Deep Learning Radiomics in Breast Cancer with Different Modalities: Overview and Future

Article *in* Expert Systems with Applications · May 2020

DOI: 10.1016/j.eswa.2020.113501

CITATION
1

READS
282

4 authors:

Pang Ting
University of Malaya
5 PUBLICATIONS **12** CITATIONS
SEE PROFILE

Jeannie Hsiu Ding Wong
University of Malaya
79 PUBLICATIONS **334** CITATIONS
SEE PROFILE

Wei Lin Ng
University of Malaya
9 PUBLICATIONS **42** CITATIONS
SEE PROFILE

Chee Seng Chan
University of Malaya
152 PUBLICATIONS **2,437** CITATIONS
SEE PROFILE

Expert Systems with Applications 158 (2020) 113501

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Review

Deep learning radiomics in breast cancer with different modalities: Overview and future

Ting Pang [a,b], Jeannie Hsiu Ding Wong [c], Wei Lin Ng [c], Chee Seng Chan [a,*]

[a] Center of Image and Signal Processing, Faculty of Comp. Sci. and Info. Tech., University of Malaya, Kuala Lumpur 50603, Malaysia
[b] Center of Network and Information, Xinxiang Medical University, Xinxiang 453000, China
[c] Department of Biomedical Imaging, Faculty of Medicine, University of Malaya, Kuala Lumpur 50603, Malaysia

ARTICLE INFO

Article history:
Received 1 November 2019
Revised 1 March 2020
Accepted 29 April 2020
Available online 11 May 2020

Keywords:
Deep learning radiomics
Computer aided diagnosis
Breast cancer
Medical imaging modality

ABSTRACT

Recent improvements in deep learning radiomics (DLR) extracting high-level features form medical imaging could promote the performance of computer aided diagnosis (CAD) for cancer. Breast cancer is the most frequent cancer among women and prospective achievements have been reported by CAD systems based on deep learning methods for breast imaging. In this paper, we aim to provide a comprehensive overview of the recent research efforts on DLR in breast cancer with different modalities and propose the future directions in this field. First, we respectively summarize and analyze the dataset, architecture, application and evaluation on DLR for breast cancer with three main imaging modalities, i.e., ultrasound, mammography, magnetic resonance imaging. Especially, we provide a survey on deep learning architectures exploited in breast cancer, including discriminative architectures and generative architectures. Then, we propose some potential challenges along with future research directions as references to the clinical treatment management and decision making utilizing such breast cancer CAD systems.

1. Introduction

1.1. Different imaging modalities in breast cancer

Breast cancer is cancer that develops and spreads in the breast tissues. Female breast cancer makes up more than 99% of all cases of breast cancer (Institute, 2019). According to the World Health Organization (WHO) report, breast cancer is the most frequent cancer among women and the rates are increasing in nearly every region. Although some risks can be reduced through preventive strategies, breast cancer is diagnosed very late in low- and middle-income countries (Organization, 2020). Therefore, early detection and treatment of breast cancer is beneficial for improving the survival rate (Institute, 2019). Accurate and precise diagnosis in medical imaging examination is vital for assisting early detection and treatment of breast cancer (Lee et al., 2010). The commonly used types of medical imaging modalities for breast tissues involves ultrasound, mammogram and magnetic resonance imaging (MRI) (Berg et al., 2012) (seen in Fig. 1).

Mammography uses low-dose X-ray for breast examination and it is routinely exploited for breast cancer screening (Tang et al., 2009). With high sensitivity to calcification, mammographic examination is far better at detecting microcalcifications and clusters of calcifications, which are very important characterizations of breast cancer (Horsch et al., 2006). Ultrasound which uses non-ionizing radiation has been proven to be an important adjunct scanning tool to mammography for breast cancer (Corsetti et al., 2011; Drukker et al., 2004). Ultrasound is more sensitive in dense breasts and effective in distinguishing the cysts from solid masses, which are limited in mammography (Kelly et al., 2010). MRI is an emerging tool with the highest sensitivity among breast imaging techniques. MRI supports multi-planar scanning and 3D reconstruction, which can better display the size, shape, location of breast lesions (Honda et al., 2016). Besides, MRI is a valuable tool for screening on high-risk individuals, diagnosing occult cases, staging and assessing response to chemotherapy (Saslow et al., 2007).

Routinely, a radiologist analyse the medical images and produce a report based on a standardized reporting, such as Breast Imaging Reporting and Data System (BI-RADS) by the American College of Radiology (D'Orsi et al., 2018). However, the shortage of qualified radiologists and increasing number of images to review may lead the drop of mental focus of radiologists, further leading to the

* Corresponding author.
E-mail addresses: wva180024@siswa.um.edu.my (T. Pang), jeannie.wong@ummc.edu.my (J.H.D. Wong), wei.lin@ummc.edu.my (W.L. Ng), cs.chan@um.edu.my (C.S. Chan).

https://doi.org/10.1016/j.eswa.2020.113501
0957-4174/

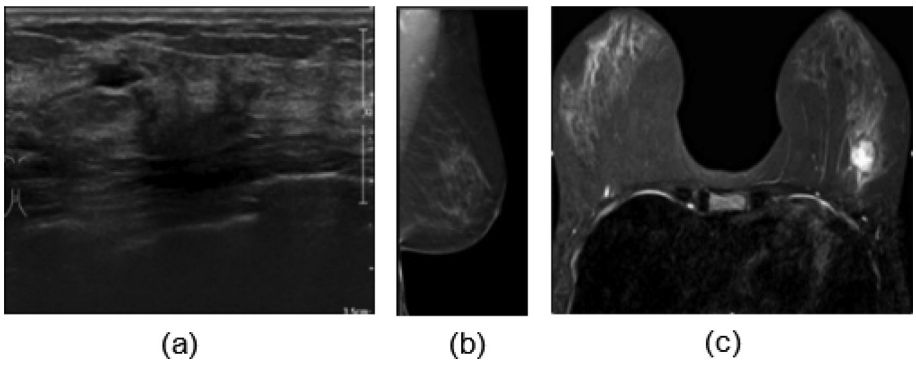

Fig. 1. The examples of medical imaging modalities for breast tissues: (a) ultrasound, (b) mammogram, and (c) MRI.

possible drop in diagnostic accuracy and efficiency for breast cancer (Horsch et al., 2006). The gold standard of diagnosis of breast cancer is histopathological examination via a biopsy, which is invasive and requires additional costs. Fortunately, now with the help of computer-aided diagnosis (CAD) system, it is possible to improve the efficiency for biomedical imaging diagnosis and provide precise treatment for patients. CAD for breast cancer can be used to assist radiologists to diagnose (such as detect and locate breast cancer, classify it as benign or malignant, predict cancer risk), tumor prognosis and treatment selection.

1.2. Deep learning radiomics

Radiomics is a recent progress and promotes medical imaging have a natural extension to quantitative imaging, which includes large amounts of information, for better CAD of cancer (Lambin et al., 2012). Unlike the previous CAD systems that directly proposed a single diagnostic result, radiomics centres on the process of extracting large amounts of quantitative features (high-dimensional data/information) from radiographic images (Gillies et al., 2015). The features can reflect the underlying pathophysiology of medical images. In this regard, mining and making full use of these features is effective for cancer diagnosis, prognosis, genomics (radiogenomics (Gillies et al., 2015)), and etc. Previously, radiomics has been developed widely via hand-crafted features method. Hand-crafted method mainly extract and select features (such as texture, shape, wavelet, geometric) from the regions of interest (ROI) after pre-processing with manual segmentation, which is done by the radiologists (Yassin et al., 2018) (seen in Fig. 2). For receiving the final clinical results, it additionally uses conventional machine learning methods, such as k-means, support vector machines (SVM), random forest, naive Bayes, adaboost (Suzuki, 2012; Rouhi et al., 2015). Although hand-crafted features-based radiomics has shown to be useful in cancer domains (Liu et al., 2019; Beura et al., 2015; Chen et al., 2017), the whole procedure using traditional methods is tedious and requires long processing time.

To date, deep learning has become the most popular technology for computer vision (CV) tasks (Lynch and Liston, 2018). Deep learning, consisting of multiple simple but nonlinear units (e.g. convolution, pooling, activation, dropout) to deeply abstract and learn features from the images, is an end-to-end processing method (LeCun et al., 2015). Inspired by the wide application of deep learning in identification of natural images, researchers in the medical domain also started to explore the applications of deep learning (Greenspan et al., 2016; Ting et al., 2019; Litjens et al., 2017). Naturally, deep learning radiomics (DLR) for cancer area is an emerging field (Afshar et al., 2019; Chaunzwa et al., 2018; Esteva et al., 2017; Dhungel et al., 2017). The front layers of DLR try to extract low-level features (e.g., edges), and the middle and subsequent layers focus on middle-level and high-level features (e.g., parts and objects) (Jiao et al., 2016), which are directly used to cancer domain obtaining immediate results (seen in Fig. 3). Therefore, compared to hand-crafted radiomics, DLR does not need

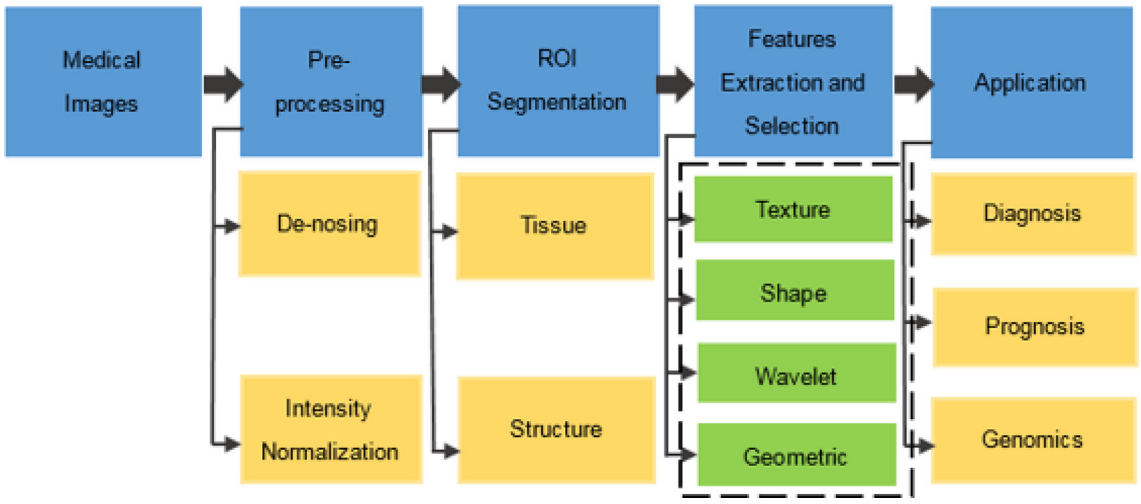

Fig. 2. The procedure of hand-crafted radiomics. Radiomics features are denoted in the dotted line box.

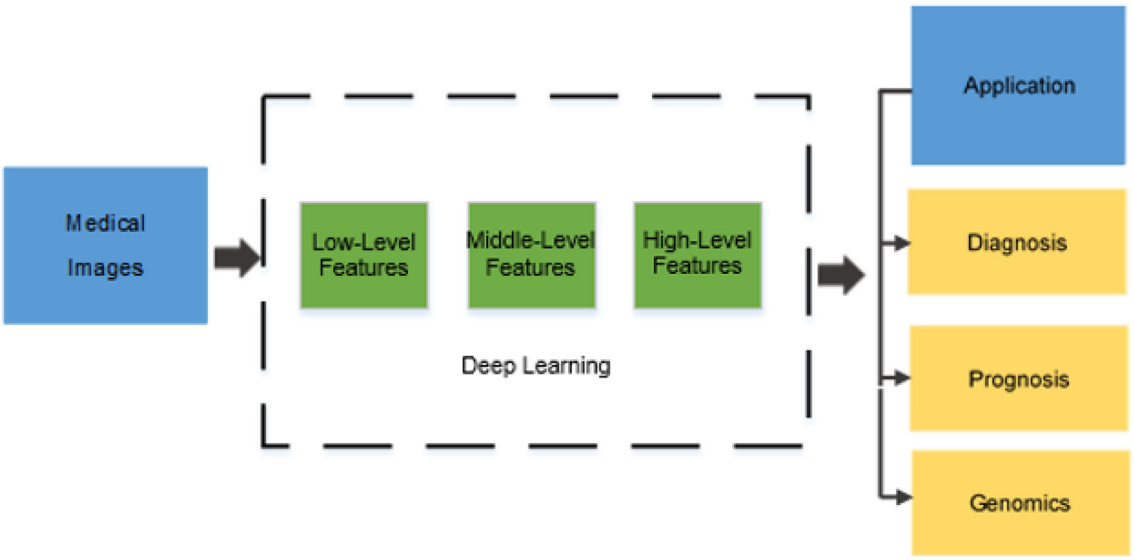

Fig. 3. The procedure of deep learning radiomics. Radiomics features are denoted in the dotted line box.

design the complex hand-crafted features, making it less labour intensive and efficient. Basically, architecture based on deep learning is the core of DLR. However, complicated deep learning modules, which are difficult to be interpreted, often appear as black boxes to users.

1.3. Contributions of the paper

The main purpose of this manuscript is to present a systematic review of studies on deep learning-based radiomics for breast cancer in different imaging modalities. Through this survey, it can build a panorama as guideline with which readers quickly grasp the field of DLR in breast cancer. Besides, this survey provides suggestions on the clinical treatment management for breast cancer exploiting deep learning based CAD systems. This survey lays the foundations of innovation to increase the richness of this field.

To summarize, the contributions of this work are in three-folds: 1) Organize and explain the current works in detail to make the deep learning based radiomics interpretable. 2) Focus on clinical value of deep learning based CAD systems to provide suggestions for clinical decision making and reduce the workload of radiologists. 3) Provide comprehensive references and some new trends to researchers in this field.

The structure of this paper is: In Section 2, we summarize and analyse the present DLR applied in breast cancer from the aspects of dataset, architecture, application and evaluation via the retrieved studies. The focus is on the deep learning frameworks developed in three breast imaging modalities, i.e., ultrasound, mammography, MRI. In Section 3, potential challenges and future directions are discussed to provide references for the further studies in this area. Finally, in Section 4, short conclusions are given.

2. Overview and analysis

When selecting the papers, studies were retrieved by searching the databases (Web of Science and Google Scholar) concerning deep learning radiomics and breast cancer since 2015 to June 2019 (the last search). This work mainly focuses on the 188 technical papers involving in three main imaging modalities, i.e. mammography, ultrasound and MRI. It can be seen from Fig. 4 that studies on DLR for breast cancer significantly double increased, as the deep learning widespread with a continuous rising trend.

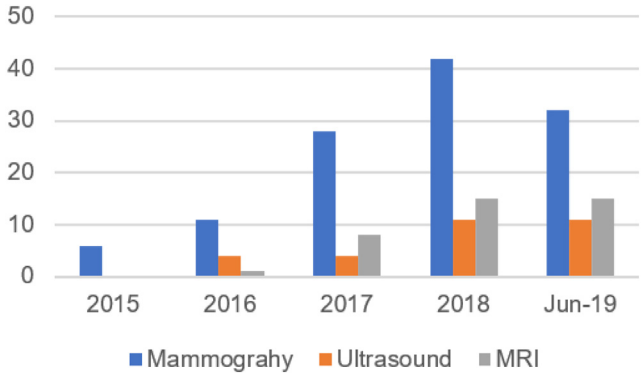

Fig. 4. The number of papers for DLR on three breast modalities from 2015 to June 2019 (the last search).

Besides, mammography accounts for the largest portion, as public available databases are almost mammogram images. After a deep investigation, we will illustrate them from datasets, architectures based on deep learning, goals for practical application and evaluations.

2.1. Dataset

Deep learning requires large amount of data to train the model with superior performance, thus DLR for breast cancer also demands a lot of breast cancer radiological images. The public dataset and private dataset utilized in retrieved studies are illustrated as below.

2.1.1. Public dataset

The current public datasets greatly promote development on DLR for breast cancer. Most commonly used databases predominantly comprise of mammographic images from the United States and Europe (Hamidinekoo et al., 2018). Such as Digital Database for Screening Mammography (DDSM) (Heath et al., 2001), Mammographic Image Analysis Society (MIAS) (Suckling et al., 2015), fBancoWeb (Bruno and Homero, 2011), Wisconsin Breast Cancer Dataset (WBCD) (William et al., 1995), INbreast (Moreira et al., 2012), Breast Cancer Digital Repository (BCDR) (Lopez et al.,

2012). About 60% of the DLR work used mammography from these public datasets, where DDSM takes up the biggest fraction due to its large numbers and comprehensive annotation. On the other hand, there is a lack of public database providing ultrasound and MRI images for breast cancer DLR studies. The significance of DLR using ultrasound and MRI images, and the total differences of cancer or tissues presented on the three imaging modalities (seen in Fig. 5) indicate the urgent requirement of public datasets for ultrasound and MRI breast cancer imaging.

The incidence of breast cancer has increased over the years, and the cancer patterns and imaging technology has also changed rapidly, which increases the complexity of the problem. The image database, which was available for training of DLR, may not be representative to the images produced by new digital mammography, ultrasound and MRI scanners. This leads to the poor reproducibility or application of the pre-trained DLR models on real clinical cases. Chen et al. (2018) pre-trained a deep learning model on DDSM dataset and transferred to INbreast dataset. It respectively achieved validation of area under the curve (AUC) of receiver operation characteristics (ROC) as 0.82 (DDSM), 0.93 (INbreast), but only 0.70 on the in-house collected dataset. It reveals the present available mammography datasets may be out of the date and not fully represent the current clinical imaging features.

Another issue, which may impinge on the application of DLR models trained on public dataset, is that there may be inherent bias between the image data used for training and those used for validation. For example, Asian women tends to have smaller and denser breast compared to Caucasian women. However, most publicly available datasets are from Caucasian or black women. Hence, it would not be surprising if the DLR models are not able to achieve high accuracy for a significantly different women population. In addition, breast cancer has distinctive features at every age stage (Kelly et al., 2010). To provide the correct training dataset for clinically relevant images, establishing the breast cancer databases of different regions at different ages, which sufficiently contain pathological heterogeneity and coexisting benign-malignance and reflect the current clinical situation, is essential.

2.1.2. Private dataset

Collecting private medical images are difficult due to patient confidentiality and data privacy concerns, not to mention the laborious effort in properly indexing, storing, and annotating the images. Image attributes including cropped image size, format, data source, number of samples for training and testing affect the final results. Han et al. (2017) indicated breast lesions in their ultrasound imaging dataset with the margin size (distance between lesion and edge of ROI) as 180 pixel has the best classification performance. Kooi et al. (2017), Chiang et al. (2018), Gallego-Ortiz and Martel (2019) and Zhang et al. (2018) each provided detail dataset production with the strict criterion on mammography, ultrasound and MRI. Imaging annotation (e.g., image segmentation and data classification), which is primarily carried out by radiologists with the professional knowledge, is time-consuming, but it is nonetheless a crucial part of the process. Scholars are exploring to utilize the unsupervised deep learning methods (seen in Section 2.2) to reduce this labour. Data augmentation (e.g. spatial translating, rotating and vertical and horizontal flipping), which increases the training dataset, has been explored to reduce possible over-fitting of DLR models (Li et al., 2017). Deep learning based on generative adversarial network (seen in Section 2.2) has also recently gained popularity as one of the augmentation methods.

2.2. Architecture based on deep learning

Deep learning is commonly divided into supervised learning and unsupervised learning. Supervised deep learning is a process whereby knowledge is gained from labeled dataset. The most famous and basic model is the convolutional neural networks (CNN) consisting of multi convolutional layers (a set of 2D

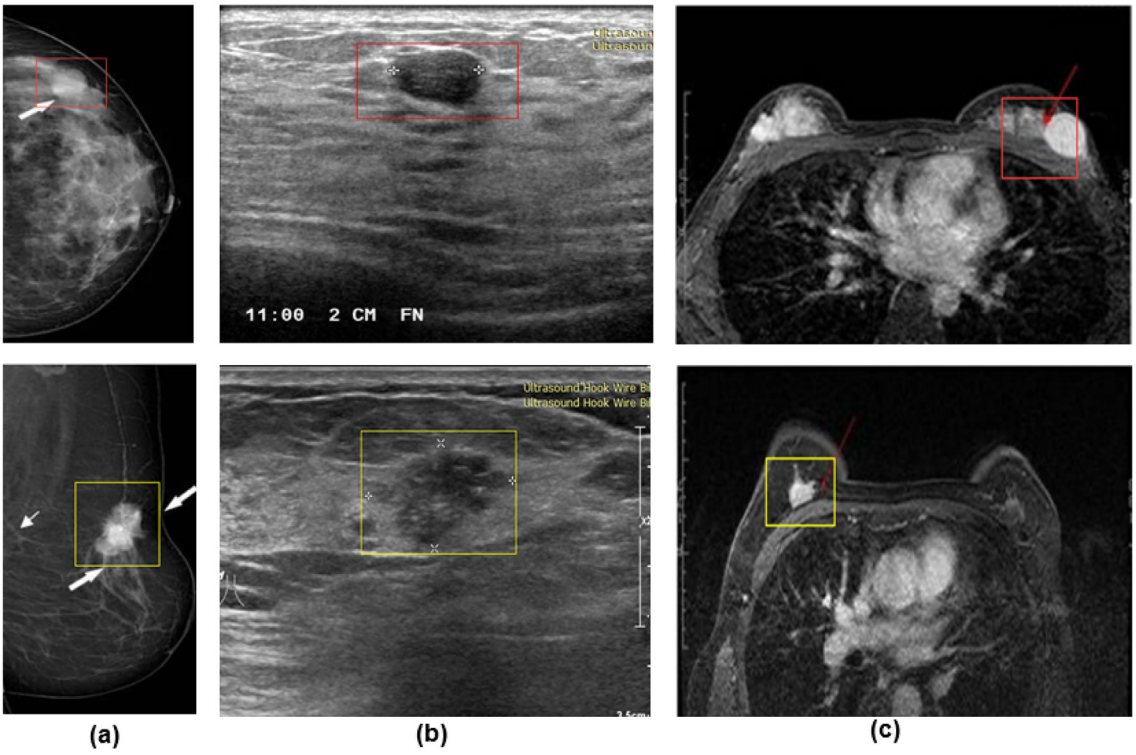

Fig. 5. The samples of benign (top row) and malignant (bottom row) breast masses on mammography (a), ultrasound (b), and MRI (c).

rectangular filters), pooling layers (reduce the spatial size), fully connected layers and activation functions (e.g. ReLU (rectified linear unit), tanh, softmax, sigmoid) (Krizhevsky et al., 2012; Gupta, 2017). There are some classical CNN models wining top computer vision challenges, e.g. AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisserman, 2014), GoogleNet (Ballester and Araujo, 2016), ResNet (He et al., 2016). Another important supervised learning model is recurrent neural networks (RNN). RNN is specifically used to process sequential data as it can remember former computations with loops and memories (Williams and Zipser, 1989). Therefore, it has been reported to be particularly useful in processing 3D volumetric images, such as MRI image slices. Unsupervised deep learning obtains knowledge form unlabeled dataset, including typical auto-encoder networks (AE), generative adversarial networks (GAN) and deep belief networks (DBN). AE has an encoder automatically transferring each of the images into a latent space, and a decoder that reconstructs the images using the learned features to grasp the core information of raw features (Ng et al., 2011). Inspired by game theory works, GAN learn the rich features and high-dimensional distributions of data in the model, then generate artificial image for data augmentation via a pair of networks in competition (Goodfellow et al., 2014; Creswell et al., 2018; Douzas and Bacao, 2018). DBN comprises of a stack of restricted Boltzmann machine (RBM) (Salakhutdinov and Hinton, 2009), seen as a visible layer, and many hidden layers (top two hidden layers contain random relationship), between which the generative probability union distribution is modelled to extract the deep features of visible layer (Hinton, 2009).

In the nomenclature of deep learning, supervised learning is also classified as discriminative architecture, while the unsupervised learning is classified as generative architecture. The discriminative DLR is directly applied to cancer through networks which extracts features from medical images, while the generative DLR generates new data by learning the features from data distributions for the cancer tasks. Currently, there are two different types of machine learning methods, reinforcement learning (RL) and transfer learning (TL), which are often used in deep learning. RL is formed by an agent that learns an optimal policy for better decision by receiving rewards from the environment at one state and it is concerned with sequential decision problems (Kaelbling et al., 1996). Deep learning networks based on RL enable the agent to learn the deep features from the raw images (Li, 2017). Maicas et al., 2017 developed a deep RL breast lesion detection framework. To speed up detection time, an agent learns policy via translation and scale, and decides the next moving action based on deep feature of the current bounding box. As another important method in machine learning, TL migrates the pre-trained deep learning model to the new model by fine-tune (Shin et al., 2016). To overcome the limitations of small number of medical images, TL is largely applied to DLR to extract features learned from the large labeled natural data set (e.g., ImageNet (Krizhevsky et al., 2012)) into the breast cancer imaging domain (Chen et al., 2018; Samala et al., 2017; Huynh et al., 2016; Huynh et al., 2016; Hadad et al., 2017). The architectures in DLR are seen as Fig. 6.

Fig. 7 shows the architectures of DLR based on different deep learning methods in the three imaging modalities for breast cancer. It can be seen that CNN is the most commonly used comparing with the rest of the algorithms. Next, we describe in detail the architectures utilized in DLR for breast cancer in mammography, ultrasound, MRI.

2.2.1. Mammography

The first implementation of DLR in mammography for breast cancer was done by Arevalo et al., 2015. They applied CNN architecture with two convolutional layers and two pooling layers, one fully connected layer to discover good features for breast mass classification. Compared to the hand crafted radiomics method, CNN architecture showed an increasement from 79.9% to 86% in terms of AUC. The architecture in the initial development of deep learning was basic and simple. However, with the wide spread interest and use of deep learning, more scholars designed different architectures with deeper CNN in radiomics of mammographic imaging for improving diagnosis of breast cancer (Kooi et al., 2017; Wang et al., 2018; Qiu et al., 2017; Mordang et al., 2016). A typical DLR architecture based on CNN for breast cancer classification is seen in Fig. 8. Except for the self-designed CNN, Ertosun and Rubin (2015) tried three well known CNN models on breast mass classification task, the validation accuracy results show that GoogleNet (85%) was higher than AlexNet (84%) and VGG (82%). Though VGG net features outperform those of GoogleNet in other multiple tasks (Tsochatzidis et al., 2019), it needs much more parameters (140 M) to evaluate. The recent popular standard CNN model is ResNet, which has very deep layers and skip-connection scheme to prevent gradients vanishing. Fathy and Ghoneim (2019) directly tested ResNet on DDSM database for classifying breast cancer and reported performance, which had a great improvement than previous classical CNN. Al-masni et al. (2018) proposed a DLR architecture based on regional deep learning technique, i.e. ROI-based CNN, called You Only Look Once (YOLO) (Redmon et al., 2016) for breast mass detection and classification. YOLO is fast and simple, and it unifies the features of entire mammograms into a single neural network to predict each bounding box so that if can detect and classify breast cancer at the same time. Gastounioti et al. (2018) combined hand-crafted methods with CNN into a hybrid framework to learn the meta-features, which are used to calculate the breast parenchymal complexity for breast cancer risk prediction. The final results show that informative interactions between patterns exist in the deep feature maps and hand-crafted maps, so that the association is better than single method to breast cancer diagnosis. Likewise, Alkhaleefah and Wu (2018) presented hybridization between CNN and SVM classifiers with the more robust breast cancer classification results than single CNN. Hence, deep learning model combination with traditional methods is emerging as a promising DLR architecture.

For deep AE method, Taghanaki et al. (2017) first adopted it for breast cancer classification in mammography. They establish a novel AE model and each layer uses sigmoid function apart from the last layer of the encoder using a linear transform function. Novel optimization methods, two-objectives (mean squared reconstruction error (Xu, 1993) and mean classification error (Smyth, 1996)), are introduced to the model to extract more useful features for improving the performance of breast cancer classification. Kallenberg et al. (2016) set an architecture for mammographic density scoring, consisting of a sparse AE (Ranzato and Szummer, 2008) within a convolutional framework to deeply learn abstract features in an unsupervised way, called convolutional sparse autoencoder (CSAE). A novel sparsity regularization method is presented to control the capacity of models. AE with CNN is able to extract better hierarchical features of spatial information.The schematic DLR architecture based on AE for breast cancer classification is shown in Fig. 9.

Due to data protection issues and complex annotated labels by the radiologists, collecting a large number of medical images for the purpose of training deep learning models is challenging (Zhang et al., 2017). Guan and Loew (2019) used GANs to generate samples for data augmentation, which are combined with original data into CNN for breast lesion classification training task. Moreover, the added data can reduce overfitting. Wu et al. (2018) also trained a class-conditional GANs to learn the high-level features by synthesizing high-resolution mammogram patches for data augmentation. The experiments on a ResNet classifier showed that GAN-augmented artificial mammographic images produced higher

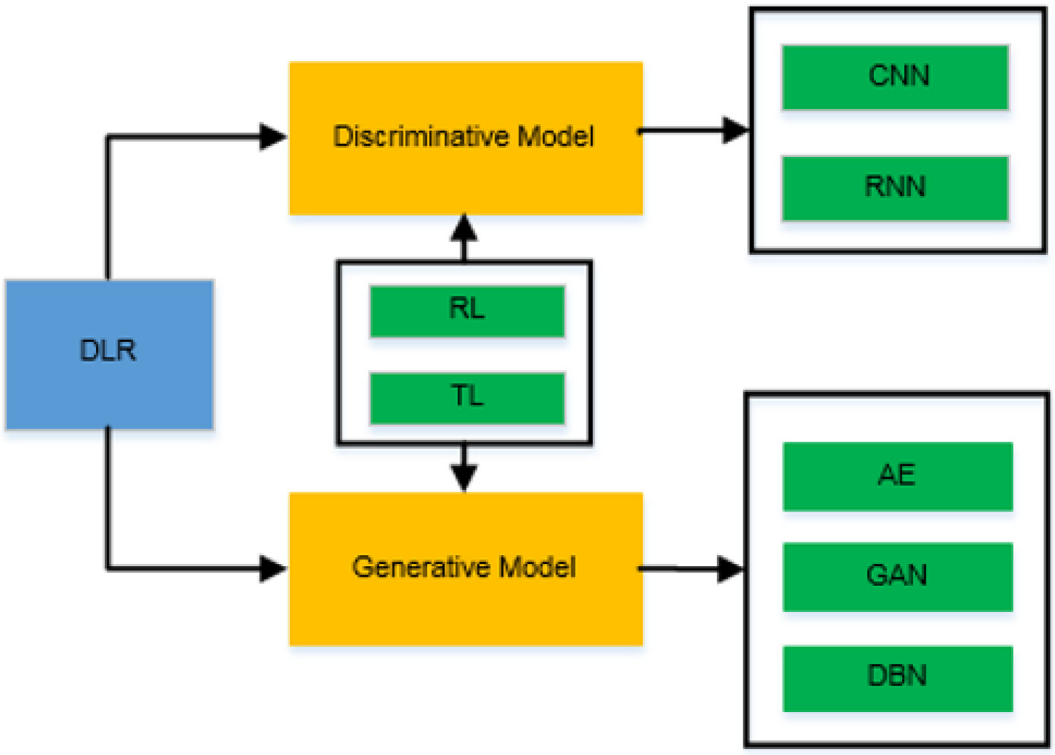

Fig. 6. The architectures in DLR.

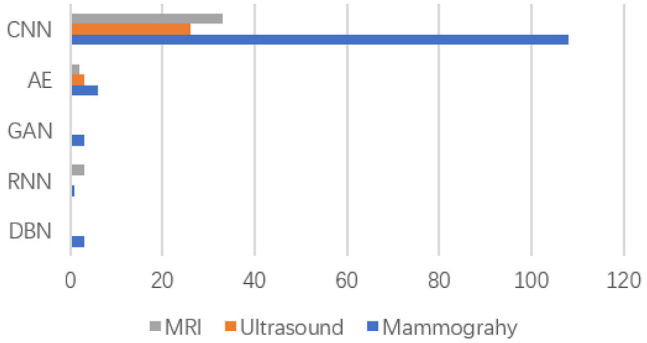

Fig. 7. The number of papers for architectures based on different deep learning methods for breast cancer.

AUC through comparing with the data augmentation using the affine transformation. Hence, DLR utilizing GAN becomes an emerging method for data augmentation to improve breast cancer diagnosis. The model of GAN generating breast cancer samples are shown in Fig. 10.

DLR with DBN is seen as an important direction in cancer area. Dhungel et al. (2015) proposed a deep learning architecture using DBN for detection of masses from mammograms. It starts with a multi-scale DBN and Gaussian mixture model (GMM) classifier, which is used to select candidate regions and then it uses a cascade R-CNN classifier with a cascade random forest classifier, followed by a post-processing based on connected component analysis (CCA) to detect the final regions. Although it achieves higher true positive rate, the whole architecture is too complicated to perform. Then Abdel-Zaher and Eldeib (2016) designed a scheme using unsupervised phase of DBN that learns features followed by supervised path of backpropagation neural network (BPNN) (Hecht-Nielsen, 1992) to achieve higher breast cancer classification accuracy. The reason for improvement is that DBN phase initializes BPNN to search objective function near a good-local optimum. The architecture only based DBN is shown in Fig. 11.

2.2.2. Ultrasound

Xiao et al., 2018 compared three standard CNN methods, inception V3 (Szegedy et al., 2016), ResNet50 and Xception (Chollet, 2017) transferring on same breast ultrasound dataset for lesion classification. Inception V3 was shown to be able to manage

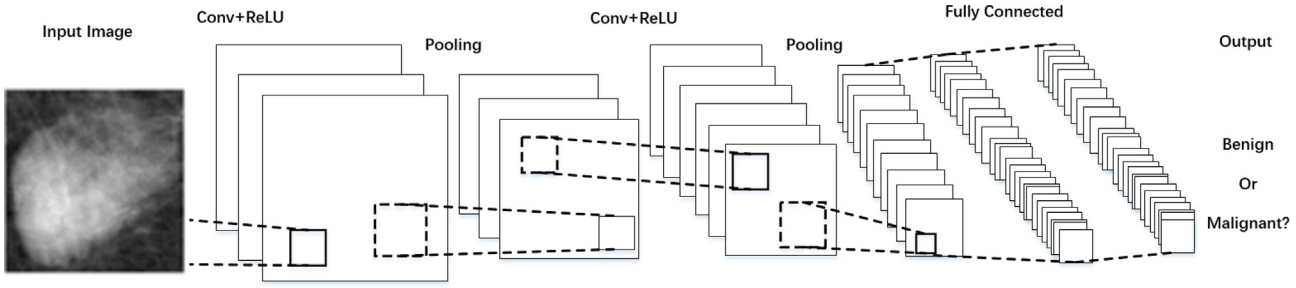

Fig. 8. A typical DLR architecture based on CNN for breast cancer classification. Conv: Convolution.

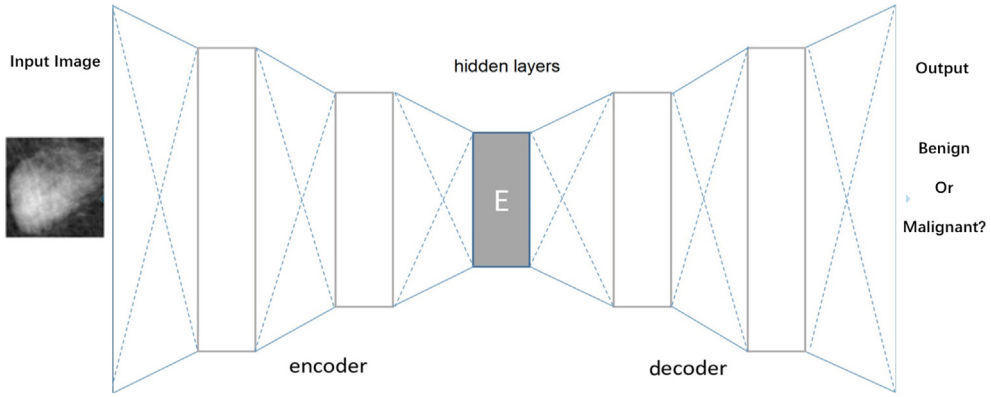

Fig. 9. The schematic DLR architecture based on AE for breast cancer classification.

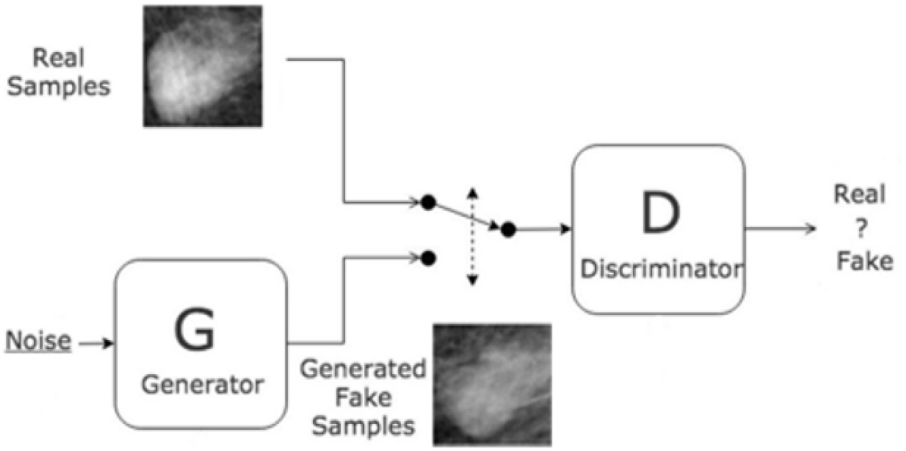

Fig. 10. The model of GAN generating breast cancer samples.

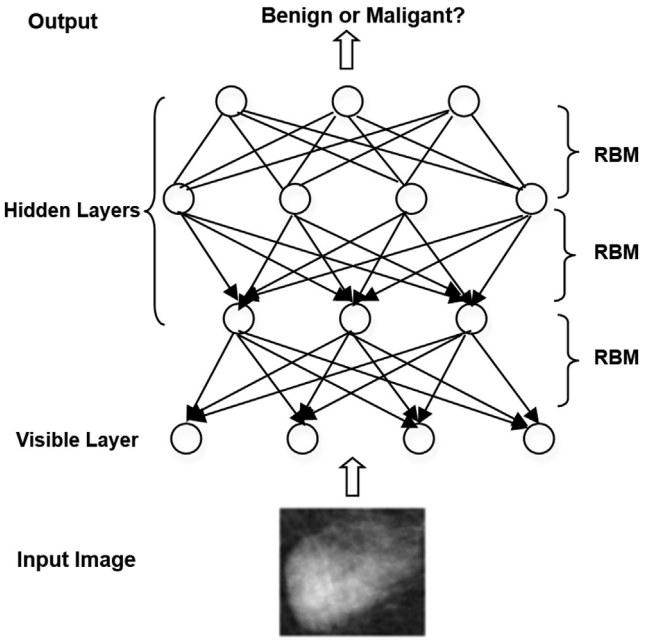

Fig. 11. The DLR architecture based on DBN for breast cancer classification.

memory more efficiently and performed the best. As other designed CNN architectures for classification of breast lesions in ultrasound are similar to mammogram, we will focus on the automatic detection of tumors in ultrasound imaging. Inspired by the famous detection network Fast R-CNN (Girshick, 2015); Li et al., 2018 added an spatially constrained layer to make the detection output not only rely on the input label but also on the topological residing regions, which are neglected by annotations. Fast R-CNN consists of a detector and region proposal network (RPN), and the basic framework for breast tumor localization on ultrasound is shown in Fig. 12. Shin et al. (2018) proposed a hybrid semi-supervised deep learning network also based on Fast R-CNN (used a strongly annotated dataset) and multiple-instance learning instance (Maron and Lozano-Pérez, 1998) (used a weekly annotated dataset). Besides, Chiang et al. (2018) designed a detection system based on 3-D CNN for 2-D slices (3-D volume) combined with a new scheme called prioritized candidate aggregation. The 3-D CNN with 3-D kernel size and stride was used to select the tumor candidates with higher estimated probability. Then the prioritized candidate aggregation scheme was designed to alleviate the overlapped candidates based on estimated tumor probability.

For deep AE method in breast ultrasound, Cheng et al., 2016 and Lee et al., 2018 both set a stacked denoising autoencoder (SDAE) (Vincent et al., 2008) deep learning framework, which is equipped with the mechanism that automatic features exploration and noise avoidance to support the discrimination of breast ultrasound lesions. SDAE is suitable for small-sized dataset and has robust distinguish as benign or malignant breast lesions. But Lee et al. (2018) added intensity inhomogeneity correction algorithm and significantly improved the performance. Integrating deep learning network with other methods is an emerging DLR architecture in breast cancer (Truhn et al., 2018).

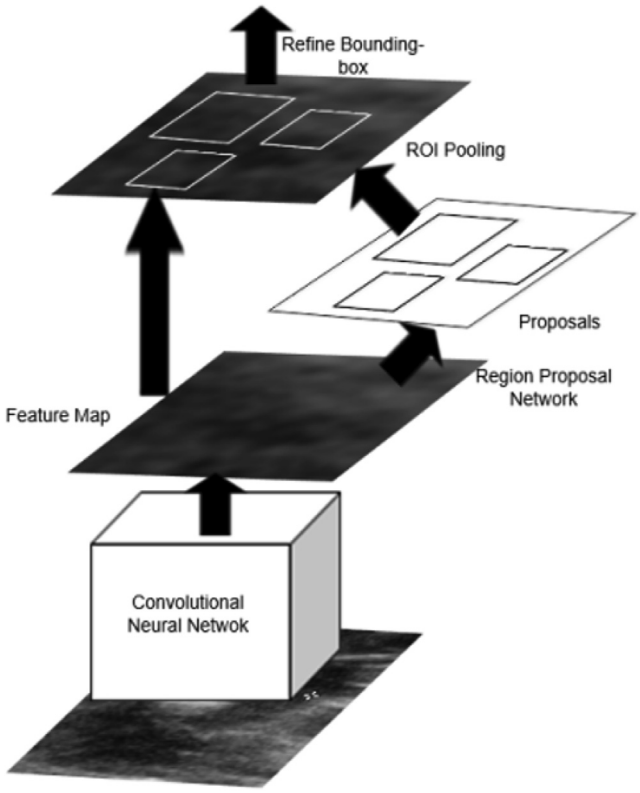

Fig. 12. The DLR architecture based on Fast R-CNN for breast cancer detection in ultrasound.

2.2.3. Magnetic resonance imaging

Deep learning using CNN in MRI imaging with breast cancer is also similar with previous mammography and ultrasound for 2D images (Hiramatsu et al., 2017; Gallego-Ortiz and Martel, 2019; Dalmış et al., 2018; Ha et al., 2019), but radiologists always judge on MRI with temporal sequences of 3D images. Hence, a few papers reported utilization of RNN, which is able to remember the previous sequence and deal with the extracted features through a convolutional neural network from MRI sequences for breast cancer. Long-short-term-memory (LSTM) (Hochreiter and Schmidhuber, 1997) including input, forgetting and output gate can learn the information long time ago and it is an advanced method of RNN for better DLR. Antropova et al. (2018) performed LSTM on sequences of features, which are extracted by VGG from breast MRI for lesions classification, and yielded an AUC 0.85 compared with single-time (0.81). The basic DLR architecture based on RNN for MRI is shown in Fig. 13. For better performance, Zheng et al. (2018) added a dense block in the CNN-RNN unit combining with semantic constrains from diagnostic report to balance the huge parameters. Based on the effectiveness of RNN used in MRI sequences for classification, we can attempt to apply it for other breast cancer applications of MRI.

2.3. Clinical application

Different imaging modalities are able to image specific attributes of breast lesions. Application of DLR for breast cancer detection and classification has been explored in various aspects including diagnosis, prognosis, therapy response, genomics and etc., seen in Fig. 14. DLR is most commonly applied in the classification of breast mass in all three modalities of imaging.

2.3.1. Mammography

In this few years, there are vast development of DLR in mammography, and most implementations are in the field of detection or classification of breast masses (Chen et al., 2018; Kooi et al., 2017; Wang et al., 2018; Dhungel et al., 2015; Li et al., 2018). Microcalcification, which is more sensitive in mammogram, is much smaller and low contrast comparing with the background of images (Tang et al., 2009). It means that deep learning for detection and classification of microcalcification is necessary, but less papers (Mordang et al., 2016; Bekker et al., 2016; Wang et al., 2016) are about this area, as annotation for microcalcification is a tougher work. Breast density is validated as an important risk factor for breast cancer and mammography has high accuracy for cancer detection of women with dense breasts (Freer, 2015). DLR has proved its effectiveness in the discrimination of breast density for clinical predictive risk of breast cancer (Gastounioti et al., 2018; Kallenberg et al., 2016; Lehman et al., 2018).

2.3.2. Ultrasound

As an adjunct to mammography, ultrasound imaging has high accuracy of breast masses with easier annotations for masses (Corsetti et al., 2011). All the DLR studies retrieved in breast ultrasound try to learn deep features to detect and classify the breast masses (Shin et al., 2018; Hiramatsu et al., 2017; Zhou et al., 2018; Byra et al., 2019; Huang et al., 2019).

2.3.3. Magnetic resonance imaging

Similar to mammography and ultrasound, most implementations for DLR in MRI focus on the detection and classification for breast masses (Dalmış et al., 2018; Antropova et al., 2018; Rasti et al., 2017; Antropova et al., 2017). Neoadjuvant chemotherapy (NAC) is used in breast cancer treatment before surgery to improve the outcome. MRI appears to provide the best prediction of the tumor response post NAC (Yeh et al., 2005). Some scholars have proved that the effectiveness of DLR for discrimination of NAC response in MRI imaging to assistant the radiologists (Ha et al., 2019; Ha et al., 2019; Adoui et al., 2019). Molecular subtype has been proved that it can help to choose patients for preoperative breast MRI imaging, hence classifying molecular subtype in MRI is necessary (Grimm et al., 2014). Scholars in recent works have tested and verified the feasibility of DLR in classification of molecular subtype of breast tumor with MRI (Ha et al., 2019; Zhu et al., 2018).

2.4. Evaluation

The effectiveness and accuracy of DLR models needs to be evaluated via clinically relevant decision-making tools. In this section, we present the technical details needed to train DLR models and the measurement index commonly used for evaluation of DLR models.

2.4.1. Technical detail

We first need use workstation with graph processing unit (GPU) to train the deep learning model on breast cancer imaging datasets. Usually, DLR is implemented by programming language (Python, Matlab, Java, etc.) on different frameworks (Theano, Tensorflow, Pytorch, Caffe, etc.), which contain deep learning packages. Before training, we should initialize the batch size (each training takes batchsize samples in the training set), iteration (one iteration is equal to training once with batchsize samples), epoch (one epoch is equal to one training with all samples in the training set), learning rate (it determines whether and when the objective function converges to the local minimum), etc and then test or validate the models on small dataset.

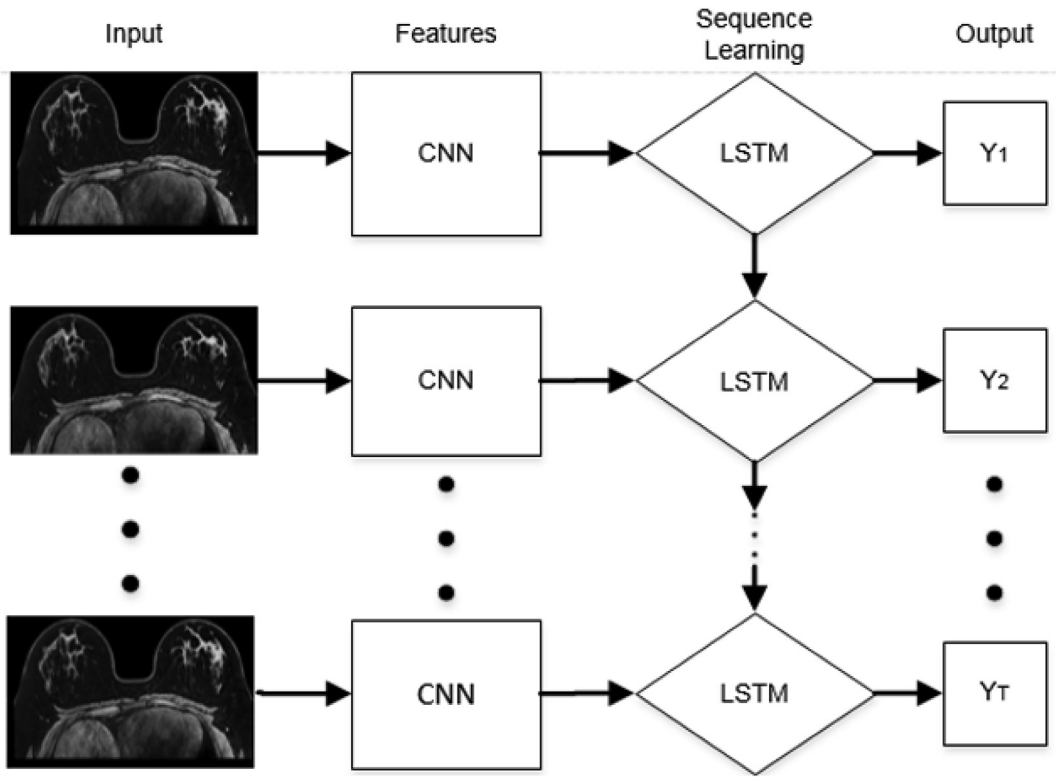

Fig. 13. The DLR architecture based on CNN-RNN for breast cancer classification in MRI.

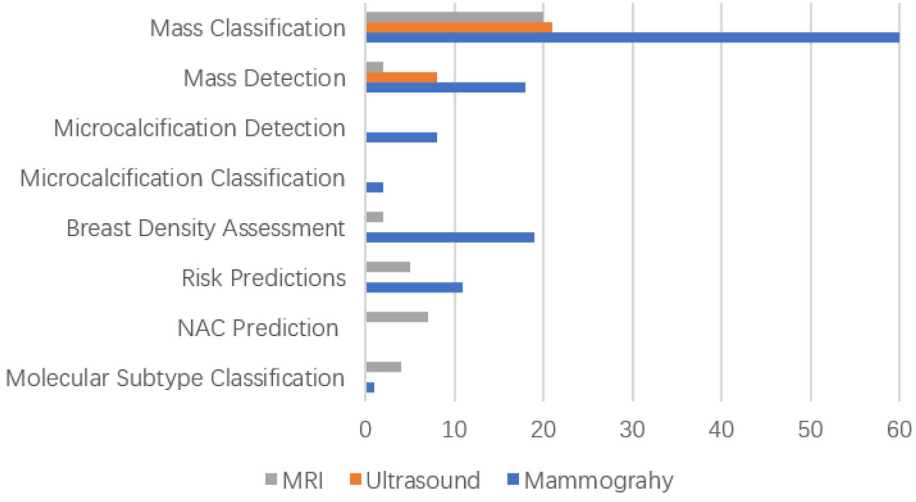

Fig. 14. The number of papers for DLR used in different tasks. (NAC: Neoadjuvant chemotherapy).

2.4.2. Measurement index

The common measurement indicators applied to evaluate the performance of DLR in breast cancer are Accuracy (accuracy of the testing outcomes), Specificity (true negative rate), Sensitivity (true positive rate) and Area Under Curve (AUC) of Receiver Operating Characteristic (ROC), as illustrated in the following formulas (Litjens et al., 2017). Besides, baseline methods are always tested on the datasets to further evaluate the deep learning architecture.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Specificity = \frac{TN}{FP + TN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

where TP denotes true positive (the number of correct malignancy), FP denotes false positive (the number of incorrect malignancy), FN denotes false negative (the number of incorrect benign), TN denotes true negative (the number of correct benign). AUC is the area under the ROC, which is one curve with (1-specificity) for x-axis and sensitivity for y-axis.

DLR for breast cancer has received great progress under academics' hard working. The detailed implementation information of some representative papers covered in this work with three image modalities is listed in the end of Section 2, respectively as Table 1 (mammography), Table 2 (ultrasound) and Table 3 (MRI).

Table 1
Studies conducted for breast cancer in mammography based on DLR.*Conv: Convolution; *Pool: Pooling; *Fc: Fully Connected; *AUC: Area Under the Curve (ROC); *Acc: Accuracy; *Sn: Sensitivity; *Sp: Specificity.

References	Data	Input Size (Pixels)	Application	Architecture	Outcome
(Arevalo et al., 2015)	Benign: 426 Malig- nant: 310 (BCDR)	150 × 150	Mass Classification	CNN: Conv2 + Pool2 + Fc	AUC: 0.86
(Ertosun and Rubin, 2015)	Total: 2420 (DDSM)	256 × 256	Mass Detection	AlexNet, VGG, GoogLeNet	Acc: 0.84, 0.82, 0.85
(Dhungel et al., 2015)	Total: 410 (DDSM, INbreast)	264 × 264	Mass Detection	Cascade CNN + Random Forest	Sp: INbreast 0.96 ± 0.03, DDSM 0.75
(Ertosun and Rubin, 2015)	Benign: 372 Malig- nant: 333 (private)	–	Macrocalcification Classification	Deep Neural Networks	Acc: 0.787, AUC: 0.929
(Abdel-Zaher and Eldeib, 2016)	Total: 683 (WBCD)	–	Mass Classification	DBN	Acc: 0.9968
(Mordang et al., 2016)	Positive: 17009 Negative: 45 million (private)	13 × 13	Macrocalcification Detection	CNN: Conv4 + Pool2 + Fc3	Sn: 0.6914±0.0041
(Wang et al., 2016)	Benign: 677Malignant: 323 (private)	–	Macrocalcification Detection	SDAE	Acc: 0.873
(Kallenberg et al., 2016)	Total: 1576 (private)	24 × 24	Breast Density Assesment	Sparse AE	AUC: 0.61
(Huynh et al., 2016)	Total: 219 (private)	256 × 256	Mass Classification	CNN + TL + SVM(pre-trained)	AUC: 0.86
(Kooi et al., 2017)	Training: 39872 Testing: 18182 (private)	250 × 250	Mass Detection	CNN: Conv5 + MaxPool5 + Fc2	AUC: 0.929
(Taghanaki et al., 2017)	Total: 949 (private)	128 × 128	Mass Classification	AE	Acc: 0.9845
(Samala et al., 2017)	Benign: 1397 Malignant: 1057(private)	128 × 128	Mass Classification	CNN + TL (pre-trained)	AUC: 0.82±0.02
(Qiu et al., 2017)	Total: 560 (private)	512 × 512	Mass Classification	CNN: Conv3 + MaxPool3 + Fc	AUC: 0.790±0.019
(Wang et al., 2018)	Total: 736 (private)	299 × 299	Mass Classification	CNN + RNN	Acc: 0.85, AUC: 0.89
(Alkhaleefah and Wu, 2018)	Total: 5147 (private)	227 × 227	Mass Classification	CNN + SVM	Acc: 0.92, Sp: 0.86
(Chen et al., 2018)	DDSM, INbreast, Private Dataset	–	Mass Classification	CNN + TL (pre-trained)	AUC: DDSM 0.82, INbreast 0.93, Private 0.70
(Alkhaleefah and Wu, 2018)	Benign: 300 Malignant: 300 (DDSM)	448 × 448	Mass Detection + Classification	YOLO	Detection Acc: 0.997, Classification Acc: 0.97
(Gastounioti et al., 2018)	Total: 318 (private)	63 × 63	Parenchymal Complexity Assessment	CNN + Texture	Sn: 0.81, Sp: 0.98, AUC: 0.90
(Li et al., 2018)	Total: 352 (private)	64 × 63	Mass Detection	CNN	Sn: 0.80
(Wu et al., 2018)	Benign: 8648 Malignant: 1832 (DDSM)	256 × 256	Mass Classification	GAN + ResNet	AUC: 0.896
(Guan and Loew, 2019)	DDSM	–	Mass Detection	GAN + CNN	–
(Lehman et al., 2018)	Total: 8677 (private)	–	Breast Density Assessment	CNN	Acc: 0.94
(Fathy and Ghoneim, 2019)	Normal: 2340 Cancer: 1592 (DDSM)	250 × 250	Mass Detection + Classification	ResNet + Class Activation Map	Classification Sn: 0.998, Sp: 0.821, AUC: 0.96 Detection AUC: 0.9367

3. Challenges and directions

Though the current studies have built a substantial basic for DLR research and application in breast cancer with different imaging modalities, this section will summarize the issues that are still unresolved and outline the future research trends in this area for a better development.

3.1. Balanced dataset

Deep learning has proven to be very promising and effective in big data analytics. However, the available datasets are imbalanced for DLR in breast cancer domain. First, it lacks of public databases involving the three different image modalities, which contain pathological heterogeneity and coexisting benign-malignance in various population. Second, the number, size and format of private datasets are arbitrary and obscure. Another big challenge is annotation, for which clinical radiologists are not always available as the task may be labor intensive and time consuming. Imbalanced dataset of training networks leads to lopsided breast cancer diagnosis.

Future works involve not only establishing public breast cancer databases, but we can make private datasets mainly to solve the problems of limited medical images and complicated annotations. There are various ways where private datasets are helpful for clinical practices including multi-source data, combination of different imaging modalities and combination with diagnostic and histopathological reports. (1) Multi-source data. Mordang et al. (2016) compared CNN on different microcalcification datasets from multi-vendor and received various performances. Bekker et al. (2016) and Carneiro et al. (2015) both used mammography tumor datasets from multi-view, i.e., cranio-caudal (CC) mediolateral oblique (MLO). Byra et al. (2019) introduced a matching layer to convert the ultrasound mass images from grayscale to red, green and blue (RGB) to more efficiently utilize the discriminative power of CNN. Hiramatsu et al. (2017) utilized a breast volume scanner data and Zhou et al. (2018) processed a breast mass dataset with shear-wave elastography for helpful identification. (2) Combination of different imaging modalities. Hadad et al. (2017) designed cross-modal that first trained on mammography images then with transfer learning to identify breast masses on MRI images. (3) Combination with diagnostic reports and histopathological reports. Wang et al. (2018) first applied diagnosis reports with semantic clinic outlines about the mass in mammography that makes interpretable image features extraction.

3.2. Interpretable deep architecture

Along with the complex radiology images, it would be hard to design deep models suitable for the medical data. Hence,

Table 2
Studies conducted for breast cancer in ultrasound based on DLR.*Conv: Convolution; *Pool: Pooling; *Fc: Fully Connected; *AUC: Area Under the Curve (ROC); *Acc: Accuracy; *Sn: Sensitivity; *Sp: Specificity.

References	Data	Input Size (Pixels)	Application	Architecture	Outcome
(Huynh et al., 2016)	Total: 2393 (private)	–	Mass Classification	CNN + TL + SVM (pre-trained)	AUC: 0.88
(Shin et al., 2016)	Dataset 1: 5624 Dataset 2: 163 (DDSM)	–	Mass Detection + Classification	Fast R-CNN + MIL	Confidence Interval(CI): 0.95, Correct Localization (CorLoc): 0.8–0.845
(Cheng et al., 2016)	Total: 1400 (private)	–	Mass Classification	S-DAE	Acc: 0.824±0.045,Sn: 0.77±0.080, Sp: 0.857 ±0.068
(Han et al., 2017)	Benign: 4254 Malignant: 3154 (private)	255 × 255	Mass Classification	GoogleNet	Acc: 0.91, Sn: 0.86, Sp: 0.96
(Hiramatsu et al., 2017)	Breast Volume (private)	–	Mass Detection	Deep CNN	Sp: 0.90
(Xiao et al., 2018)	Benign: 1370 Malignant: 688 (private)	299 × 299	Mass Classification	InceptionV3, ResNet50, Xception	Acc: InceptionV3 0.8513, ResNet 0.8494, Xception 0.8406
(Zhou et al., 2018)	Benign: 222 Malignant: 318 (private)	480 × 320	Mass Classification	Conv13 + ReLU15 + MaxPool4 + Fc3 + Softmax	Acc: 0.958, Sn: 0.962, Sp: 0.957
(Li et al., 2018)	Training: 6727 Testing1881 (private)	224 × 224	Mass Detection	Faster R-CNN	Acc: 0.935
(Byra et al., 2019)	Total: 882 (private)	128 × 128	Mass Classification	CNN + TL (pre-trained)	AUC: 0.89
(Huang et al., 2019)	Total: 2238 (private)	–	Mass Detection + Classification	CNN	Acc on BI-RADS 3: 0.998, 4A: 0.94, 4B: 0.734, 4C: 0.922 5: 0.876
(Chiang et al., 2018)	Total: 171 (private)	32 × 32 × 32	Mass Detection	3D CNN	Sn: 0.95

Table 3
Studies conducted for breast cancer in MRI based on DLR.*Conv: Convolution; *Pool: Pooling; *Fc: Fully Connected; *AUC: Area Under the Curve (ROC); *Acc: Accuracy; *Sn: Sensitivity; *Sp: Specificity.

References	Data	Input Size (Pixels)	Application	Architecture	Outcome
(Maicas et al., 2017)	Training: 72 Testing: 69 (private)	100 × 100 × 50	Mass Detection	DRL	Sn: 0.8, Time: 92±21 s
(Rasti et al., 2017)	Benign: 59 Malignant: 53 (private)	32 × 32	Mass Classification	Mixture Ensemble CNN	Acc: 0.9639, Sn: 0.9773, Sp: 0.9487
(Hadad et al., 2017)	Total: 3537 (private)	–	Mass Classification	VGG + TL (pre-trained)	Acc: 0.93
(Antropova et al., 2017)	Benign: 191 Malignant: 449 (private)	–	Segmentation + Classification	AlexNet	AUC: 0.91
(Li et al., 2018)	Benign: 66 Malignant: 77 (private)	–	Mass Classification	3D CNN	Acc: 0.781, Sn: 0.744, Sp: 0.823
(Dalmış et al., 2018)	Benign: 224 Malignant: 161 (private)	39 × 39 × 39	Mass Detection	3D CNN	Sn: 0.6429±0.0537
(Antropova et al., 2018)	Total: 703 (private)	–	Mass Classification	CNN + LSTM	AUC: 0.85
(Ha et al., 2019)	Total: 127 (private)	–	NAC Prediction	CNN: Conv10 + MaxPool4 + Fc + Dropout0.5	Acc: 0.95, Sn: 0.93, Sp: 0.77, AUC: 0.93
(Zheng et al., 2018)	–	–	Small Mass Classification	CNN + LSTM	Acc: 0.847
(Ha et al., 2019)	Total: 127 (private)	–	NAC Prediction	CNN	Acc: 0.877±0.006
(Truhn et al., 2018)	Benign: 787 Malignant:507 (private)	–	Mass Classification	ResNet	AUC: 0.88
(Adoui et al., 2019)	Training: 80% Validiation: 20%(private)	–	NAC Prediction	CNN	–
(Ha et al., 2019)	Total: 216 Patients (private)	–	Molecular Subtype Classification	CNN: 14 Layers	Acc: 0.7
(Zhu et al., 2018)	–	–	Molecular Subtype Classification	CNN + SVM	AUC: 0.64
(Gallego-Ortiz and Martel, 2019)	Benign: 267 Malignant: 110 (private)	–	Mass Classification	AE + Multi-layer Perceptron (MLP)	AUC: 0.81±0.1

non-interpretable deep learning networks are common challenges for radiomics. Capability of deep learning in dealing with heterogeneous data resources brings more opportunities for interpretable DLR in breast cancer together with diagnostic and histopathological reports. Besides, image captioning, which merges CV and natural language processing (NLP), has achieved large progress (Vinyals et al., 2015; Tan and Chan, 2019). Basically, the paradigm is a typical encoder-decoder architecture utilizing the visual features obtained from CNN (encoder) to generate the descriptions of given images through RNN (decoder). Inspired by image captioning, automatic generation of medical imaging report emerges in medical domain (Shin et al., 2016; Zhang et al., 2017). This hybrid deep learning architecture, which employs the text features to express the radiomics features, is able to better interpret the diagnostic results for

breast cancer (Jing et al., 2017; Wang et al., 2018). The first is to allow them to understand the impacting factors on the learning procedure of deep network within breast cancer images. The second is to pay attention on the learning attributes, weights and activations of the models on the breast cancer images. Furthermore, the attentional mechanism selecting critical information from the wide range of current objective is used to illustrate the correct deep learning in the imaging area (Xu et al., 2015). Attention weights not only show the inner workings of the architecture but also can highlight the essential message (e.g., calcification, spiculate edge) contributing to the explainable results. Integrating with the hand-crafted features and conventional machine learning is also anther two researching directions for better performance, see Section 2.2.

Additionally, deep features engineering of breast cancer imaging has not been fully studied. As mentioned before, making annotations including cancer class, lesion category, small size calcification on breast cancer images is a complicated and tedious process. Hence, unsupervised learning without the need of labelling is believed to be a crucial direction for DLR. As in Fig. 7, unsupervised deep learning methods still have a wide research gap for DLR in breast cancer domain. Besides, many clinical decision problems are by nature sequential (Jonsson, 2019), which is the characteristics of reinforcement learning processing data, and RL makes up the shortage of lacking the feedback mechanism in deep learning. Therefore, RL combining with deep learning architecture (deep reinforcement learning) has the potential of application in medicine area (Gottesman et al., 2019). However, to our best of knowledge, only one paper by Maicas et al. (2017) utilized RL in breast cancer area, which means applying RL into DLR for CAD of breast cancer is another promising research direction.

3.3. Clinical application

Medical decision play an important role in management and care of patients. An error in medical imaging reporting can bring detrimental consequences. To improve the accuracy of medical reports, we need to focus on the practical application value of the DLR. Although most studies tried to improve the final performance, they did not make a full use of the available information. We should consider the characteristics of breast cancer in different imaging (such as detailed malignant lesion features on ultrasound: angular margin, hypo echogenicity, internal vascularity, posterior shadowing). Besides, there are some risk factors of breast cancer including age, family history, genetic of patients usually considered by doctors, however, they are neglected in DLR. It can be another direction we can explore. For example, MRI improves the sensitivity of screening in women with a family or genetic predisposition that can be researched utilizing deep learning.

As mentioned before in Fig. 14, applications of DLR on prognosis, treatment selection and genomics still have a wide research gap in the breast cancer domain. However, it is specifically important as these areas have deep relationships with clinical diagnosis and treatment, which can reduce the death rate. Therefore, we should pay more attentions to the deep features obtained by DLR, which can be seen as data or information for further studies. Besides, multi-task learning (such as the current fusion of detection and classification) is a future direction. Radiomics features in a deep neural framework can weaken overfitting through several tasks gathering the shared learning features at a time.

3.4. Unified evaluation

The technical details for experiments were not described in detailed in most of the studies. Measurement indicators and baseline methods were selected arbitrarily, leading to an evaluation that lacks of standardization. Researchers mostly focus on the accuracy, sensitivity or specificity of the DLR network proposed. Designing a unified and practical evaluation criterion architecture is necessary, especially for the DLR in breast cancer.

3.5. Interdisciplinary background

While there are great interests in DLR for breast cancer, progresses are impinged without the close collaboration with different expertise. To many radiologists who lack technical knowledge in computer science, designing and coding for deep learning models are extremely challenging. On the other hand, the engineering and computer science specialists, who have little professional medical knowledge, cannot comprehensively understand information of radiology scanning and complex clinical applications. Hence, better communication and closer working culture would be very essential to propel DLR to much higher and clinically useful application in breast cancer with medical imaging.

4. Conclusion

Breast cancer is a serious threat to women throughout the world and improvement of the current situation of CAD systems for breast cancer is a big concern. The explosive growth of deep learning in radiomics makes it possible that not only extracts useful features, but also take full advantage of large data for better breast cancer diagnostic and precision medicine. In this paper, we surveyed the recent studies on the subject of applying deep learning radiomics for breast cancer with different imaging modalities. It is organized from the aspects of dataset, architecture, application and evaluation. The focus is on the deep learning frameworks developed in three breast imaging modalities, i.e., ultrasound, mammography, MRI. Besides, potential challenges and future directions are discussed to provide references for the further studies in this area. With the potential directions of DLR-based CAD systems in breast cancer analyzed in this paper, there are vast opportunities for developments in research and clinical applications. Furthermore, this research can provide future references for DLR in other medical domains. We hope that radiologists, biomedical researchers and computer science researchers can work together to bring this field to the next level.

Due to the tremendous concerns about breast cancer and rapid development of DLR, many research efforts about DLR for breast cancer have been published so far. However, it is quite difficult to summarize all the research contributions in a single research article. This work has tried to provide a holistic approach to the state-of-the-art findings about the breast cancer imaging utilizing DLR-based CAD systems. In order to more specifically understand the procedure of clinical applications of these CAD systems of breast cancer, we will further study on private dataset. First, we will attempt to establish a deep learning radiomics model based on GAN for data augmentation in breast cancer classification with CNN. Second, we will try to propose a radiomics-reporting network with text attention for improving interpretability of deep learning radiomics in breast cancer diagnosis.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. *Expert Systems with Applications, 46*, 139–144. https://doi.org/10.1016/j.eswa.2015.10.015.

Afshar, P., Mohammadi, A., Plataniotis, K. N., Oikonomou, A., & Benali, H. (2019). From handcrafted to deep-learning-based cancer radiomics: Challenges and opportunities. *IEEE Signal Processing Magazine, 36*(4), 132–160. https://doi.org/10.1109/MSP.2019.2900993.

Al-masni, M. A., Al-antari, M. A., Park, J.-M., Gi, G., Kim, T.-Y., Rivera, P., Valarezo, E., Choi, M.-T., Han, S.-M., & Kim, T.-S. (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolo-based cad system. *Computer Methods and Programs in Biomedicine, 157*, 85–94. https://doi.org/10.1016/j.cmpb.2018.01.017.

Alkhaleefah, M. &Wu, C. -C. (2018). A hybrid cnn and rbf-based svm approach for breast cancer classification in mammograms. In 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 94–899). IEEE. https://doi.org/10.1109/SMC.2018.00159.

Antropova, N., Huynh, B., & Giger, M. (2017). Performance comparison of deep learning and segmentation-based radiomic methods in the task of distinguishing benign and malignant breast lesions on dce-mri. *Medical imaging 2017: Computer-aided diagnosis* (10134, pp. 101341G). International Society for Optics and Photonics. https://doi.org/10.1117/12.2255582.

Antropova, N., Huynh, B., & Giger, M. (2018). Recurrent neural networks for breast lesion classification based on dce-mris. *Medical imaging 2018: Computer-aided diagnosis* (10575, pp. 105752M). International Society for Optics and Photonics. https://doi.org/10.1117/12.2293265.

Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., & Lopez, M.A.G. (2015). Convolutional neural networks for mammography mass lesion classification. In 2015 37th Annual international conference of the IEEE engineering in medicine and biology society (EMBC), pages 797–800. IEEE, https://doi.org/10.1109/EMBC.2015.7318482.

Ballester, P., & Araujo, R. M. (2016). On the performance of googlenet and alexnet applied to sketches. *Thirtieth AAAI Conference on Artificial Intelligence*. https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12278/11712.

Bekker, A. J., Greenspan, H., & Goldberger, J. (2016). A multi-view deep learning architecture for classification of breast microcalcifications. In 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI) (pp. 726–730). IEEE. https://doi.org/10.1109/ISBI.2016.7493369.

Berg, W. A., Zhang, Z., Lehrer, D., Jong, R. A., Pisano, E. D., Barr, R. G., Böhm-Vélez, M., Mahoney, M. C., Evans, W. P., Larsen, L. H., et al. (2012). Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *Jama, 307*(13), 1394–1404. https://doi:10.1001/jama.2012.388.

Beura, S., Majhi, B., & Dash, R. (2015). Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing, 154*, 1–14. https://doi.org/10.1016/j.neucom.2014.12.032.

Bruno, R. N. M., & Homero, S. (2011). Online mammographic images database for development and comparison of cad schemes. *Journal of Digital Imaging, 24*, 500–506. https://link.springer.com/article/10.1007%2Fs10278-010-9297-2.

Byra, M., Galperin, M., Ojeda-Fournier, H., Olson, L., O'Boyle, M., Comstock, C., & Andre, M. (2019). Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion. *Medical Physics, 46*(2), 746–755. https://doi.org/10.1002/mp.13361.

Carneiro, G., Nascimento, J. & Bradley, A. P. (2015). Unregistered multiview mammogram analysis with pre trained deep learning models. In International Conference on Medical Image Computing and Computer Assisted Intervention (pp. 652–660). Springer. https://doi.org/10.1007/978-3-319-24574-4_78.

Chaunzwa, T., Xu, Y., Mak, R., Christiani, D., Lanuti, M., Shafer, A., Dia, N. & Aerts, H. (2018). Image based lung cancer phenotyping with deep-learning radiomics. In Medical physics (Vol. 45, pp. E165–E165). Wiley River ST, Hoboken 07030–5774, NJ USA.

Chen, B., Zhang, R., Gan, Y., Yang, L., & Li, W. (2017). Development and clinical application of radiomics in lung cancer. *Radiation Oncology, 12*(1), 154. https://doi.org/10.1186/s13014-017-0885-x.

Chen, Q., Liu, J., Luo, K., Zhang, X. & Wang, X. (2018). Transfer deep learning mammography diagnostic model from public datasets to clinical practice: A comparison of model performance and mammography datasets. In 14th International workshop on breast imaging (IWBI 2018) (Vol. 10718, pp. 1071813). International Society for Optics and Photonics. https://doi.org/10.1117/12.2317411.

Cheng, J.-Z., Ni, D., Chou, Y.-H., Qin, J., Tiu, C.-M., Chang, Y.-C., Huang, C.-S., Shen, D., & Chen, C.-M. (2016). Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans. *Scientific Reports, 6*, 24454. https://doi.org/10.1038/srep24454.

Chiang, T.-C., Huang, Y.-S., Chen, R.-T., Huang, C.-S., & Chang, R.-F. (2018). Tumor detection in automated breast ultrasound using 3-d cnn and prioritized candidate aggregation. *IEEE Transactions on Medical Imaging, 38*(1), 240–249. https://doi.org/10.1109/TMI.2018.2860257.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1251–1258). http://openaccess.thecvf.com/content_cvpr_2017/html/Chollet_Xception_Deep_Learning_CVPR_2017_paper.html.

Corsetti, V., Houssami, N., Ghirardi, M., Ferrari, A., Speziani, M., Bellarosa, S., Remida, G., Gasparotti, C., Galligioni, E., & Ciatto, S. (2011). Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: Interval breast cancers at 1 year follow-up. *European Journal of Cancer, 47*(7), 1021–1026. https://doi.org/10.1016/j.ejca.2010.12.002.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine, 35*(1), 53–65. https://doi.org/10.1109/MSP.2017.2765202.

Dalmış, M. U., Vreemann, S., Kooi, T., Mann, R. M., Karssemeijer, N., & Gubern-Mérida, A. (2018). Fully automated detection of breast cancer in screening MRI using convolutional neural networks. *Journal of Medical Imaging, 5*(1). https://doi.org/10.1117/1.JMI.5.1.014502 014502.

Dhungel, N., Carneiro, G. & Bradley, A. P. (2015). Automated mass detection in mammograms using cascaded deep learning and random forests. In International conference on digital image computing: Techniques and applications (DICTA) (pp. 1–8). IEEE. https://doi.org/10.1109/DICTA.2015.7371234.

Dhungel, N., Carneiro, G., & Bradley, A. P. (2017). A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Medical Image Analysis, 37*, 114–128. https://doi.org/10.1016/j.media.2017.01.009.

Douzas, G., & Bacao, F. (2018). Effective data generation for imbalanced learning using conditional generative adversarial networks. *Expert Systems with Applications, 91*, 464–471. https://doi.org/10.1016/j.eswa.2017.09.030.

Drukker, K., Giger, M. L., Vyborny, C. J., & Mendelson, E. B. (2004). Computerized detection and classification of cancer on breast ultrasound. *Academic Radiology, 11*(5), 526–535. https://doi.org/10.1016/S1076-6332(03)00723-2.

D'Orsi, C., Bassett, L., Feig, S., et al. (2018). *Breast imaging reporting and data system (BI-RADS)*. Reston: American College of Radiology. https://books.google.com.my/books?id=Zj9FDwAAQBAJ&lpg=PT83&ots=C34Xnl2XMy&dq=Breast%20Imaging%20Reporting%20and%20Data%20System%202013&lr&hl=zh-CN&pg=PT83#v=onepage&q=Breast%20Imaging%20Reporting%20and%20Data%20System%202013&f=false.

El Adoui, M., Drisis, S., & Benjelloun, M. (2019). Predict breast tumor response to chemotherapy using a 3d deep learning architecture applied to dce-mri data. In *International work-conference on bioinformatics and biomedical engineering* (pp. 33–40). Springer. https://doi.org/10.1007/978-3-030-17935-9_4.

Ertosun, M. G. & Rubin, D. L. (2015). Probabilistic visual search for masses within mammography images using deep learning. In 2015 IEEE international conference on bioinformatics and biomedicine (BIBM) (pp. 1310–1315). IEEE. https://doi.org/10.1109/BIBM.2015.7359868.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115. https://www.nature.com/articles/nature21056?TB_iframe=true&width=914.4&height=921.6.

Fathy, W. E., & Ghoneim, A. S. (2019). A deep learning approach for breast cancer mass detection. *International Journal OF Advanced Computer Science and Applications, 10*(1), 175–182. https://doi.org/10.14569/ijacsa.2019.0100123.

Freer, P. E. (2015). Mammographic breast density: Impact on breast cancer risk and implications for screening. *Radiographics, 35*(2), 302–315. https://doi.org/10.1148/rg.352140106.

Gallego-Ortiz, C., & Martel, A. L. (2019). A graph-based lesion characterization and deep embedding approach for improved computer-aided diagnosis of nonmass breast mri lesions. *Medical Image Analysis, 51*, 116–124. https://doi.org/10.1016/j.media.2018.10.011.

Gastounioti, A., Oustimov, A., Hsieh, M.-K., Pantalone, L., Conant, E. F., & Kontos, D. (2018). Using convolutional neural networks for enhanced capture of breast parenchymal complexity patterns associated with breast cancer risk. *Academic Radiology, 25*(8), 977–984. https://doi.org/10.1016/j.acra.2017.12.025.

Gillies, R. J., Kinahan, P. E., & Hricak, H. (2015). Radiomics: Images are more than pictures, they are data. *Radiology, 278*(2), 563–577. https://doi.org/10.1148/radiol.2015151169.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440–1448). http://openaccess.thecvf.com/content_iccv_2015/html/Girshick_Fast_R-CNN_ICCV_2015_paper.html.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 2672–2680. http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

Gottesman, O., Johansson, F., Komorowski, M., Faisal, A., Sontag, D., F., D. -V. & L., C. A. (2019). Guidelines for reinforcement learning in healthcare. Nature Medicine, 25:16–18, https://doi.org/10.1038/s41591-018-0310-5.

Greenspan, H., Van Ginneken, B., & Summers, R. M. (2016). Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging, 35*(5), 1153–1159. https://doi.org/10.1109/TMI.2016.2553401.

Grimm, L. J., Johnson, K. S., Marcom, P. K., Baker, J. A., & Soo, M. S. (2014). Can breast cancer molecular subtype help to select patients for preoperative mr imaging? *Radiology, 274*(2), 352–358. https://doi.org/10.1148/radiol.14140594.

Guan, S., & Loew, M. (2019). Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. *Journal of Medical Imaging, 6*(3). https://doi.org/10.1117/1.JMI.6.3.031411 031411.

Gupta, D. (2017). Fundamentals of deep learning – activation functions and when to use them? https://www. analyticsvidhya. com/blog/2017/10/fundamentals-deep-learning-activationfunctions-when-to-use-them.

Ha, R., Chang, P., Mutasa, S., Karcich, J., Goodman, S., Blum, E., Kalinsky, K., Liu, M. Z., & Jambawalikar, S. (2019). Convolutional neural network using a breast mri tumor dataset can predict oncotype dx recurrence score. *Journal of Magnetic Resonance Imaging, 49*(2), 518–524. https://doi.org/10.1002/jmri.26244.

Ha, R., Chin, C., Karcich, J., Liu, M. Z., Chang, P., Mutasa, S., Van Sant, E. P., Wynn, R. T., Connolly, E., & Jambawalikar, S. (2019). Prior to initiation of chemotherapy, can we predict breast tumor response? deep learning convolutional neural networks approach using a breast mri tumor dataset. *Journal of Digital Imaging, 32*(5), 693–701. https://doi.org//10.1007/s10278-018-0144-1.

Ha, R., Mutasa, S., Karcich, J., Gupta, N., Van Sant, E. P., Nemer, J., Sun, M., Chang, P., Liu, M. Z., & Jambawalikar, S. (2019). Predicting breast cancer molecular subtype with mri dataset utilizing convolutional neural network algorithm. *Journal of Digital Imaging, 32*(2), 276–282. https://doi.org/10.1007/s10278-019-00179-2.

Hadad, O., Bakalo, R., Ben-Ari, R., Hashoul, S., & Amit, G. (2017). *Classification of breast lesions using cross-modal deep learning* (2017, pp. 109–112). IEEE. https://doi.org/10.1109/ISBI.2017.7950480.

Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., & Zwiggelaar, R. (2018). Deep learning in mammography and breast histology, an overview and future trends. *Medical Image Analysis, 47*, 45–67. https://doi.org/10.1016/j.media.2018.03.006.

Han, S., Kang, H.-K., Jeong, J.-Y., Park, M.-H., Kim, W., Bang, W.-C., & Seong, Y.-K. (2017). A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Physics in Medicine & Biology, 62*(19), 7714. https://iopscience.iop.org/article/10.1088/1361-6560/aa82ec/meta.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Heath, M., Bowyer, K., Kopans, D., Moore, R. & Kegelmeyer, W. (2001). The digital database for screening mammography. In Proceedings of the 5th international workshop on digital mammography (pp. 212–218). Medical Physics Publishing. https://www3.nd.edu/ kwb/Heath_EtAl_IWDM_2000.pdf.

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In Neural networks for perception (pp. 65–93). Elsevier, https://doi.org/10.1016/B978-0-12-741252-8.50010-8.

Hinton, G. E. (2009). Deep belief networks. *Scholarpedia, 4*(5), 5947. https://doi.org/10.4249/scholarpedia.5947.

Hiramatsu, Y., Muramatsu, C., Kobayashi, H., Hara, T., & Fujita, H. (2017). Automated detection of masses on whole breast volume ultrasound scanner: False positive reduction using deep convolutional neural network. *Medical imaging 2017: Computer-aided diagnosis* (Vol. 10134, pp. 101342S). International Society for Optics and Photonics. https://doi.org/10.1117/12.2254581.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Honda, E., Nakayama, R., Koyama, H., & Yamashita, A. (2016). Computer-aided diagnosis scheme for distinguishing between benign and malignant masses in breast dce-mri. *Journal of Digital Imaging, 29*(3), 388–393. https://doi.org/10.1007/s10278-015-9856-7.

Horsch, K., Giger, M. L., Vyborny, C. J., Lan, L., Mendelson, E. B., & Hendrick, R. E. (2006). Classification of breast lesions with multimodality computer-aided diagnosis: Observer study results on an independent clinical data set. *Radiology, 240*(2), 357–368. https://doi.org/10.1148/radiol.2401050208.

Huang, Y., Han, L., Dou, H., Luo, H., Yuan, Z., Liu, Q., Zhang, J., & Yin, G. (2019). Two-stage cnns for computerized bi-rads categorization in breast ultrasound images. *Biomedical Engineering Online, 18*(1), 8. https://doi.org/10.1186/s12938-019-0626-5.

Huynh, B., Drukker, K., & Giger, M. (2016). Mo-de- 207b-06: Computer-aided diagnosis of breast ultrasound images using transfer learning from deep convolutional neural networks. *Medical Physics, 43*(6Part30). https://doi.org/10.1118/1.4957255. 3705–3705.

Huynh, B. Q., H., L. & Giger, M. L. (2016b). Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. Journal of Medical Imaging, 3(3):034501, https://doi.org/10.1117/1.JMI.3.3.034501.

Jiao, Z., Gao, X., Wang, Y., & Li, J. (2016). A deep feature based framework for breast masses classification. *Neurocomputing, 197*, 221–231. https://doi.org/10.1016/j.neucom.2016.02.060.

Jing, B., Xie, P. & Xing, E. (2017). On the automatic generation of medical imaging reports. arXiv preprint arXiv:1711.08195, https://arxiv.org/abs/1711.08195.

Jonsson, A. (2019). Deep reinforcement learning in medicine. *Kidney Diseases, 5*, 18–22. https://doi.org/10.1159/000492670.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research, 4*, 237–285. https://doi.org/10.1613/jair.301.

Kallenberg, M., Petersen, K., Nielsen, M., Ng, A. Y., Diao, P., Igel, C., Vachon, C. M., Holland, K., Winkel, R. R., Karssemeijer, N., et al. (2016). Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE Transactions on Medical Imaging, 35*(5), 1322–1331. https://doi.org/10.1109/TMI.2016.2532122.

Kelly, K. M., Dean, J., Comulada, W. S., & Lee, S.-J. (2010). Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. *European Radiology, 20*(3), 734–742. https://doi.org/10.1007/s00330-009-1588-y.

Kooi, T., Litjens, G., Van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., den Heeten, A., & Karssemeijer, N. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Medical Image Analysis, 35*, 303–312. https://doi.org/10.1016/j.media.2016.07.007.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R. G., Granton, P., Zegers, C. M., Gillies, R., Boellard, R., Dekker, A., et al. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer, 48*(4), 441–446. https://doi.org/10.1016/j.ejca.2011.11.036.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436. https://doi.org/10.1038/nature14539.

Lee, C. H., Dershaw, D. D., Kopans, D., Evans, P., Monsees, B., Monticciolo, D., Brenner, R. J., Bassett, L., Berg, W., Feig, S., et al. (2010). Breast cancer screening with imaging: Recommendations from the society of breast imaging and the acr on the use of mammography, breast mri, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. *Journal of the American College of Radiology, 7*(1), 18–27. https://doi.org/10.1016/j.jacr.2009.09.022.

Lee, C.-Y., Chen, G.-L., Zhang, Z.-X., Chou, Y.-H., & Hsu, C.-C. (2018). Is intensity inhomogeneity correction useful for classification of breast cancer in sonograms using deep neural network? *Journal of Healthcare Engineering, 2018*. https://doi.org/10.1155/2018/8413403.

Lehman, C. D., Yala, A., Schuster, T., Dontchos, B., Bahl, M., Swanson, K., & Barzilay, R. (2018). Mammographic breast density assessment using deep learning: Clinical implementation. *Radiology, 290*(1), 52–58. https://doi.org/10.1148/radiol.2018180694.

Li, H., Weng, J., Shi, Y., Gu, W., Mao, Y., Wang, Y., Liu, W., & Zhang, J. (2018). An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images. *Scientific Reports, 8*(1), 6600. https://doi.org/10.1038/s41598-018-25005-7.

Li, J., Fan, M., Zhang, J. & Li, L. (2017). Discriminating between benign and malignant breast tumors using 3d convolutional neural network in dynamic contrast enhanced-mr images. In Medical imaging 2017: Imaging informatics for healthcare, research, and applications (Vol. 10138, pp. 1013808). International Society for Optics and Photonics. doi: 10.1117/12.2254716.

Li, Y. (2017). Deep reinforcement learning: An overview. arXiv preprint arXiv:1701.07274, https://arxiv.org/pdf/1701.07274.pdf.

Li, Y., Chen, H., Zhang, L. & Cheng, L. (2018b). Mammographic mass detection based on convolution neural network. In 2018 24th International conference on pattern recognition (ICPR) (pp. 3850–3855). IEEE. https://doi.org/10.1109/ICPR.2018.8545557.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005.

Liu, Z., Wang, S., Dong, D., Wei, J., Fang, C., Zhou, X., Sun, K., Li, L., Li, B., Wang, M., et al. (2019). The applications of radiomics in precision diagnosis and treatment of oncology: Opportunities and challenges. *Theranostics, 9*(5), 1303. https://doi.org/10.7150/thno.30309.

Lopez, M. G., Posada, N., Moura, D. C., Pollán, R. R., Valiente, J. M. F., Ortega, C. S., et al. (2012). Bcdr: A breast cancer digital repository. In *15th International conference on experimental mechanics*.

Lynch, C. J., & Liston, C. (2018). New machine-learning technologies for computer-aided diagnosis. *Nature Medicine*, 1. https://doi.org/10.1038/nature14539.

Maicas, G., Carneiro, G., Bradley, A. P., Nascimento, J. C., & Reid, I. (2017). Deep reinforcement learning for active breast lesion detection from dce-mri. In *International conference on medical image computing and computer-assisted intervention* (pp. 665–673). Springer.

Maron, O., & Lozano-Pérez, T. (1998). A framework for multiple-instance learning. *Advances in Neural Information Processing Systems*, 570–576. http://papers.nips.cc/paper/1346-a-framework-for-multiple-instance-learning.pdf.

Mordang, J.-J., Janssen, T., Bria, A., Kooi, T., Gubern-Mérida, A., & Karssemeijer, N. (2016). Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. In *International workshop on breast imaging* (pp. 35–42). Springer. https://doi.org/10.1007/978-3-319-41546-8_5.

Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). Inbreast: Toward a full-field digital mammographic database. *Academic Radiology, 19*(2), 236–248. https://doi.org/10.1016/j.acra.2011.09.014.

National Cancer Institute (2019a). Breast cancer. https://www.cancer.gov/types/breast/patient/male-breast-treatment-pdq/. Accessed 4 February 2020.

National Cancer Institute (2019b). Breast cancer treatment. https://www.cancer.gov/types/breast/patient/breast-treatment-pdq/. Accessed 4 February 2020.

Ng, A. et al. (2011). Sparse autoencoder. CS294A Lecture notes, 72(2011):1–19. http://ailab.chonbuk.ac.kr/seminar_board/pds1_files/sparseAutoencoder.pdf.

Qiu, Y., Yan, S., Gundreddy, R. R., Wang, Y., Cheng, S., Liu, H., & Zheng, B. (2017). A new approach to develop computer-aided diagnosis scheme of breast mass classification using deep learning technology. *Journal of X-ray Science and Technology, 25*(5), 751–763. https://doi.org/10.3233/XST-16226.

Ranzato, M., & Szummer, M. (2008). Semi-supervised learning of compact document representations with deep networks. In *Proceedings of the 25th international conference on machine learning* (pp. 792–799). https://doi.org/10.1145/1390156.1390256. ACM.

Rasti, R., Teshnehlab, M., & Phung, S. L. (2017). Breast cancer diagnosis in dce-mri using mixture ensemble of convolutional neural networks. *Pattern Recognition, 72*, 381–390. https://doi.org/10.1016/j.patcog.2017.08.004.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788). https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf.

Rouhi, R., Jafari, M., Kasaei, S., & Keshavarzian, P. (2015). Benign and malignant breast tumors classification based on region growing and cnn segmentation. *Expert Systems with Applications, 42*(3), 990–1002. https://doi.org/10.1016/j.eswa.2014.09.020.

Salakhutdinov, R., & Hinton, G. (2009). Deep boltzmann machines. *Artificial Intelligence and Statistics*, 448–455. http://proceedings.mlr.press/v5/salakhutdinov09a/salakhutdinov09a.pdf.

Samala, R. K., Chan, H.-P., Hadjiiski, L. M., Helvie, M. A., Cha, K. H., & Richter, C. D. (2017). Multi-task transfer learning deep convolutional neural network: Application to computer-aided diagnosis of breast cancer on mammograms. *Physics in Medicine & Biology, 62*(23), 8894. https://doi.org//10.1088/1361-6560/aa93d4.

Saslow, D., Boetes, C., Burke, W., Harms, S., Leach, M. O., Lehman, C. D., Morris, E., Pisano, E., Schnall, M., Sener, S., et al. (2007). American cancer society guidelines for breast screening with mri as an adjunct to mammography. *CA: A Cancer Journal for Clinicians, 57*(2), 75–89. https://doi.org/10.3322/canjclin.57.2.75.

Shin, H. C., Roberts, K., Lu, L., Demner-Fushman, D., Yao, J., & Summers, R. M. (2016). Learning to read chest X-rays: Recurrent neural cascade model for automated image annotation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2497–2506). https://doi.org/10.1109/CVPR.2016.274.

Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging, 35*(5), 1285–1298. https://doi.org/10.1109/TMI.2016.2528162.

Shin, S. Y., Lee, S., Yun, I. D., Kim, S. M., & Lee, K. M. (2018). Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. *IEEE Transactions on Medical Imaging, 38*(3), 762–774. https://doi.org/10.1109/TMI.2018.2872031.

Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, https://arxiv.org/abs/1409.1556.

Smyth, P. (1996). Bounds on the mean classification error rate of multiple experts. *Pattern Recognition Letters, 17*(12), 1253–1257. https://doi.org/10.1016/0167-8655(96)00105-5.

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I. & et al. (2015). Mammographic image analysis society (mias) database v1.21. https://www.repository.cam.ac.uk/handle/ 1810/250394.

Suzuki, K. (2012). Machine learning in computer-aided diagnosis: Medical imaging intelligence and analysis: Medical imaging intelligence and analysis. IGI Global, https://books.google.com.my/books?id=2V5aRuebneMC&lpg=PR1&ots=Q5tcbaroON&dq=Machine%20Learning%20in%20Computer-Aided%20Diagnosis%3A%20Medical%20Imaging%20Intelligence%20and%20Analysis&lr&hl=zh-CN&pg=PR1#v=onepage&q=Machine%20Learning%20in%20Computer-Aided%20Diagnosis:%20Medical%20Imaging%20Intelligence%20and%20Analysis&f=false.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818–2826). https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html.

Taghanaki, S. A., Kawahara, J., Miles, B., & Hamarneh, G. (2017). Pareto-optimal multi-objective dimensionality reduction deep auto-encoder for mammography classification. *Computer Methods and Programs in Biomedicine, 145*, 85–93. https://doi.org/10.1016/j.cmpb.2017.04.012.

Tan, Y. H., & Chan, C. S. (2019). Phrase-based image caption generator with hierarchical lstm network. *Neurocomputing, 333*, 86–100.

Tang, J., Rangayyan, R. M., Xu, J., El Naqa, I., & Yang, Y. (2009). Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances. *IEEE Transactions on Information Technology in Biomedicine, 13*(2), 236–251. https://doi.org/10.1109/TITB.2008.2009441.

Ting, F. F., Tan, Y. J., & Sim, K. S. (2019). Convolutional neural network improvement for breast cancer classification. *Expert Systems with Applications, 120*, 103–115. https://doi.org/10.1016/j.eswa.2018.11.008.

Truhn, D., Schrading, S., Haarburger, C., Schneider, H., Merhof, D., & Kuhl, C. (2018). Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast mri. *Radiology, 290*(2), 290–297. https://doi.org/10.1148/radiol.2018181352.

Tsochatzidis, L., Costaridou, L., & Pratikakis, I. (2019). Deep learning for breast cancer diagnosis from mammograms a comparative study. *Journal of Imaging, 5*(3), 37. https://doi.org/10.3390/jimaging5030037.

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on machine learning* (pp. 1096–1103). https://doi.org/10.1145/1390156.1390294. ACM.

Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3156–3164). https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Vinyals_Show_and_Tell_2015_CVPR_paper.html.

Wang, H., Feng, J., Zhang, Z., Su, H., Cui, L., He, H., & Liu, L. (2018). Breast mass classification via deeply integrating the contextual information from multi-view data. *Pattern Recognition, 80*, 42–52. https://doi.org/10.1016/j.patcog.2018.02.026.

Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., & Li, L. (2016). Discrimination of breast cancer with microcalcifications on mammography by deep learning. *Scientific Reports, 6*, 27327. https://doi.org/10.1038/srep27327.

William, H. W., Street, N. W. & Olvi, L. M. (1995). Breast cancer wisconsin (diagnostic) data set. https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29.

Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation, 1*(2), 270–280. https://doi.org/10.1162/neco.1989.1.2.270.

World Health Organization (2020). Breast cancer. https://www.who.int/cancer/detection/breastcancer/en/. Accessed 4 February 2020.

Wu, E., Wu, K., Cox, D., & Lotter, W. (2018). Conditional infilling gans for data augmentation in mammogram classification. In *Image analysis for moving organ, breast, and thoracic images* (pp. 98–106). Springer. https://doi.org/10.1007/978-3-030-00946-5_11.

Xiao, T., Liu, L., Li, K., Qin, W., Yu, S., & Li, Z. (2018). Comparison of transferred deep neural networks in ultrasonic breast masses discrimination. *BioMed Research International, 2018*. https://doi.org/10.1155/2018/4605191.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., & Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. *International Conference on Machine Learning*, 2048–2057. http://proceedings.mlr.press/v37/xuc15.pdf.

Xu, L. (1993). Least mean square error reconstruction principle for self-organizing neural-nets. *Neural Networks, 6*(5), 627–648. https://doi.org/10.1016/S0893-6080(05)80107-8.

Yassin, N. I., Omran, S., El Houby, E. M., & Allam, H. (2018). Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer Methods and Programs in Biomedicine, 156*, 25–45. https://doi.org/10.1016/j.cmpb.2017.12.012.

Yeh, E., Slanetz, P., Kopans, D. B., Rafferty, E., Georgian-Smith, D., Moy, L., Halpern, E., Moore, R., Kuter, I., & Taghian, A. (2005). Prospective comparison of mammography, sonography, and mri in patients undergoing neoadjuvant chemotherapy for palpable breast cancer. *American Journal of Roentgenology, 184*(3), 868–877. https://doi.org/10.2214/ajr.184.3.01840868.

Zhang, J., Saha, A., Zhu, Z., & Mazurowski, M. A. (2018). Hierarchical convolutional neural networks for segmentation of breast tumors in mri with application to radiogenomics. *IEEE Transactions on Medical Imaging, 38*(2), 435–447. https://doi.org/10.1109/TMI.2018.2865671.

Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D. P. & Chen, D. Z. (2017a). Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In International conference on medical image computing and computer-assisted intervention (pp. 408–416). Springer. https://doi.org/10.1007/978-3-319-66179-7_47.

Zhang, Z., Xie, Y., Xing, F., McGough, M., & Yang, L. (2017). Mdnet: A semantically and visually interpretable medical image diagnosis network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6428–6436).

Zheng, H., Gu, Y., Qin, Y., Huang, X., Yang, J. & Yang, G. -Z. (2018). Small lesion classification in dynamic contrast enhancement mri for breast cancer early detection. In International conference on medical image computing and computer-assisted intervention (pp. 876–884). Springer. https://doi.org/10.1007/978-3-030-00934-2_97.

Zhou, Y., Xu, J., Liu, Q., Li, C., Liu, Z., Wang, M., Zheng, H., & Wang, S. (2018). A radiomics approach with cnn for shear-wave elastography breast tumor classification. *IEEE Transactions on Biomedical Engineering, 65*(9), 1935–1942. https://doi.org/10.1109/TBME.2018.2844188.

Zhu, Z., Albadawy, E., Saha, A., Zhang, J., Harowicz, M. R., & Mazurowski, M. A. (2018). Breast cancer molecular subtype classification using deep features: Preliminary results. *Medical imaging 2018: Computer-aided diagnosis* (volume 10575, pp. 105752X). International Society for Optics and Photonics. https://doi.org/10.1117/12.2295471.