Medical Image Analysis 71 (2021) 102049

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Survey paper

Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: A review

Jun Bai[a], Russell Posner[b], Tianyu Wang[a], Clifford Yang[b,c], Sheida Nabavi[a,*]

[a] Department of Computer Science and Engineering, University of Connecticut, 371 Fairfield Way, Storrs, CT 06269, USA
[b] University of Connecticut School of Medicine, 263 Farmington Ave. Farmington, CT 06030, USA
[c] Department of Radiology, UConn Health, 263 Farmington Ave. Farmington, CT 06030, USA

ARTICLE INFO

Article history:
Received 14 September 2020
Revised 11 February 2021
Accepted 19 March 2021
Available online 3 April 2021

Keywords:
Digital breast tomosynthesis
Deep learning
Review

ABSTRACT

The relatively recent reintroduction of deep learning has been a revolutionary force in the interpretation of diagnostic imaging studies. However, the technology used to acquire those images is undergoing a revolution itself at the very same time. Digital breast tomosynthesis (DBT) is one such technology, which has transformed the field of breast imaging. DBT, a form of three-dimensional mammography, is rapidly replacing the traditional two-dimensional mammograms. These parallel developments in both the acquisition and interpretation of breast images present a unique case study in how modern AI systems can be designed to adapt to new imaging methods. They also present a unique opportunity for co-development of both technologies that can better improve the validity of results and patient outcomes.

In this review, we explore the ways in which deep learning can be best integrated into breast cancer screening workflows using DBT. We first explain the principles behind DBT itself and why it has become the gold standard in breast screening. We then survey the foundations of deep learning methods in diagnostic imaging, and review the current state of research into AI-based DBT interpretation. Finally, we present some of the limitations of integrating AI into clinical practice and the opportunities these present in this burgeoning field.

1. Introduction

Breast cancer is the second most common malignancy in women both in the US and worldwide, and causes a significant disease burden on the general population; about one in eight women will be diagnosed with it at some point in her life (Society, 2020). Like many cancers, however, early diagnosis has been repeatedly demonstrated to reduce overall disease burden and mortality. As such, breast cancer is one of the few cancer types that asymptomatic women are regularly screened for, so that cancers can be identified before clinical signs appear. The classical imaging study used for breast cancer screening has been two-view mammography — currently called full-field digital mammography (FFDM) — in which two 2D X-ray views of the breast are taken and interpreted by a radiologist. A significant body of research has shown that FFDM has had a significant effect on breast cancer morbidity

Abbreviations: ADMM, Alternating Direction Method of Multipliers; AGD, Average Glandular Dose; AUC, Area under ROC curve; BI-RADS, Breast Imaging-Reporting and Data Systems; BM3D, Block-Matching and 3D Filtering; CAD, Computer-Aided Detection; CDR, Cancer Detection Rates; CNN, Convolutional Neural Network; CT, Computed Tomography; DBT, Digital Breast Tomosynthesis; DLTG, Dictionary Learning with Temporal Gradient; DTL, Double Transfer Learning; FBP, Filtered Back Projection; FDA, U.S. Food and Drug Administration; FFDM, Full-Field Digital Mammography; FN, False Negative; FOM, Figured-of-Merit; FP, False Positive; FPN, Feature Pyramid Network; FROC, Free-response ROC; FWHMs, The Full Width at Half Maximum; GGGAN, Gradient Guided cGANs; HaarPSI, Haar Wavelet-Based Perceptual Similarity Index; IoU, Intersection over Union; IR, reconstruction algorithm; kt-SLR, kt sparse and Low-Rank; L+S, Low-Rank Plus Sparse Matrix Decomposition; mAP, Mean Average Precision; MI-RF, Multi-Instance Random Forests; MIL, Multiple-Instance Learning; MIoU, Mean Intersection over Union; MIP, Maximum Intensity Projections; MIP-HM, Maximum Intensity Projections applied histogram matching; ML, Machine Learning; MRI, Magnetic Resonance Image; MSE, Mean Squared Error; NMS, Non-Maximum Suppression; PVs, Projection Views; RMSE, Relative Mean Square Error; ROC, Receiver Operating Characteristic; ROI, Regions of Interest; RPN, Regional Proposal Network; RR, Recall Rates; s2D, Synthetic 2D; SGD, Stochastic Gradient Descent; SSIM, Structural Similarity Index Measure; STL, Single Transfer Learning; TCIA, The Cancer Imaging Archive; TN, True Negative; TP, True Positive; VOI, Volume of Interest.

* Corresponding author.
E-mail address: sheida.nabavi@uconn.edu (S. Nabavi).

https://doi.org/10.1016/j.media.2021.102049

and mortality with a greater than 20% reduction in breast cancer deaths as a result of regular screening (Tabár et al., 2011). Over the past decade or so, digital breast tomosynthesis (DBT), a three-dimensional imaging technique has emerged as the new gold standard for digital mammography because of its superior ability to capture previously undetected cancers (Vedantham et al., 2015).

The past decade has also been marked by a revolution in the ability of computers using machine learning (ML) methods to solve a host of problems in medical imaging. While relatively simplistic computer-aided detection (CAD) systems have existed in mammography for several decades now, such systems have traditionally been limited in their utility and capability. By contrast, the development of new deep learning methods has been demonstrated to detect malignancies on traditional mammograms at the level of, or even better than trained radiologists. Despite these gains, the specter of computer detection systems to be able to diagnose independently in a clinical trial has not come to fruition, and current focus has remained to the notion of ML systems supporting the radiologist rather than acting as independent diagnostician.

However, one major impediment to the adoption of the next generation of CAD systems may be that most imaging modalities have been rigorously tested and validated and exist within a fairly rigid medical management and delivery model. As a result, most deep learning systems for interpreting images must be “tacked on” to existing workflows. Digital breast tomosynthesis is different in this regard. While clinical trials have shown the diagnostic superiority of DBT over FFDM overall (Ciatto et al., 2013; Haas et al., 2013; Skaane et al., 2018; 2019; Greenberg et al., 2014; Sharpe et al., 2015; McDonald et al., 2016; Nam et al., 2015), the precise ways to efficiently acquire, reconstruct, interpret, and report DBT studies are still being developed by different physicians and practices. Hence, DBT presents a unique opportunity for the co-development of these best practices in tandem with novel ML/AI-based CAD systems from the ground up, and we believe that DBT will be one of the imaging modalities which will seamlessly integrate deep learning technology into all levels of the DBT acquisition and analysis workflow. In recent years, there are lots of studies about the state of art in DBT not only in radiology field but also in AI field (Vedantham et al., 2015).

In this review, we will examine the current state of both DBT as a clinical imaging study and deep-learning based DBT analysis and decision support systems. Along the way, we will present particular challenges and opportunities in both aspects and how the two technologies (DBT and deep learning) can help get us closer to the long-promised integration of modern AI with medical imaging technology.

2. Digital breast tomosynthesis

2.1. Technique

The methodology of DBT is quite similar to that of traditional FFDM. In FFDM, the breast is compressed between a pair of paddles with the X-ray detector below, while the X-ray emitter is placed above the paddle and oriented perpendicular to the detector. Two images of each breast are typically taken, the craniocaudal (CC) and mediolateral oblique (MLO) views (Fig. 1A). Because the breast is compressed significantly and is directly against the detection plate, these images can achieve resolutions and contrast ranges far greater than most conventional radiography techniques with a relatively low radiation dose (called average glandular dose or AGD).

However, one of the drawbacks of traditional FFDM is that the 2D view can create tissue artefacts from the summation of healthy dense breast tissue itself which may mimic malignant lesions (Poplack et al., 2007). Digital breast tomosynthesis aims to solve this problem by capturing multiple images obtained by rotating the X-ray emitter over the compressed breast to obtain a series of projection views (PVs). Reconstruction and angle options vary, but common instruments use these PVs to build a z-Stack of parallel slices in the same standard CC and MLO views (Fig. 1B) (Vedantham et al., 2015). This system represents a compromise: although the range of PVs is limited, the necessary resolution and contrast are largely maintained by keeping the breast fixed and compressed against the detector.

2.2. DBT as a replacement for traditional FFDM

Over the past two decades, several clinical trials have demonstrated that DBT increases cancer detection rates (CDR) while decreasing recall rates (RR) when compared to FFDM. Early studies demonstrated this improvement when comparing DBT with FFDM to FFDM alone (Ciatto et al., 2013; Haas et al., 2013; Skaane et al., 2018; 2019). A series of retrospective studies demonstrated that DBT was more effective as a stand-in replacement for FFDM altogether (Greenberg et al., 2014; Sharpe et al., 2015; McDonald et al., 2016), with DBT being able to find cancers which were entirely missed on FFDM (Nam et al., 2015).

Given these consistent findings on clinical trials, one would imagine that DBT implementation and workflows would be similarly consistent across radiology practices. The reality is that individual practice adoption and interpretation workflows vary widely (Gao et al., 2017; Ebuoma et al., 2015; Freer et al., 2017; Miglioretti et al., 2019). While many factors, such as cost and demographics influence the adoption of DBT, this heterogeneity in usage presents a unique opportunity for AI systems to help develop DBT best practices from the ground up. In our discussion, we will broadly focus on two main problem-domains - cancer detection and image reconstruction - with contextualized clinical examples of each.

2.3. Computer-aided detection

The concept and implementation of CAD systems in mammography date back several decades. In fact, both the first U.S. Food and Drug Administration (FDA)-approved and the first Medicare-reimbursed CAD systems in radiology as a whole were for mammography screening. Such tools were typically built by devising numerical and statistical representations of common breast abnormalities such as microcalcifications, spiculations, and masses which were then hand-tuned to the specific mammographic instrumentation. Early trials showed promise for CAD, especially in the double-reading paradigm for FFDM common in Europe (Cupples et al., 2012; Doi, 2007; Dean and Ilvento, 2012). Ultimately, however, large-scale studies have suggested that the value of traditional CAD systems is equivocal at best, and CAD has fallen out of favor in many breast clinics (Lehman et al., 2015).

Over the past decade, as DBT has begun to replace FFDM in many clinics, insurance companies have been less willing to pay a premium for CAD tools, leading to a relative drop-off in their use in the breast clinic (Nishikawa and Bae, 2018). Nonetheless, two key factors have driven a renaissance in CAD systems: reading times and deep learning. Compared to FFDM, radiologist interpretation of DBT studies takes about twice as long, limiting the ability to screen as many patients due to reader fatigue. In conjunction with modern image analysis techniques, deep-learning CAD (AI-CAD) systems have proven to be useful in reducing reading times without significantly affecting sensitivity, specificity, RRs or CDRs, producing reductions in reading times from 14–40% (Chae et al., 2019; Balleyguier et al., 2017; Benedikt et al., 2017).

While at least one AI-CAD system is currently FDA-approved for use in the clinic, the overall added benefit is unclear; prior studies explicitly include caveats that systems can identify masses but not

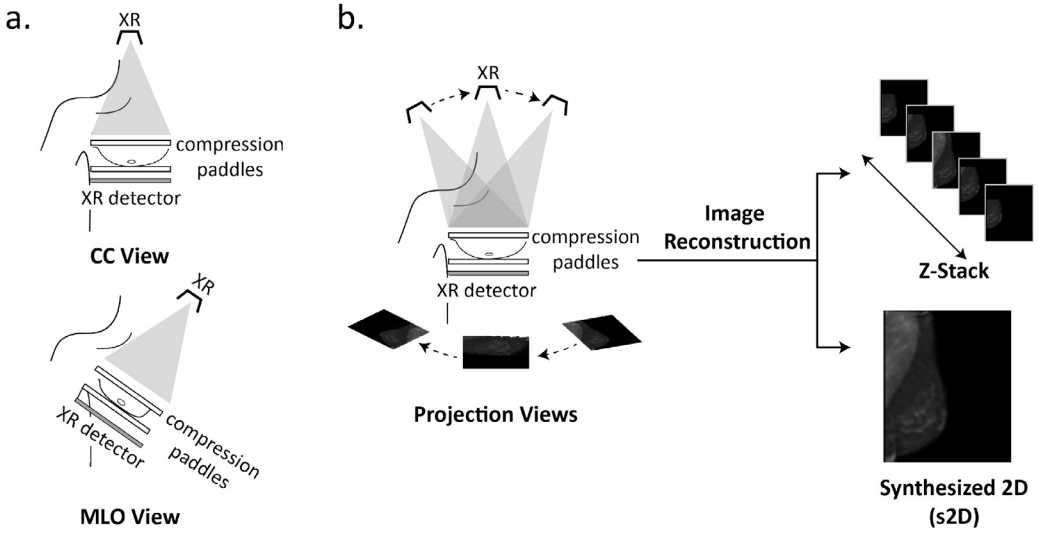

Fig. 1. a. In a standard 2D mammogram, the X-ray (XR) emitter takes an ordinary radiograph captured by a digital XR detector plate. Two views are captured, the craniocaudal (CC) and mediolateral oblique (MLO). b. In DBT imaging, the XR emitter travels in an arc over the breast so that the detector captures multiple images (projection views). These images are fed into a reconstruction algorithm to produce a volumetric "z-Stack" and a synthesized 2D mammogram to be reviewed by the radiologist. DBT also typically captures both the CC and MLO views (only CC pictured).

other potential signs of malignancy such as microcalcifications and architectural distortions (Balleyguier et al., 2017; Benedikt et al., 2017).

2.4. Synthetic mammography

Whether using DBT or FFDM, effective cancer detection in mammography depends on the use of one or more prior imaging studies to identify new breast changes which may indicate malignancy (Hakim et al., 2015; Hayward et al., 2016). In order to allow readers to compare DBT studies with prior FFDM images, a variety of techniques have been developed to generate "synthetic 2D" (s2D) mammograms. Although the DBT technology would seem to include ordinary FFDM images as part of its acquisition, each DBT image is actually relatively low resolution. Unlike FFDM, each DBT PV uses a lower radiation dose and does not use an anti-scatter grid to filter out reflected and/or refracted x-rays. While there were some initial hurdles involved, several recent studies have confirmed that DBT-s2D has similar CDRs with reduced RRs when compared to a combined study using DBT and FFDM (DBT-FFDM) or FFDM alone while also reducing AGD (Lai et al., 2018; Aujero et al., 2017; Choi et al., 2019; Freer et al., 2017).

However, the development of s2D mammography illustrates one of the key open questions in the optimization of DBT interpretation: namely, what is the ideal synthetic image (for both humans and computers)? Current FDA-approved s2D systems (e.g. C-View by Hologic and V-Preview by GE) do not actually attempt to construct a true approximation of an original FFDM image, but rather use information from the collected PVs to create an "enhanced" 2D view which is meant to guide the reader through the DBT study (Smith, 0000; GE Corporation, 2021). In clinical practice, such approximations create their own set of unique advantages and artifacts which the reader must be aware of (Zuckerman et al., 2017).

2.5. Translating to the computational domain

Analysis of DBT volumes is a distinctly different task from many commonly-studied image analysis tasks, so understanding how the goals of such work translate over to the deep learning domain is critical.

For example, accurate reading of studies requires very high resolution (> 1 M pixels), high grayscale dynamic range and contrast, incorporation of prior studies and medical records, and specialized training in image interpretation and clinical decision-making. In contrast, popular deep learning models currently used for image classification have inputs that are roughly an order of magnitude less in both spatial resolution and image contrast (Table 1).

Another challenge unique to the interpretation of breast cancer screening is the low prevalence of malignancy: only about 0.6% of screening mammograms contain malignancies. Because deep learning algorithms depend on large training data sets, it is unknown whether models should be trained using real-world data — with inherent low number of malignancies — or enriched data sets. With less than 1% of screening studies containing true malignancies, a system trained on population-level data will see relatively few malignant studies and may tend to under-diagnose. On the other hand, using training data which is enriched with malignancies may train a model to over-diagnose cancers; a common problem in current breast cancer screening programs (Miller et al., 2014; Kalager et al., 2012).

While these and other concerns factor into the setup and training of deep learning models, the general problems to be addressed by any deep learning model are relatively straightforward. Broadly, one can break the problem of clinically useful malignancy detection into a hierarchy of sub-problems:

1. Determining if a study contains an abnormality concerning for malignancy.
2. Locating the center of such an abnormality for either review by a human reader or for a follow-up study or biopsy.
3. Classifying the abnormality itself.
4. Making follow-up clinical decisions based on the identified and classified abnormality (if present).

In machine learning terms, the first task can be considered either an "anomaly detection" or a "classification" problem. The second task is one of "object localization" or "segmentation," while the third is a classification problem on a sub-region of the entire image. The fourth task is a very interesting problem, which may be considered to be a multi-class classification problem. This task has not been studied in great detail, and will be reserved for our discussion of future challenges and opportunities. With these sub-

Table 1
Deep learning models used in biomedical image analysis.

	Model	Input size	Layers	Total parameters	Modality	Organ	Modality reference	Evaluation metrics
Image classification	LeNet (LeCun et al., 1998)	32×32	6	60,000	CT X-ray MRI	chest breast kidney	Khan and Yong (2017); Gastounioti et al. (2018)	
	AlexNet (Razzak et al., 2017)	$224 \times 224 \times 3$	8	61 M	CT X-ray MRI	chest breast kidney heart	Gastounioti et al. (2018); Lakhani and Sundaram (2017); Lévy and Jain (2016)	Accuracy ROC F1 Precision Recall
	GoogleLeNet (Szegedy et al., 2015)	$224 \times 224 \times 3$	22	4 M	X-ray MRI	breast heart	Gastounioti et al. (2018)	
	VGG16 (Simonyan and Zisserman, 2015)	$224 \times 224 \times 3$	41	138 M	X-ray	breast	Shen et al. (2019)	
	ResNet(2015) (He et al., 2015)	$224 \times 224 \times 3$	34/50/101	Over 23 M	X-ray	chest breast	Shen et al. (2019)	
	Model	**Module**	**Learning**	**Loss function**	**Modality**	**Organ**	**Modality reference**	**Evaluation matrix**
Object detection	faster-RCNN (Ren et al., 2016)	Convnet/ RPN	SGD (Stochastic gradient descent)	log loss + bbox regression	X-ray CT	chest brain breast	Ribli et al. (2018); Rahmat et al. (2019); Ezhilarasi and Varalakshmi (2018)	Mean Average Precision Intersection over Union
	YOLO (Redmon et al., 2016)	Convnet	SGD	sum squared error + bbox regression + object confidence + background confidence	X-ray	chest breast	Al-masni et al. (2018); Ozturk et al. (2020); Cao et al. (2019)	
	RetinaNet (Lin et al., 2017)	Feature pyramid network	SGD	Focal loss	X-ray CT	liver breast	Zlocha et al. (2019); Lotter et al. (2019)	
	Model	**Backbone/ Base model**	**Type**	**Loss function**	**Modality**	**Organ**	**Modality reference**	**Evaluation matrix**
Image segmentation	mask-RCNN (He et al., 2018)	R-CNN	Instance	log loss + smoth L1 + pixel wise cross entropy	CT X-ray MRI	brain breast, liver	Thang et al. (2021, 2021)	Pixel accuracy Intersection over Union Mean-IoU Dice coefficient
	U-Net (Ronneberger et al., 2015)	Autoencoder	Semantic	pixel wise cross entropy	X-ray CT	breast liver	Lai et al. (2020); Chlebus et al. (2018)	
	attention U-Net (Oktay et al., 0000)	Autoencoder	Semantic	pixel wise cross entropy	X-ray CT	liver breast	Li et al. (2019)	
	FCN (Long et al., 2015)	VGG	Instance	pixel wise cross entropy	CT	liver	Ben-Cohen et al. (2016)	

problems in mind, we turn our attention to the current tools and techniques which have been used to solve these tasks.

3. Current deep learning methods

Deep learning models are a subset of machine learning models focused on allowing a model to design itself to extract features for the underlying task. Features of an object are object's characteristics, such as edges, lines, and shapes that can be quantified and therefore represent the object. The quantified features are used to analyze the object and perform a particular computational task. The learning process is achieved by assigning weights or parameters of a model after many iterations such that the model is optimized for a given task — also called training.

In the field of mammography, traditional CAD systems rely primarily on non-deep learning methods, in which features are hand-engineered to calculate quantities such as the degree of calcification, spiculation, or distortion. In this setup, conventional classification methods such as regression models are fitted to identify normal, benign, and malignant mammogram images.

By contrast, the underlying principle of supervised deep learning (training requires ground truth label) is giving the computer many labeled images and then allowing it to determine its own set of classifying features that may or may not correspond to human-interpretable features.

The "depth" of these systems corresponds to the system's capability to combine basic values, such as pixel locations and intensities, into sophisticated, non-linear features to better extract local and global attributes from images. In medical image analysis, deep learning has been successfully applied to many different modalities, including Computed Tomography (CT), Magnetic Resonance Image (MRI), X-ray and conventional radiographs (Razzak et al., 2018).

In deep learning breast cancer diagnostic, the term "cancer detection" is an umbrella term, referring to several different tasks including:

Classification. Identifying the probability of the entire image as "normal", "benign" or "malignant." Some methods even seek to identify the precise type of a malignancy's probability.

Object detection. Locating any suspicious or abnormal regions of an image and presenting them to the end-user.

Segmentation. Labeling individual pixels as normal or pathological by combining object detection and classification.

3.1. Deep learning - building blocks

The advantages of deep learning compared to classical ML are two-fold. First, they have many tunable parameters that allow them to derive much richer and more complex representations of objects. Second, they have a highly modular structure, with a set of interchangeable blocks which makes them easy to design and implement.

There are many types of deep learning models including Deep Neural Networks (DNNs), Autoencoders (AEs), Deep Belief Networks (DBNs), Deep Convolution Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Generative Adversarial Networks (GANs). All these models can be applied to the tasks of image classification, object localization/detection, and image segmentation. However, because they encode the spatial relationships among pixels in an image, CNN-based deep learning models are mostly used in image analysis. A number of popular deep learning models used in biomedical image analysis and their general structures are summarized in Table. 1. For the classification task, CNN (Fig. 2a) is the major model used in medical image cancer classification. U-Net (Fig. 2b) and Faster RCNN (Fig. 2c) are the models often used in medical image segmentation and object detection (they both are CNN-based models).

3.1.1. Layer

The major components of a CNN-based model are convolution (LeCun et al., 1998), dropout (Srivastava et al., 2014), batch normalization (Ioffe and Szegedy, 2015), dense (fully connected) and pooling layers (Ranzato et al., 2007) as explained in Table. 2. Although the main operation of CNNs is convolution, additional network components (layers) have made the overall model more tractable, efficient, and accurate.

3.1.2. Loss functions

Training supervised deep neural networks depends critically on the choice of loss functions (Table 3). A loss function measures the difference between the output from the network and the underlying ground truth. The choice of the loss function directly impacts the model generalization capability and performance. In general, loss functions can be categorized into two main groups: classification loss functions and regression loss functions. In classification tasks, the goal is to predict to which class a sample belongs. Therefore, classification loss measures the error between the predicted classes and ground truth classes. In regression, the goal is to use input samples to predict continuously-valued outputs. Regression loss functions measure the error between these predicted numerical values and the values of the ground truth.

Depending on the deep learning models and their tasks, the loss function can be selected from regression and classification loss functions or a combination of both. Deep learning DBT image analysis application's loss function for cancer detection can be categorized into three types: classification loss, object detection loss, and segmentation loss (the detailed information of these loss functions are described in the supplementary material file). The commonly used loss functions for the aforementioned three types are summarized in Table 3.

3.1.3. Evaluation metrics

The common evaluation metrics for classification are accuracy, false positive rate, false negative rate, precision, recall, F1 score, Receiver Operating Characteristic (ROC) curve, Area under ROC curve (AUC), Precision-Recall (PR) curve, Average Precision (AP) and mean Average Precision (mAP). For object detection models, the frequent evaluation metrics are mAP, Intersection over Union (IoU), precision, recall, and Non-Maximum Suppression (NMS). Intersection over Union evaluates the overlap between a predicted candidate bounding box and a ground truth bounding box. Ideally, the predicted candidate bounding box should perfectly overlap with the ground truth bounding box, such that IoU is equal to 1. NMS is a technique to select the best bounding box over an object for mAP evaluation. In other words, the mAP is computed after NMS. The common evaluation metrics for segmentation models are pixel accuracy, mean pixel accuracy, Mean Intersection over Union (MIou) and Dice Score. For non-classification tasks, particularly image generation and denoising, the most commonly-used metrics include Structural Similarity Index Measure (SSIM), and Contrast-to-Noise Ratio (CNR).

3.2. Representative models

The previously-outlined components and methods are mixed and matched in a multitude of ways to generate useful deep learning models. In order to understand the principles behind the novel model structures that have been applied in DBT analysis, it is useful to understand the details of a few representative models for each diagnostic task, as most of the surveyed models are variations of these core types.

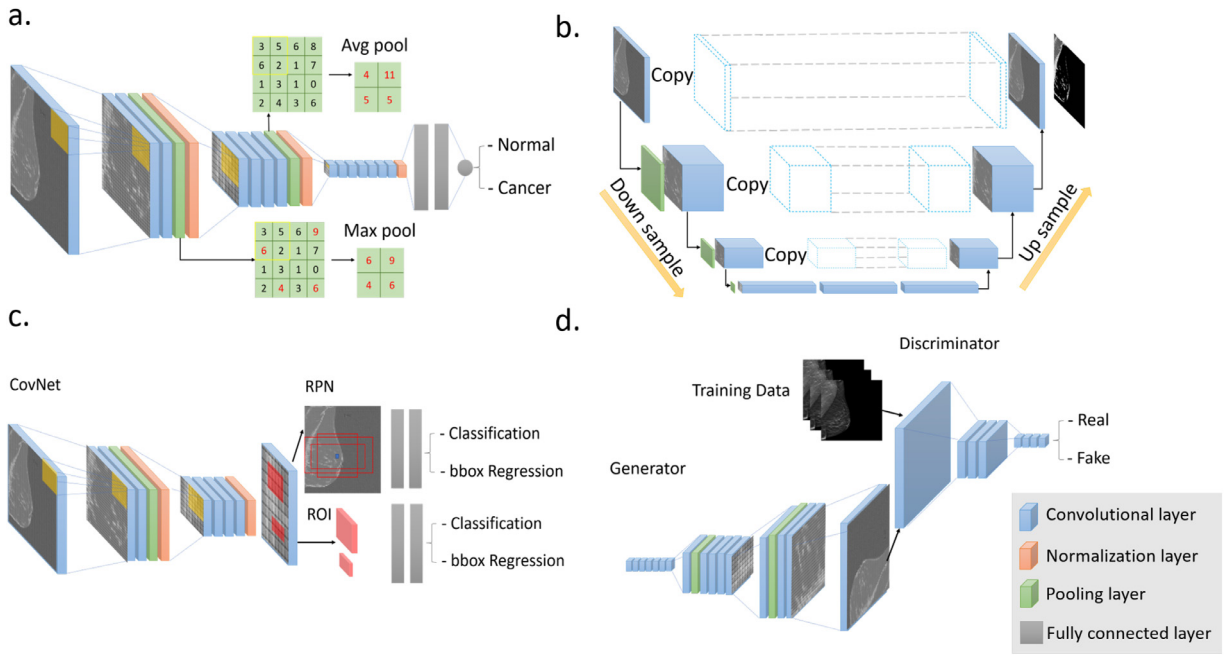

Fig. 2. Structure of deep learning models. a. CNN model-is often used for the classification task. The model typically builds by convolutional blocks that contain convolutional layers, pooling layers, and normalization layers. The outputs of a CNN model are class probabilities. b. U-Net model is often used for segmentation tasks. In the down sample stage, heights and weights of feature dimensions are decreasing, and depth is increasing. In the up sample stage, the feature heights and weights are increasing, and depth is decreasing. The output of this model is the mask of the segmented tumor. c. Faster RCNN model-is often used for object detection tasks. The CovNet typical CNN model which generate features and feed into RPN. Region Proposed Network outputs a set of proposed objects which feed to parallel branches for classification and bounding box (bbox) regression. d. GAN-is often used to generate simulated data. The generator produces simulated data which will be determined by the discriminator as real or fake.

Table 2
Core components of a CNN-based deep learning model.

	Hyperparameters	Input	Output	Type	Function
Convolution Layer	Kernel size/strides/ padding	3D tensor	3D tensor	1D conv/2D conv/ 3D conv	Extract features
Pooling Layer	Pool size/padding	3D tensor	Reduced 3D tensor	Max pooling/average pooling/global average pooling	Extract features/reduce dimension
Normalization Layer	Momentum/ epsilon/beta	Arbitrary	same shape as input	Batch norm/layer norm/instance norm/group norm	Standardize input
Dense Layer	Number of nodes	n dimension tensor	n dimension tensor	–	Regular fully connected layer
Dropout Layer	Rate	Arbitrary	same shape as input	–	Prevent model over fit
Activation Layer	Activation function	Arbitrary	same shape as input	ReLU/sigmoid/Tanh/ softmax/leaky ReLU/binary step	Applies activation function to previous layer

3.2.1. Classification - CNN

The most established algorithm — among a variety of deep learning models — for medical image classification is CNN (Gastounioti et al., 2018; Abdelhafiz et al., 2019; Lakhani and Sundaram, 2017; Shen et al., 2019). Fig. 2.a shows a schematic diagram of a basic CNN model. A CNN model uses grid-patterned data, such as images, as inputs and learns features from low to high levels based on trained weights (LeCun et al., 1998). Regardless of the specific task, using deep learning and the development of CNNs has played a pivotal role in medical image analysis (LeCun et al., 1998).

More specifically, convolution layers are fundamental for CNNs. Convolution layers perform feature extraction by applying convolution process through filter kernels defined by height, weight and depth shifted over each layer. The size of a filter kernel, typically, is 3×3, 5×5 or 7×7. Pooling layers are used for down sampling the feature maps after the convolution layer, in order to reduce the dimension of features. The most favored pooling operation is max pooling which is used to select the maximum activated neuron from previous feature maps as input to the next layer. A common alternative pooling operation is average pooling, which uses the average activation values of neighboring neurons as input to the next layer. A fully connected layer, also known as a dense layer, connects all neurons between two layers. To reduce the high-dimensional output from the preceding layer, the features extracted from the last convolution layer are flattened to a vector representation which is the input to the next dense layer.

Finally, the fully connected layer will be mapped to output, and a probabilistic distribution activation function such as the sigmoid function will be applied to predict the output classes' probabilities.

One major challenge associated with employing CNNs for analyzing medical images is the dependency of CNNs on a large number of training images. However, the sizes of medical image data sets are often small, and labeling data is time-consuming — require significant time and effort from trained experts. In recent research, one of the well known topics is transfer learning which stores the learned weights from task A and uses it to train another task B to

Table 3
Loss functions of commonly used deep learning models.

	Name	Equation	Variable definition
Image classification	Cross-Entropy	$l(y,\hat{y}) = -\sum_i^n y_i \log \hat{y}_i$	• n number of classes • y is ground truth (GT) classes • $\hat{y}$ is predicted classes probability
	Binary cross-entropy (log loss)	$l(y,\hat{y}) = -(y\log(\hat{y}) + (1-y)\log(1-\hat{y}))$	
Object detection	Smooth L_1 loss	$smoothL_1 = \begin{cases} 0.5f^2 & \lvert f\rvert \le 1 \\ \lvert f\rvert - 0.5 & otherwise \end{cases}$	• f is difference between predicted value and ground
	faster-RCNN	$l(y_m, t_m) = L_{cls}(y_m, y_m^*) + \lambda y_m^* L_{reg}(t_m, t_m^*)$	• m is anchor index • y_m is predicted probability of anchor i • y_m^* is 1 if anchor is positive; else 0 • t_m is coordinates of predicted bounding box (bbox) • t_m^* is bbox GT of positive anchor
	YOLO	$l_1 = \lambda_{coord}\sum_{i=0}^{s^2}\sum_{j=0}^{B}\mathbb{L}_{ij}^{obj}[(x_i - \hat{x_i}^2) + (y_i - \hat{y}_i)^2]$ $+\lambda_{coord}\sum_{i=0}^{s^2}\sum_{j=0}^{B}\mathbb{L}_{ij}^{obj}[(\sqrt{w_i} - \sqrt{\hat{w}_i}^2) + (\sqrt{h_i} - \sqrt{\hat{h}_i}^2]$ $+\sum_{i=0}^{s^2}\mathbb{L}_{ij}^{obj}\sum_{j=0}^{B}[((p_i(n) - \hat{p}_i(n))^2]$ $+\sum_{i=0}^{s^2}\sum_{j=0}^{B}\mathbb{L}_{ij}^{obj}[(C_i - \hat{C_i}^2)]$ $+\lambda_{coord}\sum_{i=0}^{s^2}\sum_{j=0}^{B}\mathbb{L}_{ij}^{noobj}[(C_i - \hat{C_i}^2)]$	• λ is weighted parameter for balancing • x_i, y_i is bbox's center coordinate GT • $\hat{x_i}, \hat{y_i}$ is bbox's center coordinate prediction • w_i, h_i is bbox's width and height GT • $\hat{w_i}, \hat{h_i}$ is bbox's width and height GT • n is classes • C_i is confidence score of if there is object • $\hat{C_i}$ is 1 if there is object, else 0 • p_i is class prediction, $\hat{p_i}$ is class GT • s number of cells in bbox • B number of bbox
	RetinaNet	$FL(p_t) = -(1-p_t)^{\gamma}\log(p_t)$	• p_t is prediction if $y = 1$, else p_t is $(1 - prediction)$
Image segmentation	mask-RCNN	$l = L_{cls} + L_{box} + L_{mask}$	• L_{cls} is log loss same as faster RCNN • L_{box} is smooth L_1 same as faster RCNN • L_{mask} is pixel-wise cross entropy
	U-Net	$E = \sum_{x\in\Omega}\omega(x)\log(p_{l(x)}(x))$	• x is pixel position • l is truth label of each pixel • ω weight for important pixels

enhance the deep neural networks' performance (Yosinski et al., 2014). Many well-known models trained on natural image data sets are available for transfer learning such as LetNet(LeCun et al., 1998), AlexNet, VGG, and GoogleNet (which are spatial exploitation based CNNs), ResNet, Inception-V3, and Inception-V4 (which are Depth based CNNs), DenseNet (which is multi-path based CNNs), Residual Attention Neural Network, and the Convolutional Block Attention network (which are attention-based CNNs) (Khan et al., 2020).

3.2.2. Object detection - faster RCNN

One typical medical image analysis task is object detection, and Faster RCNN is the most common model for this task (Fig. 2c.). The Faster RCNN model contains a few components: ConvNet, region proposed network (RPN), and multi-task branches. The ConvNet, which is a particular CNN model, outputs a set of feature maps. The RPN takes the ROI selected based on anchors over the feature maps generated from ConvNet, and outputs a set of proposed objects. In other words, the ROI from ConvNet will feed into two parallel branches. One branch — typically employs few fully connected layers — is for bounding box (bbox) regression, and the other branch is for classification of the objects inside the bounding boxes. Anchors are proposed bounding boxes around a pixel by particular scales (size of ROI) and aspect ratios ($width_{ROI}/height_{ROI}$). For example, if a feature map dimension is $n \times m$, and there are k anchor boxes. In total, RPN will propose $n \times m \times k$ ROI.

3.2.3. Segmentation - U-Net

U-net is a deep learning model widely used in medical image analysis for segmentation (Ronneberger et al., 2015). It has a similar structure as an autoencoder network (Rumelhart and McClelland, 1987) and similar components as CNNs. There are two stages for U-net: down sampling and up sampling as shown in Fig. 2b. The down sampling stage, also known as contracting path, follows the typical structure of CNNs for learning the features and making a copy of features for the up sampling stage. In the up sampling stage, which is also known as expansive path, the copied features from contracting path are concatenated and expanded to get to the dimension of the original inputs.

3.2.4. Generative modeling - GAN

In addition to analyzing images, one of the more recent breakthroughs in deep learning is the development of models which can generate new images based on a set of training data. As we will discuss, one major application of these models in DBT is for non-diagnostic tasks such as image reconstruction and artefact removal; however, several such models have also been cleverly applied to perform diagnostic tasks in DBT studies. In order to understand the nuances of such studies, we outline the mechanism behind one classical generative model, the generative adversarial network (GAN).

The GAN was first introduced in 2014 (Goodfellow et al., 2014) to generate realistic new photographs that have identical realistic characteristics as the training data set. A GAN model (Fig. 2d) is a generative model, which is mostly used for generating instances of data that have the same distribution as the original data. The GAN model consists of two parts, the generator and the discriminator, which are both deep networks. The generator is used to generate synthetic samples, while the discriminator is used to classify the fake samples from the real ones. A well trained GAN model can generate plausible data that cannot be distinguished by the discriminator. A GAN model is prevalent for generating simulated medical images. It has been widely applied to different types of medical image modalities such as X-ray, CT and MRI (Lei et al., 2019; Shin et al., 2018; Nie et al., 2018; Frid-Adar et al., 2018; Bi et al., 2017; Yi et al., 2019; Yang et al., 2018; Wu et al., 2019; Yu et al., 2020; Zhang et al., 2019). Several variations of the GAN have been proposed for different purposes. One of them is the conditional GAN. The conditional GAN(cGAN) is the extension of GAN that added extra information as a condition to both generator and

Table 4
Summary of deep learning DBT studies: only deep learning applied on DBT are included.

Reference	Transfer learning	Augmentation	Dataset*	Data type*	# data	ROI/Patch/Image*	Size	Evaluation metric	Best results
Classification									
Fotin et al. (2016)		✓	*pvt*	*STA*	344	*R*	256 × 256	Sensitivity	0.93
Kim et al. (2016)		✓	*pvt*	*STA*	160	*R*	32 × 32	AUC	0.847
Yousefi et al. (2018)		✓	*pvt*	*STA*	87	*I*	256 × 256	AUC	0.87
Samala et al. (2018)	✓	✓	*pvt*	*STA*	324	*R*	128 × 128	AUC	0.90
Zhang et al. (2018)	✓	✓	*pvt*	*STA*	3290	*I*	832 × 832	AUC	0.6632
Liang et al. (2019)	✓	✓	*pvt*	*STA*	1124	*I*	1024 × 1024	AUC	0.97
Zhang et al. (2019)	✓	✓	*pvt*	*STA*	3290	*I*	1024 × 1024	AUC	0.854
Mendel et al. (2019)	✓		*pvt*	*s2D PRJ*	78	*R*	512 × 512	AUC	0.89
Singh et al. (2020)	✓	✓	*pvt*	*STA*	68,311	*P*	512 × 512	AUC	0.847
Li et al. (2020a)	✓	✓	*pvt*	*STA*	927	*R*	256 × 256	AUC	0.99
Matthews et al. (2020)	✓	✓	*pvt*	s2D	78,445	*I*	416 × 320	AUC	0.97
Rodriguez-Ruiz et al. (2018)			*pvt*	*RCS*	956	*P*	29 × 29	AUC	0.88
Object detection and segmentation									
Fan et al. (2019)		✓	*pvt*	*STA*	182	*R, I*	–	AUC	0.96
Lotter et al. (2019)	✓	✓	*pvt*	*STA*	24,253	*I*	–	AUC	0.945
Lai et al. (2020)	✓	✓	*pvt*	*STA*	87	*P*	–	AUC	0.859
Swiecicki et al. (2020)			*pvt*	*STA*	19,230	*P*	128 × 128	MSE	0.0333
Buda et al. (2021)			*pub*	*STA*	22,032	*P*		Sensitivity	0.60 at 2 FP per DBT volume
Fan et al. (2020)			*pvt*	*STA*	364	*P*	256 × 256	Sensitivity	0.90 at 0.80 FP per breast
Li et al. (2020b)			*pvt*	*STA*	265	*I*	–	Sensitivity	0.80 at 1.95 FP per volume
Denoising, reconstruction and synthesis									
Jiang et al. (2019)		✓	*pvt*	*STA*	1077	*R*	512 × 512	SSIM	0.629 ± 0.196
Teuwen et al. (2020)			*pvt*	*PRJ*	50	*I*	–	SSIM	0.89 ± 0.01
Sahu et al. (2019)			*pvt*	*PRJ*	30	*P*	200 × 200	HaarPSI	0.383
Wu et al. (2020)			*pvt*	*PRJ*	176	*R*	–	SSIM	$0.0.724 \pm 0.098$
Michielsen et al. (2020)			*pvt*	*PRJ*	52	*I*	–	MSE	101×10^{-6}
Su et al. (2020)			*pvt*	*PRJ*	1000	*I*	512 × 1024	FWHMs	12.9*mm*
Gao et al. (2020)	✓		*pvt*	*PRJ*	9	*P*	32 × 32	CNR	improved 0.097

pvt* is private database,pub* is public database, **STA* is z-Stack, **PRJ* is projection,**RCS* is reconstruction, **I* is image, **R* is ROI, **P* is patch

discriminator. The additional information could be any kind, such as class labels, images, or words (Mirza and Osindero, 2014).

4. DBT deep learning studies

In the previous section, we introduced the well-known deep learning models used in medical image analysis. In the following subsections, we review studies that employed deep learning in DBT for assisting cancer diagnostic tasks and non-diagnostic tasks. The studies that applied deep learning in DBT are summarized in Table 4.

4.1. Diagnostic applications

For deep learning DBT diagnostic, studies can be grouped into two types: classification and localization. Because of the model selection and computation resources, the researchers often design their studies to use patches or whole images for classification. Localization tasks typically can be grouped into object detection tasks and segmentation tasks. In this section, we review studies that classify images and locate suspicious tumors.

4.1.1. Classification

Recently, deep learning has been used for mammogram classification to power the accuracy of tumor or cancer detection. Many studies have employed various types of deep learning architectures (not limited to CNN) for breast cancer detection and classification (Abdelhafiz et al., 2019). In this section, we review studies that have used deep learning for DBT image classification.

The study by Fotin et al. (2016) compared the boosted decision tree method with CNN approaches in cancer (positive or negative mass and positive or negative architectural distortion) classification using DBT. The CNN model used in their study is identical to AlexNet but takes input dimensions as 256 × 256 (ROI). They compared the CNN model sensitivity with that of the boosted decision tree. Because they had few biopsy proven malignancies, they included suspicious images annotated by radiologists as cancer. They compared results using the suspicious lesion data set, and the biopsy proved malignancy data set. The results showed that the CNN (using suspicious included data, sensitivity: 0.893; and using proved malignancy data sensitivity: 0.930) performed better compared with the boosted decision tree (using suspicious included data, sensitivity: 0.832; and using proved malignancy data, sensitivity: 0.852).

Samala et al. (2018) used two stages transfer learning in their proposed model for classifying DBT images as mass or normal. In the first stage, the authors employed the pre-trained AlexNet model, trained using ImageNet, and fine-tuned the model using FFDM images. In the second stage, the pre-trained weights learned from FFDM images are transferred as initials weights for training the model using DBT z-Stack images. Rather than using the CNN directly as a classifier, the CNN at the second stage is used for feature extraction. Then, a feature selection method and random forest classifier are applied to the extracted features for classification (Samala et al., 2018). In the first stage of fine-tuning AlexNet, 19632 augmented ROI patches from 2454 mass lesions were used. In the second stage of training the model, 9120 ROI patches from 228 mass lesions from DBT z-Stack images were used. They compared the performance of the proposed model, when changing the size of model parameters, in terms of AUC. They also applied network pruning to reduce the complexity of the model. They showed that after pruning the model, the number of trainable neurons reduced by 87.2%, the number of trainable parameters reduced by 34.4%, and the number of multiplications and additions decreased by 95.5%. The result showed an AUC of 0.90 with pruning the model and an AUC of 0.89 without pruning.

Mendel et al. (2019) used ImageNet pre-trained VGG19 for feature extraction in FFDM and DBT images for malignant and benign classification. Features were selected from the VGG19 model after each max pooling layer. Average pooling was performed for feature dimension reduction. To avoid redundant features, Leave-One-Out Step-wise feature selection was performed to obtain the most frequently selected features. Then, the Leave-One-Out Support Vector Machine (SVM) classifier was used to estimate the likelihood of malignancy. In this study, 78 lesion images, of which 30 are malignant, and 48 are benign were used. The ROI patches (dimensions: 512×512) were extracted from FFDM images, reconstructed 2D mammography images and DBT z-Stack slice images. The AUC metric was used to compare the performance of the model using FFDM images (CC and MLO view AUC $= 0.81 \pm 0.05$, CC view AUC $= 0.76 \pm 0.05$ and MLO view AUC $= 0.76 \pm 0.05$), 2D reconstructed images (CC and MLO view AUC $= 0.86 \pm 0.04$, CC view AUC $= 0.81 \pm 0.05$ and MLO view AUC $= 0.88 \pm 0.04$) and DBT z-Stack slice images (CC and MLO view AUC $= 0.89 \pm 0.04$, CC view AUC $= 0.74 \pm 0.05$ and MLO view AUC $= 0.83 \pm 0.04$). The best AUC in mass/architectural distortion and calcification detection is for DBT z-Stack slice images in CC and MLO view (AUC $= 0.98 \pm 0.01$ and AUC $= 0.97 \pm 0.03$, respectively).

In another work by Singh et al. (2020), the authors proposed a method for adapting a deep learning model trained from FFDM images to be used for DBT images. In their study, FFDM and DBT images were labeled into four classes. The classes are normal (no remarkable finds), benign (notable finds are benign), high-risk (biopsied determined tissue types are highly possible to develop as cancer), and malignant (biopsy proved malignant finds). The model — ResNet with 29 layers — was designed to have patches with dimensions of 512×512 as input. They first trained the model with FFDM, then, the authors trained the model with 2D Maximum Intensity Projections(MIP) of DBT images. They applied histogram matching on MIP images to match them to the FFDM images. Two fine-tuning methods were used in this study: (1) fine-tuning the last two fully connected layers, and (2) fine-tuning just the optimal layers for fine-tuning. The two fine-tuning methods were compared using histogram matched MIP images and original MIP images. The results showed that fine-tuning the last two layers with MIP-HM has the best performance in terms of AUC (AUC $= 0.847$).

In a study by Li et al. (2020a), they compared several cancer classification models (malignant, benign, or normal) for FFDM and DBT images. They extracted ROI from FFDM (dimensions: 256×256) and DBT (dimensions: $256 \times 256 \times 16$). They studied three transfer learning models to analyze DBT and FFDM images: Double Transfer Learning (DTL), Mixture of DBT and FFDM (MIX), and Single Transfer Learning (STL). The DTL model used two-stage transfer learning: the first stage used ImageNet pre-trained VGG, and the second stage used FFDM to train the model again. The MIX model used ImageNet pre-trained VGG and employed a mix of FFDM and DBT images for training. The STL model used ImageNet pre-trained VGG. The results showed that the MIX model has the best performance in term of AUC (malignant AUC $= 0.917$, benign AUC $= 0.951$, and normal AUC $= 0.990$) when applied to DBT images.

In the paper by Kim et al. (2016), they proposed a hierarchical model — latent bilateral feature representation of asymmetric breast tissue — for classifying masses. The main idea of their model is classifying masses based on asymmetry between left and right breasts. In their method, first, volume registration was applied to DBT's main view and lateral view (left and right breast). Then, the volume of interest (VOI) transform was applied. After the VOI transformation, they used a 3D CNN for extracting bilateral feature representation. The features (extracted by the CNN) of the main view and lateral view VOI were concatenated and are input into a fully connected layer (dimension: 512) for mass classification. The results showed that their proposed model performed better (AUC $= 0.847$) than the hand-crafted feature classifier (AUC $= 0.826$).

The article by Yousefi et al. (2018), proposed a model using a CNN as feature selection and a multi-instance random forests (MI-RF) model as classifier. In the data pre-processing module, they denoised DBT z-Stack images and augmented images. In addition, the pectoral muscle was removed from images. The pre-processed z-Stack images from each DBT were inputted to a CNN, which has three convolutional and two fully connected layers followed by ReLU activation, max pooling, and batch normalization layers. The convolutional layers used in this model have 16, 32, and 64 filter's depth with the kernel dimension of 5×5. The learned features from z-Stack images are output from the deep CNN. Then, an MI-RF is used to classify images as benign or malignant. The ML-RF used in this study is formulated based on a multi-instance learner and randomized trees. Their proposed DCNN MI-RF model's accuracy is 86.6%, sensitivity is 87.5%, specificity is 87.5%, and AUC is 0.87.

In the paper by Zhang et al. (2018), the authors developed several CNN models for FFDM classification and DBT classification. The data set has 3018 negative and 272 positive mammogram screens (contained both FFDM and DBT images) in this study. They used data augmentation and transfer learning in their work. Over all proposed models for FFDM, the 2D-T2-Alex showed the best performance (AUC of 0.7274). The 2D-T2-Alex model extracted features (feature dimensions: $25 \times 25 \times 256$) from pre-trained AlexNet, then employed shallow CNN to classify normal or cancer. The shallow CNN contains one convolutional layer with kernel size of 1×1 and filter's depth as 256, followed with two 1024 fully connected layers. In the DBT classification, 3D-T2-Alex showed the best performance (AUC of 0.6632). The 3D-T2-Alex model used the same network structure as the 2D-T2-Alex model, but used DBT images as input (for each exam, they selected a subset of DBT frames).

The same group built up on their prior work mentioned above explored alternative ways to incorporate entire DBT volumes into pre-trained models designed for 2D data. They explored ways to "fuse together" the 2D slice images in the reconstructions (Zhang et al., 2019). Specifically, they proposed a model that consists of an early fusion stage and a late fusion stage. In the early fusion stage, two classification methods have been tested: (i) classification using features extracted from pseudo 2D mammogram generated by averaging all slices, and (ii) classification using features extracted from the reconstructed 2D dynamic image which is an RGB image summarizing information and appearance from a 3D sequence of images or z-Stack. In the late fusion stage, pre-trained AlexNet is used to extract features from DBT z-Stack images. Then, the features are consolidated into a fixed sized feature map for binary classification as benign or malignant. In this study, the authors compared the performances of the early fusion stage and the late fusion stage against their 3D-AlexNet model proposed in their previous study.

They also compared the results of the late fusion stage when using different feature pooling methods including max pooling, average pooling, and minimum pooling. The results showed that the late fusion stage when using max pooling and AlexNet transfer learning provides the best performance with an AUC of 0.854. The AUC is improved by 22.80% from that of their prior model (AUC $= 0.663$).

Continuing previous work, the group then investigated whether training a CNN model using both FFDM and DBT data from the same patient can improve model's diagnostic performance (Liang et al., 2019). Their method can be divided into four stages. The first stage is data pre-processing. In this stage, 2D dynamic images are reconstructed from DBT z-Stack images. Each view of z-Stack images forms one dynamic image which collects all z-

Stack image variations. In the second stage, the reconstructed 2D dynamic images and FFDM images are input to a deep learning model for feature extraction. In the third stage, the AlexNet, ResNet, DenseNet, and SqueezeNet models are used as the deep learning (or backbone) network. The extracted features are input to the fourth stage, which has three classifiers. The first classifier classifies features from dynamic images. The second classifier classifies features that are concatenated from dynamic images and their matching FFDM images. And the third classifier classifies features from FFDM images. An ensemble method, which gives each classifier a weight and sums all weighted classifier results, was employed for the final output. A private data set containing 415 benign patients and 709 malignant patients were used in this study. All the data used in this study were biopsy proved. The results showed that their ensemble model using the AlexNet model had the best performance with an AUC of 0.97.

Matthews et al. (2020) proposed a CNN model to transfer FFDM trained domain to DBT s2D images. The proposed model in this study was pre-trained ResNet and a data set with 78445 s2D images was used. The result showed that the AUC of the proposed model is 0.97.

Rodriguez-Ruiz et al. (2018) employed a CNN model to classify calcification using reconstructed images when employing different DBT reconstruction algorithms (filtered back projection and enhanced multiple parameter iterative reconstruction). They used a shallow network containing four convolutional layers and three fully connected layers. The input dimensions of this model is $29 \times 29 \times 3$. The AUC of the CNN model trained with images that are reconstructed using the filtered back projection method is 0.857, and the AUC of the model trained with images that are reconstructed using the enhanced multiple parameters iterative method is 0.880.

4.1.2. Localization

In addition to classification, another important tasks in analyzing biomedical images are cancer localization (Fig. 2c and b). Various detection and segmentation methods have been proposed for 3D medical images. However, few studies have been published for cancer localization and segmentation for DBT images. Marking abnormalities on DBT slices requires experienced radiologists to review each frame from the DBT z-Stacks and identify tumor locations, which is extremely time consuming considering the number of slices per DBT scanning. Further, the public database for DBT images is very limited. In this section, we will review the object detection and segmentation models proposed for DBT images.

In the paper by Lotter et al. (2019), the authors proposed a three-stage model. In the first stage, a pre-trained ResNet model was trained using FFDM patches for classification (mass, calcifications, focal asymmetry, architectural distortion, and normal). In the second stage, the trained ResNet was used as a backbone model for RetinaNet. In this stage, full FFDM images were used for lesion localization. The third stage consists of two parts: (1) In part A, multiple-instance learning (MIL) — used class likelihoods of multiple bounding boxes for an image — was used to classify FFDM images, and (2) in part B, MIL was used to classify optimized 2D images condensed from DBT z-Stack (ROI extracted from stage two model using z-Stack). The AUC of their proposed model is 0.945.

Fan et al. (2019) introduced a CNN based deep learning method using Faster RCNN for mass detection in DBT images. In this work, the authors proposed three modules: (1) module A is for preprocessing DBT z-Stack images; (2) module B uses the Faster RCNN model (RCNN-based model) for efficient mass detection; and (3) module C uses a deep CNN (DCNN-based model) for reducing False Positives (FPs) and for comparing the RCNN-based model with the DCNN-based model.

In module B, the DBT z-Stack images were passed into the Faster RCNN model one by one to generate bounding boxes for each image (slice) with a score to show the likelihood of mass (confidence score of detected mass). The bounding box of a mass was identified by merging all the bounding boxes from the same set of DBT z-Stack images that have overlapping ratios (intersection over union ratios) greater than 0.5. The probability of detected mass was assigned by the highest confidence score for each image in one set. In module C, the CNN-based model used pre-screening steps for identifying candidate ROI of masses by ranking the 3D gradient field convergence maps' gradient values and selecting the top 10% ROI. For reducing FP rates, these top rank ROI were input into the DCNN with 22 layers. The classification score was obtained using the maximum prediction scores from the top five ROI. To compare the results between the RCNN-based and DCNN-based models, the authors used free-response ROC (FROC) curves that showed a figures-of-merit (FOM) of 0.4374, and a p-value of 0.0011, which suggests that the RCNN-based CAD model performs better compared to the DCNN-based model.

More recently, the same group of researchers compared this 2D-slice–based system to a fully 3D volumetric model (Fan et al., 2020). In their study, they designed a 3D version of the Mask-RCNN model in order to perform segmentation on the same reconstructed volumes from their previous study. The proposed model takes patches with the dimension of $256 \times 256 \times 64$ as input. In their model, ResNet-Feature Pyramid Network (ResNet-FPN) was used as the backbone. The ResNet-FPN extracted different scales of features. In the Feature Pyramid Network (FPN), different scaled features were combined. The Bounding boxes of input images were generated from RPN. Segmentation masks were produced from the mask branch from each ROI using a FCN. Comparing this model to their prior work, they saw better performance (sensitivity of 90% with 0.8 FPs per breast) compared with both 2D-Mask-RCNN (sensitivity of 90% with 1.24 FPs per breast) and Faster RCNN (sensitivity of 90% with 02.38 FPs per breast).

Lai et al. (2020) proposed a method that employed U-Net to segment DBT images. There are six stages in their method: data pre-processing, patch extraction, data augmentation, U-Net segmentation, voting stage, and post-processing. In the data preprocessing stage, image contrast was enhanced by applying a top-hat transform. In the patch extraction stage, images were split into patches to increase the number of data. For data augmentation, patches were rotated by 90 degrees. They used a 23 layers U-Net model for segmentation. After segmentation, the final prediction used one of three voting techniques: (1) majority voting in, by which the final image label was selected from majority class labels of predicted patches, (2) maximum probability in, by which the final image label was selected from the label of the predicted patch with the highest probability, and (3) sum of probabilities in, which predicted patch probabilities are summed and the largest probability over all classes as final prediction class is selected. In the post-processing stage, segmentation was predicted from U-Net, in which less than 50 voxels in volumes were removed as volumetric constraints. They compared the performance of their model to linear discriminant analysis, support vector machine, CNN, and neural network in terms of AUC. Their model out performed the other methods and showed an AUC of 0.859.

In the study by Buda et al. (2021), the authors employed a deep learning model to detect masses and architectural distortions using a data set containing 5610 cases from 5060 patients. 5129 of cases are normal (no abnormal findings), 280 are not biopsy-proven cases (abnormal finds in the report but not biopsy-proven as abnormal), 112 are biopsy-proven benign, and 89 are biopsy-proven cancer (at least one mass or architectural distortion finds). They developed a single-stage deep learning model to detect masses and architectural distortions. The research group makes the data set (from Duke Health System) publicly available at The Cancer Imaging Archive (TCIA) website

(https://www.cancerimagingarchive.net). The study applied data preprocessing by window-leveling images (adjust the image appearance to highlight specific structures), downscaling each slice, removing skin, and extracting the breast region. The proposed model is a single-stage CNN (they employed YOLO) for 2D object detection with DenseNet (Huang et al., 2017). The authors divided each input image into 96×96 patches. The network outputs a confidence score containing the center point of a bounding box for each cell (each cell predicts only one bounding box). During training, for positive samples (cancer), they only selected slices that contain abnormal tumors. The authors tested different loss functions in the experiment, including binary cross-entropy, weighted binary cross-entropy, focal loss, and reduced focal loss. The results showed that their model performed best using focal loss, with a sensitivity of 60% at 2 FPs per DBT volume.

More recently, the same group proposed a unique method for anomaly detection using GANs (Swiecicki et al., 2020). In this study, the authors designed a GAN to generate "normal" breast tissue images using normal DBT data. After training, the model blocks out patches in the test data in a sliding-window fashion and has the GAN attempts to fill in the patches. In theory, if there is a significant deviation between the generated patch and the original patch, then the region has a high probability not be "normal." Though it is a preliminary study, early results indicate that differences between lesions and their GAN-generated counterpart are visually conspicuous, and the average pixel intensity difference between the two is about twice that seen with normal patches.

While most of the previous studies have attempted to identify masses and microcalcifications (both more easily circumscribable anomalies), special mention is deserved to the recent work of Li et al. (2020b), in which the authors designed a model specifically to detect architectural distortions using mammary glands distribution as prior information. The mammary glands distribution includes Gabor magnitude, Gabor orientation field, and convergence map. Gabor magnitude and Gabor orientation field are obtained by using the Gabor filter. The convergence map is obtained by convergence measure, which measures convergent pixels that are oriented towards the center circle in nested two loops. That prior information and DBT slices fed into a Faster RCNN model to predict 2D candidates from each slice. 3D candidates of each DBT volume were finally generated based on fused 2D candidates using a 3D aggregation scheme. The 3D aggregation scheme employed density-based spatial clustering to cluster 2D candidates. The 2D binary candidate masks belonged to the same cluster are concatenated together. The results showed that their proposed model has a sensitivity of 80% with 1.95 FPs per volume.

4.2. Non-diagnostic applications

Besides directly improving the detection accuracy of the cancer detection models, enhancing the consistency and quality of PVs, synthetic images and reconstruction images will also assist the diagnostic process. Many traditional image reconstruction algorithms have been widely applied to medical images such as filtered back projection, algebraic reconstruction, statistical reconstruction, iterative sparse asymptotic minimum variance, and learned iterative reconstruction. In addition, to reconstruct images as a stack of parallel planes, a variety of competing reconstruction techniques have been developed (Suryanarayanan et al., 2001; Zhou et al., 2015; Zhu et al., 2020; Krammer et al., 2019). However, DBT reconstruction using deep learning is still an open area. Recently, a few studies have explored developing deep learning models to reconstruct medical images. In this section, we review studies focusing on deep learning for DBT denoising, generating DBT s2D images and DBT reconstruction images.

4.2.1. Denoising

In the paper by Sahu et al. (2019), a GAN was used for denoising the DBT projections. The proposed model contains a generator that learns the distribution of data and discriminator, which predicts the label of input images (fake or real). The proposed model takes patches (200×200) of DBT projections images. To evaluate their model, they used the Haar Wavelet-Based Perceptual Similarity Index (HaarPSI), which measures the similarity of two images in the range of [0,1]. A HaarPSI closer to 1, the better the similarity will be. They compared their denoise DBT projection results with the Mean Squared Error (MSE) loss based CNN denoise model, K-SVD, and Block-Matching and 3D Filtering (BM3D) denoising algorithms. The results showed that the proposed GAN model has the best performance in terms of HaarPSI (0.383) compared with the MSE loss CNN (0.367), KSVD (0.295), and BM3D (0.2851) models.

Gao et al. (2020) have developed a CNN denoising system for microcalcifications which is applied to PVs: they show that after reconstruction using simultaneous algebraic reconstruction technique, this denoising system actually improves the contrast of microcalcifications, theoretically enhancing human detection of these lesions. The authors used 9 projection images (generated patches with dimension as 32×32) in their study. The patches was grouped into subtle, medium, and obvious depends on the microcalcifications' nominal sizes. The authors showed the CNR values of subtle, medium, and obvious after denoising was improved by 6.7%, 9.7%, and 9.5% compared with before denoising.

4.2.2. Generate s2D

The study by Jiang et al. (2019) proposed using a GAN to synthesize realistic DBT s2D using reconstructed DBT volumes. This research aims to replace FFDM with DBT s2D. In the paper, a Gradient Guided cGANs (GGGAN) was proposed. The GGGAN used gradient of s2D and FFDM as additional inputs to the GGGAN's discriminator. In other words, the gradient feature extracted from an s2D paired with the s2D to compose a fake pair of images, and gradient feature extracted from an FFDM image paired with the FFDM image to compose a real pair. These real and fake pairs were used to train the discriminator. A U-net model was used for GGGAN's generator model. To generate s2Ds, the input of the U-net model is multi-scaled DBT images, and output is simulated images as s2Ds. The data set used in this study contains 1077 pairs of DBTs and FFDMs. To evaluate their model, they employed three radiologists to compare the quality of s2Ds generated from GGGAN with s2Ds generated from cGANs using FFDM images. One thousand pairs of DBTs and FFDMs were randomly selected. All of the examinations by the radiologists showed that the s2Ds generated from GGGAN and cGAN have better quality compared with the real FFDM images, where GGGAN can generate better quality s2Ds compared to cGAN.

4.2.3. Image reconstruction

Teuwen et al. (2020) proposed a deep learning model DBToR for DBT reconstruction from PVs. The DBToR — containing a couple of reconstruction blocks — takes projection images and the breast thickness as initial inputs. The DBToR is a CNN-based model that used the L_2 loss function. The data set they used for experiments were phantom projections and breast CT. For noise-free phantom projection data, they compared the SSIM of the proposed model (0.89 ± 0.01) with those of the Maximum Likelihood for Transmission (MLTR) (0.83 ± 0.042) and Learned Primal-Dual (LPD) (0.38 ± 0.18) models. For breast CT, the SSIM of the proposed model and MLTR are 0.93 and 0.82, respectively.

Building upon their previous work, Michielsen et al. (2020), proposed reconstruction methods that use a low-resolution deep learning model to predicted reconstruction as initialization for

a high-resolution iterative algorithm. The low-resolution DBToR-X reconstruction method, which is a memory-optimized version of DBToR (Teuwen et al., 2020), was used as initialization with MLTR algorithm to produce the final DBT reconstruction. The author compared the proposed model MSE with that of vanilla MLTR (without initialization). The results showed the MSE of DBToR-X initialized MLTR is 101×10^{-6} and the MSE of vanilla MLTR is 177×10^{-6}.

In the study by Wu et al. (2020), they proposed a method that used decoupled unrolled network to improve the contrast and in-depth resolution of DBT reconstruction images. The authors also proposed a novel ROI loss function for adding attention to the region of microcalcifications. The decoupled unrolled network employed U-net to replace the penalty function of the iterative reconstruction algorithm (IR) and unrolled the proximal gradient descent to finite iterations. The network loss used weighted summation of whole-volume L_2 loss and microcalcification ROI L_2 loss. The experiment used 176 DBT of realistic breast phantoms. To evaluate the model's performance, the authors compared the performance of the proposed model with and without employing IR in terms of Relative Mean Squared Error (RMSE) and SSIM. The results showed that the proposed model using the ROI loss has the best performance ($\text{RMSE} = 0.041 \pm 0.005$, $\text{SSIM} = 0.724 \pm 0.098$) compared with using IR ($\text{RMSE} = 0.103 \pm 0.015$, $\text{SSIM} = 0.656 \pm 0.123$).

Another study by Su et al. (2020) proposed a novel method DIR-DBTnet, which employed deep learning with the standard iterative reconstruction algorithm. The authors employed the alternating direction method of multipliers (ADMM) — divides complex objective function into several subproblems — to solve DBT reconstruction problem. The authors used CNN extracted information as regularizer in ADMM, and iterated parameters as learnable variables. The network structure contains three parts: the reconstruction module, denoising module, and multiplier update module. The denoising module employed a CNN to generate denoised images. In the experimental study, simulated DBT sinogram-label pairs data were used.

5. Discussion

Several deep-learning–based CAD systems for DBT are currently available, and many new systems will be developed. However, there are still numerous challenges to be overcome. As Wang et al. (2020b) recently demonstrated, even peer-reviewed, published models for FFDM classification fail when applied to different data sets, even when those data sets include acquisitions using similar equipment. Nonetheless, the trained radiologist can adapt when looking at different data sets, indicating that high-performing deep learning models are still missing the "key" features which distinguish disease from normal.

5.1. The role of reconstruction

The concept of image reconstruction (i.e. tomosynthesis) in DBT is unique when compared to other common imaging modalities. FFDMs are fundamentally single X-ray images bounded by physical constraints. By contrast, fully-3D methods such as CT represent a well-defined inverse problem with an optimal (but not necessarily easy to compute) reconstruction. DBT PVs have a limited angle range, resulting in less information regarding abnormalities. As a result, there is no deterministic optimal reconstruction: in fact, it is unclear even what an "optimal" reconstruction should look like. While detection models operating on reconstructed or s2D images have shown success, further development of reconstruction methods can pose new challenges for such detection models.

As previously noted, several detection models use the original PVs, which it may be one way to produce a reconstruction-method–invariant detection model. Interestingly, it appears that this push-and-pull of detection and reconstruction may be leading to the co-development of two different but equally important applications of deep learning for DBT analysis: low-level models can use "raw" data for detection purposes, while higher-level models can develop reconstructions optimized for human-eye–based detection and diagnosis. The composition of these two levels of deep-learning models may produce an optimal result for patients. Low-level detection models can operate on data that has not been pre-processed but may be harder for humans to interpret, and then these possible lesions can be projected onto reconstructions optimized for scrutiny by radiologists and surgeons. As has been noted e.g. Pisano, 2020), one major concern related to AI detection models is their interpret-ability that their underlying algorithms make it difficult for physicians to communicate findings to patients and for both patients and physicians to trust these findings. This paired model structure may mitigate some of these issues by displaying these detected latent lesions more clearly on an visually-optimized image to facilitate better communication and understanding.

5.2. How many dimensions?

Because tomosynthesis lies in this quasi-3D domain, the question of whether to use 2D or 3D models is central to creating effective diagnostic tools. In the studies we reviewed in this work, most studies employed 2D model using DBT slices as independent images to overcome the problem of limited computational resources. However, this independent assumption does not take advantage of DBT volume data, which is corresponding to tissue relation between each slice. Building a 3D model which takes volume as input and learns underlying features between each slice is necessary. However, unlike MRI, DBT volumes are high resolution. Using high resolution data to train a 3D model requires huge computation resources that are often unpractical. Moreover, resizing a DBT volume is not recommended since scaling down the resolution may result in losing the characteristics of the tumors as can be seen in Fig. 4.

All of this information is also important as researchers wrangle with 3D-CNN models which are significantly more complex and computationally demanding. However, because 3D CNNs rely on reconstructed images, it remains unclear whether these systems actually will outperform 2D CNNs on the s2D of the same reconstructed images, with conflicting results thus far (Zhang et al., 2018; Fan et al., 2020). Preliminary results indicate that not only large training sets, but also full-resolution (not downsampled) inputs greatly improve diagnostic capability (Seyyedi et al., 2020).

5.3. Domain-crossing

One change can help improve clinical implementation is the use of a common language for evaluating and comparing models. As noted in Table 5, clinicians and computer scientists evaluate models in fundamentally different ways. While many diagnostic tests use metrics such as sensitivity, specificity and ROC integrals to diagnostic evaluation, screening imaging is not diagnostic imaging. Since high specificity is required for diagnostic, it is unlikely that tissue biopsy will be replaced as the gold standard for cancer diagnosis in the near future. As a result, in practice, the question is whether to recall the patient for further workup, which impose additional cost and anxiety or not (Fig. 7).

We believe that developing a common set of metrics for both clinical and computational evaluation will go a long way toward allowing for the objective comparison of different AI-CAD models. Some of these differences are being bridged already. For example, (likely because of its clear commercial applications) optimizing reading time has been a particularly good example of this domain-crossing to improve clinical practice. How-

Table 5
Some of the most commonly-used metrics to compare ML models. While the classical objective metrics of screening test performance are useful for comparing ML models (especially when using a binary classifier), adoption of any model depends on improving patient and staff quality metrics as well.

Classical metrics	
Sensitivity	The ability to detect abnormalities of any significance $\frac{TP}{TP+FN}$
Specificity	The ability to discriminate between cancerous and non-cancerous lesions $\frac{TN}{TN+FP}$
Area under ROC curve (AUC)	An approximation of the probability that a TP will be viewed as "more abnormal" than a TN
Practical metrics	
Recall rate	Percentage of patients who require additional imaging or procedures following a screening mammogram
Cancer detection rate	Percentage of cancers diagnosed in all patients (after recall)
PPV1	Positive predictive value of abnormal screening mammogram (Fraction of recalls leading to proven diagnosis of malignancy)
Average Glandular Dose (AGD)	Amount of radiation delivered to sensitive tissue
Reading time	Time spent by physician reading mammogram
Cost	Includes fixed costs as well as technician and radiologist time

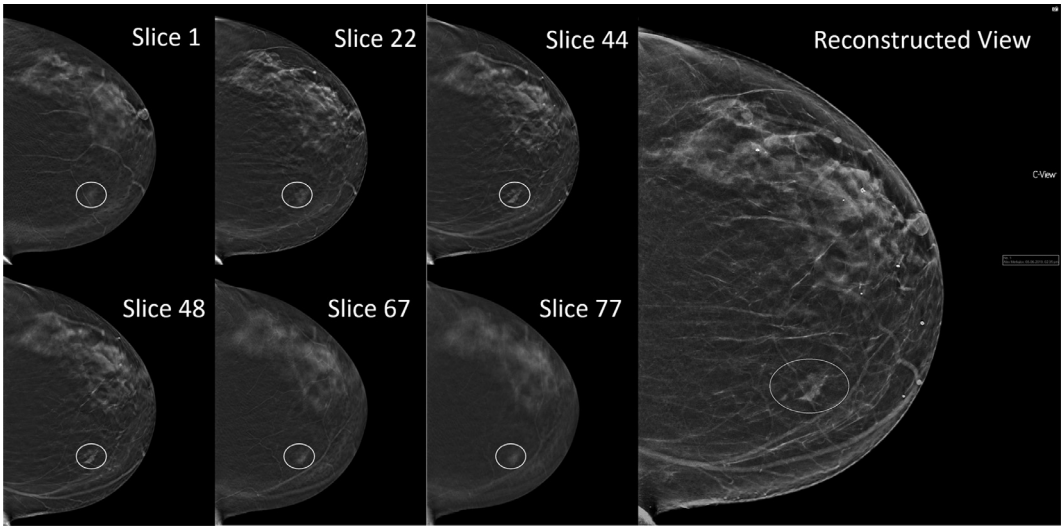

Fig. 3. DBT selected slices from a z-Stack and reconstructed 2D view. The white circle indicates the location of the malignant mass.

ever, we still are seeing improvement on this front. As an example, Seyyedi et al. (2020) used a novel set of classifiers: "normal," "benign," and "further workup needed," which appears to better align with clinical objectives.

5.4. Data curation and availability

One major DBT-specific challenge is proper data curation and labeling. However, this does not take into account the volumetric nature of reconstructions (or the overlapping regions in PVs), which both add significantly to the complexity and labor of curating data. Marking tumors in 3D is certainly a harder task, and the best way to draw boundaries is unclear, especially when many of the aforementioned models using overlapping boundary boxes on each 2D-slice for detection.

Looking only at 2D slices, it is still unclear whether models operate better using abnormalities labeled using bounding boxes or tightly-drawn margins of lesions (Wang et al., 2020a). This challenge is compounded by the additional dimensions and images in DBT. As Fig. 3 shows, malignant tumor boundaries present differently depending on how close the slice is to the center of the tumor. Hence, labeling DBT slices at pixel level remains a challenge.

DBT studies easily require more storage than FFDM by an order of magnitude or more. While most of the aforementioned studies used small, potentially non-generalizable data from private data sets, we remain optimistic on the data availability front. The recent release of the first publicly-accessible annotated DBT dataset will allow many more researchers to experiment with different curation techniques, and we applaud Buda et al. (2021) for making such a valuable resource widely available. Hopefully others will follow this effort, allowing a broader community of researchers to explore all of these avenues in greater detail.

5.5. Lesion classification

As previously discussed, one demonstrated advantage of DBT is its ability to distinguish small lesions which may obscured in the projections obtained using FFDM. According to the Breast Imaging-Reporting and Data System (BI-RADS), cancerous lesions are identified using four hallmarks: (1) the presence of asymmetry, (2) a mass, (3) microcalcifications or (4) architectural distortion (D'Orsi, 2014). Different methods work better for each type of abnormality, although most work has focused on masses and microcalcifications, which are easier to label and bound than the other two types of abnormalities. However, identifying the distortions which are not as easy to annotate is one of the real strengths of 3D imaging. Li et al. (2020b) reasonably claim that FFDM is unable to make such structure maps because of the significant amount of overlapping tissue, suggesting that AI-CAD systems for DBT may be able to find diagnoses that were previously impossible to find using FFDM. As such, there is an important role for general anomaly detection methods which are insensitive to the type of abnormality and which do not necessarily attempt to localize a particular lesion.

Intuitively, the identification of "smoother" masses seems to require a different approach when compared to the "sharp" delineation of microcalcifications, although some new methods show some promise for handling general anomalies.

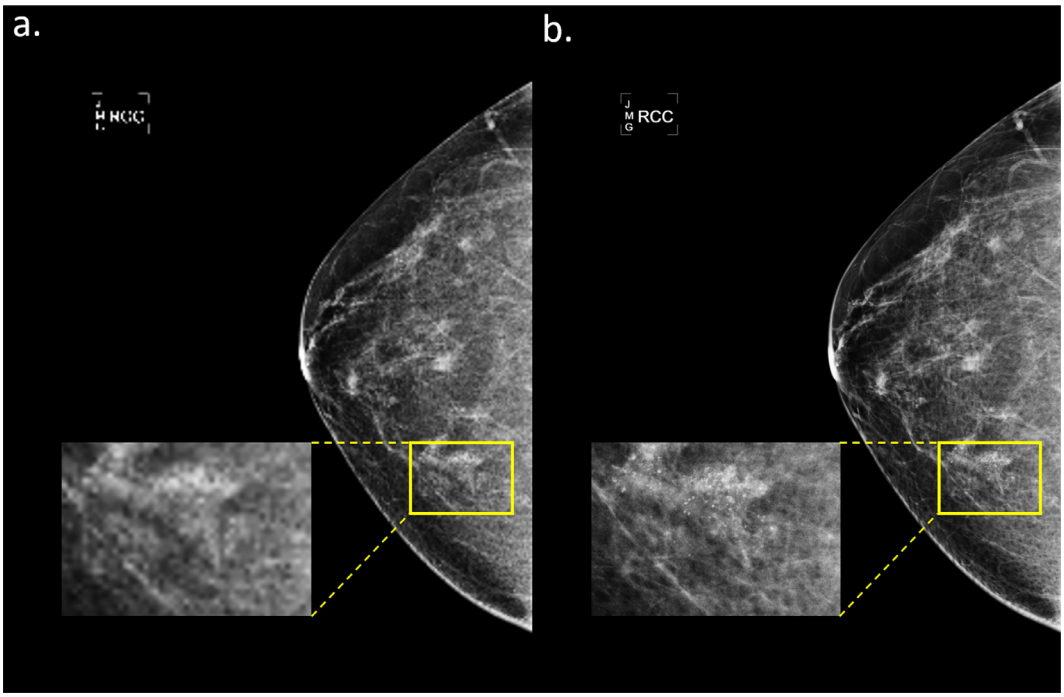

Fig. 4. Mammograms with microcalcification in two different resolutions. The yellow box indicates the location of malignant microcalcification. a. image resize of 332×256 b. image with original dimension as 3328×2560. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

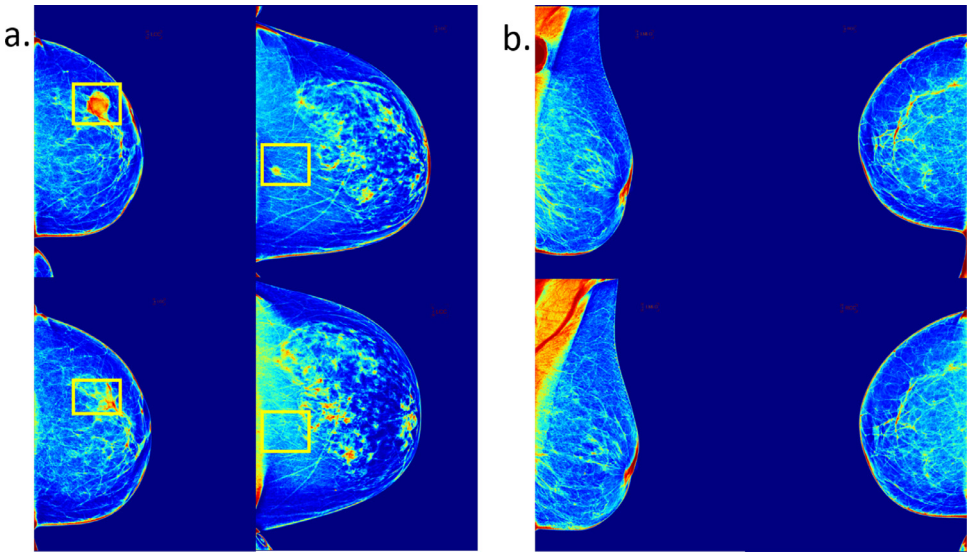

Fig. 5. The heat map of previous year exam vs. current year exam. Top row image are current year examinations and bottom row image are previous year examinations a. Mass tumor comparison between the previous year and current year. Yellow box indicates mass identified by radiologists. b. Normal breast mammogram comparison between the previous year and current year. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

These systems appear to be successful based on these small preliminary studies, and we look forward to future results based on such detection methods in the future.

5.6. Prior images and previously occult tumors

The availability and incorporation of prior imaging studies with current ones is an essential component of radiology practice. This is certainly true in mammography, and models using prior normal FFDM images appear to improve diagnostic accuracy (Park et al., 2019). In clinical practice, radiologists always compare the previous year's exam with the current year's exam to observe tissue changes for diagnosing cancer (as Fig. 5 demonstrated tissue changes from the previous year exam to the current year exam in both cancer and normal cases). Studies can further integrate related prior knowledge in the form of patients' clinical information to improve diagnostic accuracy, and previous year mammogram examination has not been studied for developing deep learning models for analyzing DBT images.

While more work on such multimodal methods is needed as Fig. 6 illustrates, it is also interesting to think about how such work may expand the frontiers of breast imaging. As an example, Arefan et al. (2020) built a model which attempts to estimate short-term cancer risk using normal prior FFDM images. In addition to better estimating cancer risk, it is possible that such studies can also be used to find previously occult tumors and improve

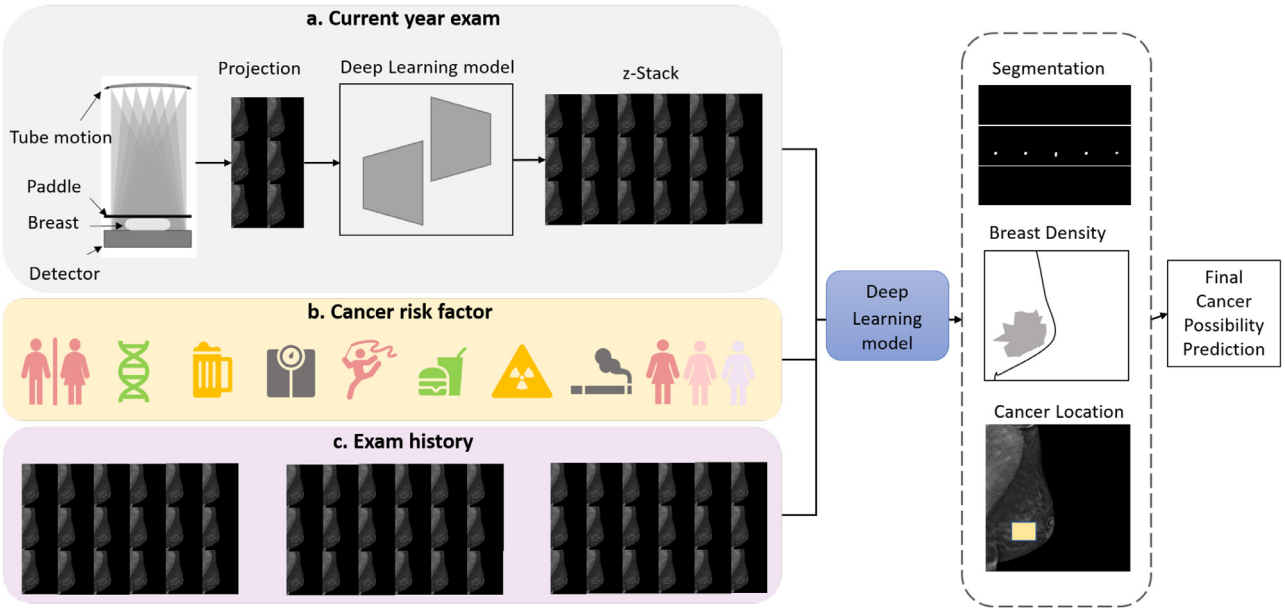

Fig. 6. Comprehensive CAD system. a. An embedding deep learning model in the current year exam for high-quality DBT reconstruction. b. and c. integrate cancer risk factor and exam history as input to a comprehensive deep learning model to predict cancer segmentation, breast density, cancer location, and the final possibility of breast cancer.

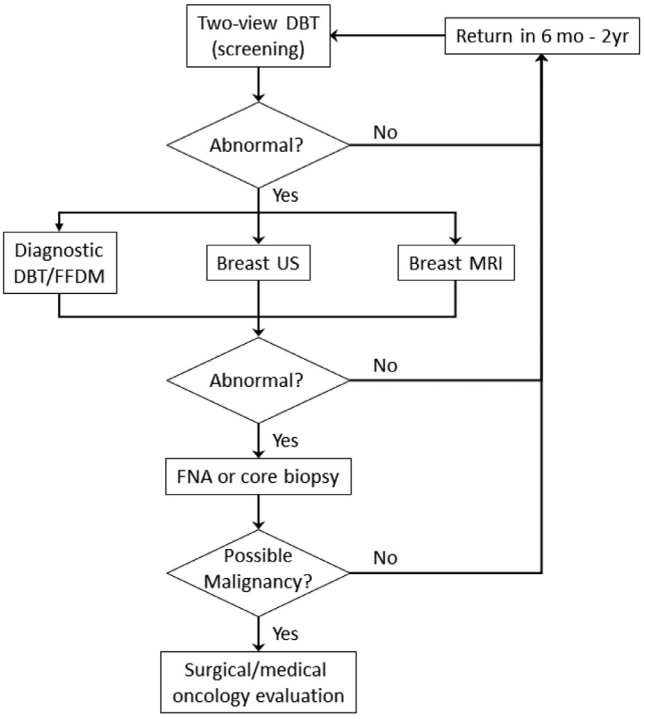

Fig. 7. An abbreviated flowchart outlining the key decisions and procedures leading from screening mammography to definitive cancer diagnosis. Despite effective standardization tools, the American College of Radiology still leaves the medical decision-making to the radiologist (ACR, 2018), and practices can differ (see e.g. Ebuoma et al., 2015).

early diagnosis. Some of these images may in fact not be truly normal, and may contain abnormal features that only a sophisticated deep learning model might extract. We look forward to further research in this area.

6. Conclusions

Closing the gap between the aims of the computer scientist and the clinician will be essential to the integration of deep learning models into clinical workflows. One of the key distinctions between the computational researcher's goals and the clinician's goals is the determination of "benign" and "malignant". In reality, these designations represent a spectrum between fully inert and highly lethal, and decisions are made by taking a patient's entire case and medical history into account.

One can argue that determining the level of lethality from a simple mammogram or DBT study may not be the true endpoint, as it would still be prudent to employ other imaging methods and biopsy techniques prior to taking a patient to the operating room.

For FFDMs, deep-learning–based detection models have been shown to perform with near-human accuracy (McKinney et al., 2020); as more studies and data become available, there is no reason to believe that this should be any different for DBT.

However, it is also evident that the most effective models - specifically their components and structures - differ from those used for the analysis of conventional (i.e. non-diagnostic) images. Nonetheless, one can see that the use of deep learning for automatic breast cancer detection in DBT can have a clear benefit for healthcare throughput and cost reduction.

The ways in which deep learning models have been adapted for this emerging technology comprise an important case study into how AI-based tools can adapt to new, yet-undiscovered imaging technologies. Open problems specific to DBT such as optimal reconstruction methods for human interpretation demonstrate how the next generation of deep learning models must be developed hand-in-hand with the next generation of imaging technologies. With many recent successes in the application of deep learning models to DBT studies, we look forward to discovering even more novel ways of seamlessly bridging the gap between the computational and clinical domains.

Declaration of Competing Interest

Authors declare that they have no conflict of interest.

CRediT authorship contribution statement

Jun Bai: Conceptualization, Investigation, Writing - original draft, Writing - review & editing. **Russell Posner:** Investigation, Writing - original draft. **Tianyu Wang:** Investigation, Writing -

original draft. **Clifford Yang:** Writing - review & editing. **Sheida Nabavi:** Funding acquisition, Writing - review & editing, Supervision.

Acknowledgments

This work is supported by a grant from the University of Connecticut Research Excellence Program, PIs: Nabavi and Yang; and Jun Bai's Cigna Graduate Fellowship from University of Connecticut.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2021.102049

References

Abdelhafiz, D., Yang, C., Ammar, R., Nabavi, S., 2019. Deep convolutional neural networks for mammography: advances, challenges and applications. BMC Bioinform. 20 (S11), 281. doi:10.1186/s12859-019-2823-4. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2823-4

Al-masni, M., Al-antari, M., Park, J.-M., 2018. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Comput. Methods Prog. Biomed. 157, 85–94. doi:10.1016/j.cmpb.2018.01.017. Publisher: Elsevier, https://www.sciencedirect.com/science/article/pii/S0169260717314980

Arefan, D., Mohamed, A.A., Berg, W.A., Zuley, M.L., Sumkin, J.H., Wu, S., 2020. Deep learning modeling using normal mammograms for predicting breast cancer risk. Med. Phys. 47 (1), 110–118. doi:10.1002/mp.13886. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mp.13886

Aujero, M.P., Gavenonis, S.C., Benjamin, R., Zhang, Z., Holt, J.S., 2017. Clinical performance of synthesized two-dimensional mammography combined with tomosynthesis in a large screening population. Radiology 283 (1), 70–76. doi:10.1148/radiol.2017162674. Publisher: Radiological Society of North America, http://pubs.rsna.org/doi/full/10.1148/radiol.2017162674

Balleyguier, C., Arfi-Rouche, J., Levy, L., Toubiana, P.R., Cohen-Scali, F., Toledano, A.Y., Boyer, B., 2017. Improving digital breast tomosynthesis reading time: a pilot multi-reader, multi-case study using concurrent computer-aided detection (CAD). Eur. J. Radiol. 97, 83–89. doi:10.1016/j.ejrad.2017.10.014. http://www.sciencedirect.com/science/article/pii/S0720048X17304072

Ben-Cohen, A., Diamant, I., Klang, E., Amitai, M., Greenspan, H., 2016. Fully convolutional network for liver segmentation and lesions detection. In: Carneiro, G., Mateus, D., Peter, L., Bradley, A., Tavares, J.M.R.S., Belagiannis, V., Papa, J.P., Nascimento, J.C., Loog, M., Lu, Z., Cardoso, J.S., Cornebise, J. (Eds.), Deep Learning and Data Labeling for Medical Applications, 10008. Springer International Publishing, Cham, pp. 77–85. doi:10.1007/978-3-319-46976-8_9. Series Title: Lecture Notes in Computer Science., http://link.springer.com/10.1007/978-3-319-46976-8_9

Benedikt, R.A., Boatsman, J.E., Swann, C.A., Kirkpatrick, A.D., Toledano, A.Y., 2017. 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. doi:10.2214/AJR.17.18185. Publisher: American Roentgen Ray Society, http://www.ajronline.org/

Bi, L., Kim, J., Kumar, A., Feng, D., Fulham, M., 2017. Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs). In: Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment, pp. 43–51. doi:10.1007/978-3-319-67564-0_5. https://link.springer.com/chapter/

Buda, M., Saha, A., Walsh, R., Ghate, S., Li, N., Świçicki, A., Lo, J. Y., Mazurowski, M. A., 2021. Detection of masses and architectural distortions in digital breast tomosynthesis: a publicly available dataset of 5060 patients and a deep learning model. arXiv:2011.07995[cs, eess]

Cao, Z., Duan, L., Yang, G., Yue, T., Chen, Q., 2019. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. BMC Med. Imaging 19 (1), 1–9. doi:10.1186/s12880-019-0349-x. Number: 1 Publisher: BioMed Central. https://bmcmedimaging.biomedcentral.com/articles/

Chae, E.Y., Kim, H.H., Jeong, J.-w., Chae, S.-H., Lee, S., Choi, Y.-W., 2019. 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), 2518–2525. doi:10.1007/s00330-018-5886-0.

Chlebus, G., Schenk, A., Moltz, J.H., Ginneken, B.v., Hahn, H.K., Meine, H., 2018. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. Sci. Rep. 8 (1), 1–7. doi:10.1038/s41598-018-33860-7. Number: 1 Publisher: Nature Publishing Group. https://www.nature.com/articles/s41598-018-33860-7

Choi, Y., Woo, O.-h., Shin, H.-s., Cho, K.R., Seo, B.K., Choi, G.-Y., 2019. Quantitative analysis of radiation dosage and image quality between digital breast tomosynthesis (DBT) with two-dimensional synthetic mammography and full-field digital mammography (FFDM). Clin. Imaging 55, 12–17. doi:10.1016/j.clinimag.2019.01.014. http://www.sciencedirect.com/science/article/pii/S0899707119300142

Ciatto, S., Houssami, N., Bernardi, D., Caumo, F., Pellegrini, M., Brunelli, S., Tuttobene, P., Bricolo, P., Fantó, C., Valentini, M., Montemezzi, S., Macaskill, P., 2013. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol. 14 (7), 583–589. doi:10.1016/S1470-2045(13)70134-7. http://www.sciencedirect.com/science/article/pii/S1470204513701347

Cupples, T.E., Cunningham, J.E., Reynolds, J.C., 2012. Impact of computer-aided detection in a regional screening mammography program. Am. J. Roentgenol. doi:10.2214/AJR.04.1300. https://www.ajronline.org/

Dean, J.C., Ilvento, C.C., 2012. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. Am. J. Roentgenol. doi:10.2214/AJR.05.0111. https://www.ajronline.org/doi/full/10.2214/AJR.05.0111

Doi, K., 2007. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput. Med. Imaging Graph. 31 (4-5), 198–211. doi:10.1016/j.compmedimag.2007.02.002.

D'Orsi, C., 2014. 2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System. American College of Radiology. https://books.google.com/books?id=fxUrnwEACAAJ

Ebuoma, L.O., Roark, A.A., Sedgwick, E.L., 2015. Practical considerations for integrating digital breast tomosynthesis into clinical practice. J. Am. Coll. Radiol. 12 (9), 944–946. doi:10.1016/j.jacr.2015.05.003. http://www.sciencedirect.com/science/article/pii/S1546144015003907

Ezhilarasi, R., Varalakshmi, P., 2018. Tumor detection in the brain using faster R-CNN. In: 2018 2nd International Conference on 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 388–392. doi:10.1109/I-SMAC.2018.8653705.

Fan, M., Li, Y., Zheng, S., Peng, W., Tang, W., Li, L., 2019. Computer-aided detection of mass in digital breast tomosynthesis using a faster region-based convolutional neural network. Methods 166, 103–111. doi:10.1016/j.ymeth.2019.02.010. http://www.sciencedirect.com/science/article/pii/S1046202318303608

Fan, M., Zheng, H., Zheng, S., You, C., Gu, Y., Gao, X., Peng, W., Li, L., 2020. Mass detection and segmentation in digital breast tomosynthesis using 3D-mask region-based convolutional neural network: a comparative analysis. Front. Mol. Biosci. 7. doi:10.3389/fmolb.2020.599333. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7686533/

Fotin, S. V., Yin, Y., Haldankar, H., Hoffmeister, J. W., Periaswamy, S., 2016. Detection of soft tissue densities from digital breast tomosynthesis: comparison of conventional and deep learning approaches. San Diego, California, United States, p. 97850X. http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2217045. 10.1117/12.2217045

Freer, P.E., Riegert, J., Eisenmenger, L., Ose, D., Winkler, N., Stein, M.A., Stoddard, G.J., Hess, R., 2017. Clinical implementation of synthesized mammography with digital breast tomosynthesis in a routine clinical practice. Breast Cancer Res. Treat. 166 (2), 501–509. doi:10.1007/s10549-017-4431-1.

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., Greenspan, H., 2018. Synthetic data augmentation using GAN for improved liver lesion classification. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 289–293. doi:10.1109/ISBI.2018.8363576.

Gao, M., Samala, R.K., Fessler, J.A., Chan, H.-P., 2020. Deep convolutional neural network denoising for digital breast tomosynthesis reconstruction. In: Medical Imaging 2020: Physics of Medical Imaging. International Society for Optics and Photonics, p. 113120Q. doi:10.1117/12.2549361. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11312/113120Q/Deep-convolutional-neural-network-denoising-for-digital-breast-tomosynthesis-reconstruction/10.1117/12.2549361.short

Gao, Y., Babb, J.S., Toth, H.K., Moy, L., Heller, S.L., 2017. Digital breast tomosynthesis practice patterns following 2011 FDA approval: a survey of breast imaging radiologists. Acad. Radiol. 24 (8), 947–953. doi:10.1016/j.acra.2016.12.011. http://www.sciencedirect.com/science/article/pii/S1076633217300132

Gastounioti, A., Oustimov, A., Hsieh, M.-K., Pantalone, L., Conant, E.F., Kontos, D., 2018. Using convolutional neural networks for enhanced capture of breast parenchymal complexity patterns associated with breast cancer risk. Acad. Radiol. 25 (8), 977. doi:10.1016/j.acra.2017.12.025. Publisher: NIH Public Access. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6026048/

GE Corporation,. Tomosynthesis and Synthetic 2D Images-A 3D Dose-Equivalent Solution for Screening Mammography. Technical Report. ht*tp://ww*w3.gehealthcare.com.pa//media/downloads/us/product/product-categories/mammography/seno-pristina/adapt%20white%20paper%20jb47137us(1).pdf?Parent=%7B294FC96F-E720-4924-9486-254F92000F7B%7D.

Greenberg, J.S., Javitt, M.C., Katzen, J., Michael, S., Holland, A.E., 2014. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. Am. J. Roentgenol. 203 (3), 687–693. doi:10.2214/AJR.14.12642. Publisher: American Roentgen Ray Society. https://www.ajronline.org/doi/full/10.2214/AJR.14.12642

Haas, B.M., Kalra, V., Geisel, J., Raghu, M., Durand, M., Philpotts, L.E., 2013. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. Radiology. https://pubs.rsna.org/doi/abs/10.1148/radiol.13130307

Hakim, C.M., Catullo, V.J., Chough, D.M., Ganott, M.A., Kelly, A.E., Shinde, D.D., Sumkin, J.H., Wallace, L.P., Bandos, A.I., Gur, D., 2015. Effect of the availability of prior full-field digital mammography and digital breast tomosynthesis images on the interpretation of mammograms. Radiology 276 (1), 65–72. doi:10.1148/radiol.15142009. Publisher: Radiological Society of North America. http://pubs.rsna.org/doi/full/10.1148/radiol.15142009

Hayward, J.H., Ray, K.M., Wisner, D.J., Kornak, J., Lin, W., Joe, B.N., Sickles, E.A., 2016. Improving screening mammography outcomes through comparison with multiple prior mammograms. AJR Am. J. Roentgenol. 207 (4), 918–924. doi:10.2214/AJR.15.15917.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative Adversarial Networks. arXiv e-prints, arXiv:1406.2661.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2018. Mask R-CNN. arXiv:1703.06870[cs],

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. arXiv:1502.01852[cs],

Huang, G., Liu, Z., Maaten, L.V.D., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269. doi:10.1109/CVPR.2017.243. ISSN: 1063-6919

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv:1502.03167[cs]

Jiang, G., Lu, Y., Wei, J., Xu, Y., 2019. Synthesize mammogram from digital breast tomosynthesis with gradient guided cGANs. In: Medical Image Computing and Computer Assisted Intervention MICCAI 2019. Springer, Cham, pp. 801–809. doi:10.1007/978-3-030-32226-7_89. https://link.springer.com/chapter/10.1007/978-3-030-32226-7_89

Kalager, M., Adami, H.-O., Bretthauer, M., Tamimi, R.M., 2012. Overdiagnosis of invasive breast cancer due to mammography screening: results from the norwegian screening program. Ann. Internal Med. 156 (7), 491–499. doi:10.7326/0003-4819-156-7-201204030-00005. Publisher: American College of Physicians. http://www.acpjournals.org/doi/10.7326/0003-4819-156-7-201204030-00005

Khan, A., Sohail, A., Zahoora, U., Qureshi, A.S., 2020. A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. doi:10.1007/s10462-020-09825-6. http://link.springer.com/10.1007/s10462-020-09825-6

Khan, S., Yong, S.-P., 2017. A deep learning architecture for classifying medical images of anatomy object. In: 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 1661–1668. doi:10.1109/APSIPA.2017.8282299.

Kim, D.H., Kim, S.T., Ro, Y.M., 2016. 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), pp. 927–931. doi:10.1109/ICASSP.2016.7471811.

Krammer, J., Zolotarev, S., Hillman, I., Karalis, K., Stsepankou, D., Vengrinovich, V., Hesser, J., M. Svahn, T., 2019. Evaluation of a new image reconstruction method for digital breast tomosynthesis: effects on the visibility of breast lesions and breast density. BJR 92 (1103), 20190345. doi:10.1259/bjr.20190345. Publisher: The British Institute of Radiology. https://www.birpublications.org/doi/abs/10.1259/bjr.20190345

Lai, X., Yang, W., Li, R., 2020. dbt masses automatic segmentation using U-Net neural networks. Comput. Math. Methods Med. 2020, 1–10. doi:10.1155/2020/7156165. https://www.hindawi.com/journals/cmmm/2020/7156165/

Lai, Y.-C., Ray, K.M., Lee, A.Y., Hayward, J.H., Freimanis, R.I., Lobach, I.V., Joe, B.N., 2018. Microcalcifications detected at screening mammography: synthetic mammography and digital breast tomosynthesis versus digital mammography. Radiology 289 (3), 630–638. doi:10.1148/radiol.2018181180. Publisher: Radiological Society of North America. http://pubs.rsna.org/doi/full/10.1148/radiol.2018181180

Lakhani, P., Sundaram, B., 2017. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 284 (2), 574–582. doi:10.1148/radiol.2017162326. http://pubs.rsna.org/doi/10.1148/radiol.2017162326

LeCun, Y., Bottou, L., Bengio, Y., Ha, P., 1998. Gradient-Based Learning Applied to Document Recognition, 46.

Lehman, C.D., Wellman, R.D., Buist, D.S.M., Kerlikowske, K., Tosteson, A.N.A., Miglioretti, D.L., 2015. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern. Med. 175 (11), 1828–1837. doi:10.1001/jamainternmed.2015.5231. https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2443369

Lei, Y., Harms, J., Wang, T., Liu, Y., Shu, H.-K., Jani, A.B., Curran, W.J., Mao, H., Liu, T., Yang, X., 2019. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Med. Phys. 46 (8), 3565–3581. doi:10.1002/mp.13617. https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.13617

Li, S., Dong, M., Du, G., Mu, X., 2019. Attention dense-U-Net for automatic breast mass segmentation in digital mammogram. IEEE Access 7, 59037–59047. doi:10.1109/ACCESS.2019.2914873. Conference Name: IEEE Access

Li, X., Qin, G., He, Q., Sun, L., Zeng, H., He, Z., Chen, W., Zhen, X., Zhou, L., 2020. Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification. Eur. Radiol. 30 (2), 778–788. doi:10.1007/s00330-019-06457-5.

Li, Y., He, Z., Lu, Y., Ma, X., Guo, Y., Xie, Z., Xu, Z., Chen, W., Chen, H., 2020. Deep learning on mammary glands distribution for architectural distortion detection in digital breast tomosynthesis. Phys. Med. Biol. doi:10.1088/1361-6560/ab98d0. http://iopscience.iop.org/article/10.1088/1361-6560/ab98d0

Liang, G., Wang, X., Zhang, Y., Xing, X., Blanton, H., Salem, T., Jacobs, N., 2019. Joint 2D-3D breast cancer classification. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 692–696. doi:10.1109/BIBM47256.2019.8983048.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017. Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV) doi:10.1109/iccv.2017.324.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440.

Lotter, W., Diab, A. R., Haslam, B., Kim, J. G., Grisot, G., Wu, E., Wu, K., Onieva, J. O., Boxerman, J. L., Wang, M., Bandler, M., Vijayaraghavan, G., Sorensen, A. G., 2019. Robust breast cancer detection in mammography and digital breast tomosynthesis using annotation-efficient deep learning approach. arXiv:1912.11027[cs, eess],

Lévy, D., Jain, A., 2016. Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks. arXiv e-prints, arXiv:1612.00542

Matthews, T. P., Singh, S., Mombourquette, B., Su, J., Shah, M. P., Pedemonte, S., Long, A., Maffit, D., Gurney, J., Hoil, R. M., Ghare, N., Smith, D., Moore, S. M., Marks, S. C., Wahl, R. L., 2020. A multi-site study of a breast density deep learning model for full-field digital mammography and digital breast tomosynthesis exams. arXiv:2001.08383[cs, eess]

McDonald, E.S., Oustimov, A., Weinstein, S.P., Synnestvedt, M.B., Schnall, M., Conant, E.F., 2016. Effectiveness of digital breast tomosynthesis compared with digital mammography: outcomes analysis from 3 years of breast cancer screening. JAMA Oncol. 2 (6), 737–743. doi:10.1001/jamaoncol.2015.5536. Publisher: American Medical Association http://jamanetwork.com/journals/jamaoncology/fullarticle/2491465

McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.C., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F.J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C.J., King, D., Ledsam, J.R., Melnick, D., Mostofi, H., Peng, L., Reicher, J.J., Romera-Paredes, B., Sidebottom, R., Suleyman, M., Tse, D., Young, K.C., Fauw, J.D., Shetty, S., 2020. International evaluation of an AI system for breast cancer screening. Nature 577 (7788), 89–94. doi:10.1038/s41586-019-1799-6. http://www.nature.com/articles/s41586-019-1799-6

Mendel, K., Li, H., Sheth, D., Giger, M., 2019. Transfer learning from convolutional neural networks for computer-aided diagnosis: a comparison of digital breast tomosynthesis and full-field digital mammography. Acad. Radiol. 26 (6), 735–743. doi:10.1016/j.acra.2018.06.019. http://www.sciencedirect.com/science/article/pii/S1076633218303337

Michielsen, K., Moriakov, N., Teuwen, J., Sechopoulos, I., 2020. Deep Learning-based Initialization of Iterative Reconstruction for Breast Tomosynthesis. arXiv:2009.01538[physics]

Miglioretti, D.L., Abraham, L., Lee, C.I., Buist, D.S.M., Herschorn, S.D., Sprague, B.L., Henderson, L.M., Tosteson, A.N.A., Kerlikowske, K., 2019. Digital breast tomosynthesis: radiologist learning curve. Radiology 291 (1), 34–42. doi:10.1148/radiol.2019182305. Publisher: Radiological Society of North America. http://pubs.rsna.org/

Miller, A.B., Wall, C., Baines, C.J., Sun, P., To, T., Narod, S.A., 2014. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. BMJ 348. doi:10.1136/bmj.g366. Publisher: British Medical Journal Publishing Group Section: Research http://www.bmj.com/content/348/bmj.g366

Mirza, M., Osindero, S., 2014. Conditional Generative Adversarial Nets. arXiv:1411.1784[cs, stat]

Nam, K.J., Han, B.-K., Ko, E.S., Choi, J.S., Ko, E.Y., Jeong, D.W., Choo, K.S., 2015. Comparison of full-field digital mammography and digital breast tomosynthesis in ultrasonography-detected breast cancers. Breast 24 (5), 649–655. doi:10.1016/j.breast.2015.07.039. http://www.sciencedirect.com/science/article/pii/S0960977615001770

Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., Shen, D., 2018. Medical image synthesis with deep convolutional adversarial networks. IEEE Trans. Biomed. Eng. 65 (12), 2720–2730. doi:10.1109/TBME.2018.2814538.

Nishikawa, R.M., Bae, K.T., 2018. Importance of better human-computer interaction in the era of deep learning: mammography computer-aided diagnosis as a use case. J. Am. Coll. Radiol. 15 (1), 49–52. doi:10.1016/j.jacr.2017.08.027. https://www.jacr.org/article/S1546-1440(17)31021-9/abstract

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., Rueckert, D.,. Attention U-Net: Learning Where to Look for the Pancreas, 10.

Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Acharya, U.R., 2020. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput. Biol. Med. 121, 103792. doi:10.1016/j.compbiomed.2020.103792. Publisher: Elsevier. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7187882/

Park, J., Phang, J., Shen, Y., Wu, N., Kim, S. G., Moy, L., Cho, K., Geras, K. J., 2019. Screening Mammogram Classification with Prior Examshttps://arxiv.org/abs/1907.13057v1.

Pisano, E.D., 2020. AI shows promise for breast cancer screening. Nature 577 (7788), 35–36. doi:10.1038/d41586-019-03822-8. http://www.nature.com/articles/d41586-019-03822-8

Poplack, S.P., Tosteson, T.D., Kogel, C.A., Nagy, H.M., 2007. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography : American Journal of Roentgenology : Vol. 189, No. 3 (AJR). Am. J. Roentgenol. 189 (3), 616–623. https://www-ajronline-org.online.uchc.edu/doi/full/10.2214/AJR.07.2231

RadiologyA.C., 2018. ACR Practice Parameter for the Performance of Screening and Diagnostic Mammography. ACR Practice Parameters. https://www.acr.org/-/media/ACR/Files/Practice-Parameters/Screen-Diag-Mammo.pdf

Rahmat, T., Ismail, A., Aliman, S., 2019. Chest X-ray Image Classification Using Faster R-CNN, 12.

Ranzato, M., Huang, F.J., Boureau, Y.-L., LeCun, Y., 2007. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1–8. doi:10.1109/CVPR.2007.383157. Event-place: Minneapolis, MN, USA. http://ieeexplore.ieee.org/document/4270182/

Razzak, M. I., Naz, S., Zaib, A., 2017. Deep Learning for Medical Image Processing: Overview, Challenges and Future. arXiv e-prints, arXiv:1704.06825

Razzak, M.I., Naz, S., Zaib, A., 2018. Deep learning for medical image processing: overview, challenges and the future. In: Dey, N., Ashour, A.S., Borra, S. (Eds.), Classification in BioApps: Automation of Decision Making. Springer International Publishing, Cham, pp. 323–350. doi:10.1007/978-3-319-65981-7_12.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, pp. 779–788. doi:10.1109/CVPR.2016.91.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497[cs]

Ribli, D., Horváth, A., Unger, Z., Pollner, P., Csabai, I., 2018. Detecting and classifying lesions in mammograms with deep learning. Sci. Rep. 8 (1), 1–7. doi:10.1038/s41598-018-22437-z. Number: 1 Publisher: Nature Publishing Group. https://www.nature.com/articles/s41598-018-22437-z

Rodriguez-Ruiz, A., Teuwen, J., Vreemann, S., Bouwman, R.W., Engen, R.E.v., Karssemeijer, N., Mann, R.M., Gubern-Merida, A., Sechopoulos, I., 2018. New reconstruction algorithm for digital breast tomosynthesis: better image quality for humans and computers. Acta Radiol. (Stockholm, Sweden : 1987) 59 (9), 1051. doi:10.1177/0284185117748487. Publisher: SAGE Publications. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6088454/

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597[cs]

Rumelhart, D. E., McClelland, J. L., 1987. Learning Internal Representations by Error Propagation. pp. 318–362.

Sahu, P., Huang, H., Zhao, W., Qin, H., 2019. Using virtual digital breast tomosynthesis for de-noising of low-dose projection images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1647–1651. doi:10.1109/ISBI.2019.8759408.

Samala, R.K., Chan, H.-P., Hadjiiski, L.M., Helvie, M.A., Richter, C., Cha, K., 2018. Evolutionary pruning of transfer learned deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis. Phys. Med. Biol. 63 (9), 095005. doi:10.1088/1361-6560/aabb5b. https://iopscience.iop.org/article/10.1088/1361-6560/aabb5b/meta

Seyyedi, S., Wong, M. J., Ikeda, D. M., Langlotz, C. P., 2020. SCREENet: A Multi-view Deep Convolutional Neural Network for Classification of High-resolution Synthetic Mammographic Screening Scans. arXiv:2009.08563[cs, eess]

Sharpe, R.E., Venkataraman, S., Phillips, J., Dialani, V., Fein-Zachary, V.J., Prakash, S., Slanetz, P.J., Mehta, T.S., 2015. Increased cancer detection rate and variations in the recall rate resulting from implementation of 3D digital breast tomosynthesis into a population-based screening program. Radiology 278 (3), 698–706. doi:10.1148/radiol.2015142036. Publisher: Radiological Society of North America. http://pubs.rsna.org/doi/full/10.1148/radiol.2015142036

Shen, L., Margolies, L.R., Rothstein, J.H., Fluder, E., McBride, R., Sieh, W., 2019. Deep learning to improve breast cancer detection on screening mammography. Sci. Rep. 9. doi:10.1038/s41598-019-48995-4. Publisher: Nature Publishing Group. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6715802/

Shin, H.-C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M., 2018. Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: Simulation and Synthesis in Medical Imaging. Springer, Cham, pp. 1–11. doi:10.1007/978-3-030-00536-8_1. https://link.springer.com/chapter/10.1007/978-3-030-00536-8_1

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556[cs]

Singh, S., Matthews, T. P., Shah, M., Mombourquette, B., Tsue, T., Long, A., Almohsen, R., Pedemonte, S., Su, J., 2020. Adaptation of a deep learning malignancy model from full-field digital mammography to digital breast tomosynthesis. arXiv:2001.08381[cs, eess]

Skaane, P., Bandos, A.I., Niklason, L.T., Sebuødegård, S., Østerås, B.H., Gullien, R., Gur, D., Hofvind, S., 2019. Digital mammography versus digital mammography plus tomosynthesis in breast cancer screening: the Oslo tomosynthesis screening trial. Radiology 291 (1), 23–30. doi:10.1148/radiol.2019182394. Publisher: Radiological Society of North America http://pubs.rsna.org/doi/10.1148/radiol.2019182394

Skaane, P., Sebuødegård, S., Bandos, A.I., Gur, D., Østerås, B.H., Gullien, R., Hofvind, S., 2018. Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo tomosynthesis screening trial. Breast Cancer Res. Treat. 169 (3), 489–496. doi:10.1007/s10549-018-4705-2.

Smith, A.,. Synthesized 2D Mammographic Imaging, 13http://www.lowdose3d.com/images/C-View-White-Paper.pdf.

Society, A. C., 2020. How Common Is Breast Cancer? Breast Cancer Statistics. https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (56), 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html

Su, T., Deng, X., Wang, Z., Yang, J., Chen, J., Zheng, H., Liang, D., Ge, Y., 2020. DIR-DBTnet: Deep iterative reconstruction network for 3D digital breast tomosynthesis imaging. arXiv:2008.08322[physics]

Suryanarayanan, S., Karellas, A., Vedantham, S., Baker, S.P., Glick, S.J., D'Orsi, C.J., Webber, R.L., 2001. Evaluation of linear and nonlinear tomosynthetic reconstruction methods in digital mammography. Acad. Radiol. 8 (3), 219–224. doi:10.1016/S1076-6332(03)80530-5.

Swiecicki, A., Buda, M., Saha, A., Li, N., Ghate, S.V., Walsh, R., Mazurowski, M.A., 2020. Generative adversarial network-based image completion to identify abnormal locations in digital breast tomosynthesis images. In: Medical Imaging 2020: Computer-Aided Diagnosis. International Society for Optics and Photonics, p. 1131428. doi:10.1117/12.2551379. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11314/1131428/Generative-adversarial-network-based-image-completion-to-identify-abnormal-locations/

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Boston, MA, USA, pp. 1–9. doi:10.1109/CVPR.2015.7298594. http://ieeexplore.ieee.org/document/7298594/

Tabár, L., Vitak, B., Chen, T.H.-H., Yen, A.M.-F., Cohen, A., Tot, T., Chiu, S.Y.-H., Chen, S.L.-S., Fann, J.C.-Y., Rosell, J., Fohlin, H., Smith, R.A., Duffy, S.W., 2011. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 260 (3), 658–663. doi:10.1148/radiol.11110469. http://pubs.rsna.org/doi/10.1148/radiol.11110469

Teuwen, J., Moriakov, N., Fedon, C., Caballo, M., Reiser, I., Bakic, P., García, E., Diaz, O., Michielsen, K., Sechopoulos, I., 2020. Deep learning reconstruction of digital breast tomosynthesis images for accurate breast density and patient-specific radiation dose estimation. arXiv:2006.06508[physics]

Thang, N.D., Dung, N.V., Duc, T.V., Nguyen, A., Nguyen, Q.H., Anh, N.T., Cuong, N.N., Linh, L.T., Hanh, B.M., Phu, P.H., Phuong, N.H., 2021. Building a X-ray database for mammography on vietnamese patients and automatic detecting ROI using mask-RCNN. Soft Comput. Biomed. Appl. Relat. Top. 315–329. doi:10.1007/978-3-030-49536-7_27. Publisher: Springer, Cham. https://link.springer.com/chapter/10.1007/978-3-030-49536-7_27

Vedantham, S., Karellas, A., Vijayaraghavan, G.R., Kopans, D.B., 2015. Digital breast tomosynthesis: state of the art. Radiology 277 (3), 663–684. doi:10.1148/radiol.2015141303. Publisher: Radiological Society of North America, http://pubs.rsna.org/doi/10.1148/radiol.2015141303

Wang, L., Zheng, C., Chen, W., He, Q., Li, X., Zhang, S., Qin, G., Chen, W., Wei, J., Xie, P., Zhou, L., Wang, X., Zhen, X., 2020. Multi-path synergic fusion deep neural network framework for breast mass classification using digital breast tomosynthesis. Phys. Med. Biol. 65 (23), 235045. doi:10.1088/1361-6560/abaeb7. Publisher: IOP Publishing

Wang, X., Liang, G., Zhang, Y., Blanton, H., Bessinger, Z., Jacobs, N., 2020. Inconsistent performance of deep learning models on mammogram classification. J. Am. Coll. Radiol. 17 (6), 796–803. doi:10.1016/j.jacr.2020.01.006. http://www.sciencedirect.com/science/article/pii/S1546144020300284

Wu, D., Kim, K., Li, Q., 2020. Digital breast tomosynthesis reconstruction with deep neural network for improved contrast and in-depth resolution. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 656–659. doi:10.1109/ISBI45749.2020.9098661. ISSN: 1945-8452

Wu, H., Jiang, X., Jia, F., 2019. UC-GAN for MR to CT image synthesis. In: Artificial Intelligence in Radiation Therapy. Springer, Cham, pp. 146–153. doi:10.1007/978-3-030-32486-5_18. https://link.springer.com/chapter/10.1007/978-3-030-32486-5_18

Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Xu, Z., Prince, J., 2018. Unpaired brain MR-to-CT synthesis using a structure-constrained CycleGAN. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 174–182. doi:10.1007/978-3-030-00889-5_20. https://link.springer.com/chapter/10.1007/978-3-030-00889-5_20

Yi, X., Walia, E., Babyn, P., 2019. Generative adversarial network in medical imaging: a review. Med. Image Anal. 58, 101552. doi:10.1016/j.media.2019.101552. http://www.sciencedirect.com/science/article/pii/S1361841518308430

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. MIT Press, Montreal, Canada, pp. 3320–3328.

Yousefi, M., Krzyżak, A., Suen, C.Y., 2018. Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning. Comput. Biol. Med. 96, 283–293. doi:10.1016/j.compbiomed.2018.04.004. http://www.sciencedirect.com/science/article/pii/S0010482518300799

Yu, B., Wang, Y., Wang, L., Shen, D., Zhou, L., 2020. Medical image synthesis via deep learning. In: Deep Learning in Medical Image Analysis, pp. 23–44. doi:10.1007/978-3-030-33128-3_2. https://link.springer.com/chapter/10.1007/978-3-030-33128-3_2

Zhang, X., Zhang, Y., Han, E.Y., Jacobs, N., Han, Q., Wang, X., Liu, J., 2018. Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks. IEEE Trans. NanoBioscience 17 (3), 237–242. doi:10.1109/TNB.2018.2845103. Conference Name: IEEE Transactions on NanoBioscience

Zhang, Y., Wang, X., Blanton, H., Liang, G., Xing, X., Jacobs, N., 2019. 2D convolutional neural networks for 3D digital breast tomosynthesis classification. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1013–1017. doi:10.1109/BIBM47256.2019.8983097.

Zhou, W., Lu, J., Zhou, O., Chen, Y., 2015. Evaluation of back projection methods for breast tomosynthesis image reconstruction. J. Digit. Imaging 28 (3),

338–345. doi:10.1007/s10278-014-9736-6. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4441696/

Zhu, F., Liu, Y., Wen, H., Wang, T., Fang, J., Tang, S., 2020. Optimization of back projection with different geometry configurations of X-ray sources in digital breast tomosynthesis: a Monte Carlo simulation study. Optik 202, 163604. doi:10.1016/j.ijleo.2019.163604. http://www.sciencedirect.com/science/article/pii/S0030402619315025

Zlocha, M., Dou, Q., Glocker, B., 2019. Improving RetinaNet for CT Lesion Detection with Dense Masks from Weak RECIST Labels. https://www.groundai.com/project/improving-retinanet-for-ct-lesion-detection-with-dense-masks-from-weak-recist-labels/1.

Zuckerman, S.P., Maidment, A.D.A., Weinstein, S.P., McDonald, E.S., Conant, E.F., 2017. Imaging with synthesized 2D mammography: differences, advantages, and pitfalls compared with digital mammography. Am. J. Roentgenol. 209 (1), 222–229. doi:10.2214/AJR.16.17476. Publisher: American Roentgen Ray Society, http://www.ajronline.org/doi/full/10.2214/AJR.16.17476