ELSEVIER

Seminars in
NUCLEAR
MEDICINE

Application of Deep Learning in Breast Cancer Imaging

Luuk Balkenende, MSc,[*,†] Jonas Teuwen, PhD,[†,‡] and Ritse M. Mann, MD, PhD[*,†]

This review gives an overview of the current state of deep learning research in breast cancer imaging. Breast imaging plays a major role in detecting breast cancer at an earlier stage, as well as monitoring and evaluating breast cancer during treatment. The most commonly used modalities for breast imaging are digital mammography, digital breast tomosynthesis, ultrasound and magnetic resonance imaging. Nuclear medicine imaging techniques are used for detection and classification of axillary lymph nodes and distant staging in breast cancer imaging. All of these techniques are currently digitized, enabling the possibility to implement deep learning (DL), a subset of Artificial intelligence, in breast imaging. DL is nowadays embedded in a plethora of different tasks, such as lesion classification and segmentation, image reconstruction and generation, cancer risk prediction, and prediction and assessment of therapy response. Studies show similar and even better performances of DL algorithms compared to radiologists, although it is clear that large trials are needed, especially for ultrasound and magnetic resonance imaging, to exactly determine the added value of DL in breast cancer imaging. Studies on DL in nuclear medicine techniques are only sparsely available and further research is mandatory. Legal and ethical issues need to be considered before the role of DL can expand to its full potential in clinical breast care practice.

Introduction

With 685,000 deaths, breast cancer is the leading cause of cancer mortality worldwide among women in 2020.[1] Breast cancer imaging plays a major role in reducing this excessive number of deaths. Screening programs have been set up to detect breast cancer early, enabling easier treatment and higher survival rates than to late-stage detected cancers.[2] Moreover, breast cancer imaging techniques are essential for the monitoring and evaluation of cancer treatment.

The screening programs, together with a rising amount of breast cancer incidences and consequential treatments, have been placing a high workload on (breast) radiologists over the last 2 decades.[3] Artificial intelligence (AI) in the form of computer-aided detection/diagnosis (CAD) systems is used to assist radiologists and cut back workload. Conventional CAD systems are based upon traditional machine learning (ML) techniques; pre-defined (hand-crafted) features are the input to the system.[4] However, large retrospective studies have shown that these conventional CAD systems do not improve diagnostic accuracy.[5]

Deep learning (DL), a subset of ML, operates directly on the image data, thereby defining (learning) appropriate features without human interference. In computer vision, outstanding achievements have been obtained with DL after the breakthrough of deep convolutional neural networks (CNNs) in 2012.[6] Since then, DL techniques and, in particular, CNNs were rapidly implemented in medical image analysis[4] and used in breast radiology to improve and elevate the conventional CAD systems to surpass the radiologist performance level.

*Department of Radiology, Netherlands Cancer Institute (NKI), Amsterdam, The Netherlands.
†Department of Medical Imaging, Radboud University Medical Center, Nijmegen, The Netherlands.
‡Department of Radiation Oncology, Netherlands Cancer Institute (NKI), Amsterdam, The Netherlands.
Address reprint requests to Ritse M. Mann, MD, PhD, Radboud University Medical Center, Department of Medical Imaging, Geert grooteplein 10, 6525 GA Nijmegen, The Netherlands. E-mail: ritse.mann@radboudumc.nl

 https://doi.org/10.1053/j.semnuclmed.2022.02.003

In recent years, many applications and submitted papers using DL in breast cancer imaging are still optimizing detection and diagnosis. However, the role of AI and DL has exceeded the early objective of CAD systems to aid radiologists only in these 2 tasks. Nowadays, DL is used in a plethora of different tasks, such as image reconstruction and generation, cancer risk prediction and prediction and assessment of therapy response. This review gives an overview of the current role of DL in the various tasks for different breast imaging modalities, and provides some insight in the future role of these DL techniques in clinical practice.

Introduction to Breast Imaging Modalities

Commonly, breast imaging is performed using digital mammography (DM), digital breast tomosynthesis (DBT), ultrasound (US), magnetic resonance imaging (MRI), or a combination of the mentioned. In DM, X-rays are sent through the breast and collected by a digital X-ray detector, creating a 2-dimensional (2D) image of the breast. It is a fast and easy technique. However, it suffers from the problem of tissue superposition. Especially in dense breasts (breasts with a high density of fibroglandular tissue), the chance that fibroglandular tissue masks lesions are high.[7] To overcome this problem partially, 2 DM scans are made: the craniocaudal view and the mediolateral oblique view.

In DBT, multiple X-ray scans are taken from slightly different positions, resulting in a (partial) tomographic image with a reduced superposition effect compared to DM. A negative aspect of DBT is its more difficult interpretation, resulting in an increased reading time.[8]

When performing US, sound waves are sent through the breast while at the same time backscattered waves are detected. From the detected waves, a US image is constructed. Therefore, US uses no ionizing radiation, which is an important advantage. US images are shown in real-time, meaning that a US procedure needs to be performed by a radiologist and is directly evaluated. However, it is difficult to document the entire breast with ultrasound. Furthermore, compared to DBT and DM, US images can be hard to interpret and the frequent detection of benign breast lesions leads to many unnecessary biopsies.

MRI uses a strong magnetic field, together with short-pulsed radio waves, to excite the water molecules inside specific places of the breast/human body, which transmit radio waves themselves when falling back to their ground state. These radio waves are detected and transformed into a 3D MRI scan. MRI is the most sensitive modality for breast cancer compared to DM, DBT, and US.[9] However, breast MRI requires intravenous contrast administration and is a relatively expensive method.

Introduction to AI and DL

Artificial intelligence is the umbrella term for the ability of algorithms to do tasks that require a form of intelligence. A subset of AI is ML: algorithms belonging to ML gain knowledge (learn) from data to improve their performance. Data fed into an ML program can be represented in 2 forms: as features or as raw data. Features are variables in data that can be quantified, like lesion length, while raw data in breast cancer imaging are the DM/US/MRI scans.

If raw data is the input-form, the algorithms need to find features themselves. Although those learned features generally result in a better performance than using handcrafted features, learning features is a problematic task for such algorithms.[10] This problem can be solved by using a subset of ML techniques, collectively known as DL. In DL, features are expressed in terms of other, simpler features. DL algorithms are normally referred to as deep neural networks (DNNs), as they consist out of multiple (deep) layers of connected neurons. A special type of DNNs are CNNs (Fig. 2). CNNs are specifically designed to find useful features in images and are widely used in breast cancer image analysis.

To compare the performance of DL networks with human standards, different metrics are used for different tasks. In

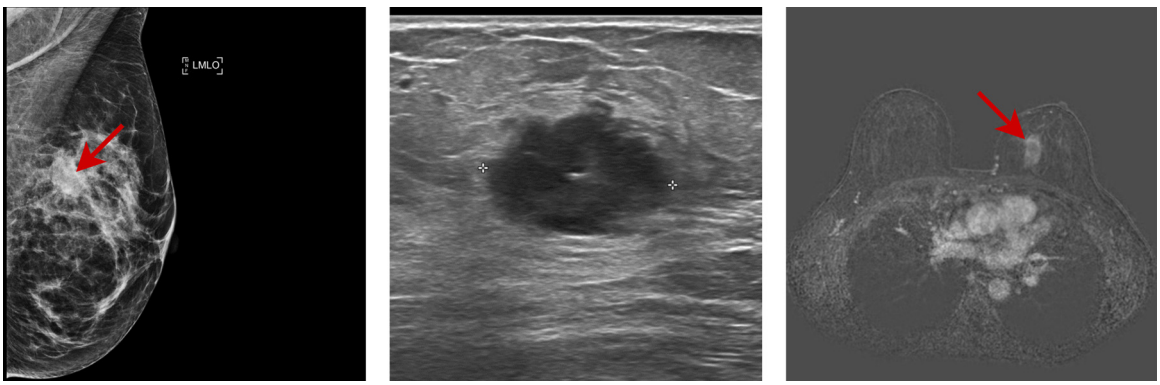

Figure 1 Digital mammography, ultrasound and magnetic resonance scans of a 45 y old woman with an infiltrating ductal carcinoma in the left breast, indicated with the red arrows. Left: digital mammography, middle: ultrasound, right: magnetic resonance.

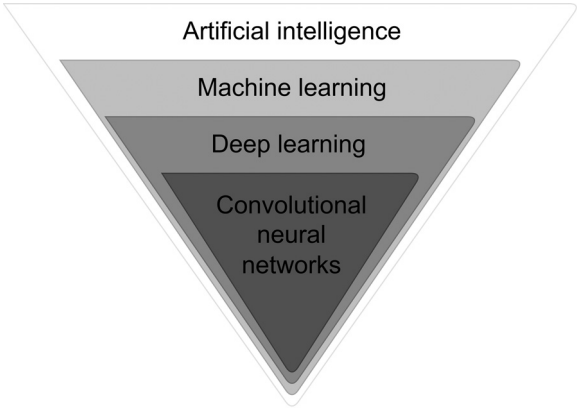

Figure 2 A Venn diagram showing the relationship between the different subsets of artificial intelligence.

classification, the metrics are based on receiver operating characteristic analysis. Here, accuracy, sensitivity, specificity, and area under the curve (AUC) play an important role. Accuracy gives the percentage of correctly classified samples, sensitivity the probability of a positive (thus malignant) output by the model/radiologist given that the sample is malignant, specificity the likelihood of a negative (thus benign) output by the model/radiologist given that the sample is benign, and AUC the average sensitivity for all possible values of the specificity.

In the case of segmentation/detection, the intersection over union and Dice similarity coefficient (DSC) are mostly used as metrics. These metrics both give a measure of the amount of spatial overlap between 2 samples. In detection, where the task is to draw an as-small-as-possible bounding box around the malignant lesions, the intersection over union/DSC score is the amount of spatial overlap between the ground truth box and the predicted box. In segmentation, which outlines the contour of the malignant lesion, the IoU/DSC score is calculated between this predicted contour and the ground truth contour.

Digital Mammography and Digital Breast Tomosynthesis

The first CAD algorithms that implemented DL used CNNs to distinguish between the benign and malignant nature of suspicious candidate lesions.[11-13] However, these candidates, also called region of interests (ROIs), were first extracted from whole DM or DBT scans by hand[13] or with conventional CAD methods,[12] without the use of DL methods. Kooi et al.[11] first extracted features directly from image patches by using a previously developed algorithm for mammographic lesions, using a random forest classifier to select the candidates. On those candidates, a CNN was trained. When comparing the results between the CNN and the reference system based on a conventional CAD method, the CNN had a non-significant higher AUC (0.929 vs 0.91, $P = 0.2$).

To fully exploit DL, it can also be used for candidate selection in whole images, thus consequently changing the input of the CAD systems from small patches containing suspicious regions to using the full DM and DBT scans[14-16] (Fig. 3). For example, Al-Masni et al.[15] implemented a CAD system based on YOLO, an algorithm that is capable of simultaneously performing classification and detection. They obtained an AUC of 0.965 on classifying masses on DM images as malignant or benign, showing that it can indeed be beneficial to use DL

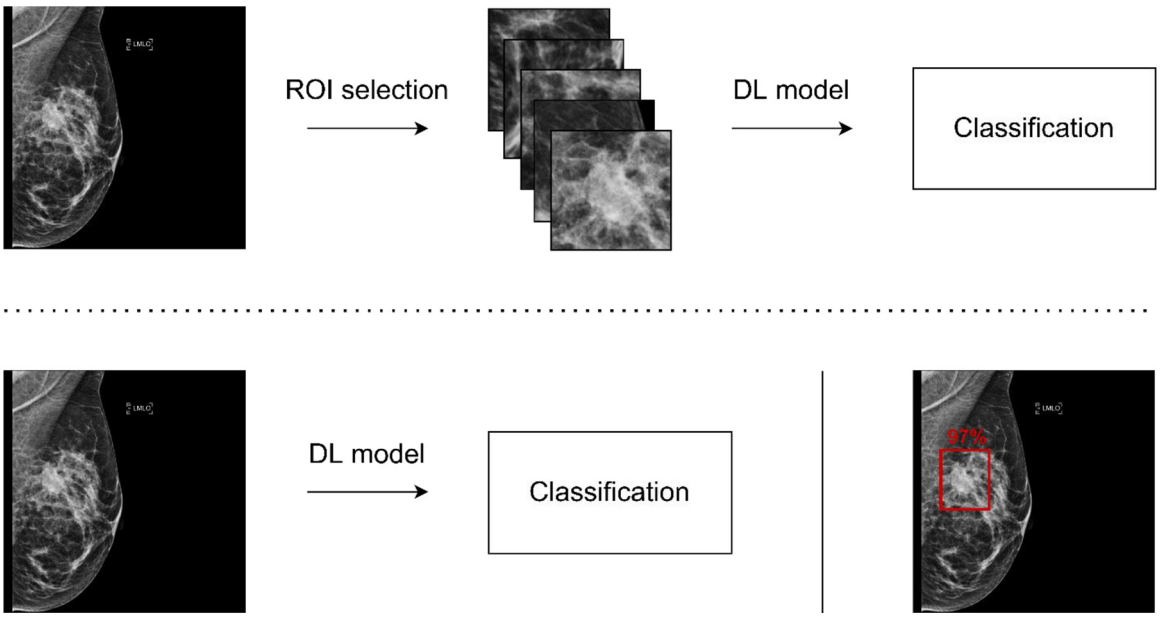

Figure 3 Two methods of using deep learning (DL) models for classification. Top: region of interests (ROIs) is selected first by hand or with conventional methods after which the DL model classifies the ROIs. Bottom: The DL model directly classifies the mammogram without ROI pre-selection. Bottom right: An example of the outcome of a DL model which classifies this lesion as malignant, with a certainty of 96%.

directly on mammograms instead of patches of suspicious lesions. When they compared the classification accuracy of their network with the accuracy of a CNN trained on suspicious patches, similar results are obtained: 97,0% (YOLO-based) vs 96,7%.

To make well-considered decisions about malignancy, radiologists not only use the image of the breast containing a suspicious lesion but also look at bilateral differences (differences between the left and right breast) and temporal differences (differences between last and previous scans). In an attempt to improve DL CAD systems, several authors created models which were capable of handling multiple input images (to include previous scans) and/or finding differences between the left and right breasts.[17-21] Kim et al.,[17] using DL and bilateral difference to classify DBT volumes, obtained an AUC of 0.847 with their 3D multi-view DCNN for classifying masses. This was an improvement relative to a CAD with hand-crafted features, which achieved an AUC of 0.826 ($P = 0.0102$). Yang et al.[21] even included the possibility to detect ipsilateral (within the same breast) differences by using different views of the same breast as input. Their network achieved an AUC of 0.962 on their in-house dataset, which was a higher performance than single-view networks on the same dataset (AUC: 0.90-0.91).

A big downside of DL is its need for large datasets to achieve satisfactory performance. However, transfer learning, learning by using images from a different modality or field, can be used to overcome this problem. Huynh et al.[13] were 1 of the first to use transfer learning in breast cancer imaging by using a network that was already trained on ~1M non-medical images from ImageNet. This type of transfer learning is nowadays an almost common practice due to its proven benefit. Similarly, due to the large similarities between digitized screen-film mammograms (SFM), DM scans, and DBT scans, DCNNs can be trained on a combination of these modalities, creating larger datasets and resulting in better performing CADs. For example, it was shown that using transfer learning between digitized SFM and DM, improved the AUC of DM classification statistically significantly.[22,23] In another study, it was shown that the AUC could be significantly increased from 0.81 to 0.90 ($P < 0.05$) by training on both SFM and DBT images vs training on only DBT images.[24]

Although some papers compare their DL CAD results with conventional CAD, the best way to evaluate new DL algorithms is by comparing them with the highest standard there is: radiologists. Some papers already included such comparisons in their initial evaluations,[11,19] while other researchers set up special multi-reader multi-case studies specifically to evaluate already existing AI systems.[25-28] The outcome of these reader studies is the same: AI is statistically non-inferior to radiologists for detecting and classifying breast lesions on DM and DBT images. For example, Romero-Martin et al.[28] investigated the stand-alone performance of an AI system for breast cancer screening on both DM and DBT scans. Their results show that the AI system can achieve comparable sensitivity as radiologists in a DM screening setting, while decreasing the recall rate. The performance of the AI system on DBT scans was found to be non-inferior to radiologists, although recall rate increased.

These studies show the great potential for AI in DM and DBT. It however also questions whether AI could totally take over the radiologists' work or that it should be used as support. This support can be as an indicator of suspicious lesions (ie, providing CAD-marks)[29-32] or for preselection (reducing the amount of normal exams that need to be read)[33,34] to improve radiologist performance and reduce workload. Rodriguez-Ruiz et al.[29] implemented AI support by providing the radiologists with an interactive tool that revealed the local cancer likelihood score calculated by the AI when clicked upon a region within the breast. This resulted in an improved AUC (ΔAUC: 0.02, $P = 0.002$), increased sensitivity (+0.03%, $P = 0.046$) and similar specificity and reading time. To decrease reading time, Yala et al.[34] reported a normal exam reduction by using an AI system that triaged ~20% of the test mammograms as cancer-free. This reduction in workload also improved specificity (+0.07%, $P = 0.002$) while maintaining a non-inferior sensitivity.

Next to (support in) detection and classification, in the last few years AI has also been used to develop more accurate risk prediction models.[35,36] Whereas in the past most risk prediction models were based on amongst others mammographic density, these models use the actual mammograms to estimate whether a patient would develop cancer in the future. For example, Dembrower et al.[35] achieved an AUC of 0.65 compared to density-based models (0.57-0.60, $P < 0.001$), showing that DL models can be useful in breast cancer risk prediction.

However, the role of AI in DM and DBT continues to develop. As outside the medical world new DL models are developed, researchers are seeing possibilities for these new models to change breast cancer imaging. One of these models is the generative adversarial network (GAN), which can generate new images which have similar distributions as the images in the training set if trained properly. For example, GANs are now used in breast cancer imaging to generate synthetic digital mammograms (SDM) from DBT scans to decrease radiation dose[37] or to detect very small microcalcifications (at most 14 pixels wide) for which it is difficult to extract efficient features.[38]

Ultrasound

Similar to DM and DBT, the first use of DL in US was for classifying tumors as either benign or malignant.[39-41] For example, in 2017, Han et al.[40] trained a modified GoogLeNet CNN on ROIs of suspicious lesions in US images, which achieved already outstanding performance with an AUC of 0.960. A conventional CAD based on hand-crafted features achieved an AUC of 0.90 on the same dataset. DL was later used to classify directly on the whole US image.[42-44] This resulted in similar AUCs compared to ROI classification, ranging from 0.84 to 0.95.

However, datasets of US are in general smaller than DM sets, mainly due to the fact that US is not as extensively used for screening purposes. Apart from transfer learning, a well-known other or complementary method to enhance performance with small datasets is data augmentation. Data augmentation is the process of creating `extra' data from existing data by rotating, flipping, shifting, etc. the available data. Actually, most mentioned research in this review makes use of a form of data augmentation. However, one should be careful which types of augmentation to use in US. As Byra et al.[41] pointed out, performing image rotation or shift in the longitudinal direction can alter known attributes of breast masses, and thus lead to a potential decrease in classification performance. This is due to the fact that lesions that are "taller than wide" (ie, vertically oriented) are more often malignant than lesions that are "wider than tall".

Instead of classifying suspicious lesions between benign and malignant, radiologists classify lesions using the Breast Imaging-Reporting and Data System (BI-RADS).[45] This classification system ranges from BI-RADS 1 (no suspicious findings) to BI-RADS 5 (>95% probability of malignancy), thus incorporating a probability of malignancy and with that a management recommendation. For lesions classified as BI-RADS 3 watchful waiting is deemed appropriate as the probability of malignancy is less than 2%, whereas lesions classified as BI-RADS 4 and 5 in general should be biopsied. Despite this system, the classification in breast US suffers from inter-observer variability, and miss-classification of lesions. Therefore, CADs are developed to help radiologists select the appropriate BI-RADS classification. Some of these are based upon pathological classification, whereas others are based upon radiologists assessment.[46,47] However, the latter approach is a bit questionable: The training data for these CADs already includes inter-observer variability, and potential structural misclassification of the likelihood of malignancy.

In US, besides lesion classification, DL is also used for lesion detection. For detection, relevant images are fed into the CAD system that subsequently determines whether a lesion is present. This can be done using normal US images[48] or 3D Automated Breast US scans.[49,50] The latter has the advantage of capturing standardly the entire breast. In normal US imaging, however, an experienced US doctor needs to manually store and select images for CAD analysis, which implies that the lesion is in fact usually already detected. This is thus dependent on the availability of experienced US operators and likely leads to missed suspicious regions. Zhang et al.,[51] however, developed a lightweight (low on GPU usage) CAD system that can detect lesions in real-time during US scanning, obtaining a 24 frames per second (fps) detection rate on US equipment.

Detection and classification can, as in DM, be combined into 1 single network.[47,52,53] A very elaborate experimental study was conducted to explore which well-known CNN-based DL architectures achieved the highest performance in US detection and classification.[52] Here, DenseNet appeared to be a good model with an accuracy of up to 85% for full images classification and 87.5% on pre-defined ROIs. Alternatively, Shin et al.[53] used semi-supervised learning to train a different network, based upon another well-known model called Faster R-CNN, for the 2 tasks. Results show that training a network with both sparsely and highly annotated images can greatly boost performance compared to training on only the (same amount of) highly annotated images.

DL can also be used to perform segmentation of lesions in handheld US images,[54-57] and 3D Automated Breast US images, to document their size and extent.[58,59] A common network used for segmentation is the U-net,[60] a U-shaped network especially developed for the segmentation of biomedical images (Fig. 4). It is shown that the U-net also delivers in breast cancer imaging, as half of the earlier mentioned references regarding segmentation use a (modified) U-net. However, newly developed segmentation networks outside the biomedical imaging domain, like the earlier mentioned YOLO network, also show their great potential for this field. Such networks are approaching clinical practice as they would detect lesions, segment them and classify them as benign or malignant.

In analogy to the DL CADs system for DM, the effect of CADs in the interpretation of US images also needs to be evaluated. Various studies validate the performance of S-detect, a CAD system of which the model of Han et al.[40] is a DL component of.[61-63] All studies show that a DL CAD system can improve the performance of radiologists. This improvement is logically mostly due to an increased specificity as the system is used to help classify already detected lesions. However, compared to DM, studies are mostly small single center evaluations, and consequently there is a need for larger retrospective/multi-reader/multi-case studies and prospective validation to obtain clear results.

Upon observing a suspicious lesion in the breast, it is common for radiologists to check for metastasis in the axillary lymph nodes with US imaging. Therefore, DL is also used to classify lymph nodes in such axillary US images.[64] Analyzing lymph nodes of breast cancer patients with a DL based network was shown to yield an increased accuracy compared to expert radiologists from 77.9% to 87,0%, and may thus help selecting the most suspicious lymph nodes for biopsy. It is also known that breast US characteristics can be associated with lymph node metastasis, opening the possibility for DL to predict lymph node metastasis directly from observing US images of breast lesions only.[65,66] Zhou et al.[66] trained a well-known CNN called InceptionV3 on US breast cancer images and compared its results with 5 radiologists, which too only had access to the breast cancer images. The network obtained both a higher sensitivity (85% vs 73%) as a higher specificity (73% vs 63%), although not statistically significant ($P > 0.05$) due to a small test set. In this way, DL can provide an alternative for checking lymph nodes for metastasis during examinations, lowering the total duration of these examinations.

DL can also be used for the prediction of molecular subtypes of breast cancer[67] and the prediction of response to neoadjuvant chemotherapy (NAC).[68] Predicting molecular subtypes can reduce unnecessary biopsies when the predictive ability of the model is at a high level and may be used to detect radiological-pathological discordance. Zhang et al.[67] obtained AUC values of 0.864, 0.811 and 0.837 for the

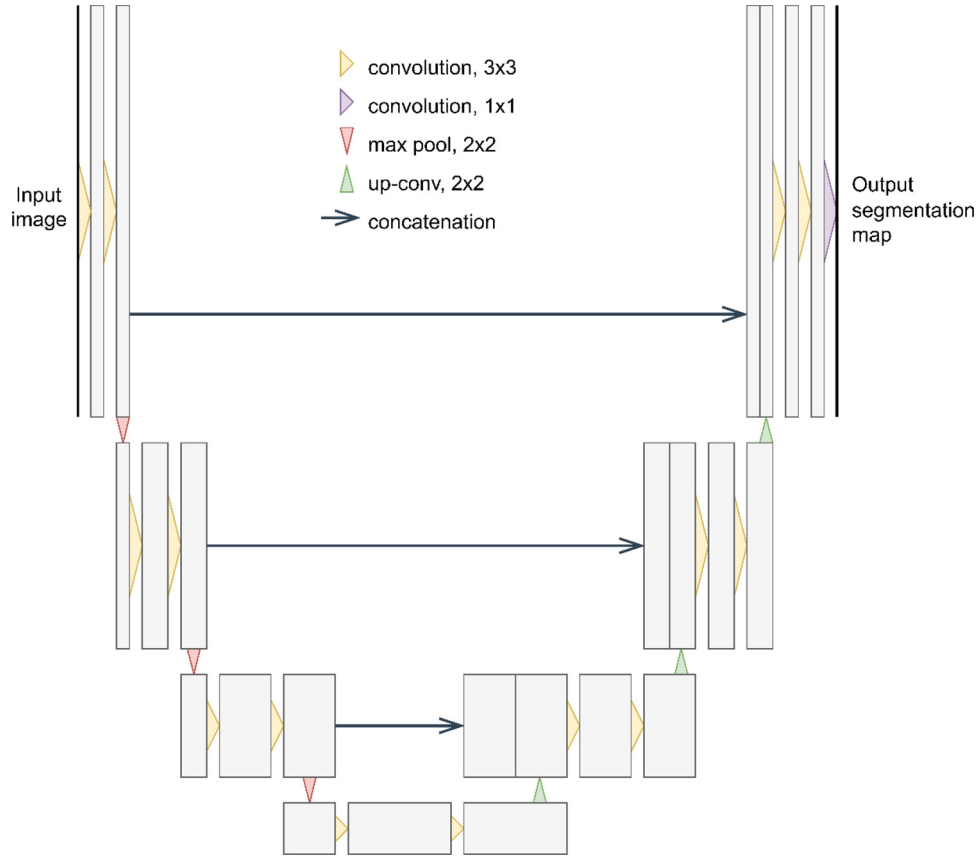

Figure 4 Schematic of U-net architecture. Grey boxes denote multi-channel feature maps; width indicates the number of feature maps, length the size of the image (in 2D). The triangles/arrow show the different operations, note that up-convolution (green triangle) halves the number of feature maps.

triple-negative, HER2 (+) and HR (+) subtype predictions, respectively, using standard ultrasound images of cancers as input. The early prediction of NAC can be crucial for the patient's health, for example, to stop ineffective chemotherapy treatment as early as possible. Byra et al.[68] developed 2 CNN models for NAC prediction: an *a priori* model and a Siamese model for prediction after the first or second course of NAC. The *a priori* model only sees the pre-NAC US images, while the Siamese model simultaneously looks at the pre-NAC and post NAC (after first or second course of NAC) US images. They obtained an AUC of 0.797 and 0.847, respectively.

GANs are also being used in US imaging. For example, GANs are used to reconstruct high-resolution images from low-resolution images,[69] lowering the 3D acquisition time and thus the burden on the patient. Furthermore, Fujioka et al.[70] implemented a GAN to generate synthetic US images with and without tumors, which can be used for strengthening DL models and educating radiologists.

Magnetic Resonance Imaging

In magnetic resonance imaging (MRI), similar to DM, DBT and US, DL is mostly used to perform or aid in classification, detection, and segmentation of breast lesions. However, the big difference between these modalities and MRI is the dimensionality. Whereas DM, DBT, and US produce 2D images, MRI creates 3D scans. In addition, MRI sequences that also observe changes over time, for instance, the inflow or outflow of contrast agents (dynamic contrast-enhanced (DCE) MRI), further expand the dimensionality to 4D. Problems can arise when using DL models created outside the medical world on these 3D or 4D MRI scans, as most of those models are developed to operate on 2D images.

Various approaches have been introduced to overcome these problems. The most commonly used approach is to transform the 3D images into 2D ones, enabling the use of standard 2D DL models. This can be done by either dividing the 3D MR image into 2D slices or by using the maximum intensity projection (MIP). However, many standard DL models are developed for color images, for example, images with 3 channels for red, green and blue (RGB). This means that the input to these models is a 3D image, in which the third dimension is created by the 3 color channels. As MR scans are greyscale images, they consist of only 1 color channel, and thus 3 slices or MIPs can be used as 1 input image. This creates the possibility to have a semi 3D MRI input image with 3 consecutive slices, or to include multiple post-contrast slices or MIPs in 1 input image. Other approaches include the use of actual 3D MRI scans and modify already existing 2D DL models to deal with 3D data, or use models

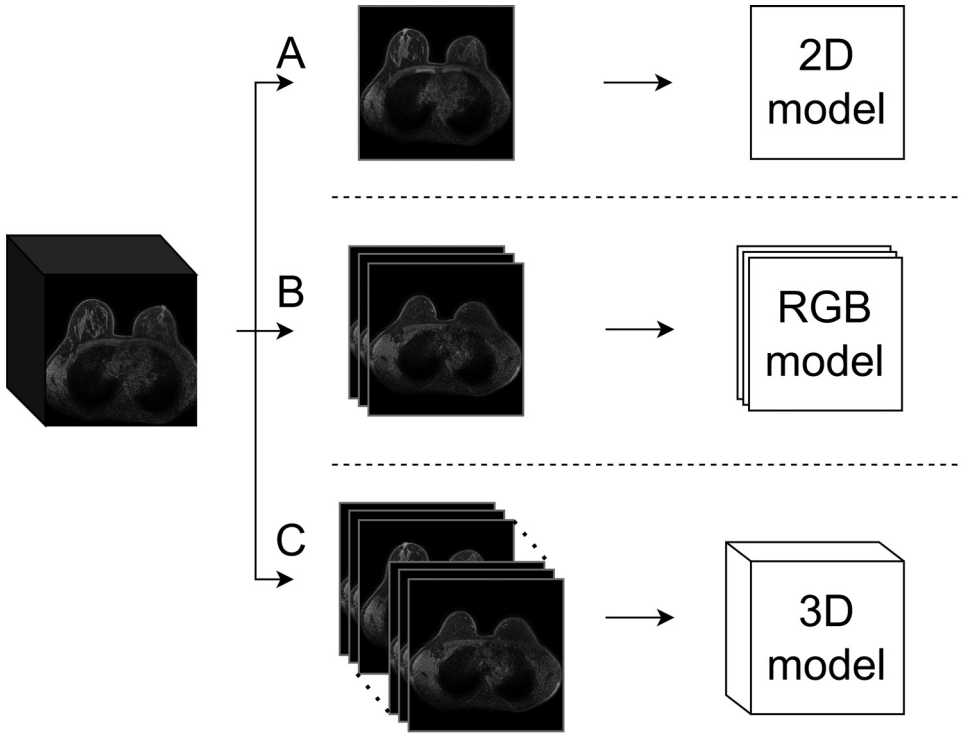

Figure 5 Different approaches to apply deep learning (DL) to 3D magnetic resonance (MR) scans. (A) 2D model approach. One 2D slice is input to the 2D DL model. (B) semi 3D (RGB) model approach. Three 2D slices are the input to the RGB model. (C) 3D model approach. Whole MR scan is directly fed into the 3D model.

like DenseNet that are specifically designed to work with 3D data (Fig. 5). It is expected that AI algorithms that can leverage the real 3D or even 4D input of breast MRI may increase performance over techniques that first apply dimension reduction, albeit so far this hasn't been shown.

Research into lesion classification in MRI has in fact made use of all of the above-mentioned approaches. Several groups used 2D slices of the ROIs as input.[71-73] They obtained AUCs ranging from 0.908 to 0.991. Other studies used MIPs as input,[74,75] finding AUC values of 0.88 and 0.895, respectively. In research where the 3 RGB channels were used to include several post-contrast slices,[76-78] the researchers obtained AUCs ranging from 0.84 to 0.92. Finally, some studies used the actual 3D MRI scans finding AUC values of 0.852 and 0.859, respectively.[78,79] Areas under the curve for the different studies are similar and may even seem to decline when moving from simple 2D approaches towards more sophisticated 3D techniques, however, it needs to be noted that these values cannot be compared with each other, as all of the mentioned research made use of different datasets. However, studies using various kind of techniques did compare their own results with radiologists interpretation.[75,77-79] In general, AI models had an improved specificity relative to radiologists' performance, and an equal or slightly lower sensitivity.

Next to lesion classification, DL is also been used in MRI to classify axillary lymph node metastasis.[80-82] Two of these studies used positron emission tomography (PET), as ground truth instead of biopsy results. They reasoned that a biopsy, while definitive as ground truth, leaves behind artifacts such as needle marks or biopsy clips, which could undesirably move the DL algorithm in the direction of a malignant classification.[81,82] By using PET as ground truth, which is also in clinical practice regularly used to determine the presence of pathological lymph nodes, they obtained accuracies of 84,8% and 88,5% compared to radiologists' performance of 78,0% and 65,8% respectively.

As in the other modalities, detection and segmentation of lesions are the other 2 tasks in which DL is already widely developed. Maicas et al.[83] were 1 of the first to build a DL-based lesion detection model for MRI. They created a network based on deep Q-networks (DQNs), which learned how to iteratively change an initial large bounding box to a smaller box containing a lesion, if this lesion exists in the initial bounding box of course. With their DQN, they achieved a similar performance compared to conventional exhaustive search methods, while significantly reducing the computation time with a factor >4. Ayatollahi et al.[84] took a different approach by modifying RetinaNet, a network specifically designed to detect small objects in images, for lesion detection on 4D MR scans. They achieved a sensitivity of 95% at 4 false positive per breast.

However, most research into lesion detection and segmentation implemented the earlier mentioned U-net.[85-88] For example, Zhang et al.[86] implemented a multi-stage U-net and filled the 3 RGB channels with slices of the breast mask, post-contrast and subtraction images, obtaining a DSC of 72 for lesion segmentation. Alternatively, Piantadosi et al.[87] used the 3 channels to include the first and second post-contrast slides next to the pre-contrast slide as input, obtaining a

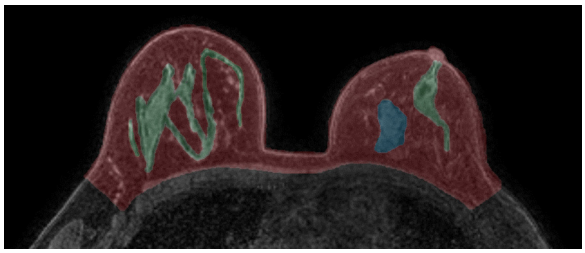

Figure 6 Example of segmented breasts in magnetic resonance (MR) scan of the patient from Figure 1. Red region: whole-breast segmentation. Green region: fibroglandular tissue. Blue region: Malignant lesion.

DSC of 61 for segmentation. In another approach with the U-net, Lu et al.[88] merged 4 different MRI sequences, namely T1-weighted, T2-weighted, diffusion weighted and dynamic-contrast-enhanced MRI, to achieve a DSC of 0.865.

It is important to note that in half of the above-mentioned detection and segmentation methods for MRI first a proper breast mask was used. These breast masks are segmentations of the whole breast region, which, when multiplied with an MRI scan, only pass-through pixels that belong to the breasts, while pixels outside the breasts are set to zero. These masks serve 2 purposes. Firstly, in the DL models for segmentation and detection, they can guide the models to focus on the right place. Secondly, the masks can be used to calculate the mammographic density, which is an important indicator for an increased risk of breast cancer. It is, therefore, no surprise that DL also plays a role in whole breast segmentation.[89-91] In these studies, Dice scores between 96 and 99 were obtained, showing that whole breast segmentation is an easier task than lesion segmentation.

Mammographic density is the percentage of fibroglandular tissue (FGT) in the breast region. This can be calculated when one has the breast mask and the segmented FGT. DL is also used to segment the FGT,[92-94] enabling the possibility to calculate the mammographic density without human input. For example, Dalmış et al.[92] compared their DL method using a U-net with older atlas-based methods, respectively finding DSCs of 0.850 vs 0.671. In another research, Parekh et al.[95] combined all of the tasks stated above, and segmented the breast, FGT, and lesion (if present) with 1 DL model, after which a standard ML classifier classified the lesion as benign or malignant. Their network obtained a DSC of 0.89 for segmenting lesions and a classification AUC value of 0.90 (Fig. 6).

Next to the 3 main tasks, namely, classification, detection, and segmentation, DL in MRI is also commonly used for the prediction of various variables. Multiple studies are investigating the possibility of predicting the breast cancer molecular subtype.[96-100] In one of these studies, Sun et al.[99] tried to distinguish malignant lesions into 2 kinds of breast cancer subtypes, namely luminal or non-luminal. With their CNN based on the well-known Inception-v3 structure, they obtained an AUC of 0.958. The study of Liu et al.[100] is another study worth mentioning, as they tried to predict the recurrence of cancer (within 5 years) next to predicting the molecular subtype (HER2+). An AUC value of 0.781 was found for molecular subtype prediction, while the prediction of recurrence had an even higher AUC value of 0.828.

Likewise, DL can be used to predict the response to neoadjuvant chemotherapy (NAC). This has great potential in reducing the physical and financial burden of treatments, as a good NAC response prediction may be used to personalize the NAC before the treatment has even started. There exist studies that use a combination of DCE-MRI scans from before and early in the NAC,[101,102] and studies that only use DCE-MRI scans from before NAC treatment.[103,104] Braman et al.[104] performed a retrospective multi-center study across 5 institutions to predict the pathological complete response (pCR) to NAC for HER2+ breast cancer patients. They found AUC values of 0.77 and 0.85 for 2 different testing datasets, which is a statistically better performance than conventional pCR NAC prediction methods, showing the potential for DL in these kinds of tasks.

GANs are also found to be useful in MRI. Since there exist differences in noise distribution and intensity values between MR scans from different vendors, models trained on 1 brand of scanners may struggle to generalize to another brand, causing a decrease in performance. To normalize the MR scans of different vendors, Modanwal et al.[105] implemented a GAN which learned the bidirectional mapping between MRI scans of 2 different vendors. They showed the potential role of DL in MRI normalization by obtaining DSCs of 0.98 (in both directions). Mori et al.[106] investigated a different type of GAN called pix2pix,[107] to solve another problem present in MRI, namely fat suppression in DCE-MRI images. The current fat suppression method is often inhomogeneous, mostly caused by an inhomogeneous magnetic field inside the MRI scanner. They generated synthetic fat-saturated T1-weighted images from non-contrast-enhanced T1-weighted images by training pix2pix to learn the mapping between those 2 kinds of images. Two breast radiologists scored the synthetic images on a scale of 1 (excellent) to 5 (poor), with 3 being equal to current fat-saturated T1-weighted images. On average, the synthetic images got a score of 3.12, showing again the potential role of DL in another area of breast cancer imaging.

Nuclear Medicine Imaging

Nuclear medicine imaging techniques like Positron Emission Tomography (PET) or scintigraphy are deemed less suitable for early-stage breast cancer evaluation as compared to the 4 earlier discussed imaging modalities, that is, DM, DBT, US and MRI. However, nuclear medicine techniques have added value for detection and classification of (extra-)axillary lymph nodes and distant staging.[108] It is therefore no surprise that DL is also implemented in this imaging area, albeit only marginally.

The quantification of the whole-body metabolic tumor volume (MTV) from PET/CT can give a measure of the tumor burden. If this task could be performed by a DL model, it greatly reduces the manual labor as in clinical practice all tumors need to be delineated to obtain the MTV. Weber

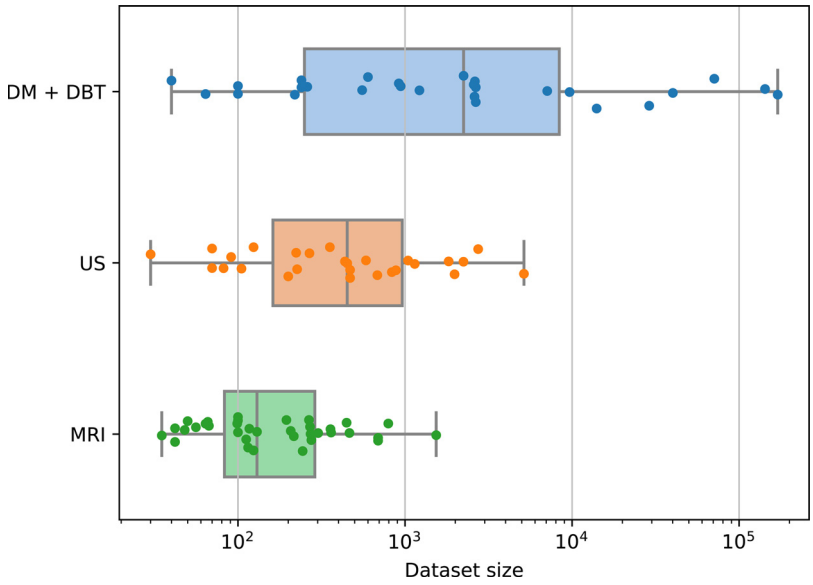

Figure 7 Dataset sizes per breast imaging modality for all research included in this review. DBT, digital breast tomosynthesis; DM, digital mammography; US, ultrasound; MRI, magnetic resonance imaging.

et al.[109] tested if a CNN, trained to obtain the MTV for lymphoma and lung cancer patients, could also accurately detect and segment lesions in whole-body PET/CT scans of breast cancer patients. They found that there was a statistically significant correlation between the DL derived and manually segmented MTV, showing the generalizability of this neural network. However, the DL model was having trouble detecting all lesions, although it detected 92% of PERCIST measurable lesions, the sensitivity for all detected lesions was only 39%.

Breast cancer and prostate cancer are the cause of most cases of bone metastasis.[110] Therefore, Papandrianos et al.[111] proposed a robust CNN for diagnosing breast cancer metastasis on whole-body scintigraphy scans. They obtained a classification accuracy of 92.5% by labelling the whole-body scans as either benign or malignant. In addition, they compare their proposed network with well-known CNN architectures, from which DenseNet achieved the highest accuracy of 95%.

Furthermore, DL and nuclear medicine techniques are combined to improve tasks which are also implemented in the other breast cancer imaging techniques. Li et al.[112] created a 3D CNN model to assist clinicians in diagnosing axillary lymph node metastasis on PET/CT images. With the aid of their network, the sensitivity of the clinicians increased on average with 7.8 percentage points, while the specificity of the clinicians remained unaffected (at ~99.0%). The DL model on its own, however, was still outperformed by both clinicians. In another study, Choi et al.[113] showed that PET/MRI scans can be used to train a DL model to predict pathological responses to NAC in patients with advanced breast cancer. Here, improved AUC values were found when comparing conventional indicators of response to NAC (eg, maximum standardized uptake value, metabolic tumor volume, etc.) to the CNN based on Alexnet. It needs to be said that their dataset was very limited, with only 6 responders and fifty non-responders.

Perspectives on AI and DL in Clinical Breast Imaging Practice

Although the roles of AI and DL in the different imaging modalities are comparable, the difference in the size of the datasets, and level of validation and evaluation of the studies differ much. Whereas DM and DBT studies nowadays have large datasets in the order of (tens of) thousands of patients from different medical centers, MRI studies do only rarely exceed 500 patients and are often from a single center. The size of US datasets is in between the other 2 modalities dataset sizes, in the order of some thousands of patients (Fig. 7). This of course favors the AI performance in DM and DBT studies, as generally larger datasets and data from different centers lead to better performing and better generalizing DL models. And again, when looking at the level of validation and evaluation per modality, DM and DBT studies are leading. Whereas there already exist multiple large retrospective and multi-reader studies for the evaluation of DL CAD systems for DM and DBT, these do exist less for US, while, to the best of our knowledge, these do not yet exist for MRI.

DL research in US and MRI needs to invest in generating larger and more diverse datasets to elevate from proof-of-concept models to systems that are ready for large multi-reader multi-case studies, as is now the case for DM/DBT studies. However, this does not mean that all DM/DBT models are already sufficiently tested for implementation into clinical practice. As found by Gur et al.,[114] the performance of radiologists can differ between case studies and the real clinical setting, the so-called 'laboratory' effect. It is thus

necessary to start up large-scale prospective trials to evaluate the performance of DL systems and the effect they have on human performance in clinical practice before one can undoubtedly show that DL systems can have a beneficial role in breast cancer imaging.

An obvious solution to overcome small datasets and grow large ones quickly, is to share data across medical centers. Unfortunately, this is often not possible due to patient-privacy policies. Approaches such as federated learning, where data remains locally, but the algorithm travels,[115] or even swarm learning, where all participants contribute to both case collection and algorithm development,[116] are poised to overcome this problem. However, thus far such approaches have not been widely implemented. This also induces another problem: validation of the exact results of DL studies in breast cancer imaging is generally not possible as (training) data cannot be shared. However, Norgeot et al.[117] argued that exact validation of the results is of less importance compared to direct assessment of clinical impact of AI systems and rapid replication of the study. Therefore, they set up a checklist, indicating the minimal requirements for the transparent reporting of clinical AI studies. Other organizations provided similar checklists, that may enable more robust and consistent AI studies in the future, which is essential before they can be widely used.

At last, next to performance evaluation, legal and ethical issues need to be answered as well. Can final decisions be made by the DL CAD systems? Who has the liability for incorrect DL decisions? Will AI aid give a (negative) bias to radiologists? What is the public opinion of DL decision tools? Can DL CAD algorithms accurately clarify how they came to their decision? It is clear that these questions need to be discussed before DL models can be widely implemented in real clinical settings.

Conclusion

There are numerous possibilities for DL algorithms to elevate breast cancer imaging to a higher level. For all common breast cancer imaging modalities (eg, DM/DBT, US and MRI) many studies exist that use AI. Studies range from classification to segmentation, and from prediction to image generation. DL studies in MRI and PET/CT show a slightly greater variety in research directions compared to DM/DBT and US, mainly due to a stronger focus on prediction of therapy response and outcome. DM/DBT studies are further developed due to larger dataset sizes and larger validation studies. However, it is clear that large trials still need to be conducted and legal and ethical issues need to be considered before the role of AI and DL will expand to its full potential in the clinical breast cancer imaging practice.

References

1. Sung H, Ferlay J, Siegel RL, et al: Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer J Clin 71(3):209-249, 2021. https://doi.org/10.3322/caac.21660
2. Saadatmand S, Bretveld R, Siesling S, et al: Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173 797 patients. BMJ 351:h4901, 2015. https://doi.org/10.1136/bmj.h4901
3. Wing P, Langelier MH: Workforce shortages in breast imaging: impact on mammography utilization. Am J Roentgenol 192(2):370-378, 2009. https://doi.org/10.2214/AJR.08.1665
4. Litjens G, Kooi T, Bejnordi BE, et al: A survey on deep learning in medical image analysis. Med Image Anal 42:60-88, 2017. https://doi.org/10.1016/j.media.2017.07.005
5. Lehman CD, Wellman RD, Buist DSM, et al: Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Inter Med 175(11):1828-1837, 2015. https://doi.org/10.1001/jamainternmed.2015.5231
6. Krizhevsky A, Sutskever I, Hinton GE: Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst: 25, 2012
7. Mainprize JG, Alonzo-Proulx O, Jong RA, et al: Quantifying masking in clinical mammograms via local detectability of simulated lesions. Med Phys 43(3):1249-1258, 2016. https://doi.org/10.1118/1.4941307
8. Skaane P, Bandos AI, Gullien R, 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 23(8):2061-2071, 2013. https://doi.org/10.1007/s00330-013-2820-3
9. Lehman CD, Schnall MD: Imaging in breast cancer: Magnetic resonance imaging. Breast Cancer Res 7(5):215, 2005. https://doi.org/10.1186/bcr1309
10. Goodfellow I, Bengio Y, Courville A: Deep Learning. Cambridge, MA: MIT Press, 2016
11. Kooi T, Litjens G, van Ginneken B, et al: Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 35:303-312, 2017. https://doi.org/10.1016/j.media.2016.07.007
12. Samala Ravi K, Chan Heang-Ping, Hadjiiski Lubomir M, Cha Kenny, Helvie Mark A: Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis. In: In: Proc. SPIE 9785, Medical Imaging 2016: Computer-Aided Diagnosis, 97850Y (24 March 2016) 2016. https://doi.org/10.1117/12.2217092
13. Huynh BQ, Li H, Giger ML: Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. JMI 3(3):034501, 2016 https://doi.org/10.1117/1.JMI.3.3.034501
14. Lotter W, Sorensen G, Cox D, et al: A multi-scale cnn and curriculum learning strategy for mammogram classification. In: Cardoso M, et al (ed): Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer, Cham, 10553, 2017. https://doi.org/10.1007/978-3-319-67558-9_20
15. Al-Masni MA, Al-Antari MA, Park J-M, et al: Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Comput Methods Programs Biomed 157:85-94, 2018. https://doi.org/10.1016/j.cmpb.2018.01.017
16. Zhang X, Zhang Y, Han EY, et al: Classification of Whole Mammogram and Tomosynthesis Images Using Deep Convolutional Neural Networks. IEEE Trans Nanobioscience 17(3), 2018. https://doi.org/10.1109/TNB.2018.284510
17. Kim DH, Kim ST, Ro YM: Latent feature representation with 3-D multi-view deep convolutional neural network for bilateral analysis in digital breast tomosynthesis. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP):927-931, 2016. https://doi.org/10.1109/ICASSP.2016.7471811
18. Kooi T, Karssemeijer N: Classifying symmetrical differences and temporal change for the detection of malignant masses in mammography using deep neural networks. J Med Imaging (Bellingham) 4(4), 2017. https://doi.org/10.1117/1.JMI.4.4.044501
19. Wu N, Phang J, Park J, et al: Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Trans Med Imaging 39(4), 2020. https://doi.org/10.1109/TMI.2019.2945514

20. Loizidou K, Skouroumouni G, Pitris C, et al: Digital subtraction of temporally sequential mammograms for improved detection and classification of microcalcifications. Eur Radiol Exp 5(1), 2021. https://doi.org/10.1186/s41747-021-00238-w
21. Yang Z, Cao Z, Zhang Y, et al: MommiNet-v2: Mammographic multi-view mass identification networks. Med Image Anal 73:102204, 2021. https://doi.org/10.1186/s41747-021-00238-w
22. Samala RK, Chan H-P, Hadjiiski LM, et al: Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. Phys. Med. Biol. 62(23), 2017. https://doi.org/10.1088/1361-6560/aa93d4
23. Agarwal R, Diaz O, Lladó X, et al: Automatic mass detection in mammograms using deep convolutional neural networks. JMI 6 (3):031409, 2019. https://doi.org/10.1117/1.JMI.6.3.031409
24. Samala RK, Chan H-P, Hadjiiski L, et al: Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. Med Phys 43(12):6654-6666, 2016. https://doi.org/10.1118/1.4967345
25. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al: Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists. J Natl Cancer Inst 111(9), 2019. https://doi.org/10.1093/jnci/djy222
26. McKinney SM, Sieniek M, Godbole V, et al: International evaluation of an AI system for breast cancer screening. Nature 577(7788), 2020. https://doi.org/10.1038/s41586-019-1799-6
27. Kim H-E, Kim HH, Han B-K, et al: Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. The Lancet Digital Health 2(3), 2020. https://doi.org/10.1016/S2589-7500(20)30003-0
28. Romero-Martín S, Elías-Cabot E, Raya-Povedano JL, et al: Stand-alone use of artificial intelligence for digital mammography and digital breast tomosynthesis screening: A retrospective evaluation. Radiology 2021:211590. https://doi.org/10.1148/radiol.211590
29. Rodríguez-Ruiz A, Krupinski E, Mordang J-J, et al: Detection of breast cancer with mammography: Effect of an artificial intelligence support system. Radiology 290(2), 2019. https://doi.org/10.1148/radiol.2018181371
30. Benedikt RA, Boatsman JE, Swann CA, et al: Concurrent computer-aided detection improves reading time of digital breast tomosynthesis and maintains interpretation performance in a multireader multicase study. Am J Roentgenol 210(3):685-694, 2018. https://doi.org/10.2214/AJR.17.18185
31. Conant EF, Toledano AY, Periaswamy S, et al: Improving accuracy and efficiency with concurrent use of artificial intelligence for digital breast tomosynthesis. Radiol Artif Intell 1(4):e180096, 2019. https://doi.org/10.1148/ryai.2019180096
32. Chae EY, Kim HH, Jeong J, et al: Decrease in interpretation time for both novice and experienced readers using a concurrent computer-aided detection system for digital breast tomosynthesis. Eur Radiol 29 (5), 2019. https://doi.org/10.1007/s00330-018-5886-0
33. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al: Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. Eur Radiol 29(9), 2019. https://doi.org/10.1007/s00330-019-06186-9
34. Yala A, Schuster T, Miles R, et al: A deep learning model to triage screening mammograms: A simulation study. Radiology 293(1), 2019. https://doi.org/10.1148/radiol.2019182908
35. Dembrower K, Liu Y, Azizpour H, et al: Comparison of a deep learning risk score and standard mammographic density score for breast cancer risk prediction. Radiology 294(2), 2020. https://doi.org/10.1148/radiol.2019190872
36. Yala A, Lehman C, Schuster T, et al: A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 292(1), 2019. https://doi.org/10.1148/radiol.2019182716
37. Jiang G, Wei J, Xu Y, et al: Synthesis of mammogram from digital breast tomosynthesis using deep convolutional neural network with gradient guided cGANs. IEEE Trans Med Imaging 40(8), 2021. https://doi.org/10.1109/TMI.2021.3071544
38. Zhang F, Luo L, Sun X, Zhou Z, Li X, Yu Y, Wang Y: Cascaded generative and discriminative learning for microcalcification detection in breast mammograms. In: In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition:12578-12586, 2019
39. Shi J, Zhou S, Liu X, et al: Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset. Neurocomputing 194:87-94, 2016. https://doi.org/10.1016/j.neucom.2016.01.074
40. Han S, Kang H-K, Jeong J-Y, et al: A deep learning framework for supporting the classification of breast lesions in ultrasound images. Phys. Med. Biol. 62(19), 2017. https://doi.org/10.1088/1361-6560/aa82ec
41. Byra M, Galperin M, Ojeda-Fournier H, et al: Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion. Med Phys 46(2), 2019. https://doi.org/10.1002/mp.13361
42. Becker AS, Mueller M, Stoffel E, et al: Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: A pilot study. Br J Radiol 91(1083), 2018. https://doi.org/10.1259/bjr.20170576
43. Fujioka T, Kubota K, Mori M, et al: Distinction between benign and malignant breast masses at breast ultrasound using deep learning method with convolutional neural network. Jpn J Radiol 37(6), 2019. https://doi.org/10.1007/s11604-019-00831-5
44. Tanaka H, Chiu S-W, Watanabe T, et al: Computer-aided diagnosis system for breast ultrasound images using deep learning. Phys Med Biol 64(23), 2019. https://doi.org/10.1088/1361-6560/ab5093
45. D'Orsie CJ, Mendelson EB, Morris EA: ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology, 2013
46. Ciritsis A, Rossi C, Eberhard M, et al: Automatic classification of ultrasound breast lesions using a deep convolutional neural network mimicking human decision-making. Eur Radiol 29(10), 2019. https://doi.org/10.1007/s00330-019-06118-7
47. Huang Y, Han L, Dou H, et al: Two-stage CNNs for computerized BI-RADS categorization in breast ultrasound images. Biomed Eng Online 18 (1), 2019. https://doi.org/10.1186/s12938-019-0626-5
48. Yap MH, Pons G, Martí J, et al: Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Inform 22(4), 2018. https://doi.org/10.1109/JBHI.2017.2731873
49. Chiang T-C, Huang Y-S, Chen R-T, et al: Tumor detection in automated breast ultrasound using 3-D CNN and prioritized candidate aggregation. IEEE Trans Med Imaging 38(1), 2019. https://doi.org/10.1109/TMI.2018.2860257
50. Moon WK, Lee Y-W, Ke H-H, et al: Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. Comput Methods Programs Biomed 190:105361, 2020. https://doi.org/10.1016/j.cmpb.2020.105361
51. Zhang X, Lin X, Zhang Z, et al: Artificial intelligence medical ultrasound equipment: Application of breast lesions detection. Ultrason Imaging 42(4–5), 2020. https://doi.org/10.1177/0161734620928453
52. Cao Z, Duan L, Yang G, et al: An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. BMC Med Imaging 19(1), 2019. https://doi.org/10.1186/s12880-019-0349-x
53. Shin SY, Lee S, Yun ID, et al: Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. IEEE Trans Med Imaging 38(3), 2019. https://doi.org/10.1109/TMI.2018.2872031
54. Kumar V, Webb JM, Gregory A, et al: Automated and real-time segmentation of suspicious breast masses using convolutional neural network. PLOS ONE 13(5), 2018. https://doi.org/10.1371/journal.pone.0195816
55. Yap MH, Goyal M, Osman FM, et al: Breast ultrasound lesions recognition: end-to-end deep learning approaches. J Med Imaging (Bellingham) 6(1), 2019. https://doi.org/10.1117/1.JMI.6.1.011007
56. Qu X, Shi Y, Hou Y, et al: An attention-supervised full-resolution residual network for the segmentation of breast ultrasound images. Med Phys 47(11), 2020. https://doi.org/10.1002/mp.14470

57. Wang K, Liang S, Zhong S, et al: Breast ultrasound image segmentation: A coarse-to-fine fusion convolutional neural network. Med Phys 48(8), 2021. https://doi.org/10.1002/mp.15006
58. Lei Y, He X, Yao J, et al: Breast tumor segmentation in 3D automatic breast ultrasound using mask scoring R-CNN. Med Phys 48(1), 2021. https://doi.org/10.1002/mp.14569
59. Pan P, Chen H, Li Y, et al: Tumor segmentation in automated whole breast ultrasound using bidirectional LSTM neural network and attention mechanism. Ultrasonics 110:106271. https://doi.org/10.1016/j.ultras.2020.106271, 2021
60. Ronneberger O, Fischer P, Brox T: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds): Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science, 9351, 2015. https://doi.org/10.1007/978-3-319-24574-4_28
61. Choi JS, Han B-K, Ko ES, et al: Effect of a deep learning framework-based computer-aided diagnosis system on the diagnostic performance of radiologists in differentiating between malignant and benign masses on breast ultrasonography. Korean J Radiol 20(5), 2019. https://doi.org/10.3348/kjr.2018.0530
62. Park HJ, Kim SM, Yun BLa, et al: A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of breast masses on ultrasound. Med (Baltimore) 98(3):e14146, 2019. https://doi.org/10.1097/MD.0000000000014146
63. Xiao M, Zhao C, Zhu Q, et al: An investigation of the classification accuracy of a deep learning framework-based computer-aided diagnosis system in different pathological types of breast lesions. J Thorac Dis 11(12), 2019. https://doi.org/10.21037/jtd.2019.12.10
64. Coronado-Gutiérrez D, Santamaría G, Ganau S, et al: Quantitative ultrasound image analysis of axillary lymph nodes to diagnose metastatic involvement in breast cancer. Ultrasound Med Biol 45(11), 2019. https://doi.org/10.1016/j.ultrasmedbio.2019.07.413
65. Zheng X, Yao Z, Huang Y, et al: Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun 11(1):1236, 2020. https://doi.org/10.1038/s41467-020-15027-z
66. Zhou L-Q, Wu X-L, Huang S-Y, et al: Lymph node metastasis prediction from primary breast cancer us images using deep learning. Radiology 294(1), 2020. https://doi.org/10.1148/radiol.2019190372
67. Zhang X, Li H, Wang C, et al: Evaluating the accuracy of breast cancer and molecular subtype diagnosis by ultrasound image deep learning model. Front Oncol 11:623506, 2021. https://doi.org/10.3389/fonc.2021.623506
68. Byra M, Dobruch-Sobczak K, Klimonda Z, et al: Early prediction of response to neoadjuvant chemotherapy in breast cancer sonography using siamese convolutional neural networks. IEEE J Biomed Health Inform 25(3), 2021. https://doi.org/10.1109/JBHI.2020.3008040
69. Dai X, Lei Y, Wang T, et al: Self-supervised learning for accelerated 3D high-resolution ultrasound imaging. Med Phys 48(7), 2021. https://doi.org/10.1002/mp.14946
70. Fujioka T, Kubota K, Mori M, et al: Virtual interpolation images of tumor development and growth on breast ultrasound image synthesis with deep convolutional generative adversarial networks. J Ultrasound Med 40(1), 2021. https://doi.org/10.1002/jum.15376
71. Rasti R, Teshnehlab M, Phung SL: Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks. Pattern Recogn 72:381-390, 2017. https://doi.org/10.1016/j.patcog.2017.08.004
72. Zhou J, Zhang Y, Chang K-T, et al: Diagnosis of benign and malignant breast lesions on DCE-MRI by using radiomics and deep learning with consideration of peritumor tissue. J Magn Reson Imaging 51(3), 2020. https://doi.org/10.1002/jmri.26981
73. Feng H, Cao J, Wang H, et al: A knowledge-driven feature learning and integration method for breast cancer diagnosis on multi-sequence MRI. Magn Reson Imaging 69:40-48, 2020. https://doi.org/10.1016/j.mri.2020.03.001
74. Antropova N, Abe H, Giger ML: Use of clinical MRI maximum intensity projections for improved breast lesion classification with deep convolutional neural networks. *J Med Imag*ing (Bellingham) 5(1), 2018. https://doi.org/10.1117/1.JMI.5.1.014503
75. Fujioka T, Yashima Y, Oyama J, et al: Deep-learning approach with convolutional neural network for classification of maximum intensity projections of dynamic contrast-enhanced breast magnetic resonance imaging. Magn Reson Imaging 75:1-8, 2021. https://doi.org/10.1016/j.mri.2020.10.003
76. Antropova N, Huynh BQ, Giger ML: A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. Med Phys 44(10):5162-5171, 2017. https://doi.org/10.1002/mp.12453
77. Truhn D, Schrading S, Haarburger C, et al: Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast MRI. Radiology 290(2), 2019. https://doi.org/10.1148/radiol.2018181352
78. Zhou J, Luo L-Y, Dou Q, et al: Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. J Magn Reson Imaging 50(4), 2019. https://doi.org/10.1002/jmri.26721
79. Dalmis MU, Gubern-Mérida A, Vreemann S, et al: Artificial intelligence—based classification of breast lesions imaged with a multiparametric breast MRI protocol with ultrafast DCE-MRI, T2, and DWI. Invest Radiol 54(6), 2019. https://doi.org/10.1097/RLI.0000000000000544
80. Ha R, Chang P, Karcich J, et al: Axillary lymph node evaluation utilizing convolutional neural networks using MRI dataset. J Digit Imaging 31(6):851-856, 2018. https://doi.org/10.1007/s10278-018-0086-7
81. Ren T, Cattell R, Duanmu H, et al: Convolutional neural network detection of axillary lymph node metastasis using standard clinical breast MRI. Clin Breast Cancer 20(3), 2020. https://doi.org/10.1016/j.clbc.2019.11.009
82. Ren T, Lin S, Huang P, et al: Convolutional neural network of multiparametric MRI accurately detects axillary lymph node metastasis in breast cancer patients with pre neoadjuvant chemotherapy. Clin Breast Cancer 2021. https://doi.org/10.1016/j.clbc.2021.07.002. pp. S1526-8209(21)00179—8
83. Maicas G, Carneiro G, Bradley AP, Nascimento JC, Reid I: Deep reinforcement learning for active breast lesion detection from DCE-MRI. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins D, Duchesne S (eds): Medical image computing and computer assisted intervention – MICCAI 2017, Lecture Notes in Computer Science, 10435, 2017. https://doi.org/10.1007/978-3-319-66179-7_76
84. Ayatollahi F, Shokouhi SB, Mann RM, et al: Automatic breast lesion detection in ultrafast DCE-MRI using deep learning. Med. Phys.: 15156, 2021. https://doi.org/10.1002/mp.15156
85. Dalmış MU, Vreemann S, Kooi T, et al: Fully automated detection of breast cancer in screening MRI using convolutional neural networks. *J Med Im*aging (Bellingham) 5(1):014502, 2018. https://doi.org/10.1117/1.JMI.5.1.014502
86. Zhang J, Saha A, Zhu Z, et al: Hierarchical convolutional neural networks for segmentation of breast tumors in MRI with application to radiogenomics. IEEE Trans Med Imaging 38(2):435-447, 2019. https://doi.org/10.1109/TMI.2018.2865671
87. Piantadosi G, Marrone S, Galli A, Sansone M, Sansone C: DCE-MRI breast lesions segmentation with a 3TP U-Net deep convolutional neural network. In: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS):628-633, 2019. https://doi.org/10.1109/CBMS.2019.00130
88. Lu W, Wang Z, He Y, Yu H, Xiong N, Wei J: Breast cancer detection based on merging four modes MRI using convolutional neural networks. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP):1035-1039, 2019. https://doi.org/10.1109/ICASSP.2019.8683149
89. Piantadosi G, Sansone M, Sansone C: Breast segmentation in MRI via U-Net deep convolutional neural networks. In: 2018 24th International Conference on Pattern Recognition (ICPR):3917-3922, 2018. https://doi.org/10.1109/ICPR.2018.8545327
90. Xu X, Fu L, Chen Y, et al: Breast region segmentation being convolutional neural network in dynamic contrast enhanced MRI. Annu Int Conf IEEE Eng Med Biol Soc 2018:750-753, 2018. https://doi.org/10.1109/EMBC.2018.8512422

91. Piantadosi G, Sansone M, Fusco R, et al: Multi-planar 3D breast segmentation in MRI via deep convolutional neural networks. Artif Intell Med 103:101781, 2020. https://doi.org/10.1016/j.artmed.2019.101781
92. Dalmış; MU, Litjens G, Holland K, et al: Using deep learning to segment breast and fibroglandular tissue in MRI volumes. Med Phys 44 (2):533-546, 2017. https://doi.org/10.1002/mp.12079
93. Nam Y, Park GE, Kang J, et al: Fully automatic assessment of background parenchymal enhancement on breast MRI using machine-learning models. J Magn Reson Imaging 53(3), 2021. https://doi.org/10.1002/jmri.27429
94. Huo L, Hu X, Xiao Q, et al: Segmentation of whole breast and fibroglandular tissue using nnU-Net in dynamic contrast enhanced MR images. Magn Reson Imaging 82:31-41, 2021. https://doi.org/10.1016/j.mri.2021.06.017
95. Parekh VS, Macura KJ, Harvey SC, et al: Multiparametric deep learning tissue signatures for a radiological biomarker of breast cancer: Preliminary results. Med Phys 47(1), 2020. https://doi.org/10.1002/mp.13849
96. Ha R, Mutasa S, Karcich J, et al: Predicting breast cancer molecular subtype with MRI dataset utilizing convolutional neural network algorithm. J Digit Imaging 32(2), 2019. https://doi.org/10.1007/s10278-019-00179-2
97. Zhu Z, Albadawy E, Saha A, et al: Deep learning for identifying radiogenomic associations in breast cancer. Comput Biol Med 109:85-90, 2019. https://doi.org/10.1016/j.compbiomed.2019.04.018
98. Zhang Y, Chen J-H, Lin Y, et al: Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers. Eur Radiol 31(4), 2021. https://doi.org/10.1007/s00330-020-07274-x
99. Sun R, Meng Z, Hou X, et al: Prediction of breast cancer molecular subtypes using DCE-MRI based on CNNs combined with ensemble learning. Phys. Med. Biol. 66(17):175009, 2021. https://doi.org/10.1088/1361-6560/ac195a
100. Liu G, Mitra D, Jones EF, et al: Mask-guided convolutional neural network for breast tumor prognostic outcome prediction on 3D DCE-MR images. J Digit Imaging 34(3), 2021. https://doi.org/10.1007/s10278-021-00449-y
101. El Adoui M, Drisis S, Benjelloun M: Predict breast tumor response to chemotherapy using a 3D deep learning architecture applied to DCE-MRI data. Bioinform Biomed Eng Cham: 33-40, 2019. https://doi.org/10.1007/978-3-030-17935-9_4
102. Qu Y-H, Zhu H-T, Cao K, et al: Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using a deep learning (DL) method. Thorac Cancer 11(3):651-658, 2020. https://doi.org/10.1111/1759-7714.13309
103. Huynh BQ, Antropova N, Giger ML: Comparison of breast DCE-MRI contrast time points for predicting response to neoadjuvant chemotherapy using deep convolutional neural network features with transfer learning. In: Medical Imaging 2017: Computer-Aided Diagnosis, 10134:207-213, 2017. https://doi.org/10.1117/12.2255316
104. Braman N, Adoui ME, Vulchi M, et al: Deep learning-based prediction of response to HER2-targeted neoadjuvant chemotherapy from pre-treatment dynamic breast MRI: A multi-institutional validation study. arXiv:2001.08570 [cs, eess, q-bio, stat] 2020. Accessed December 15, 2021 http://arxiv.org/abs/2001.08570
105. Modanwal G, Vellal A, Mazurowski MA: Normalization of breast MRIs using cycle-consistent generative adversarial networks. Comput Methods Programs Biomed 208:106225, 2021. https://doi.org/10.1016/j.cmpb.2021.106225
106. Mori M, Fujioka T, Katsuta L, et al: Feasibility of new fat suppression for breast MRI using pix2pix. Jpn J Radiol 38(11), 2020. https://doi.org/10.1007/s11604-020-01012-5
107. Isola P, Zhu JY, Zhou T, Efros AA: Image-to-image translation with conditional adversarial networks. In: In: Proceedings of the IEEE conference on computer vision and pattern recognition:1125-1134, 2017
108. Ming Y, Wu N, Qian T, et al: Progress and future trends in PET/CT and PET/MRI molecular imaging approaches for breast cancer. Front Oncol 10:1301, 2020. https://doi.org/10.3389/fonc.2020.01301
109. Weber M, Kersting D, Umutlu L, et al: Just another 'Clever Hans'? Neural networks and FDG PET-CT to predict the outcome of patients with breast cancer. Eur J Nucl Med Mol Imaging 48(10):3141-3150, 2021. https://doi.org/10.1007/s00259-021-05270-x
110. Macedo F, Ladeira K, Pinho F, et al: Bone metastases: An overview. Oncol Rev 11(1):321, 2017. https://doi.org/10.4081/oncol.2017.321
111. Papandrianos N, Papageorgiou E, Anagnostis A, et al: A deep-learning approach for diagnosis of metastatic breast cancer in bones from whole-body scans. Appl Sci 10(3), 2020. https://doi.org/10.3390/app10030997
112. Li Z, Kitajima K, Hirata K, et al: Preliminary study of AI-assisted diagnosis using FDG-PET/CT for axillary lymph node metastasis in patients with breast cancer. EJNMMI Res 11(1):10, 2021. https://doi.org/10.1186/s13550-021-00751-4
113. Choi JH, Kim H-A, Kim W, et al: Early prediction of neoadjuvant chemotherapy response for advanced breast cancer using PET/MRI image deep learning. Sci Rep 10(1), 2020. https://doi.org/10.1038/s41598-020-77875-5
114. Gur D, Bandos AI, Cohen CS, et al: The 'Laboratory' effect: Comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. Radiology 249(1):47-53, 2008. https://doi.org/10.1148/radiol.2491072025
115. Rieke N, Hancox J, Li W, et al: The future of digital health with federated learning. npj Digit. Med. 3(1):1-7, 2020. https://doi.org/10.1038/s41746-020-00323-1
116. Warnat-Herresthal S, Schultze H, Shastry KL, et al: Swarm Learning for decentralized and confidential clinical machine learning. Nature 594 (7862):265-270, 2021. https://doi.org/10.1038/s41586-021-03583-3
117. Norgeot B, Quer G, Beaulieu-Jones BK, et al: Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med 26(9):1320-1324, 2020. https://doi.org/10.1038/s41591-020-1041-y