Strategies to reduce DBT reading time

Bringing tomosynthesis to screening

siemens-healthineers.com/mammography

SIEMENS Healthineers

Digital Breast Tomosynthesis in Screening – Approaches to Reduce Reading Time

Ioannis Sechopoulos, PhD

Department of Radiology and Nuclear Medicine, Radboud University Medical Center, Nijmegen, The Netherlands and Dutch Expert Center for Screening (LRCB), Nijmegen, The Netherlands

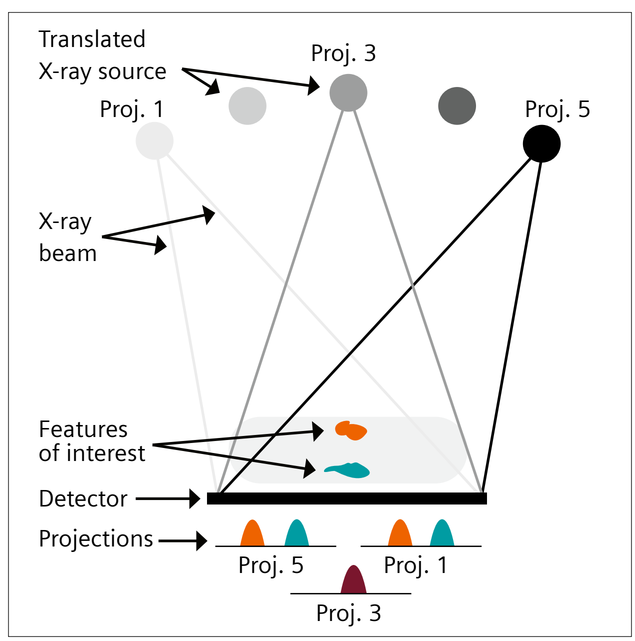
Abstract

The development of digital breast tomosynthesis (DBT) for the detection of breast cancer has resulted in many trials showing that an improvement in detection is possible with DBT. However, these trials have also shown that reading DBT images is considerably slower than reading standard digital mammography (DM) cases. Not surprisingly, it takes longer to read a stack of 50 image slices than one standard mammogram. This increase in reading time for DBT interpretation limits its introduction in large screening programs, such as the regional or national screening programs commonly found in Europe. Therefore, for the better part of this decade, multiple efforts and research have taken place to demonstrate the feasibility of different time-saving strategies when reading DBT exams. As a result, it now seems more feasible than ever that the reading time in DBT could be reduced to match or even be lower than that of DM. For these strategies to be introduced in every-day use, however, some additional studies are needed. Here, we review the strategies proposed up to now to reduce the time required for interpretation of DBT cases in breast cancer screening, and discuss the current limitations in knowledge regarding some of these interpretations.

Introduction

Since digital mammography (DM) is a two-dimensional imaging modality, mammograms of the three-dimensional breast suffer from the phenomenon of tissue superposition. That is, tissues that are separated only vertically in the breast during compression are projected to the same location in the mammogram. This can result in either normal tissues resembling a malignant finding, lowering specificity, or normal tissue masking a real finding, lowering sensitivity. Of course, the higher the proportion of the breast that is composed of dense fibroglandular tissue, the higher the risk of superposition. To ameliorate this effect, currently a mammographic examination, especially for screening for breast cancer, involves the acquisition of two views; the cranio-caudal (CC) and the medio-lateral oblique (MLO) views. However, this is not a perfect solution, since the loss of performance due to this effect, especially in dense breasts, persists.

Digital breast tomosynthesis (DBT) was introduced mostly to reduce this problem of tissue superposition. DBT involves the acquisition of multiple low-dose mammography-like projections from various angles over a limited angular range around the compressed breast (Fig. 1). These projections are then used to reconstruct a pseudo-3D volume depicting the breast tissue distribution [1–3]. This pseudo-3D volume is enough to reduce the impact of tissue superposition despite limited vertical spatial resolution, and results in improved clinical performance compared to DM [4–10].
However, a single DBT image typically consists of a stack of ~50 slices for a breast with a typical thickness under compression of about 50 mm. This increases the amount of information generated by DBT to be reviewed by the interpreting radiologist, which results in a reading time



1 Schematic of a digital breast tomosynthesis acquisition, showing a geometry equal to that used in mammography, but with the X-ray source rotating around the compressed breast, acquiring a projection image at each position. The changing X-ray source position results in different projection images, with the features in the breast changing location in the images depending on their vertical location.

that has been repeatedly reported to be double that of DM [11]. This increase in the demand of radiologist resources is one of the most important challenges needing to be overcome before DBT could be introduced in large population screening programs as a replacement of DM. However, several alternative acquisition and reading strategies may be useful in optimizing the interpretation of DBT-based screening. This would allow for the potential of DBT, and its promise of improved outcomes, to finally be introduced in high-volume screening programs without a substantial increase in the expenditure of healthcare resources. The strategies and alternatives that have been proposed or are being investigated can be divided into two categories: alternative strategies to reduce the number of images that need to be interpreted, and strategies to read DBT faster.

Reduction of images to be read

As mentioned, currently a screening DM examination consists of the acquisition of two views per breast. The main reason for this is the attempt to ameliorate the effects of tissue superposition. Since this effect is, to a great extent, solved by DBT, then perhaps it is feasible to not acquire two views of each breast, and therefore, only acquire MLO DBT views during screening. If this were the case, the MLO view would be the chosen one due to it being the view with the largest tissue coverage.

The Malmö Breast Tomosynthesis Screening Trial (MBTST) involved the comparison of the screening performance of such a DBT acquisition strategy (MLO view-only), to that of two-view DM [5, 6]. In this prospective screening trial involving almost 15,000 cases, the use of single-view DBT resulted in an increase in the cancer detection rate of 34% over that obtained in the two-view DM arm; from 6.5 to 8.7 cancers per 1,000 women screened [6]. This strategy also resulted in an important increase in the recall rate of 44% (from 2.5% to 3.6%). However, the baseline recall rate was very low to begin with, and the DBT review did not include the use of prior images, an effective tool that is known to reduce recall rate substantially [12]. In a retrospective observer study, Rodriguez Ruiz et al. compared the detection performance resulting from interpreting single-view DBT to that of single-view DBT + single-view DM, two-view DBT + two-view DM, and two-view DM only [13]. Although the retrospective, enriched case set nature of this study of course involved fewer cases than that in the MBTST, this trial allowed for the evaluation of multiple acquisition strategies, with all cases of all strategies interpreted by all participating radiologists. The authors did not detect any difference in performance among the four acquisition strategies. Therefore, it seems feasible that single-view DBT could be used for screening for breast cancer. However, both of these studies used the same wide-angle DBT system. Therefore, the generalizability of these results to DBT imaging performed with narrower-angle systems remains to be evaluated.

In European population screening programs, the most common standard is that all cases are double read by two different breast radiologists. Two other prospective trials, as part of their investigation into screening DBT, tested the hypothesis that the reduction in superposition effects with DBT results in images being easier to interpret, and therefore double reading not yielding as large an improvement as with DM. In the STORM trial, Houssami et al. determined that single-reading of DM+DBT still resulted in an increase of over 40% in the cancer detection rate and a 26% reduction in recall rate, compared to double-reading DM alone [14]. An important improvement in performance was also detected by Romero Martin et al. as part of the prospective DBT trial in Cordoba, Spain [15]. In that study, the increase in the cancer detection rate with single-reading of DBT with a synthetic mammogram (a mammogram-like image generated from the DBT data) was over 20% compared to that with DM alone, while recall rate was reduced by over 40%.

With the introduction of AI-based automated systems that seem to be approaching, if not already have matched, human performance in interpreting breast images, both DM and DBT (16,17), it is now feasible to think that an AI system could be used to interpret all images, and that only the ones picked out as being more suspicious would need to be reviewed by a breast radiologist. This concept of triaging of normal cases has been investigated by a number of different research groups, all, for now, on DM images, having found that an important reduction in caseload can be achieved (ranging from 20% to 90%), with no loss in overall performance [18–20]. Given the similarity in the results of studies that have compared the stand-alone performance of such AI systems for DM and DBT, it could be expected that the same performance when using these systems for triaging of normal cases would be achievable. However, before such triaging could be introduced in the screening realm, its impact on large-scale screening programs would have to be evaluated prospectively with real screening prevalence. This is especially important since it could be expected that the radiologists' behavior will be affected when facing a case set that has been through triaging by an AI system. Therefore, prospective clinical trials that gauge this impact are necessary.

Faster reading of images

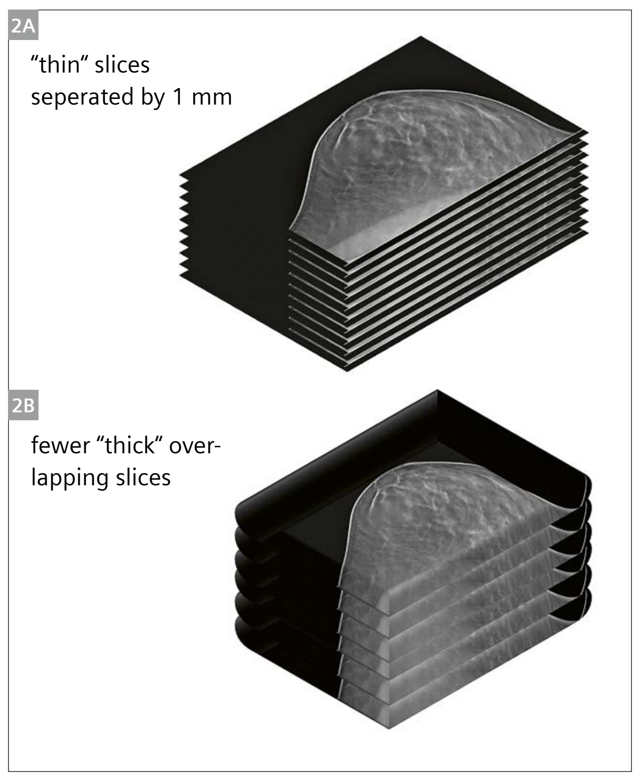
The three strategies discussed above aim to reduce the number of images that are acquired or need interpretation by a breast radiologist. Once this number has been optimized, it would be beneficial to also minimize the time spent in interpreting each of these images. For this, two strategies have been proposed, the use of slabbing, and the synthetic image-driven interpretation of the case.

To understand the motivation for presenting the reconstructed DBT volume as a few slabs instead of many thin slices, it should be noted that the spatial resolution in the vertical direction in DBT is very poor. Given the narrow angles subtended during a complete DBT acquisition (the tube movement of the widest angle DBT system covers an angular range of 50°), the DBT "volume" is actually composed of highly non-isotropic voxels. The signals in the vertical direction included in one voxel can be considerably farther away than the 1 mm often mis-quoted as the slice thickness. In DBT, the slices are reconstructed 1 mm apart from each other, but this does not mean that they are 1 mm thick. In fact, information from 5 mm or more away from the center of the slice may be included in a DBT slice [21]. Therefore, it could be logical that instead of dividing up a typical 50 mm thick breast into fifty 1 mm slices, such a breast could be depicted with considerably fewer, but thicker, slabs. Interpretation of these thicker slabs could be expected to take less time than interpretation of many more thinner slices. However, it should be ensured that all data required to produce a tomosynthesis volume is available in post-processing, meaning the reader can still choose to see the 1mm slices after scrolling through the 8 mm slabs. In a pair of studies evaluating this hypothesis, it was found that using 2 mm thick slabs resulted in a reduction of 20% in the reading time, while still rendering all lesions visible [22, 23]. In another study, Agasthya et al. compared the reading time and performance when radiologists interpreted 8 mm slabs that overlapped by 3 mm to that of interpreting the regular slices (Fig. 2) [24]. The use of the slabbing technique resulted in equal detection performance with a 30% reduction in the reading time.

Another reading strategy that could substantially reduce the reading time per image is using the synthetic mammogram as the primary image for detection, instead of the reconstructed DBT stack. Under such a scenario, the DBT stack would not be reviewed by the radiologist to detect suspicious findings. Rather, the interpreting radiologist would review the synthetic mammogram, and, if any suspicious area is detected, he/she would, if needed, review that area in the DBT stack to determine if that is, indeed, a finding that needs to be recalled, or an innocuous consequence of tissue superposition or other effect on the synthetic mammogram. An early study evaluating the feasibility of such an approach was performed by Murphy et al., finding that although 13% of the cancers included in the study would have been downgraded in suspicion, they still warranted recall, and therefore they would not have been missed [25]. It should be pointed out, however, that this is, as of now, not yet the intended use of the synthetic mammogram, and there are still probably many improvements that are needed in the generation of these images before they can be reliably used as the primary source for detection of actionable findings. However, with the advent of improved algorithms for constructing these synthetic images, probably in the future with AI having a role in this aspect, it can be expected that this could be a viable strategy in the future, especially for the interpretation of "easier" cases.

Conclusion

It can be expected that all or a combination of these time-saving strategies, be they to acquire fewer images, have fewer images be interpreted by breast radiologists thanks to their interpretation by stand-alone AI systems, and/or by reading each image faster, could result in DBT-based screening requiring the same, or fewer, resources as current DM-based screening, while resulting in improved lesion detection performance. For some of these strategies there is still a lot of evidence that needs to be gathered, or algorithms that need improvement, although some of them seem to be closer to implementation. In either case, demands on breast radiologists to reduce the time to make DBT screening in large population programs a reality seems feasible, soon.



2 Comparison between two stacks of images: (**2A**) thin slices separated by 1 mm; (**2B**) thick partially overlapping slabs. For example, combining 8 slices together with an overlap of three slices results in a five-fold reduction in the number of images in a stack.

References

1 Niklason LT, Christian BT, Niklason LE, Kopans DB, Castleberry DE, Opsahl-Ong BH, et al. Digital tomosynthesis in breast imaging. Radiology. 1997 Nov;205(2):399–406.
2 Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. Med Phys. 2013 Jan;40(1).
3 Sechopoulos I. A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications. Med Phys. 2013 Jan;40(1).
4 Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast Cancer Screening Using Tomosynthesis in Combination With Digital Mammography. JAMA. 2014 Jun 25;311(24):2499.
5 Lång 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 Malmö Breast Tomosynthesis Screening Trial, a population-based study. Eur Radiol. 2016;26(1):184–90.
6 Zackrisson S, Lång K, Rosso A, Johnson K, Dustler M, Förnvik D, et al. One-view breast tomosynthesis versus two-view mammography in the Malmö Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. The Lancet Oncology. 2018 Nov;19(11):1493–503.
7 Gilbert F, Tucker L, Gillan M, Willsher P, Cooke J, Duncan K, et al. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme. Health Technol Assess. 2015 Jan 1;19(4).
8 Skaane P, Bandos AI, Niklason LT, Sebuødegård S, Østerås BH, Gullien R, et al. Digital Mammography versus Digital Mammography Plus Tomosynthesis in Breast Cancer Screening: The Oslo Tomosynthesis Screening Trial. Radiology. 2019 Feb 19;182394.
9 Bernardi D, Macaskill P, Pellegrini M, Valentini M, Fantò C, Ostillio 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. The Lancet Oncology. 2016;17(8):1105–13.
10 Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. The Breast. 2013;22(2):101–8.
11 Skaane P, Bandos AI, Gullien 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 Aug;23(8):2061–71.
12 Roelofs AAJ, Karssemeijer N, Wedekind N, Beck C, van Woudenberg S, Snoeren PR, et al. Importance of Comparison of Current and Prior Mammograms in Breast Cancer Screening. Radiology. 2007 Jan 1;242(1):70–7.
13 Rodriguez-Ruiz A, Gubern-Merida A, Imhof-Tas M, Lardenoije S, Wanders A, Andersson I, et al. One-View Digital Breast Tomosynthesis as a Stand-Alone Modality for Breast Cancer Detection: Do We Need More? Eur Radiol. 2018;28(5):1938–48.
14 Houssami N, Macaskill P, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading – Evidence to guide future screening strategies. European Journal of Cancer. 2014 Jul;50(10):1799–807.
15 Romero Martin S, Raya Povedano JL, Cara Garcia M, Santos Romero AL, Pedrosa Garriguet M, Alvarez Benito M. Prospective study aiming to compare 2D mammography and tomosynthesis + synthesized mammography in terms of cancer detection and recall. From double reading of 2D mammography to single reading of tomosynthesis. Eur Radiol. 2018 Jan 2; 28(6):2484-91.y
16 Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. J Natl Cancer Inst. 2019 Sep 1;111(9):916–22.
17 Conant EF, Toledano AY, Periaswamy S, Fotin SV, Go J, Boatsman JE, et al. Improving Accuracy and Efficiency with Concurrent Use of Artificial Intelligence for Digital Breast Tomosynthesis. Radiology: Artificial Intelligence. 2019 Jul 1;1(4):e180096.
18 Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Teuwen J, Broeders M, Gennaro G, et al. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. Eur Radiol. 2019 Sep 1;29(9):4825–32.
19 Yala A, Schuster T, Miles R, Barzilay R, Lehman C. A Deep Learning Model to Triage Screening Mammograms: A Simulation Study. Radiology. 2019 Aug 6;293(1):38–46.
20 Kyono T, Gilbert FJ, van der Schaar M. Improving Workflow Efficiency for Mammography Using Machine Learning. Journal of the American College of Radiology. 2019 May 30 17(1PA):56-63.21 Rodríguez-Ruiz A, Castillo M, Garayoa J, Chevalier M. Evaluation of the technical performance of three different commercial digital breast tomosynthesis systems in the clinical environment. Physica Medica. 2016 Jun 1;32(6):767–77.
21 Rodríguez-Ruiz A, Castillo M, Garayoa J, Chevalier M. Evaluation of the technical performance of three different commercial digital breast tomosynthesis systems in the clinical environment. Physica Medica. 2016 Jun 1;32(6):767–77.
22 Dustler M, Andersson M, Förnvik D, Timberg P, Tingberg A. A study of the feasibility of using slabbing to reduce tomosynthesis review time. Proc SPIE. 2013;8673:86731L-86731L – 6.
23 Petersson H, Dustler M, Tingberg A, Timberg P. Evaluation of the possibility to use thick slabs of reconstructed outer breast tomosynthesis slice images. Proc of SPIE 2016 ;9787:97871M.
24 Agasthya GA, D'Orsi CJ, Holbrook A, Ho C, Piraner M, Newell M, et al. Reduction in digital breast tomosynthesis interpretation time by slabbing of the reconstructed slices. In: European Congress of Radiology. Vienna, Austria; 2016.
25 Murphy MC, Coffey L, O'Neill AC, Quinn C, Prichard R, McNally S. Can the synthetic C view images be used in isolation for diagnosing breast malignancy without reviewing the entire digital breast tomosynthesis data set? Irish Journal of Medical Science. 2018 9 2018 Nov;19(11):1493-1503.

Contact

Ioannis Sechopoulos, PhD
Associate Professor
Department of Radiology and Nuclear Medicine Radboud University Medical Center
P.O. Box 9101, Route 766
6500 HB Nijmegen
The Netherlands, and
Dutch Expert Center for Screening (LRCB)
Nijmegen, The Netherlands
ioannis.sechopoulos@radboudumc.nl

Artificial Intelligence to Help Radiologists in the Early Detection of Breast Cancer with Mammography and Breast Tomosynthesis

Alejandro Rodríguez-Ruiz, PhD and Nico Karssemeijer, PhD

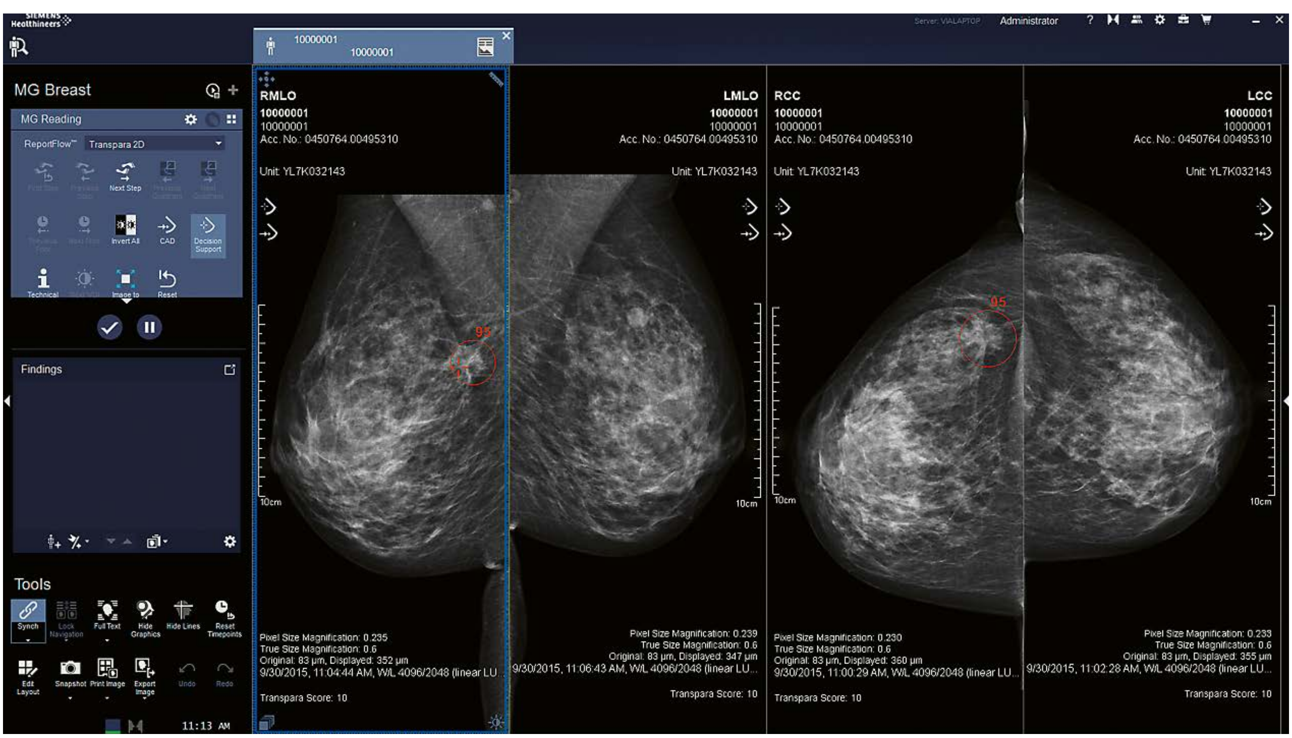
ScreenPoint Medical BV, Nijmegen, The Netherlands

Siemens Healthineers and ScreenPoint Medical are partners committed to developing artificial intelligence based applications for breast imaging. This collaborative arrangement also includes the acquisition of a strategic minority stake in ScreenPoint Medical by Siemens Healthineers.

From CAD to AI systems

Since the 1990s, computer-aided detection (CAD) systems have been developed to automatically detect and mark suspicious breast lesions in mammograms, aiming to prevent overlooking of cancers especially in screening programs. Unfortunately, despite the wide implementation of these systems in clinical practice, no studies to date have found that mammography screening cost-effectiveness improves when radiologists use CAD systems [1]. This could be ascribed to two main limitations of these traditional systems: their low specificity (high false positive rate), and their simplistic radiologist-computer interaction by simply displaying marks [2]. Consequently, such low specificity also precludes the use of traditional CAD as a stand-alone reader for screening mammography.

However, the era of traditional CAD as the only possibility to support radiologists reading mammograms could be coming to an end, due to the rise of a new type of systems based on high-accuracy artificial intelligence (AI) algorithms. The success of novel machine learning algorithms based on deep learning convolutional neural networks is rapidly elevating the field of AI for medical imaging [3]. For mammography, AI systems hold the promise to succeed where traditional CAD failed [4, 5].



1 Example of the Transpara® user interface in *syngo*.via (Siemens Healthineers) featuring decision support (circled area in mammogram with likelihood of cancer score, in this case 95, for an area which after biopsy was confirmed as an invasive ductal carcinoma) and the Transpara® Score (bottom of the image viewport).

In recent years, several deep learning-based algorithms for automated analysis of mammograms have been investigated, some of which have already shown very promising stand-alone detection results in experimental scenarios [6, 7]. The high-performance level of these new AI algorithms can allow the development of systems that can provide radiologists with an enhanced level of support, not simply displaying marks, but that can go deeper into diagnostic decisions such as determining the risk of a lesion representing cancer or confidently determining which screening exams do not contain any suspicious abnormality.

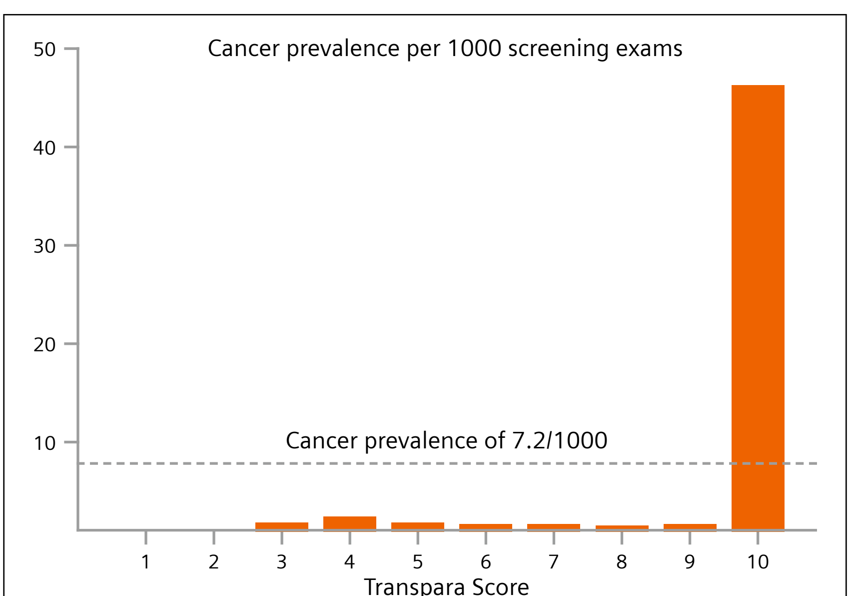
Improving reader accuracy – with focus on lowering the number of mammographically-detected cancers missed at screening – and reducing workload without compromising quality are the aims of most of the latest breast imaging AI systems. In this paper, we summarize the initial clinical evidence conducted with one of the first developed AI systems for mammography and breast tomosynthesis: Transpara®.

What is Transpara® AI? Features and evidence-based validated performance

Transpara® (ScreenPoint Medical BV, Nijmegen, The Netherlands) is a deep learning-driven AI system developed after decades of research in breast imaging and automated detection of lesions in mammograms at the Radboud University in Nijmegen. This AI system is FDA cleared for 2D and CE marked for 2D mammography (DM) and breast tomosynthesis (DBT), and can be used in different reading workstations, as shown in one example in Figure 1. This AI system was designed to aid radiologists reading mammograms, by exploiting the latest developments in deep learning algorithms in combination with deep knowledge of mammography imaging physics and radiological patterns of breast cancer.

The AI system automatically detects breast cancer lesions in DM and in DBT exams from most mammography vendors [6, 8]. It has been trained with millions of examples of breast cancer, benign abnormalities, and normal tissue, all validated by biopsy results or follow-up exams. These images originate from a large multi-center and multi-vendor database representing a wide variation of techniques encountered in mammography practices. The results of the computations are presented to the user in two different features:

- **Interactive decision support:** during reading, users can query a mammographic region using a pointer. As a response, Transpara® provides a region-specific level of suspicion (range 1–100, 100 meaning the highest suspicion for malignancy) as a second opinion. Additionally, suspicious regions-of-interest can be also automatically marked to reduce potential oversight errors with significantly less false positives than traditional CAD systems.
- **Exam-based Transpara® score:** based on all the individual findings, each exam receives a score ranging from 1 to 10, depicting the increasing risk that the exam contains cancer. The screening mammograms are equally divided across score categories (10% in each), meaning that cancer prevalence is much higher in category 10 than in the rest (see Figure 2). If no potential abnormalities are found, a low score is assigned. The highest scores are assigned to exams with suspicious findings. Exam-scores are possible given the high performance of deep learning algorithms, and where not available with traditional CAD systems.



2 Distribution of the Transpara® Score (version 1.6.0) in a consecutive screening population of 12,245 screening 2D mammograms acquired with a Siemens Healthineers MAMMOMAT Revelation with 88 screen-detected cancers (cancer detection rate = 7.2/1000). In a screening setting, the AI system places 10% of the screening exams on each category, but the cancer prevalence is significantly higher in the highest category 10.

According to two recent comprehensive studies using multi-center independent data [9, 10], the stand-alone breast cancer detection of the AI system (versions 1.3.0 and 1.4.0) has been demonstrated to be as good as that of radiologists.

This independent evaluation data originated from eleven sites across the USA and Europe, adding up to around 3,000 exams with over 700 biopsy-proven mammograms with cancer. The mammograms were acquired with devices of four different mammography vendors (Siemens Healthineers, Hologic, Philips, General Electric). Each exam was read by several radiologists, where in a total of 115 radiologists were included in this study. As a result, the AI system stand-alone interpretation of mammograms was compared to more than 30,000 radiologists' interpretations.

The breast cancer detection performance of Transpara® in DM was compared to the performance of radiologists in terms of area under the receiver operating characteristic curve (AUC) using a predefined non-inferiority margin of 0.05. In the first study, the AUC of AI was non-inferior to the average AUC of 101 radiologists (0.841 vs. 0.814, AI had 0.027 higher AUC, 95% CI of AUC difference = [-0.003,0.055]). Similarly, In the second study, the AUC of AI was non-inferior to the average AUC of 14 radiologists (0.887 vs. 0.866, AI had 0.021 higher AUC, 95% CI of AUC difference = [-0.021,0.063]). Interestingly, the AI system achieved a similar sensitivity as humans but at a higher specificity, emphasizing its potential use to discriminate normal cases as good as the best radiologists.

How does AI impact radiologists' performance in 2D mammography and DBT

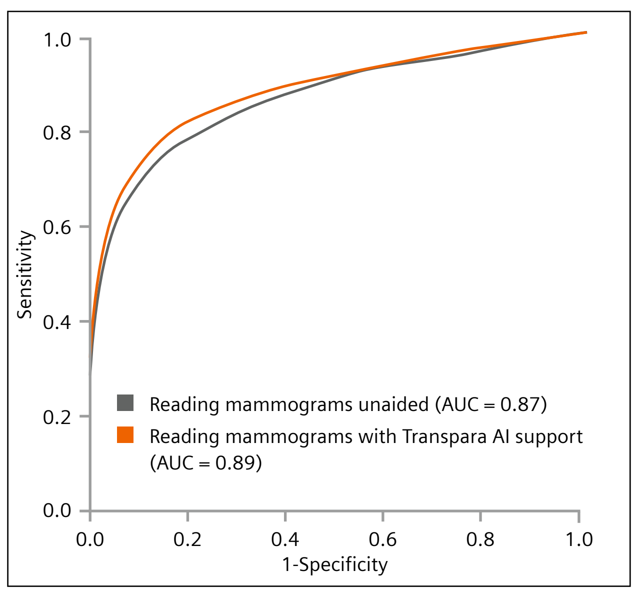
In 2018, a study published in *Radiology* [9] showed that Transpara® is the first AI-based software designed to assist radiologists reading mammograms that makes them more accurate without slowing them down.

In this fully-crossed multi-center retrospective reader study, a sample of 240 screening mammograms (of which 100 where screen-detected cancer, and 40 false positive recalls) were interpreted by 14 radiologists in the USA, once with and once without Transpara® AI (version 1.3.0) in two distinct sessions. For each mammogram, the radiologists provided a forced Breast Imaging Reporting and Data System (BI-RADS) score and a level of suspicion (1–100, 100 meaning high suspicion of cancer). When reading with AI support, radiologists could benefit of all the features of the device as indicated above. The mammograms were from two different vendors (Siemens Healthineers MAMMOMAT Inspiration and Hologic Selenia Dimensions), and radiologists had on average 10 years of experience with breast cancer screening.

The impact of concurrent use of AI in radiologists' performance was analyzed in terms of accuracy (measured via AUC of the radiologists), specificity, sensitivity, and average reading time per mammogram.

On average, the radiologists' AUC was higher with AI support than with unaided reading (0.89 vs. 0.87, respectively; statistically significant, P = 0.002). For some radiologists, the improvement was up to 5% in terms of AUC. As seen in Figure 3, the increased performance was observed in the middle part of the ROC curve, suggesting that the AI system improves the evaluation of equivocal cases, where a second opinion is needed the most. In terms of recalls using the BI-RADS scoring, sensitivity increased with AI support (86% vs. 83% P = 0.046), whereas specificity trended toward improvement (79% vs. 77%, P = 0.06). The improvement in AUC was observed independently in all sub-analysis by lesion type, breast density, and mammography vendor. Another very important finding was that all radiologists trended to improve their accuracy with AI support, regardless of their experience, reducing the inter-reader variability (Fig. 4).

For the second endpoint of the study, reading time per screening mammogram, remained similar when using AI (3 seconds difference, 2%, P = 0.15). This was not the case when using traditional CAD systems, where reading time was higher [11].



3 Average receiver operating characteristic curves of the 14 radiologists reading 2D mammograms unaided, and with support from Transpara® AI. The area under the curve (AUC) is reported within parentheses.
Adapted from the publication in Radiology [9].

Recently, a new observer study with Transpara® shows that radiologists also improve their cancer detection in DBT exams when using AI for support while simultaneously reading time is reduced.

The DBT study was performed by 9 radiologists (4–23 years of experience) who read DBT exams with synthetized 2D mammograms (Insight 2D) acquired with a Siemens Healthineers MAMMOMAT Inspiration device. Radiologists improved their accuracy performance in DBT when concurrently using AI (AUC + 0.041, P= 0.001, from 0.820 to 0.861), while reading time was on average -20% lower when using the system, down to approximately 30 seconds per DBT volume. Finally, the stand-alone performance of the AI system in DBT images was also found to be comparable to that of an average radiologist (95% CI of the difference = -0.038, 0.078). These findings are in line with the results in DM for the same system and with other results in literature for DBT [12].

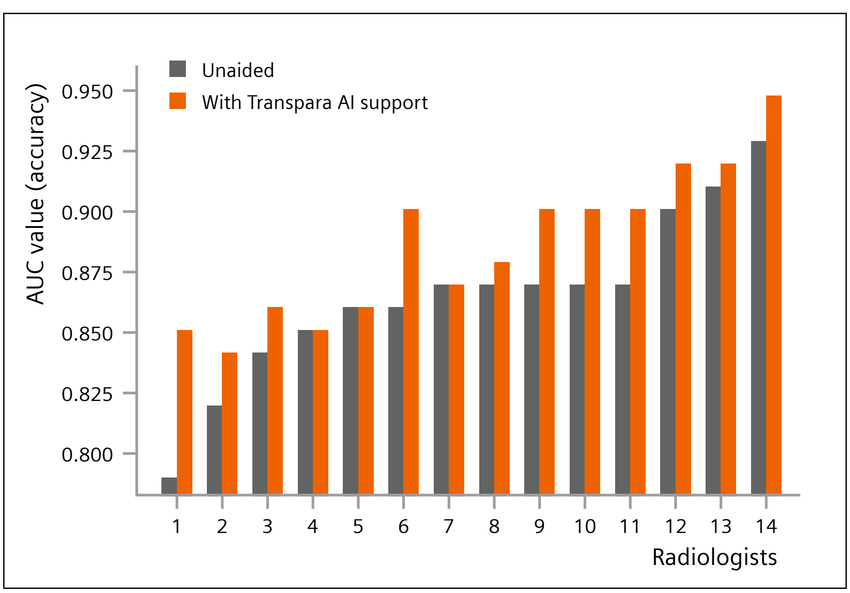
How can AI optimize the efficiency of screening programs?

Early studies indicate that AI can improve radiologists' performance reading mammograms. In a screening setting, the concurrent use of AI has therefore the potential to positively impact in terms of more homogeneous reading performance across sites, reduction in false negative and false positive assessments.

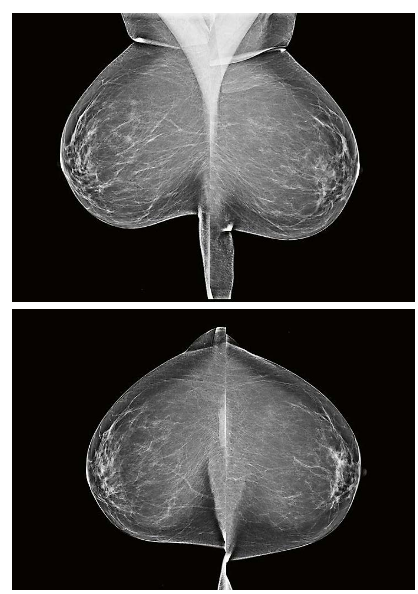
But beyond its concurrent use, given the radiologist-comparable stand-alone accuracy detecting breast cancer in mammograms, AI could potentially be used as an effective independent reader of the screening process, or as a triaging tool for screening mammograms [13].

Given that more than 95% of screening exams do not contain any abnormality, it could be hypothesized that AI can filter out a large proportion of these normal exams automatically. Preliminary studies indicate that current AI could confidently label up to 50% of screening mammograms, with little error (2%–7%) [13, 14] (see Figure 5 for an example). In screening settings where double reading of screening mammograms is performed, this 50% of screening mammograms could, for example, undergo single reading instead of double reading. Having one reader in the loop for these exams could ensure that the cancers in the group are not automatically missed, and also potentially reduce the false positive recalls of the program: overall improving the positive predictive value. This is suggested by study published in *European Radiology* in 2019 [13], where it was observed that the decrease in sensitivity when not reading those mammograms with lowest AI scores is amply compensated by the increase in specificity, because less false positive assessments would be done (AUC reading only cases with scores 6–10 was non-inferior with non-inferiority margin 0.05 to the AUC reading all cases).

Triaging screening exams with AI could allow readers in some settings to focus on the cases with higher cancer prevalence (see Figure 2) when they are more attentive,



4 Area under the receiver operating characteristic curve (AUC) of each individual radiologists, reading 2D mammograms unaided and with Transpara® AI support [9].



5 Example of a highly likely normal screening mammogram, labelled as Transpara® AI with an exam score of 1, which was confirmed normal with two years follow-up round.

potentially also reducing the time that takes to recall women for diagnostic work-up. When considering the introduction of DBT for screening, using AI becomes more important to reduce workload given the increased reading time with DBT with respect to DM (up to twice as long) [15, 16].

Conclusion

Scientific studies are beginning to show convincing evidence that new generation breast AI systems can reach human-like performance and enhance the ability of radiologists to accurately detect breast cancer. In contrast to traditional CAD, these systems can be concurrently used and hold a great potential to reduce screening workload by acting as second reader, or by automatically labelling a large number of normal examinations with high negative predictive value. It is expected that with the continuous development in the field of AI some systems will soon begin to outperform most radiologists in a routine task such as mammography screening. This will enable more cost-effective screening scenarios in which the role of the human reader will change significantly. Before implementation, new screening methods involving AI should be validated thoroughly, while QA procedures for AI products have to be implemented to ensure safety and reliability of breast screening.

References

1 Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL, et al. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. JAMA Unrestricted Med. 2015;175(11):1828-37.
2 Nishikawa RM, Bae KT. Importance of Better Human-Computer Interaction in the Era of Deep Learning: Mammography Computer-Aided Diagnosis as a Use Case. Journal of the American College of Radiology. 2018;15(1):49-52.
3 Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Medical Image Analysis. 2017;42:60-88.
4 Bahl M. Detecting Breast Cancers with Mammography: Will AI Succeed Where Traditional CAD Failed? : Radiological Society of North America; 2018.
5 Trister AD, Buist DS, Lee CI. Will machine learning tip the balance in breast cancer screening? JAMA oncology. 2017;3(11):1463-4.
6 Kooi T, Litjens G, van Ginneken B, Gubern-Merida A, Sanchez CI, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal. 2017;35:303-12.
7 Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. Investigative radiology. 2017;52(7):434-40.
8 Mordang J-J, Janssen T, Bria A, Kooi T, Gubern-Mérida A, Karssemeijer N, editors. Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. International Workshop on Digital Mammography; 2016: Springer, Cham.
9 Rodríguez-Ruiz A, Krupinski E, Mordang J-J, Schilling K, Heywang-Köbrunner SH, Sechopoulos I, et al. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. Radiology. 2018:181371.
10 Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. JNCI. 2019;111(9).
11 Astley S, Gilbert FJ. Computer-aided detection in mammography. Clinical radiology. 2004;59(5):390-9.
12 Conant EF, Toledano AY, Periaswamy S, Fotin SV, Go J, Boatsman JE, et al. Improving Accuracy and Efficiency with Concurrent Use of Artificial Intelligence for Digital Breast Tomosynthesis. Radiology: Artificial Intelligence. 2019;1(4):e180096.
13 Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Teuwen J, Broeders M, Gennaro G, et al. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. European Radiology. 2019;29(9):4825-4832.
14 Yala A, Schuster T, Miles R, Barzilay R, Lehman C. A deep learning model to triage screening mammograms: a simulation study. Radiology. 2019:182908.
15 Zackrisson S, Lång K, Rosso A, Johnson K, Dustler M, Förnvik D, et al. One-view breast tomosynthesis versus two-view mammography in the Malmö Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. The Lancet Oncology. 2018;19(11):1493-503.
16 Rodriguez-Ruiz A, Gubern-Merida A, Imhof-Tas M, Lardenoije S, Wanders AJT, Andersson I, et al. One-view digital breast tomosynthesis as a stand-alone modality for breast cancer detection: do we need more? European Radiology. 2017;28(5):1938-1948.

Professor
Nico Karssemeijer

Alejandro
Rodríguez-Ruiz

Contact

Alejandro Rodríguez-Ruiz, PhD
ScreenPoint Medical BV
Toernooiveld 300
6525 EC Nijmegen
The Netherlands
Tel. +31 24 202 0020
alejandro.rodriguezruiz@screenpointmed.com

Artificial intelligence provides decision support at the press of a button

Interview with Ritse Mann

Radboud University Medical Center, Nijmegen, The Netherlands

"It's like having an additional colleague at the press of a button."

Ritse Mann, MD, PhD
Radiologist
Radboud University Medical Center, Nijmegen,
The Netherlands

In breast cancer screening programs, such as in the program in the Netherlands, a high volume of mammography data is acquired. Radiologists have to evaluate hundreds of images every day with precision and often under time pressure. In collaboration with ScreenPoint Medical, Siemens Healthineers offers radiologists smart support with the aid of artificial intelligence (AI). "It's like having an additional colleague at the press of a button," says radiologist Ritse Mann of Radboud University Medical Center (Radboudumc).

The radiologists of the Mammapolyclinic at Radboud University Medical Center evaluate a steady stream of images every week. In doing so, they rely on solutions from Siemens Healthineers. This choice is made very deliberately, Ritse Mann explains. "Our scientific collaboration with Siemens Healthineers enables us to keep innovating."

Tomosynthesis

For diagnostic imaging, the Mammapolyclinic at Radboud umc uses tomosynthesis, in which 3D images are acquired of the breast. This technique provides a higher depth resolution (or "higher diagnostic accuracy") than conventional 2D mammograms, but it also generates a greater number of images for evaluation. Mann: "As opposed to one mammogram, you have sixty tomosynthesis slices. In practice, that means that we need twice as much reading time to evaluate a 'tomo'."

Interactive support

The "extra time" is partly made up for using Transpara® (ScreenPoint Medical) an application for interactive decision support fully integrated into the *syngo*.Breast Care reading solution offered by Siemens Healthineers. Transpara screens the images like a virtual radiologist, which enables human radiologists to make their evaluations faster – 15 to 20 percent faster, as various studies show.[1] As a result, radiologists have more time for more complex cases.

Decision support

One of the functions of Transpara software is Decision Support or Region Analysis, which enables radiologists to make more precise evaluations of lesions and calcifications, while helping to reduce the number of false positives. Transpara works interactively. If a radiologist sees an anomaly in a mammogram or tomosynthesis, he or she can click on the suspicious region. The software then shows, based on a score of 1 to 95, how high the chance is that a malignancy is present. Mann: “If a spot scores low, I know that it is likely to be a benign anomaly and I can ignore it.”

Decisive in gray areas

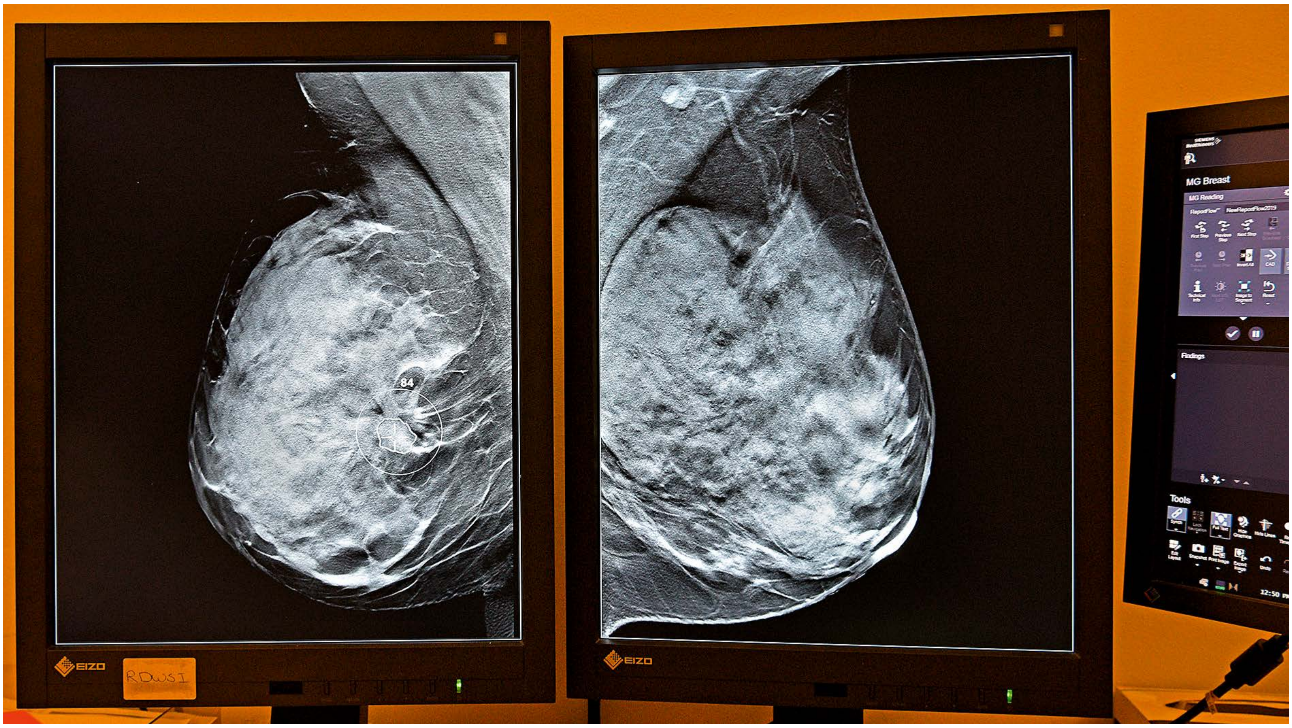
Mann: “You can roughly group the images that we evaluate into three categories: those that clearly show cancer, those that clearly do not show cancer, and the category where I cannot say for certain what I am seeing. Is it a lesion? Do I need to do something about it?” Particularly in these gray areas, Transpara has great value, says Mann. “In this category, it is nice to have a colleague take a look, for an extra check of the images. We have a relatively small clinic, so I do not always have a breast radiologist sitting next to me. But I always have Transpara at my disposal. At the press of a button, Transpara gives me a second opinion on what I think I saw – or what I maybe saw only in my head.” Mann is convinced of the reliability of the AI software. “The performance matches that of a good radiologist.”

Exam score

Transpara is not just like an additional radiologist performing a second check in the hospital. With the score feature, mammograms are evaluated for the chance of malignancies even before the radiologist looks at them. Scores of 1 to 5 mean a very low risk, while a 10 signals the highest chance of a malignant anomaly. Mann: “Let’s say the Exam score is 3. Then I know pretty much for certain: the chance of there being a visible cancer is virtually zero. I do take a look at the images, but it doesn’t yield much.” This is why Transpara can help provide better patient care, although the impacts, according to Mann, are visible particularly across the board, mainly for the imaging institute, and not so much at the individual level.

Especially at this score level, Mann sees a great potential for time savings, because in this case Transpara is the first “radiologist” looking at the images, and with a very low score a second reader - which is required in many breast cancer screening programs - could become superfluous. Thus, it is a tool that appears to have great potential, above all, for screening. “Because screening is becoming more precise, fewer women have to come back to the hospital for further analysis while the same sensitivity is maintained.”

1 Breast screening at Radboud University Medical Center in Nijmegen, The Netherlands.



2 Interactive decision support is fully integrated into the reading solution.

Artificial intelligence

Artificial intelligence is a relatively new field in breast care, with few companies active in it so far. Mann: "Transpara is one of the few genuine AI applications in the market. Many other technologies are still in the research phase and are not yet available in practice." This is despite the fact that AI can play a very big role in breast cancer care. "Especially in screening, an application like Transpara can reduce costs. For example, if Transpara indicates that a woman has a low risk of breast cancer, then you have to wonder whether you still need an actual radiologist for the second check. The situation now is that two radiologists look at every mammogram no matter what. So, there is a lot of potential to save on staff." Nevertheless, Mann realizes that many steps must be taken before that. "From an ethical standpoint alone, it is still hard to say to a woman: the computer has evaluated the images and confirmed that there is nothing to worry about. The patient expects a 'real' doctor to look at the image. At least for now. But AI learns extremely fast. The more images we can use to let the system fine-tune itself, the higher the quality of the system and the care will be. Incidentally, it does raise new questions from the standpoint of privacy. If we can find answers to them, the quality and efficiency of breast diagnostics will make even greater leaps forward."

References:

[1] A. Rodriguez-Ruiz, K. Lång, A. Gubern-Merida, M. Broeders, G. Gennaro, P. Clauser, T. Helbich, m. Chevalier, T. Tan, T. Mertelmeier, W. Wallis, I. Andersson, S. Zackrisson, R. Mann, I. Sechopoulos. **Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists.** Journal of the National Cancer Institute 2019;111(9):djy222. https://doi.org/10.1093/jnci/djy222
Conclusion: Transpara® achieved a cancer detection accuracy comparable to an average breast radiologist in this retrospective setting.

[2] A. Rodriguez-Ruiz, E. Krupinski, J. Mordang, K. Schilling, S. Heywang-Kobrunner, I. Sechopoulos, R. Mann. **Detection of breast cancer using mammography: Impact of an Artificial Intelligence support system.** Radiology 2019;290 (2), 305-314. https://doi.org/10.1148/radiol.2018181371
Conclusion: Radiologists improved their cancer detection performance at mammography when using Transpara® for support, without requiring additional reading time.

[3] A. Rodriguez-Ruiz, K. Lång, A. Gubern-Merida, J. Teuwen, M. Broeders, G. Gennaro, P. Clauser, T. Helbich, M. Chevalier, T. Tan, T. Mertelmeier, W. Wallis, I. Andersson, S. Zackrisson, I. Sechopoulos, R. Mann. **Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study.** European Radiology 2019;29:4825. https://doi.org/10.1007/s00330-019-06186-9R
Conclusion: It is possible to automatically pre-select exams using Transpara® to significantly reduce the breast cancer screening reading workload without affecting screening outcomes.

[4] A Rodríguez-Ruiz, K. Lång, A. Gubern-Mérida, M. Broeders, G. Gennaro,
P. Clauser, T. Helbich, T. Mertelmeier, M. Chevalier, T. Tan, M. G. Wallis, I. Andersson, S. Zackrisson, R. M. Mann, I. Sechopoulos. **Using AI as a pre-screening tool to replace double with single reading for likely normal mammography cases. A simulation study on the impact on sensitivity, specificity, and workload.**
Conclusion: In a simulation study, replacing double reading by single reading for mammograms labeled as most likely normal by Transpara® shows potential to reduce workload in screening, with minimal effect in sensitivity and a moderate increase in specificity.

[5] M Halling Brown, A Rodriguez-Ruiz, N Karssemeijer, M G Wallis, K C Young. **Artificial Intelligence for breast cancer detection in mammographic screening: does it detect the cancers that matter? And can it detect cancers earlier?**
Conclusion: Transpara® has the highest sensitivity for high grade and invasive cancers. It has potential to detect high grade cancers 3 years earlier. Additionally, It could be used to discriminate 50 of the screening population as being almost certainly normal with less than <1% error.

The clinical experts receive financial support from Siemens Healthineers for collaborations.

Siemens Healthineers Headquarters
Siemens Healthcare GmbH
Henkestr. 127
91052 Erlangen, Germany
Phone: +49 9131 84-0
siemens-healthineers.com

Published by Siemens Healthcare GmbH · online · 9428 1020 · ©Siemens Healthcare GmbH, 2020