Computers in Biology and Medicine 153 (2023) 106554

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Computer-aided breast cancer detection and classification in mammography: A comprehensive review

Kosmia Loizidou *, Rafaella Elia, Costas Pitris

KIOS Research and Innovation Center of Excellence, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, Cyprus

ARTICLE INFO

Keywords:
Mammography
Computer-aided detection
Breast cancer
Machine learning
Review article

ABSTRACT

Cancer is the second cause of mortality worldwide and it has been identified as a perilous disease. Breast cancer accounts for ~20% of all new cancer cases worldwide, making it a major cause of morbidity and mortality. Mammography is an effective screening tool for the early detection and management of breast cancer. However, the identification and interpretation of breast lesions is challenging even for expert radiologists. For that reason, several Computer-Aided Diagnosis (CAD) systems are being developed to assist radiologists to accurately detect and/or classify breast cancer. This review examines the recent literature on the automatic detection and/or classification of breast cancer in mammograms, using both conventional feature-based machine learning and deep learning algorithms. The review begins with a comparison of algorithms developed specifically for the detection and/or classification of two types of breast abnormalities, micro-calcifications and masses, followed by the use of sequential mammograms for improving the performance of the algorithms. The available Food and Drug Administration (FDA) approved CAD systems related to triage and diagnosis of breast cancer in mammograms are subsequently presented. Finally, a description of the open access mammography datasets is provided and the potential opportunities for future work in this field are highlighted. The comprehensive review provided here can serve both as a thorough introduction to the field but also provide indicative directions to guide future applications.

1. Introduction

According to the American Cancer Society, in 2022, there will be an estimated 1,918,030 new cancer cases and 609,360 deaths in the United States alone [1]. Breast cancer accounts for 19% of those cases and 30% of all female cancers. Furthermore, breast cancer incidence rates continue to increase by about 0.5% per year since the mid-2000s. Mammography, the radiological inspection of both breasts, in combination with a physical examination, provide effective population screening for breast cancer. As a result, the mortality of the disease has dropped by 42% since 1989 [2]. Breast cancer arises from the abnormal proliferation of cells in the breast tissue, forming various types of lesions. These include asymmetries between the left and right breasts, distortion of the normal architecture of the tissue, appearance of Micro-Calcifications (MCs) and masses of various sizes and shapes [3]. Benign abnormalities are most often harmless and do not require biopsy. However, they are followed mammographicaly since, under unknown circumstances, they could spread to the surrounding tissue or apply detrimental pressure to nearby vital structures such as blood vessels or nerves. Malignant abnormalities are unstable and associated with breast cancer [4]. When a suspicious abnormality is identified in a mammogram, its severity is always confirmed by biopsy.

The key screening tool for breast cancer is mammography, which uses low-energy X-rays to detect any abnormalities of the breast [5]. Two images of each breast are taken, forming two different projections: the Cranio-Caudal (CC) and the Medio-Lateral Oblique (MLO). In the CC mammogram, the image is taken from above, whereas in the MLO, the projection is from the side at an angle so that the pectoral muscle is visible (Fig. 1) [6]. The ratio of non-dense tissue to dense tissue can be measured and a level of breast density can be assigned to each mammogram. Levels of density are described using the Breast Imaging Reporting and Data System (BI-RADS) as: (a)-almost entirely fatty, (b)-scattered areas of fibroglandular density, (c)-heterogeneously dense and (d)-extremely dense (Fig. 2). The higher the density, the more challenging it is to assess the images, due to the similar appearance of dense normal and abnormal tissues. According to the American Cancer Society, the sensitivity of mammography is reduced by ~30% [7] resulting in an increased risk of breast cancer [3].

Expert radiologists review the acquired mammograms to determine whether there are any signs of malignancy, followed by appropriate

* Corresponding author.
E-mail addresses: cloizi01@ucy.ac.cy (K. Loizidou), relia001@ucy.ac.cy (R. Elia), cpitris@ucy.ac.cy (C. Pitris).

https://doi.org/10.1016/j.compbiomed.2023.106554
Received 14 November 2022; Received in revised form 13 December 2022; Accepted 11 January 2023
Available online 13 January 2023

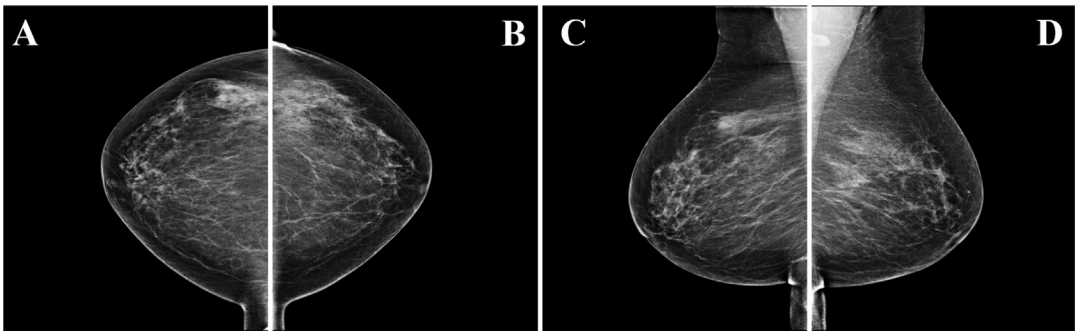

Fig. 1. Example of the available views from a mammographic session: **(A)** Right CC view. **(B)** Left CC view. **(C)** Right MLO view. **(D)** Left MLO view.

Fig. 2. Example of the four levels of mammographic density as defined by the BI-RADS. **(A)** a-Almost entirely fatty. **(B)** b-Scattered areas of fibroglandular density. **(C)** c-Heterogeneously dense. **(D)** d-Extremely dense.

disease management in the case of positive findings. MCs and masses are common findings in abnormal mammograms. MCs are microscopic deposits of calcium that often appear in the breast. They appear as bright spots in the mammogram due to the higher X-ray attenuation coefficient of calcium compared to normal tissue [8]. Most MCs are benign and do not require any intervention (Fig. 3A). However, Micro-Calcification Clusters (MCCs) are considered precursors to cancer (Fig. 3C). Benign MCs are usually larger, rounder, fewer in number and have homogeneous sizes and shapes. MCs, suspicious for early cancer, are clustered, small, with irregular shapes and sizes and branching in orientation. Another worrisome finding in a mammogram is a mass. A breast mass is associated with a localized swelling, protuberance, or lump inside the breast. Masses typically appear in a mammogram as relatively dense regions. They can be radiologically classified as benign or suspicious, depending on key parameters such as size, perimeter, density, gradient, texture, etc. [9,10]. While a benign mass is round, smooth and has a well-defined boundary, a suspicious mass has a spiculated, rough and blurry boundary (Fig. 4C) [8]. Radiologists assign one of seven assessment categories to each mammography study: 0-needs additional imaging evaluation and/or prior mammograms for comparison, 1-negative, 2-benign, 3-probably benign, 4-suspicious for malignancy (which can be further divided into 4A-low, 4B-moderate and 4C-high suspicion for malignancy), 5-highly suggestive of malignancy and 6-known biopsy-proven malignancy [11]. When a suspicious abnormality is identified, its severity and malignancy is confirmed by biopsy [12].

Current protocols require evaluation of the mammogram by two radiologists (and a third, if consensus is not reached), which is an indication of the challenges faced, even by experts, when attempting to identify probable abnormalities in a mammogram. Normal breast perturbations or benign lesions can be misconstrued as breast cancer since they can mimic malignant abnormalities [15]. Classification of masses is one of the most challenging tasks, not only because of the wide variation in their size and shape, but, also, because of their low contrast. Moreover, masses are usually surrounded and/or enclosed by other structures, such as muscle, blood vessels and normal tissue [10]. In some cases, cancer might be present inside a mammogram but missed by the radiologists due to subtle features that are difficult to perceive, or, due to the high intensity of the breast, which reduces the sensitivity of mammography [16]. The false positive rate of screening mammograms, performed in the United States in 2013, was ~61% [17].

To address the challenge of the radiological assessment of mammograms, Computer-Aided Diagnosis (CAD) systems are being developed with automatic or semi-automatic tools to aid the radiologists with the detection and classification of breast lesions [9]. Their main objective is to identify subtle abnormalities that might otherwise be missed [18]. However, this is a very complex task since the lesions tend to have: (a) a small size, which varies from 0.1 to 1 mm, (b) different shapes and distributions, and (c) low contrast with respect to normal breast tissue. Template matching and segmentation approaches are often ineffective, especially in cases where suspicious areas are obscured by dense tissue or thicker than usual skin [19]. The classification of breast abnormalities as benign or malignant is also very challenging. It often leads to a significant number of False Positives (FPs) and limits the clinical applicability of CAD systems [20].

This review examines the recent literature on the automated identification and/or classification of mammographic lesions, using both conventional feature-based Machine Learning (ML) as well as Deep Learning (DL) techniques. The first part is devoted to the detection and classification of MCs and masses, based only on the most recent mammogram. Subsequently, detection and classification of breast lesions using mammograms from two sequential screenings is covered. Finally, the available Food and Drug Association (FDA) approved CAD systems for triage and diagnosis are presented, followed by a brief description of the open-access mammography datasets.

1.1. Inclusion/Exclusion criteria and review structure

The bibliographic literature was thoroughly searched to identify relevant studies. The search was performed in Google Scholar, using the following keywords: “breast cancer diagnosis”, “breast cancer detection”, “breast cancer classification”, “breast mass diagnosis”, “breast mass detection”, “breast mass classification”, “breast

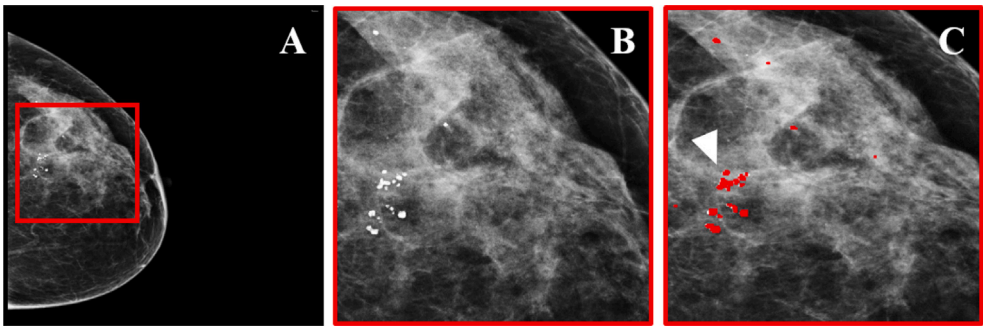

Fig. 3. (A) Mammogram of a woman with Micro-Calcifications (MCs) and a Micro-Calcification Cluster (MCC). **(B)** Zoomed region marked by the red square in **A**, showing MCs and a MCC. **(C)** The region in **B** with precise marking of MC and MCC locations, as annotated by two expert radiologists. The arrow in **C** points to the suspicious MCC.
Source: Re-print with permission from [13].

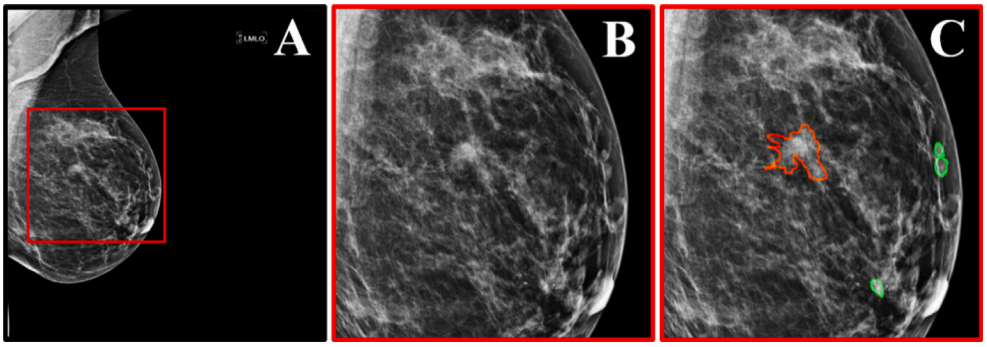

Fig. 4. (A) Mammogram of a woman with benign and suspicious masses **(B)** Zoomed region marked by the red square in **A**, showing masses. **(C)** The region in **B** with precise marking of mass locations (green for benign, red for suspicious), as annotated by two expert radiologists.
Source: Re-print with permission from [14].

micro-calcification diagnosis", "breast micro-calcification detection", "breast micro-calcification classification", "mammography", "mammogram analysis", "CAD systems for breast cancer", "deep learning in mammography", "Convolutional Neural Networks (CNNs) in mammography". The search was limited to papers published in English, between 2000 and 2022, but also included the first paper published in each category irrespective of the year. The most relevant and highly cited articles were manually selected. Articles were excluded if the analysis included additional data from screening methods other than mammography, such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and Positron Emission Tomography (PET), or if any information critical to the techniques' performance was missing from the article. Furthermore, articles were excluded based on the type of abnormality considered (only masses and MCs were considered in this review) and the dataset size (preliminary studies with small number of subjects were excluded), resulting in a total of 90 papers.

Various review articles related to cancer diagnosis and medical image analysis, in general, as well as some reviews devoted to breast cancer diagnosis using only mammography, already exist (Cheng et al. 2003 [19]; Rangayyan et al. 2007 [8]; Doi, 2007 [21]; Oliver et al. 2010 [9]; Ganesan et al. 2016 [5]; Litjens et al. 2017 [22]; Hamidinekoo et al. 2018 [23]; Gardezi et al. 2019 [24]; Zou et al. 2019 [25]; Ramadan, 2020 [26]; Abdelrahman et al. 2021 [27]; Painuli et al. 2022 [28]). However, most of these reviews are general, they do not provide some of the information which is necessary for a meaningful assessment, and consider dissimilar studies, whose results cannot be directly compared. For example, in some cases, BC detection is compared to BC classification, despite the two being different tasks. In addition, various types of breast abnormalities are considered together. In this review article, the various methodologies were subdivided into specific categories, based not only on their machine learning approach but also the various other characteristics mentioned before, in order to provide insightful comparisons and all the available details in a more comprehensive review.

The publications selected were divided into various groups based on their target and approach. First, they were split into two major categories, according to the targeted breast abnormality: (a) MCs or (b) masses. Subsequently, the articles were further divided into two sub-categories based on their main objective: (a) detection of the abnormalities or (b) classification of the abnormalities. Finally, each sub-category, was split according the ML approach: (a) not employing ML, (b) feature-based ML or (c) DL. In feature-based ML, the studies were further grouped based on the classifier used. For each category, the most recent and most successful approaches are discussed. Grouping the studies based on the abnormality (MCs or masses), main objective (detection or classification), and ML approach (feature-based or DL), along with several other parameters (dataset, classifier, validation method), provides the most complete overview of the literature but also identifies the key advantages and disadvantages of each approach. Given the large number of possible approaches, direct comparison of all the studies is quite complex and comprehensive conclusions are challenging to draw. There is also no established or straightforward way to identify the "most successful" algorithm. Other factors complicating the direct comparison of the various studies are the differences in the validation method and the reporting of the results, which will be further discussed in the next session.

2. Detection and classification of breast cancer in mammograms

2.1. Overview

Automated detection and classification of breast cancer in mammograms most often follows three basic steps: (1) pre-processing, using various filters and enhancement techniques, (2) detection of the abnormalities, including accurate segmentation and (3) classification of the

Table 1
An overview of the state-of-the-art pre-processing techniques for mammogram analysis.

Methods	Overview	Reference
Conventional	Unsharp masking	Chan et al. (1987) [29]
	Optimal adaptive neighborhood processing	Dhawan et al. (1986) [30]
	Sobel filter	Kim et al. (1997) [31] Xie et al. (2016) [12] Basile et al. (2019) [32]
	Histogram Equalization	Karssemeijer (1993) [33] Rouhi et al. (2015) [34] Arora et al. (2020) [35]
	Contrast Limited Adaptive Histogram Equalization	Agrawal et al. (2018) [36] Al-antari et al. (2018) [37]
	Morphological operations	Charan et al. (2018) [38]
	Thresholding	Dheeba et al. 2014 [39] Al-masni et al. (2018) [40] Cai et al. (2019) [41]
Feature-based	Wavelet transform	Laine et al. (1994) [42] Chowdhury and Khatun (2012) [43]
	ϕ-transform	Laine et al. (1994) [42]
	Hexagonal wavelet transform	Laine et al. (1994) [42]
	Fuzzy set theory	Cheng, Lui and Freimanis (1998) [44]
	Round high-emphasis technique	Kegelmeyer, Hernandez and Logan (1993) [45]
	Fractal modeling	Li, Liu and Lo (1997) [46]
Deep Learning	Convolutional Neural Networks	Dong et al. (2015) [47] Kim et al. (2016) [48] Umehara et al. (2017) [49] Jiang et al. (2020) [50]

detected region as normal, benign, suspicious or malignant, depending on the goal of the study.

Pre-processing is critical for the success of CAD algorithms and, conversely, if performed incorrectly it can lead to significantly diminished performance. The goal of pre-processing of a mammogram is to effectively improve the contrast and eliminate the unnecessary details inside the breast, including the background and pectoral muscle in the MLO views, without removing areas that might be important for the diagnosis. Table 1 lists the state-of-the-art pre-processing techniques applied to mammograms. The most common conventional pre-processing techniques are resizing, enhancement or formatting of the mammogram. Normalization, which adjusts the range of pixel intensity values, is also very widely used. As mentioned before, breast abnormalities usually have higher intensity compared to their surrounding tissue, thus, intensity-based techniques, such as Sobel filtering, histogram equalization, thresholding etc., can be applied either individually or in various combinations. These approaches effectively help to distinguish the high intensity abnormal regions from the background. Feature-based techniques, on the other hand, process and enhance the mammograms based on the characteristics of each region. Specifically, they can be used to increase the contrast of suspicious areas, or remove background structures and noise, based on the textural features of the Region of Interest (ROI). Additionally, mammograms can be re-constructed using wavelets, or enhanced by calculating the first derivative or local statistics. Finally, DL can also be exploited for the pre-processing of a mammogram instead of conventional filtering techniques, to enhance the ROIs.

For effective detection, the abnormalities must be accurately segmented. Segmentation is the identification and separation of the abnormalities (masses or MCs) from the background of a mammogram. The accuracy of the segmentation also affects the results of the classification, since the features used for the identification of malignant and benign abnormalities, are usually extracted from segmented region. Several segmentation approaches have been proposed depending on the type of the abnormality [26] and are divided into four major categories: (a) threshold-based, (b) region-based, (c) edge-based and (d) clustering segmentation [24]. After segmentation of the ROIs, many algorithms use ML to improve their results by discarding areas false identified as lesions. Ideally, the success of the segmentation should be compared with the findings of expert radiologists. However, very few datasets have detailed annotations (i.e. outlines of each lesions) available. Most, just indicate whether or not there is a lesion in the mammogram.

Various features have been shown to be significant in the segmentation and classification of breast abnormalities. Table 2, illustrates the most widely used features by category. These features have been applied to both MCs and masses although the two are anatomically and radiologically different. Shape is a particularly important radiological feature since benign masses are often rounder with smooth and well-defined boundaries, compared to malignant masses, which have spiculated boundaries. Breast abnormalities also exhibit higher intensity, compared to the background and other regions of the mammograms. Thus, intensity features can also be extracted to reflect such differences. Furthermore, the texture of a breast lesion can provide diagnostically useful information which can be quantified using the First-Order Statistics [55] and Gray Level Co-occurrence Matrix features [66]. Given the very large number of features that can be extracted from each region, feature selection procedures are employed to choose the best combination that will yield the most accurate classification. With the application of feature selection, the irrelevant and redundant features are removed and only the most important one are used by the classifiers. Several feature selection techniques can be applied, including feature importance [67], Minimum Redundancy Maximum Relevance (MRMR) [68], hypothesis t-test [69], SelectKBest [67], Multivariate ANalysis Of VAriance (MANOVA) [70] and more.

The most relevant features of each ROI are introduced into a classifier for the classification of the ROI as normal, benign, suspicious or malignant, depending on the objective of each study. Various classifiers have been exploited for the identification of breast abnormalities, including Support Vector Machine (SVM), Multi-Layer Perceptron

Table 2
An overview of the state-of-the-art features extracted for the diagnosis of breast abnormalities in mammograms.

Category	Features	Reference
Shape	area, compactness, density, perimeter, extent, eccentricity, orientation, solidity	Woods et al. (1993) [51] Davies and Dance (1990) [52] Yu and Guan (2000) [53] Singh and Kaur (2018) [54]
First-Order Statistics	entropy, kurtosis, skewness, smoothness, standard deviation, variance	Kumar and Gupta (2012) [55] Bekker et al. (2016) [56] Fanizzi et al. (2020) [57]
Gray Level Co-Occurrence Matrix	contrast, correlation, energy, homogeneity	Woods et al. (1993) [51] Zyout and Abdel-Qader (2011) [58] Jian, Sun and Luo (2012) [59] Milosevic et al. (2014) [60]
Gabor filter	features extracted from Gabor filter bank processed image	Rogova, Stomper and Ke (1999) [61] Bhangale, Desai and Sharma (2000) [62]
Wavelet	features created using the wavelet transform	Woods et al. (1993) [51] Strickland, Hahn and Baig (1996) [63] Mencattini et al. (2008) [64] Ghasemzadeh, Azad and Esmaeili (2019) [65]

Table 3
An overview of the state-of-the-art classifiers used for the classification of breast abnormalities in mammograms.

Method	Classifier	Reference
Feature-based	SVM	Papadopoulos et al. (2005) [71] Jian, Sun and Luo (2012) [72] Suhail et al. (2018) [73] Berbar (2018) [74]
	MLP	Tahmasbi et al. (2011) [75] Danala et al. (2018) [76]
	RF	Fanizzi et al. (2020) [57]
	ANN	Bekker et al. (2016) [56] Hu et al. (2017) [77]
Deep Learning	GoogleNet	Sert et al. (2017) [78]
	ImageNet	Sert et al. (2017)) [78]
	CNN	Chen et al. (2019) [79] Cai et al. (2020) [80] Chakravarthy and Rajaguru (2021) [81]
	InceptionResNet-V2	Al-antari et al. (2020) [82]

(MLP), Random Forest (RF), Artificial Neural Network (ANN) and others. Feature-based ML has several advantages, including the ability to unambiguously recognize which features contribute positively to the classification and can, thus, be used as biomarkers. In addition, unlike DL there is no need for large volumes of data, which is, sometimes, difficult to collect for medical applications. However, all the data undergo the same processing, making the algorithm less robust. In studies where human intervention is required, human bias can be introduced, although, in some studies, it was unavoidable.

Since 2015, the introduction of DL shifted the interest from extracted features to automatic feature identification, selection and classification of breast abnormalities. Deep Learning techniques have a number of advantages, compared to conventional feature-based classifiers and they have been widely applied in biomedical sciences for the detection of breast cancer [25]. In DL, the human intervention is eliminated, since feature extraction and selection are incorporated as an internal part of the network. Hence, these approaches are more robust and can handle more complex data. A key advance in DL, is the introduction of Convolutional Neural Networks (CNNs), which have been very successful in various classification tasks [83]. Some well-known CNN architectures for tissue segmentation and classification include AlexNet, GoogleNet, VGGNet, ResNet, DenseNet and EfficientNet. The disadvantages of DL include the large amount of data required for proper training and the uncertainty in the diagnostic significance of the image features introduced by the many deep layers. This uncertainty, reduces the explainability of the results of the deep neural networks, which makes the evaluation of the importance of the imaging information and the assessment of the role of each feature in the classification performance quite difficult. Table 3 summarizes the most popular classifiers for the classification of breast abnormalities.

For an accurate evaluation of the classifier's performance, the datasets are usually divided into training and test sets. The training set is used to fit and create the model, while the test set, is used as an input to evaluate the model's performance. The trained models are usually validated before testing. There are various established methodologies for validation, however the majority of the studies in the literature use k-fold cross-validation, with the number of k varying from 1 to 30. K-fold cross-validation is the random extraction of a number of samples from the training set and their use to evaluate the performance of the model repeated several times for different sample sets. However, there are significant differences between studies in their cross-validation approach, which can adversely affect their reported results. Some perform cross-validation by dividing the sets per patient, others per image, or some per ROI. However, all the ROIs or images of the same patient should be included either in the training or in the validation set, but not in both, to avoid bias. Introducing information from the same patient in both training and test sets, as in per image or per ROI cross-validation, artificially increases the algorithms' performance. Thus, validation should be performed per patient to provide results that will be more representative of real world applications. Another

Table 4
Comparison of the state-of-the-art algorithms for the detection of MCs in mammograms, without the use of machine learning.

Reference	Database	Type of images	Dataset	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
Chan et al. (1987) [29]	Restricted	Digitized	6 images	–	–	–	80/-	–	1 FPi
Mehdi et al. (2017) [84]	Open Access (MIAS)	Digitized	322 images	–	–	–	97.1/-	0.92	0.5 FPi
Basile et al. (2019) [32]	Open Access (BCDR)	Digital	364 images	–	–	–	91.2/-	–	2.9 FPi
Touil et al. (2020) [85]	Open Access (INbreast)	Digital	40 images	–	–	–	88.6/-	–	–

factor complicating the comparison of the various approaches described in the literature, is the reporting of the results. The most commonly used evaluation measurements include sensitivity, specificity, accuracy and Area under the Receiver Operating Characteristic Curve (AUC). To completely describe an algorithm, all the performance evaluation measures must be provided in the manuscript. Providing only some information, i.e. the accuracy and AUC, can be misleading since a combination of high specificity and low sensitivity, can still provide high accuracy. However, low sensitivity is associated with a high number of false negative detections, which is unacceptable in clinical applications.

2.2. Detection of micro-calcifications (MCs)

Various algorithms have been proposed specifically for the detection of MCs [19,86,87]. In this review they are divided into three broad categories based on the detection approach: (a) without ML, (b) with feature-based ML, (c) with DL.

2.2.1. Methods not employing Machine Learning (ML)

MC detection is indeed possible without ML, as demonstrated by one of the first papers related to computer-aided MC detection, published in 1987 by Chan et al. [29]. The algorithm proposed was based on a difference-image technique, in which signal-suppressed images were subtracted from signal-enhanced images, to successfully eliminate the background. Monte Carlo methods were then applied, to generate simulated MCs that were superimposed on normal mammographic backgrounds. Detection was performed using 6 digitized mammograms. Overall, the sensitivity reached 80%, with 1 False Positive detection per Image (FPi) [29]. Other, more recent, publications also describe MC detection without the use of ML [32,84,85]. The highest sensitivity (97.1%) was achieved by Medhi et al. 2017, using the MIAS open access database, which contains 322 digitized mammograms [84]. Frequency analysis, along with spatial, automatic, non-linear, stretching and Shannon Entropy-based wavelet coefficient thresholding, were used for the identification of the MCs. The state-of-the-art in techniques for the detection of MCs without the use of ML are provided in Table 4.

2.2.2. Feature-based Machine Learning (ML)

After the pre-processing and accurate segmentation of an ROI in a mammogram, several features can be extracted in order to classify the ROI as normal, benign, suspicious or malignant MC. As explained before, these features are selected according to their capacity to correctly identify the MCs and include various characteristics (shape, intensity, texture, etc.) of the ROI. SVMs are the most popular classifiers for the diagnosis of MCs using feature-based ML. Several studies employed SVMs to separate ROIs as true MCs and avoid the erroneous detections leading to FPs [88–92]. These studies employed different pre-processing and segmentation approaches and they were trained and tested using various different datasets (Fig. 5), including open access [89,91], restricted [88] and in some cases both [90,92]. Karale et al. 2020, achieved the highest sensitivity (100%), using mammograms from three different datasets, with 2.34, 1.57, and 0.61 FPi for each dataset [92]. A total of 717 mammograms, both digitized and digital, were pre-processed with a multi-scale 2D non-linear energy operator to enhance the contrast between the MCs and the background. Various features were extracted from the ROIs including area, compactness, shape, intensity and others. Synthetic Minority Over-sampling TEchnique (SMOTE) and majority class under-sampling based on data distribution were applied to augment the dataset and improve the classification performance. Furthermore, Principal Component Analysis (PCA) was performed to eliminate FPs by removing the vascular MCs which were not dimmed of clinical significance. It is important to mention that, for the validation, the data were split into 5-folds, such that all the mammograms of the same patient were either in the training set or in the test set, but not in both.

Milocevic et al. 2014, applied a segmentation technique to 322 digitized mammograms, obtained from an open access database (mini-MIAS), based on the discrete wavelet transform and Sobel filtering [60]. In total, 20 textural features were extracted and the MCs were identified with both SVM and k-Nearest Neighbor (k-NN) using 4-fold cross-validation. The results of the k-NN were better, compared to the SVM (65.2% vs. 56.5% sensitivity). The most promising classifiers, which have gained significant attention for the detection of MCs, are ANNs. Melekoodappattu and Subbian, 2019, achieved 99% accuracy using a hybridized Extreme Learning Machine (ELM), for automatic MC detection in 184 digitized images from a restricted dataset [93]. The unnecessary details of the mammograms were removed at the pre-processing stages and multi-scale features, such as Speeded-Up Robust Features (SURF) and Gabor filter features, were extracted using a feature generation model. However, only the accuracy is available for this study and the cross-validation method is not known either. The state-of-the-art techniques for the detection of MCs using feature-based ML are provided in Table 5.

2.2.3. Deep Learning (DL)

Since 2015, various DL approaches were pursued for the automated detection of MCs, using either deep networks or CNNs [25]. The group of Mordang et al. 2016, was one of the first to implement CNNs for the automatic detection of MCs in multi-vendor mammography. A restricted dataset was created specifically for this study, containing 1606 digital images, and data augmentation was employed to increase the size of the dataset. Augmentation was performed by flipping (horizontally and vertically) and rotating (90, 180, and 270 degrees) the positive images. A CNN was trained on a small subset of the dataset and, then, applied to the complete dataset to remove the samples that were easier to classify. Subsequently, a second CNN was trained on a larger dataset, containing samples that were harder to categorize. Overall, the sensitivity reached 70%, with 80%–20% validation per image [94].

In recent years, several groups have applied DL to MC detection with varies success. However, the results of DL can markedly vary when the algorithms are applied to different datasets with variations in the cross-validation approach. Although, as mentioned before, it is hard to directly compare most studies, some DL approaches are worth examining further. Various algorithms were developed using the open access database of the MIAS for MC detection [95,96]. Zhang and Wang, 2019,

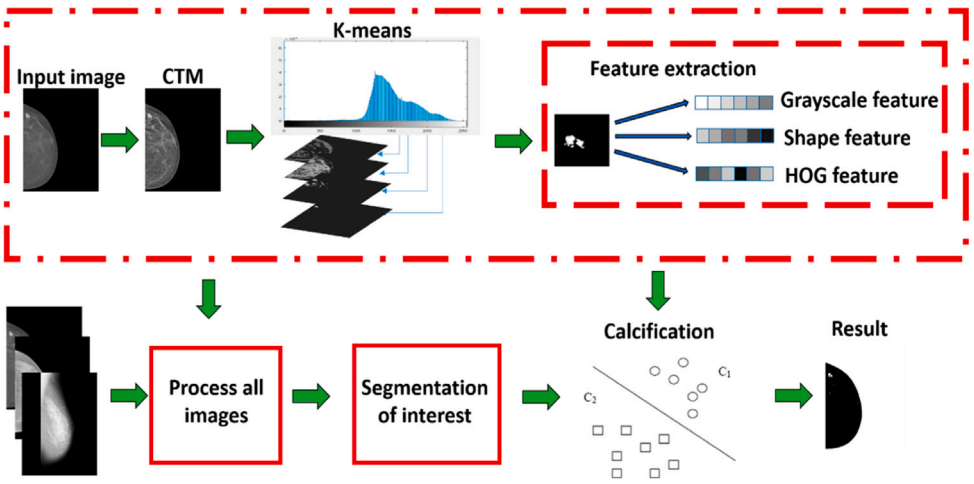

Fig. 5. Flowchart of the CAD algorithm for the detection of breast cancer.
Source: Re-print with permission from [91].

Table 5
Comparison of the state-of-the-art algorithms for the detection of MCs in mammograms, using feature-based machine learning.

Reference	Database	Type of images	Dataset	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
El-Naqa et al. (2002) [88]	Restricted	Digitized	76 images	SVM	10-fold CV (per image)	–	94/-	–	1 FPi
Milosevic et al. 2014[60]	Open Access (MIAS)	Digitized	48 images	k-NN	4-fold CV (per image)	66.7	65.2/68	–	–
Melekoodappattu and Subbian (2019) [93]	Restricted	Digitized	184 images	ANN	leave-one-out CV (per ???[a])	99	–	–	–
Li et al. [89]	Open Access (INbreast)	Digital	410 images	SVM	150 training ROIs (per ROI)	–	80.6/99.99	–	–
Karale et al. (2019) [90]	Both (DDSM INbreast PGIMER-IITKGP)	Both	197 images 410 images 110 images	SVM	5-fold CV (per patient)	–	100/-	–	2.6 FPi 1.8 FPi 0.7 FPi
Cai et al. (2019) [41]	Open Access (DDSM)	Digitized	400 images	Adaptive SVM	300-100 (per image)	94	85.4/93.8	–	–
Karale et al. (2020) [92]	Both (DDSM INbreast PGIMER-IITKGP)	Both	197 images 410 images 110 images	SVM	stratified 5-fold CV (per patient)	–	100/-	–	2.3 FPi 1.6 FPi 0.6 FPi

[a]No information was provided on how the data were divided in training and test sets (per ROI/per image/per patient).

achieved high accuracy (97.2%) using a combination of a deep, fine-grained, cascade-enhanced, network and a multi-scale, feature fusion, algorithm [95]. A total of 1288 digitized mammograms were included, with 1030 used for training and 258 for testing. Savelli et al. 2019, and Hakim et al. 2021, exploited the INbreast open access database, which contains full-field digital mammograms, to develop an algorithm for the detection of MCs using CNNs [97,98]. Hakim et al. 2021, created an automated approach, including pre-processing to enhance the image quality, followed by training of a segmentation network and a trained network [98]. On the other hand, Savelli et al. 2019, employed an ensemble of CNNs to eliminate the FPs and improve the performance of the algorithm [97]. Data augmentation was applied in both cases. Hakim et al. 2021, achieved higher sensitivity compared to Savelli et al. 2019, (88.1% vs. 83.5%) [97]. However, only the latter, correctly divided the data into training and test set by assigning the patches belonging to the same image to the same set. Hossain et al. 2019, used an open access dataset (DDSM) for the detection of MCs using DL [99]. That database contains 2620 scanned film mammograms, which have been digitized and are available online. A modified U-net segmentation network reached 98.2% accuracy, using 60% of the dataset for training, 20% for validation and the remaining 20% for testing but per image rather than per patient.

Various groups have opted to collect new custom datasets, customized to the specific need of their study [100–102]. In those cases, comparison between the studies is not straightforward since the datasets are not publicly available and can be significantly different. Valvano et al. 2019, achieved 99.2% accuracy using a custom dataset of 283 digitized and digital mammograms [102]. Data augmentation was applied and the detection was performed with two consecutive CNN blocks: the detector and the segmentator. The dataset was divided into training and test set using the images but not the patients. Marasinou et al. 2021, combined a custom dataset (141 digital mammograms) with the INbreast open access database (410 digital mammograms) [103]. The goal was to create a new algorithm for the segmentation of breast MCs using the Hessian difference of Gaussian regression. Data augmentation was applied and the data were correctly divided into training and test sets using the cases and not the images or ROIs. Unfortunately, the performance of this method was low (74.4% sensitivity), compared to other studies. The overall state-of-the-art for the detection of MCs in mammograms using DL is provided in Table 6.

2.3. Classification of Micro-calcifications (MCs)

Classification of breast MCs as benign or malignant is a challenging task even for trained radiologists [86,87]. Over the years several

Table 6
Comparison of the state-of-the-art algorithms for the detection of MCs in mammograms, using deep learning.

Reference	Database	Type of images	Dataset	Data augmentation	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
Mordang et al. (2016) [94]	Restricted	Digital	490 images 1044 images 72 images	YES	CNN	80%–20% (per image)	–	70.3/71 70.1/71 67.3/71	–	–
Wang, Nishikawa and Yang (2017) [100]	Restricted	Both	521 images 188 images	YES	CNN	167-67-158 (per patient)	–	90/-	–	0.7 FPi
Wang and Yang (2018) [101]	Restricted	Both	521 images 188 images	YES	DNN	167-67-158 (per patient)	–	80/-	–	1 FP/cm^2
Zhang and Wang (2019) [95]	Open Access (MIAS)	Digitized	1288 images	NO	DNN	1030-258 (per image)	97.2	–	–	–
Valvano et al. (2019) [102]	Restricted	Both	283 images	YES	CNN	231-25-27 (per image)	98.2	–	–	–
Hossain (2019) [99]	Open Access (DDSM)	Digitized	2620 images	YES	CNN	60%–20%–20% (per ROI)	98.2	98.4/-	–	–
Savelli et al. (2019) [97]	Open Access (INbreast)	Digital	410 images	YES	CNN	2-fold CV (per image)	–	83.5/-	–	–
Ramadan (2020) [96]	Open Access (MIAS)	Digitized	322 images	YES	CNN	222-75-25 (per image)	92.1	91.4/96.8	0.95	–
Hakim et al. (2021) [98]	Open Access (INbreast)	Digital	354 images	YES	CNN	3 sets of 20 images (per image)	90.3	88.1/-	–	–
Marasinou et al. (2021) [103]	Both (INbreast)	Digital	410 images 141 images	YES	CNN	51-17-18 (per patient)	–	74.4/-	–	0.4 FP/cm^2

techniques were proposed for automating their classification. In order to provide a better understanding of the field this review focuses on classification methodologies which can be broadly divided in feature-based ML and DL, with the same caveats as before when it comes to direct comparisons between them.

2.3.1. Feature-based Machine Learning (ML)

One of the first studies attempting to classify MCs (Shen et al. 1994), employed shape features to measure the roughness of MC contours and incorporated them into a k-NN classifier to distinguish between benign and malignant. A restricted dataset was created, containing 18 cases of digitized film mammograms and used leave-one-out validation to achieve 100% accuracy [104]. However, no details were disclosed regarding the validation (per patient vs. image vs. ROI).

Over the years, SVMs [71–73], k-NNs [105,106] and Logistic Regression (LR) [56,107] have been widely popular for the classification of benign vs. malignant MCs. The highest classification performance using SVMs was reported by Jian et al. 2012 [72] (100% accuracy vs. 96% accuracy by Suhail et al. 2018 [73] and 0.81 AUC by Papadopoulos et al. 2005 [71]). For the development of the algorithm, both an open access (MIAS) and a restricted dataset were employed, containing only digitized film mammograms. The MCs were segmented by combining space and frequency domain techniques. Texture features, based on the wavelet (Haar, DB4, DT-CWT) transform were extracted and, after PCA, a new feature vector was created for the classification. However, leave-one-out cross-validation was performed per image and not per patient. In contrast, the group of Suhail et al. 2018, achieved 96% accuracy, using digitized mammograms from the DDSM database, but with 10-fold per patient cross-validation.

Bekker et al. 2016, [56], along with Chakravarthy and Rajaguru, 2019 [81], investigated LR as a tool for the efficient classification of MCs. Both drew their images from open access databases containing digitized mammograms. Chakravarthy and Rajaguru, 2019, utilized five different wavelet families, the firefly algorithm, ELM and least-square-based non-linear regression classifiers, and managed to achieve the highest performance (98.4% vs. 69.5% accuracy) [81]. Another popular classifier for MCs is k-NN. Chen et al. 2015, used two open access databases (MIAS and DDSM) along with a restricted dataset, containing both digitized and digital mammograms, and reached 95%, 86% and 89% accuracy for each, respectively [105]. These results were significantly better compared to George et al. 2019, who used three open access datasets (MIAS, DDSM and OPTIMAM), with both digitized and digital mammograms (90%, 86.5% and 76.7% accuracy) [106]. Chen et al. 2015, managed to further improve the performance by analyzing the topology and connectivity of individual MCs within a cluster, using multi-scale morphology. However, for the cross-validation, leave-one-out and 10-fold per ROI were used. Other classifiers such as ANN [77], ensembles of classifiers [108] and RF [57], have also been tested for MC classification. Alam et al. 2019, produced the most promising results using stack generalization in a 10-fold cross-validation scheme per image, and a combination of morphological, texture, and distribution features (0.98 AUC) [108]. The results of Hu, Yang and Gao, 2017, using ELM (0.92 AUC) [77] and Fannizzi et al. 2022, using RF (0.92 AUC) [57] also showed promised. A summary of the best results of various studies, along with the dataset and classifier used can be found in Table 7.

2.3.2. Deep Learning (DL)

Since 2016, DL has been widely used for the classification of MCs as benign or malignant, using either deep networks or CNNs [109]. The group of Bekker et al. 2016, was one of the pioneers in the development of DL for the classification of breast MCs [110]. Their approach was based on two view-level decisions, created by two NNs, followed by a single-neuron layer. An open access database (DDSM) was utilized, containing 1410 clusters from digitized images, and 10-fold cross-validation was applied but no information was provided on how the data were divided in training and test sets. The results were satisfactory with 78.7% accuracy. Sert et al. 2017, [78] and Chakravarthy and Rajaguru, 2021, [81] also developed CNNs for the classification of MCs as benign or malignant. The first employed only the DDSM database, with 1755 digitized mammograms, whereas the second utilized the CBIS-DDSM, MIAS and INbreast, with 1071 digitized and digital mammograms. Both groups used data augmentation to increase the sample size. Chakravarthy and Rajaguru, 2021, reached the highest classification performance with 97.2%, 98.1%, 98.3% accuracy for each dataset (vs. 94.3% by Sert et al. 2017, [78]) when applying 5-fold cross-validation per image. They used CNNs (GoogleNet, ImageNet) after various pre-processing steps (contrast scaling, dilation and cropping) and decision fusion using ensemble of networks (Fig. 6).

Several groups worked with their own datasets, collected from various hospital around the world, rather than using open access data [41,79,111]. A comparative assessment of those studies is not possible,

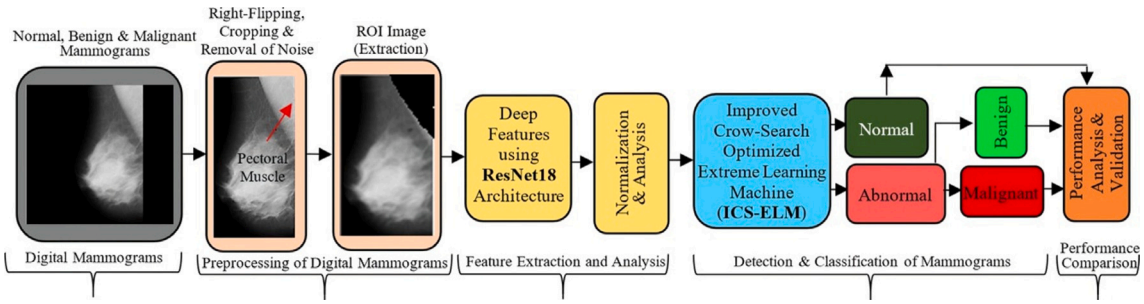

Fig. 6. Proposed workflow for detection and classification of breast MCs using deep learning .
Source: Re-print with permission from [81].

Table 7
Comparison of the state-of-the-art algorithms for the classification of MCs in mammograms, using feature-based machine learning.

Reference	Database	Type of images	Dataset	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
Shen, Rangayyan and Desautels (1994) [104]	Restricted	Digitized	18 patients	k-NN	leave-one-out (per ROI)	100	–	–	–
Papadopoulos et al. (2005) [71]	Open Access (Nijmegen MIAS)	Digitized	40 images 20 images	SVM	2-fold CV (per ???[a])	–	–	0.79 0.81	–
Jian, Sun and Luo (2012) [72]	Both (MIAS)	Digitized	50 images	SVM	leave-one-out CV (per image)	100	–	–	–
Chen at al. (2015) [105]	Both (MIAS DDSM)	Both	20 ROIs 300 ROIs 25 ROIs	k-NN	leave-one-out CV (per ROI)	95 86 96	–	0.96 0.96 0.90	–
Bekker et al. (2016) [56]	Open Access (DDSM)	Digitized	1410 ROIs	EM-LR	10-fold CV (per ???)	69.5	68.1/69.7	0.75	–
Hu et al. (2017) [77]	Open Access (Nijmegen MIAS DDSM))	Both	103 ROIs 26 ROIs 150 ROIs	Extreme Learning Machine	leave-one-out CV (per ???)	–	92/93 100/92	0.99 0.99 0.92	–
Suhail et al. (2018) [73]	Open Access (DDSM)	Digitized	288 ROIs	Scalable LDA	10-fold CV (per patient)	96	–	0.95	–
Alam et al. (2019) [108]	Open Access (OMI-DB DDSM MIAS)	Both	286 patients 280 images 24 patients	Stack Generalization	10-fold CV (per image)	96.7	–	0.98	–
Chakravarthy and Rajaguru (2019) [81]	Open Access (MIAS)	Digitized	80 images	LSNLR	90%–10% (per ???)	98.4	–	–	–
George et al. (2019) [106]	Open Access (MIAS DDSM OPTIMAM)	Both	20 images 289 ROIs 286 ROIs	k-NN	10 rounds of 10-fold CV (per ???)	90 86.5 76.7	–	0.9 0.89 0.78	–
Fanizzi et al. (2020) [57]	Open Access (BCDR)	Digital	260 ROIs	RF	100 rounds of 10-fold CV (per ???)	88.5	89.1/88	0.92	–

[a]No information was provided on how the data were divided in training and test sets (per ROI/per image/per patient).

since the performance of a CNN directly depends on the input image data. The group of Chen et al. 2019, developed a Cascade Adaboost CNN that reached 90.2% accuracy with 80% of the dataset being used in the training and the remaining 20% in the test set, without data augmentation [79]. Their dataset included 4140 digital images and yielded better results compared to similar studies (90.2% accuracy vs. 87.3% by Wang et al. 2016 [111] and 88.6% by Cai et al. 2019 [41]). Furthermore, some groups incorporated both open access and restricted datasets in order to increase their datasets [80,112]. For example Cai et al. 2020, proposed a promising algorithm for the classification of MCs on full field digital mammograms based on DL method based on neutrosophic boosting [41], using a total of 676 digital images. The network achieved 93.7% accuracy (ROI-based validation with 1421 training, 454 validation and 546 testing ROIs). The performance is higher than the study by Li et al. 2021, with 90% and 87% accuracy, but the later performed the validation on a separate group of images [112]. A comparison of the state-of-the-art techniques for the classification of MCs in mammograms using DL can be found in Table 8.

2.4. Detection of breast masses

Various research groups have been developing algorithms specifically for the detection of breast masses in mammograms [8,9,113]. As in the case of MCs, the studies can be divided into three major categories according to their specific ML approach: (a) detection without ML, (b) feature-based ML, and (c) DL.

2.4.1. Methods not employing machine learning

One of the first literature accounts of detection of breast masses without ML was presented by Lai et al. 1989 [114]. They combined criteria used by clinicians, such as shape, brightness, contrast and density uniformity of tumor areas for the detection of circumscribed

Table 8
Comparison of the state-of-the-art algorithms for the classification of MCs in mammograms, using deep learning.

Reference	Database	Type of images	Dataset	Data augmentation	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
Bekker et al. (2016) [110]	Open Access (DDSM)	Digitized	1410 images	NO	Multi-view Neural Network (MV-NN)	10-fold CV (per ???[a])	78.7	78.8/78.7	0.89	–
Wang et al. (2016) [111]	Restricted	Digitized	1204 patients	NO	DNN	10-fold CV (per ???)	87.3	93/82	0.87	–
Sert et al. (2017) [78]	Open Access (DDSM)	Digitized	1755 images	YES	CNN (GoogleNet, ImageNet)	1405-350 (per image)	94.3	94	–	–
Cai et al. (2019) [41]	Restricted	Digital	990 images	YES	CNN	10-fold CV (per ???)	88.6	88.4/86.9	0.94	–
Chen et al. (2019) [79]	Restricted	Digital	2200 images 1940 images	NO	CNN (Cascade AdaBoost)	80%–20% (per ???)	90.2	–	–	–
Cai et al. (2020) [80]	Both (INbreast)	Digital	338 patients 23 patients	YES	CNN	1421-454-546 (per ROI)	93.7	–	0.93	–
Chakravarthy and Rajaguru (2021) [81]	Open Access (CBIS-DDSM MIAS INbreast)	Both	570 images 322 images 179 images	YES	Improved Crow-Search-based Extreme Learning Machine (ICS-ELM)	5-fold CV (per image)	97.2 98.1 98.3	–	–	–
Li et al. (2021) [112]	Both (DDSM)	Both	577 images 3568 images	YES	Fully Connected Depthwise Separable CNN	5-fold CV (per ???)	90 87	99/82 97/83	0.86 0.87	–

[a]No information was provided on how the data were divided in training and test sets (per ROI/per image/per patient).

masses in 24 film mammograms from a restricted dataset. Their method was based on template matching and it could detect suspicious areas in mammograms, independent of their size, orientation and position, with 1.7 FPi.

2.4.2. Feature-based Machine Learning (ML)

Feature-based ML approaches, for the detection of masses heavily rely on SVMs [115–117]. Several studies used the same open access dataset (DDSM) for training and testing. The group of Chu et al. 2015, achieved the best performance (81.4% accuracy), using 474 digitized mammograms [115]. The proposed algorithm followed five basic steps: pre-processing (using morphological enhancement), segmentation (based on a simple linear iterative clustering method), pre-screening of suspicious regions (using rule-based classification), potential lesion contour refinement (based on distance regularized level set evolution), FP reduction (based on feature extraction) and SVM classification. After segmentation, two thirds of the available masses were used in the training and the remaining masses and normal regions were used in the test set.

Another important classifier that is widely used for the detection of masses in mammograms is ANN [39,118]. A very promising study was conducted by Mohanty et al. 2019, using two open access datasets (DDSM and MIAS) with a total of 1814 digitized mammograms [118]. The model first utilized the discrete Tchebichef transform to extract the features from the ROIs, and then, using a combination of PCA and Linear Discriminant Analysis (LDA), the dimension of each feature vector was reduced. Finally, the reduced features were classified using an ELM with the samples being split using stratified 5-fold cross-validation per image, to avoid over-fitting. Compared to Dheeba et al. 2014, who used a restricted dataset with 216 digitized mammograms [39], Mohanty et al. 2019, achieved 100% and 99.5% accuracy for each dataset (vs. 96.7% of the Dheeba et al. 2014).

Various other classifiers, such as LDA [119], k-NN [120], Naive Bayes (NB) [121], AdaBoost [122] and RF [123], have also been employed in mass detection. Tai et al. 2014 [119], Chakraborthy et al. 2018 [121], and Dhahbi et al. 2018 [120], used the same open access dataset (DDSM). The highest classification performance was obtained by Tai et al. 2014, (90.3% sensitivity and 0.98 AUC), using 358 digitized mammograms but hold-out validation per ROI [119]. The group used local and discrete texture features for the effective detection of masses. Despite the fact that the algorithm of Chakraborthy et al. 2018, using an iterative method of high-to-low intensity thresholding controlled by radial region growing, had a lower performance (85% sensitivity and 1.4 FPi), the cross-validation was performed per patient [121]. Both Dhahbi et al. 2015, and Eltoukhy et al. 2018, utilized the MIAS dataset, along with another open access database (DDSM for Dhahbi et al. 2015 and IRMA for Eltoukhy et al. 2018) [39,122]. Eltoukhy et al. 2018, achieved better performance (93.3% and 90.6% accuracy for the IRMA and MIAS databases respectively) by applying feature extraction based on the curvelet transform and moment theory, in a 10-fold cross-validation scheme but no information was provided regarding the train/test split. Dhahbi et al. 2015, reached 91.3% accuracy when extracting highly significant features (exact Gaussian–Hermite moments), using leave-one-image-out cross-validation. The overall state-of-the-art of the techniques for the detection of masses using feature-based ML is provided in Table 9.

2.4.3. Deep Learning (DL)

Recently, DL has become very popular for the automatic detection of breast abnormalities, using either deep NNs or CNNs [24]. The group of Dhungel et al. 2015, was one of the first to apply DL to mass detection [124]. Two open access databases were employed, DDSM-BCDR and INbreast, containing both digitized and digital mammograms, along with data augmentation to increase the sample size. The first stage classifier contained a multi-scale deep belief network that selected the suspicious regions and then, using a two-level cascade of deep CNNs, the regions were further processed. Finally, the regions that survived the cascade of RF classifiers, were combined using connected component analysis. For the cross-validation, 5-fold cross-validation per patient was applied, and 96% and 75% sensitivity was achieved for each dataset, with 4.8 and 1.2 FPi, respectively. Various studies also took advantage of open access databases. The group of Al-masni et al. 2018, obtained highly accurate results (99.7% accuracy) using the DDSM database with augmentation and a regional DL technique including an ROI-based CNN based on You Only Look Once (YOLO) (Fig. 7) [40]. To avoid any bias in training and testing,

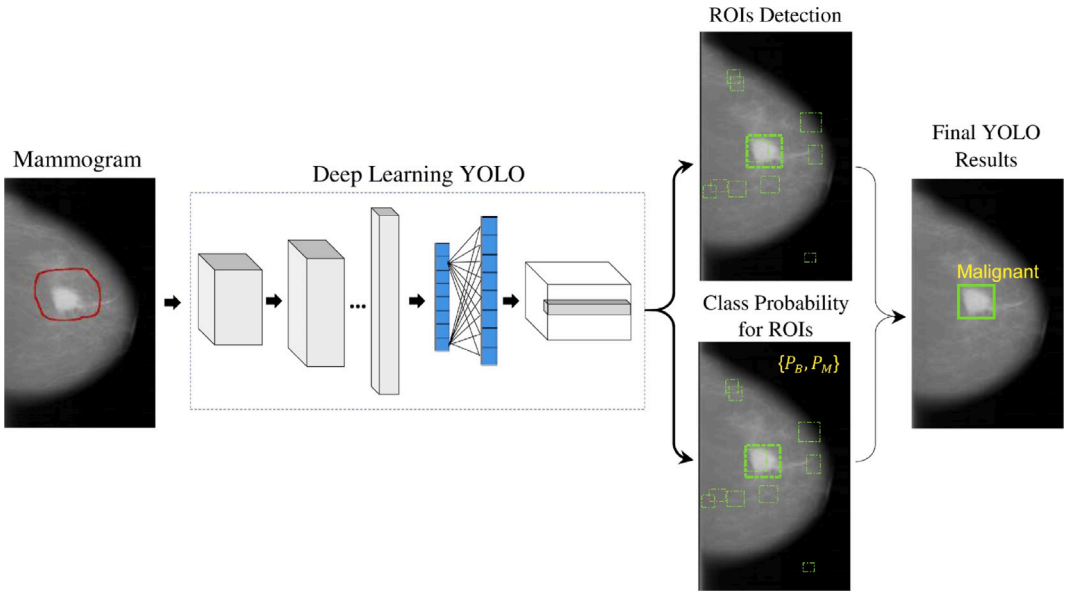

Fig. 7. Scheme of the YOLO-based CAD system for the detection and classification of breast masses using mammograms.
Source: Re-print with permission from [40].

Table 9
Comparison of the state-of-the-art algorithms for the detection of breast masses in mammograms, using feature-based machine learning.

Reference	Database	Type of images	Dataset	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
Tai, Chen and Tsai (2014) [119]	Open Access (DDSM)	Digitized	358 images	LDA	hold-out (per ROI)	–	90.3/-	0.98	–
Dheeba et al. 2014 (2014) [39]	Restricted	Digitized	216 images	Particle Swarm Optimized Wavelet Neural Network (PSOWNN)	1064 ROIs for training (per ROI)	93.7	94.2/92.1	0.97	–
Chu et al. (2014) [115]	Open Access (DDSM)	Digitized	474 images	SVM	66%–33% (per ROI)	81.4	93.4/78.2	–	0.84 FPi
de Nazare Silva et al. (2015) [116]	Open Access (DDSM)	Digitized	599 images	SVM	499-100 (per image)	83.5	92.3/82.2	0.80	1.1 FPi
Dhahbi, Barhoumi and Zagrouba (2015) [120]	Open Access (mini-MIAS)	Digitized	252 images	k-NN	leave-one-out CV (per image)	91.3	–	0.99	–
De Sampaio et al. (2015) [117]	Open Access (DDSM)	Digitized	1727 images	SVM	80%–20% (per image)	–	93/- 83.7/-	–	0.15 FPi 0.19 FPi
Chakraborty, Midya and Rabidas (2018) [121]	Open Access (DDSM)	Digitized	1300 images	NB	10-fold CV (per patient)	–	85/-	–	1.4 FPi
Eltoukhy et al. (2018) [122]	Open Access (IRMA MIAS)	Digitized	1516 ROIs 256 ROIs	AdaBoost	10-fold CV (per ???[a])	93.3 90.6	–	0.96 0.89	–
Dhahbi et al. (2018) [123]	Open Access (DDSM)	Digitized	10168 ROIs	RF	stratified 10-fold CV (per ???)	81.1	–	–	–
Mohanty et al. (2019) [118]	Open Access (MIAS DDSM)	Digitized	314 images 1500 images	Improved Grey Wolf Optimization-based ELM (IGWO-ELM)	stratified 5-fold CV (per image)	100 99.5	–	1 0.99	–

[a]No information was provided on how the data were divided in training and test sets (per ROI/per image/per patient).

the parameters of the system were optimized using only the training set and the final system performance was evaluated using the test set. The training and test sets were defined using 5-fold cross-validation, but per image instead of per patient.

Various pre-trained models were also used with moderate success. Jung et al. 2018, combined INbreast with the restricted GURO dataset, with a total of 632 digital mammograms, along with data augmentation, for a mass detection model based on RetinaNet [124]. Five-fold cross-validation was applied per image, and 98% sensitivity along with 1.3 FPi were reported. Ribli et al. 2018, proposed a CAD system based on one of the most successful object detection frameworks, Faster R-CNN [125]. In total, 3877 both digitized and digital mammograms were collected from two open access databases (DDSM and INbreast) and one restricted dataset. The algorithm was trained using both the DDSM and the restricted dataset and tested on INbreast. The results were promising with 0.95 AUC. The overall state-of-the-art of techniques for the detection of masses in mammograms using DL is provided in Table 10.

Table 10
Comparison of the state-of-the-art algorithms for the detection of breast masses in mammograms, using deep learning.

Reference	Database	Type of images	Dataset	Data augmentation	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
Dhungel et al. 2015 [124]	Open Access (DDSM-BCDR INbreast)	Both	115 cases 79 cases	YES	CNN (cascade of DL and RF)	5-fold CV (per patient)	–	96/- 75/-	–	1.2 FPi 4.8 FPi
Dhungel et al. (2017) [126]	Open Access (INbreast)	Digital	410 images	YES	CNN	5-fold CV (per patient)	90	–	–	1 FPi
Charan et al. (2018) [38]	Open Access (MIAS)	Digitized	322 images	NO	CNN	70%–30% (per image)	65	–	–	–
Ribli et al. (2018) [125]	Both (DDSM INbreast)	Both	2620 images 847 images 410 images	YES	R-CNN	trained on DDSM and restricted — tested on INbreast (per image)	90	–	0.95	0.3 FPi
Agrawal et al. (2018) [36]	Open Access (MIAS)	Digitized	322 images	NO	Voting	pre-trained CNN on open access databases (per ???[a])	80	–	–	–
Jung et al. (2018) [127]	Open Access (INbreast)	Digital	410 images 222 images	YES	CNN (RetinaNet)	5-fold CV (per patient)	–	98/-	–	1.3 FPi
Gao et al. (2018) [128]	Both (INbreast)	Digital	49 patients 89 patients	NO	Shallow-Deep CNN (SD-CNN)	leave-one-out CV (per ???)	90	83/94	0.92	–
Al-masni et al. (2018) [40]	Open Access (DDSM)	Digitized	600 images	YES	CNN (YOLO)	5-fold CV (per image)	99.7	100/94	0.97	–
Shen et al. (2019) [129]	Open Access (CBIS-DDSM INbreast)	Both	2478 images 410 images	YES	CNN	85%–15% stratified (per patient)	–	86.1/80.1 86.7/96.1	0.91 0.98	–
Zeiser et al. (2020) [130]	Open Access (DDSM)	Digitized	7989 images	YES	CNN (U-NET)	70%–10%–20% (per patient)	85.9	92.3/80.5	0.86	–
Aly et al. (2020) [131]	Both (INbreast)	Digital	410 images	YES	CNN (YOLO-V3)	5-fold CV (per image)	89.4	–	–	–

[a]No information was provided on how the data were divided in training and test sets (per ROI/per image/per patient).

2.5. Classification of breast masses

Several groups have developed algorithms for the automatic classification of breast masses as benign or malignant [9,113]. As with previous CAD tasks, it is very challenging to compare the various approaches due to significant deviations in their validation and evaluation techniques. In this review, the various studies were divided into two major categories according to their specific ML approach: (a) feature-based ML, and (b) DL.

2.5.1. Feature-based Machine Learning (ML)

SVMs [74,132,133], ANNs [134,135], MLPs [75,76] and Ensembles [136,137], have been extensively used for the classification of benign vs. malignant masses. In the majority of the studies, the available open access databases were utilized. With SVMs, Berbar, 2018, obtained the highest classification performance, using the DDSM and MIAS open access datasets, with 98.7% and 97.9% accuracy for each dataset respectively [74]. That approach was based on the extraction of seven textural features, using three hybrid methods: Wavelet-CT1, Wavelet-CT2 and ST-GLCM feature merging. In total, 1315 digitized mammograms were included and 2-fold cross-validation was implemented, but it is not specified if it was performed per patient, image or ROI. ANNs have also been applied to mass detection by Xie et al. 2016, who took advantage of two open access datasets (MIAS and DDSM) in order to create a breast mass classification system based on an ELM [135]. The results were promising (96% and 95.7% accuracy on each dataset with 5-fold cross-validation per ROI), and superior to those of Kim and Er, 2004 (70% accuracy) [134]. However, the later used generalized dynamic fuzzy NNs, 343 digitized mammograms from the DDSM dataset and leave-one-out validation per patient. Tahmasbi et al. 2011, proposed an algorithm for the classification of benign and malignant masses based on Zernike moments and MLPs [75]. The classification accuracy obtained, using 322 digitized mammograms from the MIAS database, was 96.4%. The data were divided in 3 groups, 40% as a training set, 30% as a validation set and 30% as a test set. These results were superior to the study of Danala et al. 2018, who used MLP along with a restricted dataset from 111 patients but with leave-one-out cross-validation per patient (78.4% accuracy) [76]. Seryasat and Haddadnia, 2018, evaluated a new ensemble learning framework for mass classification in mammograms [137]. Both the mini-MIAS and DDSM datasets were exploited with a total of 507 digitized mammograms. The algorithm achieved 0.96 and 0.94 AUC (vs. 0.93 by Choi et al. 2016 [136]) for the mini-MIAS and DDSM databases respectively, using 10-fold cross-validation, but the authors did not provide any details if the validation was per patient, per image or per ROI. A summary of feature-based ML approaches can be found in Table 11.

2.5.2. Deep Learning (DL)

Deep learning has been extensively applied to the classification of breast masses as benign or malignant [26]. The group of Arevalo et al. 2015, was one of the first to introduce CNNs for mammographic mass lesion classification [138]. The proposed algorithm replaced the feature extraction and selection steps with a CNN. An open access dataset (BCDR-F03) was exploited, with 763 digitized mammograms. The dataset was divided into training (60%) and test (40%) sets, following a stratified approach per patient, yielding satisfactory results (0.86 AUC). Jiao et al. 2016, and Al-masni et al. 2018, obtained similar classification accuracy (96.7% and 97%) using DL and 600 digitized mammograms from the DDSM database. The group of Al-masni et al. 2018, employed data augmentation, along with Fully Connected Neural Networks (FC-NNs) and 5-fold cross-validation per image [40]. On the other hand, the group of Jiao et al. 2016, combined intensity information and deep features, along with an SVM for the

Table 11
Comparison of the state-of-the-art algorithms for the classification of breast masses in mammograms, using feature-based machine learning.

Reference	Database	Type of images	Dataset	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
Lim and Er (2004) [134]	Open Access (DDSM)	Digitized	343 images	Generalized Dynamic Fuzzy NN (GDFNN)	leave-one-out (per patient)	70	–	0.87	–
Tahmasbi et al. (2011) [75]	Open Access (MIAS)	Digitized	322 images	MLP	40%–30%–30% (per ???[a])	96.4	-/-	0.98	–
Xie et al. (2016) [135]	Open Access (MIAS DDSM)	Digitized	322 images 2620 patients	NN (Extreme Learning Machine)	5-fold CV (per ROI)	96 95.7	96.3/94.3 94.9/97.2	0.97 0.97	–
Choi et al. (2016) [136]	Open Access (DDSM)	Digitized	2743 ROIs 514 ROIs	Ensemble	5 rounds of 2-fold CV (per ROI)	–	–	0.93	–
da Rocha et al. (2016) [132]	Open Access (DDSM)	Digitized	1155 ROIs	SVM	5 rounds of 80%–20% (per ???)	88.3	85/91.9	0.88	–
Danala et al. (2018) [76]	Restricted	Digital	111 patients	MLP	leave-one-out CV (per patient)	78.4	80.8/72.7	0.85	–
Seryasat and Haddadnia (2018) [137]	Open Access (mini-MIAS DDSM)	Digitized	55 images 240 images	Ensemble	5 rounds of 10-fold CV (per ???)	94.8 92	-/- -/-	0.96 0.94	–
Berbar (2018) [74]	Open Access (DDSM MIAS)	Digitized	1024 images 291 images	SVM	2-fold CV (per ???)	98.7 97.9	99.2/- 96.1/-	0.98 0.88	–
de Brito Silva (2020) [133]	Open Access (DDSM)	Digitized	794 ROIs	SVM	80%–20% (per ROI)	90.2	91/89.9	0.96	–

[a]No information was provided on how the data were divided in training and test sets (per ROI/per image/per patient).

final classification, and 2-fold cross-validation but did not provide any information on how the training and test sets were created (i.e. per patient/image/ROI) [139]. Al-Antari et al. 2020, combined the DDSM and INbreast open access datasets [82]. Three popular DL classifiers (feedforward CNN, ResNet-50 and InceptionResNet-V2) were modified and evaluated, with the InceptionResNet-V2 reaching 97.5% and 95.3% accuracy, for each dataset respectively (Fig. 8). In total, 1010 both digitized and digital mammograms were used, and stratified 5-fold cross-validation per patient was applied. These results were significantly improved, compared to the 88% accuracy of Arora et al. 2020, who exploited a modified version of DDSM (CBIS-DDSM) containing 2620 digitized images and a deep ensemble transfer learning neural network classifier [35].

There are several more studies in the literature dealing with mass classification which are even more challenging to compare due to differences in datasets, type of images, and validation schemes used [37, 101,140]. Abdel-Zaher and Eldeib, 2016, reached 99.7% accuracy using 683 digitized samples (Wisconsin breast cancer original database), with data augmentation, deep belief networks and 54.9%–45.1% training vs. test data partition but no information was provided whether it was performed per patient, image or ROI [140]. Wang et al. 2018, obtained 85.2% accuracy using 736 digitized mammograms (BCDR-F03 open access dataset), a Deep Neural Network based on Multi-View data (called MV-DNN) and 50%–10%–40% training, validation and test data partition, per patient [101]. Gnanasekaran et al. 2020, used 1940 digitized and digital mammograms from the DDSM, MIAS but also added a restricted dataset. The proposed CNN model consisted of eight convolutional, four max-pooling and two fully connected layers. For the validation, 80% of the images were used for training and the remaining 20% for testing [141]. Their results were superior (98.3% accuracy) to pre-trained AlexNet and VGG16 networks. A summary of DL approaches for mass classification can be found in Table 12.

2.5.3. Explainable Deep Learning (DL)

Deep learning can provide highly accurate classification but the reasoning behind the resulting decisions is not obvious. For DL to achieve clinical acceptance, it is important for DL algorithms to be interpretable and explainable. Interpretability, also called transparency, can be described as the degree at which a human observer can understand the model. Explainability, can be viewed as the basic characteristics of a model, describing its internal procedures, that clarify or "explain" its decisions [142]. Barnett et al. 2021, created a very promising algorithm (83% accuracy) for interpretable ML-based mammography, to predict whether a lesion is malignant or benign. Their dataset included 1136 newly collected digital mammograms and data augmentation was also applied to increase the size of the dataset. For validation, the dataset was divided into 73% training, 12% validation and 15% testing, ensuring no patient overlap between the test set and the other sets [143].

3. Detection and classification of breast cancer using sequential mammograms

As shown in the previous sections, detection and classification of breast cancer, using the most recent mammogram of a patient, can be effective. However, radiologists routinely compare recent and prior images in order to more effectively identify abnormalities that have developed between the screenings. New abnormalities, which are changing rapidly, are more likely to be suspicious. On the contrary, regions that remain unchanged are, usually, benign and harmless [157]. When prior information is available for direct comparison, abnormalities can be detected at an earlier stage and the radiologists feel more confident of their assessment [148]. The group of Varela et al. 2005, assessed the importance of including prior knowledge in the classification of masses as benign or malignant. Five senior and one resident radiologist, assessed 198 digitized mammograms with and without the use of prior images. The results showed that inclusion of prior images significantly improved the classification (0.8 vs. 0.76 AUC) [158].

Registration is very important when comparing prior and recent mammograms since the location of an abnormality must be correctly linked to the corresponding location in the prior image. The main challenge is to find the optimal transformation function that aligns the two mammograms. Image registration techniques are challenging to apply to mammograms due to significant variations between screenings

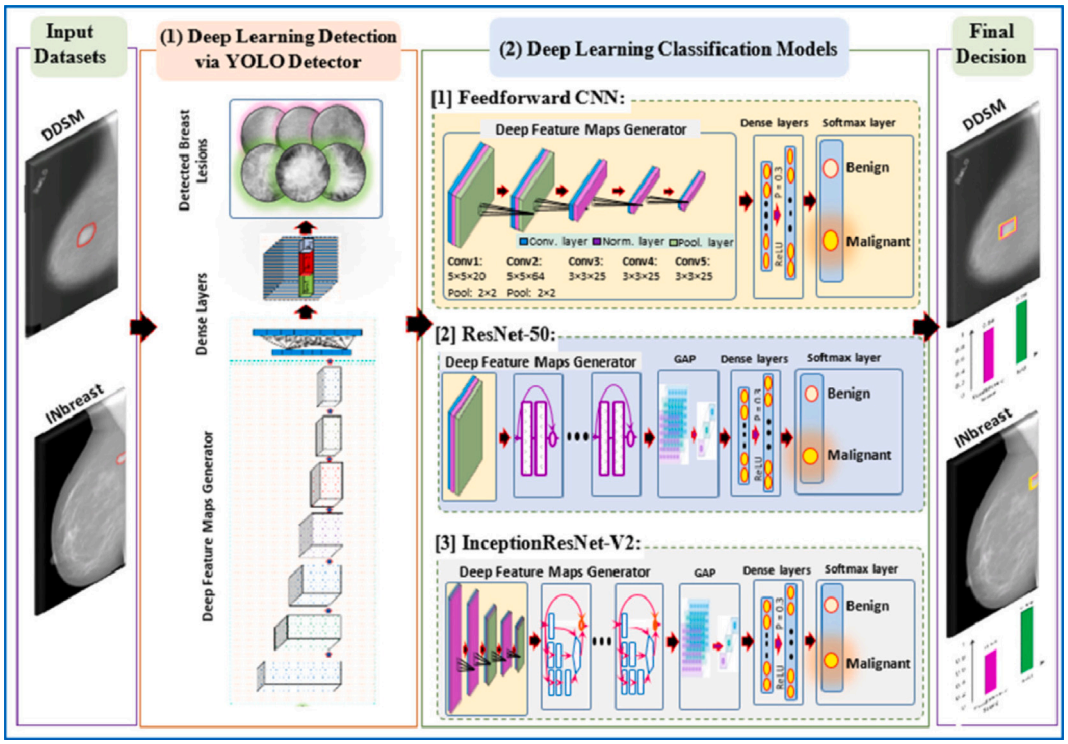

Fig. 8. Schematic diagram of the proposed deep learning integrated computer-aided diagnosis (CAD) framework for breast lesions detection and classification. *Source:* Re-print with permission from [82].

Table 12
Comparison of the state-of-the-art algorithms for the classification of breast masses in mammograms, using deep learning.

Reference	Database	Type of images	Dataset	Data augmentation	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
Arevalo et al. (2015) [138]	Open Access (BCDR-F03)	Digitized	736 images	NO	CNN	60%–40% (per patient)	–	–	0.86	–
Jiao et al. (2016) [139]	Open Access (DDSM)	Digitized	600 images	NO	SVM	2-fold CV (per ???[a])	96.7	–	–	–
Abdel-Zaher and Eldeib (2016) [140]	Open Access (Wisconsin database)	Digitized	683 samples	NO	CNN (Deep Belief Networks)	54.9%–45.1% (per ???)	99.7	100/99.5	–	–
Al-antari et al. (2018) [37]	Open Access (INbreast)	Digital	410 images	YES	CNN	75%–6.25%–18.75% (per image)	95.6	97/92.4	0.95	–
Wang et al. (2018) [101]	Open Access (BCDR-F03)	Digitized	736 images	YES	Deep NN based on Multi-View data	50%–10%–40% (per patient)	85.2	–	0.89	–
Al-masni et al. (2018) [40]	Open Access (DDSM)	Digitized	600 images	YES	FC-NN	5-fold CV (per image)	97	100/94	0.96	–
Arora et al. (2020) [35]	Open Access (CBIS-DDSM)	Digitized	2620 images	NO	DNN	–	88	91/-	0.88	–
Gnanasekaran et al. (2020) [141]	Both (MIAS DDSM)	Digitized	322 images 1416 images 202 images	NO	CNN	80%–20% (per image)	98.3	–	0.98	–
Al-antari et al. (2020) [82]	Open Access (DDSM INbreast)	Both	600 images 410 images	NO	CNN (InceptionResNet-V2)	5-fold CV (per patient)	97.5 95.3	–	–	–
Barnett et al. (2021) [143]	Restricted	Digital	1136 images	YES	DNN (Interpretable DL model)	73%–12%–15 (per patient)	83	–	–	–

[a]No information was provided on how the data were divided in training and test sets (per ROI/per image/per patient).

arising from breast tissue changes, variations in breast compression and operating factors at the time of imaging [159]. Registration methods vary on the extend of the image information used in the procedure. In global registration, all the pixels of a region are utilized for the identification of the corresponding location in a prior image. On the other hand, local registration techniques, only use some of the pixels. Registration techniques also vary with regards to the features used. "Intensity-based" techniques utilize pixel intensity, whereas "feature-based" approaches utilize geometrical structures extracted from the images. Very often intensity-based methods are global and feature-based

Table 13
A list of the state-of-the-art techniques for the registration of mammograms.

Registration category	Reference	Details
Global	van Engeland et al. (2003) [144]	joint probability distribution of the intensity
Local	Vujovic and Brzakovic (1997) [145]	modified monotony operator and accumulator matrix
	Sanjay-Gopal et al. (1999) [146]	based on the associated information between the breast regions
	Marti, Zwiggelaar and Rubi (2001) [147]	correspondence between linear structures
	Hadjiiski et al. (2001) [148]	local affine transformation and nonlinear simplex optimization
	Filev et al. (2008) [149]	corresponding local search area
	Ma et al. (2015) [150]	spatial relations and graph matching
Hybrid	Wirth, Narhan and Gray (2002) [151]	similarity measure and a point-based spatial transformation
	Timp and Karssemeijer (2006) [152]	center of mass alignment and feature space mapping
	Abdel-Nasser, Moreno and Puig (2016) [153]	curvilinear coordinates to manage both global and local deformations
	Li, Chen and Zhang (2018) [154]	global coarse registration and local fine registration
	Sharma et al. (2019) [155]	combined data-driven clustering and deformable image registration
	Mendel et al. (2019) [156]	B-splines registration and multi-resolution registration

methods are local. "Hybrid" methods combine more than one feature-space [160]. Several image registration techniques for breast images have been proposed. Table 13 provides a list of some representative approaches [160,161].

3.1. Temporal analysis

Temporal analysis was the first technique developed for the comparison of sequential mammograms and has already been applied to breast cancer diagnosis. Usually, the abnormality is detected on the most recent mammogram and, using registration algorithms, the corresponding location is identified in the prior mammogram as well. Features are extracted from both images and then those features are subtracted, to create a new feature vector to be used for the classification.

Various studies have assessed the effectiveness of temporal analysis for the diagnosis of breast masses [148,152,162–167] and MCs [168]. However, each one uses a different restricted dataset, collected specifically for the purposes of the particular study, since none of the open access databases include sequential mammograms. Most studies use either feature-based ML or DL. One of the first groups to use feature-based ML and temporal analysis was Hadjiiski et al. 2001. They developed a method for the classification of masses as benign or malignant using temporal analysis of 140 digitized mammogram pairs [148]. LDA was used along with leave-one-patient-out cross-validation, and the performance significantly improved with the addition of information from the prior mammograms (0.88 vs. 0.82 AUC). Timp and Karssemeijer, 2006, also evaluated interval changes between two consecutive mammographic screening rounds to improve the detection of masses. The detection performance, when evaluated on a dataset containing 2873 digitized pairs, also showed an improvement after including temporal features (0.72 vs 0.71 AUC) [152]. Subsequently, Timp and Karssemeijer, 2007, evaluated the effect of one independent reading with CAD and independent double readings, proving that the performance was higher with CAD using SVM and 20-fold cross-validation (0.83 vs. 0.81) [163]. Bozek et al. 2014 and Ma et al. 2015, also performed temporal analysis of breast masses, in 60 and 95 mammogram pairs respectively, leading to significantly improved performance (0.9 AUC) [164,165].

Since 2017 various DL approaches were introduced for the automated diagnosis of breast masses, using CNNs and temporal analysis [25]. Kooi and Karssemeijer, 2017, investigated how the incorporation of symmetry and temporal context information could increase the accuracy of the detection of malignant masses. A gradient boosted tree classifier was trained using 18 366 temporal ROIs and 16-fold cross-validation per patient, and the results showed that temporal information slightly improved the performance of the approach (0.88 vs. 0.87 AUC) [166]. Zheng et al. 2018, proposed a CAD system based on CNNs (VGG-19) and follow-up scans, for breast cancer screening. The algorithm was tested on a dataset containing 96 digital pairs. The mammograms were examined by three cascading object detectors to identify the suspicious regions. The performance of the CAD system was very promising with 92.8% sensitivity and 99.1% specificity but the validation was per image [167].

Temporal analysis has also been used for the classification of benign and malignant MCs, by Hadjiiski et al. 2002 [168]. Sixty-five digitized temporal pairs were collected and several statistical and morphological features were extracted from each MC. The results indicated that the use of temporal analysis improved the performance of MC classification (0.87 vs. 0.81 AUC). Despite the promising results, temporal analysis offers no benefit, over using just the most recent mammogram, when the findings are new, with no traces of the abnormality in the prior screening.

3.2. Subtraction of temporally sequential mammograms

To overcome this limitation, Loizidou et al. 2021, introduced the concept of subtraction of temporally sequential mammograms for the automatic detection and classification of breast abnormalities. The group used both global and local features to register the entire breast area of the recent and prior mammograms. This allowed direct subtraction of the images, effectively removing unchanged regions from the recent mammogram and significantly enhancing subtle recent changes. Loizidou et al. 2021, collected a new dataset with 100 digital pairs to evaluate the use of temporal subtraction to the diagnosis and radiological classification of MCs. The dataset contained precise annotations of each individual MC (benign and suspicious), by two radiologists (Fig. 9) and it is publicly available [169]. Pre-processing, image registration, and post-processing took place along with ML, to eliminate the FPs (normal tissue misclassified as MCs). Subsequently, the true MCs were classified as BI-RADS benign or suspicious. For comparison, the detection and classification was also performed using only the most recent images, without temporal subtraction. Classifying the MCs as BI-RADS benign vs. suspicious, using leave-one-patient-out, resulted in higher accuracy with temporal subtraction compared to using only the most recent mammogram (90.3% vs. 82.7%) [13]. Subsequently, the same group tested a similar approach for the detection and classification of breast masses. The performance of the algorithm was evaluated on a new publicly available dataset of 80 digital pairs [170] and temporal analysis was also performed for comparison. Ninety-six features were extracted and, with the use of ANNs and leave-one-patient-out cross-validation, the accuracy of the classification of the masses as benign or suspicious increased from 92.6%, when using temporal analysis, to 98%, with temporal subtraction [14]. Tables 15, 16 provide a comparison between the state-of-the-art algorithms that used sequential mammograms for the diagnosis of breast masses.

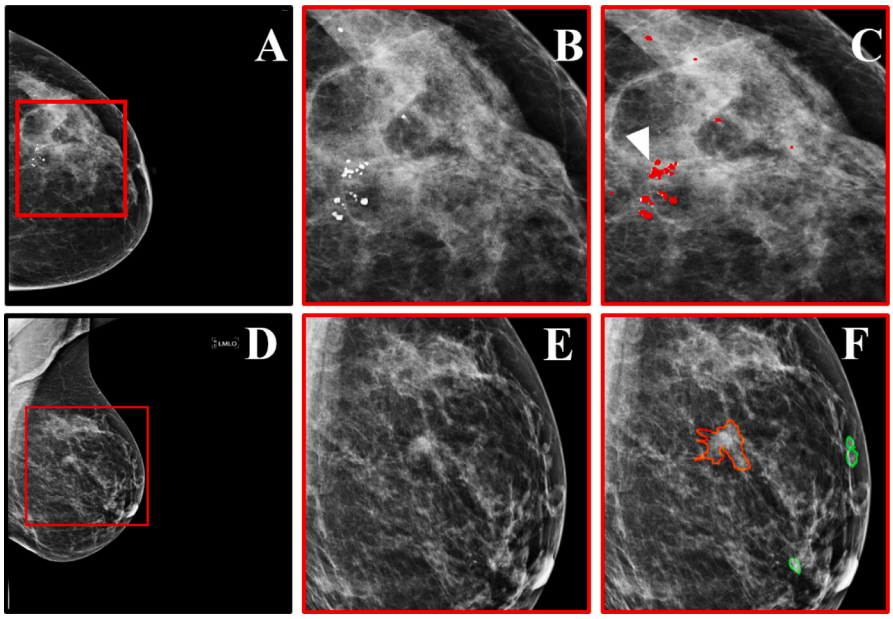

Fig. 9. Dataset examples **(A, D)** Mammographic views of women with benign and suspicious abnormalities. **(B, F)** Zoomed regions marked by the red squares in **A and D**, showing abnormalities. **(C, F)** The regions in **B ,E** with precise marking of abnormalities locations, as annotated by two expert radiologists.
Source: Re-print with permission from [13,14].

4. CAD systems for the Diagnosis of Breast Cancer

Despite the intense research interest, only 8 CAD systems have been approved by the FDA, for the detection and classification of breast cancer (Table 14). Three of those FDA-approved tools are related to triage in mammography, addressing a major challenge especially for radiologists working under high workload conditions. Saige-Q, developed by DeepHealth, obtained FDA approval in 2021 and, since, received considerable attention. Saige-Q is a workflow system that processes mammograms and Digital Breast Tomosynthesis (DBT) examinations, using Artificial Intelligence (AI) and ML, acting as a prioritization tool for the interpreting radiologists. This software helps the user prioritize (triage) the cases by automatically indicating if an image requires further assessment due to suspicious findings. According to the FDA documentation, the AUC was 0.97 for mammography and 0.99 for DBT in studies performed across multiple clinical sites, in two different states, in the United States of America [171].

One of the most widely cited, FDA-approved, tools for the detection and classification of breast abnormalities, is Transpara^{TM}, developed by ScreenPoint Medical [172]. Transpara^{TM} is a DL-driven AI system, designed for automated breast cancer detection in mammography and DBT. It is a radiological, computer-assisted, detection and diagnosis software that guides the radiological assessment [173]. A level of suspicion, between 1 and 100, is assigned to each ROI, representing the possibility that cancer is present, with 100 indicating the highest suspicion. Subsequently, the scores from the four images of a patient are combined to create a comprehensive, examination-based, score. This score can take values between 1 and 10, with 10 indicating the highest probability that cancer is present. During development, training, validation and testing, sizes varies depending on the available version. The mammograms were collected using equipment from four different vendors, Hologic, Siemens, GE Healthcare and Philips Healthcare.

Since 2019, various studies further assessed Transpara. Rodríguez-Ruiz et al. 2019, compared the performance of radiologists reading mammograms unaided vs. reading mammograms with the support of Transpara^{TM} (v.1.3.0) [182]. The results confirmed that their performance improved with the support of an AI system (Fig. 10). In the same year, Rodríguez-Ruiz et al. 2019, tested the newest version of Transpara^{TM} (v.1.4.0) [183]. While the performance of the stand-alone system was higher than the older version (0.84 vs. 0.81 AUC), it was always lower when compared to the best radiologists. This was attributed partly to the fact that the radiologists utilized prior mammograms, when available. Sasaki et al. 2020, used Transpara^{TM} (v.1.3.0) as a stand-alone system for the detection of breast cancer [184]. The diagnostic performance was significantly lower compared to that of the radiologists (0.71 vs. 0.82 AUC). Raya-Povedano et al. 2021, investigated whether Transpara^{TM} (v.1.6.0) could reduce the workload in breast cancer screening without compromising the sensitivity [185]. AI with DBT resulted in 72.5% less workload and 16.7% lower recall rates compared to double reading of DBT images. However, the group had some concerns related to the clinical impact of the AI algorithms since the readers are blinded to both prior examinations and the assessments of the radiologists. van Winkel et al. 2021, investigated whether Transpara^{TM} (v.1.6.0), can increase the accuracy of the radiologists' wide-angle DBT reading [186]. Indeed the AUC was higher (0.86 vs. 0.83) and the reading time per DBT exam was reduced. However, the stand-alone AI system was not statistically better than the average radiologist.

No FDA-approved CAD tools that include analysis of temporally sequential mammograms are available at this time. The introduction of such systems could be a significant improvement in the development of clinically useful technologies.

5. Open access mammography datasets

Open access datasets enable the development of new algorithms and make it easier to compare the performance of a new approaches with the state-of-the-art. However, some open access datasets contain outdated images or are not completely open-access, requiring approval to access. Table 17 summarizes the most widely used open access datasets, as well as important details for each one, such as the year they were published, the number of images, the image format, the resolution (bits/pixel), the available image views (MLO and/or CC), the annotations available and more.

Each dataset has its own advantages and limitations and the choice of dataset depends on the specific needs of a study. The most widely used database is the Digital Database of Screening Mammography (DDSM) [187]. It includes a large number of screenings with CC and

Table 14
Food and Drug Administration (FDA) approved applications for screening mammography.

Tool	Company	Country	Application	Date of approval
cmTriage [174]	CureMetrix	United States	Triage	08/03/2019
HealthMammo [175]	Zebra Medical Vision	Israel	Triage	16/07/2020
Saige-Q [171]	DeepHealth	United States	Triage	16/04/2021
MammoScreen [176]	Therapixel	France	Detection and Classification	25/03/2020
Genius AI Detection [177]	Hologic	United States	Detection and Classification	15/11/2020
ProFound AI Software [178,179]	iCAD	United States	Detection and Classification	12/03/2021
Transapra 1.7.0 [180]	ScreenPoint Medical B.V.	Netherlands	Detection and Classification	02/06/2021
INSIGHT MMG [181]	Lunit	Korea	Detection and Classification	17/11/2021

Table 15
Comparison of the state-of-the-art algorithms for the diagnosis of breast masses from sequential mammograms, using feature-based machine learning.

Reference	Database	Type of images	Dataset	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
Hadjiiski et al. (2001) [148]	Restricted	Digitized	140 pairs	LDA	leave-one-out (per patient)	–	–	0.88 temporal 0.82 single	–
Timp and Karssemeijer (2006) [152]	Restricted	Digitized	2873 pairs	ANN	11-fold CV (per patient)	–	–	0.72 temporal 0.71 single	–
Timp, Varela and Karssemeijer (2007) [162]	Restricted	Digitized	465 pairs	SVM	20-fold CV (per ???[a])	–	–	0.77 temporal 0.74 single	–
Timp, Varela and Karssemeijer (2010) [163]	Restricted	Digitized	198 pairs	SVM	20-fold CV (per ???)	–	–	0.83 temporal 0.81 single	–
Bozek et al. (2014) [164]	Restricted	Digital	60 pairs	LDA	leave-one-out CV (per ???)	–	–	0.90 temporal 0.77 single	–
Ma et al. (2015) [165]	Restricted	Digitized	95 pairs	SMV LDA	5-fold CV (per ???)	–	–	0.9 temporal 0.88 single	–
Loizidou et al. (2022) [14]	Open Access	Digital	80 pairs	ANN	leave-one-out CV (per patient)	–	–	0.98 sub-traction 0.92 analysis	–

[a]No information was provided on how the data were divided in training and test sets (per ROI/per image/per patient).

Table 16
Comparison of the state-of-the-art algorithms for the diagnosis of breast masses from sequential mammograms, using deep learning.

Reference	Database	Type of images	Dataset	Data augmentation	Classifier	Validation method	Results ACC [%]	Results SEN/SPEC [%]	Results AUC	Results other
Kooi and Karssemeijer (2017) [166]	Restricted	Digital	18366 pairs	YES	CNN (shallow gradient boosted tree)	16-fold CV (per patient)	–	–	0.88 temporal 0.87 single	–
Zheng, Yang and Merkulov (2018) [167]	Restricted	Digital	96 pairs	NO	CNN (VGG-19)	10 × 75% – 25% (per image)	–	92.8/99.1	–	0.004 FPi

Fig. 10. Mammograms in 62-year-old woman without cancer, who was recalled (BI-RADS score, ≥ 3) by 12 of 14 radiologists when reading unaided and by seven of 14 readers when using AI system for support. Outlined areas and scores are shown as in viewer of AI system.
Source: Re-print with permission from [182].

MLO mammographic views for each patient. Key limitations of this database are the use of outdated, digitized, film scans, instead of digital mammograms or DBT, and the lack of an Application Programming Interface (API). Various subsets of the DDSM have been further developed over the years, including the Curated Breast Imaging Subset of the DDSM (CBIS-DDSM), which includes benign and malignant cases associated with MCs and masses [188]. The advantages of this dataset are improved annotations, with bounding boxes, and standardized train and test sets available for each abnormality. Another subset of the DDSM, is the DDSM-BCRP which is similar to the CBIS-DDSM, but includes a smaller number of cases [189]. The main disadvantage of this subset is a minimal consensus of its ground truth annotations. The Mammographic Image Analysis Society (MIAS) is another widely used database, containing 322 digitized mammograms [190]. Unfortunately, the data are limited to the MLO view only and the films were digitized at low resolution. INbreast is a highly cited database developed in Portugal in 2011 and includes digital mammograms from 115 patients [191]. For each patient both the MLO and CC views are available. However, INbreast is restricted and the researchers must contact the authors directly to gain access. The Breast Cancer Digital Repository (BCDR) is an open access dataset that is sub-divided into two different repositories, a Film Mammography-based Repository (BCDR-FM) and a Full Field Digital Mammography-based Repository (BCDR-DM) [192]. Mammograms are available both for the CC and the MLO view, and auxiliary patient data (prior surgery, lesion characteristics, biopsy status) are also available. Although it is not commonly used in the literature, anyone can contribute new data through the website. The OPTIMAM mammography image database is an extremely large database, with almost 3 million mammographic images [193]. The majority of these images have been annotated using bounding boxes. The database has an open-source API for easy image retrieval in Python, but administrator approval is required for access. However, all the data originated from patients in the United Kingdom (UK) and some images are not annotated.

In the majority of the open access databases, the mammograms are film-based and digitized at low resolution. Also, the annotations are not detailed, with some databases providing only the coordinates of the abnormality or, at most, the bounding box that contains it. In addition, there are no sequential mammograms in any of the above databases. The Sequential Digital Mammograms (SDM) database is a newly developed set of sequential mammograms which includes both the prior and recent CC and MLO views of each patient. For every mammogram, precise annotation of each individual abnormality, as defined by two expert radiologists, is also provided (Fig. 11). The dataset is open source but, unfortunately, only some of the suspicious cases have biopsy confirmation. Separate datasets contain both MCs [169] and masses [170].

6. Discussion

This review covers the recent progress in the automated detection and/or classification of abnormalities in mammograms. The studies included were selected based on their historical significance, number of citations, and/or the effectiveness of the algorithms proposed. For an effective comparison, the studies were divided into two major categories based on the breast abnormality under investigation: (a) MCs and (b) masses. The articles were further sub-divided based on their goal: (a) detection and (b) classification of the abnormalities, and subsequently, each sub-category was further divided based on the type of ML used.

Although the main steps in the development of CAD systems, are the same (i.e. pre-processing, detection and classification), there are several approaches to implement each step. Pre-processing is meant to reduce the noise, enhance the contrast, and eliminate areas which are not diagnostically relevant. The various pre-processing methodologies in the literature use either conventional image processing techniques, feature-based approaches, or DL. Common pre-processing techniques include normalization, Sobel filtering, histogram equalization and thresholding. However, due to the numerous variations in available filters and other methods, it is difficult to compare the published studies and come to a definitive conclusion as to the most suitable or most effective pre-processing approach.

Accurate segmentation is the first step in the development of a successful detection algorithm and, thus, various techniques have been proposed for effective segmentation. A breast abnormality can be fully identified and segmented using shape, intensity and textural features.

Table 17
Comparison of the most commonly used open access mammography datasets.

Name	Origin	Year	File access	Number of cases	Number of images	Image format	Resolution (bits/pixel)	Image mode	Available view	Type of abnormality	Image categories	BI-RADS	Annotation
DDSM [187]	USA	1999	Open	2620	10480	.LJPEG	8 or 16	Digitized	CC/MLO	ALL	Normal Benign Malignant	YES	Contour points of the ROI
MIAS [190]	UK	2003	Open	161	322	.PGM	8	Digitized	MLO	ALL	Normal Benign Malignant	NO	Center and radius of a circle around ROI
INbreast [191]	Portugal	2011	Approval from authors	115	410	.DICOM	14	Digital	CC/MLO	Masses Calcifications Distortions Asymmetries	Normal Benign Malignant	YES	Contour points of the ROI
BCDR-FM [192]	Portugal	2012	Open (requires registration)	1125	3703	.TIFF	8	Digitized	CC/MLO	ALL	Normal Cancer	YES	Precise lesion locations and mass coordinates, detailed segmentation outlines
BCDR-DM [192]	Portugal	2012	Open (requires registration)	1042	3612	.TIFF	14	Digital	CC/MLO	ALL	Normal Cancer	YES	Precise lesion locations and mass coordinates, detailed segmentation outlines
CBIS-DDSM [188]	USA	2017	Open	1566	10239	.DICOM	8 or 16	Digitized	CC/MLO	Masses Calcifications	Benign Malignant	YES	ROI segmentation and bounding boxes
DDSM-BCRP [189]	USA	2000	Open	179	716	.TIFF	12	Digitized	CC/MLO	Masses Calcifications	Benign Malignant	YES	Contour points of the ROI
OPTIMAM [193]	UK	2020	Approval from authors	–	2889312	.DICOM	12 or 16	Digital	CC/MLO	ALL	Normal Benign Malignant	YES	Rectangular around the boundaries of the ROI
SDM-MCs [169]	Cyprus	2021	Open	100	400	.DICOM	12	Digital	CC/MLO and priors	Calcifications	Normal Benign Suspicious	YES	Precise annotation of each micro-calcification
SDM-Masses [170]	Cyprus	2022	Open	80[a]	320	.DICOM	12	Digital	CC/MLO and priors	Masses	Normal Benign Suspicious	YES	Precise annotation of each mass

[a]In progress.

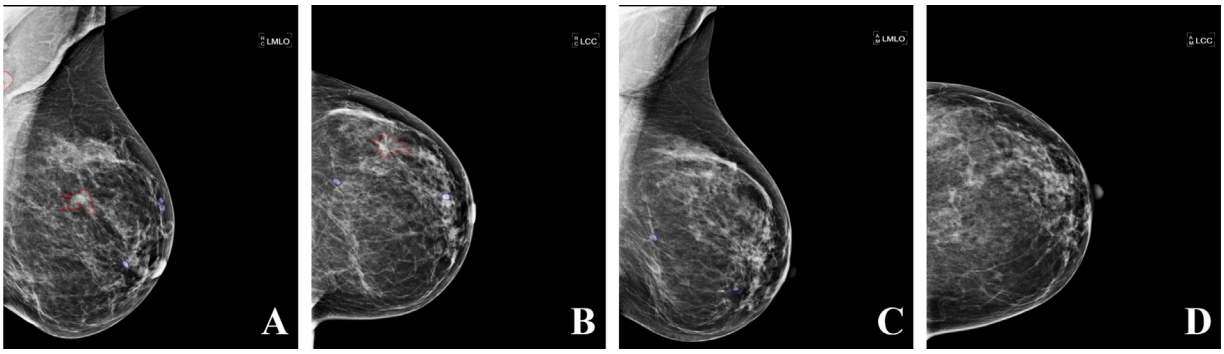

Fig. 11. Dataset example from the Sequential Digital Mammograms (SDM) database with precise marking of mass locations, as annotated by two expert radiologists (green for benign and red for suspicious). **(A)** Most recent MLO view. **(B)** Most recent CC view. **(C)** Prior MLO view. **(D)** Prior CC view.

However, there are numerous such features that can be extracted from each ROI. Selection of the best combination of features is usually required and can be performed using a variety of techniques, each leading to a different subset of features being selected, depending on its underlying principle. However, even a difference in a single feature can significantly increase or decrease the performance. Since the features selected in each study are never exactly the same, the widely varying studies in the literature are not directly comparable.

Subsequently, the features selected are incorporated into a classifier for the classification of the regions as normal, benign, suspicious or malignant, depending on the study. Several classifiers have been exploited for this task. Deep learning and CNNs, have also been extensively used for all stages of automatic feature extraction, selection and classification. The comparison of studies that use DL is even more challenging, due to the high number of parameters that must be tuned and which significantly affect the performance of the network. The decision to use feature-based ML or DL depends solely on the number of the available data and the main goal of the study. Data augmentation is often necessary to increase the size of the data. However, augmentation has been, so far, limited to making geometric transformations of the data which introduces bias. The bias is even worse when the augmented ROIs and images of a single patient are included in both the training and test sets.

A critical parameter that must be taken into account when comparing published results, is the method used for validation. The majority of the studies in the literature use k-fold cross-validation, with the number of k varying from 1 to 30. However, it is important to consider how the training and test sets are created. Validation should be performed per patient, i.e. the training and test sets should not both contain images or ROIs of the same patient. This avoid bias and assures that, as in real clinical scenarios, the test patients are truly unknown to the system. Unfortunately, in many studies the validation is performed per image or, even worse, per ROI. Even if the same datasets, features and classifiers are used in two studies, the validation approach can significantly change the performance, providing an undue advantage due to bias. The best practise for groups developing computer-based algorithms for the detection and/or classification of breast abnormalities in mammograms, is to utilize open access datasets, if possible, with digital up-to-date mammograms and detailed ground truth annotations, for the results to be comparable between studies. In addition, validation should be performed per patient and all metrics of success (i.e. sensitivity, specificity, accuracy and AUC) should be reported. New datasets should also be made openly available to assure that future studies prove the repeatability of novel results.

The studies, presented in this review, have the potential to significantly decrease the performed biopsies, minimize the detected error rates and assist the radiologists in the diagnosis of breast abnormalities, still, most of the proposed algorithms are far from real world application. Their performance is encouraging but still not reliable enough to be accepted as a standalone clinical tool by the clinical community. One of the main reasons is the un-explainability of the results of the ML models. Deep learning algorithms utilize medical images as input and return a decision which, while accurate, does not provide a reasonable rationale. Thus, explainable AI algorithms are required. Explainable AI, refers to a set of methods that allow humans, in this case clinicians, to comprehend and trust the results provided by ML algorithms. When CAD systems are acting as second readers, or play an important part in the decision-making process, the clinicians should both understand and trust the results. Another important factor limiting the clinical applicability of approaches reviewed here are the data used. The unavailability of large-scale, publicly available, databases forces researchers to collect private data, resulting in various fragmented datasets with different properties and imbalanced classes. Hence, the results reported in the literature, despite their great performance, cannot be generalized and often are limited to a second reader role in clinical practise.

7. Conclusions

This review has concentrated only on mammography for the detection of MCs and masses in the breast. For a comprehensive analysis of breast cancer detection and classification, other imaging modalities such as DBT, ultrasound, and MRI, must be considered. In addition, the shear volume of published articles, especially regarding the identification and classification of masses, which are considered diagnostically more important, does not allow incorporation of all available articles in a single review. Unfortunately, some technical details from some studies, e.g. how validation was performed or complete metrics of success, are not included in this review, since they were not reported in all articles. From the current state of the art, included in this review, it is evident that CAD algorithms must be further investigated and validated with more clinical data, in order to unequivocally demonstrate their efficacy, value, and impact on breast cancer diagnosis. It is also important, for the research community, to consider best practices that promote consolidation, repeatability, and explainability, if CAD systems for the detection of breast abnormalities are to gain a firm clinical footing. Despite its limitations, this comprehensive review can serve both as a thorough introduction to the field but also provide indicative guidelines for the design and execution of future studies.

Declaration of competing interest

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

Acknowledgments

The publication of this paper is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 739551 (KIOS CoE) and the Government of the Republic of Cyprus through the Cyprus Deputy Ministry of Research, Innovation and Digital Policy.

References

[1] J. Ferlay, M. Laversanne, M. Ervik, F. Lam, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram, F. Bray, Global cancer observatory: Cancer tomorrow. Lyon, France: International agency for research on Cancer, 2022, https://gco.iarc.fr/tomorrow. (Accessed 24 October 2022).

[2] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2022, CA: Cancer J. Clin. 72 (1) (2022) 7–33.

[3] D.A. Spak, J. Plaxco, L. Santiago, M. Dryden, B. Dogan, BI-RADS® fifth edition: A summary of changes, Diagn. Interv. Imaging 98 (3) (2017) 179–190.

[4] Breast anatomy and how cancer starts | about breast cancer, 2020, https://nbcf.org.au/about-national-breast-cancer-foundation/about-breast-cancer/what-you-need-to-know/breast-anatomy-cancer-starts/. (Accessed 20 September 2020).

[5] K. Ganesan, U.R. Acharya, C.K. Chua, L.C. Min, K.T. Abraham, K.-H. Ng, Computer-aided breast cancer detection using mammograms: a review, IEEE Rev. Biomed. Eng. 6 (2013) 77–98.

[6] S. Beura, Development of Features and Feature Reduction Techniques for Mammogram Classification (Ph.D. thesis), Department of Computer Science and Engineering National Institute of Technology Rourkela, 2016.

[7] Dense breasts - fact sheets - yale medicine, 2022, https://www.yalemedicine.org. (Accessed 24 October 2022).

[8] R.M. Rangayyan, F.J. Ayres, J.L. Desautels, A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs, J. Franklin Inst. B 344 (3–4) (2007) 312–348.

[9] A. Oliver, J. Freixenet, J. Marti, E. Perez, J. Pont, E.R. Denton, R. Zwiggelaar, A review of automatic mass detection and segmentation in mammographic images, Med. Image Anal. 14 (2) (2010) 87–110.

[10] X. Liu, Z. Zeng, A new automatic mass detection method for breast cancer with false positive reduction, Neurocomputing 152 (2015) 388–402.

[11] H. Zonderland, R. Smithuis, BI-RADS for mammography and ultrasound 2013 updated version, Radiol. Assist. (2014).

[12] W. Xie, Y. Li, Y. Ma, Breast mass classification in digital mammography based on extreme learning machine, Neurocomputing 173 (2016) 930–941.

[13] K. Loizidou, G. Skouroumouni, C. Pitris, C. Nikolaou, Digital subtraction of temporally sequential mammograms for improved detection and classification of microcalcifications, Eur. Radiol. Exp. 5 (1) (2021) 1–12.

[14] K. Loizidou, G. Skouroumouni, C. Nikolaou, C. Pitris, Automatic breast mass segmentation and classification using subtraction of temporally sequential digital mammograms, IEEE J. Transl. Eng. Health Med. (2022).

[15] X. Castells, I. Torá-Rocamora, M. Posso, M. Román, M. Vernet-Tomas, A. Rodríguez-Arana, L. Domingo, C. Vidal, M. Baré, J. Ferrer, et al., Risk of breast cancer in women with false-positive results according to mammographic features, Radiology 280 (2) (2016) 379–386.

[16] S. Banik, R.M. Rangayyan, J.L. Desautels, Detection of architectural distortion in prior mammograms, IEEE Trans. Med. Imaging 30 (2) (2010) 279–294.

[17] H.D. Nelson, M. Pappas, A. Cantor, J. Griffin, M. Daeges, L. Humphrey, Harms of breast cancer screening: systematic review to update the 2009 US preventive services task force recommendation, Ann. Intern. Med. 164 (4) (2016) 256–267.

[18] C.D. Lehman, R.D. Wellman, D.S. Buist, K. Kerlikowske, A.N. Tosteson, D.L. Miglioretti, Diagnostic accuracy of digital screening mammography with and without computer-aided detection, JAMA Intern. Med. 175 (11) (2015) 1828–1837.

[19] H.-D. Cheng, X. Cai, X. Chen, L. Hu, X. Lou, Computer-aided detection and classification of microcalcifications in mammograms: a survey, Pattern Recognit. 36 (12) (2003) 2967–2991.

[20] R.M. Nishikawa, M. Kallergi, C.G. Orton, Computer-aided detection, in its present form, is not an effective aid for screening mammography, Med. Phys. 33 (4) (2006) 811–814.

[21] K. Doi, Computer-aided diagnosis in medical imaging: historical review, current status and future potential, Comput. Med. Imaging Graph. 31 (4–5) (2007) 198–211.

[22] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.

[23] A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, R. Zwiggelaar, Deep learning in mammography and breast histology, an overview and future trends, Med. Image Anal. 47 (2018) 45–67.

[24] S.J.S. Gardezi, A. Elazab, B. Lei, T. Wang, Breast cancer detection and diagnosis using mammographic data: Systematic review, J. Med. Internet Res. 21 (7) (2019) e14464.

[25] L. Zou, S. Yu, T. Meng, Z. Zhang, X. Liang, Y. Xie, A technical review of convolutional neural network-based mammographic breast cancer diagnosis, Comput. Math. Methods Med. (2019).

[26] S.Z. Ramadan, Methods used in computer-aided diagnosis for breast cancer detection using mammograms: a review, J. Healthc. Eng. 2020 (2020).

[27] L. Abdelrahman, M. Al Ghamdi, F. Collado-Mesa, M. Abdel-Mottaleb, Convolutional neural networks for breast cancer detection in mammography: A survey, Comput. Biol. Med. 131 (2021) 104248.

[28] D. Painuli, S. Bhardwaj, et al., Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review, Comput. Biol. Med. (2022) 105580.

[29] H.P. Chan, C.J. Vyborny, H. MacMahon, C.E. Metz, K. Doi, E.A. Sickles, Digital mammography. ROC studies of the effects of pixel size and unsharp-mask filtering on the detection of subtle microcalcifications, Invest. Radiol. 22 (7) (1987) 581–589.

[30] A.P. Dhawan, G. Buelloni, R. Gordon, Enhancement of mammographic features by optimal adaptive neighborhood image processing, IEEE Trans. Med. Imaging 5 (1) (1986) 8–15.

[31] J.K. Kim, J.M. Park, K.S. Song, H.W. Park, Adaptive mammographic image enhancement using first derivative and local statistics, IEEE Trans. Med. Imaging 16 (5) (1997) 495–502.

[32] T. Basile, A. Fanizzi, L. Losurdo, R. Bellotti, U. Bottigli, R. Dentamaro, V. Didonna, A. Fausto, R. Massafra, M. Moschetta, et al., Microcalcification detection in full-field digital mammograms: A fully automated computer-aided system, Phys. Medica 64 (2019) 1–9.

[33] N. Karssemeijer, Adaptive noise equalization and image analysis in mammography, in: Biennial International Conference on Information Processing in Medical Imaging, Springer, 1993, pp. 472–486.

[34] R. Rouhi, M. Jafari, S. Kasaei, P. Keshavarzian, Benign and malignant breast tumors classification based on region growing and CNN segmentation, Expert Syst. Appl. 42 (3) (2015) 990–1002.

[35] R. Arora, P.K. Rai, B. Raman, Deep feature–based automatic classification of mammograms, Med. Biol. Eng. Comput. (2020) 1–13.

[36] S. Agrawal, R. Rangnekar, D. Gala, S. Paul, D. Kalbande, Detection of breast cancer from mammograms using a hybrid approach of deep learning and linear classification, in: 2018 International Conference on Smart City and Emerging Technology, ICSCET, IEEE, 2018, pp. 1–6.

[37] M.A. Al-antari, M.A. Al-masni, M.-T. Choi, S.-M. Han, T.-S. Kim, A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification, Int. J. Med. Inform. 117 (2018) 44–54.

[38] S. Charan, M.J. Khan, K. Khurshid, Breast cancer detection in mammograms using convolutional neural network, in: 2018 International Conference on Computing, Mathematics and Engineering Technologies, ICoMET, IEEE, 2018, pp. 1–5.

[39] J. Dheeba, N.A. Singh, S.T. Selvi, Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach, J. Biomed. Inform. 49 (2014) 45–52.

[40] M.A. Al-Masni, M.A. Al-Antari, J.-M. Park, G. Gi, T.-Y. Kim, P. Rivera, E. Valarezo, M.-T. Choi, S.-M. Han, T.-S. Kim, Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system, Comput. Methods Programs Biomed. 157 (2018) 85–94.

[41] H. Cai, Q. Huang, W. Rong, Y. Song, J. Li, J. Wang, J. Chen, L. Li, Breast microcalcification diagnosis using deep convolutional neural network from digital mammograms, Comput. Math. Methods Med. 2019 (2019).

[42] A.F. Laine, S. Schuler, J. Fan, W. Huda, Mammographic feature enhancement by multiscale analysis, IEEE Trans. Med. Imaging 13 (4) (1994) 725–740.

[43] M.M.H. Chowdhury, A. Khatun, Image compression using discrete wavelet transform, Int. J. Comput. Sci. Issues (IJCSI) 9 (4) (2012) 327.

[44] H.-D. Cheng, Y.M. Lui, R.I. Freimanis, A novel approach to microcalcification detection using fuzzy logic technique, IEEE Trans. Med. Imaging 17 (3) (1998) 442–450.

[45] L.N. Kegelmeyer, J.A.M. Hernandez, C.M. Logan, Automated analysis for microcalcifications in high-resolution digital mammograms, in: Medical Imaging 1993: Image Processing, Vol. 1898, International Society for Optics and Photonics, 1993, pp. 472–480.

[46] H. Li, K.R. Liu, S.-C. Lo, Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms, IEEE Trans. Med. Imaging 16 (6) (1997) 785–798.

[47] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2015) 295–307.

[48] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.

[49] K. Umehara, J. Ota, T. Ishida, Super-resolution imaging of mammograms based on the super-resolution convolutional neural network, Open J. Med. Imaging 7 (4) (2017) 180–195.

[50] Y. Jiang, J. Li, Generative adversarial network for image super-resolution combining texture loss, Appl. Sci. 10 (5) (2020) 1729.

[51] K.S. Woods, C.C. Doss, K.W. Bowyer, J.L. Solka, C.E. Priebe, W.P. Kegelmeyer Jr., Comparative evaluation of pattern recognition techniques for detection of microcalcifications in mammography, Int. J. Pattern Recognit. Artif. Intell. 7 (06) (1993) 1417–1436.

[52] D. Davies, D. Dance, Automatic computer detection of clustered calcifications in digital mammograms, Phys. Med. Biol. 35 (8) (1990) 1111.

[53] S. Yu, L. Guan, A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films, IEEE Trans. Med. Imaging 19 (2) (2000) 115–126.

[54] B. Singh, M. Kaur, An approach for classification of malignant and benign microcalcification clusters, Sādhanā 43 (3) (2018) 39.

[55] V. Kumar, P. Gupta, Importance of statistical measures in digital image processing, Int. J. Emerg. Technol. Adv. Eng. 2 (8) (2012) 56–62.

[56] A.J. Bekker, M. Shalhon, H. Greenspan, J. Goldberger, Multi-view probabilistic classification of breast microcalcifications, IEEE Trans. Med. Imaging 35 (2) (2016) 645–653.

[57] A. Fanizzi, T. Basile, L. Losurdo, R. Bellotti, U. Bottigli, R. Dentamaro, V. Didonna, A. Fausto, R. Massafra, M. Moschetta, et al., A machine learning approach on multiscale texture analysis for breast microcalcification diagnosis, BMC Bioinformatics 21 (2) (2020) 1–11.

[58] I. Zyout, I. Abdel-Qader, Classification of microcalcification clusters via pso-knn heuristic parameter selection and glcm features, Int. J. Comput. Appl. 31 (2) (2011) 34–39.

[59] W. Jian, X. Sun, S. Luo, Computer-aided diagnosis of breast microcalcifications based on dual-tree complex wavelet transform, Biomed. Eng. Online 11 (1) (2012) 96.

[60] M. Milosevic, D. Jankovic, A. Peulic, Segmentation for the enhancement of microcalcifications in digital mammograms, Technol. Health Care 22 (5) (2014) 701–715.

[61] G.L. Rogova, P.C. Stomper, C.-C. Ke, Microcalcification texture analysis in a hybrid system for computer-aided mammography, in: Medical Imaging 1999: Image Processing, Vol. 3661, International Society for Optics and Photonics, 1999, pp. 1426–1434.

[62] T. Bhangale, U.B. Desai, U. Sharma, An unsupervised scheme for detection of microcalcifications on mammograms, in: Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101), Vol. 1, IEEE, 2000, pp. 184–187.

[63] R.N. Strickland, H.I. Hahn, L.J. Baig, Wavelet methods for combining CAD with enhancement of mammograms, in: Medical Imaging 1996: Image Processing, Vol. 2710, International Society for Optics and Photonics, 1996, pp. 888–904.

[64] A. Mencattini, M. Salmeri, R. Lojacono, M. Frigerio, F. Caselli, Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing, IEEE Trans. Instrum. Meas. 57 (7) (2008) 1422–1430.

[65] A. Ghasemzadeh, S.S. Azad, E. Esmaeili, Breast cancer detection based on Gabor-wavelet transform and machine learning methods, Int. J. Mach. Learn. Cybern. 10 (7) (2019) 1603–1612.

[66] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. (6) (1973) 610–621.

[67] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[68] C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, J. Bioinform. Comput. Biol. 3 (02) (2005) 185–205.

[69] P. Diehr, D.C. Martin, T. Koepsell, A. Cheadle, Breaking the matches in a paired t-test for community interventions when the number of pairs is small, Stat. Med. 14 (13) (1995) 1491–1504.

[70] L. Stahle, S. Wold, Multivariate analysis of variance (MANOVA), Chemometr. Intell. Lab. Syst. 9 (2) (1990) 127–141.

[71] A. Papadopoulos, D.I. Fotiadis, A. Likas, Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines, Artif. Intell. Med. 34 (2) (2005) 141–150.

[72] W. Jian, X. Sun, S. Luo, Computer-aided diagnosis of breast microcalcifications based on dual-tree complex wavelet transform, Biomed. Eng. Online 11 (1) (2012) 1–12.

[73] Z. Suhail, E.R. Denton, R. Zwiggelaar, Classification of micro-calcification in mammograms using scalable linear Fisher discriminant analysis, Med. Biol. Eng. Comput. 56 (8) (2018) 1475–1485.

[74] M.A. Berbar, Hybrid methods for feature extraction for breast masses classification, Egypt. Inform. J. 19 (1) (2018) 63–73.

[75] A. Tahmasbi, F. Saki, S.B. Shokouhi, Classification of benign and malignant masses based on Zernike moments, Comput. Biol. Med. 41 (8) (2011) 726–735.

[76] G. Danala, F. Aghaei, M. Heidari, T. Wu, B. Patel, B. Zheng, Computer-aided classification of breast masses using contrast-enhanced digital mammograms, in: Medical Imaging 2018: Computer-Aided Diagnosis, Vol. 10575, International Society for Optics and Photonics, 2018, p. 105752K.

[77] K. Hu, W. Yang, X. Gao, Microcalcification diagnosis in digital mammography using extreme learning machine based on hidden Markov tree model of dual-tree complex wavelet transform, Expert Syst. Appl. 86 (2017) 135–144.

[78] E. Sert, S. Ertekin, U. Halici, Ensemble of convolutional neural networks for classification of breast microcalcification from mammograms, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2017, pp. 689–692.

[79] K.-C. Chen, C.-L. Chin, N.-C. Chung, C.-L. Hsu, Combining multi-classifier with CNN in detection and classification of breast calcification, in: International Conference on Biomedical and Health Informatics, Springer, 2019, pp. 304–311.

[80] G. Cai, Y. Guo, W. Chen, H. Zeng, Y. Zhou, Y. Lu, Computer-aided detection and diagnosis of microcalcification clusters on full field digital mammograms based on deep learning method using neutrosophic boosting, Multimedia Tools Appl. 79 (23) (2020) 17147–17167.

[81] S.S. Chakravarthy, H. Rajaguru, Automatic detection and classification of mammograms using improved extreme learning machine with deep learning, IRBM (2021).

[82] M.A. Al-Antari, S.-M. Han, T.-S. Kim, Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms, Comput. Methods Programs Biomed. 196 (2020) 105584.

[83] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[84] M.Z. Mehdi, N.G.B. Ayed, A.D. Masmoudi, D. Sellami, R. Abid, An efficient microcalcifications detection based on dual spatial/spectral processing, Multimedia Tools Appl. 76 (11) (2017) 13047–13065.

[85] A. Touil, K. Kalti, P.-H. Conze, B. Solaiman, M.A. Mahjoub, Automatic detection of microcalcification based on morphological operations and structural similarity indices, Biocybern. Biomed. Eng. 40 (3) (2020) 1155–1173.

[86] S. Punitha, S. Ravi, M.A. Devi, J. Vaishnavi, Computer-aided mammography techniques for detection and classification of microcalcifications in digital mammograms, Int. J. Imag. Min. 3 (1) (2018) 48–66.

[87] A. M.N. Kumar, A. M.N. Kumar, H. S. Sheshadri, Computer aided detection of clustered microcalcification: A survey, Curr. Med. Imaging 15 (2) (2019) 132–149.

[88] I. El-Naqa, Y. Yang, M.N. Wernick, N.P. Galatsanos, R.M. Nishikawa, A support vector machine approach for detection of microcalcifications, IEEE Trans. Med. Imaging 21 (12) (2002) 1552–1563.

[89] H. Li, R. Mukundan, S. Boyd, A novel application of multifractal features for detection of microcalcifications in digital mammograms, in: Annual Conference on Medical Image Understanding and Analysis, Springer, 2019, pp. 26–37.

[90] V.A. Karale, J.P. Ebenezer, J. Chakraborty, T. Singh, A. Sadhu, N. Khandelwal, S. Mukhopadhyay, A screening CAD tool for the detection of microcalcification clusters in mammograms, J. Digit. Imaging 32 (5) (2019) 728–745.

[91] S. Cai, P.-Z. Liu, Y.-M. Luo, Y.-Z. Du, J.-N. Tang, Breast microcalcification detection algorithm based on contourlet and ASVM, Algorithms 12 (7) (2019) 135.

[92] V.A. Karale, T. Singh, A. Sadhu, N. Khandelwal, S. Mukhopadhyay, Reduction of false positives in the screening CAD tool for microcalcification detection, Sādhanā 45 (1) (2020) 1–11.

[93] J.G. Melekoodappattu, P.S. Subbian, A hybridized ELM for automatic micro calcification detection in mammogram images based on multi-scale features, J. Med. Syst. 43 (7) (2019) 1–12.

[94] J.-J. Mordang, T. Janssen, A. Bria, T. Kooi, A. Gubern-Mérida, N. Karssemeijer, Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks, in: International Workshop on Breast Imaging, Springer, 2016, pp. 35–42.

[95] X. Zhang, Z. Wang, A microcalcification cluster detection method based on deep learning and multi-scale feature fusion, J. Supercomput. 75 (9) (2019) 5808–5830.

[96] S.Z. Ramadan, Using convolutional neural network with cheat sheet and data augmentation to detect breast cancer in mammograms, Comput. Math. Methods Med. 2020 (2020).

[97] B. Savelli, C. Marrocco, A. Bria, M. Molinara, F. Tortorella, Combining convolutional neural networks for multi-context microcalcification detection in mammograms, in: International Conference on Computer Analysis of Images and Patterns, Springer, 2019, pp. 36–44.

[98] A. Hakim, P. Prajitno, D. Soejoko, Microcalcification detection in mammography image using computer-aided detection based on convolutional neural network, in: AIP Conference Proceedings, Vol. 2346, AIP Publishing LLC, 2021, 040001.

[99] M.S. Hossain, Microcalcification segmentation using modified u-net segmentation network from mammogram images, J. King Saud Univ. Inf. Sci. (2019).

[100] J. Wang, R.M. Nishikawa, Y. Yang, Global detection approach for clustered microcalcifications in mammograms using a deep learning network, J. Med. Imaging 4 (2) (2017) 024501.

[101] J. Wang, Y. Yang, A context-sensitive deep learning approach for microcalcification detection in mammograms, Pattern Recognit. 78 (2018) 12–22.

[102] G. Valvano, G. Santini, N. Martini, A. Ripoli, C. Iacconi, D. Chiappino, D. Della Latta, Convolutional neural networks for the segmentation of microcalcification in mammography imaging, J. Healthc. Eng. 2019 (2019).

[103] C. Marasinou, B. Li, J. Paige, A. Omigbodun, N. Nakhaei, A. Hoyt, W. Hsu, Segmentation of breast microcalcifications: A multi-scale approach, 2021, arXiv preprint arXiv:2102.00754.

[104] L. Shen, R.M. Rangayyan, J.L. Desautels, Application of shape analysis to mammographic calcifications, IEEE Trans. Med. Imaging 13 (2) (1994) 263–274.

[105] Z. Chen, H. Strange, A. Oliver, E.R. Denton, C. Boggis, R. Zwiggelaar, Topological modeling and classification of mammographic microcalcification clusters, IEEE Trans. Biomed. Eng. 62 (4) (2014) 1203–1214.

[106] M. George, Z. Chen, R. Zwiggelaar, Multiscale connected chain topological modelling for microcalcification classification, Comput. Biol. Med. 114 (2019) 103422.
[107] S. Sannasi Chakravarthy, H. Rajaguru, Detection and classification of microcalcification from digital mammograms with firefly algorithm, extreme learning machine and non-linear regression models: A comparison, Int. J. Imaging Syst. Technol. 30 (1) (2020) 126–146.
[108] N. Alam, E. R.E. Denton, R. Zwiggelaar, Classification of microcalcification clusters in digital mammograms using a stack generalization based classifier, J. Imaging 5 (9) (2019) 76.
[109] Y. Jiménez-Gaona, M.J. Rodríguez-Álvarez, V. Lakshminarayanan, Deep-learning-based computer-aided systems for breast cancer imaging: A critical review, Appl. Sci. 10 (22) (2020) 8298.
[110] A.J. Bekker, H. Greenspan, J. Goldberger, A multi-view deep learning architecture for classification of breast microcalcifications, in: 2016 IEEE 13th International Symposium on Biomedical Imaging, ISBI, IEEE, 2016, pp. 726–730.
[111] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, L. Li, Discrimination of breast cancer with microcalcifications on mammography by deep learning, Sci. Rep. 6 (1) (2016) 1–9.
[112] K.U. Rehman, J. Li, Y. Pei, A. Yasin, S. Ali, T. Mahmood, Computer vision-based microcalcification detection in digital mammograms using fully connected depthwise separable convolutional neural network, Sensors 21 (14) (2021) 4854.
[113] S. Zahoor, I.U. Lali, M.A. Khan, K. Javed, W. Mehmood, Breast cancer detection and classification using traditional computer vision techniques: A comprehensive review, Curr. Med. Imaging (2020).
[114] S.-M. Lai, X. Li, W. Biscof, On techniques for detecting circumscribed masses in mammograms, IEEE Trans. Med. Imaging 8 (4) (1989) 377–386.
[115] J. Chu, H. Min, L. Liu, W. Lu, A novel computer aided breast mass detection scheme based on morphological enhancement and SLIC superpixel segmentation, Med. Phys. 42 (7) (2015) 3859–3869.
[116] J. de Nazaré Silva, A.O. de Carvalho Filho, A.C. Silva, A.C. De Paiva, M. Gattass, Automatic detection of masses in mammograms using quality threshold clustering, correlogram function, and SVM, J. Digit. Imaging 28 (3) (2015) 323–337.
[117] W.B. De Sampaio, A.C. Silva, A.C. de Paiva, M. Gattass, Detection of masses in mammograms with adaption to breast density using genetic algorithm, phylogenetic trees, LBP and SVM, Expert Syst. Appl. 42 (22) (2015) 8911–8928.
[118] F. Mohanty, S. Rup, B. Dash, B. Majhi, M. Swamy, A computer-aided diagnosis system using Tchebichef features and improved grey wolf optimized extreme learning machine, Appl. Intell. 49 (3) (2019) 983–1001.
[119] S.-C. Tai, Z.-S. Chen, W.-T. Tsai, An automatic mass detection system in mammograms based on complex texture features, IEEE J. Biomed. Health Inf. 18 (2) (2013) 618–627.
[120] S. Dhahbi, W. Barhoumi, E. Zagrouba, Breast cancer diagnosis in digitized mammograms using curvelet moments, Comput. Biol. Med. 64 (2015) 79–90.
[121] J. Chakraborty, A. Midya, R. Rabidas, Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns, Expert Syst. Appl. 99 (2018) 168–179.
[122] M.M. Eltoukhy, M. Elhoseny, K.M. Hosny, A.K. Singh, Computer aided detection of mammographic mass using exact Gaussian–Hermite moments, J. Ambient Intell. Humaniz. Comput. (2018) 1–9.
[123] S. Dhahbi, W. Barhoumi, J. Kurek, B. Swiderski, M. Kruk, E. Zagrouba, False-positive reduction in computer-aided mass detection using mammographic texture analysis and classification, Comput. Methods Programs Biomed. 160 (2018) 75–83.
[124] N. Dhungel, G. Carneiro, A.P. Bradley, Automated mass detection in mammograms using cascaded deep learning and random forests, in: 2015 International Conference on Digital Image Computing: Techniques and Applications, DICTA, IEEE, 2015, pp. 1–8.
[125] D. Ribli, A. Horváth, Z. Unger, P. Pollner, I. Csabai, Detecting and classifying lesions in mammograms with deep learning, Sci. Rep. 8 (1) (2018) 1–7.
[126] N. Dhungel, G. Carneiro, A.P. Bradley, A deep learning approach for the analysis of masses in mammograms with minimal user intervention, Med. Image Anal. 37 (2017) 114–128.
[127] H. Jung, B. Kim, I. Lee, M. Yoo, J. Lee, S. Ham, O. Woo, J. Kang, Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network, PLoS One 13 (9) (2018) e0203355.
[128] F. Gao, T. Wu, J. Li, B. Zheng, L. Ruan, D. Shang, B. Patel, SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis, Comput. Med. Imaging Graph. 70 (2018) 53–62.
[129] L. Shen, L.R. Margolies, J.H. Rothstein, E. Fluder, R. McBride, W. Sieh, Deep learning to improve breast cancer detection on screening mammography, Sci. Rep. 9 (1) (2019) 1–12.
[130] F.A. Zeiser, C.A. da Costa, T. Zonta, N. Marques, A.V. Roehe, M. Moreno, R. da Rosa Righi, Segmentation of masses on mammograms using data augmentation and deep learning, J. Digit. Imaging 33 (4) (2020) 858–868.
[131] G.H. Aly, M. Marey, S.A. El-Sayed, M.F. Tolba, YOLO based breast masses detection and classification in full-field digital mammograms, Comput. Methods Programs Biomed. (2020) 105823.
[132] S.V. da Rocha, G.B. Junior, A.C. Silva, A.C. de Paiva, M. Gattass, Texture analysis of masses malignant in mammograms images using a combined approach of diversity index and local binary patterns distribution, Expert Syst. Appl. 66 (2016) 7–19.
[133] T.F. de Brito Silva, A.C. de Paiva, A.C. Silva, G. Braz Júnior, J.D.S. de Almeida, Classification of breast masses in mammograms using geometric and topological feature maps and shape distribution, Res. Biomed. Eng. 36 (2020) 225–235.
[134] W.K. Lim, M.J. Er, Classification of mammographic masses using generalized dynamic fuzzy neural networks, Med. Phys. 31 (5) (2004) 1288–1295.
[135] W. Xie, Y. Li, Y. Ma, Breast mass classification in digital mammography based on extreme learning machine, Neurocomputing 173 (2016) 930–941.
[136] J.Y. Choi, D.H. Kim, K.N. Plataniotis, Y.M. Ro, Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography, Expert Syst. Appl. 46 (2016) 106–121.
[137] O.R. Seryasat, J. Haddadnia, Evaluation of a new ensemble learning framework for mass classification in mammograms, Clin. Breast Cancer 18 (3) (2018) e407–e420.
[138] J. Arevalo, F.A. González, R. Ramos-Pollán, J.L. Oliveira, M.A.G. Lopez, Convolutional neural networks for mammography mass lesion classification, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2015, pp. 797–800.
[139] Z. Jiao, X. Gao, Y. Wang, J. Li, A deep feature based framework for breast masses classification, Neurocomputing 197 (2016) 221–231.
[140] A.M. Abdel-Zaher, A.M. Eldeib, Breast cancer classification using deep belief networks, Expert Syst. Appl. 46 (2016) 139–144.
[141] V.S. Gnanasekaran, S. Joypaul, P.M. Sundaram, D.D. Chairman, Deep learning algorithm for breast masses classification in mammograms, IET Image Process. 14 (12) (2020) 2860–2868.
[142] A.B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115.
[143] A.J. Barnett, F.R. Schwartz, C. Tao, C. Chen, Y. Ren, J.Y. Lo, C. Rudin, A case-based interpretable deep learning model for classification of mass lesions in digital mammography, Nat. Mach. Intell. 3 (12) (2021) 1061–1070.
[144] S. van Engeland, P. Snoeren, J. Hendriks, N. Karssemeijer, A comparison of methods for mammogram registration, IEEE Trans. Med. Imaging 22 (11) (2003) 1436–1444.
[145] N. Vujovic, D. Brzakovic, Establishing the correspondence between control points in pairs of mammographic images, IEEE Trans. Image Process. 6 (10) (1997) 1388–1399.
[146] S. Sanjay-Gopal, H.-P. Chan, T. Wilson, M. Helvie, N. Petrick, B. Sahiner, A regional registration technique for automated interval change analysis of breast lesions on mammograms, Med. Phys. 26 (12) (1999) 2669–2679.
[147] R. Marti, R. Zwiggelaar, C. Rubin, Automatic mammographic registration: towards the detection of abnormalities, in: S T1U Conference on Medical Image Understanding and Analysis, 2001, pp. 149–152.
[148] L. Hadjiiski, B. Sahiner, H.-P. Chan, N. Petrick, M.A. Helvie, M. Gurcan, Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign breast masses, Med. Phys. 28 (11) (2001) 2309–2317.
[149] P. Filev, L. Hadjiiski, H.-P. Chan, B. Sahiner, J. Ge, M.A. Helvie, M. Roubidoux, C. Zhou, Automated regional registration and characterization of corresponding microcalcification clusters on temporal pairs of mammograms for interval change analysis, Med. Phys. 35 (12) (2008) 5340–5350.
[150] F. Ma, L. Yu, M. Bajger, M.J. Bottema, Incorporation of fuzzy spatial relation in temporal mammogram registration, Fuzzy Sets and Systems 279 (2015) 87–100.
[151] M.A. Wirth, J. Narhan, D.W. Gray, Nonrigid mammogram registration using mutual information, in: Medical Imaging 2002: Image Processing, Vol. 4684, International Society for Optics and Photonics, 2002, pp. 562–573.
[152] S. Timp, N. Karssemeijer, Interval change analysis to improve computer aided detection in mammography, Med. Image Anal. 10 (1) (2006) 82–95.
[153] M. Abdel-Nasser, A. Moreno, D. Puig, Temporal mammogram image registration using optimized curvilinear coordinates, Comput. Methods Programs Biomed. 127 (2016) 1–14.
[154] C. Li, Z. Chen, H. Zhang, Multi-individual mammographic image registration based on global-local integrated transformations, in: 2018 IEEE 9th International Conference on Software Engineering and Service Science, ICSESS, IEEE, 2018, pp. 983–986.
[155] M.K. Sharma, M. Jas, V. Karale, A. Sadhu, S. Mukhopadhyay, Mammogram segmentation using multi-atlas deformable registration, bioRxiv (2019) 542217.
[156] K. Mendel, H. Li, N. Tayob, R. El-Zein, I. Bedrosian, M. Giger, Temporal mammographic registration for evaluation of architecture changes in cancer risk assessment, in: Medical Imaging 2019: Computer-Aided Diagnosis, Vol. 10950, International Society for Optics and Photonics, 2019, 1095041.
[157] F. Ma, M. Bajger, S. Williams, M.J. Bottema, Improved detection of cancer in screening mammograms by temporal comparison, in: International Workshop on Digital Mammography, Springer, 2010, pp. 752–759.
[158] C. Varela, N. Karssemeijer, J.H. Hendriks, R. Holland, Use of prior mammograms in the classification of benign and malignant masses, Eur. J. Radiol. 56 (2) (2005) 248–255.

[159] K. Marias, C. Behrenbruch, S. Parbhoo, A. Seifalian, M. Brady, A registration framework for the comparison of mammogram sequences, IEEE Trans. Med. Imaging 24 (6) (2005) 782–790.
[160] F.P. Oliveira, J.M.R. Tavares, Medical image registration: a review, Comput. Methods Biomech. Biomed. Eng. 17 (2) (2014) 73–93.
[161] Y. Guo, R. Sivaramakrishna, C.-C. Lu, J.S. Suri, S. Laxminarayan, Breast image registration techniques: a survey, Med. Biol. Eng. Comput. 44 (1–2) (2006) 15–26.
[162] S. Timp, C. Varela, N. Karssemeijer, Temporal change analysis for characterization of mass lesions in mammography, IEEE Trans. Med. Imaging 26 (7) (2007) 945–953.
[163] S. Timp, C. Varela, N. Karssemeijer, Computer-aided diagnosis with temporal analysis to improve radiologists' interpretation of mammographic mass lesions, IEEE Trans. Inf. Technol. Biomed. 14 (3) (2010) 803–808.
[164] J. Bozek, M. Kallenberg, M. Grgic, N. Karssemeijer, Use of volumetric features for temporal comparison of mass lesions in full field digital mammograms, Med. Phys. 41 (2) (2014).
[165] F. Ma, L. Yu, G. Liu, Q. Niu, Computer aided mass detection in mammography with temporal change analysis, Comput. Sci. Inf. Syst. 12 (4) (2015) 1255–1272.
[166] T. Kooi, N. Karssemeijer, Classifying symmetrical differences and temporal change for the detection of malignant masses in mammography using deep neural networks, J. Med. Imaging 4 (4) (2017) 044501.
[167] Y. Zheng, C. Yang, A. Merkulov, Breast cancer screening using convolutional neural network and follow-up digital mammography, in: Computational Imaging III, Vol. 10669, International Society for Optics and Photonics, 2018, 1066905.
[168] L.M. Hadjiiski, H.-P. Chan, B. Sahiner, N. Petrick, M.A. Helvie, M.A. Roubidoux, M.N. Gurcan, Computer-aided characterization of malignant and benign microcalcification clusters based on the analysis of temporal change of mammographic features, in: Medical Imaging 2002: Image Processing, Vol. 4684, SPIE, 2002, pp. 749–753.
[169] Breast micro-calcifications dataset with precisely annotated sequential mammograms | zenodo, 2021, http://dx.doi.org/10.5281/zenodo.5036062. (Accessed 20 September 2022).
[170] Breast masses dataset with precisely annotated sequential mammograms | zenodo, 2022, http://dx.doi.org/10.5281/zenodo.7179856. (Accessed 13 October 2022).
[171] Food and drug administration. Saige-Q, 2021, https://www.accessdata.fda.gov/cdrh_docs/pdf20/K203517.pdf. (Accessed 22 February 2022).
[172] Welcome to screenpoint medical - home, 2021, https://screenpoint-medical.com/. (Accessed 20 January 2021).
[173] Food and drug administration. Transpara, 2018, https://www.accessdata.fda.gov/cdrh_docs/pdf18/K181704.pdf. (Accessed 20 January 2021).
[174] Food and drug administration. cmTriage, 2019, https://www.accessdata.fda.gov/cdrh_docs/pdf18/K183285.pdf. (Accessed 22 February 2022).
[175] Food and drug administration. HealthMammo, 2020, https://www.accessdata.fda.gov/cdrh_docs/pdf20/K200905.pdf. (Accessed 22 February 2022).
[176] Food and drug administration. MammoScreen, 2020, https://www.accessdata.fda.gov/cdrh_docs/pdf19/K192854.pdf. (Accessed 22 February 2022).
[177] Food and drug administration. Genius AI detection, 2020, https://www.accessdata.fda.gov/cdrh_docs/pdf20/K201019.pdf. (Accessed 22 February 2022).
[178] Food and drug administration. ProFound AI software V2.1, 2019, https://www.accessdata.fda.gov/cdrh_docs/pdf19/K191994.pdf. (Accessed 22 February 2022).
[179] Food and drug administration. 510(k) premarket notification, 2021, https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K203822. (Accessed 22 February 2022).
[180] Food and drug administration. Transpara 1.7.0, 2021, https://www.accessdata.fda.gov/cdrh_docs/pdf21/K210404.pdf. (Accessed 22 February 2022).
[181] Food and drug administration. Lunit INSIGHT MMG, 2021, https://www.accessdata.fda.gov/cdrh_docs/pdf21/K211678.pdf. (Accessed 22 February 2022).
[182] A. Rodríguez-Ruiz, E. Krupinski, J.-J. Mordang, K. Schilling, S.H. Heywang-Köbrunner, I. Sechopoulos, R.M. Mann, Detection of breast cancer with mammography: effect of an artificial intelligence support system, Radiology 290 (2) (2019) 305–314.
[183] A. Rodriguez-Ruiz, K. Lång, A. Gubern-Merida, M. Broeders, G. Gennaro, P. Clauser, T.H. Helbich, M. Chevalier, T. Tan, T. Mertelmeier, et al., Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists, J. Natl. Cancer Inst. 111 (9) (2019) 916–922.
[184] M. Sasaki, M. Tozaki, A. Rodríguez-Ruiz, D. Yotsumoto, Y. Ichiki, A. Terawaki, S. Oosako, Y. Sagara, Y. Sagara, Artificial intelligence for breast cancer detection in mammography: experience of use of the ScreenPoint medical transpara system in 310 Japanese women, Breast Cancer 27 (4) (2020) 642–651.
[185] J.L. Raya-Povedano, S. Romero-Martín, E. Elías-Cabot, A. Gubern-Mérida, A. Rodríguez-Ruiz, M. Álvarez-Benito, AI-based strategies to reduce workload in breast cancer screening with mammography and tomosynthesis: A retrospective evaluation, Radiology (2021) 203555.
[186] S.L. van Winkel, A. Rodríguez-Ruiz, L. Appelman, A. Gubern-Mérida, N. Karssemeijer, J. Teuwen, A.J. Wanders, I. Sechopoulos, R.M. Mann, Impact of artificial intelligence support on accuracy and reading time in breast tomosynthesis image interpretation: a multi-reader multi-case study, Eur. J. Radiol. 31 (11) (2021) 8682–8691.
[187] M.H. Pub, K. Bowyer, D. Kopans, R. Moore, P. Kegelmeyer, The digital database for screening mammography, in: Third International Workshop on Digital Mammography, Vol. 58, 1996, p. 27.
[188] CBIS-DDSM - the cancer imaging archive (TCIA) public access - cancer imaging archive wiki, 2021, https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM. (Accessed 27 January 2022).
[189] DoD BCRP spiculated mass detection evaluation data, 2022, http://www.eng.usf.edu/cvprg/mammography/DDSM/BCRP/bcrp_mass_01.html. (Accessed 27 January 2022).
[190] P. Suckling J, The mammographic image analysis society digital mammogram database, Digit. Mammo (1994) 375–386.
[191] I.C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M.J. Cardoso, J.S. Cardoso, Inbreast: toward a full-field digital mammographic database, Academic Radiol. 19 (2) (2012) 236–248.
[192] M.G. Lopez, N. Posada, D.C. Moura, R.R. Pollán, J.M.F. Valiente, C.S. Ortega, M. Solar, G. Diaz-Herrero, I. Ramos, J. Loureiro, et al., BCDR: a breast cancer digital repository, in: 15th International Conference on Experimental Mechanics, Vol. 1215, 2012.
[193] M.D. Halling-Brown, L.M. Warren, D. Ward, E. Lewis, A. Mackenzie, M.G. Wallis, L.S. Wilkinson, R.M. Given-Wilson, R. McAvinchey, K.C. Young, OPTIMAM mammography image database: a large-scale resource of mammography images and clinical data, Radiol. Artif. Intell. 3 (1) (2020) e200103.