CHAPTER 7

ARTIFICIAL INTELLIGENCE IN BREAST CANCER DIAGNOSIS

Levent Çelik MD BA

Maltepe University, Radiology Dept. Radiologica Imaging center.

Corresponding Author: Levent Çelik, MD

Department of Radiology Maltepe University Hospital Maltepe Istanbul Turkey

Breast cancer is the leading cause of cancer death in women worldwide (1). Since life expectancy largely depends on the stage of the cancer at the time of diagnosis, screening programs have been introduced for detecting breast cancer as early as possible. Currently screening programs are performed with mammography. Although benefits of mammography based screening have been shown, it has low sensitivity in dense breasts (2). Digital breast tomosynthesis (DBT) is partly effective in solving this problem. Breast MRI is much more effective in dense breasts; but it is used only for women with increased breast cancer risk as a screening modality.

Radiologic images are data, not just pictures. It is obvious that various data processing algorithms can be used to evaluate radiological images. Computer aided detection (CAD) systems were developed in the early 1990s to increase breast cancer detection with mammography. These systems were programs where distinctive features were defined by radiologists, and these differential features were taught to computers by programmers.

The imaging information obtained in breast MRI and DBT examinations has a high dimensional and multiparametric nature, which makes reading of these images a difficult and time consuming task for radiologists. Currently, several automated tools, based on computer vision and machine learning (ML) techniques, are being developed in order to increase reading efficiency and accuracy of the radiologists.

Since 2012, we have been witnessing rapid and revolutionary changes in the fields of ML, computer vision; and consequently, medical image analysis with the advent of the algorithms named as 'deep learning'. These fields changed literally overnight when the 2012 ILSVRC ImageNet challenge was won by a Deep Convolutional Networks (CNN) algorithm (3). Deep learning methods have been improved further with explosively increasing number of studies since 2012, being the method of choice in automated image analysis. A deep learning architecture known as CNN has become dominant for processing images. The success of deep learning with CNNs for images in nonmedical fields has increased hopes for and research towards analysis of medical images. Although neural networks have been used for decades, in recent years three key factors have enabled the training of large neural networks: (a) the availability of large quantities of labeled data, (b) inexpensive and powerful parallel computing hardware, and (c) improvements in training techniques and architectures. CNNs of increasing depth and complexity have gained significant attention since 2012.

An important advantage of deep learning is that it does not require image feature identification and calculation as a first step; rather, features are identified as part of the learning process. Deep learning systems learn the distinctive features from the labeled data themselves. Therefore, a large number of correctly labeled data is needed (Fig. 1).

Artificial Intelligence (AI) is the most popular topic nowadays in all disciplines of science. Medical imaging is the most rapidly rising area of health innovation with AI. Not more than 10 years ago, the total number of publications on AI in radiology only just exceeded 100 per year. Currently, publications about AI in radiology have increased from 100-150 per year to 700-800 per year.

Mammography and Digital Breast Tomosynthesis

United States Food and Drug Administration (FDA) approved mammography CAD systems for breast cancer detection in 1998, and reimbrusement for mammography CAD has began in 2002. Early CAD systems were actually supervised ML systems, and they favored sensitivity over specificity. They were widely used in US because of medicolegal worries, but not in the rest of the world.

CAD systems were introduced as an aid for radiologists, trying to improve their performance for detection and diagnosis. It is important to minimize misses and interpretation errors of visible lesions at digital mammography, which contribute to at least 25% of detectable cancers being missed (4). The benefit of using CAD in breast cancer screening is still unclear. Most

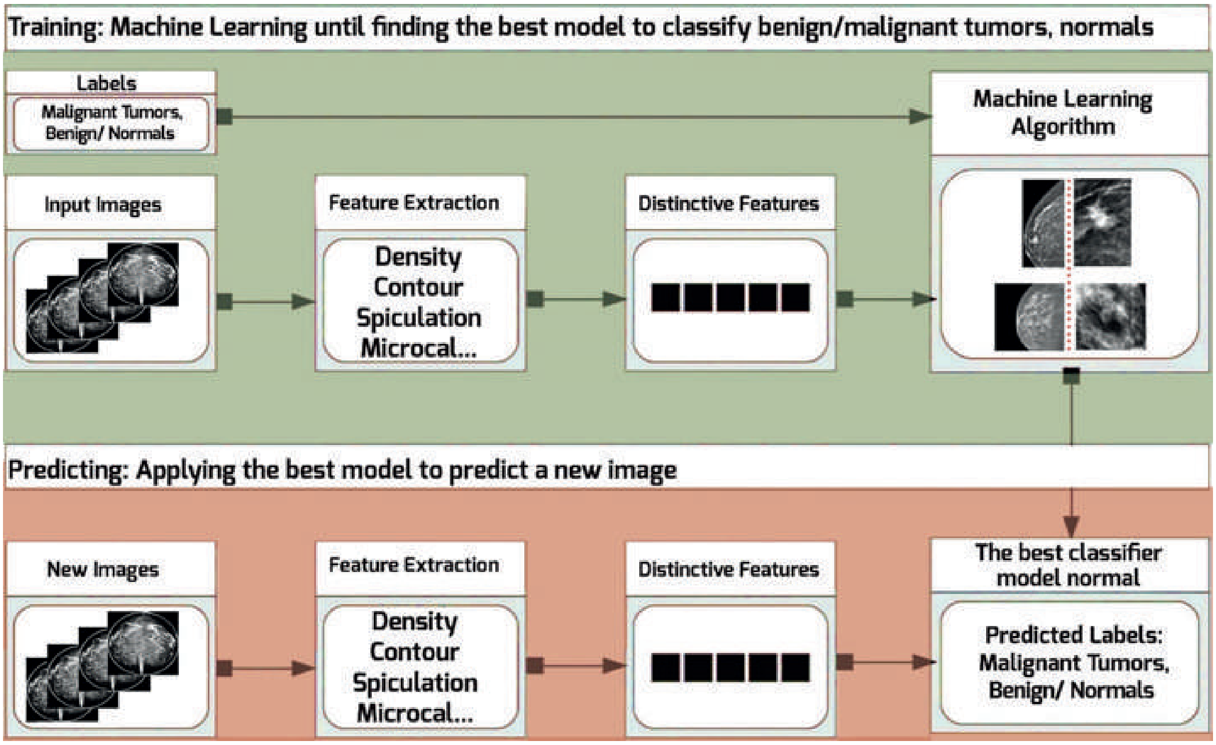


Fig. 1: A lot of accurate labeled data is required in order to train AI systems with deep learning

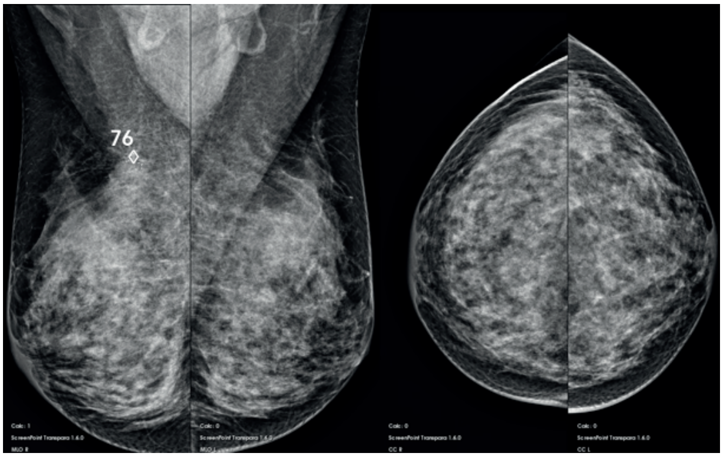
evidence shows no clear improvement in the cost-effectiveness of screening, mainly because of the low specificity of most traditional CAD systems (5). They prompt marks on true positive, but also many false positive areas on mammograms. However, substantial improvements in AI with deep CNNs are reducing the difference in performance between humans and computers in many medical imaging applications, including breast cancer detection (6). Therefore, this new generation of deep learning–based CAD systems may finally allow for an improvement in the performance of breast cancer screening programs (7).

Mammography AI systems show high-risk areas on mammography with different visual aids. Some show classic CAD marks on a mammogram for calcifications and soft tissue lesions, with a quantitative indication of the risk, while others only visualize the risk in the form of a temperature map without distinguishing features (Fig. 2).

Apart from the evolution of AI algorithms, the aid that the AI system provides can also help improve screening. Studies have shown that, using CAD concurrently as a decision support tool helps radiologists more than does the traditional approach (8). Diagnostic performance of breast radiologists were higher with support from an AI system compared with reading unaided. The average reading times per case were similar under both conditions (Fig. 3) (9).

Transfer learning is a ML method, that focuses on storing the information obtained while solving a problem, and then applying it to a different but related problem. For example, the information obtained while learning to recognize cars can be applied when trying to recognize trucks. AI programs to be developed for reading DBT will use the knowledge of AI programs developed for reading mammography. By means of transfer learning, the training of AI systems will be faster and more accurate, and will require less labeled data (13). Due to this knowledge transfer, AI programs that read DBT will be available in daily practice earlier than expected. With the spread of successful AI programs in reading DBT, tomo only breast cancer screening will probably be accelerated (Fig. 4).

DBT is a three dimensional (3D) imaging technique that has been shown to increase breast cancer-detection rates and reduce false positive results as compared with digital mammography (2D) alone (10, 11). The superimposition of tissues in 2D digital mammography has contributed to false positives. DBT reduces overlapping opacities, and thus increases lesion conspicuity while reducing recall rates. Recent studies have found that DBT is particularly beneficial in the detection of masses in women with increased breast density or heterogeneously dense breasts. A relative disadvantage is that, DBT increases reading times from 50 to 200 % due to the increased number of images (12). There is a need for optimized CAD and diagnosis systems for DBT in order to reduce radiologist evaluation time and improve efficiency. With the trend of increasing use of DBT, developers of AI-CAD systems have taken this emerging imaging technique into consideration. In March 2020, FDA approved the Screenpoint's software called Transpara in reading DBT. This artificial intelligence software has been shown to reduce reading time by 35 seconds per case while increasing radiologist ac-



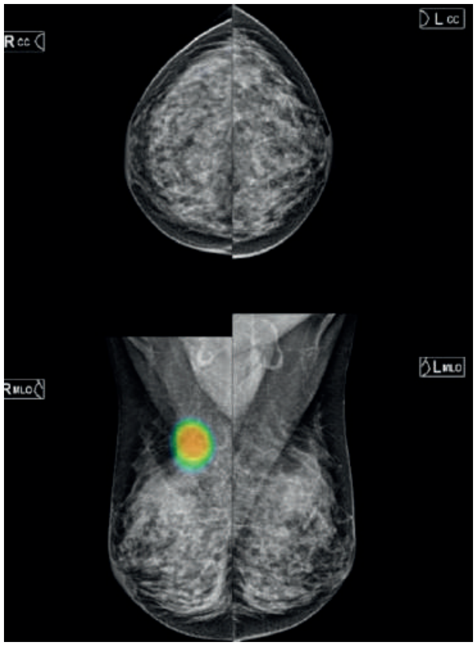


Fig. 2: A) AI program trained with deep learning for reading mammography. It shows a CAD mark on a suspicious area together with its risk score **B)** Same patient, Another AI program, trained with deep learning, shows its findings in the form of a head map.

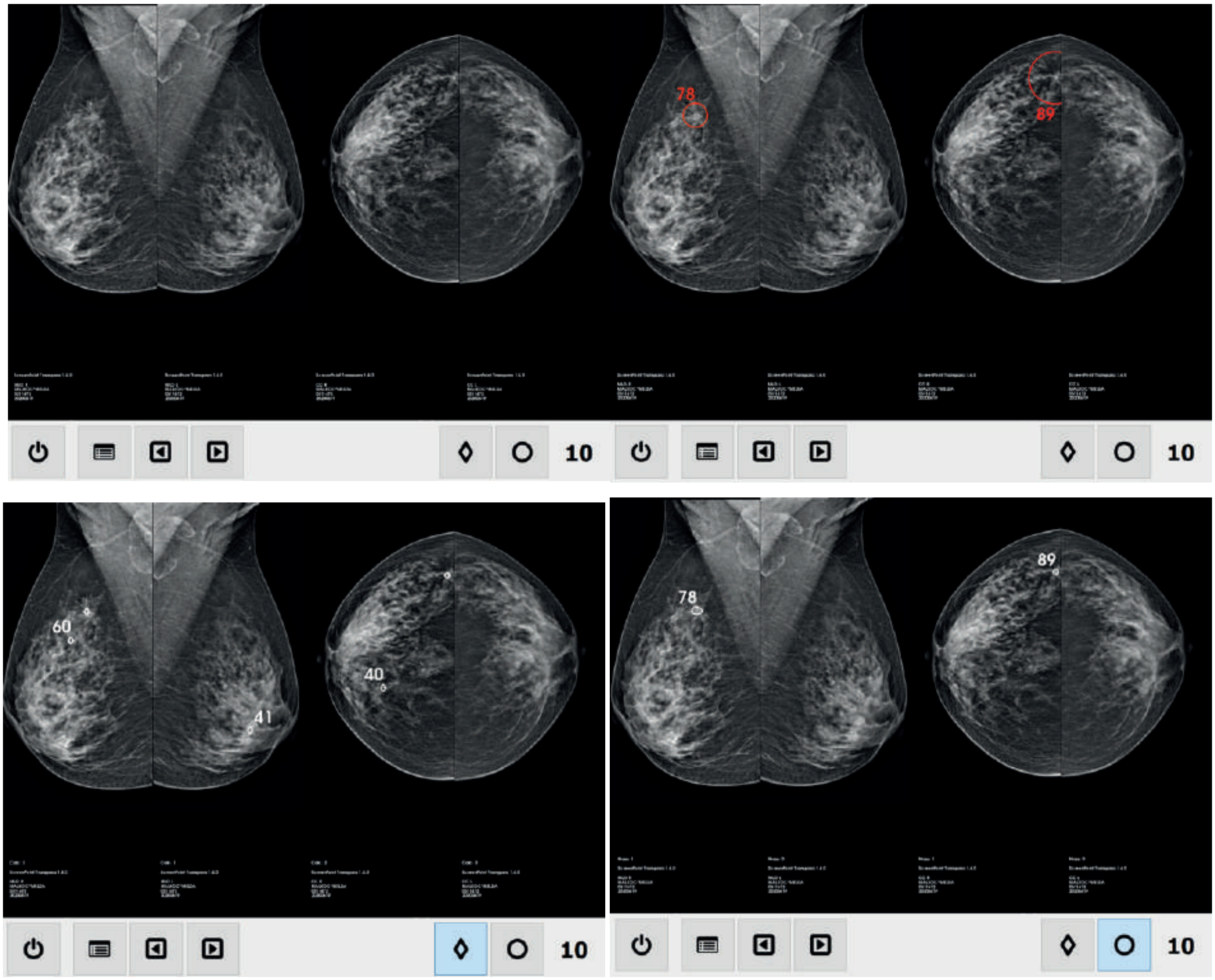


Fig. 3: An AI mammography system where the workflow works as a decision support system. **A)** System shows only risk score (Transpara score 1-10, 1 is the lowest, and 10 is the highest risk score) **B)** If the radiologist clicks on the area he suspects on the mammogram, artificial intelligence shows him the risk score on the mammography in that area and counter projection, if any. **C)** If the radiologist clicks on the microcalcification CAD marker (rhombus), the artificial intelligence shows the suspicious microcalcification areas found on the mammography with the risk score. **D)** If the radiologist clicks on the mass CAD marker (circle), the artificial intelligence shows the areas of suspicious mass found on mammography together with the risk score.

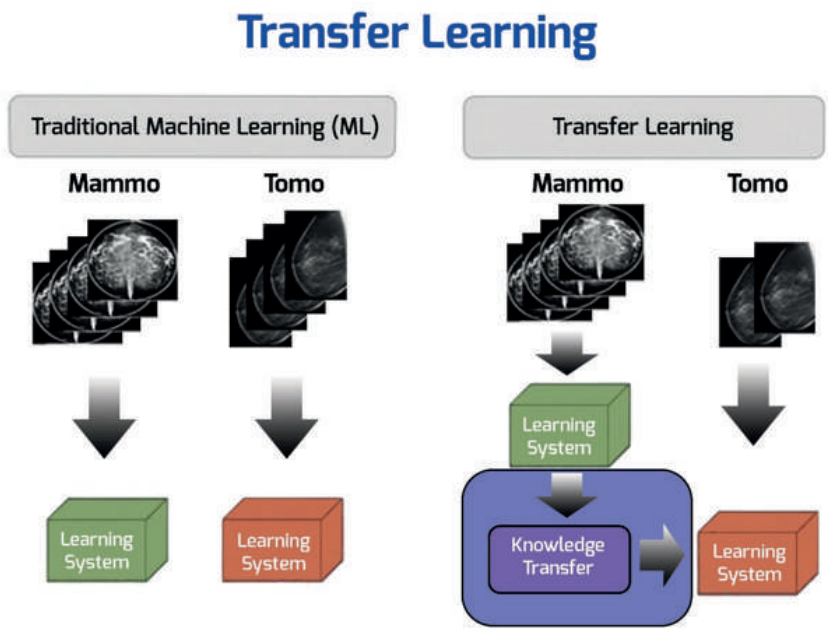


Fig. 4: Transfer learning: Artificial intelligence systems that read tomosynthesis can be trained in a shorter time with less labeled cases, using knowledge from mammography reading training.

curacy in reading DBT (Fig. 5) (https://bit.ly/3eE2Vsw, inpress).

Deep learning-based CAD softwares can correlate MLO and CC views when abnormalities are visible in both views, just as radiologists do. Major limitations of CAD systems are that, they cannot compare old and news images (temporal comparison), and most CAD systems cannot compare right and left breast images

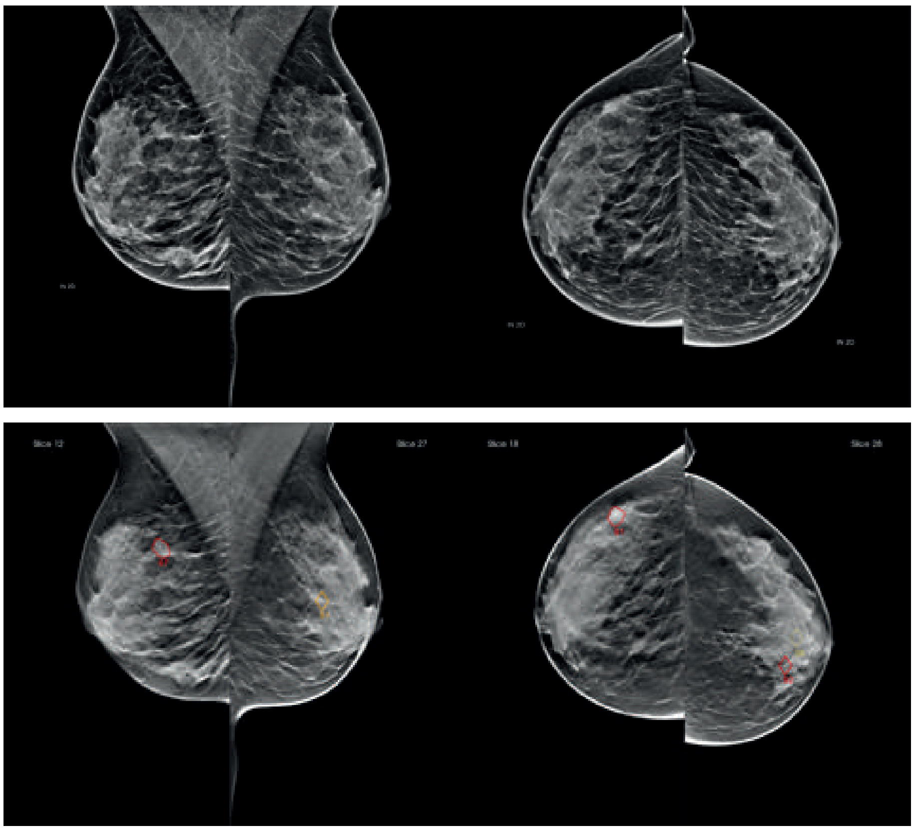


Fig. 5: A and **B)** The artificial intelligence system links findings between CC and MLO. With one click the viewer can bring up slices in CC and MLO of the same lesion (upper row synthetic mammographies).

(symmetry comparison). They examine high resolution mammography images like a patch work. They don't look at a mammogram as a whole picture. In the near future, further development of deep learning algorithms and hardwares will overcome these obstacles. Thus, AI-based CAD systems will become even better decision support systems, reaching a performance similar to or better than that of radiologists for breast cancer detection in mammography.

Another application of AI in mammography and DBT is worklist prioritization. It can prioritize cases with suspicious findings in the worklist, thus increasing efficiency of radiologists and allowing fast evaluation of cancer cases.

Breast Ultrasound

Breast ultrasound (US) can provide additional information to further characterize mammographic findings or palpable abnormalities, and to guide interventional procedures. In particular, US can increase detection of early breast cancer when used as a supplementary imaging technique in women with dense breasts. Automated breast ultrasound (ABUS) has been developed to overcome the limitations of inter-operator variability with handheld ultrasound (HHUS) and is able to generate thousands of 2D US images to obtain a 3D representation of the breast tissue. The CAD system for ABUS has been shown to decrease reading times without compromising diagnostic accuracy, with CAD-ABUS averaging a reading time of 113.4 seconds per case compared to 158.3 seconds per case using ABUS alone (14). The system is able to automatically extract features from suspicious areas of breast tissue that are larger than or equal to a diameter of 5 mm, finally generating a score of suspiciousness for each area. The CAD output includes the CAD Navigator image, which is displayed simultaneously with the original ABUS and acts as a roadmap for navigation, as well as CAD marks (coloured circles) for potentially malignant lesions. Cloud based AI programs for breast US can be integrated into PACS and used as a decision support system (15). FDA approved AI-based decision support systems for breast US are now commercial products.

Breast Magnetic Resonance Imaging

Breast magnetic resonance imaging (MRI) is known for its high sensitivity in detecting breast lesions. It has been shown that lesions that are occult in mammography and US can be detected in breast MRI (16). Despite the higher sensitivity of breast MRI, mammography remains as the standard modality for general screening of women for breast cancer, since high cost of breast MRI limits its widespread use. One of the cost-increasing factors is the acquisition of several sequences for a single breast MRI study. In a typical breast MRI acquisition protocol, after an initial T1-weighted (T1w) MRI scan is obtained, a contrast agent is administered to the patient to enhance lesions, and, subsequently, several post-contrast T1w MRI scans are obtained. CAD evaluation mostly relies on contrast media uptake dynamics. To decrease the cost and be able to facilitate the application of this imaging modality in screening, abbreviated breast MRI protocols have been introduced (17).

Performance of CAD systems in breast MRI mainly rely on dynamic features, whereas, in clinical assessment, morphology is the most vital information and dynamic information is auxiliary. Automatic evaluation of lesion morphology in a conventional CAD system is difficult, since it requires specific features to be extracted from images. Furthermore, differentiation of such features is known to be the most difficult part and main performance limiting factor of conventional computer vision systems. Recently popular deep learning methods tackle this difficulty by learning such features automatically based on examples, instead of using human-engineered features, often using CNNs (18).

Also new AI-based CAD systems use symmetry information in breast MRI arising from the differences between the contrast enhancements of the two breasts, in addition to the 3D morphological information in the candidate regions. This is important because asymmetry is an important finding in breast MRI, and it is one of the features stated in the guidelines for evaluation of breast MRI scans (19).

With the recent improvements in MR technology, novel ultrafast DCE-MRI sequences allow monitoring of the initial uptake of contrast agent, instead of imaging washout at the late phase. The high dimensional and multi-parametric nature of the information currently obtained in breast MRI makes interpretation still more complex and labor-intensive. Moreover, inter-observer variations are common. The use of CAD systems may improve diagnostic accuracy by decreasing inter-observer variations, providing support for clinical decisions and reducing the number of false-positive biopsies (20). It has been shown that, classification of benign and malignant breast lesions imaged with a multi-parametric ultrafast DCE-MRI protocol using AI techniques is at least as accurate as dedicated breast radiologists (21).

Conclusion

AI will surely impact radiology, and more quickly than other medical fields. It will change radiology practice more than anything since the discovery of x-ray by Wilhelm Roentgen. Unprecedented success of deep learning in image recognition has revived the optimism in automating image interpretation tasks at the performance level of humans. Only in the last few years, have we seen applications in various domains that reach or even surpass human performance at certain image recognition tasks, such as breast imaging (21,22). Consequently, there have been discussions about the feasibility of replacing human labor with deep learning based AI in various fields including radiology. However, in order to avoid far-fetched expectations, it is important to understand the limitations of these AI systems. ML systems, including deep learning, are specialized in solving isolated tasks, while human intelligence is able to develop understanding of various concepts and is able to combine vast amount of information from different levels and domains for performing tasks.

A malfunctioning AI system may have the opposite of its intended positive effect; a failing system can create new safety hazards. It should not be assumed that the worst-case failure of a system that includes AI is equivalent to the function of that system without AI. For instance, using AI for mammography worklist prioritization is generally considered low risk because the current state for most practices is effectively random prioritization. However, the worst-case failure for AI prioritization is not random, but reversed prioritization. Adding AI introduces new possibilities for failure.

An AI system is only as good as its inputs. The accuracy of inputs to AI systems is equally important as the AI's accuracy in interpreting those inputs. Many findings that radiologists easily label as artifacts in their daily routines may be the source of misinterpretation for AI systems reading mammography.

Radiologists, who are the most open medical group to technological developments, will include AI applications in their practice. In the near future, AI systems will probably become an integral part of breast imaging. Breast radiologists should learn all aspects of AI applications in breast imaging, and should be able to incorporate them into their practices. Breast clinicians should also be aware of the benefits and shortcomings of this new technology.

References:

1. A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," CA: a cancer journal for clinicians, vol. 61, no. 2, pp. 69–90, 2011.
2. L. Tabar, A. Gad, L. Holmberg, U. Ljungquist, K. C. P. Group, C. Fagerberg, L. Baldetorp, O. Gr¨ontoft, B. Lundstr¨om, J. M˚anson, 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," The Lancet, vol. 325, no. 8433, pp. 829–832, 1985.
3. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, 2012.
4. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology 1992;184(3):613–617.
5. Lehman CD, Wellman RD, Buist DS, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med 2015;175(11):1828–1837.
6. Kooi T, Litjens G, van Ginneken B, et al. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 2017;35:303–312.
7. Trister AD, Buist DSM, Lee CI. Will machine learning tip the balance in breast cancer screening? JAMA Oncol 2017;3(11):1463–1464.
8. Samulski M, Hupse R, Boetes C, Mus RD, den Heeten GJ, Karssemeijer N. Using computer-aided detection in mammography as a decision support. Eur Radiol 2010;20(10):2323–2330.
9. Rodríguez-Ruiz, Alejandro & Krupinski, Elizabeth & Mordang, Jan-Jurre & Schilling, Kathy & Heywang-Köbrunner, Sylvia & Sechopoulos, Ioannis & Mann, Ritse. (2018). Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. Radiology. 290. 181371. 10.1148/radiol.2018181371.
10. 7. 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;14:583e9.
11. Haas BM, Kalra V, Geisel J, et al. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. Radiology 2013;269:694e700
12. Tagliafico AS, Calabrese M, Bignotti B, et al. Accuracy and reading time for six strategies using digital breast tomosynthesis in women with mammographically negative dense breasts. Eur Radiol 2017;27: 5179e84.
13. Li X1, Qin G2, He Q1, Sun L1, Zeng H2, He Z2, Chen W2, Zhen X3, Zhou L4 Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification. Eur Radiol. 2020 Feb;30(2):778-788. doi: 10.1007/s00330-019-06457-5.
14. van Zelst JCM, Tan T, Clauser P, et al. Dedicated computer-aided detection software for automated 3D breast ultrasound; an efficient tool for the radiologist in supplemental screening of women with dense breasts. Eur Radiol 2018;28:1-11.
15. Barinov L, Jairaj A, Paster L, et al. Decision quality support in diagnostic breast ultrasound through artificial Intelligence. In: 2016 IEEE signal processing in medicine and biology symposium, SPMB 2016. Piscataway, NJ: IEEE; 2017, https://doi.org/10.1109/SPMB.2016.7846873.
16. C. D. Lehman, C. Isaacs, M. D. Schnall, E. D. Pisano, S. M. Ascher, P. T. Weatherall, D. A. Bluemke, D. J. Bowen, P. K. Marcom, D. K. Armstrong, et al., "Cancer yield of mammography, mr, and us in high-risk women: prospective multi-institution breast cancer screening study 1," Radiology, vol. 244, no. 2, pp. 381–388, 2007.
17. C. K. Kuhl, S. Schrading, K. Strobel, H. H. Schild, R.-D. Hilgers, and H. B. Bieling, "Abbreviated breast magnetic resonance imaging (mri): first postcontrast subtracted images and maximum-intensity projectiona novel approach to breast cancer screening with mri," Journal of Clinical Oncology, vol. 32, no. 22, pp. 2304– 2310, 2014.
18. C. D. Lehman, J. D. Blume, W. B. DeMartini, N. M. Hylton, B. Herman, and M. D. Schnall, "Accuracy and interpretation time of computer-aided detection among novice and experienced breast mri readers," American Journal of Roentgenology, vol. 200, no. 6, pp. W683–W689, 2013.
19. C. Kuhl, "The current status of breast mr imaging part i. choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice," Radiology, vol. 244, no. 2, pp. 356–378, 2007.
20. S. Singh, J. Maxwell, J. A. Baker, J. L. Nicholas, and J. Y. Lo, "Computer-aided classification of breast masses: performance and interobserver variability of expert radiologists versus residents," Radiology, vol. 258, no. 1, pp. 73–80, 2011.
21. Mehmet Ufuk Dalmış. Automated Analysis of Breast MRI: from Traditional Methods into Deep Learning. Doctoral Thesis to obtain the degree of doctor from Radboud University Nijmegen on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken, ISBN 978-90-9032014-4
22. B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol, et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," Jama, vol. 318, no. 22, pp. 2199–2210, 2017.