Review Articles

Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review

Afsaneh Jalalian,⁎ Syamsiah B.T. Mashohor, Hajjah Rozi Mahmud, M. Iqbal B. Saripana, Abdul Rahman B. Ramli, Babak Karasfi

Abstract

Breast cancer is the most common form of cancer among women worldwide. Early detection of breast cancer can increase treatment options and patients’ survivability. Mammography is the gold standard for breast imaging and cancer detection. However, due to some limitations of this modality such as low sensitivity especially in dense breasts, other modalities like ultrasound and magnetic resonance imaging are often suggested to achieve additional information. Recently, computer-aided detection or diagnosis (CAD) systems have been developed to help radiologists in order to increase diagnosis accuracy. Generally, a CAD system consists of four stages: (a) preprocessing, (b) segmentation of regions of interest, (c) feature extraction and selection, and finally (d) classification. This paper presents the approaches which are applied to develop CAD systems on mammography and ultrasound images. The performance evaluation metrics of CAD systems are also reviewed.

Keywords: Computer-aided detection; Computer-aided diagnosis; Breast cancer; Mammography; Ultrasound

1. Introduction

Breast cancer is the most common form of cancer among women worldwide. Early detection of breast cancer increases treatment options and patients’ survivability [1]. Although mammography is currently the most effective tool for early detection of breast cancer, it has some restrictions. On a screening mammographic examination, noncancerous lesions can be misinterpreted as a cancer (false-positive value), while cancers may be missed (false-negative value). As a result, radiologists fail to detect 10% to 30% of breast cancers [2–4]. The false-positive value indicates the percentage of lesions that were found to be cancerous and subjected to biopsy. The miss rate in mammography is increased in dense breasts where the probability of cancer is four to six times higher than in nondense breasts. In order to enhance sensitivity of mammography, complimentary modalities such as ultrasound and magnetic resonance imaging (MRI) are often recommended to achieve additional information. Recently, computer-aided detection/diagnosis (CAD) systems have been developed to reduce the expense and to improve the capability of radiologist in interpretation of medical images and differentiation between benign and malignant tissues [5–8]. The efficiency of radiologist’s interpretation can be improved in terms of accuracy and consistency in detection/diagnosis, while his/her productivity can be improved by reducing the time required for reading the images [9]. The computer outputs are derived using various techniques in computer vision to present some of the significant parameters such as the location of suspicious lesions and the likelihood of malignancy of detected lesions. Generally, CAD systems are executable on all imaging modalities and all kinds of examinations.

⁎ Corresponding author.

E-mail address: jalalian_afs@yahoo.com (A. Jalalian).
Generally, CAD systems are classified into two categories: computer-aided detection (CADe) and computer-aided diagnosis (CADx) systems. The CADe systems are developed to help the radiologist in detecting and locating the abnormal area in images, while the CADx systems are designed to diagnose and classify benign or malignant tissues. Evidence shows that CADx systems may be effective in improving the radiologist’s interpretation in false-positive value \([10–13]\). Many different types of CAD systems are produced to detect/diagnose different lesions in medical imaging, including ultrasound, computed tomography, and MRI. The organs that are mostly studied by CAD include lung \([14–16]\), colon \([17]\), and breast \([18–24]\), but other organs such as liver, brain, and vascular systems recently have also been investigated through this method \([25]\). Most of these studies report acceptable results in detection of suspicious lesions, and it is expected that, in the future, many CAD systems will be developed for clinical detection/diagnosis of cancer in different modalities.

This article reviews some of the most recent advances in breast cancer detection/diagnosis using CAD systems developed for mammography and ultrasound.

### 2. Computer-aided detection/diagnosis system in mammography

In the analysis of mammograms, detection and diagnosis of breast cancer are extremely challenging tasks due to complexities like variability in appearance of abnormalities and hiding abnormal tissues in dense breasts. As cost-effective tools, computer-aided detection/diagnosis techniques can aid the radiologist by reducing interpretation error. Used as double readers, CAD systems can increase the accuracy of radiologists’ final decision. The goal of the computer-aided detection schemes is to help radiologists to avoid overlooking abnormal features that are not visible on the screening image \([18,20,26–28]\).

Automated methods for mammographic analysis are classified in two categories:

- CADe systems, which present a computerized detection of abnormalities in mammogram images
- CADx systems, which assist the radiologist to determine the grade of abnormalities as benign or malignant

#### 2.1. CADe systems

The aim of CADe systems is to help radiologists to detect and locate abnormalities in breast screening images. The first step of CADe systems is detection of suspicious regions. The most common algorithms to identify the regions of interest (ROIs) are pixel-based or region-based methods \([29]\). The main advantage of pixel-based methods is their simple implementation, while their significant drawback is their computationally intensive process. In region-based detection techniques, ROIs are extracted by segmentation techniques. Since region-based methods consider morphology and size of masses, they have lower computing complexity than pixel-based methods \([29]\).

Some of the most frequent indications of breast cancer on mammograms are masses and microcalcification \([30]\). The most important stages of mass detection algorithms include detection of suspicious regions and classification of suspicious area as normal tissues or masses \([30]\). Masses are described by their shape and margin characteristics. The spiculated masses are the particular kind of masses that have high probability of malignancy. In such cases, calculation of edge orientation at each pixel is the most ordinary technique for finding spicule radiate at all directions. Many different ways are available to compute edge orientation such as statistical analysis of a map of pixel orientation \([31]\) and feature extraction at a multiresolution representation using wavelet transform \([32]\). Multiple threshold value is another algorithm which has been developed for detection of masses \([33]\). Researchers \([34,35]\) have developed region-based methods for mass detection on particular margin characteristics.

The second step of mass detection algorithms is classification of the suspicious region as normal tissues or masses. Radiologists look for significant characteristics in breast images to discriminate between masses and normal tissues. Researchers follow this procedure to develop a classifier to differentiate masses from normal tissues. Te Barke et al. \([36]\) identified some characteristics such as contrast, intensity, and location to distinguish between normal tissues and masses. A number of methods have been developed for this task based on the template-matching technique \([37]\) or neural network \([38]\).

Microcalcification is another significant symptom of breast cancer that radiologists look for in mammograms. Microcalcifications are tiny calcium deposits which appear as opacities with a different appearance in mammograms. The main characteristics of microcalcification are size, shape, or morphology and number of distribution. Small calcifications may be missed when they are covered by fundamental tissues of breast. Location of calcification in a region with a dense background is difficult. Calcifications depict high spatial resolutions in mammograms. Thus, techniques based on wavelet transform are powerful tools for locating the high spatial frequencies. A number of effective methods for microcalcification detection have been presented based on wavelet transform \([30,39–42]\). Other non-wavelet-based techniques, such as local area thresholding, have been applied for calcification detection \([43]\). Nishikawa et al. \([44]\) developed a method for calcification detection by combining gray-level thresholding technique and morphology erosion filter.
2.2. CADx systems

CADx systems characterize suspicious lesions to reduce the number of biopsy recommendations on benign lesions. Computer vision and artificial intelligent techniques are used to characterize an ROI as benign or malignant. To create a CADx system, the integration of various image processing operations, such as image segmentation, feature extraction, feature selection, and classification, is essential.

Segmentation is the foundation of a CADx system. Segmentation is the most rigorous stage in the computer-aided diagnosis of calcification due to small size of microcalcification. The two major categories of segmentation methods are region-growing and discrete contour models [45]. An integration of the region-growing segmentation and edge-based segmentation techniques is provided for the ROI detection in the images [46,47]. A fuzzy region-growing method has been proposed for segmentation and classification of masses based on transition information surrounding the segmented region [48]. Two extended region-growing methods based on the radial gradient and simple probabilistic models are presented for segmentation of masses [49].

Since the likelihood of malignancy depends on the shape and margin of lesions, the diagnosis tasks are designed based on extracting corresponding feature to these characteristics of masses and calcifications. These features can be categorized into texture features and morphologic features. Accurate segmentation technique has a consequential role in diagnosis algorithms which use morphologic features. Research results [50] indicate that morphology is one of the most significant clinical aspects in calcification diagnosis. Texture features are effective in discrimination between benign and malignant lesions [30]. Some common cluster features of microcalcification include standard deviation of their contrast, number of microcalcification per unit, and mean diameter of microcalcification [30].

The high numbers of features increase the computational cost and slow down the classification process. Feature selection techniques reduce the number of feature space for developing process accuracy and minimizing the computation time by eliminating redundant, irrelevant, and noisy features [51]. Feature selections are generally performed by searching algorithms such as sequential forward selection, sequential backward selection, particle swarm optimization, and genetic algorithm [51]. In some cases, a combination of search methods is used for feature selection procedure.

Classification is a machine learning technique to analyze the spatial features and organize them into desirable categories. The final section in the CAD system is the classification stage that is regarded as the heart of the method. In this stage, the ROI identification data are categorized into predefined classes which are usually considered a two-class scenario or binary classification that is commonly labeled as positive or negative classes. K-nearest neighbor (KNN) classifier is a classical approach to classify objects based on a training sample in the feature space. The similarity from previous training pattern is applied in KNN to classify the new test data [52]. Support vector machine (SVM) is one of the most powerful supervised learning that utilizes a structural risk minimization to diminish error of learning machine [53]. An SVM-based method combined with feature selection has been developed for breast cancer diagnosis [54]. SVM framework has been presented for automated detection of microcalcification in mammogram [55]. A recursive feature elimination in the framework of SVM has been developed for optical diagnosis of cancer [56].

Artificial neural network (ANN) techniques are commonly known as powerful tools inspired by human perception which are capable of modeling complex non-linear functions [57]. A prediction framework for breast cancer diagnosis based on evolutionary ANN is available in the related literature [58]. A neural network model for recognition of medical image patterns has been developed for detection of lung and breast cancer in radiography [59].

3. Computer-aided detection and diagnosis system in ultrasound

Mammography is the most effective modality in detection and diagnosis of breast cancer. However, low specificity in screening mammography may cause some unnecessary biopsy [60]. This restriction increases the cost and stress imposed on the patient. In addition, ionizing radiation of mammography endangers the patient’s health.

Presently, ultrasound imaging is one of the most effective tools as an adjunct to mammography to detect and diagnose abnormalities in the breast. Studies show that ultrasound is able to detect and discriminate benign and malignant masses with high accuracy and reduce the number of unnecessary biopsy [61–63]. Ultrasound is more sensitive for detecting invasive cancer in dense breasts [64,65]. However, it is an operator-dependent modality, and the interpretation of its images requires expertise in the part of the radiologist. In order to overcome the operator dependency and increase accurate diagnosis rate, computer-aided detection/diagnosis systems are developed for breast cancer detection and classification. Recently, several CAD systems have been proposed to reduce the influence of dependence on operator in ultrasound and increase the diagnosis sensitivity and specificity [66,67]. Many techniques such as SVM and ANN have been proposed [62,68,69] for mass detection and diagnosis.

Generally, ultrasound CAD systems for breast cancer detection and diagnosis cover four stages: (a) image processing, (b) image segmentation, (c) feature extraction and selection, and finally (d) classification. Speckle interference and low contrast are the main restrictions of ultrasound imaging [70]. Image processing techniques are involved to enhance the image and suppress speckle in the first step of ultrasound CAD systems. Speckle is a type of multiplicative noise which can make it difficult to observe
and interpret the ultrasound images. Speckle noise reduction techniques are categorized into three groups: filtering methods [71,72], wavelet domain methods [73–79], and compound approaches [80–82].

Histogram thresholding is widely used for segmentation of breast ultrasound [69,83–85]. The active contour model is a framework known as snake [86] which is applied as an edge segmentation technique. The snake model has been wildly utilized for ultrasound image segmentation [87–91]. The active contour model is used for segmentation of breast tumor on three-dimensional ultrasonic data [87,88]. Neural network is one of the popular techniques in breast segmentation of ultrasound images [92,93]. A compound method based on neural network technique and wavelet analysis has also been proposed for ultrasound image segmentation [92].

After image segmentation, feature extraction and selection are the next steps taken to reduce the volume of data processed. Features are characteristics of ROIs which will help to achieve the best result in the subsequent stage. The features of breast ultrasound images can be classified into four categories: texture, morphology, descriptor, and model-based features [94]. Texture features are calculated from ROI or whole image. Textural features have been applied in several studies [92,95,96] to discriminate benign and malignant lesions. Morphological features focus on some characteristics such as shape and margin. A morphological feature extraction technique is developed to detect the cancerous lesion in digital images [97]. Model-based feature is a specific form of ultrasound features that emphasizes on the backscattered echo from breast tissue [94]. Descriptor features are types of features that are based on the empirical classification criteria of the radiologist [94]. Most of the descriptor features can be found in the Breast Imaging Reporting and Data System [98].

The last step in CAD systems is the classification of the suspicious lesions into benign/malignant categories. ANN techniques are commonly known as powerful tools inspired by human perception which are capable of modeling complex nonlinear functions [57]. In the field of breast cancer detection and classification, ANN techniques are categorized as back propagation neural network, self-organizing map, and hierarchical neural network [62,68,69,92].

SVM is one of the most powerful supervised learning methods that utilize a structural risk minimization to diminish error of learning machine [53]. SVM aims at detecting the optimum hyper-plane in the input feature space that maximizes the distance from the maximal margin hyper-plane. In the field of breast cancer diagnosis, SVM is applied to classify benign and malignant lesions [70,99–101].

### 4. Evaluation of CAD systems

The performance of detection or diagnosis algorithms is reported as sensitivity, size of lesions, tissue density, histopathology of lesions, and the number of false-positive values per image [26,27,102–106]. Generally, receiver operating characteristics (ROC) are used to demonstrate the performance of the CAD system [107]. An ROC curve is a plot of true-positive value as a function of false-positive value. The area under the ROC curve is defined as the evaluation criteria [83]. The CAD sensitivity reported for cancer detection is over 90% [104], with higher sensitivity for detecting classification than architectural distortions or masses [102,104,106].Reportedly, the CAD system assists radiologists and increases detection sensitivity of breast cancer up to 20% [18,26]. Evaluation results of a number of CADx systems for diagnosis of mass and calcification are summarized in Tables 1 and 2. However, a fair comparison of different methods is extremely difficult as they are evaluated on various databases. The performance of detection algorithms is reported as two metrics; sensitivity (Eq. 1) and the number of false-positive values per image (Eq. 2).

\[
\text{Sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}}
\]

\[
\text{Specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}}
\]

CAD systems increase the radiologist’s accuracy and efficiency. Due to intrinsic limitations, in conventional mammography, the malignant tissues may be hidden particularly in dense breasts. If the information is inadequate to make a decision, other modalities such as ultrasound or MRI are suggested to the patient to achieve additional information. The performance of some CAD systems in ultrasound and databases used are listed in Table 3.

<table>
<thead>
<tr>
<th>Authors</th>
<th>No. image</th>
<th>Diagnosis results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinoshita et al. [108]</td>
<td>92</td>
<td>81</td>
</tr>
<tr>
<td>Sahiner et al. [109]</td>
<td>168</td>
<td>0.94</td>
</tr>
<tr>
<td>Hadjiiski et al. [110]</td>
<td>348</td>
<td>0.81</td>
</tr>
<tr>
<td>Rangayan et al. [111]</td>
<td>39</td>
<td>95</td>
</tr>
<tr>
<td>Kallergi [50]</td>
<td>60</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 1

<table>
<thead>
<tr>
<th>Authors</th>
<th>No. image</th>
<th>Diagnosis results</th>
</tr>
</thead>
<tbody>
<tr>
<td>De Santo et al. [113]</td>
<td>192</td>
<td>75.7 73.5 0.79</td>
</tr>
<tr>
<td>Tsujii et al. [114]</td>
<td>128</td>
<td>0.76</td>
</tr>
<tr>
<td>Veldkamp et al. [115]</td>
<td>280</td>
<td>0.83</td>
</tr>
</tbody>
</table>

TPF is sensitivity as defined in Eq. (1), and FPF is specificity as defined in Eq. (2).
Table 3
Performance of some selected CAD systems on ultrasound images

<table>
<thead>
<tr>
<th>Reference</th>
<th>Description</th>
<th>Performance (ROC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang and Chen [99]</td>
<td>The data set consists of two sets: 1. 88 benign and 52 malignant lesions 2. 215 benign and 35 malignant lesion</td>
<td>Az=0.97</td>
</tr>
<tr>
<td>Joo et al. [69]</td>
<td>The data set consists of two sets: 1. 300 benign and 284 malignant lesions 2. 167 benign and 99 malignant lesions</td>
<td>Az=0.95</td>
</tr>
<tr>
<td>Mogatadakala et al. [116]</td>
<td>161 benign and 43 malignant lesions</td>
<td>Az=0.91</td>
</tr>
<tr>
<td>Alam et al. [117]</td>
<td>104 benign and 26 malignant lesions</td>
<td>Az=0.95</td>
</tr>
</tbody>
</table>

5. Conclusion

This paper reviewed the literature on the use of CAD systems for breast cancer detection and diagnosis in mammography and ultrasound. The main stages of CAD system include preprocessing, segmentation of ROI, feature extraction and selection, and finally classification. Different methods for covering these stages were introduced. The evaluation metrics were also reviewed for assessment of CAD systems on mammography and ultrasound images.

References


