Original Research   **Open Access**

Comparison Between Feature-Based and Convolutional Neural Network–Based Computer-Aided Diagnosis for Breast Cancer Classification in Digital Breast Tomosynthesis

Siwa Chan[*]
Doctor, Dept. of Medical Imaging, Taichung Tzu Chi Hospital, Taiwan

Jinn-Yi Yeh
Professor, Dept. of Management Information Systems, National Chiayi University, Taiwan

Abstract

Digital breast tomosynthesis (DBT) is a promising new technique for breast cancer diagnosis. DBT has the potential to overcome the tissue superimposition problems that occur on traditional mammograms for tumor detection. However, DBT generates numerous images, thereby creating a heavy workload for radiologists. Therefore, constructing an automatic computer-aided diagnosis (CAD) system for DBT image analysis is necessary. This study compared feature-based CAD and convolutional neural network (CNN)-based CAD for breast cancer classification from DBT images. The research methods included image preprocessing, candidate tumor identification, three-dimensional feature generation, classification, image cropping, augmentation, CNN model design, and deep learning. The precision rates (standard deviation) of the LeNet-based CNN CAD and the feature-based CAD for breast cancer classification were 89.84 (0.013) and 84.46 (0.082), respectively. The T value was -4.091 and the P value was $0.00 < 0.05$, which indicate that the LeNet-based CNN CAD significantly outperform the feature-based CAD. However, there is no significantly differences between the LeNet-based CNN CAD and the feature-based CAD on other criteria. The results can be applied to clinical medicine and assist radiologists in breast cancer identification.
Keywords: Breast cancer; Digital breast tomosynthesis; Computer-aided diagnosis; Tumor detection.

1. Introduction

Analysis of breast masses from mammograms relies on the experience of radiologists to determine the location and type of tumors. However, relying entirely on manual interpretation may cause some tumors to be missed. Therefore, numerous studies have developed computer-aided diagnosis (CAD) systems to identify potential tumors in images and assist physicians in determining whether they are benign or malignant. The processes consist of reading the mammograms and performing a series of image preprocessing procedures, such as contrast enhancement, noise removal, and partial image area removal. These procedures can enhance the contrast of mammograms and make tumor detection easy. The following step applies certain algorithms to circle suspicious areas of tumors called regions of interest (ROIs). The tumor features are then calculated, such as their gray scale, shape, smoothness, and patch size. The subsequent step is feature selection, which reduces dimensionality. The stepwise feature selection method and heuristic algorithm are commonly used. Next, only the features with superior classification effects are retained to enter the machine learning process. The learned model will be used to classify various types of tumor, such as benign and malignant and tumor or normal. This method is called feature-based CAD for breast cancer classification. Two disadvantages exist. (1) Confirming whether these features are the most effective combination for tumor classification is impossible; and, (2) manual procedures are not optimally designed for hand-crafted features [1].

A deep learning model can be used to solve the aforementioned problems. Its automatic learning mechanism can generate many features for classification. At the end of 2012, this model became highly popular when deep learning based on a convolutional neural network (CNN) won an overwhelming victory in a well-known computer vision contest (ImageNet LSVRC-2010 contest). The winners trained a large, deep CNN to classify 1.2 million high-resolution images in 1000 categories. In terms of test data, they were ranked first and fifth with error rates of 17.0% and 37.5%, respectively, which were superior to state-of-the-art technology. This neural network contained 60 million parameters, 650,000 neurons, and five convolution layers. Some of the convolution layers were followed by max-pooling layers and three fully connected layers with a 1000-way softmax. They used non-saturated neurons for training and highly efficient GPUs to implement the convolution operations. To reduce overfitting at the fully connected layer, the recently developed dropout regularization method proved to be highly effective. The team also entered a variant of the model in the ILSVRC-2012 competition and achieved first place with an error rate of 15.3% and a second error rate of 26.2% [2]. MIT Technology Review named it as one of the top ten breakthrough technologies in 2013. Since then, researchers in almost all areas have begun to conduct deep learning studies. Because deep learning directly uses convolutional operations for machine learning, this approach is called CNN-based CAD [3].

*Corresponding Author

The sensitivity of mammography is often limited by the presence of overlapping dense fibroglandular tissue in the breast. The overlapping tissues depicted in mammograms may create substantial obstacles to the detection and diagnosis of abnormalities. Questionable mammography screening results frequently cause patients to experience unnecessary anxiety and incur excessive medical costs. Fortunately, digital breast tomosynthesis (DBT) is an effective imaging tool for detecting breast cancer at early stages, and it is also the only screening modality that has been proven to reduce mortality. DBT has been approved by the United States Food and Drug Administration. DBT has the potential to overcome the tissue superimposition problems that occur in traditional mammograms used for mass detection. Furthermore, it offers several exciting opportunities, including the possibility of reduced breast compression, improved diagnostic and screening accuracy, fewer recalls, and three-dimensional (3D) lesion localization [4-6]. DBT is a 3D imaging technology that involves acquiring images of a stationary compressed breast at multiple angles during a short scan. The individual images are then reconstructed into a series of thin high-resolution slices that can be displayed individually or in a dynamic cinematic mode. Figure 1 shows the procedure through which DBT images are acquired and reconstructed into 3D images [5].

Fig-1. DBT image construction [5]
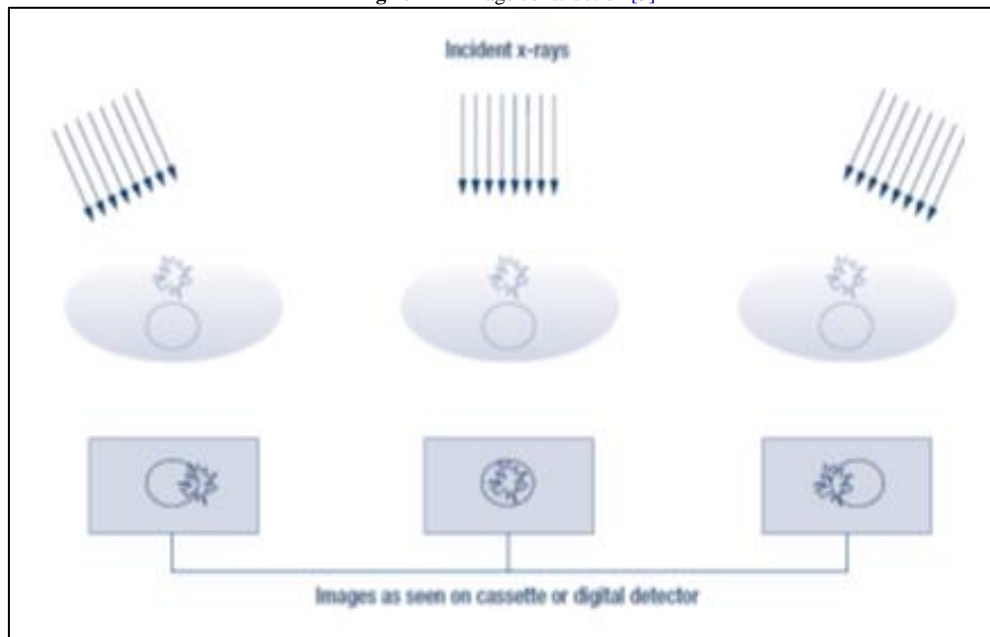

(a) DBT images acquired from different angles
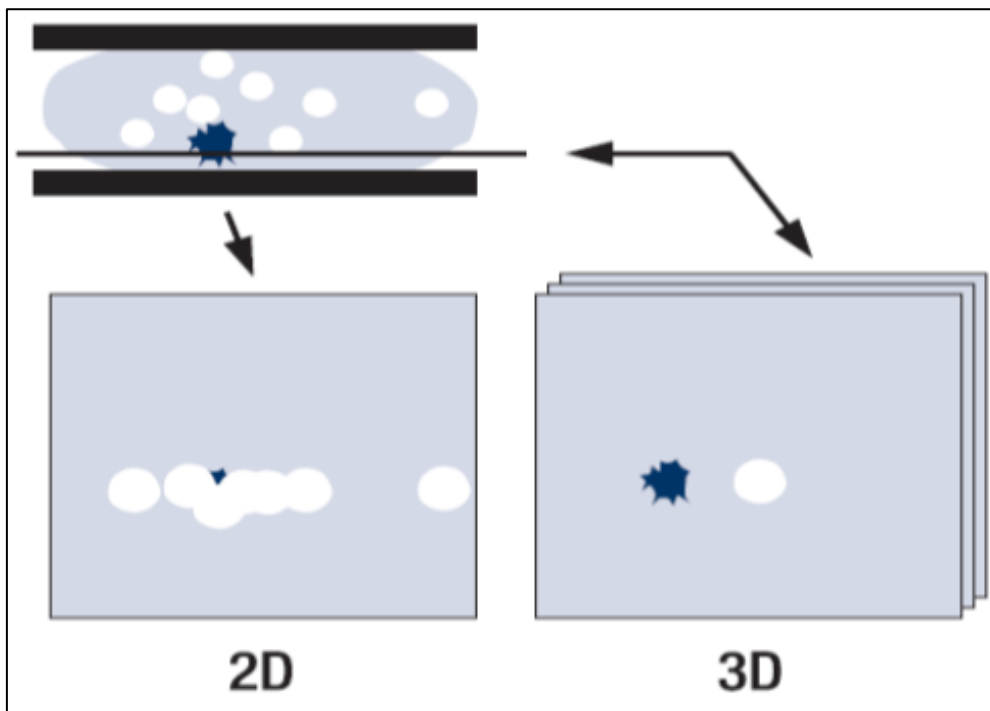
(b) DBT solves the overlap problem

DBT images can be obtained on a commercially available DBT unit (Hologic Dimensions). Images produced using this equipment have a two-dimensional (2D) pixel size of 70 μm and a reconstructed pixel size of approximately 100 μm. The typical exposure parameters are 27–35 kV_p using an X-ray; current DBT protocol is

approximately 42–78 mA$_s$. The DBT images require an additional 3.7 s per view. DBT typically acquires standard craniocaudal (CC) and mediolateral-oblique (MLO) views. Studies have shown that DBT decreases the false positive rate of mammography and may detect additional abnormalities not seen in 2D digital mammography (DM) [6]. Figure 2 presents an example from Taichung Veterans General Hospital (TVGH) in Taiwan. Figure 2 (a) shows four 2D DM images in which tumors are covered by chest and breast tissue. Figure 2 (b) depicts eight DBT slices in which tumors can be obviously identified.

Fig-2. DBT detects additional abnormalities that were not seen on 2D DMs (TVGH Case 18)

(a) 2D mammograms

(b) DBT images (from left to right) are slices 13, 15, 17, 19, 21, 23, 25, and 27
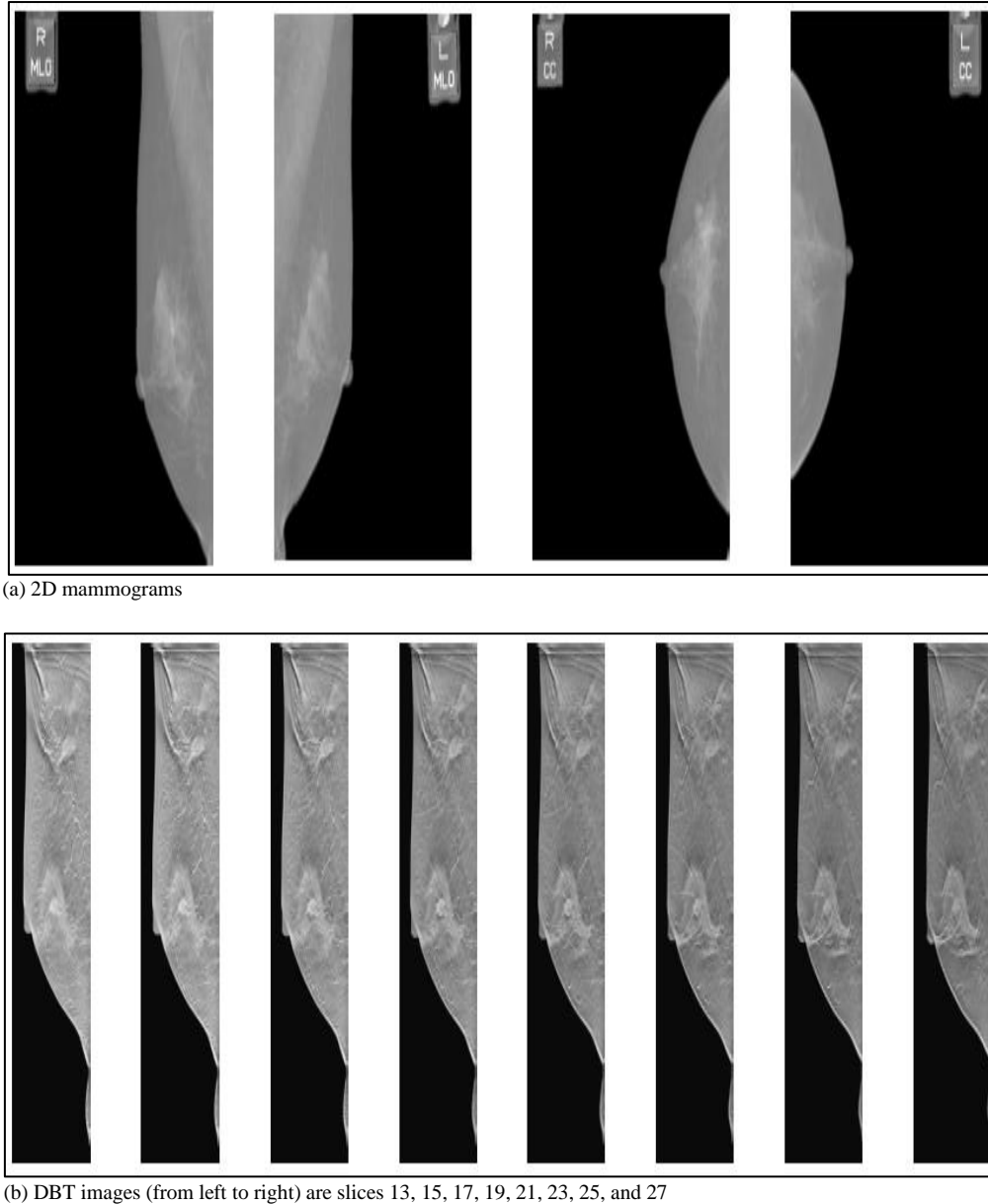
Although DBT is a promising new technique for breast cancer diagnosis, it generates numerous images, thereby creating a heavy workload for radiologists. Therefore, it is necessary to construct an automatic CAD system for DBT image analysis. This study compared feature-based CAD and CNN-based CAD for breast cancer classification in DBT images. The research methods included image preprocessing, candidate tumor identification, 3D feature generation, feature-based breast cancer classification, image cropping, augmentation, CNN model design, and deep learning. The remainder of this paper is organized as follows: Section 2 summarizes the literature review; Section 3 details the materials and methods used in the study; Section 4 presents an analysis of the experimental results; and finally, Section 5 reports the conclusions.

2. Literature Review
2.1. DBT-Related Work

In recent years, many researchers have worked on detecting breast abnormalities, but have encountered problems such as how to reduce false positives and improve the detection of abnormal tissues. In addition, DBT produces many images per scan, which require the radiologist to spend much time on inspection. Therefore, scholars have proposed breast detection methods to reduce the burden of radiologists. For example, Chan, *et al.* [7] designed a CAD system for breast mass detection in DBT images and performed a preliminary evaluation of the performance

of their system. It first screened the 3D volume of the mass candidates through gradient field analysis. Each mass candidate was segmented from the structured background, and its image features were extracted. A feature classifier was designed to differentiate true masses from normal tissues. The experimental results demonstrated the feasibility of the proposed approach for developing a CAD system for DBT images.

Schie, et al. [8], proposed a two-stage system for mass detection in reconstructed DBT volumes. In the first stage, an existing 2D CAD system was applied to slabs composed of multiple DBT slices after processing the slabs to a representation similar to that of the full-field DM training data. In the second stage, the scholars grouped detections obtained from the slabs that detected the same object and determined the 3D location of the grouped findings using one of three approaches, including one that used a set of features extracted from the DBT slabs. The study results indicated that a large benefit may exist to using 2D DB for the development of CAD for DBT and that exclusively limiting development to DBT data is not required. Kontos, et al. [9], investigated the potential advantages of DBT parenchymal texture analysis for estimating breast cancer risk. They compared the relative performance of DBT and mammogram texture features in correlation with two measures of breast cancer risk: (1) the Gail and Claus risk estimates and (2) mammographic breast density. Linear regression was performed to model the association between texture features and increasing levels of risk. Experimental results suggested that DBT parenchymal texture analysis could provide more accurate characterization of breast density patterns, which could ultimately improve the estimation of breast cancer risk.

Houssami and Skaane [10], examined the accuracy of DBT in clinical studies, which revealed some consistent findings that are summarized as follows. (1) Two-view DBT has at equal or superior accuracy to standard two-view DM, whereas one-view DBT does not have superior accuracy to standard DM. (2) The addition of DBT to standard mammography increases accuracy. (3) Improved accuracy from using DBT may be caused by increased cancer detection, reduced false positive recalls, or both. (4) Subjective interpretation of cancer conspicuity consistently found that cancers were equally or more conspicuous in DBT compared with DM. Palma, et al. [11] proposed a complete detection scheme for detecting masses and architectural distortions in DBT images. Two parts exist called channels, and each is dedicated to one type of lesion, which are then merged in a final decision step, thereby correctly handling the potential overlap between the two types of lesion. The first detection channel exploits the dense kernel nature of masses and the intrinsic imprecision of their attributes in a fuzzy approach. The second detection channel models the convergence characteristics of architectural distortions in an a contrario approach. The experimental results of 101 breasts, including 53 lesions, demonstrated the usefulness of the proposed approach, which led to high sensitivity with a reduced number of false positives, and compared favorably with existing approaches.

Qin, et al. [12], proposed an automatic method to classify fatty and glandular tissues in DBT images. First, the DBT images were preprocessed to enhance the tissue structures and decrease image noise and artifacts. Second, a global smooth filter was applied to eliminate detailed structures and enhance large-scale ones. Third, regions with similar structures were extracted and labeled using fuzzy C-means classification. Finally, each region was classified into various tissue types based on both intensity and texture features. Experimental results demonstrated the feasibility of the proposed method for classifying breast glandular and fatty tissues in DBT images. Morra, et al. [13], evaluated a commercial tomosynthesis CAD system in an independent, multicenter dataset. Their examinations were performed in three centers and included 123 patients, with 132 biopsy-proven screening-detected cancers, and 52 examinations with negative results at 1-year follow-up. In total, 111 lesions were masses or microcalcifications (72 masses, 22 microcalcifications, and 17 masses with microcalcifications), and 21 were architectural distortions. Lesions were annotated by radiologists who were aware of all available reports. CAD performance was assessed as per-lesion sensitivity and false positive results per volume in patients with negative results. Experimental results showed that the DBT CAD system allowed detection of 89% of breast cancers manifesting as masses and microcalcification clusters with an acceptable false positive rate.

Lu, et al. [14], developed CAD algorithms based on tomosynthesis images. They focused on extracting suspicious areas and providing confident scores from the screening images where spiculated masses, microcalcifications, and bilateral asymmetry occurred. In mass detection, the main techniques used are Gabor wavelet transformation, level-set segmentation, and multi-instance learning. To detect microcalcifications, they used the Laplacian of Gaussian filter in addition to preprocessing and post-analyzing. For bilateral asymmetric analysis, they applied the thin plate spline registration algorithm and a set of regional comparison metrics. Experimental results showed that their work promised initial detection results in tackling mass, calcification, and asymmetry within the framework. Samala, et al. [15], also developed novel methods for prescreening mass candidates in a CAD system for DBT images. The methods consisted of gradient field convergence features based on first-order features and eigenvalues based on second-order features from the Hessian matrix. The features from the first- and second-order analysis formed the feature vector that was input to a linear discriminant analysis (LDA) classifier to generate a candidate likelihood score; 3D objects were obtained with two-stage 3D clustering followed by active contour segmentation. Morphological, gradient field, and texture features were extracted, and feature selection was performed using stepwise feature selection. A combination of LDA and rule-based classifiers was used for false positive reduction. Experimental results showed that at breast-based sensitivities of 70% and 80%, prescreening using first- and second-order features resulted in 0.7 and 1.0 false positives/DBT, respectively.

2.2. CNN-Related Work

Deep learning methods have been widely applied to address several computer imaging tasks. The main advantages rely on avoiding the design of specific features. Some studies have focused on exploring deep learning

methods to analyze mammograms automatically. For example, Samala, *et al.* [16] developed a CAD system for mass detection in DBT volumes using a deep CNN (DCNN) with transfer learning from mammograms. They compared the performances of two CAD systems for mass detection in DBT; one used the DCNN-based approach, and the other used their previously developed feature-based approach for false positive reduction. The prescreening stage was identical in both systems, passing the same set of mass candidates to the false positive reduction stage. For the feature-based CAD system, 3D clustering and an active contour method were used for segmentation; morphological, gray level, and texture features were extracted and merged with a linear discriminant classifier to score the detected masses. For the DCNN-based CAD system, ROIs from five consecutive slices centered on each candidate were passed through the trained DCNN, and a mass likelihood score was generated. This study demonstrated that large data sets collected from mammography are useful for developing new CAD systems for DBT, alleviating the problem and effort of collecting entirely new and large data sets for the new modality.

Jiao, *et al.* [17], proposed a deep feature-based framework with a CNN and a decision mechanism for breast mass classification. The framework combined intensity information and deep features automatically extracted by the trained CNN from the original images, which were regarded as detailed descriptions of the breast mass. Then, classifiers based on the aforementioned features were used in combination to predict classes of test images. The proposed method was applied to the DDSM dataset and achieved high accuracy under two objective evaluation measures. Arevalo, *et al.* [18], presented an evaluation of CNNs to learn features for mammography mass lesions before feeding them into a classification stage. Experimental results showed that their approach outperformed the state-of-the-art representation from 79.9% to 86% in terms of area under the receiver operating characteristic (ROC) curve. Dubrovina, *et al.* [19], developed a supervised deep learning-based framework for regional classification into semantically coherent tissues. The proposed method used a CNN to learn discriminative features automatically. They also used convolutional layers instead of the traditional fully connected layers to accelerate pixelwise automatic class prediction. Experimental results showed that the approach resulted in significantly faster computations while preserving classification accuracy.

Lévy and Jain [20], applied CNNs to directly classify pre-segmented breast masses in mammograms as benign or malignant. Their work used a combination of transfer learning, careful preprocessing, and data augmentation to overcome limited training data. The proposed method achieved state-of-the-art results on the DDSM dataset, surpassing human performance, and demonstrated the interpretability of the model. Rouhi, *et al.* [21], presented two automated methods to diagnose mass types as benign or malignant in mammograms. One method used automated region growing to segment images where a threshold was obtained by a trained artificial neural network (ANN). The other method used a cellular neural network with a genetic algorithm for feature selection to segment images. The segmented tumor features consisted of intensity, texture, and shape features. ANNs were also used to classify the mammograms as benign or malignant. The proposed methods were compared with different classifiers such as random forest, naïve Bayes, support vector machine (SVM), and KNN to evaluate their performance. Experimental results applying the proposed approach to the MIAS and DDSM databases were promising. The obtained sensitivity, specificity, and accuracy were 96.87%, 95.94%, and 96.47%, respectively. Kooi, *et al.* [22], provided a head-to-head comparison between state-of-the-art mammography CAD system and a CNN system for CAD of mammographic lesions. The first system relied on a manually designed feature set, whereas the second system aimed to read mammograms independently. Both systems were trained on a large data set of approximately 45,000 images; CNN outperformed the traditional CAD system at low sensitivity and performed comparably at high sensitivity.

3. Materials and Methods

This study was approved by the Institutional Review Board of Tzu Chi Hospital. Figure 3 presents a block diagram of the proposed system. First, DBT images are read from the DBT database. Subsequently, image preprocessing is performed, which includes noise removal, intensity conversion, morphological treatment, and breast ROI extraction. Next, the candidate tumors are determined by experts through threshold methods. The subsequent step has two branches: feature-based CAD and CNN-based CAD. Feature-based CAD computes the 3D features of tumors and normal tissues. These features are intensity, shape, radial length, and chord length. Four classifiers are then applied for breast cancer classification: SVM, decision tree (DT), random forest classifier (RFC), and naïve Bayes (NB). CNN-based CAD consists of image cropping, data augmentation, CNN model selection, learning, and classification. The last step is the performance evaluation computed using the accuracy, precision, recall, and F-measure. Details of each step are described as follows.

3.1. Data Sets

DBT images of 20 practical cases were obtained from TVGH in Taiwan. All of the subjects were Asian females, comprising nine premenopausal women (mean age = 44 years; age range = 39–49 years) and 11 postmenopausal women (mean age = 57 years; age range = 51–68 years). These data were acquired with a low-dose protocol in 11 projections at angles from −25° to +25°. The number of images and the range of angles varied for each case.

3.2. Preprocessing

In this stage, the DBT images were processed to eliminate noise and non-breast areas such as artifacts and background to restrict the ROI to the breast area alone for the detection of breast cancer. The preprocessing methods were the Wiener filter, contrast-limited adaptive histogram equalization (CLAHE) enhancement, discrete wavelet

transform (DWT), morphological image processing, breast ROI extraction, and adjustment of image intensity values, all of which are described as follows.

Fig-3. Block diagram of the proposed system
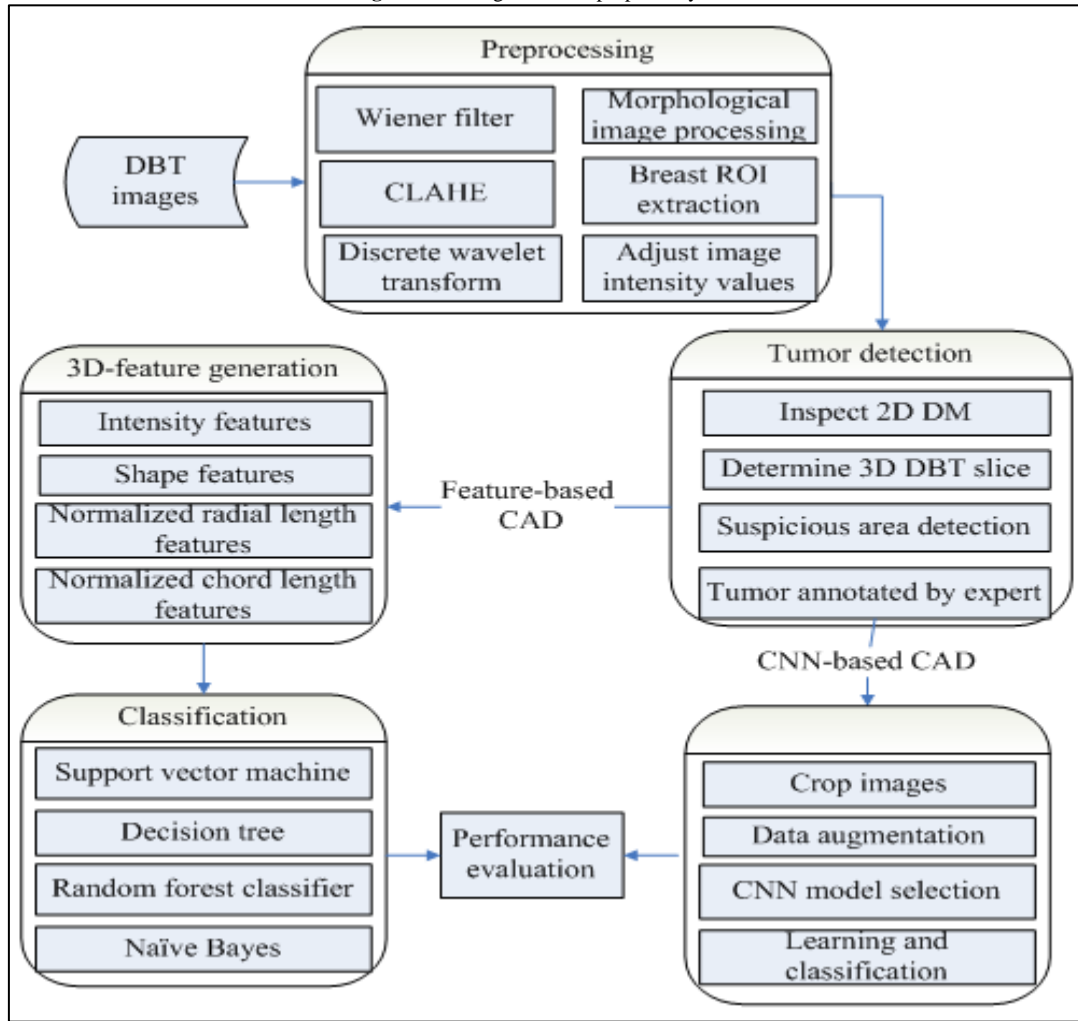
The Wiener filter is a low-pass filter; it filters a grayscale image that has been degraded by constant-power additive noise. The filter uses a pixelwise adaptive Wiener method based on statistics estimated from the local neighborhood of each pixel. CLAHE enhances the contrast of the grayscale image using operations on small regions in the image called tiles, rather than the entire image. Each tile's contrast is enhanced; thus, the histogram of the output region approximately matches the histogram specified by the distribution parameter. The neighboring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that might be present in the image.

The DWT is an implementation of the wavelet transform using a discrete set of wavelet scales and translations obeying certain defined rules. In other words, the DWT decomposes the signal into mutually orthogonal sets of wavelets, which is the main difference compared with continuous wavelet transform. In this study, we applied the DWT to remove background noise. Morphological image processing is a collection of nonlinear operations related to the shape or morphology of features in an image. The main operations of morphology are erosion and dilation. Erosion removes small-scale details from a binary image but simultaneously reduces the size of ROIs. Dilation has the opposite effect to erosion, adding a layer of pixels to both the inner and outer boundaries of regions.

In this study, we used a threshold method to extract the main breast ROI from the background. The value of this threshold was determined using the minimum value between the two most critical peaks, which are the peaks of the background and breast tissue on the histogram of the image. Adjustment of image intensity values increases the contrast of the image by mapping the values of the input intensity image to new values; thus, by default, 1% of the data are saturated at low and high intensities of the input data.

3.3. Feature-Based CAD
3.3.1. 3D Features
This study proposed 3D features including intensity, shape, normalized radial length (NRL), and normalized chord length (NCL) features, which are described as follows:
1. Intensity features

Intensity features depicted on the intensity histogram of an image provide various statistical properties. They are based only on the distribution of individual voxel values other than the interaction or co-occurrence with neighboring ones. Assume an individual object segmented from image $F(x, y, z)$ is defined as:

$$F_{obj_i}(x, y, z) = \begin{cases} 1, (x, y, z) \text{ is a voxel in object } i, \\ 0, \text{ otherwise} \end{cases}$$

(1)

The edge voxel of object i is defined as:

$$P_i(x, y, z) = \begin{cases} 1, F_{obj_i}(x, y, z) \text{ is an edge voxel of object } i, \\ 0, \text{ otherwise} \end{cases}$$

(2)

In this study, six intensity features were calculated, namely mean (F_1), standard deviation (F_2), third moment (F_3), smoothness (F_4), entropy (F_5), and uniformity (F_6). Let k and $p(k)$ denote the intensity and its probability from the object histogram defined by the segmentation of DBT images, respectively.

$$F_1 = \mu = \sum_{k=0}^{65535} k \times p(k)$$

(3)

$$F_2 = \sigma = \sqrt{\sum_{k=0}^{65535} (k - \mu)^2 \times p(k)}$$

(4)

$$F_3 = \frac{1}{\sigma^3} \sum_{k=0}^{65535} (k - \mu)^3 \times p(k)$$

(5)

$$F_4 = 1 - \frac{1}{1 + \sigma^2}$$

(6)

$$F_5 = \sum_{k=0}^{65535} p(k) \times \log_2 p(k)$$

(7)

$$F_6 = \sum_{k=0}^{65535} p^2(k)$$

(8)

3.3.2. Shape Features

Shape features provide various statistical and shape properties using the intensity histogram of an image. Assume that bac denotes the mean intensity of the tumor background. The max and min axes are the maximum and minimum axial degrees of the 3D tumor, respectively. Six shape features were extracted [23], namely foreground background ratio (F_7), foreground background difference (F_8), difference ratio (F_9), volume (F_{10}), perimeter (F_{11}), compactness (F_{12}), and elongation (F_{13}).

$$F_7 = \frac{\mu}{bac}$$

(9)

$$F_8 = \mu - bac$$

(10)

$$F_9 = \frac{\mu - bac}{\mu + bac}$$

(11)

$$F_{10} = \sum_{\forall x, \forall y, \forall z} F_{obj_i}(x, y, z)$$

(12)

$$F_{11} = \sum_{\forall x, \forall y, \forall z} P_i(x, y, z)$$

(13)

$$F_{12} = F_{11}^2 / F_{10}$$

(14)

$$F_{13} = \max axis / \min axis$$

(15)

3.3.3. NRL Features

NRL is defined as the Euclidean distance from the object center to each edge coordinate normalized by dividing by the maximum radial length [24]. For an object contour, the NRL vector is given as $R = \{r_i, 0 \leq i \leq N-1\}$, where

N is the number of edge voxels on the contour and $r_i \leq 1$. Let (x_0, y_0, z_0) be the coordinates of the centroid and (x_i, y_i, z_i) be the coordinates of the edge voxel at the ith location. The NRL is defined as:

$$d_i = \frac{\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2}}{\max(d_i)} \quad i = 1,2,\cdots,N$$

(16)

where $\max(d_i)$ is the maximum value of the radial length of the extracted region. In this study, six features were extracted from the NRL vector, namely mean (F_{14}), standard deviation (F_{15}), entropy (F_{16}), area ratio (F_{17}), and zero crossing count (F_{18}).

$$F_{14} = \frac{1}{N}\sum_{i=1}^{N} d_i$$

(17)

$$F_{15} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - F_{14})^2}$$

(18)

$$F_{16} = -\sum_{k=1}^{bin} p_k(\log(p_k))$$

(19)

$$F_{17} = \frac{1}{F_{14} \times N}\sum_{i=1}^{N}(d_i - F_{14})$$

(20)

$$F_{18} = \frac{1}{N}\sum_{i=1}^{N-1}|d_i - d_{i+1}|$$

(21)

3.3.4. NCL Features

A chord length measure L_i is defined as the length of the line segment that links a pair of boundary points normalized by the length of the longest chord. The complete set of chords for a given object comprises all possible chords drawn from every boundary pixel to every other boundary pixel. The NCL method is invariant to size, translation, and rotation and is stable with respect to noise and distortion in the shape boundary[25]. There are $K = N(N-1)/2$ unique chords of the N boundary points of an object as a sample distribution set and computed Kolmogorov–Smirnov (K–S) statistics of the chord length distribution as shape factors. Four features were computed from chord length distributions, namely mean (F_{19}), standard deviation (F_{20}), skewness (F_{21}), and kurtosis (F_{22}).

$$F_{19} = \frac{1}{K}\sum_{i=1}^{K} L_i$$

(22)

$$F_{20} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(L_i - F_{19})^2}$$

(23)

$$F_{21} = \frac{1}{F_{20}^3 \times K}\sum_{i=1}^{K}(L_i - F_{19})^3$$

(24)

$$F_{22} = \frac{1}{F_{20}^4 \times K}\sum_{i=1}^{K}(L_i - F_{19})^4$$

(25)

3.3.2. Classification Methods
3.3.2.1.SVMs

SVMs have been extensively investigated for classification, regression, and density estimation. An SVM employs nonlinear mapping to transform the original training data into data with a higher dimension. Using this new dimension, the SVM searches for the optimal linear-separating hyperplane. The SVM identifies this hyperplane using support vectors and margins [26].

3.3.2.2. DT

For classification, DT splits data into subsets such that each subset belongs to only one class. Graphically, this is equivalent to dividing the input space into regions that are as pure as possible. Boundaries separating these regions are called decision boundaries. The DT model makes classification decisions based on these decision boundaries. DT is a hierarchical structure with nodes and directed edges. A classification decision is made by traversing the DT starting with the root node. At each node, the resulting information gain determines which branch to traverse to. When a leaf node is reached, the category at the leaf node determines the classification decision. Therefore, a typical

decision learning system recursively selects attributes to test and split the data set into subsets according to the outcome of the information gain function [26].

3.3.2.3. RFC

The RFC is a classifier containing multiple DTs, and its output category is determined by the mode of the output of individual trees. Each tree is built according to the following steps. First, the number of training samples (denoted as N) is determined as well as the number of features (denoted as M). The input feature number m is then used to determine the decision result of a node on the DT, where m should be much less than M. The N training samples, sampled N times using a put-back sampling method, form a training set (that is, bootstrap sampling). Unselected samples are used as test data to evaluate the errors. For each node, m features are randomly selected, and the decision of each node in the DT is determined based on these features. Finally, the best split method is calculated according to these m features. Each tree will grow completely without pruning [27].

3.4. Naïve Bayes (NB)

NB is a family of simple probabilistic classifiers based on Bayes' theorem with assumptions of naïve independence between features. NB models assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from a finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle. All naïve Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable [26].

3.5. CNN-Based CAD

Deep learning experienced three waves of development: cybernetics from the 1940s to the 1960s, connectionism from the 1980s to the 1990s, and deep learning beginning in 2006. The framework of deep learning is generally based on ANNs. For example, the new cognitive machine proposed by Fukushima in 1980 and the standard back propagation algorithm by LeCun et al. in 1989 were used to identify handwritten zip codes. However, the training of the neural network required 3 days; thus, it could not be put to practical use. Until recently, (1) many common test data sets have been used, such as TIMIT in speech recognition and ImageNet, Cifar10, and other large data sets in image recognition; (2) GPUs appeared to greatly increase the speed of numerical and matrix calculations, resulting in a significant reduction in the execution time of machine learning algorithms; and (3) advances have been made in algorithms that promote the revival of neural networks [28].

For a computer, it is difficult to understand the meaning of the raw sensor input data, such as the representation of a character image as a collection of pixel values and the mapping of a set of pixels to the function of a marker object. Learning or evaluating such mappings seem to be difficult for computers, and in-depth learning breaks down the complex maps required to resolve them into a series of simple maps. Each simple map is described by different layers of the model. For example, input appears in the visible layer, which contains variables that we can observe. Then, a series of hidden layers are used to extract increasingly abstract features from the image. These are called hidden layers because their weight values are not in the input data series. For a given pixel, the first layer can compare the brightness of adjacent pixels to easily identify the edge. Given the first hidden layer to the edge description, the second hidden layer can be easily searched for corners and contours. These contours can be identified as a collection of edges. Given the second hidden layer's description of the corners and contours, the third hidden layer can locate specific objects. Finally, the image is described based on the parts of the object it contains, and the model can be used to identify the object [28].

In deep learning, a CNN has a positive effect on image recognition. Usually, a convolutional layer and a pooled layer are added to a traditional full-joint model. The neural network captures features through the convolutional layer. Subsequently, resampling through the pooling layer reduces the resolution of the acquired image and reduces the amount of data. The data applicable to this method usually have the same style after reducing the sampling, and the same feature has the same effect in any position within the data. CNN can improve the shortcomings of traditional machine learning. For example, before training, capturing different forms of a feature through various methods is necessary as is testing its performance on the model to determine the complicated process of effective features. Figure 4 shows the architecture of a typical DCNN. It consists of one input layer, three hidden layers, and one output layer. The input layer has a small 16-by-16-pixel image. The three hidden layers consist of two convolutional layers and one fully connected layer. The final output represents the type of image, such as a benign or malignant tumor.
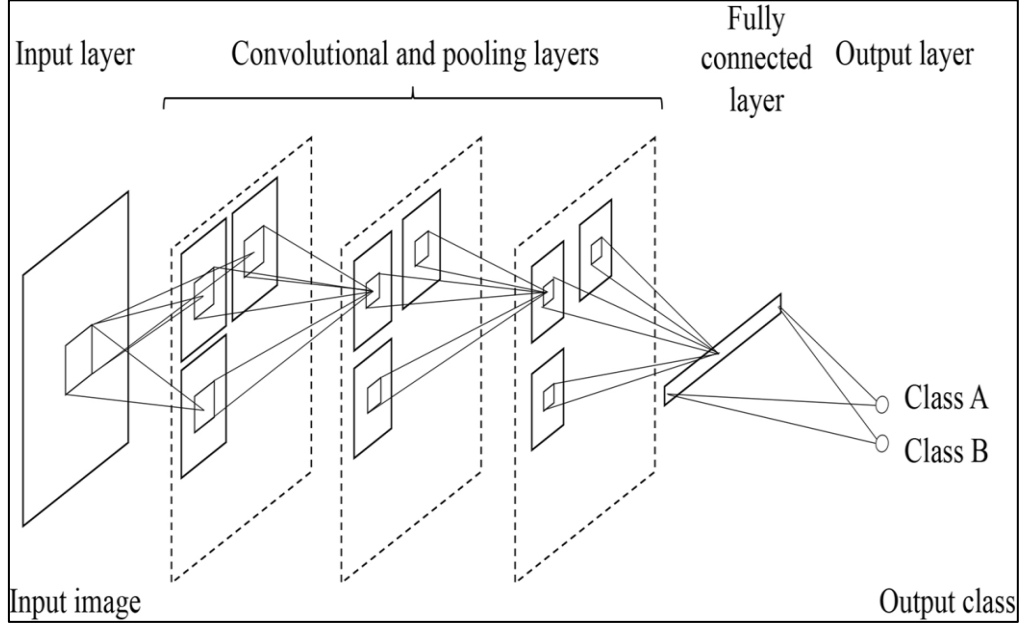
Fig-4. Architecture of a typical DCNN [3]

Figure 4 can be applied to a CNN-based CAD architecture for breast cancer classification. It contains an image input layer followed by several stages of convolution and pooling of the convolutional layer, which then connects to the fully connected layer. Finally, the fully connected layer outputs the category label. The convolutional layer is used as a feature extraction tool for learning the representation of features from the input image. The neurons in the convolutional layer are arranged as feature maps. Each neuron in the feature map has a receptive field that is connected to neighboring neurons in the previous layer and passes through a set of training weights, sometimes referred to as a filter bank [29]. Subsequently, input data and learning weights are convoluted to calculate a new feature map. The result of the convolutional operation is sent to a nonlinear activation function. The weight of all neurons in the feature map is limited to being equal. However, different feature maps have different weights in the same convolutional layer; thus, several different features can be extracted at each location. The kth output feature map Y_k can be defined as

$$Y_k = f(W_k * x) \tag{26}$$

where x is the input image, W_k is a convolution filter for k feature maps, and * is a 2D convolution operator that is used to calculate the filter model's inner product at each position of the input image.

Figure 5 shows a simple example of convolution of a 4×4 input image with a 3×3 filter. The f($\cdot$) denotes a nonlinear activation function that allows the extraction of nonlinear features. Traditionally, sigmoid and hyperbolic tangent functions have been commonly used, and recently, rectified linear units (ReLUs) have been defined as

Fig-5. Simple example of the convolution of a 4×4 input image with a 3×3 filter
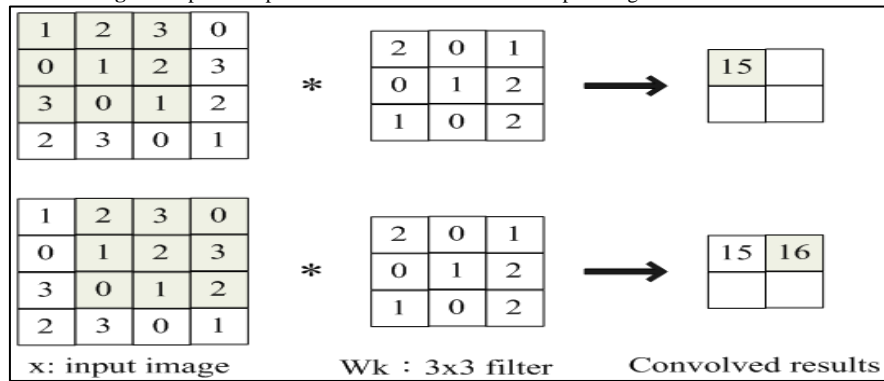
$$f(x) = \begin{cases} x, & x > 0 \\ o, & x \leq 0 \end{cases} \tag{27}$$

The purpose of the pooling layer is to reduce the spatial resolution of the feature map to achieve spatial invariance of input distortion and transformation. Initially, the average pooled aggregation layer is used to generate the average of all input values from the small neighborhood of one image to the next. However, the largest pooled aggregation layer has been used to pass the maximum value in the accepted domain to the next layer, and the largest pooling selects the largest element in each receptive field; thus,

$$Y_{kij} = \max_{(p,q) \in \mathcal{R}_{ij}} x_{kpq} \tag{28}$$

where Y_{kij} is the output of the pooling operation associated with the kth feature map, and x_{kpq} is the element contained in the pooling location (p, q) contained by the pooling region R_{ij}. Figure 6 presents an example of simple maximum pooling.

Fig-6. Example of simple maximum pooling
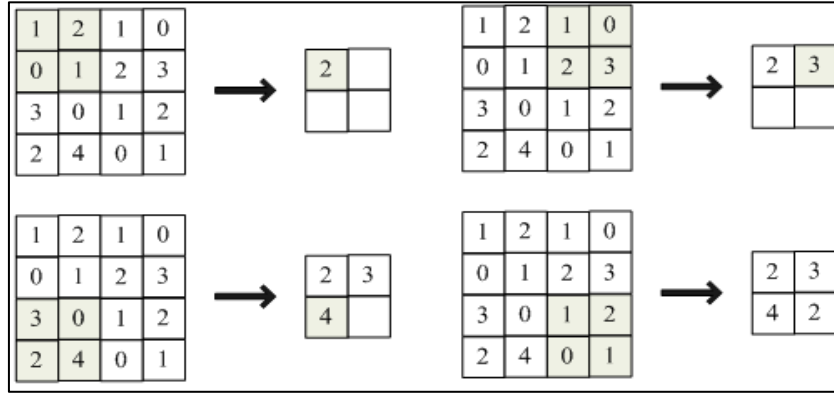
3.6. Performance Evaluation

Table 1 presents a confusion matrix for breast cancer classification. True positive (TP) represents the number of samples that both the system and the expert consider to be tumor; true negative (TN) denotes the number of samples that both the system and the expert considered to be normal; false positive (FP) is the number of samples identified as normal by the expert but classified as tumor by the system; and false negative (FN) is the number of samples identified as tumor by the expert but classified as normal by the system. The performance of the proposed method was determined using four well-known criteria: accuracy, precision, recall, and F-measurement, which are defined as follows [26].

Table-1. Confusion matrix for breast cancer classification

		Predicted class	
		Tumor	Normal
Actual class	Tumor	TP (true positive)	FN(false negative)
	Normal	FP (false positive)	TN(true negative)

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \qquad (29)$$

$$\text{Precision} = \frac{TP}{TP+FN} \qquad (30)$$

$$\text{Recall} = \frac{TP}{TP+FP} \qquad (31)$$

$$\text{F Measurement} = \frac{2*Precision*Recall}{Precision+Recall} \qquad (32)$$

4. Experimental Results

All implementations were performed on a PC with an Intel Core i7-4770 CPU @ 2.20 GHz and 16.0 GB of RAM with the operating system Microsoft Windows 10. The Python programming language, Sony Neural Network Console [30], and SPSS statistical software were used for analysis.

4.1. Preprocessing and Tumor Detection Results

The preprocessing results are demonstrated in Figure 7 using Case 17 from TVGH. Figure 7 (a) shows the original Digital Imaging and Communications in Medicine (DICOM) images. From left to right, they are slices 1, 8, 15, 22, 29, 36, and 43 of Case 17. DICOM is a standard for handling, storing, printing, and transmitting information in medical imaging; it includes a file format definition and a network communications protocol. Those images are disturbed by noise, and it is difficult to detect the tumor. Figure 7 (b) depicts DBT images after Wiener filter processing. Figure 7 (c) shows the images enhanced by CLAHE; the tumor shape can be clearly seen. Figure 7 (d) is a DBT image after the DWT. Figure 7 (e) is a DBT image in which the brightness of the breast is improved after the mathematical morphology is applied. Figure 7 (f) is a DBT image after intensity transform; the tumor shape is more prominent. Figure 7 (g) shows the binary DBT images determined through the grayscale histogram threshold method as well as by an expert. Figure 8 shows 3D feature images of tumors and normal tissues in groups of three.

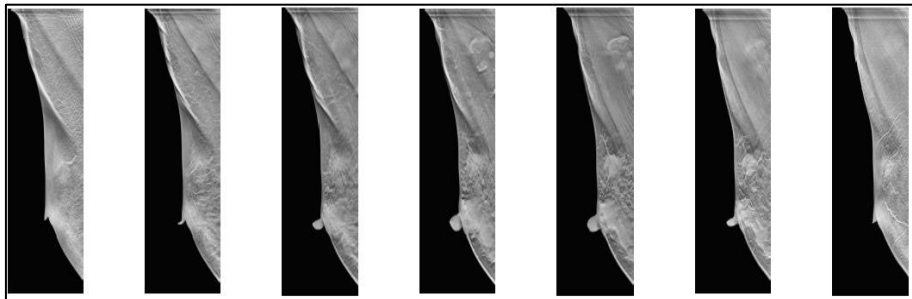**Fig-7.** Preprocessing results of Case 17 from TVGH
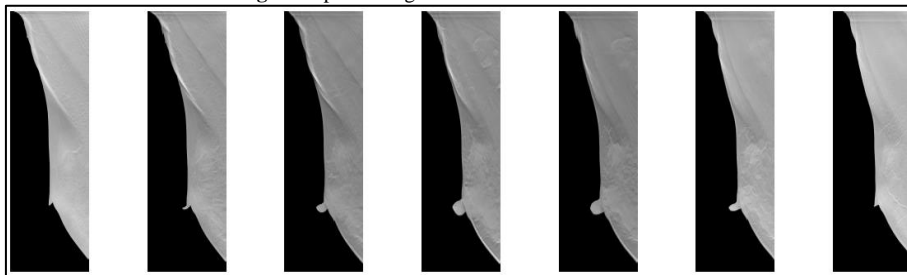
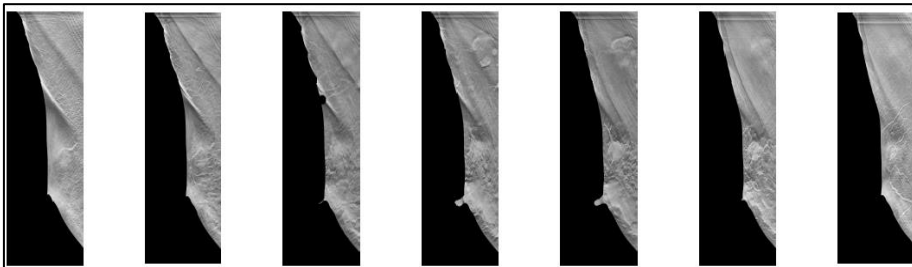(a) Original DICOM images of slices 1, 8, 15, 22, 29, 36, and 43 for Case 17
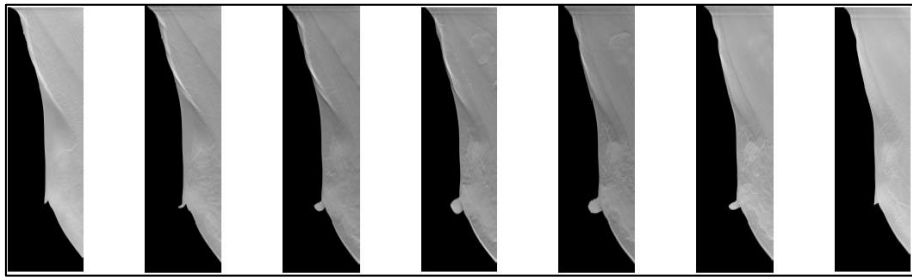
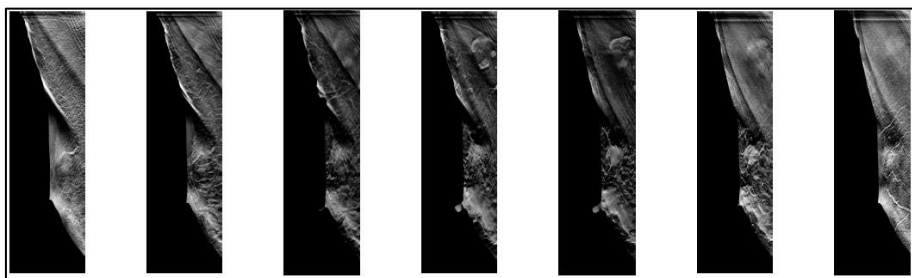(b) DBT images after Wiener filter processing
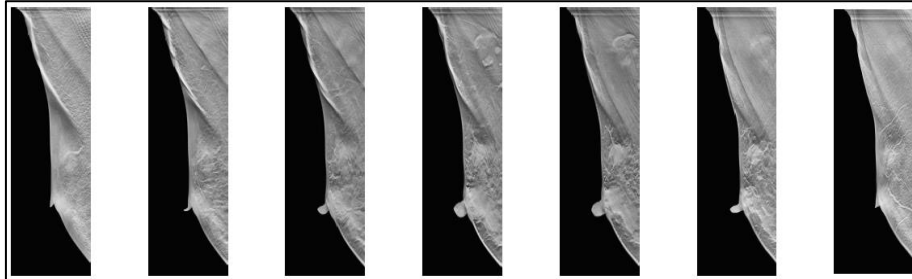
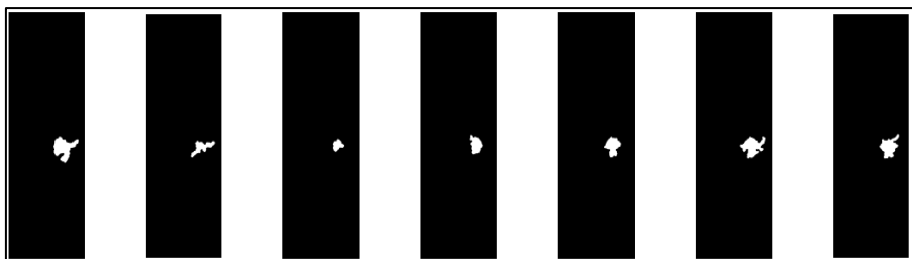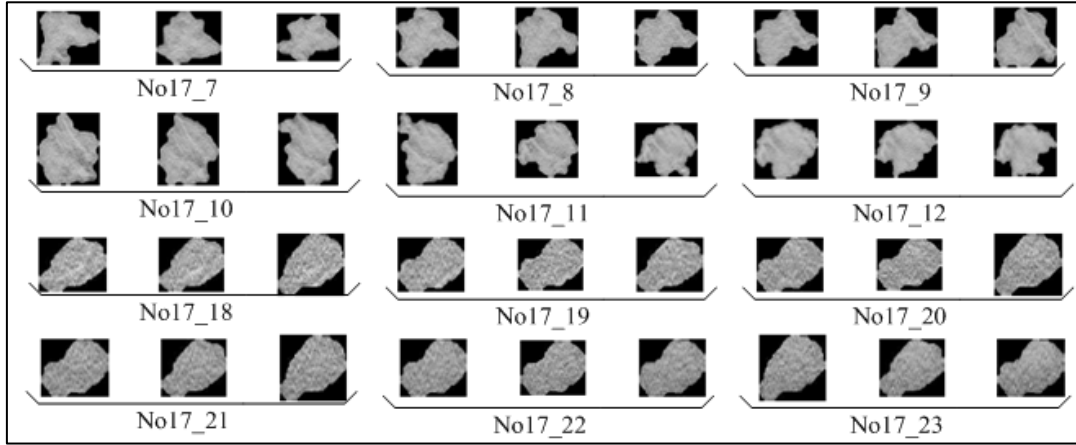(c) DBT images enhanced by the CLAHE

(d) DBT images after DWT

(e) DBT images after applying mathematical morphology

(f) DBT images after intensity transform


(g) Binary DBT images.

Fig-8. 3D feature images of tumors and normal tissues in groups of three

4.2. Feature-Based CAD Classification Results

Using 3D data, a total of 22 features were calculated for each case comprising six intensity features, six shape features, six NRL features, and four NCL features. Table 2 lists some features of tumors and normal samples. A comparison of features F_1, F_8, and F_{20} between tumor and normal types of DBT is presented in Figure 9. Obvious differences existed between cancer and normal samples.

Table-2. 3D features of tumors and normal samples

No	F_1	F_2	…	F_7	F_8	…	F_{20}	F_{21}	F_{22}	Class
No17_7	164.53	19.18	…	1.12	4629.95	…	0.19	2.19	6.79	Tumor
No17_8	166.36	19.97	…	1.16	5866.10	…	0.23	0.95	2.65	Tumor
No17_9	168.31	21.70	…	1.19	7028.14	…	0.27	0.47	1.94	Tumor
No17_10	160.99	20.84	…	1.19	6584.39	…	0.24	0.89	2.62	Tumor
No17_11	164.74	24.61	…	1.23	7932.32	…	0.21	0.56	2.90	Tumor
No17_12	170.02	29.78	…	1.24	8361.43	…	0.18	0.56	3.02	Tumor
No17_18	166.73	23.62	…	1.03	1235.15	…	0.05	13.74	205.12	Normal
No17_19	162.12	22.16	…	1.02	644.41	…	0.05	14.66	235.61	Normal
No17_20	157.27	20.38	…	1.01	412.67	…	0.05	14.89	236.86	Normal
No17_21	151.82	19.58	…	1.02	685.81	…	0.05	13.70	203.50	Normal
No17_22	150.84	18.67	…	1.03	1155.82	…	0.05	13.97	212.24	Normal
No17_23	120.32	18.56	…	1.00	127.10	…	0.22	1.55	4.33	Normal
…	…	…	…	…	…	…	…	…	…	…

Figure-9. Comparison of features with tumors and normal samples
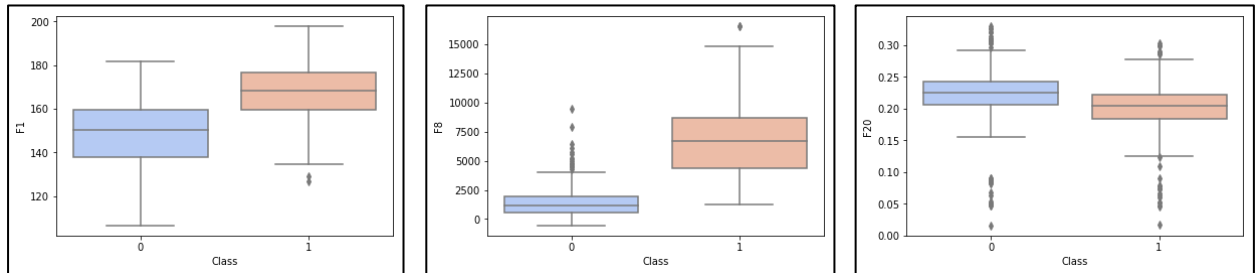
(a) Box plot of F_1 feature   (b) Box plot of F_8 feature   (c) Box plot of F_{20} feature

To evaluate the performance of the classifiers, the 327 samples (173 tumor and 154 normal) were randomly split into 70% for training and 30% for testing. This process was repeated ten times, resulting in ten training and testing data sets. The mean accuracy and standard deviation of DT, RFC, SVM, and NB with 22 features were 89.69% (0.028), 93.33% (0.021), 87.58% (0.031), and 81.11% (0.085), respectively. Other criteria are shown in Tables 3. The Wilks' lambda of the multivariate analysis was 0.313, the F value was 4.025, and the P value was $0.00 < 0.05$, which indicate that the classification methods have significant effects on the results. RFC outperformed other methods.

**Table-3.** Performance of the feature-based CAD for breast cancer classification

Methods	Accuracy % (stdev)	Precision% (stdev)	Recall% (stdev)	F-Measure% (stdev)
DT	89.69 (0.028)	86.74 (0.043)	88.89 (0.059)	89.86 (0.035)
RFC	93.33 (0.021)	90.27 (0.029)	95.40 (0.031)	93.67 (0.021)
SVM	87.58 (0.031)	84.06 (0.052)	87.40 (0.047)	87.92 (0.036)
NB	81.11 (0.085)	76.76 (0.113)	86.11 (0.069)	82.74 (0.077)

4.3. CNN-Based CAD Classification Results

This study fixed the input size of images to 3×50×50 voxels, which are cropped from the center of the ROI automatically. Examples of tumor tissues are shown in Figure 10. Examples of normal tissues are shown in Figure 11. For each image, we artificially augmented three new label-preserving samples using 90°, 180°, and 270° rotation transformations, and then vertically mirrored and rotated again. These results from 1384 tumor images and 1232 normal images were then randomly split into 70% training and 30% testing sets.

Fig-10. Examples of tumor tissues
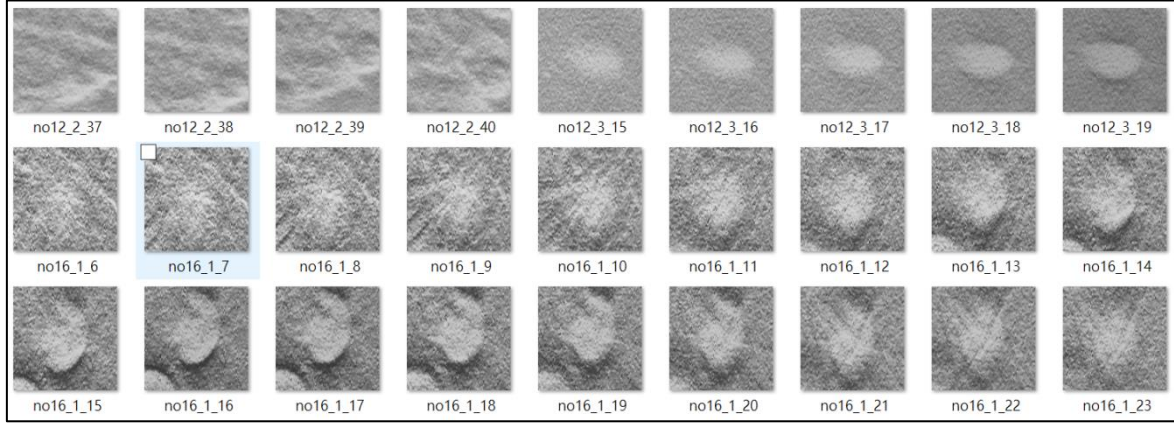
Fig-11. Examples of normal tissues
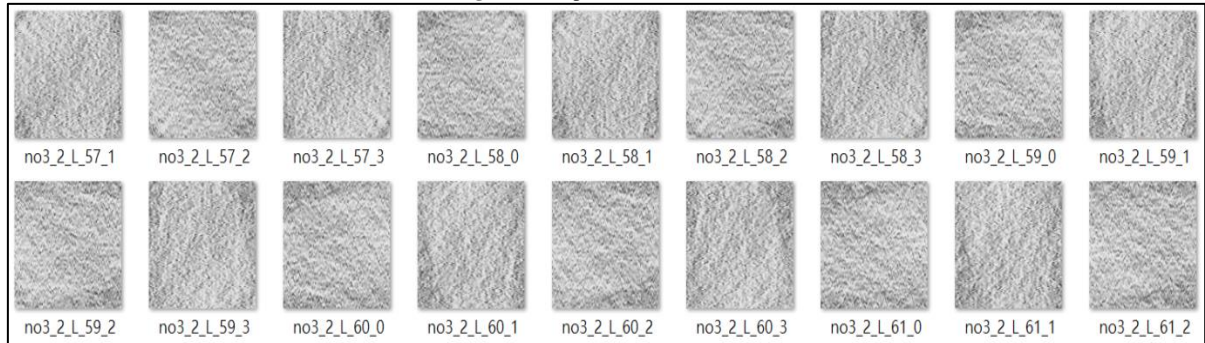
This study proposed two CNN architectures for breast cancer classification after a literature review. The first, depicted in Figure 12, has 5 × 5 local kernels as the first convolutional layer, followed by the 2 × 2 max-pooling layer with a stride of 2 × 2 pixels along with the Tanh activation function, which outputs −1 for inputs less than or equal to 0 and +1 for inputs greater than 0. The second convolutional layer and max-pooling layer have the same architecture as the first. Subsequently, two fully connected layers with sigmoid activation function are stacked to form the output layer.
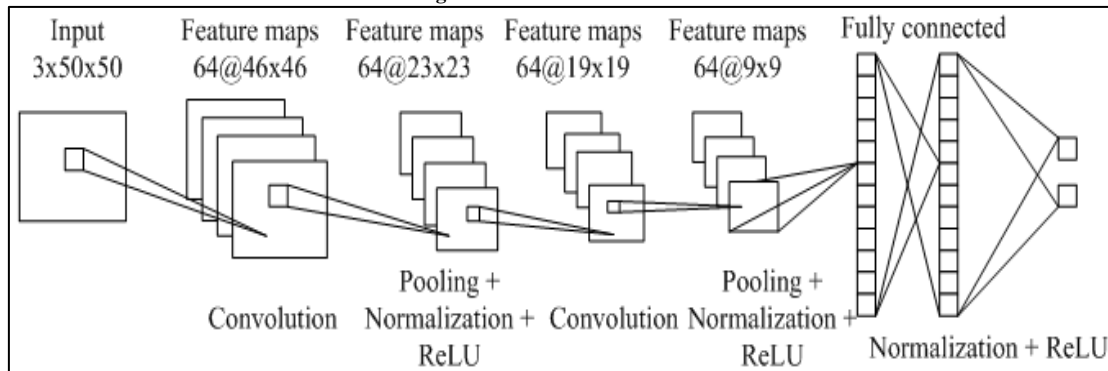
Fig-12. Basic CNN model
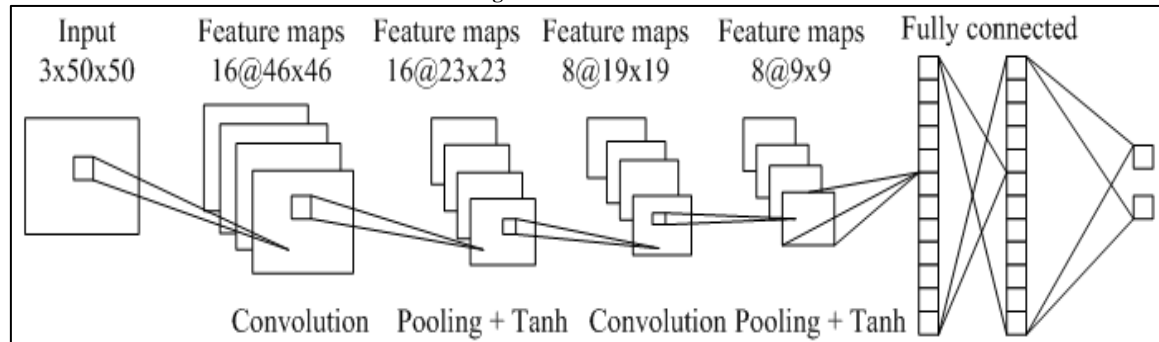
To test the performance of the basic CNN model, four kernels were selected, and images were randomly split into 70% for training and 30% for testing. Each CNN model with different kernels was performed ten times. Experimental results are listed in Table 4. The mean accuracy and standard deviation of Adadelta, Adam, and Adamax, RMSprop, were 80.61% (0.018), 84.04% (0.014), 84.91% (0.014), and 82.54% (0.012), respectively. Other criteria are shown in Table 4. The Wilks' lambda of the multivariate analysis was 0.326, the F value was 3.854, and the P value was $0.00 < 0.05$ which indicates that different kernels have significant effects on the four criteria. Among them, the Adamax kernel had the highest rank.

Table-4. Performance of the basic CNN model for breast cancer classification

Kernels	Accuracy % (stdev)	Precision% (stdev)	Recall% (stdev)	F-Measure% (stdev)
Adadelta	80.61 (0.018)	81.56 (0.023)	79.90 (0.017)	80.07 (0.018)
Adam	84.04 (0.014)	84.50 (0.015)	83.47 (0.014)	83.70 (0.014)
Adamax	84.91 (0.014)	85.00 (0.014)	84.60 (0.014)	84.70 (0.014)
RMSprop	82.54 (0.012)	83.41 (0.013)	81.84 (0.015)	82.06 (0.015)

The second architecture, LeNet-based CNN, depicted in Fig. 13, has 5×5 local kernels as the first convolutional layer, followed by a 2×2 max-pooling layer with a stride of 2×2 pixels along with the ReLU activation function, which outputs a maximum value of 0 and input. The second convolutional layer and max-pooling layer have the same architecture as the first one. Subsequently, two fully connected layers with ReLU activation function are stacked to finally add a softmax classifier. Unlike the basic CNN model, the normalization process is added before each activation function.
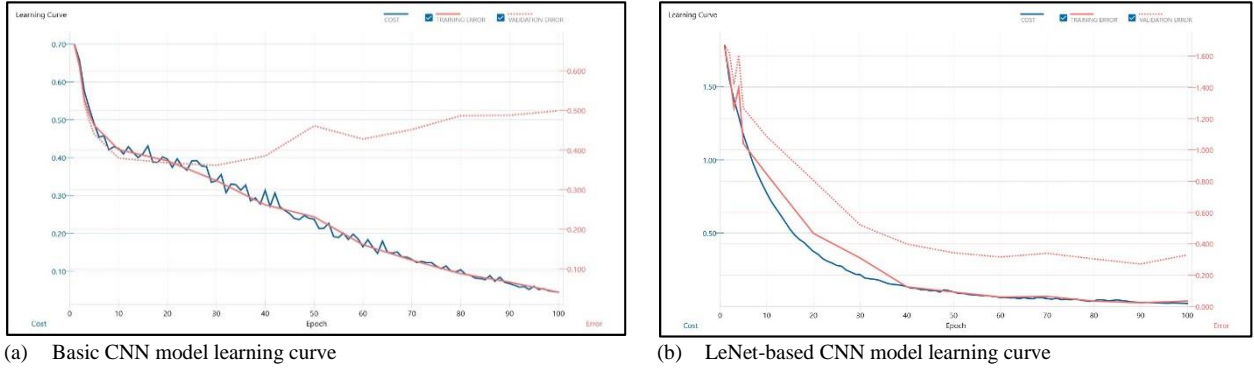
Fig-13. LeNet-based CNN model

To test the performance of the LeNet-based CNN model, four kernels were selected, and images were randomly split into 70% for training and 30% for testing. Each CNN model with different kernels was performed ten times. The experimental results are listed in Table 6. The mean accuracy and standard deviation of Adadelta, Adam, and Adamax, RMSprop, were 88.57% (0.010), 90.43% (0.011), 90.68% (0.010), and 89.00% (0.012), respectively. Other criteria are shown in Table 5. The Wilks' lambda of the multivariate analysis was 0.407, the F value was 2.95, and the P value was $0.002 < 0.05$, which indicate that different kernels have significant effects on the results. Among them, the Adamax kernel had also the highest rank.
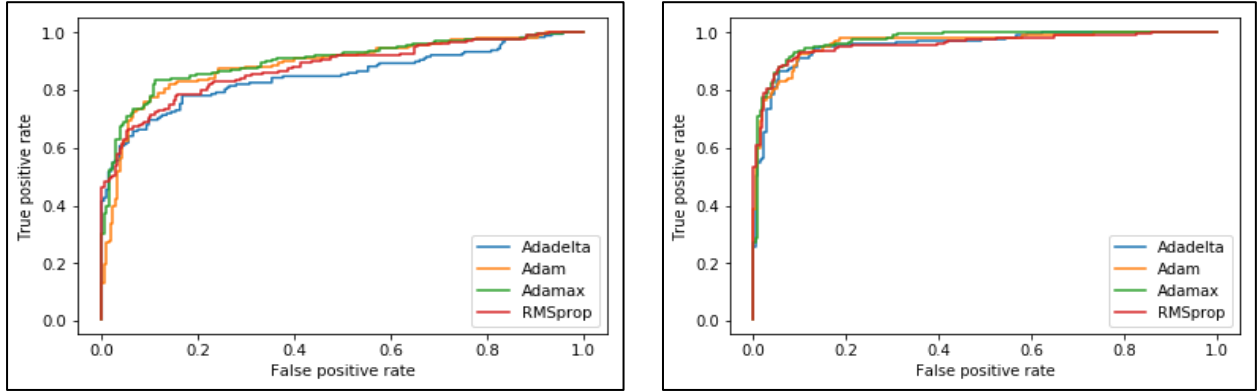
Table-5. Performance of the LeNet-based CNN model

Kernels	Accuracy % (stdev)	Precision% (stdev)	Recall% (stdev)	F-Measure% (stdev)
Adadelta	88.57 (0.010)	88.80 (0.010)	88.29 (0.009)	88.41 (0.009)
Adam	90.43 (0.011)	90.53 (0.010)	90.40 (0.011)	90.34 (0.011)
Adamax	90.68 (0.010)	90.66 (0.010)	90.80 (0.011)	90.62 (0.010)
RMSprop	89.00 (0.012)	89.36 (0.010)	88.69 (0.015)	88.82 (0.013)

The learning curves of the basic CNN model and the LeNet-based CNN model are presented in Figure 14. The major difference of the two figures is the validation error line. The basic model reaches maximum performance at approximately 30 epoch, whereas the LeNet-based model continues to improve. Figure 15 demonstrates the ROC curves of the basic CNN model and the LeNet-based CNN model.

Fig-14. CNN learning curves

(a) Basic CNN model learning curve

(b) LeNet-based CNN model learning curve

Fig-15. ROC curves

(c) Basic CNN model ROC curve

(d) LeNet-based CNN model ROC curve

Finally, multivariate tests were performed to compare the performance between the basic CNN model and the LeNet-based CNN model. The Wilks' lambda was 0.207, the F value was 71.68, and the P value was $0.00 < 0.05$, which indicate that the LeNet-based CNN model outperformed the basic CNN model in breast cancer classification.

4.4. Comparison Between Feature-Based CAD and LeNet-Based CNN CAD

Multivariate tests were also performed to compare the performance between the feature-based CAD and LeNet-based CNN CAD. The Wilks' lambda was 0.068, the F value was 256.67, and the P value was $0.00 < 0.05$, which indicate that the two methods had significant differences in terms of breast cancer classification. Table 5 presents the T-test results from the univariate analysis between the feature-based CAD and LeNet-based CNN CAD; the precision rates (standard deviation) of the LeNet-based CNN CAD and the feature-based CAD for breast cancer classification were 89.84 (0.013) and 84.46 (0.082), respectively. The T value was -4.091 and the P value was $0.00 < 0.05$, which indicate that the LeNet-based CNN CAD significantly outperform the feature-based CAD. However, there is no significantly differences between the LeNet-based CNN CAD and the feature-based CAD on other criteria.

Table-6. T test between feature-based CAD and LeNet-based CNN CAD

Crietia	Feature-based % (stdev)	LeNet-based CNN % (stdev)	t-value	P-value
Accuracy	87.93 (0.065)	89.67 (0.014)	-1.662	0.104
Precision	84.46 (0.082)	89.84 (0.013)	-4.091	0.000
Recall	89.44 (0.063)	89.55 (0.016)	-0.099	0.921
F-Measure	88.55 (0.060)	89.54 (0.014)	-1.018	0.315

5. Conclusion

This study compared feature-based CAD and CNN-based CAD in classifying breast cancer in DBT images. The feature-based CAD focused on 3D feature generation, including intensity, shape, radial length, and chord length features. Four conventional classifiers were applied to test the classification performance. The CNN-based CAD took the raw pixels of the images as inputs to learn in a hierarchical manner a set of nonlinear transformations to represent the visual content of an image. Experimental results showed that the LeNet-based CNN CAD outperformed the feature-based CAD significantly in precision. The results can be applied to clinical medicine as well as to assist radiologists in identifying breast tumors. Future studies will evaluate the proposed strategy on more images and larger CNN architectures as well as include other image modalities to enhance the representation.

Acknowledgement

References

[1]   Dhungel, N., Carneiro, G., and Bradley, A. P., 2016. "The automated learning of deep features for breast mass classification from mammograms." In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016.* pp. 106-114.

[2]   Krizhevsky, A., Sutskever, I., and Hinton, G. E., 2017. "ImageNet classification with deep convolutional neural networks." *Communications of the AM,* vol. 60, pp. 84-90.

[3]   Suzuki, K., 2017. "Overview of deep learning in medical imaging." *Radiological Physics and Technology,* vol. 10, pp. 257-273.

[4]   Park, J. M., Franken, E. A., Garg, M., Fajardo, L. L., and Niklason, L. T., 2007. "Breast tomosynthesis: Present considerations and future applications." *Radio Graphics,* vol. 27, pp. 231-240.

[5]   Smith, A., 2008. *Fundamentals of breast tomosynthesis: Improving the performance of mammography.* White Paper, Hologic Inc., WP-00007, pp. 1-8.

[6]   Cohen, S. L., Margolies, L. R., Szabo, J. R., Patel, N. S., and Hermann, G., 2014. "Introductory pictorial atlas of 3D tomosynthesis." *Clinical Imaging,* vol. 38, pp. 18–26.

[7]   Chan, H. P., Wei, J., Sahiner, B., Rafferty, E. A., Wu, T., and Roubidoux, M. A., 2005. "Computer-aided detection system for breast masses on digital tomosynthesis mammograms: preliminary experience." *Radiology,* vol. 237, pp. 1075–1080.

[8]   Schie, G., Wallis, M. G., Leifland, K., Danielsson, M., and Karssemeijer, N., 2013. "Mass detection in reconstructed digital breast tomosynthesis volumes with a computer-aided detection system trained on 2D mammograms." *Medical Physics,* vol. 40,

[9]   Kontos, D., Bakic, P. R., Carton, A. K., Troxel, A. B., Conant, E. F., and Maidment, A. D., 2009. "Parenchymal texture analysis in digital breast tomosynthesis for breast cancer risk estimation: a preliminary study." *Academic Radiology,* vol. 16, pp. 283-298.

[10]   Houssami, N. and Skaane, P., 2013. "Overview of the evidence on digital breast tomosynthesis in breast cancer detection." *The Breast,* vol. 22, pp. 101-108.

[11]   Palma, G., Bloch, I., and Muller, S., 2014. "Detection of masses and architectural distortions in digital breast tomosynthesis images using fuzzy and a contrario approaches." *Pattern Recognition,* vol. 47, pp. 2467–2480.

[12]   Qin, X., Lu, G., Sechopoulos, I., and Fei, B., 2014. "Breast tissue classification in digital tomosynthesis images based on global gradient minimization and texture features, Medical Imaging 2014, Image Processing, edited by S. Ourselin and M.A. Styner." *Proc. of SPIE,* vol. 9034, 90341, p. 2014.

[13]   Morra, L., Sacchetto, D., Durando, M., Agliozzo, S., Carbonaro, L. A., and Delsanto, S., 2015. "Breast cancer: computeraided detection with digital breast tomosynthesis." *Radiology,* vol. 277, pp. 56-63.

[14]   Lu, Y., Yousefi, M., Ellenberger, J., Moore, R. H., Kopans, D. B., Krzyżak, A., and Suen, C. Y., 2016. *3D tomosynthesis to detect breast cancer.* 5th ed ed. Handbook of Pattern Recognition and Computer Vision, pp. 371-393.

[15]   Samala, R. K., Chan, H. P., Hadjiiski, L., Helvie, M. A., Wei, J., and Cha, K., 2016. "Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography." *Med. Phys.,* vol. 43, pp. 6654-6666.

[16]   Samala, R. K., Wei, J., Chan, H. P., Hadjiiski, L., Cha, K., and Helvie, M. A., 2016. "First and second-order features for detection of masses in digital breast tomosynthesis," Medical Imaging 2016, Computer-Aided Diagnosis, edited by G.D. Tourassi and S.G. Armato III." *Proc. of SPIE,* vol. 9785, p. 2016.

[17]   Jiao, Z., Gao, X., Wang, Y., and Li, J., 2016. "A deep feature based framework for breast masses classification." *Neurocomputing,* vol. 197, pp. 221-231.

[18]   Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., and Lopez, M. G., 2015. "Convolutional neural networks for mammography mass lesion classification." In *The IEEE 37th Annual International Conference in Engineering in Medicine and Biology Society (EMBC).* pp. 797-800.

[19]   Dubrovina, A., Kisilev, P., Ginsburg, B., Hashoul, S., and Kimmel, R., 2018. "Computational mammography using deep neural networks." *Computer Methods in Biomechanics and Biomedical Engineering Imaging and Visualization,* vol. 6, pp. 243-247.

[20]   Lévy, D. and Jain, A., 2016. "Breast mass classification from mammograms using deep convolutional neural networks." *arXiv preprint arXiv,* vol. 1612, p. 00542.

[21]   Rouhi, R., Jafari, M., Kasaei, S., and Keshavarzian, P., 2015. "Benign and malignant breast tumors classification based on region growing and CNN segmentation." *Expert Systems with Applications,* vol. 42, pp. 990-1002.

[22]   Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., and Mann, R., 2017. "Large scale deep learning for computer aided detection of mammographic lesions." *Medical image analysis,* vol. 35, pp. 303-312.

[23]   Gonzalez, R. C. and Woods, R. E., 2008. *Digital image processing.* 3rd Edition ed. Prentice Hall.

[24]   Kilday, J., Palmieri, F., and Fox, M., 1993. "Classifying mammographic lesions using computerized image analysis." *IEEE Transaction on Medical Imaging,* vol. 12, pp. 664–669.

[25]   Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. E., and Alim, O. A., 1997. "Measures of acutance and shape for classification of breast Tumors." *IEEE Transaction on Medical Imaging,* vol. 16, pp. 799-810.

[26]   Kantardzic, M., 2011. *data mining: Concepts, models, methods, and algorithms*. 2nd Ed ed. Wiley-Interscience, NJ, p. 2011.

[27]   Breiman, L., 2001. "Random forests." *Machine Learning,* vol. 45, pp. 5-32.

[28]   Goodfellow, I., Bengio, Y., and Courville, A., 2016. *Deep Learning*. MIT Press.

[29]   LeCun, Y., Bengio, Y., and Hinton, G., 2015. "Deep learning." *Nature,* vol. 521, pp. 436-444.

[30]   Neural Network Console version, 2018. "1.1.6519.49966, sony network communications inc." Available: https://dl.sony.com/