Using Artificial Intelligence to Interpret Digital Breast Tomosynthesis – current performance and future perspectives

By Dr. Michal Rosen-Zvi, Yoel Shoshan & Dr. Lisa Mullen

Over the past few years, there has been a significant increase in the development of artificial intelligence (AI) technologies applied to cancer detection in breast imaging, as evidenced by the ever-increasing number of publications in the field. A recent study showed that in 2020 alone, there were more than 400 peer-reviewed articles on the subject as indexed in PubMed or available in arXiv [1]. This number of papers is double the number published in 2018. Since 2D mammography is the imaging modality that has been the most-widely used in breast cancer screening in recent years, most of the published papers deal with AI technologies whose input data are in the form of 2D mammography images. Using image analysis, the AI-derived technology has been successfully applied to identify and separate normal cases from cancerous cases. Some of the studies have complemented the image data with additional information such as the patient age or other relevant clinical data. The use-cases considered in the literature vary from cancer risk assessment [2] to so-called safety-nets for radiologists that are aimed at detecting any cancers that the radiologist may have missed [3], and more. A number of studies have shown that when AI-developed technologies are applied to the analysis of mammography images, the performance level in terms of cancer detection is the same as that as that of human radiologists (see e.g., Chorev *et al.* [4] and references therein).

AN AI SYSTEM TO FILTER OUT CANCER-FREE DBT EXAMINATIONS

More recently, digital breast tomosynthesis (DBT), a newer mammography technology for breast imaging and breast cancer screening, has been shown to overcome some of the limitations of 2D mammograms and to improve the number of cancers detected (e.g., Rahman *et al.*[5]).

The Authors

Michal Rosen-Zvi, Ph.D [1], Yoel Shoshan B.Sc.[1] and Lisa Mullen, M.D.[2]

1. AI in Healthcare Department,
IBM Research – Haifa, Mount Carmel, Israel

2. Assistant Professor, Breast Imaging Fellowship Director
Breast Imaging Division, Johns Hopkins Medicine
Baltimore, MD 21287, USA

Communicating Author
Dr. Michal Rosen-Zvi - Email: ROSEN@il.ibm.com

In the past twelve months two articles describing studies aimed at providing support to radiologists in reading DBT images by applying powerful AI-based technology to the analysis of DBT images and clinical data, have been published in the journal Radiology [6, 7]. It should be noted that datasets containing DBT images are much larger than those containing 2D mammography images, just as, in turn, datasets of 2D mammography images are much larger than those containing simple snapshots of typical subjects such as cat or dog pictures taken by regular cameras. Correspondingly, the application of AI algorithms to tomosynthesis data is very challenging and requires vast amounts of storage space, computer power, and GPU memory. It also involves the application of distributed training, fused layers, and other methods to reduce GPU memory consumption and accelerate the training process.

Both studies mentioned above involved data from about 10,000 women. In Shoshan *et al* [6] the data were collected retrospectively from women who had at least one breast tomography exam at 1 of 22 imaging sites in the USA. In the paper from Raya-Povedano *et al* [7], the data were collected retrospectively from the Córdoba Tomosynthesis Screening Trial, which was a prospective screening trial that collected consecutive examinations in women who were screened with both two-view DM and two-view DBT. Both publications describe a potential future work-flow for minimizing radiologists' work-load in interpreting DBT examinations, namely by filtering out cases that were found to be normal by the AI algorithm. Figure 1 shows a schematic summary of the two studies, highlighting key similarities and differences in design.

Three evaluation methods were applied in these studies:

The first method tested the AI model by comparing the results predicted by the AI technology with the ground truth of cancer or cancer-free cases. In addition, simulation tests were also carried out to evaluate what would have been the resulting performance of the whole system had the technology first automatically filtered out certain studies, namely cases which were estimated as being cancer-free or cases which were estimated as being very likely cancerous and so were automatically scheduled for a recall and follow-up investigation.

The second evaluation method compared the performance of the AI technology with that of human readers who operated in a setting similar to that of the AI models.

The third evaluation method included multiple readers reviewing the localization of lesions determined by the AI system and comparing findings with a suspicion score that had been established by the human readers.

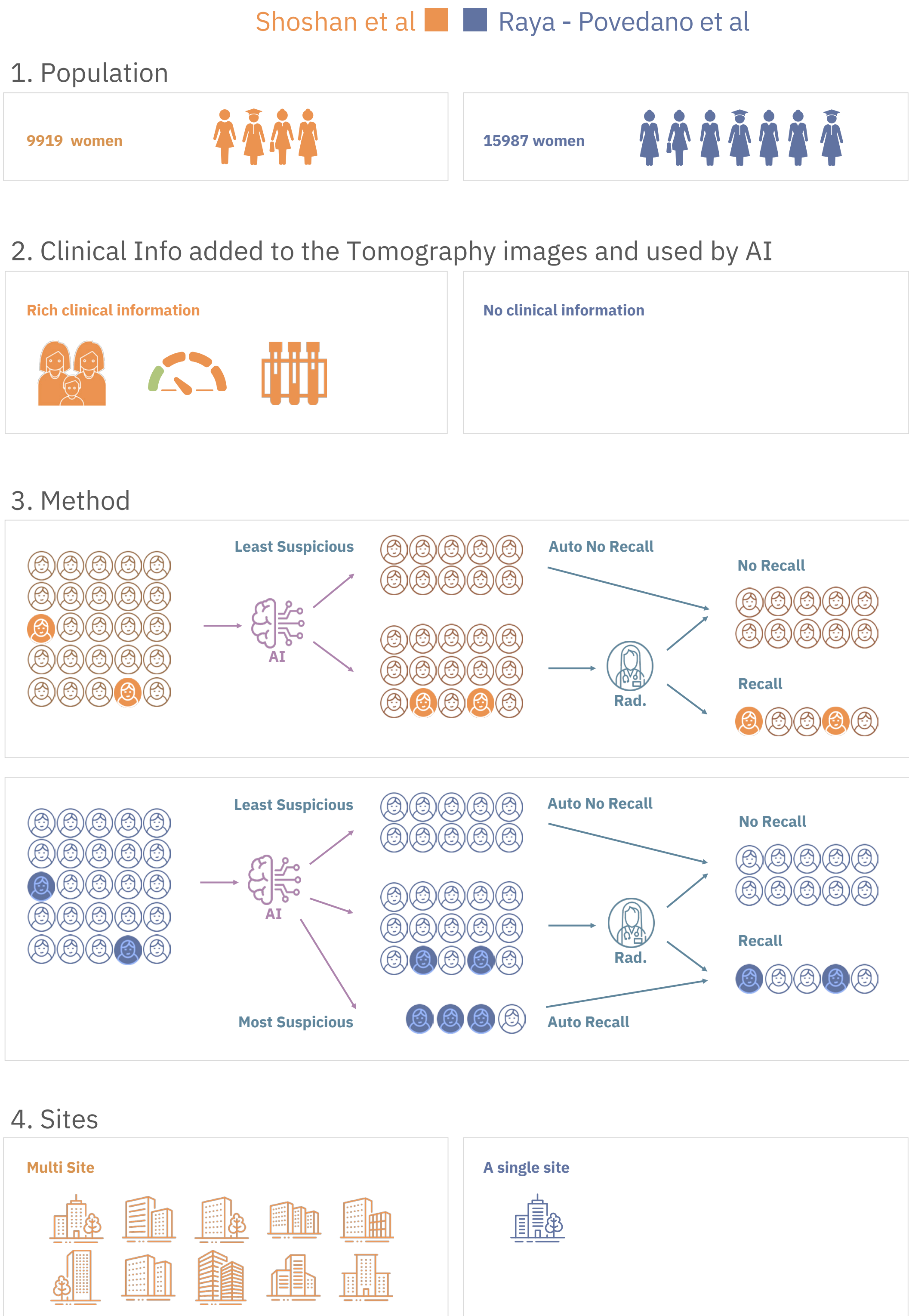

Figure 1: The figure above shows the main features of two recently published studies on the effect of the use of AI-based algorithms in the reading of DBT images. The studies are from two research groups namely Shoshan *et al.* [6] and Raya-Povedano *et al.* [7]. The information on the Shoshan *et al.* study is shown in orange icons and that from the study of Povedano *et al.* in blue icons.

Panel 1: The number of women involved in each study. Both studies involved large population cohorts.

Panel 2: In addition to DBT images, the study from Shoshan *et al.* also included non-image based clinical information for analysis by the AI-based algorithm.

Panel 3: Both studies assessed the effect on the radiologists' remaining work-load if examinations that were found by the AI approach to be the least suspicious had been excluded.

Panel 4: The Shoshan *et al.* study was multi-site whereas the Raya-Povedano *et al.* study was single site.

SUCCESSFUL APPLICATION OF AI TECHNOLOGY TO DBT

In both papers, the results are excitingly promising. The study by Raya-Povedano *et al* included 15 987 DBT examinations, of which 113 were confirmed cancer cases. In the study, the time necessary to carry out double human reading of DBT images was measured and found to be a total of 568 hours. A simulation was then performed to assess what would have been the necessary reading time if AI had been used to filter out all normal cases as well as automatically recall cancer cases. It was found that such a hybrid system would have meant that only 156 hours of human reader time would be needed, i.e. a 72.5% decrease in workload (p-value, 0.001). In addition the hybrid system showed non-inferior sensitivity compared to human readers, detecting 95 out of 113 cancers, (p-value = 0.38). With 588 recalls out of 15 987 examinations the recall rate of the hybrid system, was 16.7% lower than that of human readers (p-value = 0.001). More details and further discussion of the results are available in the original article [7].

In the study of Shoshan *et al*, the performance of the radiologists was measured by comparing the BI-RADS assessment category that was assigned to the case with the actual patient outcome. The sensitivity was 90.8% (417 of the 459 cancer cases were detected by the radiologists) while the specificity was 91.3% (4312 out of the 4723 cancer-free cases were correctly detected). In a simulation scenario, the AI technology was found to be able to reduce the screening workload by 39.6% (p-value, 0.001) while maintaining non-inferior sensitivity (90.0% vs 90.8%).

The ability of the AI model to assist radiologists by reducing their reading time workload was simulated as follows: assuming the AI technology reviews all DBT exams and removes from the radiologists' simulated worklist the cases that were estimated with high confidence as being cancer-free, what would be the effect on the radiologists' performance in reading the remaining exams? It was found that in such a scenario, the radiologist who benefitted from the AI analysis would have achieved a sensitivity of 93.6% and a specificity of 90.0%. The original study did not include localization of the identified lesions. However, in a follow-up study, the AI technology used to filter out normal cases was modified and applied to a localization task, namely to pinpoint specific 3D locations of biopsied areas in 3D tomosynthesis screening images. The technology was evaluated on the SPIE-AAPM-NCI DAIR Digital Breast Tomosynthesis Lesion Detection (DBTex) Challenge and showed equitable performance [8].

LOOKING FORWARD

In this short article, we have presented an innovative AI technology-based approach that was specifically developed to separate out cancer-free exams from those with cancer. The system also analyzes the tomography image to identify the slices that encompass suspicious lesions. The results of the evaluation of the system showed that AI technology can reach a high level of performance and complement the current work of radiologists. This holds true for both 2D mammography screening and for digital breast tomosynthesis screening.

The high performance of the system in interpreting DBT thus opens up the promise of further exploiting AI technology for novel tasks, such as the management of cases with high-risk benign lesions and the identification of prognostic breast cancer biomarkers. The first of these, namely the management of high-risk benign lesions has been extensively discussed in the literature, where it has been shown that tailored management and follow-up of patients with high-risk lesions can be a good alternative to surgical intervention for some women. An AI-based technology that interprets DBT images has the potential to support such an approach, thereby reducing over-treatment and sparing patients from unnecessary anxiety as well as reducing the high healthcare costs associated with surgery [9].

In the event that cancerous lesions are detected in a screening exam, it is vital that an accurate prognosis be assessed and treatment optimized at an early stage while of course minimizing the risks of both over- and under- treatment. In this context, AI algorithms can be trained on data that include information such as pathology findings and follow-up treatments so as to potentially allow the early stratification of patients and help support decisions for optimal treatment.

More broadly, in this article we have discussed how by leveraging DBT images and clinical data, AI-based technology could be developed to enable identification and separation of cancer-free breast exams from those containing cancer. Such technology, combined with other novel testing methods [10] could lead to a better understanding of tumor complexity and heterogeneity, and ultimately lead to the identification of prognostic biomarkers that could indicate optimal cancer treatment.

REFERENCES

1. Born J, Beymer D, Rajan D, Coy A, Mukherjee VV, Manica M, Prasanna P, Ballah D, Guindy M, Shaham D, Shah PL, Karteris E, Robertus JL, Gabrani M & Rosen-Zvi M. On the role of artificial intelligence in medical imaging of COVID-19. Patterns. 2021; 2(6): 100269.
https://doi.org/10.1016/j.patter.2021.100269

2. Yala A, Lehman C, Schuster T, Portnoi T & Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 2019; 292(1): 60-66.
https://doi.org/10.1148/radiol.2019182908

3. Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, Barkan E, Herzel E, Naor S, Karavani E & Koren G. Predicting breast cancer by applying deep learning to linked health records and mammograms. Radiology 2019; 292(2): 331-342.
https://doi.org/10.1148/radiol.2019182622

4. Chorev M, Shoshan Y, Akselrod-Ballin A, Spiro A, Naor S, Hazan A, Barros V, Weinstein I, Herzel E, Shalev V & Guindy M., The case of missed cancers: applying AI as a radiologist's safety net. In International Conference on Medical Image Computing and Computer-Assisted Intervention 2020, October. pp. 220-229. Springer, Cham.

5. Rahman WT and Helvie MA. Breast cancer screening in average and high-risk women. Best Practice & Research Clinical Obstetrics & Gynaecology. Best Pract Res Clin Obstet Gynaecol. 2021; S1521-6934(21)00171-1.
https://doi.org/10.1016/j.bpobgyn.2021.11.007.

6. Shoshan Y, Bakalo R, Gilboa-Solomon F, Ratner V, Barkan E, Ozery-Flato M, Amit M, Khapun D, Ambinder EB, Oluyemi ET, Panigrahi B, DiCarlo PA, Rosen-Zvi M & Mullen LA. Artificial Intelligence for Reducing Workload in Breast Cancer Screening with Digital Breast Tomosynthesis. Radiology, 2022 Apr; 303(1):69-77.
https://doi.org/10.1148/radiol.211105.

7. Raya-Povedano JL, Romero-Martín S, Elías-Cabot E, Gubern-Mérida A, Rodríguez-Ruiz A & Álvarez-Benito M. AI-based strategies to reduce workload in breast cancer screening with mammography and tomosynthesis: a retrospective evaluation. Radiology. 2021 Jul; 300(1): 57-65.
https://doi.org/10.1148/radiol.2021203555

8. Park J, Shoshan Y, Martí R, Gómez del Campo P, Ratner V, Khapun D, Zlotnick A, Barkan E, Gilboa-Solomon F, Chłędowski J & Witowski J. Lessons from the first DBTex Challenge. Nat Mach Intell 2021. 3, 735–736 (2021). https://doi.org/10.1038/s42256-021-00378-z

9. Catanzariti F, Avendano D, Cicero G, Garza-Montemayor M, Sofia C, Venanzi Rullo E, Ascenti G, Pinker-Domenig K & Marino MA. 2021. High-risk lesions of the breast: concurrent diagnostic tools and management recommendations. Insights into Imaging. 2021; 12(1): 1-19.
https://doi.org/10.1186/s13244-021-01005-6

10. Kashyap A, Rapsomaniki MA, Barros V, Fomitcheva-Khartchenko A, Martinelli AL, Rodriguez AF, Gabrani M, Rosen-Zvi M & Kaigala G. Quantification of tumor heterogeneity: from data acquisition to metric generation. Trends in Biotechnology 2021Trends Biotechnol 2021 Dec 28; S0167-7799(21)00267-5.
https://doi.org/10.1016/j.tibtech.2021.11.006.