer detection systems.

Acknowledgments

The authors thank Itsara Wichakam for the help with the calcification annotations on the digital breast tomosynthesis images.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: I Sechopoulos has research and speaker agreements with Siemens Healthineers (Erlangen, Germany), a research agreement with Toshiba Medical (Otawara, Japan), and is a scientific advisor of Fischer Imaging (Broomfield, USA); N Karssemeijer is one of the co-founders of Volpara Health Technologies Solutions (Wellington New Zealand), Qview Medical (Los Altos, CA, USA), and ScreenPoint Medical (Nijmegen, the Netherlands); R Mann has relationships with the following companies: ScreenPoint Medical (Nijmegen, the Netherlands), Seno Medical (San Antonio, USA), Micrima (Bristol, UK), and Transonic Imaging.

Funding

The author(s) received the following financial support for the research, authorship, and/or publication of this article: this work was partly funded by Siemens Healthineers (Erlangen, Germany).

References

1. Tabar L, Gad A, Holmberg L, et al. Reduction in mortality from breast cancer after mass screening with mammography: randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. Lancet 1985;325:829–832.
2. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet 2012;380:1778–1786.
3. Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. Med Phys 2013;40:014301.
4. Andersson I, Ikeda DM, Zackrisson S, et al. Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and BIRADS classification in a population of cancers with subtle mammographic findings. Eur Radiol 2008;18:2817–2825.
5. Skaane P, Gullien R, Bjorndal H, et al. Digital breast tomosynthesis (DBT): initial experience in a clinical setting. Acta Radiol 2012;53:524–529.
6. Ciatto S, Houssami N, Bernardi D, 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.
7. Gennaro G, Hendrick RE, Ruppel P, et al. Performance comparison of single-view digital breast tomosynthesis plus single-view digital mammography with two-view digital mammography. Eur Radiol 2013;23:664–672.
8. Gilbert FJ, Tucker L, Gillan MG, et al. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme–a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. Health Technol Assess 2015;19:i–xxv, 1–136.
9. Lång K, Andersson I, Rosso A, et al. 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.
10. Gilbert FJ, Tucker L, Young KC. Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. Clin Radiol 2016;71:141–150.
11. Skaane P. Breast cancer screening with digital breast tomosynthesis. Breast Cancer 2017;24:32–41.

12. Marshall NW, Bosmans H. Measurements of system sharpness for two digital breast tomosynthesis systems. Phys Med Biol 2012;57:7629–7650.
13. Rodriguez-Ruiz A, Castillo M, Garayoa J, et al. Evaluation of the technical performance of three different commercial digital breast tomosynthesis systems in the clinical environment. Phys Med 2016;32:767–777.
14. Rodríguez-Ruiz A, Castillo M, Garayoa J, et al. Further results on the evaluation of the performance of a digital breast tomosynthesis system in the clinical environment. Phys Med 2016;32:217.
15. Maldera A, De Marco P, Colombo P, et al. Digital breast tomosynthesis: Dose and image quality assessment. Phys Med 2017;33:56–67.
16. Wu G, Mainprize JG, Boone JM, et al. Evaluation of scatter effects on image quality for breast tomosynthesis. Med Phys 2009;36:4425–4432.
17. Sechopoulos I. A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications. Med Phys 2013;40:014302.
18. Ren B, Ruth C, Stein J, et al. Design and performance of the prototype full field breast tomosynthesis system with selenium based flat panel detector. In: Flynn MJ (ed.) *Medical Imaging 2005: Physics of Medical Imaging*. Bellingham, WA: SPIE, 2005, pp.550–561.
19. Mertelmeier T, Orman J, Haerer W, et al. Optimizing filtered backprojection reconstruction for a breast tomosynthesis prototype device. In: Flynn MJ (ed.) *Medical Imaging 2006: Physics of Medical Imaging*. Bellingham, WA: SPIE, 2006: 61420F.
20. Zhang Y, Chan H-P, Sahiner B, et al. Tomosynthesis reconstruction using the simultaneous algebraic reconstruction technique (SART) on breast phantom data. In: Flynn MJ, Hsieh J (eds) *Medical Imaging 2006: Physics of Medical Imaging*. Bellingham, WA: SPIE, 2006. 614249.
21. Sidky EY, Pan X, Reiser IS, et al. Enhanced imaging of microcalcifications in digital breast tomosynthesis through improved image-reconstruction algorithms. Med Phys 2009;36:4920–4932.
22. Jerebko AK, Mertelmeier T. Evaluation and optimization of the maximum-likelihood approach for image reconstruction in digital breast tomosynthesis. In: Samei E, Pelc NJ, editors. Medical Imaging 2010: Physics of Medical Imaging. Bellingham, WA: SPIE, 2010:76220E..
23. Ludwig J, Mertelmeier T, Kunze H, et al. A novel approach for filtered backprojection in tomosynthesis based on filter kernels determined by iterative reconstruction techniques. In: Krupinski EA (ed.) *Digital Mammography*. *IWDM 2008*. *Lecture Notes in Computer Science*. Vol. 5116, Berlin, Heidelberg: Springer, 2008, pp.612–620.
24. Abdurahman S, Dennerlein F, Jerebko A, et al., editors. Optimizing high resolution reconstruction in digital breast tomosynthesis using filtered back projection. In: Fujita H, Hara T, Muramatsu C, editors. Breast Imaging. IWDM. Lecture Notes in Computer Science, vol. 8539. Cham: Springer, 2014:520–527.
25. Abdurahman S, Jerebko A, Mertelmeier T, et al., editors. Out-of-plane artifact reduction in tomosynthesis based on regression modeling and outlier detection. In: Maidment ADA, Bakic PR, Gavenonis S, editors. Breast Imaging. IWDM 2012. Lecture Notes in Computer Science, vol. 7361. Berlin, Heidelberg: Springer, 2012:729–736.
26. Smedby O, Fredrikson M. Visual grading regression: analysing data from visual grading experiments with regression models. Br J Radiol 2010;83:767–775.
27. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–444.
28. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
29. Bassett LW. Mammographic analysis of calcifications. Radiol Clin North Am 1992;30:93–105.
30. Samala RK, Chan H-P, Hadjiiski LM, et al. Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis. In: Tourassi GD, Armato SG, editors. Medical Imaging 2016: Physics of Medical Imaging. Bellingham, WA: SPIE, 2016:97850Y.
31. Mordang J-J, Janssen T, Bria A, et al., editors. Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. In: Tingberg A, Lang K, Timberg P, editors. Breast Imaging. IWDM 2016. Lecture Notes in Computer Science, vol. 9699. Cham: Springer, 2016:35–42.
32. Hong AS, Rosen EL, Soo MS, et al. BI-RADS for sonography: positive and negative predictive values of sonographic features. Am J Roentgenol 2005;184:1260–1265.
33. Pope T Jr, Read M, Medsker T, et al. Breast skin thickness: Normal range and causes of thickening shown on film-screen mammography. J Can Assoc Radiol 1984;35: 365–368.
34. Chollet F. Keras. 2015. Available at: https://github.com/fchollet/keras.