

Deep Learning in Breast Cancer Imaging: A Decade of Progress and Future Directions

Luyang Luo^{ID}, Member, IEEE, Xi Wang^{ID}, Yi Lin^{ID}, Graduate Student Member, IEEE, Xiaoqi Ma^{ID}, Andong Tan^{ID}, Ronald Chan^{ID}, Varut Vardhanabhuti^{ID}, Winnie CW Chu^{ID}, Kwang-Ting Cheng^{ID}, Fellow, IEEE, and Hao Chen^{ID}, Senior Member, IEEE

(Methodological Review)

Abstract—Breast cancer has reached the highest incidence rate worldwide among all malignancies since 2020. Breast imaging plays a significant role in early diagnosis and intervention to improve the outcome of breast cancer patients. In the past decade, deep learning has shown remarkable progress in breast cancer imaging analysis, holding great promise in interpreting the rich information and complex context of breast imaging modalities. Considering the rapid improvement in deep learning technology

and the increasing severity of breast cancer, it is critical to summarize past progress and identify future challenges to be addressed. This paper provides an extensive review of deep learning-based breast cancer imaging research, covering studies on mammograms, ultrasound, magnetic resonance imaging, and digital pathology images over the past decade. The major deep learning methods and applications on imaging-based screening, diagnosis, treatment response prediction, and prognosis are elaborated and discussed. Drawn from the findings of this survey, we present a comprehensive discussion of the challenges and potential avenues for future research in deep learning-based breast cancer imaging.

Index Terms—Breast cancer, medical image analysis, deep learning.

I. INTRODUCTION

BREAST cancer has become the malignancy with the highest incidence rate worldwide with estimated 2.3 million new cases in 2020 [1]. Although the mortality rate has steadily decreased since 1989 [2], breast cancer remains the fifth leading cause of cancer mortality globally and the primary cause of cancer mortality in women, with an increasing incidence rate in most of the past four decades and an estimated 685,000 deaths in 2020 [2], [3].

Breast cancer can be categorized into invasive cancer and in situ cancer according to whether it spreads out or not, and invasive cancer is further divided into four stages (i.e., I, II, III, or IV) based on the spreading severity [4]. Recent statistics by the American Cancer Society showed that breast cancer survival varies significantly by stage at diagnosis. The 5-year survival rates of USA patients diagnosed during 2012–2018 were >99% for stage I, 93% for stage II, 75% for stage III, and 29% for stage IV [2]. Early detection and efficient systemic therapies are essential in reducing the mortality rate of breast cancer [1]. Breast imaging, including mammography, ultrasonography, magnetic resonance imaging, and pathology imaging, has played a crucial role in providing both macroscopic and microscopic investigations of breast cancer to guide treatment decisions.

Mammography, first performed in 1913, has been proven able to reduce breast cancer mortality rates after long-term follow-up [5]. Mammography uses low-energy X-rays to examine the breast, often by projecting the tissues into a 2D image. Organized population-based mammography screening for women is

Manuscript received 16 August 2023; revised 23 December 2023; accepted 19 January 2024. Date of publication 24 January 2024; date of current version 30 January 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62202403, in part by the Shenzhen Science and Technology Innovation Committee under Grant SGDX20210823103201011, in part by the Hong Kong Innovation and Technology Fund under Grant PRP/034/22FX, and in part by the Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone under Grant HZQB-KCZYB-2020083. (Corresponding author: Hao Chen.)

Luyang Luo, Yi Lin, Xiaoqi Ma, and Andong Tan are with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR, China (e-mail: cseluyang@ust.hk; yi.lin@connect.ust.hk; xma2325@gmail.com; atanac@connect.ust.hk).

Xi Wang is with the Zhejiang Lab, Hangzhou 311121, China, also with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, SAR 999077, China, and also with the Department of Radiation Oncology, Stanford University School of Medicine, Palo Alto, CA 94305 USA (e-mail: xiwang@cse.cuhk.edu.hk).

Ronald Chan is with the Department of Anatomical and Cellular Pathology, The Chinese University of Hong Kong, Hong Kong (e-mail: ronaldckchan@cuhk.edu.hk).

Varut Vardhanabhuti is with the Department of Diagnostic Radiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong (e-mail: varv@hku.hk).

Winnie CW Chu is with the Department of Imaging and Interventional Radiology, The Chinese University of Hong Kong, Hong Kong (e-mail: winniechu@cuhk.edu.hk).

Kwang-Ting Cheng is with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR, China, and also with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR, China (e-mail: timcheng@ust.hk).

Hao Chen is with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR 999077, China, also with the Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR 999077, China, and also with HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Shenzhen 518045, China (e-mail: jhc@cse.ust.hk).

This article has supplementary downloadable material available at <https://doi.org/10.1109/RBME.2024.3357877>, provided by the authors.

Digital Object Identifier 10.1109/RBME.2024.3357877

recommended by The World Health Organization [6], which can provide early diagnosis and improve the prognosis for potential patients [7]. Apart from screening, mammograms are also used to diagnose abnormalities such as masses, calcifications, architecture distortions, or area asymmetries. Due to the superposition nature of 2D mammograms, different views of a breast could be needed for richer information. Standard mammography views are the craniocaudal (CC) view and the mediolateral oblique (MLO) view for both breasts, which are taken directly from the above and from an angled side of the breast, respectively. Extra views might also be generated depending on practical needs. Digital breast tomosynthesis (DBT), also regarded as 3D mammography, has been introduced to provide more spatial context for detailed examination of the breasts and is emerging as the standard of breast imaging care [8].

Ultrasound imaging (sonography) uses high-frequency sound waves to view inside the body without any ionizing radiation. Since the early attempts in describing the acoustic characteristics of breast tumors [9], ultrasonic imaging has undergone a series of transformations, both in instrument design and in clinical applications. Over the past few decades, the quality of ultrasound images has been largely improved by advances in transducer design, electronics, computers, and signals. Sonography thus has become a major mode of imaging for the diagnosis of breast cancer in clinical practice [10]. Currently, breast ultrasound is widely used to distinguish cysts and solid nodules with a high specificity [11] and classify solid masses as benign or malignant when combined with mammography [12]. It has also shown usefulness in screening and detecting early-stage breast cancers [13], and is recommended for Asian women with dense breasts [14]. Due to its ease of use and real-time imaging capability, breast ultrasound has become popular in guiding breast biopsies and other interventional procedures. B-mode is the most common form of ultrasonic imaging for the breasts. Compound imaging and harmonic imaging are also increasingly applied to visualize breast lesions and reduce image artifacts. Moreover, there is growing interest in applying Colored Doppler ultrasound and contrast agents for measuring tumor blood flow and imaging tumor vascularity [15].

Breast magnetic resonance imaging (MRI) [16] takes advantage of radio waves and magnetic fields to generate more detailed information, which is often a 3D picture of the inside of the breasts. Since the invention of MRI in 1971, multiple clinical assessments have witnessed the versatility and effectiveness of breast MRI. Breast MRI has the highest sensitivity among radiological imaging techniques for breast cancer detection [16], and it is widely used as an auxiliary tool for breast-related lesion diagnosis and prognosis. Nowadays, MRI examinations are becoming the main scanning modalities for monitoring the cycle treatment response and recurrence, offering more details of the breasts without introducing ionizing radiations. Considering that the breast anatomy contains different types of tissues, fat suppression technique [17] has been introduced to suppress the signal from adipose tissue as an auxiliary step. To provide different visible foci, multiple types of sequences could be generated [18], such as T1-weighted, T2-weighted, and Diffusion-weighted MRI. Moreover, Dynamic Contrast Enhanced (DCE)-MRI has become the main clinical and

research sequence, which could provide additional information by observing the T1 changes over multiple phases after injection of the contrast agent [19]. Abbreviated breast MRI which uses single early phase DCE has been introduced as a shortened examination approach for screening breast cancers [20].

Breast pathology provides a microscopic investigation of cancers in an invasive way. In clinical practice, microscopic analysis by pathology imaging is also regarded as the gold standard for the final determination of breast cancer. A sample of the patient's breast tissue would be taken by pathologists and placed onto a microscope slide. Then, specific stains and dyes are used to identify cancer cells and confirm the presence of chemical receptors. The most common stain for breast tissue specimens is the hematoxylin-eosin stain (H&E stain) [21], which has been used for more than a century and is still the standard process for histopathological diagnosis [22]. Moreover, auxiliary techniques are often required to complete the diagnosis, such as immunohistochemistry (IHC) [23] and in situ hybridization (ISH) [24]. In the routine clinical pathology, the predictive and prognostic biomarkers estrogen receptor α (ER), progesterone receptor (PgR), human epidermal growth factor receptor 2 (HER2), and the proliferation-associated nuclear protein Ki67 are analyzed by IHC [25]. HER2 gene amplification can be further verified by ISH analysis [24].

The breast imaging-reporting and data system (BI-RADS) was proposed to categorize the overall assessment of the radiological imaging findings [26]: BI-RADS 0 refers to an incomplete examination; BI-RADS 1 refers to negative findings; BI-RADS 2 refers to benign findings; BI-RADS 3 refers to likely benign findings with <2% chance of malignancy; BI-RADS 4 has three sub-categories, 4a, 4b, and 4c, which refer to suspicious findings with 2%-10%, 10%-50%, and 50%-95% likelihood of malignancy, respectively; BI-RADS 5 refers to suspicious findings with >95% likelihood of malignancy; and BI-RADS 6 refers to pathology-proven malignancy. The radiological findings can only be used as a reference for suspicion of malignancy. Usually, patients with BI-RADS 4 or above would be recommended for a biopsy examination to determine the status of malignancy in a microscopic view.

The description of breast cancer requires interpretation of the complex and rich clinical information provided by breast imaging from the macroscopic level to the microscopic level. With the fast increase in medical data scale and the development of imaging technology, analyzing large-scale high-dimensional breast images with artificial intelligence (AI) holds great promise in improving the accuracy and efficiency of clinical procedures. Current AI is typically represented by deep learning (DL), which has made remarkable achievements over the past decade and has been widely adopted in various fields such as image or speech recognition [27]. Compared with conventional computer-aided diagnosis techniques that rely on hand-engineered features, deep learning models show great efficacy in extracting representations from high-dimensional data (e.g., images), and the performance of deep models is often better with more training data. Thus far, deep learning has also been widely studied for analyzing medical images [28] and demonstrated high performance in various fields [29]. With the convergence of AI and human performance, deep learning

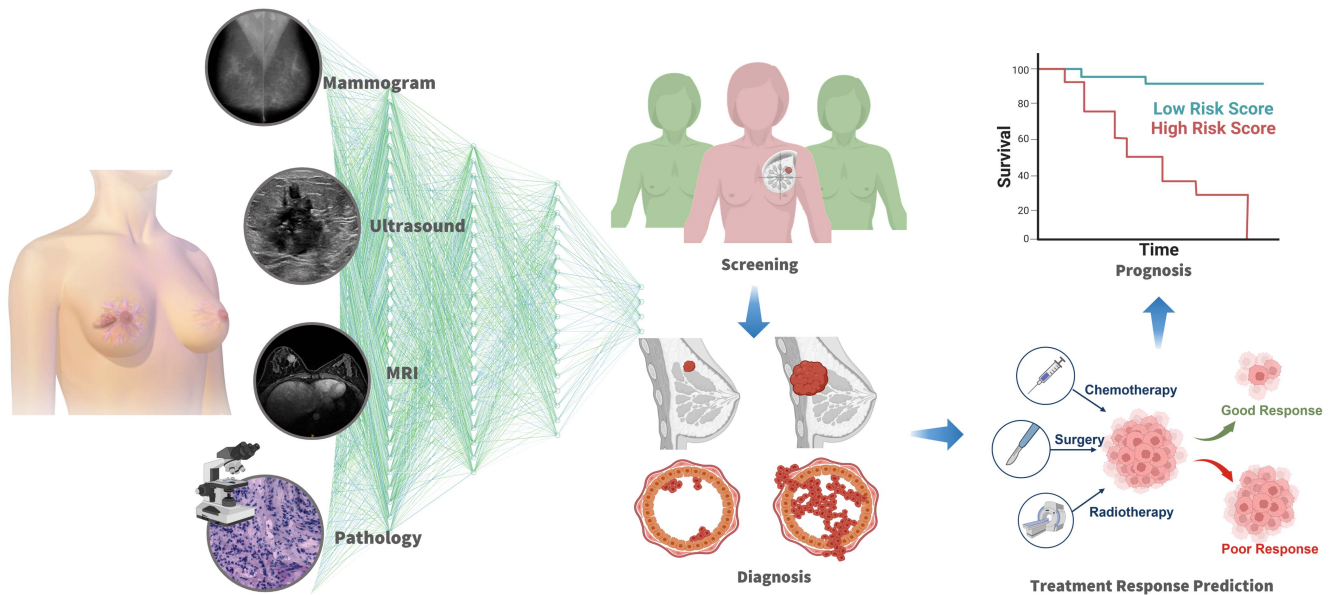


Fig. 1. Overview of deep learning in breast cancer imaging. Typical imaging techniques include mammograms, ultrasound, magnetic resonance imaging (MRI), and pathology images. Deep learning is often used for screening, diagnosis, treatment response prediction, and prognosis.

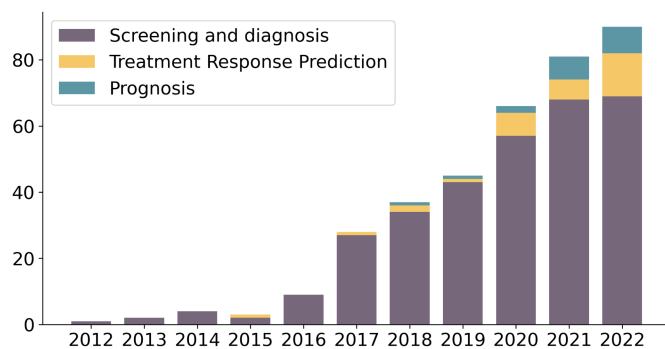


Fig. 2. Number of representative papers on deep learning for breast cancer imaging published from 2012 to 2022.

nowadays is also reshaping cancer research and personalized clinical care.

As shown in Fig. 2, deep learning-based breast imaging has had a prosperous development in the past decade. However, an extensive survey on deep learning-based breast cancer analysis is yet absent to narrate the progress in various imaging modalities over the past decade. Therefore, the main goal of this paper is to review the development of deep learning in breast cancer imaging, identify the challenges yet to be addressed in this field, and highlight potential solutions to these challenges. Specifically, this survey includes applications from screening, diagnosis, and treatment response prediction to prognosis, covering imaging modalities from mammography, ultrasound, and MRI, to pathology images, as shown in Fig. 1. Compared with previous surveys that focus on one or two specific modalities [30], [31], [32], [33], [34], this work provides a more comprehensive summary of the advances in this field. In total 366 papers from 2012 to 2022 were surveyed, covering a wide variety of applications of deep learning in breast cancer imaging.

The remainder of this work is structured as follows: In Section II, we introduce the major deep learning techniques used

in breast cancer image analysis. In Section III, we elaborate in detail on the applications of deep learning in breast cancer image analysis in four aspects: screening, diagnosis, treatment response prediction, and prognosis. In Section IV, we discuss the major challenges facing the field and highlight the future perspectives that hold promise in advancing the field. Finally, we conclude this survey in Section V. We also summarize publicly available datasets and provide a more detailed table of surveyed papers in the Supplementary Materials for interested readers.

II. DEEP LEARNING METHODS FOR BREAST CANCER ANALYSIS

This section will introduce the major deep learning techniques used in breast cancer imaging. For a more detailed review of deep learning, we refer the readers to [35]. We will first introduce the formulations and some majorly used deep learning models by categorizing breast cancer image analysis into three basic tasks, i.e., classification, detection, and segmentation, according to the output types, and a brief illustration of the deep learning models commonly utilized in each task can be found in Fig. 3. We will then introduce the widely applied deep learning paradigms, including supervised learning, semi-supervised learning, weakly-supervised learning, unsupervised learning, transfer learning, and multimodal learning.

A. Classification

Classification aims to give discrete predictions to categorize the whole inputs, e.g., 1 to indicate that a breast image contains cancer and 0 to indicate that the image does not contain cancer. A classification model can be regarded as a mapping function $f: X \rightarrow Y$, where X is the domain of images or features and $Y \in \mathbb{R}$ is usually a one-hot representation of the disease existence. Formally, given x an input, y the target output, and \hat{y} the model output, the classification models are typically optimized

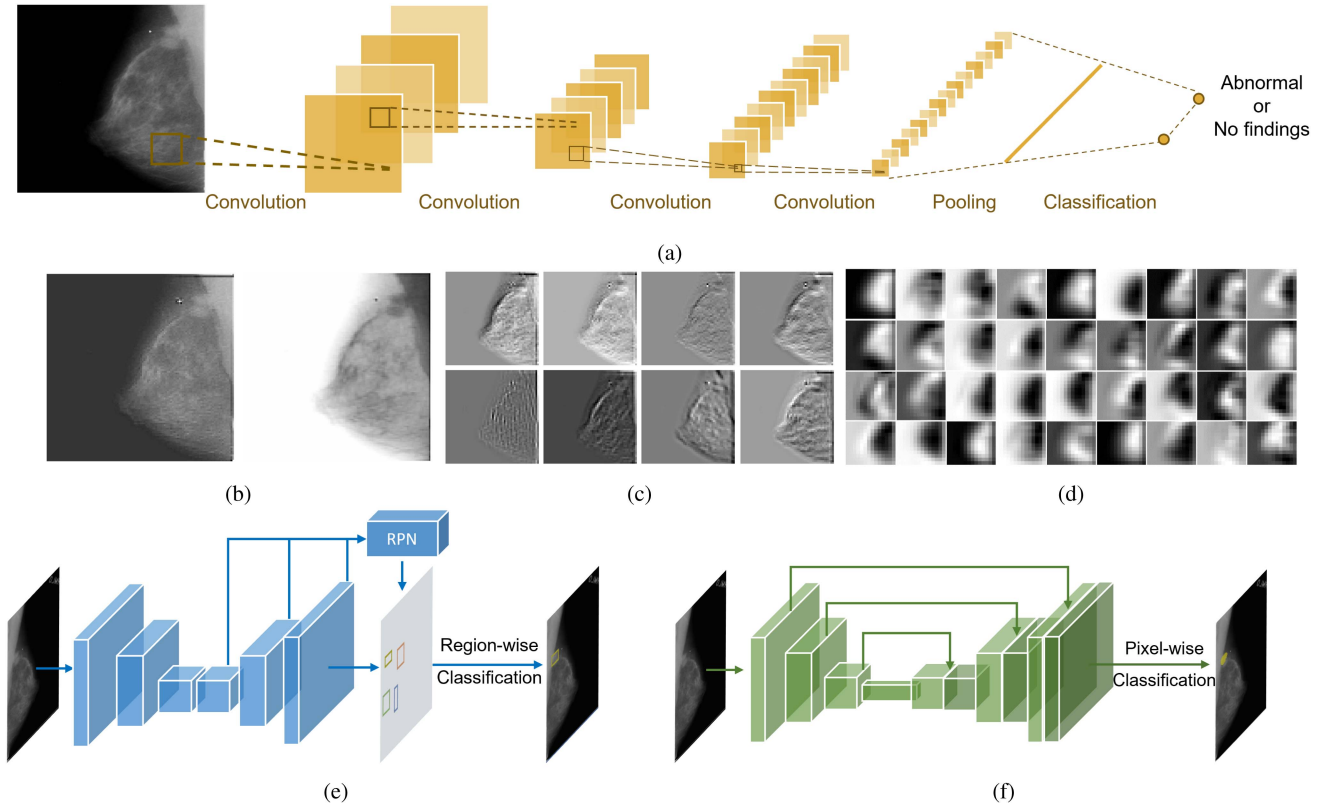


Fig. 3. Brief illustration of deep learning models, taking mammogram as an example. (a) A typical classification network that uses convolutional and pooling to downsample the image while expanding the channels of features. The final feature maps will be pooled into a feature vector, and often a fully-connected layer can be used to conduct the classification based on the feature vector. Typical feature maps extracted by a ResNet-18 pre-trained on ImageNet from layers 1, 7, and 17 are shown in (b)–(d), respectively. (e) A typical detection network. The downsampling workflow often follows the classification network. Then, the feature maps are upsampled, the multi-scale features are fed into a region proposal network (RPN) for region proposal generation, and a region-wise classification is performed to determine the final output. (f) A typical segmentation network. The downsampling workflow could follow the classification network. Then, the feature maps are upsampled several times and concatenated with the shallow-layer features. The final results are obtained based on pixel-wise classification on the largest feature map. All the models are optimized with backpropagation [38].

by minimizing the cross entropy between \hat{y} and y :

$$\mathcal{L} = -y \log \hat{y} \quad (1)$$

To model f , earlier studies would utilize artificial neural networks (ANNs) [36], [37], [38] that are constructed by several fully-connected layers and take as input hand-crafted features. Convolutional neural network (CNN) [39] gets rid of feature engineering and makes the classification problem on images fully end-to-end. In 2012, the success of AlexNet [40], a 5-layer CNN powered by graphic processing unit (GPU), kicked off the era of deep learning with its outstripping performance on the ImageNet challenge [41]. VGG [42] extended the depth of CNNs with smaller kernels and auxiliary losses. Residual networks (ResNet) [43] further deepened CNNs to hundreds of layers and conquered the gradient vanishing problem with skip connections. Apart from AlexNet, VGG, and ResNet, many other networks like Densely Connected Network (DenseNet) [44] and the Inception series [45], [46] have all been widely used in breast cancer imaging. Recently, vision transformer [47], a type of deep neural networks that are mostly based on attention mechanism [48], has also shown great potential in image processing. It is worth mentioning that classification models are often used

as a feature extractor for other tasks which will be introduced in the following sections.

B. Detection

Detection aims to predict region-wise classification results, e.g., drawing a bounding box for a recognized malignancy. Reusing $f : X \rightarrow Y$ as the mapping function of a detection model, X remains the domain of images, while Y is a set of $\{(b, y)\}$ with b the region and y the corresponding class for that region. Note that $\{(b, y)\}$ could be an empty set if there are no regions of interest (ROIs) on the image. The most commonly used formulation of b is a quadruple $\{u, v, w, h\}$, where u and v represent the center of a object box, and w and h represent the weight and height, respectively. The detection objective is often formalized as sibling tasks containing a region-wise classification loss \mathcal{L}_{cls} and a bounding box regression loss \mathcal{L}_{loc} :

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{loc}, \quad (2)$$

where \mathcal{L}_{cls} is commonly formed as a cross entropy loss, λ is a loss balancing hyper-parameter, and \mathcal{L}_{loc} is often formalized as

a smooth L1 loss as follows:

$$\mathcal{L}_{\text{loc}} = \begin{cases} 0.5 \times (t - \hat{t})^2, & \text{if } |t - \hat{t}| < 1; \\ |t - \hat{t}| - 0.5, & \text{otherwise,} \end{cases} \quad (3)$$

where \hat{t} is the model prediction, t is the transformed location representation based on b for regularized regression [49], and $|\cdot|$ represents L1-norm. There are also other choices for \mathcal{L}_{cls} (e.g., focal loss [50]) and \mathcal{L}_{loc} (e.g., the intersection over union loss [51]).

Object detection models often leverage well-trained classification networks as feature extractors. To conduct region-wise prediction on the extracted feature maps, a considerable number of deep learning architectures have been proposed, and some of the frequently used models in breast cancer imaging are covered here. Fast R-CNN [52] extracts and pools the proposals (i.e., the candidate object features) from a pre-trained CNN and conducts the sibling localization and classification tasks. Faster R-CNN [53] introduces the concept of anchors with region proposal network (RPN), which boosts the speed of detection with reference boxes on the feature maps. Apart from the mentioned two-stage methods that extract proposals and then conduct classification and regression, one-stage detectors have been proposed to further accelerate the inference speed, with classical representatives such as the YOLO series [54], [55] and RetinaNet [50].

C. Segmentation

Segmentation aims to give pixel-wise classification predictions, and the contour of objects can then be delineated. Taking the segmentation task on 2D images as an example, the input domain is $X \in \mathbb{R}^{W \times H \times C}$ and the output domain is $Y \in \mathbb{R}^{W \times H \times N}$, where W , H , C , N represents the width, height, channel, and the number of classes, respectively. Typical objectives for medical image segmentation are pixel-wise cross entropy loss and the Dice loss [56]:

$$\mathcal{L} = 1 - \frac{2 * \sum_{i=1}^M y_i \hat{y}_i + \epsilon}{\sum_{i=1}^M y_i^2 + \sum_{i=1}^M \hat{y}_i^2 + \epsilon}, \quad (4)$$

where M is the total number of pixels, y_i is the i -th pixel target, \hat{y}_i is the i -th pixel prediction, and ϵ is a hyper-parameter for numerical stability. The target y here is also called the mask of the image.

Segmentation models for medical imaging typically follow an encoder-decoder structure that first downsamples (e.g., via convolution and pooling) the input image into features and then upsamples (e.g., via deconvolution and interpolation) the features to pixel-wise predictions. FCN [57] first introduced deconvolution to a VGG classifier for image segmentation. U-Net [58] expanded the upsampling process to multiple levels of interpolation and further introduced skip connections between the encoder and decoder paths to enrich details. Later on, U-Net and its variants [59] have almost dominated the field of medical image segmentation. Further, to separate overlapped instances that belong to the same class, Mask-RCNN [60] is often adopted, which first detects object proposals on the images and then performs segmentation for each detected object.

D. Deep Learning Paradigms

There are diverse options of deep learning paradigms to apply the models to different scenarios, given the availability of the data and labels.

Supervised learning requires all training samples to be labeled exactly in the form of targeted outputs, e.g., masks for the segmentation task or bounding boxes for the detection task. Supervised learning is the most common form of deep learning, and a large proportion of studies reviewed in this paper fall into this category. However, deep learning is notoriously data-hungry and labeling medical images is time-consuming, and expertise-depending. Hence, supervised learning may not be the optimal solution for many practical medical image analysis scenarios.

Weakly-supervised learning (WSL) is applied when the given label is not in the format of the targeted output. For example, using image-level annotations for detection or segmentation. In breast cancer imaging, the most used weakly-supervised learning methods are class activation map (CAM) [61] and multiple instance learning (MIL) [62]. CAM is often used for rough detection of targeted lesions, which is computed as the feature maps weighted by corresponding gradients. Higher values on a CAM indicate the regions that contribute more to the final prediction. MIL treats an input image as a bag of instances (i.e., image patches) which is negative only when all instances are negative. The goal of MIL is often to develop a bag-level classifier, which is quite a common strategy in processing whole slide pathology images which are of giga-pixel scale. Like CAM, MIL can also be used to roughly localize the lesions by highlighting the most contributed instances.

Semi-supervised learning (SSL) can be regarded as another type of WSL, which enables utilizing a large amount of unlabeled data together with limited labeled data. Typical SSL methods are based on graph, entropy minimization, pseudo labeling, generative modeling, or consistency learning. Recently, consistency-based approaches have shown great success in SSL, which inject a regularization on the model that the predictions on different perturbed versions of a model should remain consistent.

Unsupervised learning leverages unlabeled data for model training, often aiming at clustering or dimension reduction. In the literature of deep learning-based breast cancer image analysis, two major directions in unsupervised learning have gained research attention: generative modeling and self-supervised learning. The former uses generative methods, such as the generative adversarial network (GAN) [63], to model the data distribution and generate new samples, which is also quite often used in SSL. Self-supervised learning trains a neural network on the unlabeled images to learn representations for the supervised downstream tasks [64]. Self-supervised learning has shown remarkable strength in reducing the requirement of large amounts of downstream labeled data, which holds great promise in medical image analysis.

Transfer learning aims to transfer the knowledge learned on a source domain to the target domain, which is especially useful when a target domain does not possess too much data. A common transfer learning strategy in medical imaging is pre-training the networks on large-scale natural image datasets,

such as the ImageNet [40]. Recently, domain adaptation [65] and domain generalization [66] have also gained huge attention, which mitigates the gaps between the source domain and the target domain.

Multimodal learning aims to process and summarize information from different views/modalities of a subject. Taking the examination process of breast cancer as an example, multimodal data from mammogram, ultrasound, MRI to pathology images could be generated and utilized together. Multimodal data could provide rich and complementary information from macroscopic to microscope. It is worth mentioning that multimodal data could also be yielded with a single type of imaging technique, such as multi-view mammograms or multi-sequence MRIs. Typically in breast imaging, multimodal learning focuses on the information fusing strategies, which mainly include early fusion (i.e., fuse data at the input level), joint fusion (i.e., fuse at the feature level), and late fusion (i.e., fuse at the decision level).

III. DEEP LEARNING APPLICATIONS IN BREAST CANCER

We here provide a concise review of deep learning-based applications in breast cancer imaging. One may refer to the supplementary where we also provide detailed lists of studies surveyed by us.

A. Screening and Diagnosis

Screening aims to find patients out of the examined cohort, and diagnosis aims to give a more precise description of the patients' status. Screening is often based on population-scale mammograms, and diagnosis often uses other imaging modalities. However, there is no clear demarcation between the two tasks in the context of deep learning, as a large number of studies focus on determining malignancies from normal or benign subjects. We hence introduce the deep learning-based breast cancer screening and diagnosis in a combined section.

1) Mammogram-Based Screening and Diagnosis: Classification: As a routine breast cancer screening approach, mammograms are often studied for binary classification (e.g., malignant vs benign/normal/non-malignant) or three-class classification (e.g., malignant vs benign vs normal). Studies also tried to distinguish different types of lesions such as mass or architectural distortion, the levels of breast density, or the levels of cancer risks.

The early studies relied on hand-crafted features as inputs to ANNs [67], [68]. With their remarkable success in analyzing natural images [40], CNNs were also used as a powerful feature extractor combined with other machine learning classifiers like Support Vector Machine (SVM) and Random Forest [69].

As medical data are often limited in scale, some works [70], [71] transferred existing networks with pre-trained weights from the ImageNet to the mammogram datasets. A comparative study on mammogram classification performance of different networks was reported in [72]. A comparison of the efficiency of mammogram-based classification (using 2D VGG) and DBT-based classification models (using 3D VGG) was reported by Li et al. [73]. Apart from directly using on-the-shelf models, studies further sought more effective transfer learning methods

to improve the pre-training learning process and fully utilize the learned knowledge from the pre-training dataset [74]. A comparative study on the pre-training strategies has been reported by Clancy et al. [75]. To enlarge the training data and learn more robust models, data augmentation and model ensemble have been widely used [76]. Several works also attempted to use generative models to enlarge the scale of training data [77].

Multiple instance learning (MIL) has played an important role in mammogram-based classification, as the lesions are often sparse [78]. A line of works focused on developing MIL classifiers with different aggregation strategies to summarize the final feature maps of a CNN model [78], [79]. Moreover, Lotter et al. [80] enabled the training of a RetinaNet detector [50] with both bounding box annotations and image-level supervision using the MIL strategy. For 3D DBT classification, some works utilized a 2D classifier to obtain results for each slice and fused the results with MIL for final 3D prediction [81].

A large proportion of studies proposed learning patient-level prediction from multiple input images. On the one hand, multiple views (e.g., the bilateral craniocaudal (CC), mediolateral oblique (MLO) views, etc.) are often generated for more detailed examination. Different multi-view feature fusion methods have thus been proposed [82], [83], mostly focusing on developing feature extractors to generate more semantically meaningful representations or fusing strategies to inference final results based on the heterogeneous features. On the other hand, studies also attempted to combine the information of the previously screened image (i.e., prior mammogram) and the currently examined image (i.e., current mammogram) for malignancy classification [84]. In particular, Baccouche et al. [85] proposed using GAN to generate a prior image from current mammograms and combining the two images for lesion detection.

Multi-task learning has also been studied to enhance classification performance with extra supervision from other tasks [86]. Notably, multi-task learning and multi-view inputs can both enrich the information provided to the model, and these two strategies can be simultaneously incorporated [87].

Detection: Lesion detection could specify the location and type of the tumors and further quantify cancer development.

Two-stage cancer detection pipeline was widely used, where hand-crafted [88] or network-segmented [89] lesion candidates are fed into a classification network for false positive reduction. Studies also adopted and modified off-the-shelf end-to-end detectors, such as Faster-RCNN [52] and YOLO [54], which take as input the whole mammography image and output bounding box coordinates for lesions with scores indicating the likelihoods of different lesion types [90]. Li et al. [91] also proposed a cancer detection method for 3D DBT by aggregating the 2D predictions into 3D results.

Multi-view methods are also frequently studied for breast lesion detection. Based on the same feature extractor backbone for each view, Liu et al. [92] compared and fused the features from different views to improve the detection performance on top of Mask-RCNN [60]. Graph-based reasoning was also integrated into the multi-view detection framework with graph nodes generated by pseudo landmarks [93]. Recently, transformer-based detector was also used for multi-view mass detection [94]. On

the other hand, Yang et al. [95] proposed to fuse the features of different networks, where each network is designed for a specific view.

It is also noteworthy that the feature maps extracted from a classification network can also be used to localize the lesions by CAM or the attention mechanism in a weakly supervised manner. However, quantitative evaluations are often lacking for these studies, and the attention maps are often used for qualitative interpretation purpose [78], [79]. A recent study also proposed utilizing intermediate features and CAM as pseudo labels to train a detection model [96].

Segmentation: Segmentation provides contour delineation for a more detailed description of the lesion, which often requires a considerable amount of pixel-wise annotations by experienced radiologists for model training.

There are a few studies on mass segmentation from mammograms, based on modified structures or losses with FCN [97], conditional GAN [98], [99], or U-Net-alike structures [100]. A comparative study on different segmentation models was conducted in [101].

Multiple Tasks: Some studies tried to integrate modules for different tasks to establish multi-functional computer-aided diagnosis (CAD) frameworks.

A straightforward way to a multi-task CAD system is training a set of task-specific models [102], where the studies are often based on existing solutions for each task. Object detection networks, such as Faster R-CNN, Mask R-CNN and YOLO, have also been adopted to conduct both classification and detection due to their multi-task learning nature [85], [103]. As mentioned, classification models could be used to extract detection results in a weakly-supervised manner as well [104], [105]. Moreover, unified multi-task models have been proposed in the literature, e.g., combining classification with detection [106] or segmentation [107].

Other Tasks: Many studies aim for more than developing specific methods for the aforementioned basic tasks.

A comparative study of mammogram classification performance between a deep learning model and 101 radiologists has been reported by [108], where the model showed comparable performance to the average of radiologists. Later on, a handful of studies have shown that deep learning models could help improve the radiologists' performance [109], [110], [111]. Recent study also reported that deep learning surpassed the performance of traditional risk prediction algorithms [112]. Moreover, deep models have been demonstrated capable of screening mammograms based on large or even international populations [113], [114] and finding out high-risk subjects for further examination by radiologists [115].

Despite achieving expert-comparable or expert-surpassed accuracy, recent studies reported significant performance drop when applying deep learning models to external testing mammograms [116], [117]. An increasing number of works attempted to improve the robustness of deep learning models for mammograms in the aspects of noisy label [118], adversarial attacks [119], external domain generalization [120], and privacy preserving under federated learning [121]. Efforts have also been devoted to make the models more interpretable [122].

To reduce potential side effects caused by extra imaging process, recent studies also made efforts on mammograms synthesis [123].

2) Ultrasound-Based Diagnosis: Classification: Breast ultrasound (US) images are often used for malignant lesion identification, lymph node metastasis estimation, or breast risk prediction, as an appealing non-invasive alternative to traditional invasive approaches. Similar to the studies on mammograms, well-validated classification networks like VGG, ResNet, DenseNet, etc., have also been widely adopted in breast ultrasound classification. Some studies cropped suspicious lesions in US images as the regions of interest (ROIs), which were later fed into the CNNs for malignancy classification [124], [125] and metastasis estimation [126], [127], [128]. However, a prominent drawback of ROI-based analysis is that lesion regions should be manually cropped in advance, which not only increases the annotation burden of the experts but also impedes the flexibility of these methods in real-world applications. Instead of using manual crops, Lee et al. [129] first leveraged a Mask R-CNN model to detect and segment lesions and then used a DenseNet121 model for auxiliary lymph node metastasis status prediction based on the extracted peritumoral tissues. On the other hand, researchers also attempted to analyze the whole US image without lesion candidate detection in the first place [130], [131], [132]. It is worth noting that with transfer learning and data augmentation techniques, the whole US image-based works can achieve comparable performance to the studies based on carefully extracted ROIs.

In clinical routine practice, radiologists normally make diagnostic decisions based on a comprehensive evaluation of US images in different views and a combination of different modalities such as B-mode, color Doppler, and elastography images. Therefore, studies have also attempted to fuse the complementary information from multi-view or multimodal US images, where feature-level fusion [131], [133] and output-level (predicted probability for each view) fusion [134] have both been explored.

To alleviate the overfitting problem resulting from data deficiency in training, researchers applied various strategies, such as using ImageNet pre-trained models, data augmentation [135], or transfer learning from other related tasks [136]. Moreover, GANs have also been used for generating synthetic data as a surrogate of data augmentation, which demonstrated excellent performance in the task of classification [137].

To efficiently leverage limited annotated data and a large amount of unlabeled data for training, unsupervised learning [138] and semi-supervised learning techniques [139] were also explored to enhance the discrimination capability of the model, based on techniques like autoencoder, Mean Teacher [140], and Virtual Adversarial Training (VAT) [141].

Detection: Detection of lesions is also clinically preferable for breast US diagnosis. Many studies used CNNs for candidate classification and false positive reduction after a lesion candidate extraction process [142], [143], [144]. A comparative study on a variety of state-of-the-art detection networks such as Faster R-CNN and YOLO, and classification networks such as AlexNet and DenseNet were systematically reported in [145]. To further

incorporate the large amount of data annotated at the image level, Shin et al. [146] proposed a joint weakly- and semi-supervised network based on the multi-instance learning scheme.

Segmentation: Segmentation of important ROIs from US images, such as the tumor region, major functional tissues, and breast anatomic layers, aims to provide more fine-grained and quantitative information to clinicians. Automated object segmentation from ultrasound images is quite challenging owing to speckle artifacts, low contrast, shadows, blurry boundaries, and the variance in object shapes. Recently, deep learning-based approaches, particularly FCN and U-Net, have been successfully applied to this field [147], [148].

To enhance the confidence of the hardly-predicted boundary, a number of strategies were designed in previous works [149], [150], [151]. For instance, Xue et al. [150] developed a breast lesion boundary detection module in shallow CNN layers to embed additional boundary maps of breast lesions for obtaining the segmentation result with high-quality boundaries. Other boundary-aware modules were also proposed to achieve more precise segmentation in the confusing and ambiguous boundary areas [149], [151].

Different variants of attention mechanisms were used in conjunction with the deep learning model for US segmentation, such as paying spatial attention to most specific regions in the US images and weighting the feature channels that have different semantic information [152]. In addition, saliency maps that highlight visually salient regions or objects were also explored to strengthen the network's attention to the region of interest in US images and help to boost the segmentation performance of the model [149].

Similar to the classification studies, GANs were widely applied to synthesize training data to augment the training set for improving the segmentation results [147], [153].

Multiple Tasks: A complete CAD system often requires multiple functions, such as lesion detection and classification [134], [145], [146] or lesion segmentation and classification [154], either in a manner of sequential modeling or multi-task learning. Some accomplished such goal in two steps, i.e., first detecting or segmenting the lesion regions, and then classifying the detected or segmented areas into benign or malignant classes [129], [134], [146]. By contrast, there were also methods proposed to conduct multiple tasks simultaneously in an integrated framework [155]. For instance, Zhou et al. [155] proposed a multi-task learning framework for joint segmentation and classification of tumors in 3D automated breast ultrasound images. It is composed of two sub-networks: an encoder-decoder network for segmentation and a lightweight multi-scale network for classification.

Weakly-supervised segmentation has also been explored in breast US [156], where CAM-based methods were often applied in a post-hoc manner to highlight the most discriminative regions discovered by the model.

Other Tasks: Some studies explored novel modalities other than conventional US images, such as the contrast-enhanced ultrasound videos [157] and automated whole breast ultrasound images [143], [144], [155], [158], [159], for lesion classification, detection, and segmentation, where 3D CNNs were often built

to leverage the rich temporal or spatial information for more robust learning [143], [157].

Observer study by radiologists is quite important to validate the potential of DL methods in real-world applications. There are also many works demonstrating that the deep learning approaches have achieved expert-level performance, showing promise in commercial applications [133].

To improve model robustness against noise, Cao et al. [160] built a noise filter network to prevent classification models from overfitting the noisy labels. Zou et al. [161] developed an end-to-end noisy annotation tolerance network for robust US image segmentation.

3) MRI-Based Diagnosis: As the most sensitive radiological modality for breast cancer detection, MRI examinations are applied to fine-grained diagnosis of breast cancer and provide more detailed preoperative guidance for treatment planning. We here generally categorize previous studies into classification, segmentation, detection, and other tasks.

Classification: Classification of breast MRI could be categorized from three perspectives according to their purpose: First, the basic screening task by detecting the absence of lesions which helps physicians to dismiss the normal MRI examinations. Second, a binary classification distinguishing malignant tumors from benign ones for follow-up treatment regimens occupies a large part of research. Third, fine-grained classification like predicting molecular subtypes, BI-RADS, and metastasis also facilitates detailed and further diagnosis. Currently, a large proportion of studies are based on DCE-MRI which takes several MR images at different time points after injection of contrast agent. Ultrafast MRI, which takes a short scan duration within several seconds, has also gained large research interest as patients with motion sensitivity and emergency situations could get accurate radiology examinations from ultrafast MRI promptly and get timely treatments. Moreover, learning from multi-parametric MRI enables the fusion of more comprehensive information from multiple modalities, which has also triggered exploration and improvements.

Different deep learning architectures have been proposed regarding the diverse MRI sequences. Compared with other sequences, DCE-MRI further possesses temporal information provided by the contrast agents. Long short-term memory (LSTM)-based [162] or convolutional LSTM-based [163] networks were applied to exploit the temporal features involved in DCE MRI sequences. To lower the computation cost of 3D MRI, maximum intensity projection (MIP), which projects the voxels throughout the volume onto a 2D image, has also been applied in the DCE series [164], [165], [166]. For instance, from a comparison study based on MIP and central slides of different sequences, Antropova et al. [164] witnessed the effectiveness and advantages of time and space-saving of MIP techniques.

Learning from the multi-parametric MRI combining multiple sequences and integrating their corresponding advantages is another research focus. For instance, Hu et al. [167] implemented three different fusion strategies: image fusion of DCE MIP and center slice of T2-weighted modality, feature fusion of CNN-based deep features, or classifiers output fusion for two training branches of modalities. It's worth noting that feature

fusion was found to perform significantly better in classifying between benign and malignant lesions. Ren et al. [168] compared five models based on different MRI modalities on the task of axillary lymph node metastasis prediction, where the combination of DCE + T2 inputs was found to perform best. Some studies also combined deep features with hand-crafted radiomics features [169] to improve the classification performance.

Apart from network architecture design, other techniques have also been embedded in MRI-based CAD systems. Inspired by transfer learning, using pre-trained networks [166], [170] could solve the data shortage to some extent and speed up the convergence. Also, ensemble learning frameworks [171], [172] were introduced to reduce the model uncertainty. Rasti et al. [172] designed multiple gating networks sharing the same input and fused the outputs at last. Sun et al. [171] predicted the molecular subtypes (luminal and non-luminal) based on the ensemble outputs from three sub-models trained with different post-contrast sequences.

Weakly-supervised and unsupervised learning have been studied to tackle the scarcity of labels. Liu et al. [173] classified the benign and malignant tumors from the whole slides instead of the targeted region of interest with ResNet-based networks. Sun et al. [174] utilized transfer learning strategies from the source domain with unsupervised pre-training on DCE-MRI containing benign and malignant cohorts for predicting molecular subtypes.

Detection: Compared with other modalities, studies on breast MRI-based detection is of a relatively smaller scale. Dalmics et al. [175] took the lesion candidate patch and its contralateral patch as inputs for lesion classification. Maicas et al. [176] incorporated reinforcement learning-based Deep Q network [177] with an attention mechanism for breast lesion detection, which showed accurate localization while saving more inference time. Ayatollahi et al. [178] modified a 3D RetinaNet for small breast lesion detection on ultrafast DCE-MRI.

Segmentation: U-Net-based architectures [179], [180], [181] are the most commonly used structures for breast tumor segmentation. In addition, considering the temporal information and physiological inheritance involved in the DCE-MRI, the Three Time Points (3TP) approach was introduced to help quantize the intensity change of the breasts before, during, and after injection of the contrast agent. For example, Vidal et al. [180] fused the outputs from independent U-Net branches segmenting different series combinations such as 3TP series and full series. Galli et al. [181] extracted the 3TP slices after breast masking and motion correction for lesion segmentation. Moreover, some work [182] conducted multi-stage coarse-to-fine segmentation.

Multiple Tasks: For a more complete CAD system, lots of works implement the combination of tasks simultaneously. Studies [183], [184] implemented the segmentation and classification tasks in one pipeline with sequential order for more diagnosis analysis on the segmented lesions. Zhu et al. [183] utilized VNet and Attention U-Net to segment lesions from DCE and DWI, and ResNet was then used for benign and malignant classification based on the segmentation outputs. Parekh et al. [184] utilized stacked sparse autoencoder networks for segmentation of intrinsic tissue signatures, which was followed by an SVM classifier for classification.

Some studies combined classification and weakly-supervised detection [185], [186] to achieve a comprehensive diagnosis. Zhou et al. [185] implemented the classification task on 3D densely connected networks and localized the lesions with Class Activation Map and conditional Random Dense Conditional Random Field. Luo et al. [186] proposed Cosine Margin Sigmoid Loss for learning cancer malignancy classification and Correlation Attention On-the-shelf models, such as Faster R-CNN, were also used for ROI localization first and combined with custom CNNs for lesion classification [187].

Other Tasks: Breast density estimation by 3D CNN-based regression [188] showed benefit in helping breast cancer risk prediction. Predicting the biomarkers, such as Ki-67 status, can help indicate the development of breast cancer. For example, Liu et al. [189] fused the deep features based on transfer learning CNN frameworks from multi-parametric MRI and predicted the Ki-67 status with a multilayer perceptron classifier. Moreover, to increase the accessibility of breast MRI, Chung et al. [190] generated simulated multi-parametric MRI based on 3D fully convolutional networks and validated its quality by comparing it with real scans.

4) Digital Pathology Images-Based Diagnosis: Pathology image-based diagnosis plays an irreplaceable role as the “gold standard” in cancer characterization. Deep learning-based breast pathology diagnosis has blackflourished in the past decade along with an increasing number of publicly available datasets. Here, we categorize the studies into classification, segmentation, detection, and other tasks.

Classification: The cancer type/grade assignment is an essential task in breast pathology image analysis. Conventional methods for breast pathology classification are based on hand-crafted features qualitatively designed by domain experts, such as the spatial distribution, arrangement, and individual types of discrete tissue elements or primitive shapes (e.g., nuclei, lymphocytes, or glandular structures). In the past few years, deep learning methods have been widely used to automate this process for a higher accuracy [191]. In parallel with breakthroughs in deep learning algorithms, the availability of large-scale datasets has also been a critical factor for the success of deep learning methods. One milestone is the CAMELYON16 dataset [192], which contains 400 histopathology whole slide images (WSIs) of lymph node sections. Wang et al. [193], the winner of the CAMELYON16 challenge, adopted various DL models for the patch-wise classification task. The patch-wise classification results were further aggregated to obtain the geometrical and morphological features of the whole slide image (WSI). Then the embedded features were used to classify the WSIs into metastasis or negative findings via a Random Forest classifier. Following the success, many studies have been conducted to improve the performance of patch-wise classification, including the modification of model architectures [194], [195], the use of different data preprocessing techniques [196], and attention mechanism [197]. A more recent challenge summarized multiple solutions for quantitative tumor cellularity assessment in breast cancer histology images following neoadjuvant treatment [198].

However, a major drawback of these studies is that the patch-wise classification methods usually require extensive manual annotation at the pixel/patch level by expert pathologists, which is time-consuming and labor-intensive for WSIs that are of gigapixel scales. For example, the CAMELYON16 dataset consists of 270 WSIs at $40\times$ magnification, with roughly the same number of pixels as the entire ImageNet dataset [41], which was recognized as one of the largest datasets in the field of computer vision. How to leverage coarsely annotated data (e.g., globally labeled WSIs) to effectively train a well-performed model is much more preferable as WSI-wise analysis is often the ultimate goal of pathologists.

Thus, recent studies have been conducted to develop methods that only require the WSI-level annotations, which are much more convenient and cost-effective for real-world clinical pathological practice. For example, Campanella et al. [199] proposed a deep learning method that only requires WSI-level annotations, which formulates the WSI classification task as an example of the multiple instance learning (MIL) problem. Typically, the MIL-based deep learning method consists of two stages: in the first stage, a deep neural network is used to extract the features of the instances; in the second stage, the instance-level features are aggregated to obtain the bag-level features that are fed into a bag-level classifier to yield the final prediction for bags. The follow-up studies tried to improve the MIL-based WSI classification method by various strategies, including enhancing the instance representation [200], [201], extracting more discriminative features [199], [202], [203], and improving the aggregation strategy [197], [204], [205]. How to improve the information connection between the feature extractor and the aggregator to obtain more discriminative whole-slide level representation remains an open problem, not only for breast pathology image analysis but also for all other DL studies based on WSIs. Recent studies have made efforts with techniques such as coupled iterative training [206] and end-to-end training with huge hardware (GPU) support [207].

Detection: Mitosis detection is a representative research field in pathology image analysis. The density of mitosis is used to assess the cell proliferation activity, which is a key factor for the prognosis of breast cancer [208]. Recent DL methods on automatic mitosis detection can be divided into three categories: 1) object detection-based methods; 2) two-stage methods; and 3) pixel-wise segmentation methods. The object detection-based methods [209] mainly employed popular detection frameworks such as Mask R-CNN. The localization and classification of mitosis were performed simultaneously. The second line of methods consists of two stages. In the first stage, the mitosis candidates were detected by the object detection methods, including Mask R-CNN [210], Mask R-CNN [211], RetinaNet [212], etc. In the second stage, the mitosis candidates were classified into mitosis or non-mitosis [210], [211], [213]. The follow-up methods further improved the detection performance in various aspects, including more representative features [210], network architectures [211], [213], and training strategies [214], etc. The third type of method achieve the mitosis detection task via fine-grained segmentation, which will be introduced in the following section.

Segmentation: According to the Nottingham Grading System [215], [216], the grading of breast cancer is based on the assessment of three morphological features: 1) degree of tubule or gland formation, 2) mitotic count, and 3) nuclear pleomorphism. Thus, segmentation of glands and nuclei is a fundamental yet crucial task in breast pathology image analysis. For gland segmentation, existing DL-based methods typically adopt U-Net and its variations [217]. One of the most popular research foci is to explore the boundary information to boost the segmentation performance [218].

In parallel with the progression of gland segmentation in breast pathology, segmentation of nuclei has also been well studied, which is used to extract the morphological features of the nuclei, such as size, shape, and texture. The morphological features of the nuclei assess the nuclear pleomorphism, which can be used to predict the diagnosis and prognosis of breast cancer [219]. In DL-based nuclei segmentation methods, the main challenge is to obtain accurate segmentation results for nuclei with complex shapes and overlapping. To address this challenge, studies proposed to use multi-task learning [220], [221], multi-scale learning [222], and adversarial learning [223]. Further, the other methods exploited information of the nuclear contour within the training stage. The most straightforward way is to simultaneously predict the contour and the segmentation mask [224]. In this manner, the instance segmentation results can be obtained by the post-processing of the contour and the segmentation mask. The follow-up methods boosted the performance in the aspects of pre-training [225], data augmentation [226], network architectures [227], and loss functions [228], [229].

Another challenge of nuclei segmentation is the scarce manual annotation. One WSI can contain tens of thousands of nuclei, which makes the manual annotation of the nuclei segmentation infeasible. To address this challenge, some studies aimed at developing methods that only require weak annotations, such as scribbles [230] or even point annotations [231], [232]. Based on the size and shape assumptions of the nuclei, existing weakly-supervised methods typically encode the morphological priors into the weak annotations, transforming the weak annotations into the coarse pixel-wise annotations, such as pseudo edge maps [233]. Further, the following studies proposed various techniques to eliminate the bias of the inaccurate and incomplete coarse annotations, such as self-training [231], co-training [232], and multi-task learning [233].

Other Tasks: Due to the significant variance of the staining and the imaging conditions (e.g., slide preparation and microscope scanning), the DL-based breast pathology image analysis methods could suffer from the domain shift problem. Domain shift refers to the data heterogeneity between the source and the target domains [234]. Existing methods for mitigating the domain shift problem of pathology images can be divided into three categories: 1) data augmentation, 2) domain adaptation, and 3) domain generalization. Data augmentation is a common technique to enhance the robustness of the DL models by increasing the diversity of the training data. In breast pathology analysis, the commonly used data augmentation methods is color distortion on both the RGB channels [235] and the HSI (hematoxylin,

eosin, and residual) channels [214]. For domain adaptation, the most representative method is adversarial learning. For example, some [223], [236] proposed to use the generative adversarial networks to map the images from the source domain to the target domain. The domain generalization methods aim at learning domain-invariant features, which can be used to generalize the models to unseen domains. The typical methods for domain generalization in breast pathology are feature alignment [237] and domain-invariant feature learning [238].

Recently, as one type of domain adaption, virtual staining techniques have attracted large attention in breast pathology analysis. For instance, immunohistochemical (IHC) staining reflects protein expression, which is vital for diagnosing cancers, histological classification, grading, staging, and prognosis of tumors. However, the IHC staining procedure is costly, laborious, and time-consuming. To complete the diagnosis, virtual HER2 IHC staining methods [239] were proposed to transform autofluorescence microscopic images of breast tissue sections into bright-field equivalent microscopic images, matching the HER2 IHC staining that is chemically performed on the same tissue sections.

In addition, observer studies by pathologists are also a crucial research issue in clinical practice. Several studies have shown the significant effectiveness of deep learning models compared to that of doctors and pathologists in various applications, such as lymph node metastases detection [192] and pathology-based diagnosis [240].

B. Treatment Response

Different regimens targeting at breast cancer have been well proposed for patients with pertinence based on the specific subtype of tumor, anatomic cancer stage, personal preferences and toxicity risk etc. [241]. Assessing treatment response is of significance for monitoring the progression of cancer and therapeutic effects, which could help implement further clinical decisions and improve patients' outcomes with personalized treatment plans.

Most of the studies focus on Neoadjuvant Chemotherapy (NAC) [242] response prediction. Neoadjuvant treatment, or preoperative treatment, has become a safe and often effective therapeutic choice for larger primary and locally advanced breast cancer [243], and NAC is one of the most mainstream chemotherapies presently [242]. Apart from the imaging modalities we introduced before, Computed Tomography (CT) [244], [245] is also studied for treatment response prediction.

Classification: A qualitative metric to assess NAC is whether the patients achieve pathological complete response (pCR) or not [246], which is demonstrated as an indication of a high disease-free survival rate based on the absence of cancer cells combined with the involvement of lymph nodes after treatment course [247], [248]. Therefore, a binary classification problem is formulated to identify the pCR and non-pCR of NAC treatment, which could help physicians determine further therapeutic plans.

Directly applying CNNs models [244], [249], [250], [251] for classification of NAC response is one of the most common and straightforward strategies. Moreover, replacing the last fully

connected layers of CNN models with other robust conventional classifiers such as Random Forest or Support Vector Machine [252], [253], [254] also attracts a lot of attention. The CNN modules are considered feature extractors providing image representations, which are often combined with other hand-crafted features or clinical information for training conventional classifiers at last. Non-imaging data sources such as pathological records as supplementary to imaging representations could provide more comprehensive information for better performance. Modified and improved network architectures could also achieve NAC response performance. Taleghamar et al. [255] concatenated features from two branches consisting of modified ResNet and modified residual attention network and output classification with a fully connected network. Qi et al. [245] proposed a modified 3D MultiResUnet with Gradient-weighted class Activation Map (Grad-CAM) which could mark the interesting regions during the training process, and show its improvements compared with conventional radiomics analysis.

The multimodal data has also sparked an in-depth exploration of the research on combining with multimodal learning. Joo et al. [256] took multi-parametric MRI as inputs such as T1 weighted, T2 weighted, and clinical information in parallel and concatenated features at last. In addition, clinical information, molecular information, kinetic information, etc., could also be fused into DL frameworks based on multi-stage fusion strategies [253], [254], [257], [258]. Notably, some studies [253], [258] fused the molecular types information which has been known as having a correlation with NAC response results with imaging features based on CNN models. Also, the handcrafted features extracted based on pathological knowledge with conventional algorithms such as histograms [245], [254] were also incorporated and fused into the image-based deep learning frameworks.

Unlike the diagnostic tasks which mostly take one-phase examinations, a large part of source data for predicting treatment utilizes multiple-phase scans across cycles of treatments. Researchers attempt to highlight such temporal information and differences before and after treatment. Methods of independently inputting images of different stages gained many explorations [259], [260]. For example, Xie et al. [259] proposed a framework of dual-branch CNN-based models with inputs of images extracting before and after NAC treatments, and then implemented feature fusions from convolutional blocks. Similarly, Tong et al. [260] proposed dual-branch transformers for NAC response prediction on US images. Siamese architectures-based CNNs [261], [262] were also explored to capture the differences between images before and after treatment cycles. In addition, Recurrent Neural Networks (RNN) were used for capturing the information across temporal dimensions and achieved better prediction outcomes on long-term sequential treatment cycles [263]. For multiple cycles of NAC treatments, Gu et al. [264] constructed a deep learning pipeline for step-wise prediction of different stages of treatments. A comparison study of breast DCE-MRI contrast time points for predicting NAC response has been reported by Huynh et al. [265].

Transfer learning decreases the need for a large amount of new data and speeds up the training process on similar tasks.

Some [253] utilized pre-trained networks as feature extractors for subsequent classification. Byra et al. [262] fine-tuned the Inception-ResNet-V2 CNN based on the benign and malignant classification tasks and took Siamese architecture for receiving images before and after NAC. Multi-task learning simultaneously implementing [261], [266] segmentation and treatment response prediction also attracts great attention. Liu et al. [261] proposed a Siamese multi-task network (SMTN) consisting of segmentation sub-networks and pCR predication sub-networks. Wu et al. [266] constructed three image signatures based on features extracted from segmentation networks for tumor segmentation of three treatment phases, and then the image signatures were integrated with clinical factors for pCR prediction.

Other Tasks: Identifying and classifying the biomarkers or indicators that represent the effect of treatment also play important roles in clinical practice. Aghaei et al. [257] computed kinetic image features and implemented classification of response to chemotherapy based on different fusion combinations of the features with ANN models, and then selected clinical markers according to feature analysis. For more specific assessment, some [267] developed Stroma-derived bio-marker and then obtained new clinical markers based on CNNs from histological images.

C. Prognosis

Prognosis aims to evaluate the likely outcome or course of a disease. We categorize the applications in this field into classification and other tasks.

1) Classification: *Survival prediction:* Apart from methods using a single modality, Liu et al. [268] additionally combined multiple modalities including clinical information (such as sex, age and cancer stage), multi-scale WSIs, as well as the tumor, lymphocyte, and nuclear segmentation results for survival prediction via a MobileNetv2 [269]. Wang et al. [270] combined WSI and genomic features for survival prediction using a bi-linear neural network.

Recurrence prediction: Liu et al. [271] used a 3D CNN to predict breast cancer recurrence after 5 years as well as the HER2 status based on DCE-MRI. Using the immunofluorescence images of CD8+ T lymphocytes and cancer cells, Yu et al. [272] used “deepflow” from MXNet [273] for relapse prediction for patients with triple-negative breast cancer. Ha et al. [274] proposed to use CNN to predict the Oncotype Dx recurrence score based on MRI images to provide an alternative to the invasive and expensive genetic analysis of Oncotype Dx could be avoided. Ma et al. [275] identified the bio-markers indicating recurrence for TNBC patients with NAC treatments based on radiomics analysis in segmented images by the 3D pre-trained U-nets.

In addition to single-modality methods, multi-modal methods have also been explored in this task. Kim et al. [276] first identified 32 features related to breast cancer recurrence and developed a recurrent neural network to predict the recurrence time. To predict the recurrence and metastasis of HER2 positive breast cancer in patients, Yang et al. [277] proposed to use CNN to extract feature from WSI and combine it with clinical

information via a multi-modal model. Rabinovici et al. [278] utilized the ensemble strategy on the prediction scores on parallel CNN-based models for different MRI modalities and clinical classifiers. To predict the recurrence risk of early-stage breast cancer, Nichols et al. [279] used an artificial neural network to combine pathological, clinical, and imaging variables. Specifically, the global mammographic breast density and local breast density (LBD) are used, and LBD was measured with optical spectral imaging capable of sensing regional concentrations of tissue constituents. This method demonstrated high correlation of risk prediction with Oncotype Dx recurrence score. Whitney et al. [280] used nuclear morphology features from hematoxylin and eosin (H&E) stained images to predict risks derived by the Oncotype DX test. After feature selection, the performance of multiple machine learning methods and a deep neural network are compared.

2) Other Task: The main purpose aside from prognosis prediction is the discovery of prognostic biomarkers and explore the interactions between various prognostic factors.

One line of studies uses a neural network as the first step to detect important features. Bai et al. [281] applied a neural network for the detection of tumor cells, immune cells, fibroblast cells, and others. In the end, five machine tumor-infiltrating lymphocyte variables were derived based on features. These variables were found to be independent and robust prognostic indicators.

Another line of research uses neural networks to associate features. Balkenhol et al. [282] tried to assess the tumour infiltrating lymphocytes (TILs). A CNN was applied to register unmixed multispectral images and corresponding H&E sections. It was found that for all TILs markers, the presence of a high density of positive cells correlated with improved survival. None of the TILs markers was superior to the others. Using the graph neural network (GNN), Qiu et al. [283] proposed to tape the regional interactions among existing biomarkers (tumor size, nodal status, histologic grade, molecular subtype, etc.) to reveal hidden prognostic values. Differently, Lee et al. [284] used a GNN to explore contextual features in gigapixel-sized WSIs in a semi-supervised manner to provide interpretable prognostic biomarkers.

Recently, Zuo et al. [285] combined the WSI data and genomic data to assess the prognostic outcome. This work fused the interaction between WSI and genomic features via the attention mechanism which enabled the identification of survival-associated imaging and genomic biomarkers strongly correlated with the interaction between TILs and tumors.

D. Discussion

With the increasing accuracy achieved by deep learning models, more and more studies reported methods that had high performances that are even comparable with radiologists under limited data in mammogram [108], Ultrasound [133], MRI [286], and pathology images [192]. Growing numbers of studies are seeking extra clinical applications based on DL.

For mammograms, owing to the large-scale data and the need for population-level screening, more and more FDA-cleared or

approved DL products are available on the market, such as Lunit INSIGHT MMG, Transpara, MammoScreen 2.0, etc. Several DL algorithms and products could take multi-view inputs and conduct accurate cancer classification and lesion detection at the same time. There are increasing efforts to predict risks from even normal mammograms [287], [288], [289], which would potentially increase the screening efficacy and effectiveness. We also noticed that a recent prospective study [290] based on Mia (version 2.0, Kheiron Medical Technologies) showed that using AI as an additional reader can improve the early detection of breast cancer with relevant prognostic features, with minimal to no unnecessary recalls. Compared with mammograms, ultrasound is more effective in screening for women with dense breasts (which has a high proportion in China) and is more often used for lymph node metastasis estimation and risk prediction. Despite the prosperity of the studies on breast ultrasound, there is a lack of mature DL products that can be used in clinics. Also, the scale of public breast ultrasound data is far less than that of mammograms.

Breast MRI is often used preoperatively for treatment planning, which plays a lesser role in cancer screening than mammogram and breast ultrasound. This is partially the reason why many of the public breast MRI datasets are aimed at treatment response or prognosis predictions, and studies did show that MRI is effective in outcome prediction [258], [291]. In addition, current studies also showed that breast MRI could be used to reduce false positives [286] and potentially used in screening [170], [292], [293]. Further, it is also of great interest whether the examination process could be more efficient with less time and contrastive agent [178], [190].

DL has demonstrated the capability of analyzing pathology images in various clinical tasks, such as molecular subtyping, mitosis detection, metastasis detection, fine-grained cancer classification, virtual staining, outcome prediction, etc. Recent studies are also seeking effective algorithms that can automatically grade the cancers [294], [295] according to the well-established prognostic factor, Nottingham histological grade (NHG). A recent promising study by Amgad et al. [296] showed a DL-based population-level digital histologic biomarker consistently outperformed pathologists (usually use NHG) in predicting survival outcomes, independent of tumor–node–metastasis stage and pertinent variables. Besides, it's also quite appealing to find the association between imaging features of pathology images and the genetic information [297].

Outcome prediction, i.e., treatment response prediction and prognosis prediction, is trending in the field of deep learning-based breast cancer analysis. It usually requires clinicians to summarize a diverse range of different clinical factors and their sophisticated interconnections to predict the future outcome of the patients. Owing to their remarkable capability of learning patterns from complex high-dimensional data, DL-based breast cancer imaging also shows promise in outcome prediction by providing a non-invasive alternative to conventional analysis. We found not only an increase of studies in this direction but also the practical values demonstrated by DL techniques whose results could further help clinicians make personalized treatment plans and facilitate the discovery of potential biomarkers [296].

To summarize this section, we witnessed the growing scale of research on using deep learning for breast cancer image analysis and the increasing feasibility of DL in assisting real-world clinical applications. In the next section, we will delve into details on the challenges and future directions of DL-based breast cancer imaging from a methodological perspective.

IV. CHALLENGES AND FUTURE DIRECTIONS

A. Robust Learning With Limited Data

Large-scale training data is the key to the success of deep learning. There were some efforts to make larger-scale data accessible for research purposes [298], [299]. However, most of the available public datasets are of relatively small scales, especially the modalities other than mammograms as those are not routine exams (please find our summary of the publicly available datasets in the appendix). This is a potential factor that hinders the robustness and generalizability of deep learning-based breast cancer analysis models. A plausible direction to enlarge the training data while protecting patients' privacy is federated learning (FL) with the collaboration of multiple institutes. FL allows jointly training deep learning models without sharing data among participants, which enables cooperation across clients while also preserving the patient's privacy. Synthetic data is another potential option, and a recent pioneer work provided 2,000 synthetic pathology images for nuclei segmentation studies [300].

Another challenge naturally inherited from limited training data is the limited generalization ability of the developed model, such as the results reported by Wang et al. [116]. Novel algorithms that can improve the robustness of deep learning models on unseen domains (such as domain adaptation and domain generalization) would be of interest in the field of breast cancer analysis [301], [302]. Existing approaches are majorly based on learning invariant representations from multi-source data. A recent study also found that fine-grained annotations could improve a model's generalizability without involving training data from multiple sources [303]. It is worth mentioning that domain gaps generally exist under FL scenarios as well, where the data provided by each client are naturally heterogeneous. In addition, robustness to adversarial attack is also in a need to protect the medical analysis systems from potential threats [119], [304].

With the development of medical imaging techniques, novel imaging data would also emerge. Compared with the widely and maturely studied modalities, the new modalities would need further adaptation of current deep learning algorithms. Transfer learning and domain adaptation hold promises in addressing this challenge, and a typical example is the studies that made efforts to transfer the models learned from mammograms to DBT [305].

B. Efficient Learning With Weak Labels

Apart from data, label is another important factor that drives the thriving of deep learning models, and supervised learning is the most common form of deep learning [27]. For breast cancer imaging, labels are often obtained in two ways: determined by radiologists' interpretation of the radiology images (e.g.,

BI-RADS 2 as normal, 3 as benign, 4–6 as malignant); or determined by pathological results based on biopsy or surgery, which is also the “gold standard” for cancer identification. However, the former strategy inevitably involves inter-reader variation, while the latter is invasive and may not always be available. Moreover, annotating medical images is labor-exhaustive and expertise-dependent, and hence annotations are scarce, especially fine-grained annotations like bounding boxes or segmentation masks. The medical image annotations often cannot meet the requirement of quantity and quality, and efforts have been made to develop label-efficient learning models to leverage more available data [306], [307]. Recent studies proposed omni-supervised learning to utilize different types of annotations for training a unified network [308], [309], which could potentially serve as a unified solution to label scarcity. Considering the inter-reader variation, a potential solution is developing models that are robust to noisy labels or using the calibration of multiple-reader results [310], [311]. Self-supervised learning could largely mitigate the need for experts’ annotations, which would also be a promising solution to label scarcity. Self-supervised learning pre-trains a network with only unlabeled data has shown remarkable results (e.g., surpassing the performance of ImageNet pre-trained networks) [312], [313]. Moreover, free-text report can also benefit self-supervised learning, especially in label-efficient finetuning for downstream tasks [314]. It is also worth noting that self-supervised learning may require a large amount of data, which is entangled with the challenge of limited data.

C. Multimodal Learning for Information Fusion

In this paper, we have surveyed a number of studies with multimodal learning, and most of them involve only one type of imaging data (e.g., multiple views of mammogram; B-mode ultrasound and color Doppler; different sequences of MRI). However, different imaging techniques provide various insights into the status of breast cancer patients, while current studies have not yet fully utilized the rich context provided by all different modalities generated during the clinical process. For example, mammograms are more sensitive to calcifications than ultrasounds, MRI provides more detailed spatial information, and pathology images enable observing the microscopic environment and the cell status. Also, data beyond the images, such as clinical information, molecular biomarkers, genomics, or clinical reports, could further enrich the descriptions for the patients with more structured knowledge and are proved beneficial for developing multimodal learning models [277], [283]. For prognosis prediction, multimodal learning is also essential to improve the accuracy, explore the interconnections among the information from different modalities, and facilitate novel biomarkers [283], [315], [316]. Moreover, interconnections among the multimodal data could be further explored. For instance, the study by Coudray et al. [317] showed that genes could be predicted from pathology images. The discovery of such associations could narrow down the search space of biomarkers for clinical analysis and possibly provide non-invasive alternatives to biopsy or serology tests as well.

We can witness a trend of increase in multimodal breast cancer papers, and this line of study is yet to be further explored in the near future. The key factor that hinders this line of work is the difficulty in constructing a dataset where many different modalities such as gene, age, images, therapy, clinical record, etc., for each patient are complete and collected appropriately. As a result, future multimodal studies may also pay attention to the problem of missing modalities when the provided data are incomplete [318]. We look forward to an open-source multimodal breast cancer dataset to stimulate the development of this direction.

D. Reliable and Explainable Model Learning

For breast cancer, factors such as age and race have long been important factors that affect healthcare disparity [319], and one of the reasons is that these factors are highly related to breast density which is often evaluated for breast cancer diagnosis. The statistics by Giaquinto et al. [2] also showed that Black women have the lowest survival for every stage of breast cancer diagnosis except for stage I for which survival is similar. Deep learning, unfortunately, inherits unreliable and unfair diagnoses, especially for under-represented groups [320]. Recent studies also found that most of the data supporting approval of AI products by The United States Food and Drug Administration (FDA) did not report the race or ethnicity of the patients [321]. As a result, fair deep learning has raised a lot of attention [322]. Another problem raised by group imbalance is shortcut learning [323], which refers to the phenomenon when DL models learn wrong patterns for making decisions due to spurious correlations. One of the key reasons for shortcut learning is that the dataset is not diverse enough and the minority group(s) are under-represented and under-learned. Shortcut learning is also one of the reasons that DL models perform inconsistently on the training data and external testing data. Broadly, three types of algorithms could be explored for fair and debias model learning: pre-processing, in-processing, and post-processing, which focus on fair data stratification, fair model development, and fair modification of a trained model’s output, respectively. Most existing solutions often rely on exhaustive labeling of group information, while a recent study has proposed methods without knowing the group information [324], which is more feasible when the patients’ information is protected.

Explainability is another important factor in achieving reliable and fair deep learning models. A large proportion of previous works focused on post hoc methods to interpret an already trained model and help the users, i.e., doctors and patients, to understand the decision-making process of the AI models. A typical example is using saliency maps, such as CAM, to explain which part of an image contributes most to the final prediction. Global interpretation (e.g., by Shapley Value) that summarizes the holistic decision-making rules based on the whole cohort could also help understand the learned knowledge of the models. Recently, developing ante hoc algorithms that embed the reasoning process into DL models has gained more attention. For example, Wang et al. [122] proposed a prototype-based model which conducts classification by comparing input images with

learned prototypes. Learning and reasoning based on attributes related to diseases is also a promising approach but may require additional information for labeling the attributes [325], [326]. These methods could not fully explain the DL models but have contributed to opening the “black box” to a certain level.

E. Personalized Treatment Planning

Optimal therapy for each patient depends on tumor subtype (e.g., HER2 negative, HER2 positive, and triple-negative), anatomic cancer stage, and patient preference [241]. Personalized treatment planning for breast cancer could largely improve patients’ life quality by advancing the treatment for patients with good responses and avoiding over-treatment for patients with poor responses. In this survey, we have observed an increase in studies on breast imaging-based treatment response prediction and prognosis, of which a considerable proportion utilized a single modality of breast imaging. Considering the rich context provided by the multimodal information generated during the diagnosis process, it is of great importance to develop multimodal learning algorithms to combine imaging information, medical history, and genetic profile individualized treatment planning. Meanwhile, facilitating precise biomarkers is another crucial direction, which enables identifying patients who will benefit from either escalated or de-escalated treatment [327]. There were works on constructing new biomarkers based on H&E pathology images [267], [296], [328], and we look forward to more effective novel biomarkers in the future.

F. Accountable AI Evaluation and Regulation

With the increasing performance of AI techniques, there are rising calls to establish the accountability standard of deep learning to improve the evidence of its usefulness and fully unleash its huge potential in healthcare. Currently, many works that reported high performance of DL-based medical image analysis are retrospective studies conducted on limited datasets. Future studies should also rethink the evaluation design to improve the strength of evidence for the developed AI systems. For example, external validation is now required for reporting AI performance in radiology studies [329]. The U.K. National Screening Committee further suggested that prospective studies should be required to provide further assessment of AI systems in breast screening pathway, as enriched, multi-reader, multi-case, test laboratory studies are also biased [330]. Generalizability and robustness of the AI decision support tools should be thoroughly evaluated before adoption for patient care in the clinic [331], and we’ve seen some efforts on developing DL systems with more diverse data [332], [333]. Apart from diverse and more transparent evaluation of the developed systems, involving human experts’ intervention could also possibly achieve trustable AI. Human-in-the-loop is a promising approach where doctors can participate in the training of deep learning algorithms by providing the knowledge on the labels or attributes in a medical image to rectify the models’ decision-making process [326].

Further, accountability requires stronger and clearer regulation of AI systems. AI-based software is also emerging with the increasing studies on AI-based breast cancer analysis, and the

FDA has approved several AI software indicated for breast cancer image diagnosis. However, a recent review of FDA-approved or FDA-cleared AI software for breast cancer also reported important gaps in validation approaches, based on which the FDA evidentiary regulatory was suggested to be strengthened [321]. The efforts by clinicians, researchers, engineers, ethicists, and the government are needed to collectively ground the AI technology in breast cancer analysis.

V. CONCLUSION

Breast cancer has become the most diagnosed malignancy worldwide, and breast imaging plays a significant role in breast cancer screening, diagnosis, treatment response prediction, and prognosis. With the ground-breaking development of deep learning research, emerging studies have been conducted to apply deep learning techniques to tackle breast cancer. To this end, we conducted this survey to review the deep learning-based breast cancer imaging analysis over the past decade. Specifically, screening and diagnosis has attracted most of the attention from the deep learning community, while the increase of studies for treatment response prediction and prognosis has also been observed. The findings of this paper suggest that these studies are largely determined by the available data, especially those released to the public. By discussing the challenges and exploring the potential future directions, we hope to provide novel insights to inspire readers to devote further efforts on developing the next-generation trustworthy healthcare models for breast cancer as well as other diseases.

REFERENCES

- [1] N. Harbeck and M. Gnant, “Breast cancer,” *Lancet*, vol. 389, no. 10074, pp. 1134–1150, 2017.
- [2] A. N. Giaquinto et al., “Breast cancer statistics,” *CA: A Cancer J. Clinicians*, vol. 72, no. 6, pp. 524–541, 2022.
- [3] H. Sung et al., “Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: A Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [4] M. B. Amin et al., “The eighth edition ajcc cancer staging manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging,” *CA: A Cancer J. Clinicians*, vol. 67, no. 2, pp. 93–99, 2017.
- [5] L. Nyström et al., “Long-term effects of mammography screening: Updated overview of the swedish randomised trials,” *Lancet*, vol. 359, no. 9310, pp. 909–919, 2002.
- [6] World Health Organization, “WHO position paper on mammography screening,” Feb. 2014. [Online]. Available: <https://www.who.int/publications/i/item/who-position-paper-on-mammography-screening>
- [7] M. G. Marmot et al., “The benefits and harms of breast cancer screening: An independent review,” *Brit. J. Cancer*, vol. 108, no. 11, pp. 2205–2240, 2013.
- [8] A. Chong et al., “Digital breast tomosynthesis: Concepts and clinical practice,” *Radiology*, vol. 292, no. 1, pp. 1–14, 2019.
- [9] J. Wild and D. Neal, “Use of high-frequency ultrasonic waves for detecting changes of texture in living tissues,” *Lancet*, vol. 257, no. 6656, pp. 655–657, 1951.
- [10] C. M. Sehgal et al., “A review of breast ultrasound,” *J. Mammary Gland Biol. Neoplasia*, vol. 11, no. 2, pp. 113–123, 2006.
- [11] R. J. Hooley et al., “Breast ultrasonography: State of the art,” *Radiology*, vol. 268, no. 3, pp. 642–659, 2013.
- [12] A. Kapur et al., “Combination of digital mammography with semi-automated 3D breast ultrasound,” *Technol. Cancer Res. Treat.*, vol. 3, no. 4, pp. 325–334, 2004.

- [13] W. A. Berg et al., "Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer," *Jama*, vol. 299, no. 18, pp. 2151–2163, 2008.
- [14] L. C. Leong et al., "Supplementary breast ultrasound screening in asian women with negative but dense mammograms—a pilot study," *Ann. Acad. Med.-Singap.*, vol. 41, no. 10, 2012, Art. no. 432.
- [15] C. M. Sehgal et al., "Quantitative vascularity of breast masses by doppler imaging: Regional variations and diagnostic implications," *J. Ultrasound Med.*, vol. 19, no. 7, pp. 427–440, 2000.
- [16] R. M. Mann et al., "Breast MRI: State of the art," *Radiology*, vol. 292, no. 3, pp. 520–536, 2019.
- [17] A. Kalovidouris et al., "Fat suppression techniques for breast MRI: Dixon versus spectral fat saturation for 3D t1-weighted at 3 t," *La Radiologia Medica*, vol. 122, no. 10, pp. 731–742, 2017.
- [18] M. A. Bernstein et al., *Handbook of MRI Pulse Sequences*. New York, NY, USA: Elsevier, 2004.
- [19] M. Y. Su et al., "Correlation of dynamic contrast enhancement mri parameters with microvessel density and VEGF for assessment of angiogenesis in breast cancer," *J. Magn. Reson. imaging: JMIR*, no. 4, p. 18, 2003.
- [20] Y. Gao and S. L. Heller, "Abbreviated and ultrafast breast MRI in clinical practice," *Radiographics*, vol. 40, no. 6, pp. 1507–1527, 2020.
- [21] E. Duregon et al., "Comparative diagnostic and prognostic performances of the hematoxylin-eosin and phospho-histone h3 mitotic count and ki-67 index in adrenocortical carcinoma," *Modern Pathol.*, vol. 27, no. 9, pp. 1246–1254, 2014.
- [22] E. Klaiman et al., "Enabling histopathological annotations on immunofluorescent images through virtualization of hematoxylin and eosin," *J. Pathol. Informat.*, vol. 9, no. 1, 2018, Art. no. 1.
- [23] H. Liu, "Application of immunohistochemistry in breast pathology: A review and update," *Arch. Pathol. Lab. Med.*, vol. 138, no. 12, pp. 1629–1642, 2014.
- [24] W. M. Hanna et al., "Her2 in situ hybridization in breast cancer: Clinical implications of polysomy 17 and genetic heterogeneity," *Modern Pathol.*, vol. 27, no. 1, pp. 4–18, 2014.
- [25] A. S. Coates et al., "Tailoring therapies—improving the management of early breast cancer: St gallen international expert consensus on the primary therapy of early breast cancer 2015," *Ann. Oncol.*, vol. 26, no. 8, pp. 1533–1546, 2015.
- [26] S. J. Magny et al., "Breast imaging reporting and data system," in *StatPearls [Internet]*, StatPearls publishing, 2022.
- [27] Y. LeCun et al., "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [28] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
- [29] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Med.*, vol. 25, no. 1, pp. 44–56, 2019.
- [30] J. Bai et al., "Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: A review," *Med. Image Anal.*, vol. 71, 2021, Art. no. 102049.
- [31] A. Duggento et al., "Deep computational pathology in breast cancer," in *Seminars in Cancer Biology*, vol. 72, New York, NY, USA: Elsevier, 2021, pp. 226–237.
- [32] A. Hamidinekoo et al., "Deep learning in mammography and breast histology, an overview and future trends," *Med. Image Anal.*, vol. 47, pp. 45–67, 2018.
- [33] K. J. Geras et al., "Artificial intelligence for mammography and digital breast tomosynthesis: Current concepts and future perspectives," *Radiology*, vol. 293, no. 2, 2019, Art. no. 246.
- [34] I. Sechopoulos et al., "Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art," in *Seminars in Cancer Biology*, vol. 72, New York, NY, USA: Elsevier, 2021, pp. 214–225.
- [35] I. Goodfellow et al., *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [36] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, 1943.
- [37] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychol. Rev.*, vol. 65, no. 6, 1958, Art. no. 386.
- [38] D. E. Rumelhart et al., "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [39] Y. LeCun et al., "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.
- [40] A. Krizhevsky et al., "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira et al., Eds., vol. 25, Red Hook, New York, US: Curran Associates, Inc., 2012.
- [41] J. Deng et al., "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
- [42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Representations, Comput. Biol. Learn. Soc.*, 2015, pp. 1–14.
- [43] K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [44] G. Huang et al., "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.
- [45] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
- [46] C. Szegedy et al., "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.
- [47] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations*, 2020.
- [48] A. Vaswani et al., "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [49] R. Girshick et al., "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.
- [50] T.-Y. Lin et al., "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
- [51] H. Rezaatoughi et al., "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 658–666.
- [52] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.
- [53] S. Ren et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.
- [54] J. Redmon et al., "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
- [55] A. Bochkovskiy et al., "Yolov4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [56] F. Milletari et al., "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.
- [57] J. Long et al., "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [58] O. Ronneberger et al., "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.
- [59] R. Azad et al., "Medical image segmentation review: The success of u-net," 2022, *arXiv:2211.14830*.
- [60] K. He et al., "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.
- [61] B. Zhou et al., "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921–2929.
- [62] M.-A. Carbonneau et al., "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognit.*, vol. 77, pp. 329–353, 2018.
- [63] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [64] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 4037–4058, Nov. 2021.
- [65] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, 2018.
- [66] K. Zhou et al., "Domain generalization: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4396–4415, Apr. 2023.
- [67] J. Dheeba et al., "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach," *J. Biomed. Informat.*, vol. 49, pp. 45–52, 2014.
- [68] H.-P. Chan, R.-K. Samala, L.-M. Hadjiiski, and C. Zhou, "Deep learning in medical image analysis," *Advances in Experimental Medicine and Biology*, G. Lee and H. Fujita, Eds., Springer, Cham, vol. 1213, 2020.
- [69] N. Dhungel et al., "The automated learning of deep features for breast mass classification from mammograms," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, Springer, 2016, pp. 106–114.

- [70] P. Teare et al., "Malignancy detection on mammography using dual deep convolutional neural networks and genetically discovered false color input enhancement," *J. Digit. Imag.*, vol. 30, no. 4, pp. 499–505, 2017.
- [71] A. A. Mohamed et al., "A deep learning method for classifying mammographic breast density categories," *Med. Phys.*, vol. 45, no. 1, pp. 314–321, 2018.
- [72] A. P. Adedigba et al., "Performance evaluation of deep learning models on mammogram classification using small dataset," *Bioengineering*, vol. 9, no. 4, 2022, Art. no. 161.
- [73] X. Li et al., "Digital breast tomosynthesis versus digital mammography: Integration of image modalities enhances deep learning-based breast mass classification," *Eur. Radiol.*, vol. 30, no. 2, pp. 778–788, 2020.
- [74] R. K. Samala et al., "Breast cancer diagnosis in digital breast tomosynthesis: Effects of training sample size on multi-stage transfer learning using deep neural nets," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 686–696, Mar. 2019.
- [75] K. Clancy et al., "Deep learning pre-training strategy for mammogram image classification: An evaluation study," *J. Digit. Imag.*, vol. 33, no. 5, pp. 1257–1265, 2020.
- [76] E. Sert et al., "Ensemble of convolutional neural networks for classification of breast microcalcification from mammograms," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2017, pp. 689–692.
- [77] H. Li et al., "Signed laplacian deep learning with adversarial augmentation for improved mammography diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2019, pp. 486–494.
- [78] W. Zhu et al., "Deep multi-instance networks with sparse label assignment for whole mammogram classification," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2017, pp. 603–611.
- [79] X. Shu et al., "Deep neural networks with region-based pooling structures for mammographic image classification," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2246–2255, Jun. 2020.
- [80] W. Lotter et al., "Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach," *Nature Med.*, vol. 27, no. 2, pp. 244–249, 2021.
- [81] M. Yousefi et al., "Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning," *Comput. Biol. Med.*, vol. 96, pp. 283–293, 2018.
- [82] G. Carneiro et al., "Unregistered multiview mammogram analysis with pre-trained deep learning models," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2015, pp. 652–660.
- [83] G. Carneiro et al., "Automated analysis of unregistered multi-view mammograms with deep learning," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2355–2365, Nov. 2017.
- [84] J. Bai et al., "Feature fusion siamese network for breast cancer detection comparing current and prior mammograms," *Med. Phys.*, vol. 49, no. 6, pp. 3654–3669, 2022.
- [85] A. Baccouche et al., "Early detection and classification of abnormality in prior mammograms using image-to-image translation and yolo techniques," *Comput. Methods Programs Biomed.*, vol. 221, 2022, Art. no. 106884.
- [86] S. Shams et al., "Deep generative breast cancer screening and diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2018, pp. 859–867.
- [87] M. Wimmer et al., "Multi-task fusion for improving mammography screening data classification," *IEEE Trans. Med. Imag.*, vol. 41, no. 4, pp. 937–950, Apr. 2022.
- [88] M. Suganthi and M. Madheswaran, "An improved medical decision support system to identify the breast cancer using mammogram," *J. Med. Syst.*, vol. 36, no. 1, pp. 79–91, 2012.
- [89] R. Shen et al., "Multicontext multitask learning networks for mass detection in mammogram," *Med. Phys.*, vol. 47, no. 4, pp. 1566–1578, 2020.
- [90] M. Buda et al., "A data set and deep learning algorithm for the detection of masses and architectural distortions in digital breast tomosynthesis images," *JAMA Netw. Open*, vol. 4, no. 8, pp. e2119100–e2119100, 2021.
- [91] Y. Li et al., "Architectural distortion detection based on superior–inferior directional context and anatomic prior knowledge in digital breast tomosynthesis," *Med. Phys.*, vol. 49, no. 6, pp. 3749–3768, 2022.
- [92] Y. Liu et al., "Compare and contrast: Detecting mammographic soft-tissue lesions with C2-Net," *Med. Image Anal.*, vol. 71, 2021, Art. no. 101999.
- [93] Y. Liu et al., "Act like a radiologist: Towards reliable multi-view correspondence reasoning for mammogram mass detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 5947–5961, Oct. 2022.
- [94] Z. Zhao et al., "Check and link: Pairwise lesion correspondence guides mammogram mass detection," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2022, pp. 384–400.
- [95] Z. Yang et al., "Momminet-v2: Mammographic multi-view mass identification networks," *Med. Image Anal.*, vol. 73, 2021, Art. no. 102204.
- [96] G. Liang et al., "Weakly-supervised self-training for breast cancer localization," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2020, pp. 1124–1127.
- [97] W. Zhu et al., "Adversarial deep structured nets for mass segmentation from mammograms," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, 2018, pp. 847–850.
- [98] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
- [99] Y. Li et al., "SAP-CGAN: Adversarial learning for breast mass segmentation in digital mammogram based on superpixel average pooling," *Med. Phys.*, vol. 48, no. 3, pp. 1157–1167, 2021.
- [100] N. R. Gudhe et al., "Area-based breast percentage density estimation in mammograms using weight-adaptive multitask learning," *Sci. Rep.*, vol. 12, no. 1, pp. 1–19, 2022.
- [101] R. F. Ribeiro et al., "Deep learning methods for lesion detection on mammography images: A comparative analysis," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2022, pp. 3526–3529.
- [102] P. H. Yi et al., "Deepcat: Deep computer-aided triage of screening mammography," *J. Digit. Imag.*, vol. 34, no. 1, pp. 27–35, 2021.
- [103] H. Min et al., "Fully automatic computer-aided mass detection and segmentation via pseudo-color mammograms and mask R-CNN," in *Proc. IEEE 17th Int. Symp. Biomed. Imag.*, 2020, pp. 1111–1115.
- [104] S. Hwang and H.-E. Kim, "Self-transfer learning for weakly supervised lesion localization," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2016, pp. 239–246.
- [105] R. Bakalo et al., "Classification and detection in mammograms with weak supervision via dual branch deep neural net," in *Proc. IEEE 16th Int. Symp. Biomed. Imag.*, 2019, pp. 1905–1909.
- [106] M. V. Sainz de Cea et al., "Multi-task learning for detection and classification of cancer in screening mammography," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2020, pp. 241–250.
- [107] H. Li et al., "Dual convolutional neural networks for breast mass segmentation and diagnosis in mammography," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 3–13, Jan. 2022.
- [108] A. Rodriguez-Ruiz et al., "Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists," *JNCI: J. Nat. Cancer Inst.*, vol. 111, no. 9, pp. 916–922, 2019.
- [109] N. Wu et al., "Deep neural networks improve radiologists' performance in breast cancer screening," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1184–1194, Apr. 2020.
- [110] A. Rodríguez-Ruiz et al., "Detection of breast cancer with mammography: Effect of an artificial intelligence support system," *Radiology*, vol. 290, no. 2, pp. 305–314, 2019.
- [111] M. C. Pinto et al., "Impact of artificial intelligence decision support using deep learning on breast cancer screening interpretation with single-view wide-angle digital breast tomosynthesis," *Radiology*, vol. 300, no. 3, pp. 529–536, 2021.
- [112] C. D. Lehman et al., "Deep learning vs traditional breast cancer risk models to support risk-based mammography screening," *J. Nat. Cancer Inst.*, vol. 114, no. 10, pp. 1355–1363, 2022.
- [113] K. Lång et al., "Identifying normal mammograms in a large screening population using artificial intelligence," *Eur. Radiol.*, vol. 31, no. 3, pp. 1687–1692, 2021.
- [114] S. M. McKinney et al., "International evaluation of an ai system for breast cancer screening," *Nature*, vol. 577, no. 7788, pp. 89–94, 2020.
- [115] A. Rodriguez-Ruiz et al., "Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? a feasibility study," *Eur. Radiol.*, vol. 29, no. 9, pp. 4825–4832, 2019.
- [116] X. Wang et al., "Inconsistent performance of deep learning models on mammogram classification," *J. Amer. College Radiol.*, vol. 17, no. 6, pp. 796–803, 2020.
- [117] W. Hsu et al., "External validation of an ensemble model for automated mammography interpretation by artificial intelligence," *JAMA Netw. Open*, vol. 5, no. 11, pp. e2242343–e2242343, 2022.
- [118] Y. Dgani et al., "Training a neural network based on unreliable human annotation of medical images," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, 2018, pp. 39–42.
- [119] Q. Zhou et al., "A machine and human reader study on ai diagnosis model safety under attacks of adversarial images," *Nature Commun.*, vol. 12, no. 1, pp. 1–11, 2021.

- [120] Z. Li et al., "Domain generalization for mammography detection via multi-style and multi-view contrastive learning," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2021, pp. 98–108.
- [121] H. R. Roth et al., "Federated learning for breast density classification: A real-world implementation," in *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning*. Berlin, Germany: Springer, 2020, pp. 181–191.
- [122] C. Wang et al., "Knowledge distillation to ensemble global and interpretable prototype-based mammogram classification models," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2022, pp. 14–24.
- [123] G. Jiang et al., "Synthesis of mammogram from digital breast tomosynthesis using deep convolutional neural network with gradient guided cGANs," *IEEE Trans. Med. Imag.*, vol. 40, no. 8, pp. 2080–2091, Aug. 2021.
- [124] J.-Z. Cheng et al., "Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans," *Sci. Rep.*, vol. 6, no. 1, pp. 1–13, 2016.
- [125] W. K. Moon et al., "Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 190, 2020, Art. no. 105361.
- [126] X. Zheng et al., "Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer," *Nature Commun.*, vol. 11, no. 1, pp. 1–9, 2020.
- [127] L.-Q. Zhou et al., "Lymph node metastasis prediction from primary breast cancer us images using deep learning," *Radiology*, vol. 294, no. 1, pp. 19–28, 2020.
- [128] X. Guo et al., "Deep learning radiomics of ultrasonography: Identifying the risk of axillary non-sentinel lymph node involvement in primary breast cancer," *EBioMedicine*, vol. 60, 2020, Art. no. 103018.
- [129] Y.-W. Lee et al., "Axillary lymph node metastasis status prediction of early-stage breast cancer using convolutional neural networks," *Comput. Biol. Med.*, vol. 130, 2021, Art. no. 104206.
- [130] J. Xing et al., "Using bi-rads stratifications as auxiliary information for breast masses classification in ultrasound images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 2058–2070, Jun. 2021.
- [131] R. Huang et al., "AW3M: An auto-weighting and recovery framework for breast cancer diagnosis using multi-modal ultrasound," *Med. Image Anal.*, vol. 72, 2021, Art. no. 102137.
- [132] S.-Y. Lu et al., "Safnet: A deep spatial attention network with classifier fusion for breast cancer detection," *Comput. Biol. Med.*, vol. 148, 2022, Art. no. 105812.
- [133] X. Qian et al., "Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning," *Nature Biomed. Eng.*, vol. 5, no. 6, pp. 522–532, 2021.
- [134] H. Tanaka et al., "Computer-aided diagnosis system for breast ultrasound images using deep learning," *Phys. Med. Biol.*, vol. 64, no. 23, 2019, Art. no. 235013.
- [135] S. Han et al., "A deep learning framework for supporting the classification of breast lesions in ultrasound images," *Phys. Med. Biol.*, vol. 62, no. 19, 2017, Art. no. 7714.
- [136] A. Hijab et al., "Breast cancer classification in ultrasound images using transfer learning," in *Proc. 5th Int. Conf. Adv. Biomed. Eng.*, 2019, pp. 1–4.
- [137] T. Pang et al., "Semi-supervised GaN-based radiomics model for data augmentation in breast ultrasound mass classification," *Comput. Methods Programs Biomed.*, vol. 203, 2021, Art. no. 106018.
- [138] E. Zhang et al., "Birads features-oriented semi-supervised deep learning for breast ultrasound computer-aided diagnosis," *Phys. Med. Biol.*, vol. 65, no. 12, 2020, Art. no. 125005.
- [139] X. Wang et al., "Deep virtual adversarial self-training with consistency regularization for semi-supervised medical image classification," *Med. Image Anal.*, vol. 70, 2021, Art. no. 102010.
- [140] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [141] T. Miyato et al., "Virtual adversarial training: A regularization method for supervised and semi-supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1979–1993, Aug. 2019.
- [142] M. H. Yap et al., "Automated breast ultrasound lesions detection using convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 4, pp. 1218–1226, Jul. 2018.
- [143] T.-C. Chiang et al., "Tumor detection in automated breast ultrasound using 3-D CNN and prioritized candidate aggregation," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 240–249, Jan. 2019.
- [144] Y. Wang et al., "Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 866–876, Apr. 2020.
- [145] Z. Cao et al., "An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures," *BMC Med. Imag.*, vol. 19, no. 1, pp. 1–9, 2019.
- [146] S. Y. Shin et al., "Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images," *IEEE Trans. Med. Imag.*, vol. 38, no. 3, pp. 762–774, Mar. 2019.
- [147] J. Xing et al., "Lesion segmentation in ultrasound using semi-pixel-wise cycle generative adversarial nets," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 6, pp. 2555–2565, Nov.–Dec. 2020.
- [148] W. Gómez-Flores and W. C. de Albuquerque Pereira, "A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound," *Comput. Biol. Med.*, vol. 126, 2020, Art. no. 104036.
- [149] Z. Ning et al., "SMU-Net: Saliency-guided morphology-aware U-Net for breast lesion segmentation in ultrasound image," *IEEE Trans. Med. Imag.*, vol. 41, no. 2, pp. 476–490, Feb. 2022.
- [150] C. Xue et al., "Global guidance network for breast lesion segmentation in ultrasound images," *Med. Image Anal.*, vol. 70, 2021, Art. no. 101989.
- [151] R. Huang et al., "Boundary-rendering network for breast lesion segmentation in ultrasound images," *Med. Image Anal.*, vol. 80, 2022, Art. no. 102478.
- [152] M. R. Ferreira et al., "Comparative analysis of current deep learning networks for breast lesion segmentation in ultrasound images," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2022, pp. 3878–3881.
- [153] D. Zhai et al., "ASS-GAN: Asymmetric semi-supervised GaN for breast ultrasound image segmentation," *Neurocomputing*, vol. 493, pp. 204–216, 2022.
- [154] W.-X. Liao et al., "Automatic identification of breast ultrasound image based on supervised block-based region segmentation algorithm and features combination migration deep learning model," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 4, pp. 984–993, Apr. 2020.
- [155] Y. Zhou et al., "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images," *Med. Image Anal.*, vol. 70, 2021, Art. no. 101918.
- [156] J. Kim et al., "Weakly-supervised deep learning for ultrasound diagnosis of breast cancer," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, 2021.
- [157] C. Chen et al., "Domain knowledge powered deep learning for breast cancer diagnosis based on contrast-enhanced ultrasound videos," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2439–2451, Sep. 2021.
- [158] B. Lei et al., "Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder-decoder network," *Neurocomputing*, vol. 321, pp. 178–186, 2018.
- [159] Y. Zhou et al., "Cross-model attention-guided tumor segmentation for 3D automated breast ultrasound (ABUS) images," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 1, pp. 301–311, Jan. 2022.
- [160] Z. Cao et al., "Breast tumor classification through learning from noisy labeled ultrasound images," *Med. Phys.*, vol. 47, no. 3, pp. 1048–1057, 2020.
- [161] H. Zou et al., "A robust breast ultrasound segmentation method under noisy annotations," *Comput. Methods Programs Biomed.*, vol. 209, 2021, Art. no. 106327.
- [162] H. Feng et al., "A knowledge-driven feature learning and integration method for breast cancer diagnosis on multi-sequence MRI," *Magn. Reson. Imag.*, vol. 69, pp. 40–48, 2020.
- [163] Y. Zhang et al., "Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers," *Eur. Radiol.*, vol. 31, no. 4, pp. 2559–2567, 2021.
- [164] N. O. Antropova et al., "Use of clinical MRI maximum intensity projections for improved breast lesion classification with deep convolutional neural networks," *J. Med. Imag.*, vol. 5, no. 1, 2018, Art. no. 014503.
- [165] T. Fujioka et al., "Deep-learning approach with convolutional neural network for classification of maximum intensity projections of dynamic contrast-enhanced breast magnetic resonance imaging," *Magn. Reson. Imag.*, vol. 75, pp. 1–8, 2021.
- [166] Q. Hu et al., "Improved classification of benign and malignant breast lesions using deep feature maximum intensity projection MRI in breast cancer diagnosis using dynamic contrast-enhanced MRI," *Radiol.: Artif. Intell.*, vol. 3, no. 3, 2021, Art. no. e200159.
- [167] Q. Hu et al., "A deep learning methodology for improved breast cancer diagnosis using multiparametric MRI," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020.

- [168] T. Ren et al., "Convolutional neural network of multiparametric MRI accurately detects axillary lymph node metastasis in breast cancer patients with pre neoadjuvant chemotherapy," *Clin. Breast Cancer*, vol. 22, no. 2, pp. 170–177, 2022.
- [169] H. M. Whitney et al., "Comparison of breast MRI tumor classification using human-engineered radiomics, transfer learning from deep convolutional neural networks, and fusion methods," *Proc. IEEE*, vol. 108, no. 1, pp. 163–177, Jan. 2020.
- [170] X. Jing et al., "Using deep learning to safely exclude lesions with only ultrafast breast MRI to shorten acquisition and reading time," *Eur. Radiol.*, vol. 32, no. 12, pp. 8706–8715, 2022.
- [171] R. Sun et al., "Prediction of breast cancer molecular subtypes using DCE-MRI based on CNNs combined with ensemble learning," *Phys. Med. Biol.*, vol. 66, no. 17, 2021, Art. no. 175009.
- [172] R. Rasti, "Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks," *Pattern Recognit.*, vol. 72, pp. 381–390, 2017.
- [173] M. Z. Liu et al., "Weakly supervised deep learning approach to breast MRI assessment," *Academic Radiol.*, vol. 29, pp. S166–S172, 2022.
- [174] R. Sun et al., "Transfer learning strategy based on unsupervised learning and ensemble learning for breast cancer molecular subtype prediction using dynamic contrast-enhanced MRI," *J. Magn. Reson. Imag.*, vol. 55, no. 5, pp. 1518–1534, 2022.
- [175] M. U. Dalmış et al., "Fully automated detection of breast cancer in screening MRI using convolutional neural networks," *J. Med. Imag.*, vol. 5, no. 1, 2018, Art. no. 014502.
- [176] G. Maicas et al., "Deep reinforcement learning for active breast lesion detection from dce-MRI," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2017, pp. 665–673.
- [177] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [178] F. Ayatollahi et al., "Automatic breast lesion detection in ultrafast DCE-MRI using deep learning," *Med. Phys.*, vol. 48, no. 10, pp. 5897–5907, 2021.
- [179] W. Yue et al., "Deep learning-based automatic segmentation for size and volumetric measurement of breast cancer on magnetic resonance imaging," *Front. Oncol.*, vol. 12, 2022, Art. no. 984626.
- [180] J. Vidal et al., "A u-Net ensemble for breast lesion segmentation in DCE MRI," *Comput. Biol. Med.*, vol. 140, 2022, Art. no. 105093.
- [181] A. Galli et al., "A pipelined tracer-aware approach for lesion segmentation in breast DCE-MRI," *J. Imag.*, vol. 7, no. 12, 2021, Art. no. 276.
- [182] J. Zhang et al., "Hierarchical convolutional neural networks for segmentation of breast tumors in MRI with application to radiogenomics," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 435–447, Feb. 2019.
- [183] J. Zhu et al., "Development and validation of a deep learning model for breast lesion segmentation and characterization in multiparametric MRI," *Front. Oncol.*, vol. 12, 2022, Art. no. 946580.
- [184] V. S. Parekh et al., "Multiparametric deep learning tissue signatures for a radiological biomarker of breast cancer: Preliminary results," *Med. Phys.*, vol. 47, no. 1, pp. 75–88, 2020.
- [185] J. Zhou et al., "Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in mr images," *J. Magn. Reson. Imag.*, vol. 50, no. 4, pp. 1144–1151, 2019.
- [186] L. Luo et al., "Deep angular embedding and feature correlation attention for breast MRI cancer analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2019, pp. 504–512.
- [187] Y. Chen et al., "A deep learning model based on dynamic contrast-enhanced magnetic resonance imaging enables accurate prediction of benign and malignant breast lessons," *Front. Oncol.*, vol. 12, 2022, Art. no. 943415.
- [188] B. H. van der Velden et al., "Volumetric breast density estimation on MRI using explainable deep learning regression," *Sci. Rep.*, vol. 10, no. 1, pp. 1–9, 2020.
- [189] W. Liu et al., "Preoperative prediction of ki-67 status in breast cancer with multiparametric MRI using transfer learning," *Academic Radiol.*, vol. 28, no. 2, pp. e44–e53, 2021.
- [190] M. Chung et al., "Deep learning to simulate contrast-enhanced breast MRI of invasive breast cancer," *Radiology*, vol. 306, no. 3, 2022, Art. no. 213199.
- [191] N. Hatipoglu and G. Bilgin, "Classification of histopathological images using convolutional neural network," in *Proc. 4th Int. Conf. Image Process. Theory, Tools Appl.*, 2014, pp. 1–6.
- [192] B. E. Bejnordi et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017.
- [193] D. Wang et al., "Deep learning for identifying metastatic breast cancer," 2016, *arXiv:1606.05718*.
- [194] B. E. Bejnordi et al., "Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images," *J. Med. Imag.*, vol. 4, no. 4, 2017, Art. no. 044504.
- [195] B. Kong et al., "Cancer metastasis detection via spatially structured deep network," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, Springer, 2017, pp. 236–248.
- [196] A. Cruz-Roa et al., "Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent," *Sci. Rep.*, vol. 7, no. 1, pp. 1–14, 2017.
- [197] M. Ilse et al., "Attention-based deep multiple instance learning," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2127–2136.
- [198] N. Petrick et al., "SPIE-AAPM-NCI breastpathq challenge: An image analysis challenge for quantitative tumor cellularity assessment in breast cancer histology images following neoadjuvant treatment," *J. Med. Imag.*, vol. 8, no. 3, pp. 034501–034501, 2021.
- [199] G. Campanella et al., "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Med.*, vol. 25, no. 8, pp. 1301–1309, 2019.
- [200] R. Valieris et al., "Deep learning predicts underlying features on pathology images with therapeutic relevance for breast and gastric cancer," *Cancers*, vol. 12, no. 12, 2020, Art. no. 3687.
- [201] M. Saini and S. Susan, "Vggin-Net: Deep transfer network for imbalanced breast cancer dataset," in *Proc. IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2022, pp. 752–762.
- [202] S. Mehta et al., "End-to-end diagnosis of breast biopsy images with transformers," *Med. Image Anal.*, vol. 79, 2022, Art. no. 102466.
- [203] Z. Wang et al., "Label cleaning multiple instance learning: Refining coarse annotations on single whole-slide images," *IEEE Trans. Med. Imag.*, vol. 41, no. 12, pp. 3952–3968, Dec. 2022.
- [204] B. Li et al., "Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 14318–14328.
- [205] W. Lu et al., "Slidegraph: Whole slide image level graphs to predict her2 status in breast cancer," *Med. Image Anal.*, vol. 80, 2022, Art. no. 102486.
- [206] H. Wang et al., "Iteratively coupled multiple instance learning from instance to bag classifier for whole slide image classification," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, Springer, 2023, pp. 467–476.
- [207] W. Wang et al., "When an image is worth 1,024 x 1,024 words: A case study in computational pathology," 2023, *arXiv:2312.03558*.
- [208] M. C. Balkenhol et al., "Deep learning assisted mitotic counting for breast cancer," *Lab. Investigation*, vol. 99, no. 11, pp. 1596–1606, 2019.
- [209] M. Sebai et al., "Maskmitosis: A deep learning framework for fully supervised, weakly supervised, and unsupervised mitosis detection in histopathology images," *Med. Biol. Eng. Comput.*, vol. 58, no. 7, pp. 1603–1623, 2020.
- [210] C. Li et al., "Deepmitosis: Mitosis detection via deep detection, verification and segmentation networks," *Med. Image Anal.*, vol. 45, pp. 121–133, 2018.
- [211] A. Sohai et al., "Mitotic nuclei analysis in breast cancer histopathology images using deep ensemble classifier," *Med. Image Anal.*, vol. 72, 2021, Art. no. 102121.
- [212] S. A. Nasser, "Domain generalisation for mitosis detection exploiting preprocessing homogenizers," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2021, pp. 77–80.
- [213] H. Lei et al., "Attention-guided multi-branch convolutional neural network for mitosis detection from histopathological images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 358–370, Feb. 2020.
- [214] D. Tellez et al., "Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2126–2136, Sep. 2018.
- [215] C. W. Elston and I. O. Ellis, "Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up," *Histopathol.*, vol. 19, no. 5, pp. 403–410, 1991.
- [216] E. A. Rakha et al., "Breast cancer prognostic classification in the molecular era: The role of histological grade," *Breast Cancer Res.*, vol. 12, no. 4, pp. 1–12, 2010.
- [217] H. Su et al., "Region segmentation in histopathological breast cancer images using deep convolutional neural network," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, 2015, pp. 55–58.

- [218] L. Chan et al., "Histosegnet: Semantic segmentation of histological tissue type in whole slide images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 10662–10671.
- [219] C. Lu et al., "Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers," *Lab. Investigation*, vol. 98, no. 11, pp. 1438–1448, 2018.
- [220] J. Zhao et al., "AI-net: Attention learning network based on multi-task learning for cervical nucleus segmentation," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 6, pp. 2693–2702, 2021.
- [221] S. Graham et al., "One model is all you need: Multi-task learning enables simultaneous histology image segmentation and classification," *Med. Image Anal.*, vol. 83, 2022, Art. no. 102685.
- [222] T. Wan et al., "Robust nuclei segmentation in histopathology using aspp-net and boundary refinement," *Neurocomputing*, vol. 408, pp. 144–156, 2020.
- [223] F. Mahmood et al., "Deep adversarial training for multi-organ nuclei segmentation in histopathology images," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3257–3267, Nov. 2019.
- [224] N. Kumar et al., "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.
- [225] X. Xie et al., "Instance-aware self-supervised learning for nuclei segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2020, pp. 341–350.
- [226] Y. Lin et al., "Insmix: Towards realistic generative data augmentation for nuclei instance segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2022, pp. 140–149.
- [227] Y. Zhou et al., "CIA-Net: Robust nuclei instance segmentation with contour-aware information aggregation," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, Springer, 2019, pp. 682–693.
- [228] P. Naylor et al., "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 448–459, Feb. 2019.
- [229] S. Graham et al., "Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Med. Image Anal.*, vol. 58, 2019, Art. no. 101563.
- [230] H. Lee and W.-K. Jeong, "Scribble2label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2020, pp. 14–23.
- [231] H. Qu et al., "Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3655–3666, Nov. 2020.
- [232] Y. Lin et al., "Nuclei segmentation with point annotations from pathology images via self-supervised learning and co-training," *Med. Image Anal.*, vol. 89, 2023, Art. no. 102933.
- [233] I. Yoo et al., "Pseudoedgenet: Nuclei segmentation only with point annotations," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2019, pp. 731–739.
- [234] R. Challen et al., "Artificial intelligence, bias and clinical safety," *BMJ Qual. Saf.*, vol. 28, no. 3, pp. 231–237, 2019.
- [235] K. Faryna et al., "Tailoring automated data augmentation to h&e-stained histopathology," in *Proc. Med. Imag. With Deep Learn.*, 2021.
- [236] K. de Haan et al., "Deep learning-based transformation of h&e stained tissues into special stains," *Nature Commun.*, vol. 12, no. 1, pp. 1–13, 2021.
- [237] P. Alirezazadeh et al., "Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images," *Biocybernetics Biomed. Eng.*, vol. 38, no. 3, pp. 671–683, 2018.
- [238] M. W. Lafarge et al., "Domain-adversarial neural networks to address the appearance variability of histopathology images," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Berlin, Germany: Springer, 2017, pp. 83–91.
- [239] Y. Rivenson et al., "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," *Nature Biomed. Eng.*, vol. 3, no. 6, pp. 466–477, 2019.
- [240] J. Noorbakhsh et al., "Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images," *Nature Commun.*, vol. 11, no. 1, pp. 1–14, 2020.
- [241] A. G. Waks and E. P. Winer, "Breast cancer treatment: A review," *JAMA*, vol. 321, no. 3, pp. 288–300, 2019.
- [242] H. Charfare et al., "Neoadjuvant chemotherapy in breast cancer," *J. Brit. Surg.*, vol. 92, no. 1, pp. 14–23, 2005.
- [243] A. Thompson and S. Moulder-Thompson, "Neoadjuvant treatment of breast cancer," *Ann. Oncol.*, vol. 23, pp. x231–x236, 2012.
- [244] J. H. Choi et al., "Early prediction of neoadjuvant chemotherapy response for advanced breast cancer using PET/MRI image deep learning," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020.
- [245] T. H. Qi et al., "Multi-center evaluation of artificial intelligent imaging and clinical models for predicting neoadjuvant chemotherapy response in breast cancer," *Breast Cancer Res. Treat.*, vol. 193, no. 1, pp. 121–138, 2022.
- [246] G. Von Minckwitz et al., "Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes," *J. Clin. Oncol.*, vol. 30, no. 15, pp. 1796–1804, 2012.
- [247] P. Cortazar et al., "Pathological complete response and long-term clinical benefit in breast cancer: The ctneo pooled analysis," *Lancet*, vol. 384, no. 9938, pp. 164–172, 2014.
- [248] X. Liang et al., "Machine learning with magnetic resonance imaging for prediction of response to neoadjuvant chemotherapy in breast cancer: A systematic review and meta-analysis," *Eur. J. Radiol.*, vol. 150, 2022, Art. no. 110247.
- [249] M. Z. Liu et al., "A novel CNN algorithm for pathological complete response prediction using an i-spy trial breast MRI database," *Magn. Reson. Imag.*, vol. 73, pp. 148–151, 2020.
- [250] R. Ha et al., "Prior to initiation of chemotherapy, can we predict breast tumor response? Deep learning convolutional neural networks approach using a breast MRI tumor dataset," *J. Digit. Imag.*, vol. 32, no. 5, pp. 693–701, 2019.
- [251] S. Farahmand et al., "Deep learning trained on hematoxylin and eosin tumor region of interest predicts HER2 status and trastuzumab treatment response in HER2 breast cancer," *Modern Pathol.*, vol. 35, no. 1, pp. 44–51, 2022.
- [252] M. C. Comes et al., "Early prediction of neoadjuvant chemotherapy response by exploiting a transfer learning approach on breast DCE-MRIs," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021.
- [253] B. Li et al., "Deep learning with biopsy whole slide images for pre-treatment prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer: A multicenter study," *Breast*, vol. 66, pp. 183–190, 2022.
- [254] M. Jiang et al., "Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer," *Eur. J. Cancer*, vol. 147, pp. 95–105, 2021.
- [255] H. Taleghamar et al., "Deep learning of quantitative ultrasound multi-parametric images at pre-treatment to predict breast cancer response to chemotherapy," *Sci. Rep.*, vol. 12, no. 1, pp. 1–13, 2022.
- [256] S. Joo et al., "Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer," *Sci. Rep.*, vol. 11, no. 1, pp. 1–8, 2021.
- [257] F. Aghaei et al., "Computer-aided breast mr image feature analysis for prediction of tumor response to chemotherapy," *Med. Phys.*, vol. 42, no. 11, pp. 6520–6528, 2015.
- [258] Y. Peng et al., "Pretreatment DCE-MRI-based deep learning outperforms radiomics analysis in predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer," *Front. Oncol.*, vol. 12, pp. 846775–846775, 2022.
- [259] J. Xie et al., "Dual-branch convolutional neural network based on ultrasound imaging in the early prediction of neoadjuvant chemotherapy response in patients with locally advanced breast cancer," *Front. Oncol.*, vol. 12, 2022, Art. no. 812463.
- [260] T. Tong et al., "Dual-input transformer: An end-to-end model for preoperative assessment of pathological complete response to neoadjuvant chemotherapy in breast cancer ultrasonography," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 1, pp. 251–262, Jan. 2023.
- [261] Y. Liu et al., "Early prediction of treatment response to neoadjuvant chemotherapy based on longitudinal ultrasound images of her2-positive breast cancer patients by siamese multi-task network: A multicentre, retrospective cohort study," *EClinicalMedicine*, vol. 52, 2022, Art. no. 101562.
- [262] M. Byra et al., "Early prediction of response to neoadjuvant chemotherapy in breast cancer sonography using siamese convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 3, pp. 797–805, Mar. 2021.
- [263] M. Byra et al., "Prediction of response to neoadjuvant chemotherapy in breast cancer with recurrent neural networks and raw ultrasound signals," *Phys. Med. Biol.*, vol. 67, no. 18, 2022, Art. no. 185007.

- [264] J. Gu et al., "Deep learning radiomics of ultrasonography can predict response to neoadjuvant chemotherapy in breast cancer at an early stage of treatment: A prospective study," *Eur. Radiol.*, vol. 32, no. 3, pp. 2099–2109, 2022.
- [265] B. Q. Huynh et al., "Comparison of breast DCE-MRI contrast time points for predicting response to neoadjuvant chemotherapy using deep convolutional neural network features with transfer learning," *Proc. SPIE*, vol. 10134, pp. 207–213, 2017.
- [266] L. Wu et al., "An integrated deep learning model for the prediction of pathological complete response to neoadjuvant chemotherapy with serial ultrasonography in breast cancer patients: A multicentre, retrospective study," *Breast Cancer Res.*, vol. 24, no. 1, pp. 1–17, 2022.
- [267] F. Li et al., "Predicting neoadjuvant chemotherapy benefit using deep learning from stromal histology in breast cancer," *npj Breast Cancer*, vol. 8, no. 1, pp. 1–11, 2022.
- [268] H. Liu and T. Kurc, "Deep learning for survival analysis in breast cancer with whole slide image data," *Bioinformatics*, vol. 38, no. 14, pp. 3629–3637, 2022.
- [269] M. Sandler et al., "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.
- [270] Z. Wang et al., "Gpbn: Deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction," *Bioinformatics*, vol. 37, no. 18, pp. 2963–2970, 2021.
- [271] G. Liu et al., "Mask-guided convolutional neural network for breast tumor prognostic outcome prediction on 3D DCE-MR images," *J. Digit. Imag.*, vol. 34, no. 3, pp. 630–636, 2021.
- [272] G. Yu et al., "Predicting relapse in patients with triple negative breast cancer (TNBC) using a deep-learning approach," *Front. Physiol.*, vol. 11, 2020, Art. no. 511071.
- [273] T. Chen et al., "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems," in *Proc. Neural Inf. Process. Syst., Workshop Mach. Learn. Syst.*, 2016.
- [274] R. Ha et al., "Convolutional neural network using a breast MRI tumor dataset can predict oncotype dx recurrence score," *J. Magn. Reson. Imag.*, vol. 49, no. 2, pp. 518–524, 2019.
- [275] M. Ma et al., "Radiomics features based on automatic segmented MRI images: Prognostic biomarkers for triple-negative breast cancer treated with neoadjuvant chemotherapy," *Eur. J. Radiol.*, vol. 146, 2022, Art. no. 110095.
- [276] J.-Y. Kim et al., "Deep learning-based prediction model for breast cancer recurrence using adjuvant breast cancer cohort in tertiary cancer center registry," *Front. Oncol.*, vol. 11, 2021, Art. no. 596364.
- [277] J. Yang et al., "Prediction of her2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning," *Comput. Struct. Biotechnol. J.*, vol. 20, pp. 333–342, 2022.
- [278] S. Rabinovici-Cohen et al., "Multimodal prediction of five-year breast cancer recurrence in women who receive neoadjuvant chemotherapy," *Cancers*, vol. 14, no. 16, 2022, Art. no. 3848.
- [279] B. S. Nichols et al., "Quantitative assessment of distant recurrence risk in early stage breast cancer using a nonlinear combination of pathological, clinical and imaging variables," *J. Biophotonics*, vol. 13, no. 10, 2020, Art. no. e201960235.
- [280] J. Whitney et al., "Quantitative nuclear histomorphometry predicts oncotype dx risk categories for early stage breast cancer," *BMC Cancer*, vol. 18, no. 1, pp. 1–15, 2018.
- [281] Y. Bai et al., "An open-source, automated tumor-infiltrating lymphocyte algorithm for prognosis in triple-negative breast cancer machine-read til variables in TNBC," *Clin. Cancer Res.*, vol. 27, no. 20, pp. 5557–5565, 2021.
- [282] M. C. Balkenhol et al., "Optimized tumour infiltrating lymphocyte assessment for triple negative breast cancer prognostics," *Breast*, vol. 56, pp. 78–87, 2021.
- [283] L. Qiu et al., "Intratumor graph neural network recovers hidden prognostic value of multi-biomarker spatial heterogeneity," *Nature Commun.*, vol. 13, no. 1, pp. 1–12, 2022.
- [284] Y. Lee et al., "Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning," *Nature Biomed. Eng.*, pp. 1–15, 2022.
- [285] Y. Zuo et al., "Identify consistent imaging genomic biomarkers for characterizing the survival-associated interactions between tumor-infiltrating lymphocytes and tumors," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2022, pp. 222–231.
- [286] J. Witowski et al., "Improving breast cancer diagnostics with deep learning for MRI," *Sci. Transl. Med.*, vol. 14, no. 664, 2022, Art. no. eabo4802.
- [287] D. Arefan et al., "Deep learning modeling using normal mammograms for predicting breast cancer risk," *Med. Phys.*, vol. 47, no. 1, pp. 110–118, 2020.
- [288] A. Yala et al., "Toward robust mammography-based models for breast cancer risk," *Sci. Transl. Med.*, vol. 13, no. 578, 2021, Art. no. eaba4373.
- [289] A. Yala et al., "Multi-institutional validation of a mammography-based breast cancer risk model," *J. Clin. Oncol.*, vol. 40, no. 16, pp. 1732–1740, 2022.
- [290] A. Y. Ng et al., "Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer," *Nature Med.*, vol. 29, pp. 1–6, 2023.
- [291] H. Duanmu et al., "Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using deep learning with integrative imaging, molecular and demographic data," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2020, pp. 242–252.
- [292] E. Verburg et al., "Deep learning for automated triaging of 4581 breast MRI examinations from the dense trial," *Radiology*, vol. 302, no. 1, pp. 29–36, 2022.
- [293] A. Bhowmik et al., "Automated triage of screening breast MRI examinations in high-risk women using an ensemble deep learning model," *Invest. Radiol.*, vol. 58, pp. 10–1097, 2023.
- [294] Y. Wang et al., "Improved breast cancer histological grading using deep learning," *Ann. Oncol.*, vol. 33, no. 1, pp. 89–98, 2022.
- [295] N. Wahab et al., "AI-enabled routine H&E image based prognostic marker for early-stage luminal breast cancer," *npj Precis. Oncol.*, vol. 7, no. 1, pp. 1–13, 2023.
- [296] M. Amgad et al., "A population-level digital histologic biomarker for enhanced prognosis of invasive breast cancer," *Nature Med.*, vol. 30, pp. 1–13, 2023.
- [297] R. K. Mondol et al., "HIST2RNA: An efficient deep learning architecture to predict gene expression from breast cancer histopathology images," *Cancers*, vol. 15, no. 9, 2023, Art. no. 2569.
- [298] A. D. Trister et al., "Will machine learning tip the balance in breast cancer screening?," *JAMA Oncol.*, vol. 3, no. 11, pp. 1463–1464, 2017.
- [299] M. D. Halling-Brown et al., "Optimam mammography image database: A large-scale resource of mammography images and clinical data," *Radiol.: Artif. Intell.*, vol. 3, no. 1, 2020, Art. no. e200103.
- [300] K. Ding et al., "A large-scale synthetic pathological dataset for deep learning-enabled segmentation of breast cancer," *Sci. Data*, vol. 10, no. 1, 2023, Art. no. 231.
- [301] Y. Xue et al., "Selective synthetic augmentation with histogram for improved histopathology image classification," *Med. Image Anal.*, vol. 67, 2021, Art. no. 101816.
- [302] Y. Zhang et al., "From whole slide imaging to microscopy: Deep microscopy adaptation network for histopathology cancer image classification," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2019, pp. 360–368.
- [303] L. Luo et al., "Rethinking annotation granularity for overcoming shortcuts in deep learning-based radiograph diagnosis: A multicenter study," *Radiol.: Artif. Intell.*, vol. 4, no. 5, 2022, Art. no. e210299.
- [304] G. Qi et al., "Stabilized medical image attacks," in *Proc. Int. Conf. Learn. Representations*, 2020.
- [305] R. K. Samala et al., "Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography," *Med. Phys.*, vol. 43, no. 12, pp. 6654–6666, 2016.
- [306] C. Jin et al., "Label-efficient deep learning in medical image analysis: Challenges and future directions," 2023, *arXiv:2303.12484*.
- [307] C. You et al., "Rethinking semi-supervised medical image segmentation: A variance-reduction perspective," in *Proc. Adv. Neural Inf. Process. Syst.*, 2023.
- [308] L. Luo et al., "Oxnet: Deep omni-supervised thoracic disease detection from chest x-rays," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2021, pp. 537–548.
- [309] Z. Chai et al., "Orf-Net: Deep omni-supervised rib fracture detection from chest ct scans," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2022, pp. 238–248.
- [310] W. Ji et al., "Learning calibrated medical image segmentation via multi-rater agreement modeling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 12341–12351.
- [311] Y. Wang et al., "Learning robust medical image segmentation from multi-source annotations," 2023, *arXiv:2304.00466v2*.

- [312] K. You et al., "Intra-class contrastive learning improves computer aided diagnosis of breast cancer in mammography," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, Springer, 2022, pp. 55–64.
- [313] Y. Schirris et al., "Deepsmile: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer," *Med. Image Anal.*, vol. 79, 2022, Art. no. 102464.
- [314] H.-Y. Zhou et al., "Generalized radiograph representation learning via cross-supervision between images and free-text radiology reports," *Nature Mach. Intell.*, vol. 4, no. 1, pp. 32–40, 2022.
- [315] Y. Xu and H. Chen, "Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 21241–21251.
- [316] F. Zhou and H. Chen, "Cross-modal translation and alignment for survival analysis," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 21485–21494.
- [317] N. Coudray et al., "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," *Nature Med.*, vol. 24, no. 10, pp. 1559–1567, 2018.
- [318] Y. Qiu et al., "Scratch each other's back: Incomplete multi-modal brain tumor segmentation via category aware group self-support learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 21317–21326.
- [319] C. G. Yedjou et al., "Health and racial disparity in breast cancer," *Breast Cancer Metastasis and Drug Resistance, Advances in Experimental Medicine and Biology*, A. Ahmad, Eds., Springer, Cham, vol. 1152, 2019.
- [320] L. Seyyed-Kalantari et al., "Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations," *Nature Med.*, vol. 27, no. 12, pp. 2176–2182, 2021.
- [321] K. C. Potnis et al., "Artificial intelligence in breast cancer screening: Evaluation of FDA device regulation and future recommendations," *JAMA Intern. Med.*, vol. 182, no. 12, pp. 1306–1312, 2022.
- [322] Z. Xu et al., "A survey of fairness in medical image analysis: Concepts, algorithms, evaluations, and challenges," 2022, *arXiv:2209.13177*.
- [323] R. Geirhos et al., "Shortcut learning in deep neural networks," *Nature Mach. Intell.*, vol. 2, no. 11, pp. 665–673, 2020.
- [324] L. Luo et al., "Pseudo bias-balanced learning for debiased chest x-ray classification," in *Proc. Med. Image Comput. Comput. Assist. Interv.*, 2022, pp. 621–631.
- [325] G. Zhao et al., "Diagnose like a radiologist: Hybrid neuro-probabilistic reasoning for attribute-based medical image diagnosis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7400–7416, Nov. 2022.
- [326] S. Yan et al., "Towards trustable skin cancer diagnosis via rewriting model's decision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 11568–11577.
- [327] K. Goutsouliak et al., "Towards personalized treatment for early stage her2-positive breast cancer," *Nature Rev. Clin. Oncol.*, vol. 17, no. 4, pp. 233–250, 2020.
- [328] F. Li et al., "Deep learning-based predictive biomarker of pathological complete response to neoadjuvant chemotherapy from histological images in breast cancer," *J. Transl. Med.*, vol. 19, no. 1, pp. 1–13, 2021.
- [329] D. A. Bluemke et al., "Assessing radiology research on artificial intelligence: A brief guide for authors, reviewers, and readers—from the radiology editorial board," *Radiology*, vol. 294, no. 3, pp. 487–489, 2020.
- [330] S. Taylor-Phillips et al., "Uk national screening committee's approach to reviewing evidence on artificial intelligence in breast cancer screening," *Lancet Digit. Health*, vol. 4, no. 7, pp. e558–e565, 2022.
- [331] L. Hadjiiski et al., "Aapm task group report 273: Recommendations on best practices for ai and machine learning for computer-aided diagnosis in medical imaging," *Med. Phys.*, vol. 50, no. 2, pp. e1–e24, 2023.
- [332] H. Xiang et al., "Deep learning-assisted diagnosis of breast lesions on us images: A multivendor, multicenter study," *Radiol.: Artif. Intell.*, vol. 5, no. 5, 2023, Art. no. e220185.
- [333] M. Aubreville et al., "A comprehensive multi-domain dataset for mitotic figure detection," *Sci. Data*, vol. 10, no. 1, 2023, Art. no. 484.