Neural Computing and Applications (2022) 34:1815–1836
https://doi.org/10.1007/s00521-021-06804-y

REVIEW

Deep convolutional neural networks for computer-aided breast cancer diagnostic: a survey

Parita Oza[1] · Paawan Sharma[1] · Samir Patel[1] · Pankaj Kumar[2]

Received: 21 May 2021 / Accepted: 26 November 2021 / Published online: 11 January 2022

Abstract

Advances in deep learning networks, especially deep convolutional neural networks (DCNNs), are causing remarkable breakthroughs in radiology and imaging sciences. These advances have influenced the development of computer-aided diagnosis (CAD). This study presents applications of DCNNs for computer-aided breast cancer diagnosis. We discuss the recent breakthrough, achievements, and notable advances in CAD for breast cancer. Various key and novel insights and challenges on the use of DCNNs for mammogram analysis have been presented in the paper. The latest deep learning toolkits and libraries that are available and insights for using them have been elaborated. We also point out the possible limitations in the use of DCNNs for breast cancer detection. Finally, give some ideas of future research which can address the existing limitations.

Keywords Breast cancer · Mammograms · Computer-aided diagnosis · Deep convolutional neural networks · Deep learning

1 Introduction

Breast cancer is the most commonly detected cancer among the female population worldwide. It is also the leading cause of death amongst females [1, 2]. There have been some reports of breast cancer amongst men, but we have chosen to focus on females in this survey. There are various imaging modalities for breast cancer screening and diagnosis. Mammography (the process of imaging the breast with x-rays) is one of them. It is usually preferred over other imaging modalities by medical imaging experts due to its less imaging time, better resolution, and high SNR (Signal to Noise Ratio). Due to the efforts made by the scientific community in analyzing mammograms, diagnostic accuracy has improved significantly [3] and this has contributed considerably towards increasing the survival rate through early detection of the disease [4]. For breast abnormality diagnosis, clinicians need to segment and categorize suspicious lesions as either benign (non-cancerous) or malignant (cancerous). This process of diagnosis is arduous, time-consuming, routine, and tedious for medical image analysts. Furthermore, the analysts/practitioners are overburdened with the increasing number of mammograms to analyze [5–7]. The prevalent hypothesis is that by using artificial intelligence (AI) for mammogram analysis, the process of computer-aided diagnosis (CAD) can be made interesting, effective, and free from subjectivity. Computer-aided diagnosis (CAD) for breast cancer has already started experiencing the arrival of a progressive artificial neural network that is Deep Learning (DL). Figure 1 portrays the general architecture of deep learning and its training process.

Deep convolution neural networks (DCNNs)-based computer-aided diagnosis (CAD) systems for breast cancer have been developed in the past few decades and act as a decision-making system. It also differentiates between cancerous and non-cancerous lesions by providing additional information [8]. These systems have often been considered a second reference tool to assist the radiologist in making proper clinical decisions. The CAD could improve breast lesion diagnosis's overall correctness,

✉ Parita Oza
parita.prajapati@nirmauni.ac.in;
parita.ophd19@sot.pdpu.ac.in

[1] Pandit Deendayal Energy University, Gandhinagar, India

[2] Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India

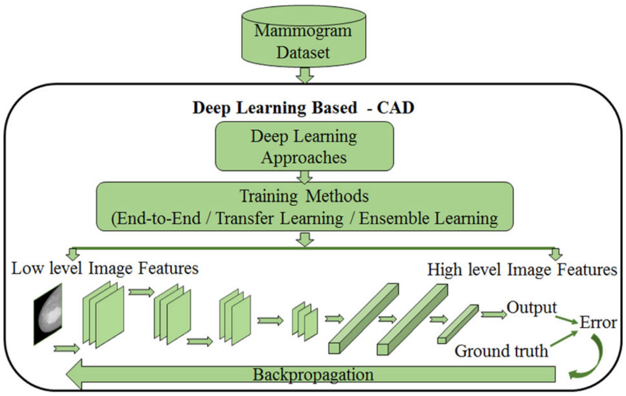

Fig. 1 Deep learning architecture and learning process for computer-aided diagnosis. Features maps are generated using multiple filters. Layers are flattened into vectors before generating the final output. Parameters of feature maps are updated by calculating errors through backpropagation

sensitivity, and specificity. This system can also significantly reduce the number of false positives due to oversight errors [7]. There have been several studies in the literature exploring the usage of CAD systems for the diagnosis of breast cancer, out of which some have already gone through clinical testing [9].

Lesion detection from mammograms is an important image processing task; the CAD system further analyzes these detected lesions. This task is very much challenging due to anatomical variability of masses like size, shape, as well as the location of a lesion in a mammogram [7, 10]. Earlier CAD systems used to do manual detection by handcrafted features to identify doubtful masses using conventional machine learning approaches [7, 11–14]. This practice of manual feature extraction has resulted in more false positives [15]. Image has many details; hence, it becomes very challenging to identify the correct set of features for detection. Deep learning approaches have overcome these weaknesses of conventional machine learning approaches by learning the features of an object during the training phase [16]. Many novel methods of lesion detection based on deep learning are introduced for the development of the CAD system [3, 17]. Figure 2 shows the generalized architecture of the CAD system for breast cancer diagnosis.

Breast mass segmentation is also an important application that extracts discriminative features of particular lesion regions [7, 17]. However, there is a solid association between masses in the breast and their irregular, low contrast and vague boundaries, which makes mass segmentation a big challenge [18]. In literature, methods like region growing, active contour, and chan-vese method have been used for mass segmentation [18–20]. But due to handcrafted features, these methods could not learn complex variations in shape [7, 19]. Few studies based on deep

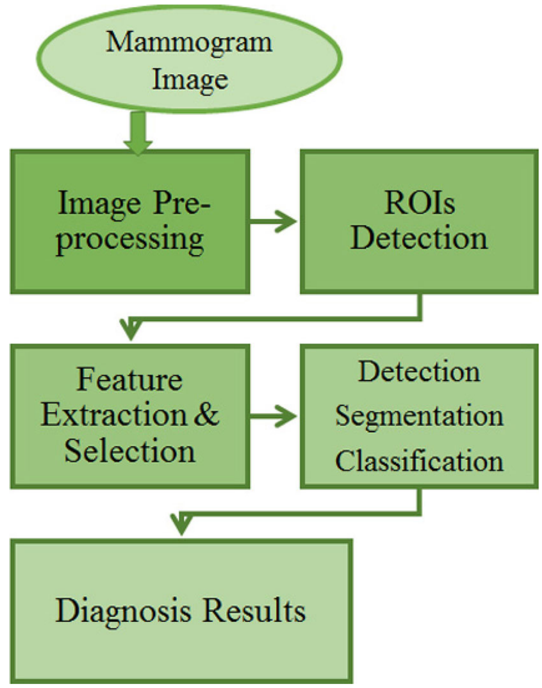

Fig. 2 Generalized architecture of CAD for breast cancer diagnosis

learning models have shown better output as compared to traditional segmentation methods by automatic extraction of high-level features for the segmentation process [5, 15, 21, 22].

Most of the CAD systems are developed for mass classification to differentiate cancer as either malignant or benign using conventional machine learning algorithms [11–14]. Deep learning approaches, especially DCNNs, have shown powerful enhancements in breast cancer classification in the past decade. These approaches can extract meaningful background and texture features, which allows the algorithm to distinguish between normal and abnormal lesions, which may vary in shape, size, as well as orientation [13]. By learning and extracting deep non-linear features from raw input images directly, these approaches attain much better classification output as compared to traditional machine learning approaches [5–7, 23].

Deep learning approaches need a substantial amount of data as compared to machine learning approaches. There is a need for a proper mammographic database for proper training, validation, and testing of deep learning-based CAD for breast cancer diagnosis. Publicly available mammogram datasets mainly include DDSM [24], MIAS [25], INBreast [26], BCDR [27], IRMA [28] and SuReMaPP [3]. These datasets include all types of abnormalities such as mass, microcalcification, and architectural distortion. Annotations are available either at pixel-level boundaries of abnormalities or at the center of suspicious regions's radius. Some datasets, such as MIAS and

SureMapp, come with files containing the class of abnormality and other important details. Digital mammogram images are originally available in DICOM format with some metadata that is not useful for deep learning methods; hence we need to extract just the image matrix out of each file in the set. Some datasets have different formats like PGM (potable grey map) and LJPEG (lossless JPEG). Image size and resolution, as well as abnormality distribution, may vary in these datasets. MLO (mediolateral oblique) and CC (craniocaudal) are two standard views that encompass regular screening mammography. MLO is the most important projection because it allows portraying most breast muscles. CC view must show the medial and external lateral portions of the breast as much as possible. Datasets like DDSM, INBreast, and IRMA are available with both these views.

This paper presents a study of hundred-plus articles from reputed journals, books, and conference proceedings out of major leading academic databases, including Scopus, Web of Science, ACM digital library, and IEEE Xplore. In the scientific literature, Shrewd and comprehensive surveys on deep learning-based breast cancer diagnosis systems are present. For example, Zou et al. [8] review the recent progress of CNN models for mammographic breast cancer diagnosis. It aims to offer many clues on using CNN for the related task in the domain. Another review by Burt et al. [9] also presents recently developed deep learning-based CAD systems for breast cancer diagnosis with their advantages concerning previously established systems. Our paper's main scope is to let readers embark on a journey through a fully comprehensive description of various pre-trained DCNN architectures and their applications for breast cancer diagnosis, such as cancer detection, classification, and segmentation. Further, we discuss different key and novel insights on the use of DCNN for breast mammograms. This paper elaborates on the latest deep learning toolkits and libraries available and insights for using them. The article also discusses possible limitations in using DCNNs for breast cancer diagnosis and future research directions.

The paper is structured as follows: In Sect. 2, we define the study criteria for the survey. Sections 3 and 4 discuss basic building blocks of DCNNs and popular DCNN architectures, respectively. Section 5 describes the contribution of DCNNs for different tasks in breast cancer diagnosis: detection, classification, and segmentation. Section 6 deals with key challenges in the imaging and diagnostics process and proposed solutions. Section 7 presents open-source tools deep learning libraries. Section 8 states research directions and opportunities. We end with the conclusion in Sect. 9.

2 Methods

The research focus is mainly on various DCNN architectures and their applications for breast cancer diagnosis. This study also focuses on various building blocks of DCNN, pre-trained DCNN architectures for mammogram images, challenges, and techniques for the improvements. Additionally, the study presents popular and publicly available mammogram datasets and deep learning libraries available to the public. Table 1 mentions the criteria for this study. Readers of this study will be able to get answers to the following research questions:

1. Which last layer activation functions are commonly used for specific classification tasks?
2. What are the most popular pre-trained DCNN models used for breast lesion or mass detection, segmentation, and classification? What are their configurations?
3. Which are the most common applications of DCNNs for breast imaging and diagnostic?
4. Does building a DCNN model with a small dataset result in any adverse effect?
5. How does an imbalanced dataset affect the performance of the DCNN model?
6. What are the key challenges for training the DCNN model and the possible solutions to handle it?
7. What are commonly used deep learning toolkits and libraries for mammogram images?
8. Are there still any open issues that need further investigation?

3 Deep convolutional neural networks (DCNNs)

The most significant advancement in the field of Artificial Intelligence is deep learning approaches, especially DCNNs. DCNN is very much useful for processing images [29]. It is designed to adopt spatial hierarchies of features from low-level to high-level patterns automatically. Deep convolutional neural networks can perform feature extraction and classification with the help of their different layers. Layers like convolutional and pooling help in

Table 1 Criteria for the Study

	Study Criteria
Keywords	BreastCancer,Mammograms,DeepLearning, DeepConvolutionalNeuralNetworks
Research Focus	DCNN Based Breast Cancer Diagnosis, Applications of DCNN in BreastCancer Diagnosis

extracting non-linear features from the image wherein fully connected layers map these extracted features into the final result called classification. Digital images are stored as a 2D array, and a small kernel is applied at each image position. This is the reason why DCNNs are most suitable for processing images since features can occur at any place in the image [8, 29, 30]. DCNN has a stack of convolutional layers, and the output of one layer passes on to the next layer. This hierarchy of layers progressively extracts more complex features. The entire network is trained using the training set of data. Training a network means minimizing the loss, which is the difference between output obtained and available ground truth. Backpropagation and gradient descent are generally used as optimization algorithms.

3.1 Integrants of DCNN

A simple convolutional network is a stack of layers, and every layer transforms from one volume of activation functions to another with the help of a differentiable function. DCNN is equipped with three layers: Convolutional Layer, Pooling Layer, and Fully connected layers (FC layers). These layers build convolutional network architecture. Figure 3 shows the basic architecture of a deep convolutional neural network.

Convolutional layers: An important building block of DCNNs is a convolutional layer. This layer has a set of feature detectors called kernels or filters, which apply across the entire image area. Each of these feature detectors has a defined group of learnable weights. These weights do not change across the image. For example, for 3×3 feature detector, we use the same nine weights for computing the activation of every neuron in output. This is called parameter sharing, and it is used to reduce memory requirements and prevent the effect of overfitting. For a feature detector with width f_w and height f_h, an input volume with the depth of D and an output volume with the depth of O, Equation 1 computes the total number of weights W in a convolutional layer.

$$W = (D * f_w * f_h + 1) * O \tag{1}$$

Hyperparameters like stride, depth, and padding are used to control the size of output volume. Stride moves a feature detector. For stride value 1, the feature detector moves one pixel at a time. This number can vary and can produce different output volumes. Depth states a number of feature detectors and each of which is observing different features in an image. For many applications, it is necessary to preserve the spatial size of output volumes precisely the same as the input volume. Padding can help here, which pad the input volume with zeros around the borderline of an image to preserve the input image size [29]. This process is called zero padding. Equation 2 can compute the spatial size of the output volume.

$$f_{out} = (f_{in} + 2P - f)/S + 1 \tag{2}$$

where,

f_{out}: Number of output features
f_{in}: Number of input features
S: Size of stride
P: Size of padding

Pooling Layer: Similar to the convolutional layer, vector to scalar transformation, which operates on every region of an image, is done by the pooling layer. Pooling layers do not have filters; it does not compute dot products with the local region of an image like in the convolutional layer. It averages the pixels in the image region called average pooling or selects the pixels with higher intensity while removing the rest of the pixels called max pooling. The idea of pooling may not seem productive as it leads to information loss. Still, it is proven that it makes DCNN invariant to variations while presenting the images, and also it diminishes background noise. Ideally, the size of images at every layer in the model is directly proportional to the computation cost of every layer. The role of pooling is to reduce the image dimensions as the network goes

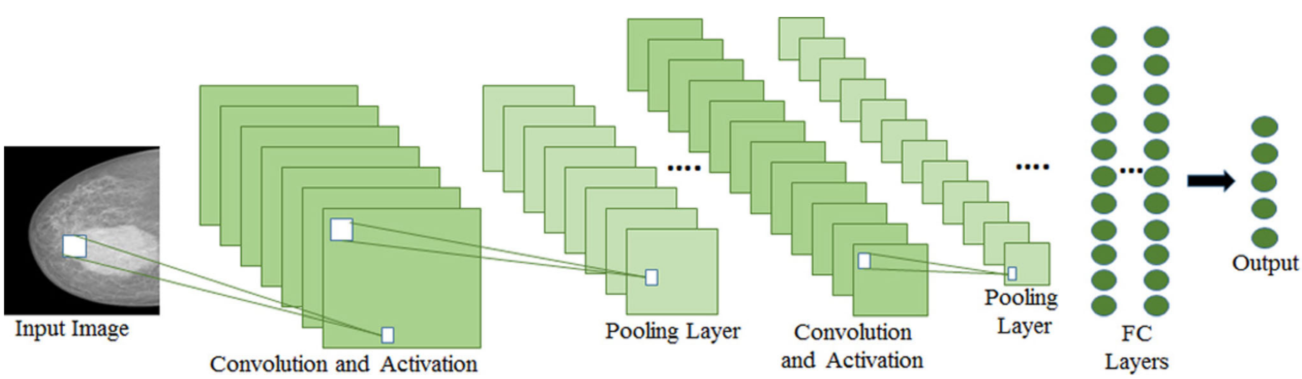

Fig. 3 Basic DCNN architectures

deeper. This could help prevent the outburst of the number of floating-point operations per second a network needs.

Activation Functions: Activation functions are used to add non-linearity into a model. Also, these functions are used to check model accuracy and determine the computational efficiency of a training model. Mathematical representation of non-linear functions like sigmoid and hyperbolic tangent is similar to biological neuron behavior [29]. Sigmoid activation may result in a vanishing gradient problem. This function squeezes an ample space of an input into a small space and normalized between 0 and 1. Hence, a considerable variation in the input will cause a minimal variation in the output. Due to this, derivatives become small. Gradients in deep neural networks are found using backpropagation through hidden layers. These may slow the learning in the initial layers of the model, which is not recommended in deep learning due to costly computations [30, 31]. Equation 3 gives the mathematical definition of the sigmoid activation function.

$$\sigma(x) = 1/(1 + e^{x}) \tag{3}$$

ReLU has become a very much popular activation function in deep learning as it can solve the vanishing gradient problem. ReLU can be mathematically defined by equation 4.

$$f(x) = max(0, x) \tag{4}$$

ReLU activation has many variants like leaky ReLU, randomized ReLU and parametric ReLU [30, 32]. Equation 5 defines Leaky ReLU. Parametric ReLU and Randomized ReLU are types of leaky ReLU. Their definition is the same as that of leaky ReLU, except for the value of a.

$$f(a, x)) = \left\{ \begin{array}{cc} x, & \text{if} x > 0 \\ ax, & x <= 0 \end{array} \right\} \tag{5}$$

where a = 0.01.

The gradient value in ReLU is zero for negative input values and one for positive input values. For values above zero, gradient value results into one, making the network learn in any way. But for values below zero, gradients turn into zero, and hence, there is no learning. This problem of ReLU is sometimes called dying ReLU [30]. In the dying ReLU problem, some ReLU neurons die for all input due to no gradient flow. This problem can be solved by leaky ReLU [30, 33]. In leaky ReLU slope is changed to 0.01 or so, which causes a leak and extends the range of ReLU. Authors of papers [31, 32] have given an in-depth explanation of various activation functions. ReLU activation is a common choice in many DCNN models [34–36].

Fully connected and last layer activation [30, 37]: Convolutional layers are used for feature extraction. Features extracted by these deeper layers are extremely abstract and also not human readable. Fully connected layers can solve this problem. The input to the fully connected layer is an output of the final convolutional or pooling layer, which is then flattened and connected to the final output layer. Fully connected layers learn how to use features extracted by deeper layers in order to properly classify the images. This layer essentially learns non-linear combinations of extracted high-level features. We can think of fully connected layers as translators between the model's language and ours. These layers act as a classifier on the top of the features extracted by convolutional layers and then assign the class probability of the input image. These layers are sometimes called dense layers, where there is a direct connection between every input and output. Also, each of these connections has learnable weights. Following every fully connected layer, there is a non-linear activation function, as discussed previously. Ideally, the number of output nodes in the last fully connected layer is the same as the number of classes for a particular application. The activation function of the last fully connected layer may vary for different applications. Table 2 presents some common last layer activation functions used for various applications.

4 Popular DCNN architectures

DCNNs have been used extensively for medical imaging tasks like detection, classification, and segmentation of various cancer diseases. Unfortunately, training a network from scratch may take many days and weeks and needs a lot of computational power. However, various pre-trained networks are already available and widely used by the research community. This section presents an overview of such networks along with their parameter-based comparison.

4.1 DCNN architectures for breast cancer detection and classification

AlexNet: AlexNet [38] is a popular DCNN architecture that has shown revolutionary results for recognizing and classifying images. This network is the winner of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 and is considered the first successful application of DCNN for the voluminous dataset. The network has a relatively simple layout as compared to modern DCNN architectures (See Fig. 4).

Authors of this framework came up with key ideas like local response normalization, ReLU activation, overlapping pooling, and data augmentation. Data augmentation will be defined later in the paper in Sect. 6.1.2. AlexNet was trained on multiple parallel GPUs to overcome

Table 2 Last layer activation functions for a specific task

	Subcategory of task	Last layerActivation
Classification Task	Binary Classification	Sigmoid Activation
	Multitask Single Class	Softmax Activation
	Multitask Multi Class	Sigmoid Activation
Regression Task	–	Linear / Identity Activation

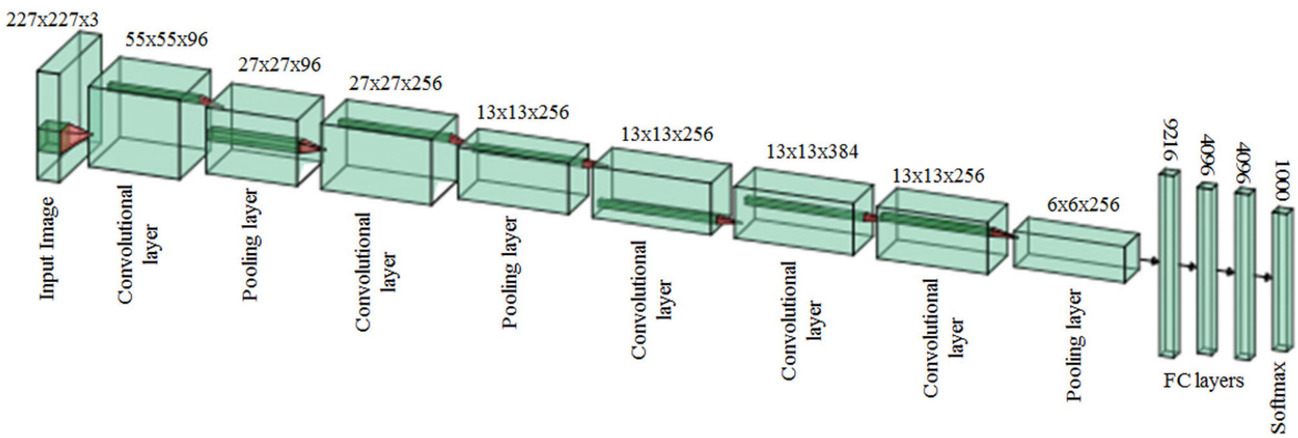

Fig. 4 AlexNet architecture with the receptive field at every layer

shortcomings of the hardware. ReLU activation is used to add non-linearity into the network. Authors have added local response normalization after the first and second convolutional layers to add generalization into the network and encourage lateral inhibition. The concept of overlapping pooling layers is used for the first time by this network to summarize the outputs of neighboring neurons in the same feature map. In AlexNet, deepness is extended to 8 layers (which was five layers in LeNet [39]) that have resulted in an overfitting problem. Authors of Alexnet have used data augmentation techniques like image translation, image reflection, and changing intensity of RGB channels to defeat the problem of overfitting. AlexNet has a very efficient learning attitude, which has made it very popular in the research community, especially for medical imaging tasks. A new era of research in the architectural expansion in DCNNs has started by this network [40].

VGGNet: After the success of AlexNet, the image recognition and classification task have speedup the exploration in the architectural expansion of DCNNs. Simple and uniform yet very effective architecture of DCNN called VGGNet is presented in [41]. Figure 5 depicts the general architecture of VGGNet.

Authors have proposed a total of five configurations of this model, out of which VGG 16 and VGG 19 are very popular. Input to this architecture is 224×224 RGB image. Authors have used 3×3 filter throughout the network, making it very simple and uniform. Authors of VGGNet have practically demonstrated that a stack of three 3×3 convolutional layers with stride 1 have the same receptive field as one 7×7 convolutional layer. This model also demonstrated that filters with such a small receptive field could result in low computational costs while decreasing the number of parameters. It is fascinating to know that after every pooling layer, the total number of filters doubles. This innovation has reduced spatial dimension while growing the depth of the network. VGGNet stood second in classification in ILSVRC'14 and first in localization. The only drawback related to VGGNet is the number of parameters (138 million). Due to which it becomes prolonged on systems with less computational power [40, 41].

GoogleNet: In 2015, Authors of GoogleNet, also called Inception-V1, came up with the idea of "going deeper with computation efficiency" [42]. This model has a total of 22 layers. This model stood first in ILSRVRC'2014 for the image classification task with the top 5 errors of 6.67%. The main innovation in the model was the introduction of inception modules in DCNN. An inception module is a block of a model that aims to estimate an optimal local sparse structure in a convolutional neural network. It permits several filter sizes (1×1, 3×3, 5×5, etc.) instead of being constrained to a single filter size in a single block, which is then concatenated and passed to the next layer in the network. However, GoogleNet has resulted in costly computation after concatenation operation. Authors have solved this problem by adding a bottleneck layer with 1×1 convolutional to reduce feature depth. GooleNet has no

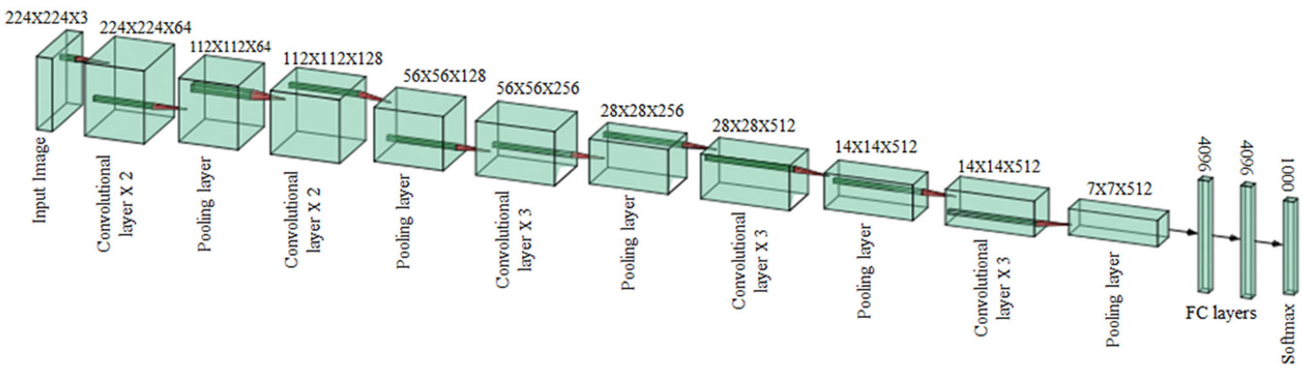

Fig. 5 VGG Architecture with the receptive field of every layer

fully connected layers at the end; instead, it has used global average pooling at the last layer to minimize the connection density. This model has a total of five million parameters, which is very few compared to AlexNet (60 million) and VGGNet(138 million). However, the heterogeneous topology of GoogleNet is its main downside. Also, the use of the bottleneck layer significantly decreases the feature map for the next layer, and sometimes it may result in loss of essential data [40, 42]. Figure 6 demonstrates the basic architecture of the inception module with split and merge concepts.

ResNet: "When a network with higher depth starts converging, a degradation problem is exposed. With a deeper network, accuracy may get saturated, and the network degrades the performance. It is important to notice here that this degradation is not caused due to overfitting." This was stated in [43, 44] and practically demonstrated in [45]. The deeper model should be at least able to perform as well as the shallow network. The authors of ResNet have shown this an optimization problem and proposed a solution by constructing a block called a residual block. The purpose of this block is to copy learned layers from the shallow network and set additional layers to identity mapping. The approach is called skip connection which allows the flow of information into the network even if some layers could not learn, i.e., the flow of information from one layer to the next's next layer. Figure 7 depicts the basic unit of ResNet architecture. ResNet has a stack of

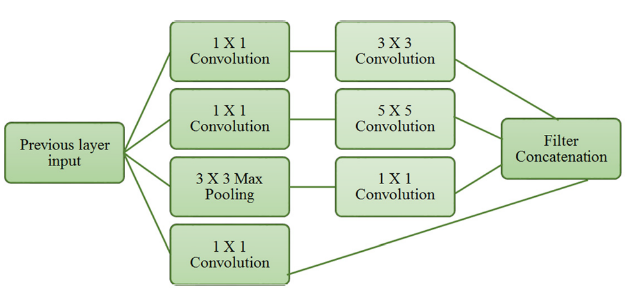

Fig. 6 GoogleNet inception module with split and merge concept

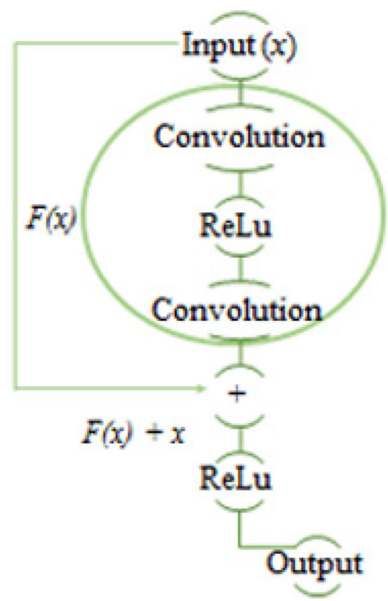

Fig. 7 Basic unit of ResNet: residual block

residual blocks, each with two 3×3 convolutional layers. A network periodically doubles the number of filters and downsample spatially using stride 2. Authors of ResNet have shown that, as compared to a plain 34 layers network, ResNet (with 50/101/152 layers) has less error on the classification process. This model was the winner of image classification in ILSVRC'15 with a top 5 error of 3.57%.

Inception V2, V3, V4 and Inception-ResNet: Inception-V2 [46], V3 [46], V4 [47] and Inception-ResNet [47], are upgraded versions of Inception-V1. Inception-V2 is an earlier version of Inception-V3, which is identical to inception-V3 but not generally used. The authors of Inception-V2 made some architectural changes like changing the loss function, optimizer, and adding batch normalization to the auxiliary layers and came up with the new version, which became popular as Inception-V3. Bottleneck layers used in Inception-V1(GoogleNet) may extremely reduce the dimensions of an input image for the next layer, resulting in the loss of important information. The main motivation behind Inception-V2/V3 was to avoid representational bottlenecks and the use of factorization techniques. Authors have provided various principles and optimization techniques for scaling up the network in a well-organized way. Two concepts of the factorization are introduced by the authors, factorization into smaller convolutions (Factorize 5×5 convolution into two 3×3 convolutions and factorize 7×7 convolutions into a series of 3×3 convolutions) and spatial factorization into asymmetric convolutions (Factorize $n \times n$ convolutions into asymmetric convolutions: $1 \times n$ and $n \times 1$ convolutions). People from Google again came up with an improved version of Inception-V3, which is Inception-V4 [47]. The main difference is the uniform choices for all inception modules for all sizes of grids. In the same paper, authors have introduced a new variant in this family called Inception-ResNet. This model combines two concepts, inception block and residual learning. Authors have practically proved that adding residual blocks in Inception-V4 has a similar generalization as in Inception-v4. Authors have shown that adding residual connections quickens the training of Inception networks considerably.

Xception [48]: One more adoption of Inception is Xception. Authors of this model have used depthwise separable convolution by replacing inception modules. Different spatial dimensions such as 1×1, 3×3 and 5×5 of inception module are replaced by single 3×3 convolution followed by 1×1 convolution to control complex computation in the network. Xception takes the hypothesis of Inception to its extreme, 1×1 convolutions take cross-feature map correlations and regular 3×3 convolutions take spatial correlations within every channel. Xception took this idea to an extreme meaning: it performs 1×1 to every input channel and then performs 3×3 to each output channel. The change adopted by Xception could not reduce the number of parameters, but it could help in more efficient learning and improve the model's performance.

You Only Look Once(YOLO): An object detection architecture based on DCNN called YOLO is proposed in [49]. This architecture is used for real-time object detection, but several studies show that YOLO is also used for breast cancer detection. GoogleNet inspires this network. There is a total of 24 convolutional layers following two fully connected layers. The inception module of GoogleNet is replaced by 1×1 reduction layer, which is following 3×3 convolutional layers. This architecture splits input images into $m \times m$ grid, and for every grid, it generates two bounding boxes and probabilities of class. Each of these bounding boxes has five predictors (x, y, w, h, and confidence). (x,y) coordinate pair represent the center of the box, w and h represents width and height, respectively. At last, confidence shows the intersection over union (IOU) between ground truth and the predicted output. YOLO is very fast and reasons globally regarding the input image while making predictions. However, there are some limitations of YOLO. For example, detecting small objects is difficult. Hence this architecture may not be successful for the detection of microcalcifications or very tiny lesions from mammograms.

Table 3 summarises pre-trained DCNN architectures. Table 4 shows the strength and limitations of these models.

4.2 DCNN architectures for breast cancer segmentation

Fully Convolutional Network (FCN): A popular DCNN-based image segmentation method used for various medical imaging tasks is presented in [50]. This architecture has many convolutional layers, taking any arbitrary image size and producing a corresponding segmentation map of the same size. This model uses both down and upsampling to get contextual information and recover spatial information, respectively. FCN uses a concept of skip connections which combine the final prediction layer with finer strides to recover the spatial information loss due to downsampling. This work is considered a breakthrough to segment images using a deep learning network on arbitrary size images. Though it is a popular model, it still has some restrictions. The model is not easily transferable to 3D images and does not consider global context information effectively.

UNet: U-Net [51] is built upon the architecture of a fully convolutional network. U-Net architecture is modified to work with less training data and yet gives accurate segmented output. A significant modification in this architecture is that upsampling operations have large feature channels to allow the network to pass on information to higher layers. The segmentation map produced by this model only holds image pixels. The benefit of this strategy is to have unified segmentation for even arbitrarily large images. The name of this architecture is due to its symmetry shape, which differs from other variants of fully convolutional networks. Figure 8 shows U-Net architecture designed with 3 parts: The contracting or downsampling path, bottleneck, and the expanding or upsampling path. The first part is typical DCNN architecture with 3×3 convolutions, each of which is following 2×2 max pooling with stride value 2. The number of feature channels doubles at every downsample step. Expanding path contains upsampling of the feature map, which follows 2×2 convolution that divides the number of feature channels, concatenation with the corresponding cropped feature map and two 3×3 convolutions followed by ReLU activation. Finally, 1×1 convolution is used at the last layer to map the feature vector to the desired class. There are total of 23

Table 3 Comparision of popular pre-trained CNN architecture

	AlexNet	VGGNet	ResNet	GoogleNet	InceptionV2/V3	InceptionV4	Inception-ResNet	Exception
Introduction Year	2012	2014	2014	2015	2015	2016	2016	2017
Image Size	227×227	224×224	224×224	224×224	299×299	299×299	299×299	299×299
Number of Layers	8	16/19	22	50/101/152	159	70	572	126
Size of Filters	3, 5, 11	3	1, 3, 5, 6	1, 3, 7	1, 3, 5	1, 3, 7	1, 3, 7	3
Regularization Technique	Data Augmentation,	Data Augmentation,	Data Augmentation,	Data Augmentation,	Data Augmentation,	Data Augmentation,	Data Augmentation,	Data Augmentation,
	Dropout	Dropout	Dropout	Batch Normalization	Dropout	Dropout	Dropout	Dropout
Last Layer Activation	Softmax	Softmax	Softmax	Softmax	Softmax	Softmax	Softmax	Softmax
NewFeatures	Overlapping Pooling, Local Response Normalization, ReLU	Uniform filtersize (3×3) throughout the network	Inception Module with split and mergeconcept	Skip Connection Residual Block	Filter Factorization	Uniforminception modules	Converting Inception modules to Residual Inception blocks	Depth wise separable convolution layers

'Dropout' will be defined later in the paper in Sect. 6.1.2

Table 4 Model Comparison based on strength and limitation

	Strength	Limitation
AlexNet	Extraction of low, medium and high-level features by using variable size filters, Use of parallel GPUs, Deep and wide network	Aliasing effect in feature map because of large size of filters
VGGNet	Homogeneous network, Effective receptive field	Large computation and a large number of parameters due to fully connected layers
ResNet	Residual learning, Low error rate for deep model	Longer training time, Over adaption of hyperparameters
GoogleNet	Use of multi-scale filters within layers, Use of bottleneck layer to reduce the number of parameters	Complex structure and information loss due to bottleneck layer
Inception V2/V3	Exploited asymmetric filters and bottleneck layer to lessen the computational cost of deep architectures	Complex architecture design, Lack of homogeneity
Inception V4	Deep hierarchies of features, Multilevel feature representation	Slow in learning
Inception-ResNet	Combined the power of residual learning and inception block	
Xception	Depth-wise separable convolution is introduced Use of cardinality to learn good abstractions	High computational cost

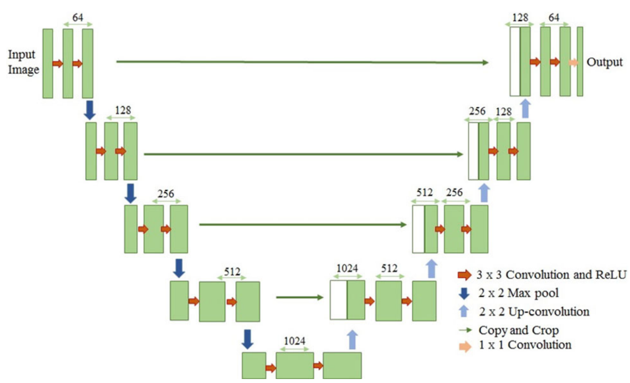

Fig. 8 Unet architecture

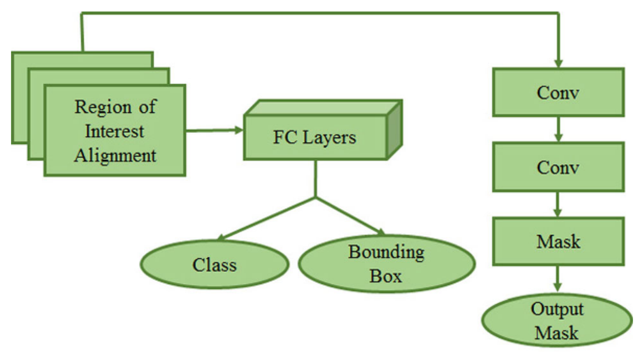

Fig. 9 Mask-RCNN architecture

convolutional layers in this network. UNet is designed and trained specially for medical images.

Mask RCNN: The Architecture of Mask RCNN presented in [52] is the combination of features of FCN, and faster RCNN [53]. The model has three branches to compute the coordinates of the bounding box, associated class, and mask for object segmentation. The loss function of mask R-CNN combines losses of all these three branches and train them together. The output of the pooling layer feeds into all three branches. First, two branches generate predictions of bounding box coordinates and associated class, and the third operates on the region of interest to predict the mask for the object. Figure 9 presents the architecture of Mask-RCNN.

5 Applications of DCNN in breast cancer diagnosis

5.1 Breast mass detection and classification

Breast mass detection and classification from mammogram images are the most popular research topics. Many algorithms have been developed for automated breast mass detection and classification, and researchers are still exploring the same. Deep Learning has solved the problems of conventional machine learning algorithms by replacing feature engineering with feature learning. This section presents various DCNN-based approaches used in the literature to detect and classify breast masses from mammogram images. Table 5 shows a summary of papers based on various DCNN architectures for the detection and classification of breast abnormalities. A team of researchers proposed a new method in [3] to detect suspicious regions in mammograms using scale invariant feature transform

Table 5 Summary of papers on breast cancer detection and classification

References	Year	Classifier/Model/ Method	Type of abnormality	Task performed	Dataset	Model Performance
[3]	2020	Scale Invariant Feature Transform, PyramidNet, AlexNet	Mass	Detection	SuReMaPP MIAS	SuReMaPP: 98% of sensitivity and 90% of specificity MIAS: 94% of sensitivity and 91% of specificity
[6]	2018	YOLO, Fully Connected Neural Networks (FCNNs)	Mass	Detection Classification	DDSM	Detection Accuracy: 99.7% Classification Accuracy: 97%
[7]	2018	YOLO, FrCN, DCNN	Mass	Detection Classification	INBreast	Accuracy: 95.64%
[15]	2017	DCNN	Mass	Detection Classification	INBreast	Accuracy: 90%
[21]	2017	DCNN	Mass	Detection	Private	Accuracy: 85.2%
[54]	2017	Semisupervised DCNN	Mass	Detection	Private	Accuracy: 82.43%
[55]	2018	Faster R-CNN	Mass	Detection	INBreast	Accuracy: 95%
[56]	2018	DCNN	Mass	Detection	DDSM	Accuracy: Non-dense region: 91% Dense region: 94.8%
[57]	2018	DCNN	Mass	Detection	Private	Accuracy: 86.81% Sensitivity: 86.6% Specificity: 87.5% AUC: 0.87
[58]	2019	GoogleNet, VGGNet, ResNet	Mass	Detection Classification	Private	Accuracy: 97.67%
[59]	2019	Deep active learning, Self-paced learning	Mass	Detection	Private	AUC: 0.92
[60]	2020	Multi depth CNN	Mass	Detection	INBreast	Sensitivity: 83.54%
[61]	2020	Faster R-CNN	Mass	Detection	Private INBreast	Private: TPR of 0.91 $\pm$ 0.06 at 1.69 FPI INBreast: TPR of 0.99 $\pm$ 0.03 at 1.17 FPI for malignant 0.85 $\pm$ 0.08 at 1.0 FPI for benign masses
[62]	2020	YOLO, InceptionResNetV2	Mass	Detection Classification	DDSM INBreast\	Accuracy: DDSM: 99.17% INBreast: 97.27%
[63]	2020	Bilateral image analysis based Convolution Neural Network	Mass	Detection	INBreast Private	INbreast: 0.88 true positive rate (TPR) with 1.12 false positives per image (FPs/I) Private: 0.85 TPR with 1.86 FPs/I. For TXMD dataset
[64]	2020	FCNN	Mass	Detection	INBreast Private	Private:AUC:0.90, TPR@2.0FPI-0.94 INBreast: AUC-0.85, TPR@2.0FPI-0.87
[65]	2020	YOLO, ResNet and Inception (for feature extraction)	Mass	Detection Classification	INBreast	Accuracy: 89.4%
[66]	2018	VGGNet, ResNet, AlexNet, GoogleNet	Mass Microcalcification	Detection	CBIS-DDSM	Accuracy: VGGNet: 92.53% AlexNet: 91.23% GoogleNet: 91.10% ResNet: 91.80%
[67]	2020	Pre-trained CNN models with global average pooling	Mass	Detection Classification	CBIS-DDSM	AUC: VGG16: 0.70 InceptionV3: 0.74 InceptionResNetV2: 0.76 Xception: 0.75 MobileNet: 0.76
[68]	2020	ResNet50, InceptionV3	Mass	Classification	DDSM	Accuracy: ResNet50: 85.71 InceptionV3: 79.6
[69]	2018	AleXNet, ResNet50	Mass	Classification	Private	AUC: AlexNet: 0.6749 ResNet50: 0.6239
[70]	2019	YOLOv3	Mass	Detection Classification	DDSM	Accuracy: 97%
[71]	2020	DCNN, AlexNet, VGGNet	Mass	Classification	DDSM MIAS Private	Proposed CNN achieves higher AUC and accuracy as compared to the pre-trained network. Accuracy: MIAS: 92.54 DDSM: 96.47 AUC: MIAS: 0.85 DDSM: 0.96

Table 5 (continued)

References	Year	Classifier/Model/Method	Type of abnormality	Task performed	Dataset	Model Performance
[72]	2020	DCNN with region based pooling structures	Mass Microcalcification	Classification	INBreast DDSM	AUC: InBreast: 0.934 DDSM: 0.838
[73]	2017	YOLO	Mass	Detection Classification	DDSM	Detection Accuracy: 96.33% Classification Accuracy: 85.52%
[74]	2017	VGGNet	Mass	Classification	IRMA	AUC:1.0
[75]	2019	AlexNet, VGGNet, GoogleNet, ResNet, InceptionV2	Mass	Classification	DDSM-400 CBIS-DDSM	ResNet outperforms rest of the pre-trained networks. Accuracy: DDSM-400: 0.859 CBIS-DDSM: 0.804\
[76]	2017	Pre-trained CNN Models	Mass Microcalcification	Classification	INBreast DDSM	AUC:0.9

descriptors (SIFT) and pre-trained DCNNs like PyramidNet and AlexNet. SIFT is an image descriptor used for image-based recognition and matching. This descriptor is invariant to geometric transformations such as scaling, rotation, and translation. The SIFT-based method preprocesses the images and extracts features while the deep learning models validate suspicious regions detected by SIFT. This model is trained and tested on two public datasets; SureMapp and MIAS. As compared to various state-of-the-art methods, this model has shown higher performance. CAD system based on YOLO algorithm for breast mass detection is presented in [6]. For mass classification, authors have used fully connected neural networks (FCNNs). Six hundred mammograms from the DDSM repository are used in this study. The authors also have used data augmentation. Results are evaluated as an average of the 5-fold cross-validation. A Fully integrated system that includes all tasks in one system such as detection, segmentation and classification is proposed in [7]. For detection of mass, the YOLO algorithm is used. Segmentation is done using a fully convolutional network, and finally, DCNN is used for the mass classification process. This system could achieve 98.96 % detection accuracy using a four-fold cross-validation test on the INBreast dataset. Authors of [15] have proposed a CAD system for breast mass detection, segmentation as well as classification. The authors have used the Bayesian optimization approach for mass detection. Deep structured output learning refined by the level set method is used for mass segmentation. For breast mass classification, pre-trained deep learning classifiers are used, which are fine-tuned based on the labels of the dataset used. This system could detect 90% of correct masses and has 85% of segmentation accuracy. The model could show the sensitivity of 0.98 and specificity of 0.7. Comparison between various CAD systems for mammogram detection and classification using manually designed feature set and DCNN is developed in [21]. The authors have also compared the results of DCNN to a group of experts in the medical imaging field and have shown that both DCNN and human reader have almost the same performance. A graph-based semi-supervised learning method based on DCNN is presented in [54]. This system needs a small portion of labeled data for the training process. From 1874 pairs of mammogram images, 3158 regions of interest (ROI), each containing a mass are extracted and used in the study. Out of which, only 100 ROIs are considered as labeled data while the rest are considered as unlabeled data. The author of [55] used R-CNN for the detection and classification of a breast mass. The suspected region generated by R-CNN is refined by fine-tuning the hyperparameters. This method could achieve better classification results on the INbreast public dataset. In [56] simple CNN is used to detect masses in mammograms. Model is built to detect mass in dense as well as non-dense regions in the breast. For non-dense breast, a model could achieve an accuracy of 91%, and for dense breast, the accuracy achieved is 94.8%. A computer-aided detection framework for mass detection from digital breast tomosynthesis is presented in [5]. This framework is built using DCNN and operates on a private dataset. DCNN is used to learn complex patterns of 2D slices in the images, and then multiple instance learning with randomized trees technique is applied to classify the images. In [58] a novel deep learning framework using transfer learning is proposed for the detection and classification of breast cancer. Using pre-trained DCNN architectures like

GoogleNet, VGGNet, and ResNet, features are extracted from images. These features are fed into the FC layer for the classification process. The proposed architecture could achieve 97.6% of accuracy. A new learning framework for the detection of mass using deep active learning and self-paced learning is shown in [59]. The author has shown that using active learning, the annotation efforts by the expert can significantly be reduced while improving the performance of the training model. Self-paced learning can also deal with ambiguous data and result in a robust model with higher generalization capabilities. The proposed framework also has reduced the need for annotations of samples by about 20% of all training data. Multi-depth CNN is used in [60] for the detection of lesions in mammogram images. To learn about various image spatial context levels, the authors have used a multi-context ensemble of CNNs. The novelty behind the proposed method was to individually train multiple depth CNNs on image patches and then combine them. A fully automated framework for mass detection in mammogram images is presented in [61]. Faster region-based convolutional neural network is used in this study. The authors used two datasets: INBreast dataset and the large-scale OPTIMUM Mammography Image Database (OMI-DB). In the study shown in [62] an integrated system for breast lesion detection and classification is proposed. The YOLO detector is used for lesion detection, and pre-trained deep learning classifiers such as feed-forward CNN, ResNet-50, and Inception ResNet-V2 are modified and used for the classification process. The authors could achieve promising results using this system on datasets like DDSM and INBreast. The authors have used a 5-fold cross-validation test to evaluate the results. A bilateral mass detection method is developed by the authors of [63] using two networks; Registration net- for registering bilateral mammogram images and Faster RCNN- for mass detection using images that are registered by the registration net. A self-supervised approach is used for registration bet to build a network for learning spatial information between bilateral mammogram images. The model achieved 0.88 true positive rates (TPR) with 1.12 false positives per image (FPs/I). For the density prediction, a task-specific network called a fully convolutional network is proposed in [64]. The network also used adversarial training for the alignment of less annotated features with well-annotated features. The author of this study has stated that the proposed training has proved to converge fast. End to end computer-aided diagnosis system based on YOLO is proposed in [65]. All three versions of YOLO are used and compared for the detection and classification process. Inception V3 and ResNet are then used to extract features to compare their classification performance against YOLO. A classification model based on deep learning methods is developed by the authors of [76]. The work showed that mammograms with multi-view and segmentation maps need not be registered for accurate results when fine-tuned deep learning models are used.

5.2 Breast mass segmentation

Mass segmentation or segmentation of any anatomical structure is the most fundamental image processing task, especially for medical imaging. Segmentation of lesions can be performed by clinicians, which is a very time-consuming task. DCNNs are widely used for this task to save time and avoid oversight errors in the diagnosis process. Figure 10 represents an example of segmentation of benign breast mass of mammogram image. First, the DCNN classifier constructs a probability map of breast mass and a patch of images for mass segmentation. Then, these probability maps and global image context are used for the refinement steps. This section discusses various approaches used in the literature for breast mass segmentation. Table 6 shows a summary of papers on breast abnormality segmentation based on various DCNN architectures.

An end-to-end mass segmentation model comprising a fully convolutional network and conditional random field is proposed in [77]. To defeat the problem of overfitting due to the small size of the dataset, authors have used adversarial training. This model could show better performance when compared with state-of-the-art techniques. A holistically nested edge detection network inspired by DCNN is developed by the authors of [78] for automatic breast muscle segmentation. The authors have modified the network to find contour-like objects in mammogram images. This framework creates a probability map that is used for the initial estimation of breast muscle boundary. These maps are then processed by extracting morphological properties for finding an actual boundary. The result produced by this framework is compared to state-of-the-art methods across four publicly available mammogram repositories. In [79], a new framework for breast mass segmentation using conditional generative adversarial

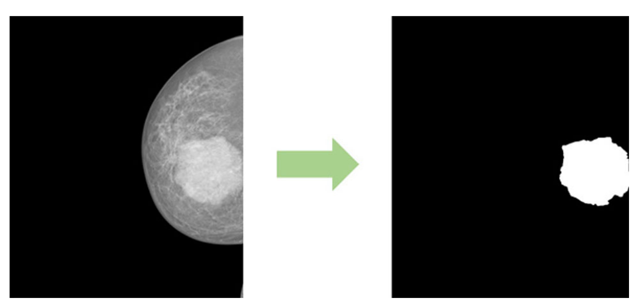

Fig. 10 Mammogram image with mass and corresponding segmented region

Table 6 Summary of papers on breast cancer segmentation

References	Year	Classifiers/Models/Methods	Task performed	Dataset	Model Performance
[15]	2017	DCNN	Mass Segmentation	INBreast,	Accuracy:85%
[77]	2018	Fully Convolutional Network	Mass Segmentation	INBreast, DDSM-BCRP	Dice: INBreast-0.9097 DDSM-BCRP-0.9130\
[78]	2019	CNN based on holistically nested edge detection	Mass Segmentation	MIAS, INBreast, BCDR, CBIS-DDSM	Jaccard Index : MIAS: 94.6 INBreast: 92.6 BCDR: 96.9 CBIS-DDSM: 95.7
[79]	2019	cGAN, Unet	Mass Segmentation	INBreast , Private	Accuracy: INBreast: 92% Private: 88.2%
[80]	2019	Unet	Microcalcification Segmentation	DDSM	F1 Measure: 98.5% Dice: 0.97
[81]	2019	PyramidNet	Mass Segmentation	DDSM, INBreast	Dice: DDSM: 91.10% INBreast: 91.69%
[82]	2019	Modified Unet	Mass Segmentation	DDSM	F1 Score:82.24+0.06
[83]	2020	Histogram Normalization Regression Architecture	Fibroglandular Tissue Segmentation.	Private	Dice:0.84
[84]	2020	Fuzzy sets and fuzzy soft sets	Mass Segmentation	MIAS	Accuracy:88.46%
[85]	2020	cGAN	Mass Segmentation	DDSM, INBreast, Private	Dice:0.94
[86]	2021	DL based crossoverNet	Mass Segmentation	DDSM,INBreast	Dice: DDSM: 0.9250 INBreast: 0.9126
[87]	2020	Unet	Mass Segmentation	INBreast, CBIS-DDSM	Dice: INBreast: 0.8164 CBIS-DDSM: 0.8216
[88]	2020	Unet	Mass Segmentation	Private	Dice: 0.918
[89]	2020	Mask-RCNN	Mass Segmentation	CBIS-DDSM	Precision: 0.65
[90]	2020	End to End CNN	Mass Segmentation	MIAS, CBIS-DDSM, INBreast	Accuracy: 99.64%
[91]	2019	Unet	Mass Segmentation	INBreast	Accuracy: 94.16%

network (cGAN) is proposed. This model learns actual mass images and also a mapping between segmentation masks and images. The authors have also introduced modified U-net architecture for breast mass segmentation. Finding microcalcifications from breast mammograms is a very challenging task as they are very tiny. A computer-based detection of microcalcification is needed for proper diagnosis and to avoid oversight errors. Work in [80] shows automated segmentation of microcalcification from mammogram images. In this approach, following various preprocessing steps, fuzzy C-means clustering divides image patches into positive and negative. Finally, U-net architecture is used to segment the microcalcification areas from the input images. Better performances are achieved by this framework than the state of the art methods. Authors of [81] developed a novel multi-level nested PyramidNet for mass segmentation. This model deals with limitations like intraclass inconsistency and inter-class distinction. PyramidNet contains encoder-decoder pair. The encoder is used to encode low-level details and high-level semantic information. The decoder is used to refine the segmentation results along boundaries of masses. In [82] dense U-net with attention gates (AGs) with encoder-decoder pair is proposed and evaluated on performance matrices like F1-score, sensitivity, specificity, and overall accuracy. In [83] framework for pectoral muscle exclusion and fibroglandular tissue, segmentation is presented. This framework is based on deep learning and histogram normalization. The authors have used regression architectures to learn segmentation parameters. The authors claimed that this framework demonstrates the same performance when compared to two expert practitioners. Intuitionistic fuzzy set and fuzzy soft set-based models for breast segmentation are suggested in [84]. The model showed a segmentation accuracy of 88.46%. cGAN is proposed in [85] to segment breast lesions from the region of interest in a mammogram image. cGAN performs even better with a limited number of training samples. Shape descriptor based on CNN trained on the DDSM dataset is used by the authors to classify the segmented mask as oval, round, lobular, and irregular. To combat the effect of overfitting, the author has used dropout by randomly setting some activation to 0. A new deep learning-based model called Crossover-Net for image segmentation is used in [86]. The features learned from vertical and horizontal directions can provide background information to enhance discriminative ability between different tissues. This observation is very well taken to develop this model. The authors also have made available the Github repository of this project at https://github.com/Qianyu1226/Crossover-Net. In [87] Modified U-net architecture is introduced as a segmentation network for generating segmentation masks. Class unbalancing problem is solved using weighted cross-entropy loss. Model is operated on public datasets such as INBreast and CBIS-DDSM and could result in a dice value of 81.64% and 82.16%, respectively. The dice score can be used to measure model performance. The authors used the dice similarity coefficient to quantitatively represent the segmentation performance of the model on the two datasets, INBreast and CBIS-DDSM.

Figure 11 shows the number of papers with a particular dataset, paper distribution based on applications of DCNNs in breast cancer diagnosis, and a number of papers with a pre-trained DCNN model.

6 Key challenges and techniques for improvements

DCNNs often need a huge amount of properly annotated data for appropriate training. The large-sized medical datasets are not readily available due to data proprietary and some privacy concerns. There is also a shortage of skilled annotation in some datasets. In some datasets, positive samples are underrepresented. This would bias a model for predicting an overrepresented sample or healthy label. Many training samples can be redundant and has no information. All these issues collectively represent challenges in training the DCNN model for breast cancer diagnosis. This section presents key challenges and probable solutions to improve the model performance.

6.1 Small dataset size and overfitting

A key challenge in deep learning is to come up with a model that can perform very well not only on training data but also when new data is presented during testing. A situation where a model learns regular patterns of training dataset but does not provide a generalized output when presented new data is called overfitting, a prevalent issue in deep learning models. Deep learning models can remember the regular patterns in training data [29, 92]. So, instead of learning important features, it may remember inappropriate noise; hence it performs poorly on the test dataset. Overfitted models cannot produce generalized output. Overfitting can be monitored by checking training and validation errors. A deep learning network must be trained using bulky (huge) training data to improve accuracy rates and avoid overfitting. Transfer learning and various regularization techniques can be the solution to this problem [92]. For example, to train DCNNs, Transfer Learning (TL) and data augmentation are proper blends for mammography

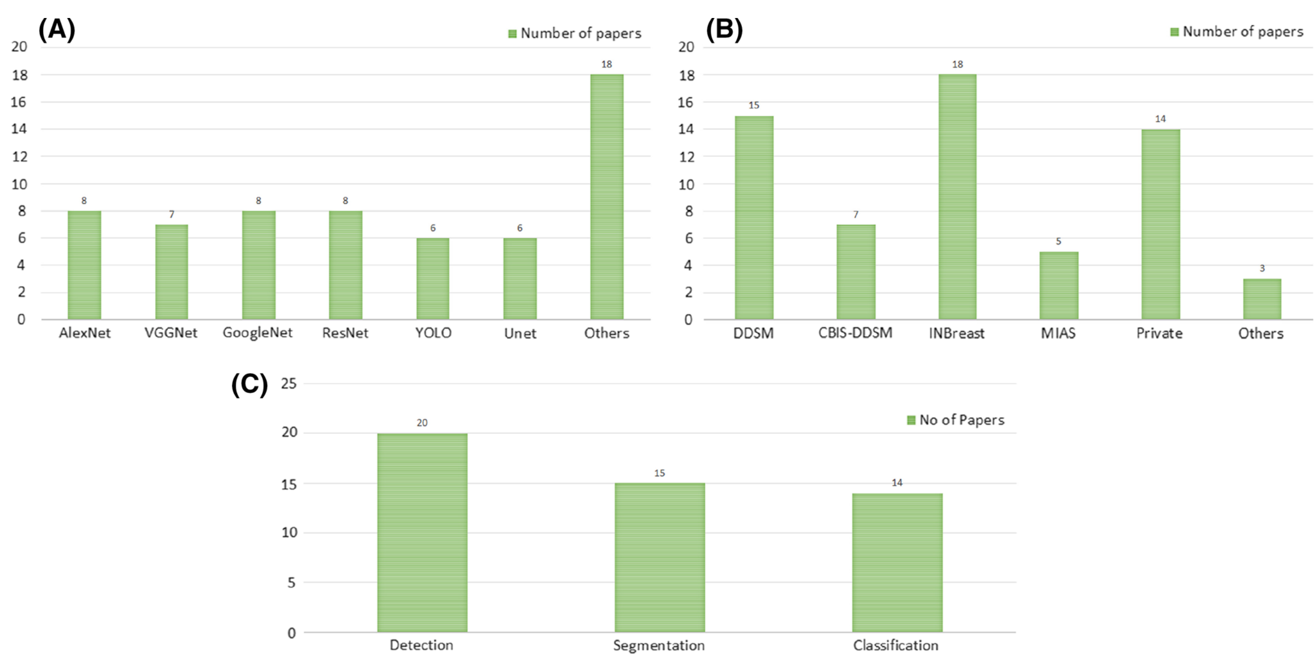

Fig. 11 Paper distribution: **a** number of papers that used particular pre-trained DCNN architecture **b** number of papers with particular datasets discussed in the paper **c** distribution of papers according to DCNN applications for breast cancer diagnosis

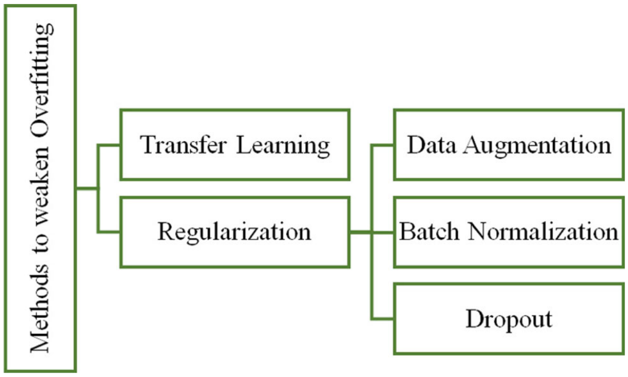

Fig. 12 Methods to weaken overfitting

datasets [30]. Figure 12 pictorially gives the summary of methods used to combat the effect of overfitting.

6.1.1 Transfer learning [TL]

Transfer learning is a technique to train a model for one application and to reuse the same model for another application. This domain adaptation is the situation where whatever is already been educated in one domain is to be used for the improvement or generalization in another domain. There are very few studies in the literature, which have built the DCNN from the scratch; most of the research work [58, 92–104] have fine-tuned the pre-trained networks or used them as only feature extractors. Training a DCNN or any other deep learning model from scratch requires lot of memory as well as computational resources. VGGNet [58, 66, 69, 74, 75, 102, 103], AlexNet [3, 66, 69, 71, 75, 93, 94, 96], ResNet [58, 62, 65, 66, 68, 69, 75, 99], and GoogleNet [58, 65–67, 75, 94, 98, 101] are most commonly used DCNN architectures for mammogram images as discussed in Sect. 4. These architectures are pre-trained using ImageNet weights.

6.1.2 Regularization

Regularization is another technique used primarily for deep neural networks, which allows generalization to unseen data even when training a model on a limited training set [104]. There are different forms of regularization like L1 regularization and L2 regularization (also called weight decay). Authors of [104] divided regularization techniques into a total of five categories; based on data, based on error function, based on network architecture, based on regularization term, and based on optimization techniques. One more survey of regularization techniques is presented in [105]. These techniques are classified into a total of 13 categories. The most commonly used regularization techniques used to build DCNN for breast cancer diagnosis are data augmentation, batch normalization, and dropout.

Overfitting occurs due to fewer training samples. A regularization technique called *Data Augmentation* is used to mitigate the effect of overfitting by modifying the training dataset. This technique applies arbitrary transformations on training data such as reflection, cropping, translation, and rotation [29]. Authors of [98, 105–107] have used the data augmentation techniques and have shown improved performance.

Batch Normalization [BN] is a technique that normalizes input data over a subset of training data. This subset of training data is called a mini-batch. First, batch normalization layers normalize the activations by deducting the mean of mini-batch and then dividing it by the standard deviation of mini-batch. BN helps speed the training process and also helps in reducing the sensitivity for the initialization of the network. BN also allows faster convergence and result in a high learning rate. For mammogram images, BN is used along with dropout [30]. BN grants a higher learning rate, and it does not care too much about initialization as stated in [108].

Dropout is another regularization technique proposed in [109]. This technique prevents the DCNN network from overfitting by randomly selecting neurons and discounting them during training time. These neurons are randomly dropped out. Meaning that their contribution to the activation of downstream neurons is removed during forwarding pass and will apply no weight update to these neurons during backward pass [29, 30]. Performance comparison of various regularization techniques is shown in [110] and the author has shown that dropout performs better than other techniques. In most studies related to mammogram images, 0.5 is the most common dropout value.

6.2 Class imbalance

There is one more barrier that makes training of image classifiers more challenging. The dataset may not have an equal number of examples for disease and non-disease. This skewed distribution of data mainly arises in many medical application, including breast imaging, where several positive samples are occurring with low frequency than that of negative samples [111]. Real-world data are typically not balanced well, which is the main reason for low generalization in deep learning and machine learning approaches. Existing architectures give the same attention to both classes, minority, as well as the majority [111, 112]. A classifier may miss-predict minority class in many imbalanced datasets, so a good solution is needed to overcome this problem. Many answers are proposed in the literature to solve the class imbalance problem, including weighted loss function, sampling methods, data augmentation, and generating synthetic images using GAN [112].

Methods to solve class imbalance problems are categorized into data-level methods and classifier/algorithm-level methods. Various sampling methods are used to reduce the class imbalance at the data level. The sampling-based process rebuilds a balanced dataset not by increasing the number of training samples but by the consideration of demonstrative proportions of examples in the class distribution [114]. Classifier/algorithm level methods are implemented with cost functions or weights. These methods modify the learner by adjusting the training process or inference algorithm used for that classifier [111–114]. A detailed description of methods for handling class imbalance problems for lesion detection is presented in [113]. A new hybrid system to solve the class imbalance problem by combining both data-level methods and algorithmic approaches is proposed by the authors of [115]. Figure 13 shows a summary of different ways to handle the class imbalance problem.

7 Toolkits and deep learning libraries

Building any deep learning network from scratch is an arduous task. It is always more convenient to make use of publicly available resources to build the network. The main reason behind the popularity of deep learning networks, specially DCNN, is the accessibility of various open-source libraries and toolkits. These toolkits provide very good GPU support [116]. In [117], authors have provided evaluation criteria for toolkit like execution and training speed, development environment, reporting, and quality of document and language interface. TensorFlow [118], and Keras [119] are the two most popular toolkits which are widely used for the training of DCNN on mammogram images. However, Caffe [120], PyTorch [121], Microsoft Cognitive Toolkit (CNTK) [122] and MXNet [123] are also used by many researchers. Table 7 depicts a summary of all these toolkits.

TensorFlow: TensorFlow is developed by the team of Google. This framework provides support for both mobile and desktop applications. Moreover, it supports interfaces like C++, R, and Python. This framework is widely adopted by researchers working in image recognition, forecasting, text classification, tagging, and natural language processing.

Keras: Keras supports both DCNN and recurrent neural networks. It is also capable of running on top of CNTK, TensorFlow, or Theano. There are many active developers and contributors across the world. There is also a good amount of documentation available for this framework.

Caffe: Caffe is supported by programming platforms like MATLAB, C, C++, Python, and command line. Caffe has a pre-trained Deepnet repository called "Caffe Model Zoo," which is available for use immediately. However, it has no support for fine granularity network layers like in TensorFlow and CNTK.

PyTorch: Pytorch also runs on python like other frameworks. PyTorch is popular amongst the deep learning community for automatic differentiation, dynamic computation graphs, and GPU support. This scientific computation package was developed by people of Facebook and used by corporations like Facebook, Google, and Twitter.

Microsoft CNTK: Microsoft CNTK is supported by interfaces like the command line, C++, and python. It provides very high scalability in terms of training DCNNs for images, speech, and textual data. When working with multiple machines, the scalability is much better as compared to TensorFlow. However, it has deficient capability on mobiles due to a lack of support of ARM architecture.

MXNet: This framework provides support for programming languages like C++, R, SCALA, Julia, Python, and many others. It allows user to code their deep learning model in any language. MXNet supports DCNN as well as recurrent neural networks. It also provides support for long- and short-term memory (LSTM) networks.

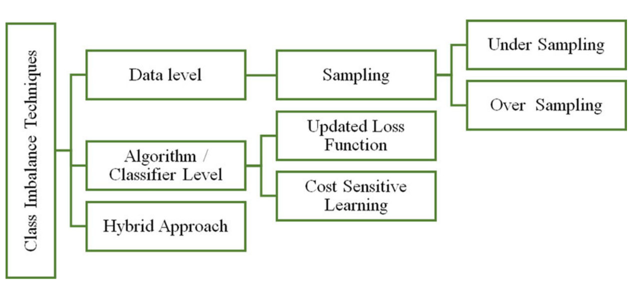

Fig. 13 Classification of methods to solve the class imbalance

8 Research directions and opportunities

Various DL and ML approaches [124, 128–130] have been used for more than a decade for breast cancer screening and diagnosis. From all of them, DCNNs have achieved outstanding accomplishments for breast cancer diagnosis and other medical diagnostic tasks. But, there are still some research challenges apart from class imbalance and small-size datasets that need further examination. This section presents such challenges.

Building Multi-Task Model: Most of the DCNNs are developed as a binary classifier where a model classifies a lesion or a mass as benign or malignant. There is a need to develop a model which can classify an input mammogram into seven categories of Breast Imaging-Reporting and Data System (BIRADS) [125]. BIRADS is a well-defined

Table 7 Comparison of deep learning toolkits for training mammogram images

	Open Source	Support for interface	Strength	Limitation
CNTK	Yes	Python, C++	Good scalability, Efficient resource usage	Significant learning curve, Low capability on mobiles
Caffe	Yes	Python, C++, C, MATLAB	Speed (Can process 60 million images with one GPU), Pre-trained deepnet "caffe model zoo"	No support for fine granularity networks
MXNet	Yes	Python, C++, R, SCALA, Julia	Support for multiple programming languages, Highly efficient, High productivity, support for LSTM	Small open source community
Keras	Yes	Python, R	High-level neural network APIs, Many active developers and contributors	Low-level backend errors
PyTorch	Yes	Python, C	Pycharm debugging tool, Support for many pre-trained models	No interface for visulaisation and monitoring
TensorFlow	Yes	Python, C++, R	Large community support, User-friendly, Tensorboard support for visulization and monitoring process, Support for many pre-trained models	Challenging debugging process

assessment scheme proposed initially by the American College of Radiology (ACR) and widely used by the radiological community for the diagnosis and reporting process. This categorization and final assessment decreased ambiguity in recommendations. DCNNs can also be developed to classify breast cancer as per TNM (Tumor, Node, Metastasis) staging [126]. TNM staging is one more classification system that is used to see the degree of spread of cancer.

Detecting Multiple Abnormalities: Most of the DCNN models built so far have been developed to locate, detect or segment only one abnormality from a mammogram, either mass or microcalcification. Therefore, there is a need for a classifier or model that can detect, segment, or classify multiple abnormalities within the same mammogram.

Need For Pre-processing Technique: Due to the low resolution and poor quality of mammogram images, there are chances that a model misclassifies the input image. Hence there is a need to preprocess those images before training. Preprocessing filters like median filter and histogram equalization are widely used to improve image quality [106]. Also, images of different mammogram datasets have different resolutions as various devices take them. Hence, preprocessing filters used for one dataset may not produce a good result on another dataset. The selection of a suitable preprocessing technique is still an open challenge.

Incomplete Annotation: Despite the fact that there are various ways for smoothing training on smaller datasets and weakening overfitting and class imbalance concerns (as stated in Sect. 6), there is still a need for adequately annotated large mammography repositories. But, building such a huge repository is a very costly process and takes a large amount of dedicated time. It may also hold privacy and legal issues. Active learning and self-paced paradigms have come up with an idea to reduce annotation efforts by the expert clinicians and get better results with very few annotated examples [59]. To defeat the problem of incomplete annotation, authors have used unsupervised pre-training and supervised fine-tuning approaches in [127]. Though the dataset may have many images, incomplete annotations do not allow us to use them all. Hence, this problem needs further investigation to come up with a good solution.

Need for more Data Augmentation Techniques: Training a DCNN model with a small size dataset is one of the most challenging tasks. As discussed in the Sect. 6, although varieties of methods such as transfer learning, data augmentation, and regularization have been widely used to deal with the problem, the problem has still remained challenging. There is still a need to develop more accurate and proper data augmentation techniques to increase the training dataset size.

9 Conclusion

Deep convolutional neural networks (DCNNs) have achieved notable advances in various domains, including medical diagnostics, and growing interest has emerged in breast cancer diagnosis. In this paper, the most common applications of DCNNs for breast cancer diagnosis are presented. We also present in this study, most popular pre-trained DCNN architectures to address these applications.

Comparative analysis of this model based on its configuration and key features are shown in the paper. The strength and limitations of these pre-trained architectures guide the researchers of this domain in selecting the appropriate model for their application. This study further shows the key challenges of DCNNs for training mammogram images. Techniques to improve the generalization of the model are also discussed. Toolkits and libraries to build a model are also presented with their strength and limitations. Although DCNNs thrived in a state of the art achievement for breast cancer diagnosis and other imaging tasks, some research challenges still need further examination.

Funding There is no funding involved in writing this article.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Bray F (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68(6):394–424
2. Cancer Facts & Figures (2020), [Online]. Available: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/ cancer-facts-figures-2020.html. Accessed on: (March 28, 2021)
3. Alessandro B et al (2020) A novel solution based on scale invariant feature transform descriptors and deep learning for the detection of suspicious regions in mammogram images. J Med Sig Sens 10(3):158
4. Debelee TG et al (2020) Survey of deep learning in breast cancer image analysis. Evolv Syst 11(1):143–163
5. Al-antari Mugahed A et al (2018) An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network. J Med Biol Eng 38(3):443–456
6. Al-Masni MA et al (2018) Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Comput Meth Prog Biomed 157:85–94
7. Al-Antari Mugahed A et al (2018) A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. Int J Med Inform 117:44–54
8. Zou L et al (2019) A technical review of convolutional neural network-based mammographic breast cancer diagnosis. Comput Math Meth Med
9. Burt Jeremy R et al (2018) Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks. British J Radiol 91(1089):20170545
10. Anna C-G et al (2016) Positive psychological functioning in breast cancer: an integrative review. Breast 27:136–168
11. Muramatsu C et al (2016) Breast mass classification on mammograms using radial local ternary patterns. Comp Biol Med 72:43–53
12. Jitendra V et al (2016) PCA-PNN and PCA-SVM-based CAD systems for breast density classification. Applications of intelligent optimization in biology and medicine. Springer, Cham, pp 159–180
13. Yassin NIR et al (2018) Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. Comp Meth Prog Biomed 156:25–45
14. Haixia L et al (2017) Breast masses in mammography classification with local contour features. Biomed Eng online 16(1):44–55
15. Neeraj D et al (2017) A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med Image Anal 37:114–128
16. Jamal GS, Safdar, et al (2019) Breast cancer detection and diagnosis using mammographic data: systematic review. J Med Int Res 21(7):e14464
17. Gustavo C et al (2017) Automated analysis of unregistered multi-view mammograms with deep learning. IEEE Trans Med Imag 36(11):2355–2365
18. Peyman R et al (2012) Mammography segmentation with maximum likelihood active contours. Med Image Anal 16(6):1167–1186
19. Ying W et al (2011) Mammographic mass segmentation: embedding multiple features in vector-valued level set in ambiguous regions. Patt Recog 44(9):1903–1915
20. Rojas DA, Nandi Asoke K (2009) Toward breast cancer diagnosis based on automated segmentation of masses in mammograms. Patt Recog 42(6):1138–1148
21. Thijs K et al (2017) Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 35:303–312
22. Dhungel N et al (2015) Deep structured learning for mass segmentation from mammograms, IEEE International Conference on Image Processing (ICIP). QC, Canada, Quebec City, pp 2950–2954
23. Zhicheng J et al (2016) A deep feature-based framework for breast masses classification. Neurocomputing 197:221–231
24. (2001) The digital database for screening mammography [Online]. Available:http://www.eng.usf.edu/cvprg/Mammography/Database.html
25. Suckling J et al (1994) The mammographic image analysis society digital mammogram database exerpta medica. Int Congr Ser 1069:375–378
26. Moreira IC et al (2012) INbreast: toward a full-field digital mammographic database. Acad Radiol 19(2):236–248
27. Lopez MG, et al (2012) BCDR: a breast cancer digital repository 15th International Conference on Experimental Mechanics
28. Oliveira Júlia EE et al (2008) Toward a standard reference database for computer-aided mammography. Med Imag 2008: Comput Aid Diag 6915
29. Rikiya Y et al (2018) Convolutional neural networks: an overview and application in radiology. Insights Imag 9(4):611–629
30. Dina A et al (2019) Deep convolutional neural networks for mammography: advances, challenges and applications. BMC Bioinform 20(11):75–94
31. Pedamonti D (2018) Comparison of non-linear activation functions for deep neural networks on MNIST classification task. arXiv preprint arXiv:1804.02763
32. Glorot X et al (2011) Deep sparse rectifier neural networks. Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics : 315-323
33. Maas Andrew L et al (2013) Rectifier nonlinearities improve neural network acoustic models. Int Conf Mach Learn Atlanta Georg USA 30:1
34. Adnan Q et al (2017) Medical image retrieval using deep convolutional neural network. Neurocomputing 266:8–20

35. Mohamed Aly A et al (2018) Understanding clinical mammographic breast density assessment: a deep learning perspective. J Dig Imag 31(4):387–392
36. Wichakam I, Vateekul P (2016) Combining deep convolutional networks and SVMs for mass detection on digital mammograms. presented in 8th IEEE International Conference on Knowledge and Smart Technology 239-44
37. Vasilev I et al (2019) Neural networks, in python deep learning: exploring deep learning techniques and neural network architectures with Pytorch, Keras. and TensorFlow, 2nd edn. Packt Publishing Ltd, BIRMINGHAM - MUMBAI, India, pp 43–44
38. Alex K, Ilya S, Hinton Geoffrey E (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inform Proc Syst 25:1097–1105
39. Yann LC et al (1998) Gradient-based learning applied to document recognition. Proceed IEEE 86(11):2278–2324
40. Asifullah K et al (2020) A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 53(8):5455–5516
41. Simonyan K, Andrew Z (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556v6 [cs.CV]. 1-14
42. Szegedy Christian et al (2015) Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition: 1-12
43. He K, Sun J (2015) Convolutional neural networks at constrained time cost. Proceedings of the IEEE conference on computer vision and pattern recognition. 5353–5360
44. Srivastava RK et al (2015) Highway networks. arXiv preprint<error l="295" c="End of data reached while scanning argument" /><hyperimage arxiv="http://arxiv.org/abs/arXiv:1505.00387v2" />
45. He K et al (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition: 770-778
46. Szegedy, Christian, et al (2016) Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition : 2818-2826
47. Szegedy C et al (2016) Inception-v4, inception-resnet and the impact of residual connections on learning. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17) pp 4278–4284
48. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition pp 1800–1807
49. Redmon, Joseph, et al (2016). You only look once: unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition: 779-788
50. Long, Jonathan, et al(2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition : 3431-3440
51. Ronneberger, Olaf, et al (2015). U-net: convolutional networks for biomedical image segmentation. Proceeding of International Conference on Medical image computing and computer-assisted intervention : 234-241
52. He, Kaiming, et al (2017). Mask r-cnn. Proceedings of the IEEE international conference on computer vision : 2980-2988
53. Ren S et al (2017) Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans Patt Anal Mach Intell 39(6):1137–1149
54. Wenqing S et al (2017) Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. Comput Med Imag Graph 57:4–9
55. Dezsø R et al (2018) Detecting and classifying lesions in mammograms with deep learning. Scient Rep 8(1):1–7
56. Diniz JO, Bandeira, et al (2016) Detection of mass regions in mammograms by bilateral analysis adapted to breast density using similarity indexes and convolutional neural networks. Comput Meth Prog Biomed 156:191–207
57. Yousefi M et al (2018) Mass detection in digital breast tomosynthesis data using convolutional neural networks and multiple instance learning. Comput Biol Med 96:283–293
58. SanaUllah K et al (2019) A novel deep learning-based framework for the detection and classification of breast cancer using transfer learning. Patt Recog Lett 125:1–6
59. Rongbo S et al (2019) Breast mass detection from the digitized X-ray mammograms based on the combination of deep active learning and self-paced learning. Fut Gen Comput Syst 101:668–679
60. Benedetta S et al (2020) A multi-context CNN ensemble for small lesion detection. Artif Intell Med 103:101749
61. Agarwal R et al (2020) Deep learning for mass detection in full field digital mammograms. Comput Biol Med
62. Al-Antari et al (2020) Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms. Comput Meth Prog Biomed. 196:105584
63. Yanfeng L et al (2020) Mass detection in mammograms by bilateral analysis using convolution neural network. Comput Meth Prog Biomed. 195:105518
64. Rongbo S et al (2020) Unsupervised domain adaptation with adversarial learning for mass detection in mammogram. Neurocomputing 393:27–37
65. Aly GH et al (2020) YOLO-based breast masses detection and classification in full-field digital mammograms. Comput Meth Prog Biomed. 200:105823
66. Xi P, Shu C, Goubran R (2018) Abnormality detection in mammography using deep convolutional neural networks, IEEE International Symposium on Medical Measurements and Applications (MeMeA). Italy, Rome, pp 1–6
67. Deep DS et al (2020) Breast cancer detection and classification using global pooling, 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, pp 1–5
68. Abdel Rahman AS et al (2020) Breast mass tumor classification using deep learning, IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, : 271-276
69. Xiaofei Z et al (2018) Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks. IEEE Trans Nanobiosci 17(3):237–242
70. Djebbar K et al (2019) Deep convolutional neural networks for detection and classification of tumors in mammograms, 6th International Conference on Image and Signal Processing and their Applications (ISPA). Algeria, Mostaganem, pp 1–7
71. Gnanasekaran VS et al (2020) Deep learning algorithm for breast masses classification in mammograms. IET Image Proc. 14(12):2860–2868
72. Xin S et al (2020) Deep neural networks with region-based pooling structures for mammographic image classification. IEEE Trans Med Imag 39(6):2246–2255
73. Al-Masni et al. (2017). Detection and classification of the breast abnormalities in digital mammograms via regional Convolutional Neural Network 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) :1230-1233
74. Gardezi SJ, Safdar, et al (2017) Mammogram classification using deep learning features IEEE International Conference on Signal and Image Processing Applications (ICSIPA) pp 485–488
75. Lazaros T et al (2019) Deep learning for breast cancer diagnosis from mammograms-a comparative study. J Imag 5(3):37

76. Carneiro G et al (2017) Deep learning models for classifying mammogram exams containing unregistered multi-view images and segmentation maps of lesions. in Deep learning for medical image analysis Academic Press. pp 321–339
77. Zhu W et al (2018) Adversarial deep structured nets for mass segmentation from mammograms. IEEE 15th International Symposium on Biomedical Imaging (ISBI) : 847–850
78. Andrik R et al (2019) Breast pectoral muscle segmentation in mammograms using a modified holistically nested edge detection network. Med Image Anal 57:1–17
79. Shen T et al (2019) Learning from adversarial medical images for X-ray breast mass segmentation. Comput Meth Prog Biomed. 180:105012
80. Hossain MDS (2019) Micro calcification segmentation using modified U-net segmentation network from mammogram images. J King Saud Univ Comput Inform Sci
81. Runze W et al (2019) Multi-level nested pyramid network for mass segmentation in mammograms. Neurocomputing 363:313–320
82. Shuyi L et al (2019) Attention dense-u-net for automatic breast mass segmentation in digital mammogram. IEEE Access 7:59037–59047
83. Pérez-Benito FJ et al (2020) A deep learning system to obtain the optimal parameters for a threshold-based breast and dense tissue segmentation. Comput Meth Prog Biomed. 195:105668
84. Ghosh Swarup KR et al (2021) A novel intuitionistic fuzzy soft set entrenched mammogram segmentation under Multigranulation approximation for breast cancer detection in early stages. Exp Syst Appl. 169:114329
85. Singh VK et al (2020) Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. Exp Syst Appl. 139:112855
86. Qian Yu et al (2021) Crossover-Net: leveraging vertical-horizontal crossover relation for robust medical image segmentation. Pattern Recog. 113:107756
87. Juan C et al (2020) A novel multi-scale adversarial networks for precise segmentation of X-ray breast mass. IEEE Access 8:103772–103781
88. Dina A et al (2020) Convolutional neural network for automated mass segmentation in mammography. BMC Bioinf 21(1):1–19
89. Yu Hui et al (2020). Deep learning-based fully automated detection and segmentation of breast mass. 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). Chengdu, China, pp 293–298
90. Soleimani H, Michailovich OV (2020) On segmentation of pectoral muscle in digital mammograms by means of deep learning. IEEE Access 8:204173–204182
91. Tianyu S et al (2019) Simultaneous segmentation and classification of mass region from mammograms using a mixed-supervision guided deep model. IEEE Sig Proc Lett 27:196–200
92. Mehreen T et al (2020) Medical image-based breast cancer diagnosis: state of the art and future directions. Exp Syst Appl. 167:114095
93. Huynh Benjamin Q et al (2016) Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. J Med Imag 3(3):034501
94. Jiang F et al (2017) Breast mass lesion classification in mammograms by transfer learning. Proceedings of the 5th International Conference on Bioinformatics and Computational Biology. ACM. pp 59–62
95. Chougrad, Hiba, et al (2017). Convolutional neural networks for breast cancer screening: transfer learning with exponential decay. arXiv preprint arXiv:1711.10752v1 [cs.CV]
96. Samala Ravi K et al (2017) Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. Phys Med Biol 62(23):8894
97. Hiba C et al (2020) Multi-label transfer learning for the early diagnosis of breast cancer. Neurocomputing 392:168–180
98. Yemini M et al (2018) Detecting masses in mammograms using convolutional neural networks and transfer learning, IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE). Israel, Eilat, pp 1–4
99. Falconí LG et al (2019). Transfer learning in breast mammogram abnormalities classification With mobilenet and nasnet, International Conference on Systems, Signals and Image Processing (IWSSIP), Osijek, Croatia, pp 109–114
100. Kamrul H et al (2020) Automatic mass classification in breast using transfer learning of deep convolutional neural network and support vector machine. IEEE Region 10 Symposium (TENSYMP) p 110:113
101. Mednikov Y et al (2018) Transfer representation learning using inception-V3 for the detection of masses in Mammography, 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). HI, USA, Honolulu, pp 2587–2590
102. Guan S, Loew M (2017) Breast cancer detection using transfer learning in convolutional neural networks, IEEE Applied Imagery Pattern Recognition Workshop (AIPR). USA, Washington, DC, pp 1–8
103. Falconí L et al (2020) Transfer learning and fine tuning in Mammogram BI-RADS classification. IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS) p 475:480
104. Kukačka J et al. (2017) Regularization for deep learning: A taxonomy. arXiv preprintarXiv:1710.10686v1 [cs.LG]
105. Reza M et al (2020) A survey of regularization strategies for deep models. Artif Intell Rev 53(6):3947–3986
106. Lu H et al (2019) The classification of mammogram using convolutional neural network with specific image preprocessing for breast cancer detection, 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD). China, Chengdu, pp 9–12
107. Zhao X et al (2018) Classification of benign and malignant breast mass in digital mammograms with convolutional neural networks, Proceedings of the 2Nd International Symposium on Image Computing and Digital Medicine . pp 47–50
108. Ioffe S, Christian S (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167v3 [cs.LG]
109. Srivastava N et al (2014) Dropout: a simple way to prevent neural networks from overfitting. The J Mach Learn Res 15(1):1929–1958
110. Smirnov EA et al. (2014) Comparison of regularization methods for imagenet classification with deep convolutional neural networks. Aasri Procedia 6:89–94
111. Johnson Justin M, Khoshgoftaar Taghi M (2019) Survey on deep learning with class imbalance. J Big Data 6(1):27
112. Mostafizur Rahman M, Davis DN (2013) Addressing the class imbalance problem in medical datasets. Int J Mach Learn Comp 3(2):224–228
113. Alessandro B et al (2020) Addressing class imbalance in deep learning for small lesion detection on medical images. Comput Biol Med. 120:103735
114. Zhang C (2019) Medical image classification under class imbalance. Graduate Theses and Dissertations. 17130. https://lib.dr.iastate.edu/etd/17130
115. Krawczyk B (2016) Learning from imbalanced data: open challenges and future directions Prog Artif Intell. 5(4):221–232
116. Geert L et al (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88

117. Erickson Bradley J et al (2017) Toolkits and libraries for deep learning. J Digit Imag 30(4):400–405
118. Abadi M et al (2016)TensorFlow: a system for large-scale machine learning. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI) pp 265–283
119. Nikhil K (2017) Introduction to keras. Apress, Berkeley, CA, Deep learning with Python, pp 97–111
120. Jia Y et al (2014) Caffe: convolutional architecture for fast feature embedding. Proceedings of the 22nd ACM international conference on Multimedia. pp 675–678
121. Paszke A et al (2017) Automatic differentiation in pytorch. 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA
122. Yu D et al (2014) An introduction to computational networks and the computational network toolkit. Microsoft Technical Report MSR-TR-2014-112
123. Chen TM et al (2015) A flexible and efficient machine learning library for heterogeneous distributed systems arXiv preprint arXiv:1512.01274
124. Pillai R et al. (2019). Review of machine learning techniques in health care In: Singh P, Kar A, Singh Y, Kolekar M, Tanwar S. (eds) Proceedings of ICRIC 2019. Lecture Notes in Electrical Engineering,Springer, Cham. 597 : 103-111
125. Liberman L, Menell JH (2002) Breast imaging reporting and data system (bi-rads). Radiol Clin 40(3):409–430
126. TNM and Staging of Breast Cancer Simplified, [Online]. Available: https://epomedicine.com/medical-students/tnm-staging-breast-cancer-simplified
127. Suk HI, Shen D (2013). Deep learning-based feature representation for AD/MCI classification In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N. (eds) Medical Image Computing and Computer-Assisted Intervention MICCAI 2013. MICCAI 2013. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 8150 : 583-590
128. Oza P, Sharma P, Patel S (2021) Machine learning applications for computer-aided medical diagnostics. In: Singh PK, Wierzchoń ST, Tanwar S, Ganzha M, Rodrigues JJPC (eds) Proceedings of Second International Conference on Computing, Communications, and Cyber-Security. Lecture Notes in Networks and Systems, Springer, Singapore. vol 203. : 377-392
129. Parita O et al (2021) A bottom-up review of image analysis methods for suspicious region detection in mammograms. J Imag 7(9):190
130. Oza P, Shah Y, Vegda M (2022) A comprehensive study of mammogram classification techniques. Cham, tracking and preventing diseases with artificial intelligence. Springer, pp 217–238