Mammogram breast cancer CAD systems for mass detection and classification: a review

Nada M. Hassan[1] · Safwat Hamad[2] · Khaled Mahar[3]

Abstract
Although there is an improvement in breast cancer detection and classification (CAD) tools, there are still some challenges and limitations that need more investigation. The significant development in machine learning and image processing techniques in the last ten years affected hugely the development of breast cancer CAD systems especially with the existence of deep learning models. This survey presents in a structured way, the current deep learning-based CAD system to detect and classify masses in mammography, in addition to the conventional machine learning-based techniques. The survey presents the current publicly mammographic datasets, also provides a dataset-based quantitative comparison of the most recent techniques and the most used evaluation metrics for the breast cancer CAD systems. The survey provides a discussion of the current literature and emphasizes its pros and limitations. Furthermore, the survey highlights the challenges and limitations in the current breast cancer detection and classification techniques.

Keywords CAD system · Breast cancer · Mammogram · Mass · Detection · Classification

Abbreviations

Acc	Accuracy
ACR BI-RADS	American College of Radiology Breast Imaging Reporting and Data System
AP	Average Precision
AUC	Area Under Curve
BCDR	Breast Cancer Digital Repository
CAD	Computer Aided (Detection/Diagnosis)

✉ Nada M. Hassan
   nadamahmoud@aast.edu

1   College of Computing and Information Technology, Arab Academy for Science and Technology, Cairo, Egypt

2   Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

3   College of Computing & Information Technology, Arab Academy for Science and Technology, Alexandria, Egypt

CBIS-DDSM	Curated Breast Imaging Subset of DDSM
CC	Craniocaudal
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DBT	Digital Breast Tomosynthesis
DDSM	Digital Database for Screening Mammography
DICOM	Digital Imaging and Communication in Medicine
DL	Deep Learning
DT	Decision Tree
DWT	Discrete Wavelet Transform
E2E	End to End learning
ELM	Extreme Learning Machine
FCN	Fully Convolutional Network
FDA	Food and Drug Administration
FFDM	Full Field Digital Mammography
FM	Film Mammography
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FRCNN	Full Resolution Convolutional Neural Network
GA	Genetic Algorithm
GAN	Generative Adversarial Network
GCO	Global Cancer Observatory
GLRLM	Gray Level Run Length Matrix
HSV	Hue Saturation Value
IDC	Invasive Ductal Carcinoma
IEEE	Institute of Electrical and Electronics Engineers
ILC	Invasive Lobular Carcinoma
IoU	Intersection over Union
JPEG	Joint Photographic Experts Group
KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
mAP	mean Average Precision
MFO	Moth Flame Optimization
MG	Mammogram
MIAS	Mammographic Image Analysis Society
MLO	Medio Lateral Oblique
MLP	Multi-Layer Perceptron
MRI	Magnetic Resonance Image
NB	Naïve Bayes
NN	Neural Network
PCA	Principal Component Analysis
PEIPA	Pilot European Image Processing Archive
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis
PSO	Particle Swarm Optimization
QDA	Quadratic Discriminant Analysis

R-CNN	Region based Convolutional Neural Network
RA	Random Forest
RIFD-CNN	Rotation Invariant Fisher Discriminative Convolutional Neural Network
ROC	Receiver Operating Characteristics
ROI	Region of Interest
RPN	Regional Proposal Network
RS	Rough Set
SFM	Screen Film Mammography
SVM	Support Vector Machine
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
VFCS	Vector Field Convolution Snake
WHO	World Health Organization
YOLO	You Look Only Once

1 Introduction

Breast cancer is classified as one of the most widespread cancer among women worldwide. According to the statistics [80] that was published in 2020 by the Global Cancer Observatory (GCO) that is affiliated with the World Health Organization (WHO); for every 100,000 persons in the world 47.8% were diagnosed with breast cancer which has the highest incidence rate for the top 10 cancer types in the world within females, and comes the second in the mortality rate after the lung cancer with 13.6% for every 100,000 persons; this means almost 29.1% died out of the 47.8% who were diagnosed with breast cancer. Figure 1 shows the distribution of new cases and deaths for the top 10 cancers among females in 2020 all over the world [80]. Breast cancer may begin in the milk ducts and this type is called Invasive Ductal Carcinoma (IDC) or starts in the milk-producing glands and this type is called Invasive Lobular Carcinoma (ILC) [18] . Many factors are considered as risk factors for breast cancer such as family history, ageing, gene changes, race, exposure to chest radiation, and obesity [13] .

The early detection of breast cancer helps in increasing the survival rate of this disease. And so, regular screening is considered one of the most important tools that can help in the early detection of this type of cancer. A mammogram is considered as one of the effective screening modalities in detecting breast cancer at early stages [54, 83], it can reveal different abnormalities in the breast even before any symptoms appear. With the significant development in machine learning and image processing techniques, several studies were proposed for breast cancer detection and classification in an attempt to create more effective Computer-Aided (Detection / Diagnosis) systems for breast cancer.

CAD systems can be categorized into two types; Computer-Aided Detection (CADe) systems and Computer-Aided Diagnosis (CADx) systems. *CADe* mainly provides localization and detection for the masses or abnormalities that appear in the medical images, and let the interpretation of these abnormalities to the radiologist. On the other hand, *CADx* provides a classification for the masses and help in the decision making of the radiologist about the identified abnormalities [14] .
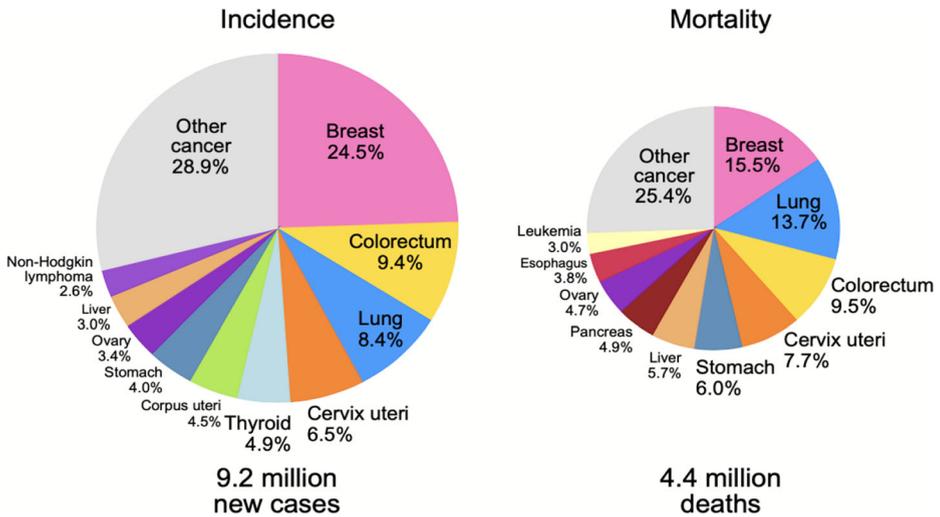
Fig. 1 Distribution of cases and deaths for the top 10 most common cancers in 2020 among females for Incidence and Mortality [80]

This review aims to cover the significant and well-known approaches which are introduced in the field of breast cancer detection and classification for masses using conventional machine learning and deep learning. Furthermore, the paper is demonstrating the evolution of the models that were introduced over the past ten years. The paper presents the current challenges and provides a discussion of the proposed models in the literature and their limitations.

This survey highlights the current screening modalities, mammogram projections, and different public mammography datasets. Also, the paper focuses on presenting a quantitative dataset-based comparison between the deep learning-based models for the most well-known and used public datasets, furthermore, the paper highlights the limitations and pros of the conventional machine learning-based and deep learning-based CAD systems.

The paper aims to answer the following questions:

- RQ1: What is the pipeline for the breast cancer CAD system, and what are the phases of developing such a system?
- RQ2: What are the breast screening modalities, and the public mammographic datasets?
- RQ3: What are the recent techniques that are used in developing CAD systems?
- RQ4: What are the evaluation measures that are currently used for breast cancer CAD systems assessment?
- RQ5: What are the limitations, challenges and future work in breast cancer detection and classification?

The paper is organized as the following; Section 2 provides the survey methodology, then section 3 gives an overview for the screening modalities and the publicly available mammography datasets, then section 4 presents the breast cancer CAD systems (conventional based and deep learning-based), followed by section 5 which demonstrates a dataset-based quantitative comparison between the deep learning-based CAD systems. Then, section 6 presents the evaluation metrics for detection and classification tasks in CAD systems, and finally, section 7 provides a discussion and conclusion. The organization of the survey is shown in Fig. 2.
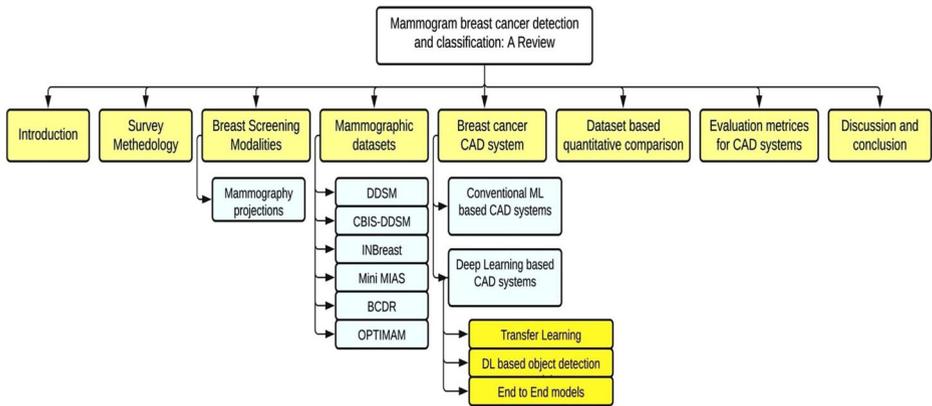
Fig. 2 Organization of the survey

2 Survey methodology

In this survey, the authors searched for articles through PubMed, Springer, Science Direct, Google Scholar, and Institute of Electrical and Electronics Engineers (IEEE). The articles that are included in this survey were published in English. The survey includes most of the published articles for mammography mass detection and classification from 2009 to earlier in 2021. Also, some articles are referenced for background context. Figure 3 shows the flow of information for the identification of the studies via databases, screening and the included studies in the literature according to Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA). In this survey, there are some exclusion criteria that were applied as the following:

C1: The papers that aren't in the form of a full research paper.
C2: Papers that aren't written in English.
C3: papers that aren't delivering results.
C4: Papers that aren't using mammographic datasets.
C5: Papers that aren't including mass detection, mass segmentation or mass classification.

3 Breast screening modalities

Different modalities are used for screening breasts such as mammogram (MG), Magnetic Resonance Imaging (MRI), Digital Breast Tomosynthesis (DBT), and ultrasound [43].

A mammogram (MG) is a non-invasive screening technique as it's an X-ray for imaging the breast tissue. It can reveal the masses and calcifications. Moreover, it is considered the most effective and sensitive screening modality as it can help in reducing the mortality rate through early detection of breast cancer even before any symptom appears.

Magnetic Resonance Imaging (MRI) is mainly depending on using strong magnets and radio waves to produce detailed pictures of inside the breast. This modality is considered to be helpful in the case of women at high risk for breast cancer.

Ultrasound uses sound waves to generate images of the internal structure of the breast. It is used for women who are at high risk for breast cancer and can't make MRI or women who are pregnant and shouldn't be exposed to the x-ray that is used
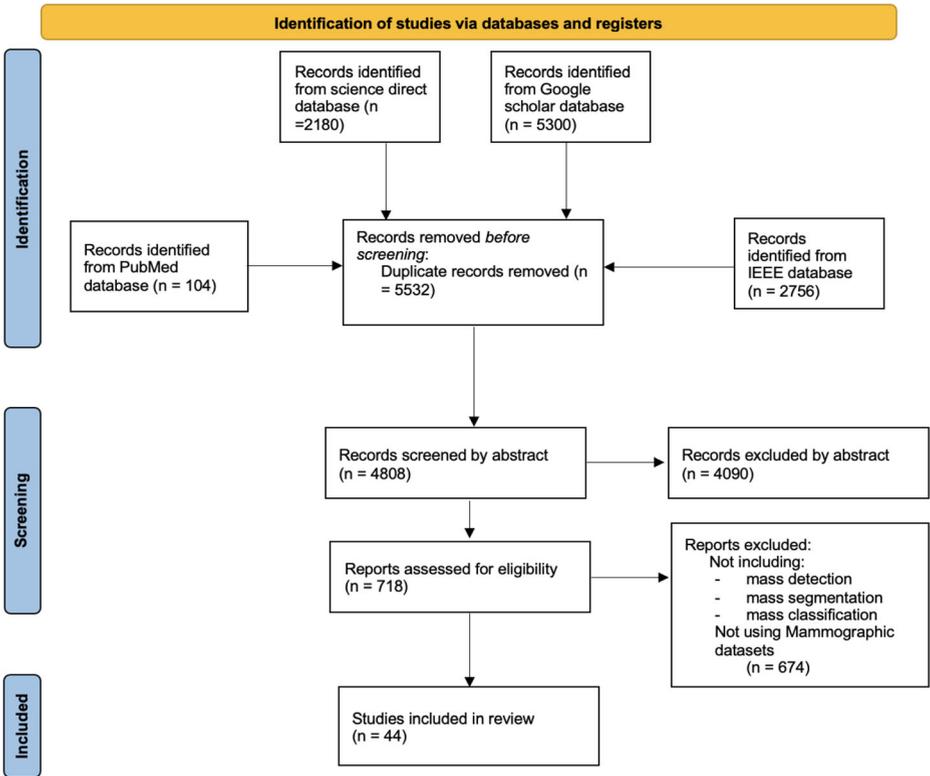
Fig. 3 Flow diagram according to preferred reporting items for systematic review and meta-analysis (PRISMA)

in MG. Also, ultrasound is very common to be used to screen women who have dense breast tissue.

Digital Breast Tomosynthesis (DBT) is a more recent technology that Food and Drug Administration (FDA) approved in 2011. DBT generates a more advanced form of mammogram that is generated through a low dose of x-ray. It is considered as 3D mammographic images that can reveal masses and calcifications in more detailed form, which can be very effective for radiologists especially with diagnosing dense breasts [9].

3.1 Mammography projections

A mammogram is considered the most effective and sensitive screening modality; MRI and ultrasound are used as a supplement for the mammogram especially with the cases that have high dense breast tissues, however, this doesn't mean they can replace the mammography [26].

There are multiple views for mammograms that are used to provide more information before detection/diagnosis. The main two views of mammograms are carnio-caudal (CC) and mediolateral oblique (MLO) as shown in Fig. 4. A CC view mammogram is taken horizontally from an upper projection at C-arm angle 0°; the breast is compressed between two paddles to reveal the glandular tissue, and the surrounding fatty tissue, also the right position of a CC view shows the outermost edge of the chest muscle. MLO view mammography is captured at a C-arm angle of 45° from the side; the breast is diagonally compressed between the paddles and
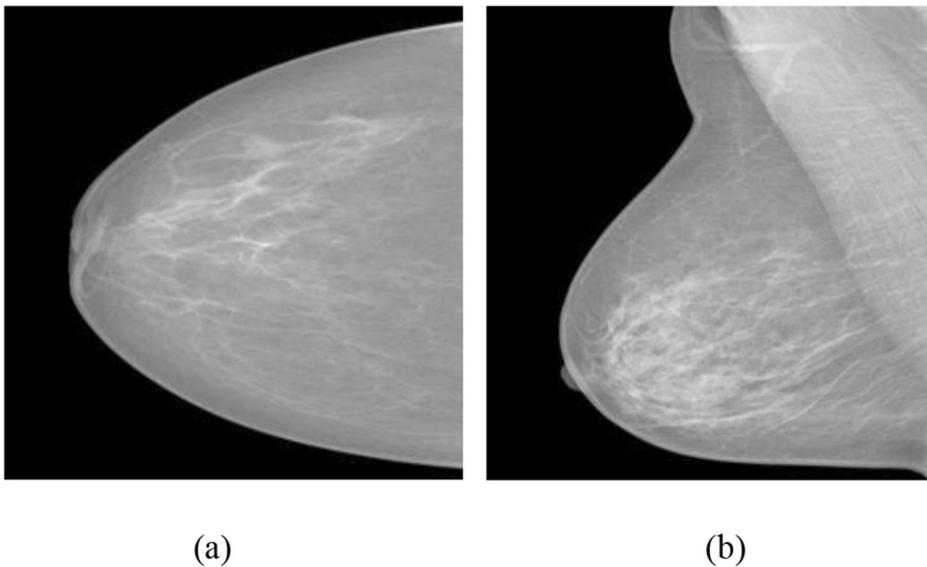
(a)   (b)

Fig. 4 (a) Carnio-Caudal (CC) projection, and (b) Medio-Lateral Oblique (MLO) projection from INBreast Dataset

accordingly this allows imaging a larger part of the breast tissue compared to other views. In addition to that, the MLO projection allows the pectoral muscles to appear in the mammographic image [11, 75].

The main two abnormalities that can be revealed in the mammograms are breast masses and calcifications. ***Breast masses*** may be cancerous or non-cancerous; the cancerous tumours appear in the mammograms with irregular edges and spikes extending from the mass. On the other hand, the non-cancerous masses often appear with round or oval shapes and well-defined edges [15].

Breast calcifications can be categorized into macrocalcifications and microcalcifications [59]. Macrocalcifications appear as large white dots on the mammogram and spread randomly over the breast, and are considered as non-cancerous cells. The microcalcifications seem as small calcium spots that look like white specks in the mammogram and they often appear in clusters. Microcalcification usually is considered as a primary indication for early breast cancer or a sign of existing precancerous cells. Figure 5 provides an illustration for the mass and calcification of an image from the INBreast dataset.

4 Mammographic datasets

Different datasets are publicly available, these datasets differ in size, resolution, image format, the type of the images (Full-Field Digital Mammography (FFDM), Film Mammography (FM), or Screen-Film Mammography (SFM)), and abnormalities' types that are included in each dataset. Table 1 compares among the publicly available datasets such as the digital database for screening mammography (DDSM), INBreast,
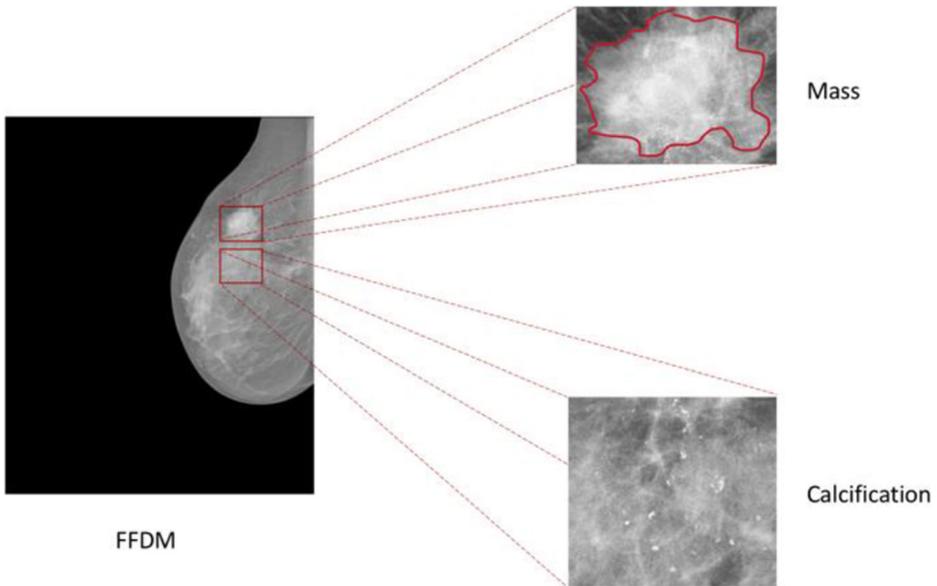
Fig. 5 Illustration for mass and calcification from INBreast Dataset [25]

Mini-MIAS, curated breast imaging subset of DDSM (CBIS-DDSM), BCDR, and OPTIMAM.

4.1 The digital database for screening mammography (DDSM)

DDSM is composed of 2620 scanned film mammography, the studies were divided into 43 volumes. There are four breast mammographic images for each case, as each breast side was captured from two projections which are Mediolateral Oblique (MLO) and Cranio-Caudal (CC) views. Also, the dataset includes the ground truth and kinds of the suspected regions at pixel-level annotations. Each case has a file that contains the date of the study, the age of the patient, the score of breast density according to the American College of Radiology Breast Imaging Reporting and Data System (ACR BI-RADS), and the size and the scanning resolution for each image. The images are in Joint Photographic Experts Group (JPEG) format with different sizes and different resolutions [39].

4.2 Curated breast imaging subset of DDSM (CBIS-DDSM)

This dataset is an enhanced version of the DDSM, it includes decompressed images, with updated mass segmentation and bounding boxes for the region of interest (ROI). The data is selected and curated by trained mammographers; the images are in Digital Imaging and Communication in Medicine (DICOM) format. The size of the dataset is 163.6GB with 6775 studies, the dataset contains 10,239 images, that consists of mammogram images with their corresponding mask images. There are CSV files attached with the dataset that provided the pathological information for the patients. The dataset has four CSV files: mass training set, mass-testing set, calcification training set and calcification testing set. The mass training set has

Table 1 Breast cancer mammographic datasets (DM: Digital Mammogram, FFDM: Full Field Digitized Mammogram, SFM: Screen Film Mammogram)

Dataset	Type	# of images	view	Format	Classes	Resolution (bit/ pixel)	Availability
DDSM [39]	DM	10,480	MLO / CC	LJPEG	Normal, benign and malignant	8–16	Publicly available
CBIS-DDSM [45]	DM	10,239	MLO / CC	DICOM	Normal, benign and malignant	10	Publicly available
INBreast [56]	FFDM	410	MLO / CC	DICOM	Normal, benign and malignant	14	Available upon request
Mini-MIAS [77]	SFM	322	MLO	PGM	Normal, benign and malignant	8	Publicly available
BCDR [50]	FM/FFDM	7315 (3703 FM - 3612 DM)	MLO / CC	TIFF	Normal, benign and malignant	8–14	Publicly available
OPTIMAM [38]	NA	>1 M	MLO / CC	DICOM	Normal, benign, malignant and interval cancer	12–16	Available upon administrator approval for access.

images for 1318 tumours, while the mass testing set has images for 378 tumours. The calcification training set has images for 1622 calcifications, and the calcification testing set has images for 326 calcifications [45].

4.3 INBreast

INBreast contains 115 cases with a total of 410 images. 90 cases were diagnosed with cancer on both breasts out of 115 cases. The dataset includes four different types of breast diseases breast mass, breast calcification, breast asymmetries, and breast distortions. The dataset contains images of (CC) and (MLO) views; the images were saved in DICOM format. Also, the dataset provides the Breast Imaging-Reporting and Data System (BI-RADS) score for breast density [56].

4.4 Mini-MIAS

The dataset includes 322 digital films, also the ground truth markings for any existing abnormality. The categories of the abnormalities that are included in the dataset are calcifications, masses, architectural distortion, asymmetry, and normal. The size of the images was reduced to become 1024 × 1024. The images are publicly available on the Pilot European Image Processing Archive (PEIPA) which belongs to the University of Essex [77].

4.5 BCDR

The BCDR is mainly divided into two mammographic repositories: (1) Film Mammography-based Repository (BCDR-FM) and (2) Full Field Digital Mammography-based Repository (BCDR-DM). BCDR repositories provide normal and abnormal cases of breast cancer with its mammography lesions outlines and related clinical data. The BCDR-FM includes 1010 cases which are for 998 females and 12 males. Furthermore, it includes 1125 studies, 3703 mammographic images in the two views MLO and CC with 1044 identified lesions. The BCDR-DM, still under construction; till now it contains 724 cases 723 cases of them are for
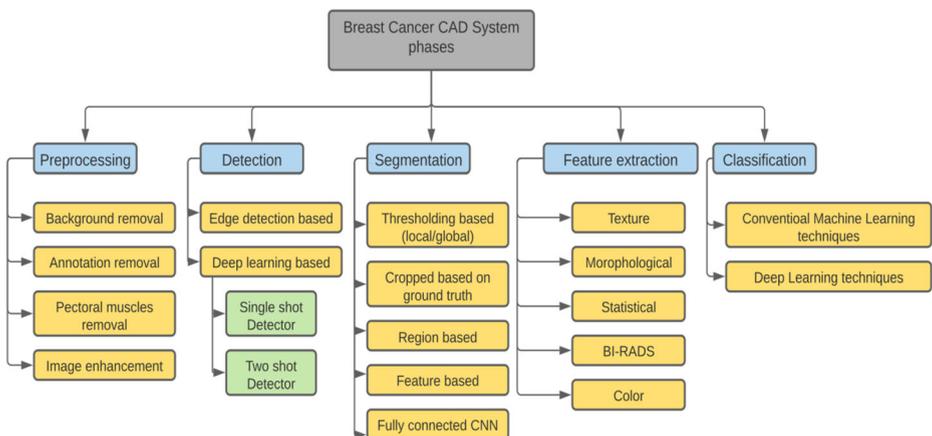

Fig. 6 Most of the used techniques in the different phases of breast cancer CAD systems

females and 1 case for a male, the repository includes 1042 studies. It provides 3612 MLO and CC mammography images and 452 identified lesions [50].

4.6 OPTIMAM

It is composed of more than 2.5 million images that were collected from three UK breast screening centres for 173,319 cases; all of the cases are women. The dataset is divided into 154,832 cases with normal breasts, 6909 cases with benign cancer, 9690 cases with identified lesions and 1888 cases have interval cancers. It provides unprocessed and processed medical images, the dataset includes the region of interest annotations and clinical data relating to the identified cancers and the interval cancers [38].

5 Breast cancer CAD systems

Through the past decades, machine learning contributes significantly to creating more reliable CAD systems for breast cancer diagnosis that can help radiologists in interpreting and reading mammograms. Many studies introduced models for breast cancer diagnosis and prognosis through mammograms, and many of these methods showed very promising performance, however, they aren't tested over a unified large database. The breast cancer CAD systems are composed of some phases that differ based on the task of the CAD system. As shown in Fig. 6 these phases can be divided into pre-processing [31, 76], mass detection, mass segmentation [8, 33, 81], feature extraction [27, 28, 44, 49] and mass classification. Also, the figure presents most of the used techniques in these different phases that can be used in a breast cancer CAD system.

Lusted was the first one to discuss the analysis of the radiographic abnormalities using computers [52] in 1955 as shown in Fig. 7. The researchers in the 60s and 70s started to work toward creating automated methods for classification and detection for the abnormalities in the medical images including the breast images. In 1987, a team from the University of Chicago introduced an automated system that can aid the radiologist in the detection of microcalcification in mammograms by providing the radiologist with analysis output of the image [17].

As shown in Fig. 7. Starting from 90's the research efforts increased toward CAD systems. In 1998 the U.S. Food and Drug Administration (FDA) approved the first CADe system, then
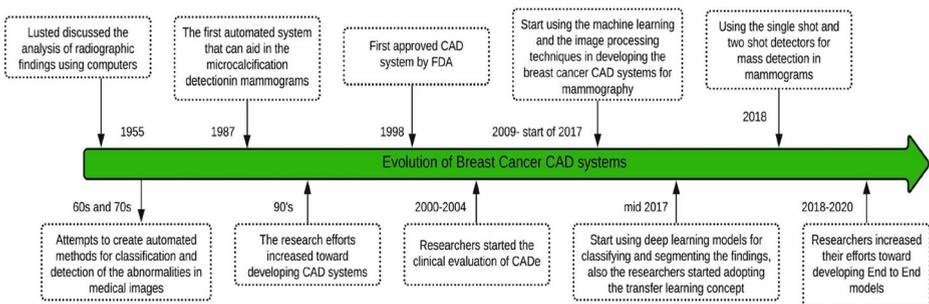

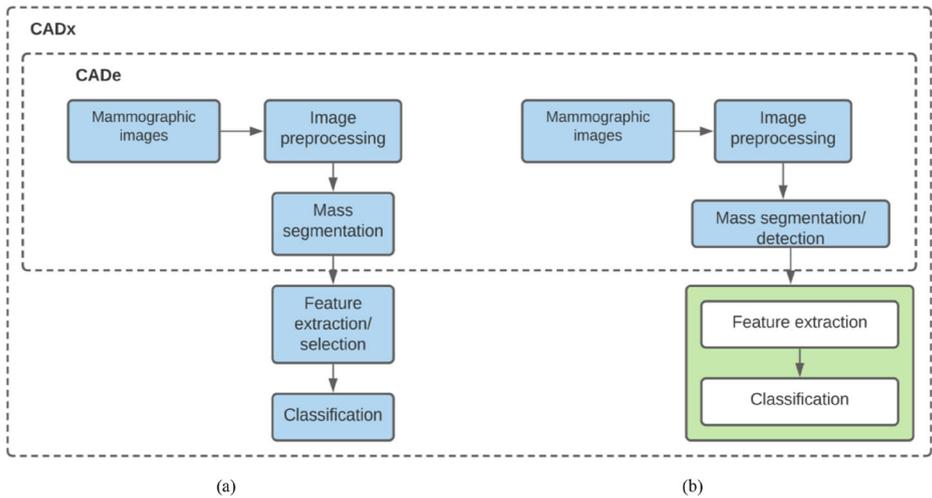**Fig. 7** Timeline for the evolution of breast cancer CAD systems

Fig. 8 Illustration for (a) Conventional machine learning based CADe/CADx system pipeline, (b) Deep learning-based CADe/CADx system pipeline

from 2000 to 2004, the researchers started to evaluate CADe [30, 37] to assess the effectiveness of the clinical use of CADe and its impact on cancer detection rate in mammography. From the beginning of 2009 to 2017, different conventional machine learning-based CAD systems were introduced to enhance abnormalities detection and classification in mammograms. With the appearance of deep learning networks, the researchers started in the middle of 2017 to adopt deep learning models and the transfer learning concept in developing more accurate mammographic CAD systems. The deep learning detection models showed very promising results at the abnormalities' detection based on the results of the proposed system in 2018 that adopted those detection models. Recently from 2018 till now, the researchers started to create end to end models for mammographic CAD systems. In this survey, the existing CAD systems are categorized into conventional CAD systems and deep learning-based CAD systems. Figure 8 illustrates the pipeline of the conventional learning-based CADe / CADx and deep learning base CADe / CADx. The pipeline of conventional machine learning started with image processing then mass segmentation, followed by feature extraction and selection and finally the classification. On the other hand, the deep learning-based CAD system pipeline goes through the same phases except for the feature extraction and classification as these two phases are done as a single phase as the deep learning models can extract the features automatically through the training phase. In the CADe systems, the process stops at the mass segmentation/detection phase.

5.1 Conventional CAD systems

Several trials and studies were proposed to develop CAD systems that can act as a second opinion or helper for the radiologists, these trials started with the use of the traditional computer vision techniques that are based on conventional machine learning and image processing techniques. This section demonstrates some of these studies with details.

Rejani, Y, and S. Thamarai Selv (2009) [65] presented an algorithm for tumour detection in mammograms, their work aimed to discuss a solution for two problems; the first one was about extracting the features that characterize the tumours and the second problem was about how to detect the masses especially the ones that have low contrast with their background. For the mammogram enhancement, they applied a Gaussian filter, top hat for eliminating the background, and Discrete Wavelet Transform (DWT). The mass region segmentation was implemented using the thresholding technique, then the morphological features were extracted from these segmented regions, and Support Vector Machine (SVM) was used for classification. Their approach achieved a sensitivity of 88.75%, however, their work needs to be tested on larger datasets as they applied their method only on 75 mammograms from the mini-MIAS dataset.

Ke, Li et al. (2010) [42] introduced a system that can detect the mass based on the texture features. They used the bilateral comparison to detect the masses and locating the Region of interest (ROI). They implemented fractal dimension and the two-dimension entropy to extract the texture features from the ROI. The ROIs were classified into a mass or normal using SVM. They run their experiment over 106 mammograms, and the results showed that their automated diagnosis method achieved a sensitivity of 85.11%.

Dong, Min, et al. (2015) [24] proposed an automated system to detect and classify the breast masses in the mammographic images. They extracted the position of the masses and the ROI using the chain codes that are provided with the DDSM dataset, then the intensity values were mapped linearly to new values based on the grey level distribution that ranges from 0 to 255. They applied the Rough Set (RS) method to apply more enhancements to the ROIs. To segment the masses from the ROIs, they used an improved Vector Field Convolution Snake (VFCS), which showed robustness to the interference of the blurry tissues. Multiple features were extracted from the segmented masses and the background of the ROIs. For classification, they applied two classifiers the first one was an optimized SVM with particle swarm optimization (PSO) and genetic algorithm (GA), while the other one was random forest (RF). They applied their experiment on DDSM and MIAS datasets. The results showed that the first method outperformed the second one with an accuracy of 97.73% on the DDSM; However, their work needs to be experimented on a larger sample data size through augmentation or using a larger dataset.

Also, Rouhi, Rahimeh, et al. (2015) [68] proposed two different methods for mass segmentation. The ROIs were cropped based on the chain codes of the DDSM dataset. Histogram equalization and median filtering were applied to reduce the noise. For segmentation, they implemented two different techniques namely region growing-based method and cellular neural-based method. They applied Genetic Algorithm (GA) with different chromosome structures and fitness functions for feature selection. The masses were classified into benign and malignant using different classifiers namely Multi-Layer Perceptron (MLP), Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM) and K- Nearest Neighbour (KNN). They run their experiment using DDSM and MIAS datasets. Their method showed high sensitivity of 96.87% for classification with the use of the second segmentation technique, however, the results showed variability as shown in Table 2 for DDSM and MIAS.

Mughal et al. (2017) [58] used texture and colour features to present a system that can detect and classify the masses in mammograms. They applied Contrast Limited Adaptive Histogram Equalization (CLAHE) for enhancing the contrast of the mammogram. Moreover, the mean filter was used as well as the wavelet transform to reduce the noise. They introduced a segmentation method which is composed of two phases, firstly they extracted the normal

Table 2 Pros and limitations of some breast cancer conventional machine learning based CAD systems

Ref	Task	Dataset	Pros	Limitations	Results Detection/Segmentation	Results Classification
Dong, Min, et al. (2015) [24]	Mass segmentation and classification	DDSM and MIAS	-The improved segmentation technique that is robust to the interference of the blurry tissues	-They used a limited sample of the data in their experiment	—	Acc=93.24%
Rouhi, Rahimeh, et al. (2015) [68]	Mass segmentation and classification	DDSM and MIAS	-The model shows high sensitivity for classification	-The results were variable on DDSM and MIAS, so they have to conduct more experiments on other datasets.	—	MIAS: Acc=90.1% DDSM: Acc=96.4%
Punitha, S. et al. (2018) [61]	Mass segmentation and classification	DDSM	-The use of the dragonfly algorithm with the region growing algorithm enhanced the segmentation result.	- They need to test their model on more datasets. - They have to test more classification models.	Jaccard index = 90%	Sensitivity=98.1%
Suhail et al. (2018) [78]	Microcalcification classification	DDSM	-They introduced novel automated feature extraction through dimensionality reduction	- They are classifying only the microcalcification - Their approach may show good results in classifying masses or the mammographic images themselves.	—	Acc=96%
Mohanty Figlu et al. (2020) [55]	Mammogram classification	MIAS, DDSM and BCDR	-Their proposed scheme classifies the mammograms in the real-time	- ROIs are manually cropped.	—	(Normal &Abnormal) DDSM: Acc= 99.92% MIAS: Acc=99.62% (Benign & malignant) MIAS: Acc=99.28% DDSM: Acc=99.63% BCDR: Acc=99.60%
Muduli Debendra et al. (2020) [57]	Mass classification	MIAS and DDSM	- Their feature reduction techniques minimized the computational time.	- They used a small sample size of data in their experiment - Their work needs to be tested on a larger sample data size.	—	MIAS (Normal and Abnormal): Acc= 99.9% DDSM (Benign and Malignant): Acc = 99.6%

breast region by highlighting the pectoral muscle to remove it. To highlight the pectoral muscle the greyscale image was transformed to RGB followed by a transformation to the hue saturation value (HSV), then each RGB value was represented with a value in a range from 0 to 1. In the second phase, they extracted the abnormal breast boundary region by creating a texture image using a function based on an entropy filter. Moreover, they used a mathematical morphology function to extract and refine the ROI. They applied mathematical expressions to extract the intensity, texture, and morphological features. Different classifiers namely SVM, decision tree, KNN, and bagging tree were used for classification. The SVM with (quadratic kernel) showed the best results as it achieved a sensitivity, specificity, and accuracy of 98.40% 97.00% 96.9% respectively for DDSM and 98.00% 97.00% 97.5% for MIAS.

Punitha, S. et al. (2018) [61] presented an automated detection method for masses in mammograms. They used the gaussian filter for smoothening the grey level variations and reduce the noise in the image. An enhanced version of the region growing method with the dragon-fly optimization technique was used for segmentation. Forty-five features were extracted from the ROIs; the Gray level co-occurrence matrices, and Gray level Run Length Matrix (GLRLM) were used for texture analysis and to extract other features. A Feed-Forward Network was used for classification, moreover, they trained this network using Back Propagation with the Levenberg Marquardt algorithm. In the experiment, they used 146 malignant cases, and 154 benign cases from DDSM. They divided these cases into a training set and testing set in which 100 images for testing and 200 images were used for training. It was shown that the use of the dragonfly with the growing region algorithm improved their segmentation results and accordingly the classification as the approach achieved Sensitivity of 98.1% Specificity of 97.8%.

In the same year Suhail et al. [78] proposed an approach to classify the existing microcalcifications in the mammographic images into benign and malignant. Their approach depends on using two stages scalable Linear Discriminant Analysis (LDA) algorithms for extracting the features and reducing the dimensionality, as the binary classification data is encoded to a one-dimensional representation of the microcalcification data. The classification was applied using five classifiers, which are K-NN, SVM, DT, Baysian Network, and ADTree. To evaluate the performance of their approach, they compared their technique (scalable LDA) with the PCA/LDA technique. The results showed the scalable LDA outperformed the PCA-LDA. The classification accuracy for SVM, Baysian Network, K-NN, DT and ADTree were 96%, 0.975, 0.972, 0.975, and 0.985, respectively. This work can be extended to classify the masses besides the microcalcification.

The Extreme Learning Machine (ELM) [23] which is a type of feedforward network with a single hidden layer were used in some studies [55, 57]. Mohanty Figlu et al. (2020) [55] proposed a CAD system for mass classification that showed high accuracy with the use of a reduced number of features. They classified mammograms into normal and abnormal, also their system provides a classification for the mass if it's benign or malignant. They used the DDSM, MIAS and the BCDR datasets to validate their proposed approach. In their approach, the chaotic maps and concept of weights are fused in the salp swarm algorithm for selecting the optimal features set and also to tune the parameters of the KELM algorithm. Their approach is mainly divided into four steps: firstly, they generated the ROI using the ground truth locations, then they extracted the tsallis entropy, energy-Shannon entropy, and renyi entropy from the ROI through the discrete wavelet transform (DWT). For the feature reduction, they applied principal component analysis (PCA) [1]; finally, they used a modified learning approach which is based on ELM for classification. Their technique achieved an accuracy of 99.62% for MIAS and 99.92% for DDSM, for normal and abnormal classification. On the other hand,

for the benign-malignant classification, it showed an accuracy of 99.28% for MIAS, 99.63% for DDSM, 99.60% for BCDR. Although their model can classify the mammograms in real-time, the manually cropped ROIs is considered a weak point in such an automated CAD system.

In [57] Muduli Debendra et al. (2020) merged the ELM with the Moth flame optimization (MFO) algorithm which is a meta-heuristic algorithm to tune the ELM network parameters (i.e., weights, the bias of hidden nodes) to resolve the problem of the ill-conditioned problem in the hidden layer of the network. Also, they applied a fusion between the PCA and LDA for feature reduction and accordingly reducing the computational time, the approach achieved an accuracy of 99.94% for MIAS and %, 99.68% for DDSM, however, they need to run their work over a larger sample set of data.

The authors summarize some of the breast cancer detection methods that are based on the conventional machine learning models in Table 2 to illustrates the pros and limitations of these studies, the task of the proposed model, the results and the used datasets.

5.2 The deep learning-based CAD system

Recently, many promising deep-learning models that are used in computer vision showed significant improvements in the CAD systems performance especially Convolutional Neural Network (CNN), transfer learning approach, and the deep learning-based object detection models. Several algorithms were proposed for the CAD systems based on the use of deep learning models.

Dhungel Neeraj et al. (2017) introduced a CAD tool for mass detection, segmentation and classification in mammographic images with minimal user intervention [22]. For mass detection, they used random forest and cascade of deep learning models, followed by hypothesis refinement. Moreover, they extracted a partial image from the detected masses after refining it by active contour models to segment the masses. They used a deep learning model for classification, the model was pre-trained on hand-crafted feature values, the work was tested on the INBreast dataset. The results showed that the system detected almost 90% of masses with 1 false-positive rate per image, while the accuracy of segmentation achieved 0.85 (Dice index), and the model reached a sensitivity of 0.98 for classification.

In the same year, Geras et al. [34] developed a Deep Convolutional Network (DCN) that can handle multiple views of screening mammography, as the network takes the CC and MLO views for each breast side of a patient. Furthermore, the model works on large high-resolution images with a size of 2600 X 2000; the model learned to predict the assessment of a radiologist and classifying the image based on Breast Imaging-Reporting and Data System (BI-RADS) [47] to "incomplete", "normal" or "benign". In their work, they investigated the impact of the size of the dataset and image resolution on the screening performance. The results showed that when the size of the training set increased the performance increases, also they found that with the original resolution, the model achieves its best performance. It was shown the model achieved a macUAC of 0.688 in a reader study that was done on a random set from the private dataset that they used in their experiments, while a committee of radiologists achieved the macUAC of 0.704.

Al-antari et al. (2018) [5] proposed a deep belief network-based CAD system. For detecting the initial suspicious regions, they used the adaptive thresholding method, which achieved an accuracy of 86%. They adopted two ways to extract the ROIs; in the first, multiple mass regions of interest were extracted as they randomly extract four non-overlapping ROIs of size

32 × 32 pixels around the center of each mass. The second technique depends on extracting the whole mass region of interest as a rectangular box placed around masses and the irregular shapes that are extracted manually. The morphological and statistical features were extracted from these ROIs to be used in classification, they applied different classifiers namely Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), Neural Network (NN), and deep belief network (DBN). The DBN outperformed the other classifiers with an accuracy of 92.86% with the first ROI extraction technique, while it achieved 90.48% with the second ROI extraction technique.

Shen et al. (2020) [71] proposed a framework that depends on adversarial learning to detect the masses in the breast mammograms in an attempt to facilitate the annotation process for the masses in the mammograms through an automated process. The framework consists of two networks, the first network is a Fully Convolutional Network (FCN) to predict the spatial density, and the second one is a domain discriminator that works as a domain transfer that utilizes the adversarial learning to align the low-annotated target domain features with the high-annotated source domain features. The FCN takes the source and the target domains as input to generate a pixel-wise heatmap for them, as every pixel in the heatmap indicates whether the corresponding input pixel relates to a mass lesion. Then the heatmap of the target domain is fed into another network that acts as a domain discriminator which is used to decrease the difference of the heatmap distribution between the source and target domains. They compared their approach with state of art approaches, their approach achieved an AUC score of 0.9083 for a private dataset, and 0.8522 for INBreast.

5.2.1 Transfer learning

Transfer learning is considered recently as one of the keys that aim to enhance the performance of the learner models. It can be defined as the concept of transferring the knowledge acquired for a task to solve a related task. Transfer learning is used widely nowadays in most of the developed CAD systems to resolve the problem of having a non-sufficient amount of data, also it reduces the computational cost and the time needed for training the models [82]. It is a well-known methodology in the deep learning discipline where the pre-trained models can be adapted to be used with other tasks such as computer vision tasks. And so, this can accelerate the computational time that is needed to develop a neural network from scratch, also transfer learning resolved the problem of the difficulty of getting vast amounts of labelled data, considering the time and effort that is needed for that [94].

Some studies recently adopted the transfer learning approach in developing their CAD systems [2, 40, 46]. Ragab et al. (2019) [62] introduced a CAD system that aims to classify mammogram masses into benign and malignant. They applied two different segmentation techniques; the first technique depended on manually cropping the ROI using a circular contour that is provided with the dataset, while the second technique adopted the thresholding and the region-based to crop automatically the ROI. The features were extracted using deep AlexNet architecture-based CNN, then these features were fed through the last fully connected layer in the CNN to SVM classifier for classification. Based on their results the second segmentation technique outperformed the first one. The best results that the model achieved were accuracy of 80.5%, AUC of 88%, and sensitivity of 77.4%, for DDSM. Moreover, the results showed that the segmentation accuracy increased to 73.6% when using samples from the CBIS-DDSM dataset, furthermore, the classification accuracy enhanced to became 87.2% with an AUC of 94%.

Ansar et al. (2020) [12] presented a MobileNet based architecture model that was able to classify the masses in the mammograms into malignant and benign with competitive performance relative to the state of art architectures and less computational cost. The proposed approach firstly detects the masses in the mammogram through classifying the mammograms into cancerous and non-cancerous using a CNN, then the cancerous ones are fed into a pre-trained MobileNet based model to be classified. They compared the performance of their model with the performance of VGG-16, AlexNet, VGG-19, GoogLeNet and ResNet-50. Their model showed competitive performance with an accuracy of 86.8% for DDSM and 74.5% for CBIS-DDSM.

5.2.2 Deep learning-based- object detection (single shot and two shot detectors)

Deep learning replaced the use of the hand-crafted features through learning automatically the most relevant image features to be used to perform a specific task. Object detection is one of the disciplines that showed very promising performance with the use of deep learning. Object detection deep learning-based techniques can be classified into two types, one-stage detectors that are based on regression or classification, and two-stage detectors that are based on regional proposals [89]. Anchor boxes are considered as the key concept behind both of those techniques, it's one of the main factors that affect the performance of the detector in detecting the objects within the image [90].

One stage detector mainly depends on taking one shot of the image to detect more than one object within the image. On the other hand, the regional proposal network (RPN) based approaches are working through two phases, one for generating the candidate region proposals, while the other stage is responsible for detecting the object for each candidate. One stage detector is much faster compared with two-stage detectors as the detection and the classification are done simultaneously over the whole image once; however, the RPN based approaches showed more accurate results [87].

Ribli Dezső, et al. (2018) [67] adopted one of the two-shot detectors named Faster R-CNN [66] to build a system that can detect, localize and classify the abnormalities in mammograms. They used the DDSM in their work, accordingly they mapped the pixel values to optical density due to the low quality of digitized film-screen mammograms, then they rescaled the pixel values to the 0–255 range. Through their experiment model, they noticed that the higher resolution images give good results. They used the INbreast dataset for testing, and a private dataset besides the DDSM dataset for training. The final layer in their model classifies the masses into benign or malignant, also the model generates a bounding box for each detected mass. Furthermore, the model provides a confidence score that indicates to which class the mass belongs. Their model achieved an AUC of 0.95 for classification and was able to detect 90% of the malignant masses in the INbreast dataset with 0.3 false-positive rate/image. The limitation of this work is that they tested their work only on INBreast due to the lack of pixel annotated publicly available datasets, so their model should be tested on larger datasets to generalize their results.

In [6, 10] Al-antari et al. used You Only Look Once (YOLO) [64] in their work for detecting masses in mammograms. In [6] (2018) they proposed a fully automated breast cancer CAD system that is based on deep learning in its three phases of mass detection, segmentation and classification. They used YOLO for detecting and localizing the masses. In the next phase, they used a Full Resolution Convolutional Network (FRCN) to segment the detected masses. Then the segmented masses were classified into benign and malignant through a pre-trained

Table 3 Pros and limitations of some breast cancer deep learning-based CAD systems

Ref	Task	Dataset	Pros	Limitations	Results Detection/ Segmentation	Classification
Geras et al. (2017) [34]	Indicating possible suspicious findings (screening) and predicting the BI-RADS of the findings.	Private dataset	- Investigating the effect of the size of the data set and image resolution on the screening performance. - Utilizing the multi-view images for each patient.	-They did not perform a systematic search for optimal hyperparameters for their model due to the limited computational resources that they used in their experiment	AUC of 0.688	—
Dhungel Neeraj et al. (2017) [22]	Detection and classification	DDSM for training and INBreast for testing and private dataset	- High sensitivity with low false-positive rate/ image (for INbreast dataset)	-Best detection performance was achieved only with the small INbreast dataset. -lack of the pixel-level annotated publicly available dataset.	-Detected 90% of the malignant masses (for INBreast dataset) with 0.3 false positive per image.	AUC=0.95 for INBreast
Al-antari et al. (2018) [5]	Detection and classification	DDSM	-DBN outperforms the LDA QDA and NN in the classification	-They used handcrafted features in the classification	Acc=86%	Acc=92.86%
Al-antari et al. (2018) [6]	Detection, segmentation and classification	INBreast	-The classification accuracy increased with segmentation which was implemented using a full resolution convolutional network (FrCN) -Their approach detects the masses through a single shot detector (YOLO) which means less information loss	- Detection of small masses. - They eliminated manually the false localized masses before segmentation, which is impractical for an automated CAD system.	Detection: Acc=98.96% Segmentation: Acc= 92.97%	Acc=95.64%
Ragab et al. (2019) [62]	Segmentation and classification	DDSM and CBIS-DDSM	-Their proposed system can detect masses and microcalcifications.	- If segmentation was done with the DL model may that enhance the accuracy	—	DDSM: Acc= 80.5% CBIS-DDSM: Acc=87.2%
Shen et al. (2019)	Detection	CBIS-DDSM and INBreast	- Their approach showed that End-to-End training fashion can achieve more accurate classification results	- The mammographic images were downsized due to GPU limitations, while some studies proved that the full resolution of modern digital mammogram images provides	CBIS-DDSM: AUC of 0.88/ image.	—

Table 3 (continued)

Ref	Task	Dataset	Pros	Limitations	Results	
					Detection/ Segmentation	Classification
[70]				better details of the ROIs which accordingly can enhance the performance.	The average AUC of the four-model improved to 0.91 with sensitivity: 86.1%, specificity: 80.1%	—
Shen et al. (2020) [71]	Detection	INBreast and private dataset	Their training strategy addressed: -The problem of Oscillation -The limitation of small batch size.	Their methodology needs to be investigated in more medical imaging scenarios.	Private dataset: AUC=0.9083 INBreast: AUC=0.8522	
Ansar et al. (2020) [12]	Classification	DDSM and CBIS-DDSM	Their approach reduced the computational cost compared to existing models	Low classification accuracy compared to other approaches.	—	DDSM: Acc= 86.8% CBIS-DDSM: Acc=74.5%

Table 4 Quantitative comparison for results of mass detection and classification techniques that used DDSM / CBIS-DDSM datasets

Ref	Method	Detection				Classification			
		AUC	Acc.	Sensitivity	Specificity	AUC	Acc.	Sensitivity	Specificity
Ertosun MG, Rubin DL. (2015) [29]	Whole-image Mass Localization	–	85%	–	–	–	85%	–	–
Lévy D, Jain A. (2016) [46]	GoogLeNet pre-trained classifier	–	–	–	–	–	92.9%	93.4%	92.4%
Platania R, et al. (2017) [60]	YOLO based ROI Classifier	–	90%	–	–	92.3%	93.5%	94%	93%
Lotter W, et al. (2017) [51]	Multi-scale CNN (patches)	–	–	–	–	92%	–	–	–
Chougrad H, et al. (2018) [21]	Pre trained InceptionV3 Classifier	–	–	–	–	98%	97.3%	–	–
Al-masni MA, et al. (2018) [10]	YOLO based Classifier	–	99.7%	–	–	–	97%	–	–
Yan Y, et al. (2019) [85]	Multiscale Autoencoder Segmentation	–	–	72.3%	99.6%	–	–	–	–
Shen L, et al. (2019) [70]	Pre-trained patch classifier (ResNet50, VGG16)	–	–	–	–	91%	–	86.1%	80.1%
Sun H, et al. (2020) [79]	Attention-guided Dense Upsampling segmentation	–	–	84.9%	–	–	–	–	–
Rahman AS, et al. (2020) [63]	Pre-trained Tumor Classifier (ResNet50)	–	–	–	–	–	85.7%	87.3%	85.7%

Table 5 Quantitative comparison for results of mass detection and classification techniques that used INBreast dataset

Ref	Method	Detection				Classification			
		AUC	Acc.	Sensitivity	Specificity	AUC	Acc.	Sensitivity	Specificity
Zhu W, et al. (2016) [92]	Deep multi-instance network	–	–	–	–	85.8%	90%	–	–
Dhungel Neeraj et al. (2017) [22]	Bounding Box CNN Detection	–	90%	–	–	76%	91%	–	–
Ribli D et al. (2018) [67]	Faster-RCNN Detection with Cross-Domain Training	–	–	90% of malignant masses	–	–	95%	–	–
Chougrad H, et al. (2018) [21]	Pre-trained InceptionV3 Classifier	–	–	–	–	97%	95.5%	–	–
Jung H, et al. (2018) [41]	RetinaNet based Detector	–	97%	–	–	–	–	–	–
Shen L, et al. (2019) [70]	Pre-trained patch classifier (ResNet50, VGG16)	–	–	–	–	98%	–	86.7%	96.1%
Sun H, et al. (2020) [79]	Attention-guided Dense Upsampling segmentation	–	–	80.8%	–	–	–	–	–
Shen R, et al. (2020) [71]	Adversarial Domain Transfer Learning	85.2%	–	–	–	–	–	–	–
Agarwal R, et al. (2020) [3]	Faster-RCNN Detection	–	–	95–71%	70%	–	–	–	–

CNN that based on AlexNet architecture. The system achieved a mass detection accuracy of 98.96%, segmentation accuracy of 92.97%, and classification accuracy of 95.64%. Moreover, in [7] (2020) they proposed the same model they introduced in [6], with some improvements in the classification, and segmentation phase. After these improvements, YOLO achieved detection accuracy of 97.27%, breast lesion segmentation accuracy of 92.97%. CNN, ResNet-50, and InceptionResNet-V2 were used for classification and achieved an average overall accuracy of 88.74%, 92.56%, and 95.32%, respectively.

Cao et al. (2021) [16] proposed a novel model for detecting breast masses in mammograms, furthermore, they proposed a new data augmentation technique to overcome the overfitting problem due to the small dataset. Their augmentation technique is based on local elastic deformation, this technique enhanced the performance of their model; however, its calculation speed is slower compared to the traditional augmentation techniques. In their approach, they firstly segment the breast to remove most of the background through Gaussian filtering and the Otsu thresholding method. Moreover, they used an enhanced version of the RetinaNet named FSAF [93] for mass detection. Each image has an average of 0.495 false-positive rate for INBreast, while for the DDSM dataset each image has 0.599 false-positive rate.

5.2.3 End to end models

The End to End (E2E) learning approach is the concept of replacing a pipeline of several modules in a complex learning system with a single model (deep neural network). E2E training approach enhances the performance of the model as it allows a single optimization criterion instead of optimizing each module separately under different criteria as in the pipelined architecture [35]. Recently different studies build their models based on the E2E training approach that showed promising results.

Shen et al. (2017) introduced in [70] a CNN based end to end model to detect and classify the masses within the whole mammographic image, moreover, in (2019) [70] they improved the work they introduced in [70] by classifying the local image patches through a pre-trained model on a labelled dataset that provides the ROI data. They initialized the weight parameters of the whole image classifier with the weight parameters of the pre-trained patch classifier. They used the two pre-trained CNN models that are Resnet50 and VGG16 to build four classification models. They used CBIS-DDSM to train the patch and the whole image classifiers, then with the use of transfer learning they transferred the whole image classifier for testing over the INbreast dataset. The patch images were classified into 5 classes which are background, benign/malignant mass, and benign/malignant calcification. In their results, the best single model that tested on CBIS-DDSM achieved an AUC of 0.88 per image, while the average AUC of the four-model was up to 0.91. Also, the INbreast dataset showed that the AUC of the best single model achieved 0.95 per image, and the average AUC of the four-model improved to be up to 0.98. In this work, they downsized the images due to GPU limitations and this led to losing some information of the ROIs, if this information was retained maybe it can differ in the performance of this approach.

Agnes et al. (2020) [4] presented a Multiscale CNN which is based on an end-to-end training strategy. The main task of their model is to classify the mammographic images into normal and malignant. Their model is mainly divided into two parts; context feature extraction and mammogram classification. The model implements a multi-level convolutional network

that can extract both high- and low-level contextual features from the image. The model achieved an accuracy of 96.47% with an AUC of 0.99 for the mini-MIAS dataset.

Table 3 Provides a summary of the pros, limitations, tasks and results for some of the deep learning-based breast cancer CAD systems.

6 Dataset based quantitative comparison

Tables 4 and 5 provides a quantitative comparison between some selected techniques that were proposed on a per- dataset basis. The selected techniques were used DDSM, CBIS-DDSM and INBreast in their work. The selected datasets are the most used ones in most of the state of art proposed models.

It can be demonstrated from Tables 4 and 5 that there's an improvement at both levels; mass detection and classification. Still, mass detection needs more work to enhance the detection of true positives to increase the sensitivity as the best sensitivity that was reached was 90%. In addition to that, the specificity which indicated the true negative still needs extensive efforts to enhance it.

7 Evaluation metrics for breast cancer CAD systems

This section presents the most used evaluation metrics for evaluating the performance of the breast cancer CAD systems. Various performance measures are used for evaluating and analyzing the performance of the CAD systems at detection and classification [91].

Intersection over Union (IoU) is one of the most used methods to evaluate the performance of the detection. IoU represents the amount of overlap between the ground truth and the predicted bounding box. It can be calculated as shown in Eq. (1).

$$\text{IoU} = \frac{A \cap B}{A \cup B}, \text{ where } A \text{ is the predicted bounding box and } B \text{ is the ground truth box} \quad (1)$$

Also, sensitivity, specificity and accuracy are used for evaluating both abnormality detection and classification. Furthermore, the confusion matrix must be taken into consideration as it represents the number of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). Table 6 provides an illustration for the structure of the confusion matrix where:

- True Positive (TP): is the number of times that the system correctly detected or classified the present masses in the mammography as positive.
- False-positive (FP): represents the number of the negative masses that are incorrectly detected or classified as positive.

Table 6 Illustration for the basic framework of the confusion matrix

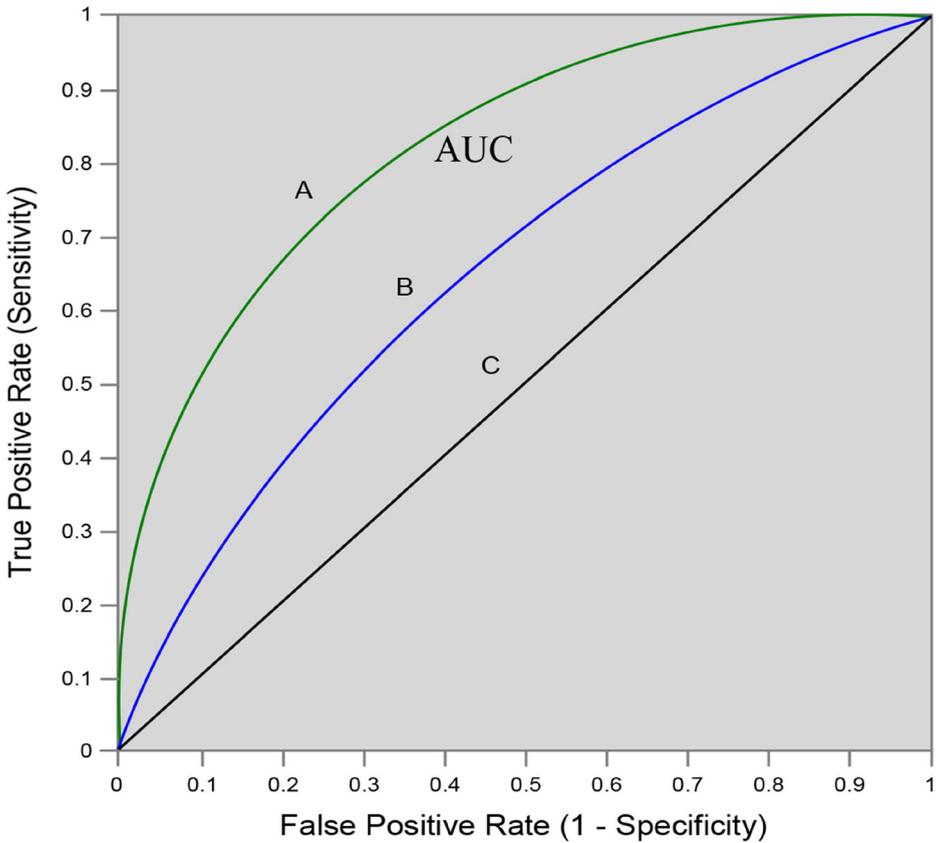| Actual class		Predicted class	
		Positive	Negative
	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Fig. 9 Receiver operating characteristics (ROC)

- True Negative (TN): is the negative masses that are correctly detected or classified within the mammogram image as negative.
- False Negative (FN): represents the masses that are existing in the mammogram and weren't detected or classified correctly.

Sensitivity, Recall or True Positive Rate (TPR) is the probability that the actual positive is correctly tested as positive masses in the mammogram, and it can be calculated as shown in Eq. (2). Specificity which is also called True Negative Rate (TNR) is the probability of the actual negative that is tested correctly as negative when no abnormalities exist in the mammogram as shown in Eq. (3).

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

Accuracy (Acc) is describing the ratio of the number of the correct predictions regarding the total number of predictions, it's mainly describing the performance of the system regarding all classes as shown in Eq. (4). In addition to these evaluation methods, the Receiver Operating Characteristics (ROC) is considered as one of the important performance measures for CAD systems as it represents the trade-off between TPR and False Positive Rate (FPR) at different classification thresholds. The ROC curve is plotted with two-axes; the y-axis for TPR against the FPR which is represented by the x-axis. The area under the ROC curve (AUC) is indicating to what extent the system can distinguish between positive and negative classes. Figure 9 shows an illustrative example for ROC with different classifiers, as each curve of the three curve lines (A, B and C respectively) represents the ROC while the AUC is the area under each curve line. The higher the value of AUC means the better the model is; For example, as shown in Fig. 9 the ROC curve of A has the higher AUC value which means that the performance of classifier A is better than B and C. The AUC value ranges from 0 to 1; the value of AUC is 0 when the model fails to predict any prediction correctly and equals to 1 when the model can distinguish correctly between all of the negatives and positives.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

F1-Score also is one of the evaluation measures that is used to evaluate the model's performance at the binary classification such as classifying the masses into benign and malignant. Calculating F1-Score depends mainly on the precision and recall, as it can be calculated using the following formula as shown in Eq. (5):

$$F1{-}Score = 2\times = \frac{precision \times recall}{precision + recall} \tag{5}$$

mAP (mean Average Precision) is another evaluation method that mostly used for evaluating the performance of the object detection models. mAP is calculating by calculating the average precision (AP) for each class then taking the average of the Average Precision (AP) for all classes. The mAP is calculated as the following in Eq. (6), where Q is the number of classes in the set and AP is the average precision for a given class q.

$$mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q} \tag{6}$$

8 Discussion and conclusion

In summary, this survey highlights the current deep learning and conventional machine learning techniques for mammographic CAD systems, datasets, and concepts. It can be demonstrated that the studies that have adopted the use of conventional machine learning techniques and algorithms showed good performance with high accuracy rates, however, although these techniques won't perform well with large datasets, and almost all of them

depend on expert crafted features. Over the last few years, the conventional ML techniques have evolved especially with the appearance of deep learning techniques.

Recently, various researchers started to use deep learning models in an attempt to create more reliable CAD systems with fewer false-positive rates. Although the review showed that Deep learning techniques showed very promising performance and significant contribution in the development of CAD systems, there are still some limitations in these techniques especially with the lack of datasets and this complicates its clinical applicability.

From Tables 2, 3, 4 and 5 the current challenges in CAD systems development can be summarized as the following: (1) Increasing the number of the mammographic images to overcome the problem of the insufficient data amount that is using in the experiments is one of the big obstacles especially when it comes to the medical images; it's very hard to find or acquire annotated mammograms at pixel level and image level. (2) Mass localization and detection are still considered a challenging task because the mass features in dense breasts seem to be like the ones in normal tissues, also these tissues mask the cancerous cells [32]. Moreover, the masses' sizes vary hugely [53], and this also makes this task more challenging especially for the small masses. (3) Reducing the false positive rate and increasing the specificity and sensitivity rates needs more work. (4) Selecting the optimal parameters for the DL models is also one of the challenges that need more investigation to build more robust CAD systems.

Due to the insufficient number of mammographic images that are included in the publicly available datasets; data augmentation techniques are needed to create synthetic mammographic images, especially with the appearance of Generative Adversarial Network (GAN) [36]. However, some researchers started to work in this direction [69, 84], but it still needs more investigation to generate large scale of mammographic images in an attempt to solve the imbalanced class problem in the mammography available datasets. Moreover, GAN can be able to generate more realistic images than the ones that are generated through the traditional augmentation techniques like rotation, flipping, cropping, translation, noise injection and colour transformation [74], and this may affect the performance positively and increase the capability of the models to detect and classify the lesions correctly.

Also, more investigation is needed for developing new data augmentation techniques that can preserve the mass features, and add a variation at the morphological level. Furthermore, other different strategies can be used to overcome the problem of the insufficient data amount such as using pre-trained models and so the pre-trained weights are transferred to initialize the network and the parameters are fine-tuned through training [72, 73].

The object detection deep learning-based models like YOLO and Faster RCNN are considered as one of the recent customizable techniques that achieved better detection accuracies and enhanced the mass detection and localization within the mammographic image, however, the small mass detection still needs more investigation, especially for the very close ones. Training these models with enough amount of data that contain more images with small masses may enhance the performance of such models at the small mass detection, also fine-tuning for the bounding boxes can help in overcoming this problem.

The positions of body or images' angles vary in mammographic masses, and so recognition of the texture to be estimated at different angles is important when performing texture analysis for the masses [19]. However, as shown in the literature many studies presented models that used the morphological features such as texture, color, and so on, some studies such as [88] recently listed the problem of absence of neighbourhood invariant components, which can't adequately react to image transformation or changes brought about by imaging points when

classifying the mammographic masses via CNN. And so, they proposed a novel approach that is based on a fusion between the rotation invariant features, texture features and the deep learning for classifying masses in mammograms. This listed problem can be considered as a new challenge that needs more investigation, as it can be extended to harness the rotation invariant features in mass detection through using Rotation Invariant Fisher Discriminative Convolutional Neural Networks (RIFD-CNN) for object detection [20].

It was shown from the review that the studies that recently started to focus on using more than one mammography view in the classification such as MLO and CC views; the use of more than one view proved effectiveness in mammogram classification more than using single-view images [48, 86]. Accordingly, utilizing the multi-view mammographic images in mass detection needs more investigation and research efforts, as it can enhance the sensitivity and specificity rates for mass detection and classification through preserving more information and features from both views. Also, the studies showed that the full image resolution can give better accuracy results [84], so developing systems that can retain the full resolution of the mammographic image are needed to minimize the information loss that occurs due to downsizing the images which affect the image quality.

Based on the aforementioned discussion, breast cancer CAD system development still needs more research efforts to solve the current challenges, especially for the DL models that suffer from the lack of annotated data, and so building deep learning models that can learn from a small size of data is considered as one of the open challenges.

This survey contributed a review of the literature of the past ten years on the state of art methodologies for breast cancer CAD systems specifically for mass detection and classification. This work aimed to help in building CAD systems that can be applied clinically to assist in breast cancer diagnosis. The review provides evaluation for some of the studies presented in the literature through presenting their pros and limitations. The survey gives an overview of the main phases of the CAD system and what are the used techniques in these phases. Moreover, it lists the breast screening modalities and the publicly available mammographic datasets, also, it provides a dataset-based quantitative comparison between the most recent techniques. In addition to that, the evaluation metrics that are used for CAD assessment are demonstrated. Furthermore, the survey presented the current challenges that need more investigation to improve the efficiency and the performance of these systems.

Declarations

Conflicts of interest/competing interests   There's no conflict of interest between the authors.

Ethics approval   Not applicable.

Consent to participate   All the authors approved their participation in this work.

Consent for publication All the authors approved to publish this paper.

References

1. Abdi H, Williams LJ (2010) Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics 2(4):433–459
2. Aboutalib SS, Mohamed AA, Berg WA, Zuley ML, Sumkin JH, Wu S (2018) Deep learning to distinguish recalled but benign mammography images in breast cancer screening. Clin Cancer Res 24(23):5902–5909. https://doi.org/10.1158/1078-0432.CCR-18-1115
3. Agarwal R, Díaz O, Yap MH, Lladó X, Martí R (2020) Deep learning for mass detection in full field digital mammograms. Comput Biol Med 121:103774. https://doi.org/10.1016/J.COMPBIOMED.2020.103774
4. Agnes SA, Anitha J, Pandian SI, Peter JD (2020) Classification of mammogram images using multiscale all convolutional neural network (MA-CNN). J Med Syst 44(1):1–9. https://doi.org/10.1007/S10916-019-1494-Z
5. Al-antari MA, Al-masni MA, Park SU et al (2018a) An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network. J Med Biol Eng 38:443–456. https://doi.org/10.1007/S40846-017-0321-6
6. Al-antari MA, Al-masni MA, Choi MT et al (2018b) 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. https://doi.org/10.1016/J.IJMEDINF.2018.06.003
7. Al-Antari MA, Al-Masni MA, Kim TS (2020) Deep learning computer-aided diagnosis for breast lesion in digital mammogram. Adv Exp Med Biol 1213:59–72. https://doi.org/10.1007/978-3-030-33128-3_4
8. Al-Bayati M, El-Zaart A (2013) Mammogram images thresholding for breast cancer detection using different thresholding methods. Adv Breast Cancer Res 02:72–77. https://doi.org/10.4236/ABCR.2013.23013
9. Ali EA, Adel L (2019) Study of role of digital breast tomosynthesis over digital mammography in the assessment of BIRADS 3 breast lesions. Egypt J Radiol Nucl Med 50(1):1–10
10. Al-masni MA, Al-antari MA, Park JM et al (2018) Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Comput Methods Prog Biomed 157:85–94. https://doi.org/10.1016/J.CMPB.2018.01.017
11. Andersson I, Hildell J, Mühlow A, Pettersson H (1978) Number of projections in mammography: influence on detection of breast disease. AJR Am J Roentgenol 130(2):349–351. https://doi.org/10.2214/ajr.130.2.349
12. Ansar W, Shahid AR, Raza B, Dar AH (2020) Breast cancer detection and localization using mobilenet based transfer learning for mammograms. In: International symposium on intelligent computing systems. Springer, Cham, pp 11–21. https://doi.org/10.1007/978-3-030-43364-2_2
13. Ataollahi M, Sharifi J, Paknahad M, Paknahad A (2015) Breast cancer and associated factors: a review. J Med Life 8:6–11
14. Baker JA, Rosen EL, Lo JY, Gimenez EI, Walsh R, Soo MS (2003) Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting architectural distortion. AJR Am J Roentgenol 181(4):1083–1088
15. Berment H, Becette V, Mohallem M, Ferreira F, Chérel P (2014) Masses in mammography: what are the underlying anatomopathological lesions? Diagnostic and Interventional Imaging 95:124–133. https://doi.org/10.1016/J.DIII.2013.12.010
16. Cao H, Pu S, Tan W, Tong J (2021) Breast mass detection in digital mammography based on anchor-free architecture. Comput Methods Prog Biomed 205:106033
17. Chan HP, Doi K, Galhotra S et al (1987) Image feature analysis and computer-aided diagnosis in digital radiography. I Automated detection of microcalcifications in mammography. Med Phys 14:538–548. https://doi.org/10.1118/1.596065

18. Chen Z, Yang J, Li S, Lv M, Shen Y, Wang B, Li P, Yi M, Zhao X, Zhang L, Wang L, Yang J (2017) Invasive lobular carcinoma of the breast: A special histological type compared with invasive ductal carcinoma. PLoS One 12(9):e0182397. https://doi.org/10.1371/journal.pone.0182397

19. Cheng HD, Shi XJ, Min R, Hu LM, Cai XP, du HN (2006) Approaches for automated detection and classification of masses in mammograms. Pattern Recogn 39:646–668. https://doi.org/10.1016/J.PATCOG.2005.07.006

20. Cheng G, Han J, Zhou P, Xu D (2018) Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection. IEEE Trans Image Process 28(1):265–278. https://doi.org/10.1109/CVPR.2016.315

21. Chougrad H, Zouaki H, Alheyane O (2018) Deep convolutional neural networks for breast cancer screening. Comput Methods Prog Biomed 157:19–30. https://doi.org/10.1016/J.CMPB.2018.01.011

22. Dhungel N, Carneiro G, Bradley AP (2017) A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med Image Anal 37:114–128. https://doi.org/10.1016/J.MEDIA.2017.01.009

23. Ding S, Zhao H, Zhang Y, Xu X, Nie R (2015) Extreme learning machine: algorithm, theory and applications. Artif Intell Rev 44(1):103–115. https://doi.org/10.1007/S10462-013-9405-Z

24. Dong M, Lu X, Ma Y, Guo Y, Ma Y, Wang K (2015) An efficient approach for automated mass segmentation and classification in mammograms. J Digit Imaging 28(5):613–625. https://doi.org/10.1007/S10278-015-9778-4

25. Du H, Feng J, Feng M (2019) Zoom in to where it matters: a hierarchical graph based model for mammogram analysis. arXiv preprint arXiv:1912.07517

26. Elmore JG, Armstrong K, Lehman CD, Fletcher SW (2005) Screening for breast cancer. JAMA : the Journal of the American Medical Association 293:1245–1256. https://doi.org/10.1001/JAMA.293.10.1245

27. Eltoukhy MM, Faye I, Samir BB (2010). Curvelet based feature extraction method for breast cancer diagnosis in digital mammogram. 2010 international conference on intelligent and advanced systems. https://doi.org/10.1109/icias.2010.5716125

28. Eltoukhy MM, Faye I, Samir BB (2012) A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation. Comput Biol Med 42:123–128. https://doi.org/10.1016/J.COMPBIOMED.2011.10.016

29. Ertosun MG, Rubin DL (2015) Probabilistic visual search for masses within mammography images using deep learning. 2015 IEEE international conference on bioinformatics and biomedicine (BIBM). https://doi.org/10.1109/bibm.2015.7359868

30. Freer TW, Ulissey MJ (2001) Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. Radiology 220:781–786. https://doi.org/10.1148/RADIOL.2203001282

31. Ganesan K, Acharya UR, Chua KC, Min LC, Abraham KT (2013) Pectoral muscle segmentation: a review. Comput Methods Prog Biomed 110(1):48–57. https://doi.org/10.1016/j.cmpb.2012.10.020

32. García-Manso A, García-Orellana CJ, González-Velasco HM, Gallardo-Caballero R, Macías-Macías M (2013) Study of the effect of breast tissue density on detection of masses in mammograms. Comput Math Methods Med 2013:1–10. https://doi.org/10.1155/2013/213794

33. George MJ, Sankar SP (2017). Efficient preprocessing filters and mass segmentation techniques for mammogram images. 2017 IEEE international conference on circuits and systems (ICCS). https://doi.org/10.1109/iccs1.2017.8326032

34. Geras KJ, Wolfson S, Shen Y, Wu N, Kim S, Kim E, Cho K (2017). High-resolution breast cancer screening with multi-view deep convolutional neural networks. arXiv preprint arXiv:1703.07047

35. Glasmachers T (2017) Limits of end-to-end learning. In Asian conference on machine learning, pp 17-32. PMLR

36. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial networks. Commun ACM 63:139–144

37. Gur D, Sumkin JH, Rockette HE et al (2004) Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. J Natl Cancer Inst 96:185–190. https://doi.org/10.1093/JNCI/DJH067

38. Halling-Brown MD, Warren LM, Ward D, Lewis E, Mackenzie A, Wallis MG, Wilkinson LS, Given-Wilson RM, McAvinchey R, Young KC (2020) OPTIMAM mammography image database: a large-scale resource of mammography images and clinical data. Radiology Artificial Intelligence 3(1):e200103. https://doi.org/10.1148/ryai.2020200103

39. Heath M, Bowyer K, Kopans D, Kegelmeyer P, Moore R, Chang K et al (1998) Current status of the digital database for screening mammography. Digital Mammography:457–460. https://doi.org/10.1007/978-94-011-5318-8_75

40. Huynh BQ, Li H, Giger ML (2016) Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. Journal of Medical Imaging (Bellingham, Wash) 3:034501. https://doi.org/10.1117/1.JMI.3.3.034501

41. Jung H, Kim B, Lee I et al (2018) Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. PLoS ONE:13. https://doi.org/10.1371/JOURNAL.PONE.0203355

42. Ke L, Mu N, Kang Y(2010). Mass computer-aided diagnosis method in mammogram based on texture features. 2010 3rd international conference on biomedical engineering and informatics. https://doi.org/10.1109/bmei.2010.5639515

43. Kissane J, Neutze JA, Singh H (2020) Breast imaging. In: Radiology fundamentals. Springer, Cham, pp 139–154

44. Kulkarni DA, Bhagyashree SM, Udupi GR (2010) Texture analysis of mammographic images. Int J Comput Appl 5:12–17. https://doi.org/10.5120/919-1297

45. Lee RS, Gimenez F, Hoogi A, Miyake KK, Gorovoy M, Rubin DL (2017) A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific Data 4(1):1–9

46. Lévy D, Jain A (2016) Breast mass classification from mammograms using deep convolutional neural networks. arXiv preprint arXiv:1612.00542

47. Liberman L, Menell JH (2002) Breast imaging reporting and data system (BI-RADS). Radiol Clin 40(3): 409–430

48. Liu X, Zhu T, Zhai L, Liu J (2016) Improvement of mass detection in mammogram using multi-view information. In: Eighth international conference on digital image processing (ICDIP 2016). https://doi.org/10.1117/12.2244627

49. Llobet R, Paredes R, Pérez-Cortés JC (2005) Comparison of feature extraction methods for breast Cancer detection. Lect Notes Comput Sci 3523:495–502. https://doi.org/10.1007/11492542_61

50. Lopez, M G, Posada N, Moura D C, Pollán R R, Valiente J M F, Ortega C S, Araújo B M F (2012) BCDR: a breast cancer digital repository. In: 15th International conference on experimental mechanics (Vol. 1215)

51. Lotter W, Sorensen G, Cox D (2017) A multi-scale CNN and curriculum learning strategy for mammogram classification. Lect Notes Comput Sci:169–177. https://doi.org/10.1007/978-3-319-67558-9_20

52. Lusted LB (1955) Medical electronics. N Engl J Med 252(14):580–585

53. Michaelson J, Satija S, Moore R, Weber G, Halpern E, Garland A, Kopans DB, Hughes K (2003) Estimates of the sizes at which breast cancers become detectable on mammographic and clinical grounds. Journal of Women's Imaging 5:3–10. https://doi.org/10.1097/00130747-200302000-00002

54. Misra S, Solomon NL, Moffat FL, Koniaris LG (2010) Screening criteria for breast cancer. Adv Surg 44: 87–100. https://doi.org/10.1016/J.YASU.2010.05.008

55. Mohanty F, Rup S, Dash B, Majhi B, Swamy MNS (2020) An improved scheme for digital mammogram classification using weighted chaotic salp swarm algorithm-based kernel extreme learning machine. Appl Soft Comput 91:106266. https://doi.org/10.1016/J.ASOC.2020.106266

56. Moreira IC, Amaral I, Domingues I et al (2012) INbreast: toward a full-field digital mammographic database. Acad Radiol 19:236–248. https://doi.org/10.1016/J.ACRA.2011.09.014

57. Muduli D, Dash R, Majhi B (2020) Automated breast cancer detection in digital mammograms: A moth flame optimization based ELM approach. Biomed Signal Process Control 59:101912. https://doi.org/10.1016/J.BSPC.2020.101912

58. Mughal B, Sharif M, Muhammad N (2017) Bi-model processing for early detection of breast tumor in CAD system. Eur Phys J Plus 132(6):1–14. https://doi.org/10.1140/epjp/i2017-11523-8

59. Nalawade YV (2009) Evaluation of breast calcifications. The Indian Journal of Radiology & Imaging 19(4): 282–286. https://doi.org/10.4103/0971-3026.57208

60. Platania R, Shams S, Yang S, Zhang J, Lee K, Park S-J (2017) Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection (BC-DROID). Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. https://doi.org/10.1145/3107411.3107484

61. Punitha S, Amuthan A, Joseph KS (2018) Benign and malignant breast cancer segmentation using optimized region growing technique. Future Computing and Informatics Journal 3:348–358. https://doi.org/10.1016/J.FCIJ.2018.10.005

62. Ragab DA, Sharkas M, Marshall S, Ren J (2019) Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ 7:e6201. https://doi.org/10.7717/PEERJ.6201

63. Rahman AS, Belhaouari SB, Bouzerdoum A, Baali H, Alam T, Eldaraa AM (2020) Breast mass tumor classification using deep learning. 2020 IEEE international conference on informatics, IoT, and enabling technologies (ICIoT). https://doi.org/10.1109/iciot48696.2020.9089535

64. Redmon J, Divvala S, Girshick R, Farhadi A (2016). You only look once: unified, real-time object detection. 2016 IEEE conference on computer vision and pattern recognition (CVPR). 10.1109/cvpr.2016.91

65. Rejani Y, Selvi S T (2009) Early detection of breast cancer using SVM classifier technique. arXiv preprint arXiv:0912.2314

66. Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39:1137–1149

67. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I (2018) Detecting and classifying lesions in mammograms with deep learning. Sci Rep 8(1):1–7. https://doi.org/10.1038/s41598-018-22437-z

68. Rouhi R, Jafari M, Kasaei S, Keshavarzian P (2015) Benign and malignant breast tumors classification based on region growing and CNN segmentation. Expert Syst Appl 42:990–1002. https://doi.org/10.1016/J.ESWA.2014.09.020

69. Shams S, Platania R, Zhang J, Kim J, Lee K, Park SJ (2018) Deep generative breast cancer screening and diagnosis. In: In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, pp 859–867. https://doi.org/10.1007/978-3-030-00934-2_95

70. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W (2019a) Deep learning to improve breast cancer detection on screening mammography. Sci Rep 9(1):1–12

71. Shen R, Yao J, Yan K, Tian K, Jiang C, Zhou K (2020) Unsupervised domain adaptation with adversarial learning for mass detection in mammogram. Neurocomputing 393:27–37. https://doi.org/10.1016/J.NEUCOM.2020.01.099

72. Shie C-K, Chuang C-H, Chou C-N, Wu M-H, Chang EY (2015) Transfer representation learning for medical image analysis. In: Annual international conference of the IEEE engineering in medicine and biology society. IEEE engineering in medicine and biology society. Annual international conference, 2015, vol 2015, pp 711–714. https://doi.org/10.1109/EMBC.2015.7318461

73. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 35:1285–1298

74. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. Journal of Big Data 6(1):1–48. https://doi.org/10.1186/S40537-019-0197-0

75. Sickles EA, Weber WN, Galvin HB, Ominsky SH, Sollitto RA (1986) Baseline screening mammography: one vs two views per breast. AJR Am J Roentgenol 147(6):1149–1153. https://doi.org/10.2214/ajr.147.6.1149

76. Slavine NV, Seiler S, Blackburn TJ, Lenkinski RE (2018). Image enhancement method for digital mammography. Medical imaging 2018: image processing. https://doi.org/10.1117/12.2293604

77. Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, Savage J (2015) Mammographic image analysis society (mias) database v1. 21

78. Suhail Z, Denton ER, Zwiggelaar R (2018) Classification of micro-calcification in mammograms using scalable linear fisher discriminant analysis. Med Biol Eng Comput 56(8):1475–1485. https://doi.org/10.1007/S11517-017-1774-Z

79. Sun H, Li C, Liu B, Liu Z, Wang M, Zheng H, Feng DD, Wang S (2020) AUNet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms. Phys Med Biol 65(5):055005

80. Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71:209–249. https://doi.org/10.3322/CAAC.21660

81. Varughese LS, Anitha J (2013) A study of region based segmentation methods for mammograms. Int J Res Eng Technol 02:421–425. https://doi.org/10.15623/IJRET.2013.0212070

82. Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. Journal of Big Data 3(1):1–40. https://doi.org/10.1186/S40537-016-0043-6

83. Welch HG, Prorok PC, O'Malley AJ, Kramer BS (2016) Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. N Engl J Med 375(15):1438–1447

84. Wu E, Wu K, Cox D, Lotter W (2018) Conditional infilling GANs for data augmentation in mammogram classification. Lect Notes Comput Sci:98–106. https://doi.org/10.1007/978-3-030-00946-5_11

85. Yan Y, Conze PH, Decencière E, Lamard M, Quellec G, Cochener B, Coatrieux G (2019) Cascaded multi-scale convolutional encoder-decoders for breast mass segmentation in high-resolution mammograms. 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC). https://doi.org/10.1109/embc.2019.8857167

86. Yuhang L et al (2021) Act like a radiologist: towards reliable multi-view correspondence reasoning for mammogram mass detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Advance online publication. https://doi.org/10.1109/TPAMI.2021.3085783

87. Zaidi SS, Ansari MS, Aslam A, Kanwal N, Asghar M, Lee B (2021) A survey of modern deep learning based object detection models. arXiv preprint arXiv:2104.11892

88. Zhang Q, Li Y, Zhao G, Man P, Lin Y, Wang M (2020) A Novel Algorithm for Breast Mass Classification in Digital Mammography Based on Feature Fusion. Journal of Healthcare Engineering 2020:2020. https://doi.org/10.1155/2020/8860011

89. Zhao Z-Q, Zheng P, Xu S, Wu X (2018) Object detection with deep learning: a review. IEEE Transactions on Neural Networks and Learning Systems 30:3212–3232
90. Zhong Y, Wang J, Peng J, Zhang L (2020) Anchor box optimization for object detection. In: 2020 IEEE winter conference on applications of computer vision (WACV). https://doi.org/10.1109/wacv45572.2020.9093498
91. Zhou XH, McClish DK, Obuchowski NA (2009) Statistical methods in diagnostic medicine, John Wiley & Sons, (Vol. 569). https://doi.org/10.1002/9780470906514
92. Zhu W, Lou Q, Vang YS, Xie X (2016) Deep multi-instance networks with sparse label assignment for whole mammogram classification. https://doi.org/10.1101/095794
93. Zhu C, He Y, Savvides M (2019) Feature selective anchor-free module for single-shot object detection. 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2019.00093
94. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, He Q (2020) A comprehensive survey on transfer learning. Proc IEEE 109(1):43–76

Springer