A Robust Deep Neural Network to Enhancement Features Extraction for Cancer Detection and Classification

Romany F. Mansour

Faculty of Science, Northern Border University, Saudi Arabia

Abstract

The exponential rise in breast cancer cases across the globe has alarmed academia-industries to achieve certain more efficient and robust Breast Cancer Computer Aided Diagnosis (BC-CAD) system for breast cancer detection. A number of techniques have been developed with focus on case centric segmentation, feature extraction and classification of breast cancer Histopathological images. However, rising complexity and accuracy often demands more robust solution. Recently, Convolutional Neural Network (CNN) has emerged as one of the most efficient techniques for medical data analysis and various image classification problems. In this paper a highly robust and efficient BC-CAD solution has been proposed. Our proposed system incorporates pre-processing, enhanced adaptive learning based Gaussian Mixture Model (GMM), connected component analysis based region of interest localization, AlexNet-DNN based feature extraction. The Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) based on feature selection which is used as dimensional reduction. One of the advantages of the proposed method is that none of the current dimensional reduction algorithms employed with SVM to perform breast cancer detection and classification. The overall results obtained signify that the AlexNet-DNN based features at fully connected layer; FC6 in conjunction with LDA dimensional reduction and SVM based classification outperforms other state-of-art techniques for breast cancer detection. The proposed BC-CAD system has been performed over real world data BreakHis having significant diversity and complexity, and therefore we suggest it to be used for other real-world applications. The proposed method achieved 96.15 for AlexNet-FC6 and 96.20 for AlexNet-FC7 in term of evaluation measures.

Keywords

Breast Cancer Detection; Computer Aided Diagnosis; Convolutional Neural Network; AlexNet-DNN; Linear Discriminant Analysis.

1. Introduction

Cancer is a type of disease involving the development and growth of abnormal cell that invades or spreads to other parts of a human body. In last few decades, cancer has emerged as one of the deadliest heath diseases claiming huge death rate. There are different types of cancers such as breast cancer, blood cancer, melanoma or skin cancer…etc. However, in last few years, breast cancer has emerged as the second most common cancer type after skin cancer in women. Literatures reveal that almost 50% of the breast cancer cases are found in the developing countries, while approximate 58% of deaths take place in less developed countries. It is estimated that around the world over 5,08,000 women died in 2011 due to breast cancer [1]. Various case studies [1] have revealed that the earlier cancer identification and its diagnosis can play vital role in providing success diagnosis or treatment. A similar survey [2] also indicated high pace rise in deaths caused due to breast cancer in 2012, which is even increasing with vast pace [2]. The study revealed that the death rate can reach up to 27 million till 2030 [3]. The exponential rise in breast cancer has alarmed academia-industries to develop certain vision based approach for earlier breast cancer detection and diagnosis. The rise in advanced computing techniques, vision based computations and decision process…etc. has given rise to a new dimension having immense potential to meet major demands. Non-deniably such development often gives a hope for better human life, health and decision process. In fact, the development of science and technology often intends to enable human life secure, healthy and productive. On the other hand, gigantically rising population across world requires optimal healthcare solutions. In fact, the traditional manual diagnosis processes are either confined or less productive to meet the demands and enable optimal healthcare solution. This as a result has motivated academia-industries to achieve more efficient computer aided diagnosis (CAD) solutions for earlier health-diagnosis. The development of Histopathological analysis and molecular biology has made vision based breast cancer CAD system more efficient [4]. Unlike traditional manual microscopic analysis based approaches vision based CAD system can be of paramount significance for breast cancer detection and classification [5].

In last few years, efforts have been made to exploit X-ray mammography and ultrasound technique to assist breast cancer detection and diagnosis; however, mammography usually exhibits poor sensitivity to the minute size cancers. In addition, it is insignificant especially for augmented breast or dense breast conditions [60]. Similarly, poor specificity and the complexity in whole breast-imaging process make ultrasound method confined and ineffective. As an enhanced solution, authors proposed Magnetic Resonance Imagining (MRI) systems that possess better soft tissue resolution capacity and enhanced sensitivity for

Manuscript received April 5, 2019
Manuscript revised April 20, 2019

breast cancer detection [6]. The MRI was method used to avoid any ionization radiation and hence it can be suitable for patients with implants. The key usefulness of MRI method is its ability of quantification of tumor volume, multi-focal and multi-centric breast cancer detection [7]. Vision based CAD systems have been found efficient to perform more time efficient and accurate breast cancer detection to assist early diagnosis. Exploring in depth of the CAD systems, it can be found that its efficacy primarily depends on the efficiency of the Region Of Interest (ROI) segmentation, feature extraction, feature mapping and classification techniques. There are a number of techniques, such as wavelet transform [8][9][10], Gabor transform [11], applied for feature extraction while classification is done using standard support vector machine (SVM) or other artificial intelligence approaches [9][12][13]. However, most of the traditional approaches are confined due to limited feature extraction efficiency and classification for lesion of different shape, size, and orientation density. The low information availability too confines the efficacy of the traditional feature extraction and cancer classification models. In addition, there is an inevitable need to develop efficient approaches to deal with high dimensional features, feature selection and classification or prediction. Considering large size unannotated data (i.e., breast cancer image data), the use of CNN as feature extraction tool can ensure better breast cancer CAD (BC-CAD) solution.

In this paper, a highly robust and efficient CNN based Deep Neural Network (DNN) model has been developed for breast cancer detection and classification. To enable an optimal solution, our proposed model encompasses various enhancements including pre-processing, ROI identification, Caffe-enet-AlexNet DNN based feature extraction, PCA and LDA based feature selection and SVM based classification. To examine the effectiveness of the proposed DNN model, a parallel BC-CAD model using Spatial Invariant Fourier Transform (SIFT) feature extraction has been developed. In addition, to perform breast cancer prediction, a two class classification model is developed; where the overall result affirms that the AlexNet-DNN with LDA feature extraction outperforms other state-of-art techniques for BC-CAD purpose.

The other sections of the presented manuscript are divided as follows: Section two presents the related work, which is followed by the discussion of the proposed AlexNet DNN based breast cancer detection and classification in Section three. Section four presents the results and discussion, while the overall research conclusion is presented in Section five.

2. Related work

This section briefs some of the key literatures pertaining to the breast cancer detection and related technologies.

Rehman et al. [12] derived a diverse feature based breast cancer detection model, where they applied phylogenetic trees, various statistical features and local binary patterns for generating certain distinctive and discriminative features to perform classification. Authors applied radial basis function based SVM for two class classification: cancerous and non-cancerous. Pramanik et al. [8] assessed the breast thermogram to perform cancer detection, where they at first applied Initial Feature Point Image (IFI) to perform feature extraction for each segmented breast thermogram. They applied DWT to perform feature extraction, which was followed by feature classification using feed-forward Artificial Neural Network (ANN). A similar effort was made in [9], where authors applied wavelet and counterlet transformation for feature extraction and SVM classifier based classification of the mammograms. Yousefi et al. [11] used multi-channel Gabor wavelet filter for feature extraction over mammograms. Authors applied Gabor filter design to perform feature extraction, which was then followed by two class classifications using Bayesian classifier. Caorsi et al. [14] proposed an ANN based radar data processing model for breast cancer detection where they emphasized not only on the detection of the cancerous tumour but also its geometric features such as depth and width. A similar effort was made in [15], where authors proposed a data-driven matched field model to assist microwave breast cancer detection. George et al. [16] developed robust and intelligent breast cancer detection and classification model using cytological images. Authors applied different classifiers including back-propagation based multilayer perceptron, Probabilistic Neural Network (PNN), SVM and learning vector quantization to perform breast cancer classification. Considering mitosis as a cancer signifier, Tashk et al. [17] developed an automatic BC-CAD model, where at first they applied 2-D anisotropic diffusion for noise removal and image morphological process. To achieve pixel-wise features from the ROI (i.e., mitosis region), authors derived a statistical feature extraction model. In addition, to alleviate the issue of misclassification of mitosis and non-mitosis objects, authors developed an object-wise Completed Local Binary Pattern (CLBP) that enabled efficient textural features extraction. The predominant novelty of the CLBP model was that it is robust against positional variation, color changes …etc. Authors applied SVM to perform feature classification. A space processing model was developed in [18] that enabled time reversal imaging approach to perform breast cancer detection. To perform malignant lesion detection, authors applied FDTD breast model that comprises dense breast tissues with differing fibro-glandular tissue composition. To achieve better accuracy, Naeemabadi et al. [13] applied LS and SMO techniques rather applying traditional SVM for classification. Li et al. [19] developed an MFSVM-FKNN ensemble classifier for

breast cancer detection, where authors applied the concept of mixture membership function. Fuzzy classifier based breast cancer detection and classification model was proposed in [20]. A bio-inspired immunological scheme was proposed for mammographic mass classification that classifies malignant tumors from the benign ones. Gaike et al. [21] at first focused on retrieving the higher order (particularly third order) features that was later processed for clustering to perform malignant breast cancer detection. Ali et al. [22] developed an automatic segmentation model using the concept of the data acquisition protocol parameter over the image statistics of DMR-IR database. A similar effort was made in [23], where authors developed a two phase BC-CAD system. In their approach at first they applied Neutrosophic Sets (NS) and optimized Fast Fuzzy C-Mean (F-FCM) algorithm for ROI segmentation, which was followed by classification to perform two class classifications, i.e. normal and abnormal tissue. Ismahan et al. [24] used a mathematical morphology concept to perform masses detection over digitized mammograms. Texture analysis based ROI detection for thermal imaging based BC-CAD was performed in [25]. Considering the significance of feature extraction for mammograms classification, Sanae et al. [10] exploited comprehensive statistical Block-Based features, which were extracted from sub-bands of the discrete wavelet transformation. Once mapping the extraction features, SVM was applied to perform classification. Unlike traditional approaches, Maken et al. [26] examined the efficacy of Multiple Instance Learning (MIL) algorithm for mammograms classification, where they developed MIL using tile-based spatio-temporal features. Rosa et al. [27] and Hatipoglu et al. [28] intended to exploit MIL in conjunction with SVM for mammograms classification. Non-deniably, MIL approaches have performed better, particularly for unannotated data; however, the likelihood of better performance with deep features cannot be ignored.

To enable more efficient feature extraction and accurate performance, DNN has gained significant attention. CNN has gained widespread recognition to extract features from the complex image data, such as mammograms, MRI medical data or even histopathological datasets. CNN techniques which are inspired biologically by the organization of human visual cortex possess robust efficiency due to its ability to learn features invariant to translation, rotation and shifting is their great advantage. Spanhol et al. [29] applied CNN approach to classify breast cancer Histopathological images. Their proposed model performed learning CNN by means of training patches generated through varied approaches. Authors obtained the best classification accuracy of 89.6% for images with 40x enlargement. To achieve this result, authors considered the patches of the size 64x64 pixels. Hatipoglu et al. [28] used CNN to classify cellular and non-cellular structures in breast cancer histopathological images. Authors applied neural network to perform classification, where they achieved the best classification accuracy of 86.88%. Recently, Wang et al. [30] used Googlenet CNN model to perform Histopathological image classification, where they achieved the accuracy of 98.4% patch classification. To detect the invasive ductal carcinoma tissue in histological images for BC-CAD, Cruz-Roa et al. [31] applied CNN feature extraction approach. Ertosun et al. [32] developed BC-CAD model using Deep Learning with three distinct CNN models so as to localize masses in mammography images. To enhance accuracy of segmentation or ROI identification, authors incorporated additional novelties such as cropping, translation, rotation, flipping and scaling techniques. Arevalo et al. [33] applied CNN based feature extraction followed by SVM based classification for BC-CAD solution. Authors achieved Receiver Operating Characteristics (ROC) of 86%. A similar work was performed in [34], where authors achieved classification accuracy of 96.7%. Russakovsky et al. [35] and Zuiderveld et al. [40] trained CNN over ImageNet to perform breast cancer classification. Authors Abdel-Zaher et al. [36] and Wang et al. [41] performed breast cancer classification, developing a classifier by means of the weights of a previously trained Deep Belief Network (DBN). They applied Levenberg Marquardt learning based ANN to perform classification. Amongst the various researches it has been realized that in addition to the ROI identification and feature extraction efficiency, selecting optimal features is equally significant to ensure efficient performance. To achieve this, dimensional reduction and feature selection measures are of utmost significance. With these objectives, Olfati et al. [37] applied Genetic Algorithm (GA) to enhanced PCA by selecting optimal principal components analysis (GA-SPCA). They applied PCA for dimension reduction, GA for feature selection [38] and SVM for classification which facilitates to achieve better performance in order to gain desired results.

The developed system is much effective as it is prepared by utilizing advanced technology but it has a risk which may create major problems due to minor technical issues. Moreover, it is suggested that they can utilize Computerized tomography (CT) scan and Positron emission tomography (PET) scan to determine actual situation of breast cancer appropriately.

3. Proposed methodology

This section primarily discusses the proposed research work and implementation model to achieve intended BC-CAD solution. This solution provide an accurate breast cancer detection to assist early diagnosis in more efficient way. My contribution is all about to make highly robust and efficient CNN based on a model named as Deep Neural

Network (DNN) model. Moreover, it has been developed for breast cancer detection and classification. In the current research work, a novel deep learning based Breast Cancer detection and classification algorithm is developed. In our work, the predominant emphasis is made on developing a novel and robust automated CAD system for Breast Cancer (BC) detection and classification. Our research method applied on multi-dimensional shape to increase the performance of the proposed BC-CAD system, where enhancement has been made for major functional processes including pre-processing, ROI detection and segmentation, feature extraction, feature selection and classification. In addition, a standard breast cancer dataset named BreaKHis [39], which is the Histopathological image data for breast cancer, has been considered for our study. To ensure better efficiency enriching input data or medical image quality is often a better solution. With this objective, the input Histopathological images are processed for noise removal and resizing. Once retrieving the suitable input images, it has been further processed for ROI segmentation, where an enhanced GMM approach has been taken into consideration. Unlike generic GMM algorithm, we have derived an adaptive learning based Gaussian approach that enables swift and accurate ROI identification. This as a result can play vital role in accurate and significant feature retrieval for optimal BC-CAD solution. Noticeably, our applied datasets BreaKHis [39] possesses key features as marked cellular atypia, mitosis, disruption of basement membranes, metastasize …etc. These features have been applied to perform two class classifications. Realizing the non-deniable fact that inaccurate ROI localization and insignificant pixel conjuncture might lead inaccurate classification accuracy, we have applied Connected Component Analysis (CCA) model over segmented region. It eliminates those image components that percept to be connected with the target ROI; however, it is insignificant toward BC-CAD classification. Once obtaining the CCA processed segmented ROI, we have processed it for feature extraction followed by feature selection and classification to achieve anticipated BC-CAD solution. Unlike existing feature extraction models, such as Gabor filter [11], Wavelet Transform [8][9][10], MIL [26-28] or even classical CNNs [28-33][35], in our research a CNN model is derived. In the proposed research work, an enhanced DNN model named AlexNet DNN has been applied to extract significant features from the Histopathological image datasets. Realizing the fact that high dimensional features often provide more significant information to make precise classification or analysis; in our research work, we have developed AlexNet-DNN to extract high dimensional features with 4096 Kernels or dimensions. The proposed AlexNet-DNN model comprises five convolutional layers in succession with three Fully Connected (FC) layers (FC6, FC7 and FC8). The feature extracted at the higher layer provides more significant information that helps in achieving accurate BC-CAD performance. Considering major classical approaches, where the probability of over-fitting and accuracy is often ignored, we have applied a supplementary model called Caffe-enet that assists AlexNet-DNN to overcome existing limitations and enables its (AlexNet-DNN) execution over general purpose computers without any need of sophisticated Graphical Processing Units (GPUs). The use of Caffe-enet in conjunction with AlexNet-DNN plays a vital role in assuring efficient feature extraction even under large scale unannotated data. It makes a suitable environment for major breast cancer detection applications. Specifically, owing to high unannotated breast cancer data, performing DNN learning and further classification is a mammoth task. Hence to alleviate such issues, in our proposed model AlexNet applied multilayered DNN architecture, where at each layer we retrieve the features for further mapping and classification. However, to achieve more efficient results, we have considered 4096-dimensional features extracted at the Fully Connected (FC) layers, FC-6 and FC-7 of the AlexNet-DNN. The detailed discussion of the developed DNN model and its implementation for BC-CAD is given in the next section. In addition, to the proposed AlexNet-DNN based BC-CAD solution, in this research we have developed a parallel feature extraction model, called Scale Invariant Fourier Transform (SIFT). We have applied Fully-Connected (FC) layers: FC6 and FC7 features, which are 4096-dimensional features and hence amounts a huge data to process. To enhance computational efficiency, in this research work, we have applied PCA and LDA techniques to perform dimension reduction or feature selection. Once processing for the dimensional reduction, the selected features are projected to polynomial kernel based classifier that performs two class classifications: Malignant and Benign. To further strengthen the efficiency of our proposed work, 10-fold cross validation approach is applied that ensures optimal classification accuracy for BC-CAD solution as shown in Fig. 1.

The overall proposed research methodology and implementation schematic is given in the next section.

4. Proposed system design

In this section, the detailed discussion of the proposed research work and implementation schematic is presented. Before discussing the proposed BC-CAD solution, introducing key terminologies is must. Table 1 presents the nomenclature of different abbreviated terms.

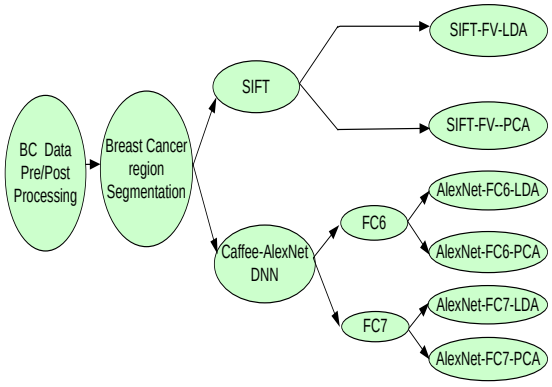


Fig. 1 Proposed BC-CAD System

A detailed discussion of the proposed research model is given, as presented in Fig. 2, as follows:

a) Breast Cancer Histopathological Image Data Collection
b) Pre-processing
c) ROI Detection and Localization
d) Feature Extraction
e) Dimensional reduction of Feature Selection, and
f) BC-CAD Two-Class Classification.
g) Classification

Which will be discussed in details:

a) Breast Cancer Histopathological Image Data collection

In this research or study, we have used standard datasets named Breast Cancer Histopathological Image Classification (BreakHis) [39] that contains a total of 9,109 Histopathological (say, microscopic) images of breast cancer tissue. The historical perspective of the BreaKHis states that the microscopic images are retrieved from 82 patients. BreakHis dataset has been built in association with the P&D Laboratory - Pathological Anatomy and Cytopathology, Parana, Brazil (http://www. prevencaoediagnose.com.br). Unlike generic singe size or single feature datasets, BreaKHis data are obtained with varied magnification factors such as 40X, 100X, 200X, and 400X. It shows the diversity of the data under consideration and the suitability of BC-CAD model with such data can make it applicable for different data types. BreakHis breast cancer dataset contains a total of 2,480 benign and 5,429 malignant Histopathological images with pixel-dimension 700X460, 3-channel RGB, 8-bit depth in each channel. The applied BreakHis datasets contains two distinct groups: benign tumors and malignant tumors. Noticeably, histologically benign signifies marked cellular atypia, mitosis…etc. In other words, it signifies a lesion having no probability of malignancy. Usually, benign tumors are stated to be "innocents", which grows gradually and often remains confined to a definite shape and size. On the other hand, malignant tumor refers the cancer. It refers a lesion that invades, destroys the neighboring structures and even spread across distant sites causing eventual death. The applied BreaKHis datasets has been obtained by means of partial mastectomy or excisional biopsy [39].

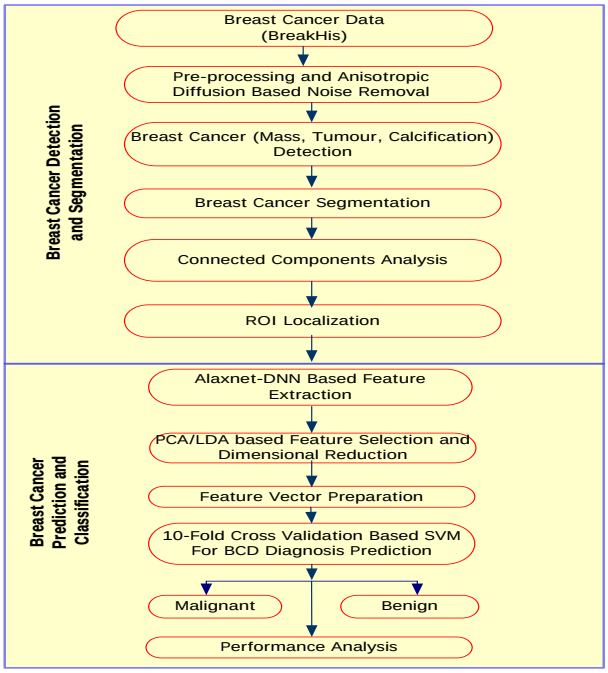


Fig 2. Proposed AlexNet-DNN based BC-CAD solution

The distribution of the data under study is given in Table 2.

Table 2: BreaKHis 1.0 is structure

Magnification	Benign	Malignant	Total
40X	652	1370	1995
100X	644	1437	2081
200X	623	1390	2013
400X	588	1232	1820
Total of Images	2480	5429	7909

Interestingly, both types of breast tumors can be sorted in distinct categories on the basis of the way the tumoral cells look under the microscope. Different types of the breast tumors can have distinct prognoses and treatment implications. Our data sets comprise the following distinct types of tumors. The table 3 shown about different types of tumors such as benign and malignant along with their examples including overall breast tumors which may or may not be cancer.

Table 3: Tumor containts (distribution) in BreaKHis 1.0 is

Benign Tumors: adenosis	Malignant Tumors
Fibroadenoma	carcinoma
phyllodes tumor	lobular carcinoma
tubular adenona	mucinous carcinoma
adenosis	papillary carcinoma

In the dataset, individual Histopathological image filename stores significant image information containing biopsy method, type of tumor, patient information (Patient ID), and magnification factor. For illustration, for an image SOB_B_TA-14-4659-40-001.png, as illustrated in Fig. 3, it signifies a benign tumor of tubular adenoma type, obtained at the magnification factor 40. The biopsy method used was SOB.

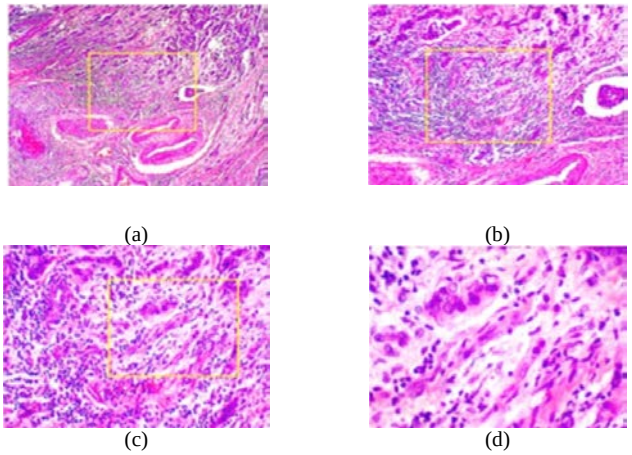


Fig 3. Snippet of the breast malignant cancer at different magnification factors: (a) 40X, (b) 100X, (c) 200X, and (d) 400X [39].

b) Pre-processing

Non-deniably, the quality of data influences the final classification accuracy and efficiency. To enable suitable data for further processing, we have performed pre-processing. In our research model, the pre-processing model is exceptionally significant to ensure precise nuclei and allied vicinity identification and characterization. In breast cancer detection, the features, such as the shape and size of nuclei, glandular shape …etc. is expected to be precisely identifiable. Performing pre-processing assures that suppression of the original image qualities required to estimate or identify texture representation and other significant features, precisely [40]. Realizing interchangeable nature of the pre-processing approaches, we have performed pre-processing before performing candidate region detection and feature extraction [40-41]. In our paper, we have followed the pre-processing instructions provided in [35][41] Once performing pre-processing, the concept region identification or segmentation has been processed. In addition to the above mentioned, in pre-processing, we have resized input breast cancer Histopathological images to 224X224. A brief of the proposed segmentation model or ROI identification approach will be explained as follows.

c) Concept Region Identification

In this section, the process of breast cancer region detection and segmentation is discussed. Unlike traditional approaches of segmentation such as Otsu's method [42], our model applies an enhanced GMM technique. Here, we segment cancer tumor region from the background so as to distinguish it from other regions over Histopathological images. GMM has exhibited better performance over other state-of-the-art techniques for background subtraction and ROI identification because of b a pixel-wise segmentation approach. In addition, GMM robustness causes to be used for segmenting ROI from the background region that can be further processed for feature extraction over the detected ROI. A brief of the proposed GMM model for ROI detection including several steps is given as follows:

- Adaptive-Learning-Based GMM for Background Subtraction

The typical Histopathological images exhibits highly complex color and textural features that makes classical segmentation approaches, such as Otsu method, static thresholding based segmentation and even generic GMM confined to enable optimal ROI detection. One of the predominant reasons is the static learning rate [43]. To deal with such limitations, in this research model, we have developed an enhanced adaptive learning based GMM model for ROI detection and segmentation. Consider y is the pixel-value at certain instant; GMM can be used to estimate Probability Density Function (PDF) of y. Noticeably, PDF can have the all connected Gaussians. Here, the PDF of the Gaussian mixture $g(y)$ with N components is estimated as:

$$g(y) = \sum_{n=1}^{N} w_n M(y; \mu_n, \varsigma_n), \qquad (1)$$

Where w_n signifies weight factor, $M(y; \mu_n, \varsigma_n)$. Here, y represents the output with respect to the values of μ_n, and ς_n at nth instant. Here y refers the normalized density of mean μ_n. In equation (1), the variable $\sum_n = \varsigma_n J$ represents the covariance matrix. In fact, this condition hypothesizes that the pixel values of the color components (i.e., red, green, and blue) in the breast cancer image are independent and have equal variances. On the contrary, it is infeasible and therefore the distribution of recently occurred pixel values in the Histopathological image can be defined in the Gaussian mixture form. In our model, the value of a new pixel is characterized by means of one of the predominant components of the mixture model, which is further applied to update the model. Stauffer et al. [43] applied this concept to estimate background where initially they initialized these variables with zero. Once detecting any similarity, such as $\|y - \mu_i\|/\varsigma_i < T_h$, with $i \in$

$[1, \dots, N]$ and $T_h(> 0)$ as a threshold, the operating variables of the GMM model are obtained as:

$$w_n(t) = (1-\beta)w_n(t-1) + \beta P_n(t) \quad (2)$$

$$\mu_n(t) = (1-\alpha)\mu_n(t-1) + \alpha y \quad (3)$$

$$\varsigma_n{}^2(t) = (1-\alpha)\varsigma_n{}^2{}_{(t-1)} + \alpha\left\|\left(y-\mu_n(t)\right)\right\|^2, \quad (4)$$

where $\begin{cases} P_n(t) = 1 & with\ matching\ element\ i \\ P_n(t) = 0 & others. \end{cases}$

If no element matches together then the element with the minimal w_n is re-initialized, and the variables (2-4) are updated as:

$$w_n = w_0, \mu_n = \mu_0, \varsigma_n = \varsigma_0. \quad (5)$$

So, α is obtained as:

$$\alpha = \beta N(x; \mu_n, \varsigma_n) \quad (6)$$

In above expression (6), β refers to the learning rate
In our proposed GMM model, we normalize $w_n$in a way that it converges toward unit value (i.e., 1). To obtain background region, authors [43] sorted Gaussians w_n/ς_n in decreasing order by using threshold τ in conjunction with the sums of the weights. As a result, this achieves a set$\{1, \dots, R\}$, where B is obtained as:

$$R = arg\min_{N_R}\left(\sum_{n=1}^{N_R \le N} w_n > \tau\right). \quad (7)$$

With decrease in variances, usually the Gaussian distribution gains increases. This as a result makes it more evident to perform background subtraction. Once estimating GMM parameters, sorting is performed from the matched mixture distribution toward most probable background. This is feasible only because only the relative value of the matched model changes. In our applied model, ordering is performed in a way that the most feasible background distributions remain on the top, while the low probability distributions move towards bottom. In such manner, the initial value of (7) is considered as the background. In (7), parameter τ signifies the result of the least section of the data responsible for the background. With the minimal possible value of τ, the background model is stated to be unimodal, where using only the most probable distribution enhances computational efficiency. On the contrary, the high τ signifies multimodal distribution, is typically caused due to the iterative background motion. In this paper, the breast cancer Histopathological images with static features are taken as input and hence the use of unimodal approach with low τ ensures computationally efficient process. In our approach, the Gaussian with the maximum w_n and minimal $\varsigma_n$refer to the background region. Thus, applying aforementioned approach the background region can be identified that consequently helps in segmenting the ROI. Fig. 4 shown an examples of real textures present in histopathological images (HE staining).

In major cases, classical GMM updates μ_n, ς_n with certain fixed learning rate α [43] are used. While, considering breast cancer detection requirement, the static learning rate approach seems questionable, particularly when there can be excessive texture variation over the breast cancer Histopathological images. To deal with such limitation, authors [44] developed an adaptive learning rate α based GMM for ROI segmentation over varying surface or texture conditions. Non-deniably, there can be certain texture or the surface conditions, where the pixel might neither be the breast can tumor (i.e., ROI) nor the background; however, it could be classified as either type (mention here the two types). This as a result may influence accuracy of the proposed BC-CAD solution. Increasing learning rate α might cause high-rate pixel feature changes [44] that consequently may make breast cancer tumor or ROI detection susceptible. Though, the adaptive learning based GMM model [43] can perform ROI identification under varying background conditions. However, the very minute and fine-connected texture or surface features still make existing approach limited. Precisely for breast cancer Histopathological images, textural or structural features, such as cellular atypia, mitosis, disruption of basement membranes, metastasize …etc. are highly intricate and minute feature that requires more efficient ROI identification model. An enhanced Gaussian mixture learning based GMM model was proposed in [44]. However, authors could not address the saturation scenario particularly during convergence. (starting from this section till the end of the paragraph, you repeat yourself would you revise it) As an enhanced solution, in this paper an adaptive learning based GMM model is developed that detects precise ROI. In our proposed approach, initially, we have decoupled learning rate α from other components such as μ_n and ς_n. We have applied a new learning rate variable called adaptive learning rate $\delta_n(t)$ that updates mean component μ_n using a probability parameter $B_n = M(y; \mu_n, \varsigma_n)$. Here, B_n states whether a pixel is a part of the nth Gaussian distribution or not. In our model, the adaptive learning rate parameter δ_n is estimated as follows:

$$\delta_n(t) = \delta_n(t-1) + \frac{N+1}{N}B_n - \frac{1}{N}\sum_{j=1}^{N} B_j. \quad (8)$$

Where, N states the number of distributions.
$\delta_n(t)$ Can provide fast and accurate Gaussian update that may enhance the system to adapt with swift textural variations. As a result, this can enable optimal ROI detection for the targeted BC-CAD solution.

Now, substituting δ_n as α in (3), it is observed that variance updating might avoid Gaussian saturation. On the other hand, the fast learning rate might cause degeneracy [44]. To deal with this issue, we have developed a semi-parametric model to estimate the variance that can effectively enable quasi-linear adaptation. It is significant for the condition where even a minute change from the average and certain reduced response might cause decisive variation. To achieve it, in our model we have applied a sigmoid function (9):

$$s_{p,q}(y,\mu_n) = p + \frac{q-p}{1+e^{-UV(y,\mu_n)}}, \qquad (9)$$

In (9), the variable $V(y,\mu_n) = (y-\mu_n)^T(y-\mu_n)$ refers to a controller called sigmoid slope controller (SSC). Now, substituting (9) in (2), we obtain the variance update as (10).

$$\varsigma_n{}^2(t) = (1-\varphi)\varsigma_n{}^2{}_{(t-1)} + \varphi s_{p,q}\big(y,\mu_n(t-1)\big), \qquad (10)$$

In our study, performing try and assess approach, we find $\varphi = 0.6$ suitable to perform better. To estimate φ, we have applied a mathematical approach, where the two real positive variables p and q are selected in such manner that B expand over one nth of the pixel range. Here, $s_{p,q}(y,\mu)_{B^+}$ caps ς_n to a range $B \in \left[\frac{p+q}{2}, r\right]$. In (9), we have used a new variable φ rather using α as applied in (2) -(4). Thus, applying above discussed proposed background subtraction model the breast cancer ROI is detected and segmented to further process connected component analysis and feature extraction. Fig. 5 (a) represents the breast cancer region (i.e. breast cancer ROI) segmentation process. In Fig. 5(b), the first column refers the original breast cancer Histopathological image from the BreakHis dataset. The other two consecutive columns signify the gray images with noise filtering as processed through [41]. Column 4 presents the subtracted BC-ROI regions and allied internal geometries.

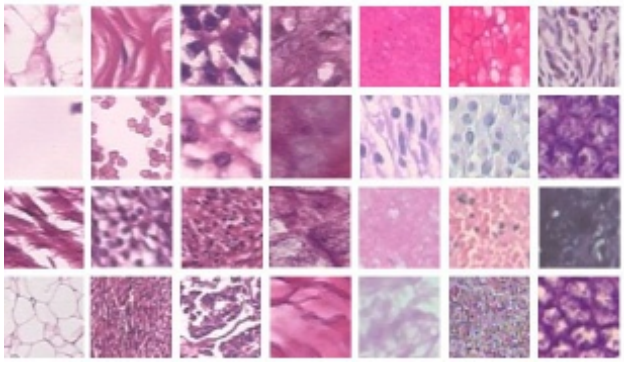
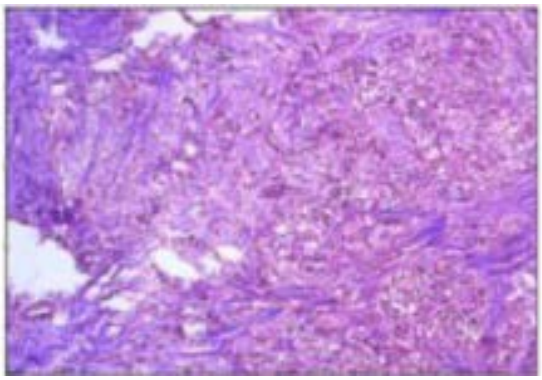


Fig. 4 Examples of real textures present in histopathological images (HE staining).



(a)

(b)

Fig. 5 (a) Example of breast malignant tumor acquired at 40× magnification and (b) 32 × 32 patch images.

d) Breast Cancer Region Localization

Once performing ROI segmentation, we have applied CCA over the segmented breast cancer region that considers the region, size, and location of the key BC features such like mitosis, marked cellular atypia, disruption of basement membranes, metastasize, and their shape, size, and proximity to the normal breast regions. In our model, we consider that the disconnected region referring to the Gaussian components belongs to the breast cancer ROI. The significant feature information such as the dimension and proximity to the BC ROI has been normalized. In our model, the normalized dimensional feature is estimated as the width of the connected component region divided by the width of the cancer regions at the centroid of the connected region. In this manner, using the normalized width it becomes easier to compare the sizes of marked cellular atypia, mitosis, disruption of basement membranes, metastasize …etc., at distinct locations on the breast cancer Histopathological image. Thus, applying above mentioned enhanced GMM and the CCA approaches of accurate BC-

ROI has been identified for further feature extraction. The following section briefs the proposed CNN based feature extraction model for BC-CAD solution.

Feature Extraction

Once identifying or localizing the BC-ROI the detected region has been projected further for feature extraction. Unlike traditional approaches, in this research work we have applied AlexNet-DNN model for feature extraction and learning. In addition, our proposed AlexNet-DNN model functions in conjunction with the Caffe-enet DNN that makes feature extraction even for large size unannotated data over general purpose computers. A brief of the proposed Caffe-enet-AlexNet-DNN model for BC-ROI feature extraction is given as follows:

AlexNet DNN

Unlike generic feature extraction techniques, such as wavelet transform or Gabor filter …etc., we have applied a recently developed and robust DNN model called AlexNet. In fact, AlexNet was at first developed by Alex Krizhevsky for color image classification, where it was trained over millions of images [61]. However, its efficacy over other approaches enabled it to be used in major image classification problems. In fact, AlexNet-DNN represents a multilayered DNN architecture that applies CNN concept to learn features over ImageNet [45]. The transferable nature of AlexNet enables it to be used for major image classification applications, including medical data analysis and classification [46]. Considering these robustness, in our model we have applied AlexNet-DNN to perform feature extraction. In our proposed model, the transferability of AlexNet has enabled the extracted features to be projected over ImageNet [45], which has been further processed for training and learning. The proposed AlexNet DNN model enables learning of the rich midlevel breast cancer image features and corresponding semantic representations that signify breast cancer perception. As already stated, to enable ease of implementation of the AlexNet DNN model over general purpose computers, we have applied Caffe-enet-DNN [47]. It does not only enable easy implementation but also ensures fast feature extraction over large size unannotated data. A typical presentation of AlexNet-DNN architecture is given as follows:

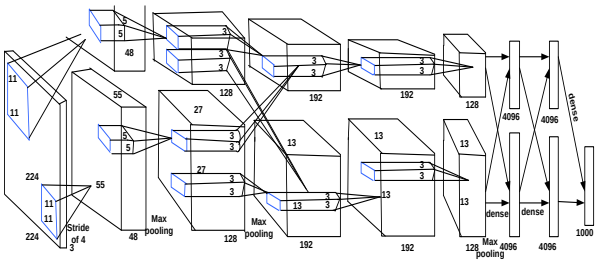


Fig. 6 AlexNet CNN architecture. Extracted from [48].

- ***Input Layer:***

In AlexNet-DNN, this layer takes data as input and generates output which is further fed as input to the convolutional layers. To assure better performance, a few transformations like feature scaling and mean-subtraction can also be performed at this layer. In our proposed model, the inputs are the segmented images while the parameters state its dimension (i.e., 32 ×32 or 64 ×64 pixels) and the total channel numbers (i.e., for RGB, 3).

- ***Convolutional Layers:***

This layer primarily convolves the input image with certain filters, each generating single feature map in the output image. In our proposed work, AlexNet-DNN with five convolutional layers has been developed. Here, the receptive fields, also known as kernels of respective size 5 × 5 are taken into consideration and the zero-padding is fixed at 2. In addition, at convolutional layer the stride is fixed at 3 which is known as long term decisive steps that are required to be put in a correct direction. In our proposed AlexNet-DNN model, the Convolution layers (CONV) have the kernels specification as: CONV1-96 kernels, CONV2-256 kernels, CONV3-384 kernels, CONV4-384 kernels and CONV5-256 kernels. Moreover, these kernels are specified as all of them has single accurate feature which is mandatory for creation of an output image.

- ***Pooling Layers:***

These layers in AlexNet usually functions for down-sampling the spatial dimension of the input provided. In our proposed model, we have incorporated one pooling-layer after each CONV layer. Here, each of these is defined to apply a 3×3 receptive field having a stride of 3. Here, receptive filed states spatial extent. Usually, the first pooling layer applies the most generic max-operation over the defined spatial extent, while the other remaining exhibit average pooling.

- ***ReLU Layers:***

In addition to the above mentioned layers, there is a layer called ReLU that typically functions as an activation function. In fact, it is a non-linear element-wise operator that functions as a layer. In AlexNet, there are three distinct ReLU layers. Provided an input y, the ReLU layer estimates the output of the neuron $q(y)$ as y if $y > 0$ and $(\delta \times y)$ if $y <= 0$. Here, δ states whether to escape the negative component by performing multiplication with it with slope (such as 0.01…) or setting it to 0. In our case, we have set $\delta = 0$ as the default case. In case, δ is not defined it behaves similar to a classical ReLU function $q(y) = \max(0, y)$. In another way, in this case the activation is performed at the zero threshold value.

- ***Fully Connected (FC) Layers:***

In AlexNet, FC layer usually treats input data as a simple vector and generates output as a single vector. In our applied AlexNet-DNN model, three FC layers are applied: FC6, FC7 and FC8; they have kernels of 4096, 4096 and

1000, respectively. In other words, FC6-4096 kernels, FC7-4096 kernels and FC8-1000 kernels are used in proposed AlexNet-DNN because three layers are applied including FC6, FC7, FC8 are combined with selected kernels to form the same. The last one layer is a fully connected output layer with softmax activation, which primarily depends on the total number of classes required for classification. Here, we have considered two class classifications which is generally known as binary classification. Hence, there are two output filters for the binary classification of the results.

Now, applying the derived DNN model, the feature extraction has been performed. A brief of the AlexNet-DNN based feature extraction is given as follows:

1) AlexNet DNN-Based BC-ROI Feature Extraction

In our BC-CAD model, AlexNet DNN model is applied for extracting 4096 dimensional features from breast cancer Histopathological images (particularly, BC-ROI localized images). For ease of implementation on general purpose computer, Caffe-enet-DNN has also been applied in conjunction with AlexNet. Here, Caffe-enet is trained over the localized BC-ROI features of the used BC dataset. A simplified model of the applied multi-layered AlexNet-DNN is given as follows:

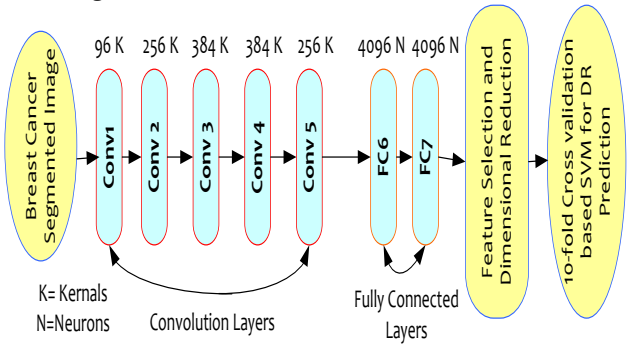


Fig. 7 Simplified AlexNet-DNN Convolutional layers

In typical AlexNet model, the features extracted at the low layer are similar to that of extracted through the classical approaches, such as Gabor information, blob features …etc. On the contrary, the features extracted at the higher layer usually comprise significant information that play vital role in targeted classification purposes (here, BC detection and classification). In our work, we have considered features at the FC6 and FC7 layers of the AlexNet DNN model. As already stated in above discussion, the features can be extracted at five CONV layers (i.e., CONV1–CONV5) and the FC layers (i.e., FC6 and FC7) of AlexNet provide extracted BC-ROI features. As depicted above, the individual CONV layer comprises multiple kernels that define a 3D filter connected to the outputs of the preceding layer. As depicted in Fig. 7, FC layers comprise multiple neurons having the real positive value and each neuron is connected to the all neurons of the previous layer. In our model, FC6 and FC7 features are considered that possess a total of 4096-dimensional features to be further processed for classification. Once retrieving the features, it is projected to the feature vector $F_V = (f_1, f_2, f_3, \ldots, f_{4096})$. This is because the extracted features have the huge quantity and hence there can be the probability that many instances could have the same significances or do not give different meaning. Hence, suppressing those insignificant features can be vital to enhance accuracy as well as computational efficiency. In our work, once obtaining the features, it has been projected further for feature selection or dimensional reduction.

In addition to the proposed AlexNet DNN based feature extraction model, we have developed a parallel approach called SIFT to extract features from the segmented BC-ROI Before discussing the feature selection or the dimensional reduction approaches, a brief of SIFT based BC-ROI features extraction is presented as follows:

2) SIFT-Based BC-ROI Feature Extraction

Scale Invariant Feature Transform (SIFT) is an algorithm in computer vision, which detects and identifies local features in an image for further decision. Typically, SIFT landmark points or the keypoints of ROI are retrieved from a reference image, which are then stored in a database. Thus, the ROI is detected in a new image by performing comparison of the individual feature from the new image to this database. It then finds the candidate matching features on the basis of Euclidean distance of their feature vectors. Thus, from the available complete set of matches, keypoint-(in subsets) agreeing the ROI or object and its position, or other features, such as scale, orientation…etc., in the new image are obtained to retrieve good matches. The estimation of constant clusters is performed rapidly by using an efficient hash table implementation of the generalized Hough transform. Each cluster of 3 or more features that agree on an object and its pose is then subject to further detailed model verification and subsequently outliers are discarded. Finally, the probability that a particular set of features indicates the presence of an object is computed, given the accuracy of fit and number of probable false matches. Object matches that pass all these tests can be identified as correct with high confidence.

To perform SIFT based feature extraction, at first SIFT descriptors have been obtained for each BC-ROI or the breast cancer Histopathological image. Here, the individual SIFT feature descriptor signifies a 128-dimensional feature vector. In our proposed work, only half of the 128-dimensional feature vectors are considered. In other words, only 64 dimensions are considered for features analysis. In our model, we have processed Fisher encoding with the 32 Gaussian distributions that as a result generates a 4096-dimensional Fisher vector equivalent to the AlexNet DNN features at FC6 and FC7.

Once, the features have been extracted from the datasets, it has been processed for the feature selection or the dimensional reduction. A brief of the proposed feature selection or the dimensional reduction approaches is given as follows.

Feature Selection and Dimensional Reduction

In this paper, we have applied two feature selection or the dimensional reduction approaches: named Principle Component Analysis (PCA) and Linear Discriminate Analysis (LDA). In major existing approaches, PCA based feature selection approach has been applied; however, the efficacy of LDA towards most discriminating features makes process more efficient. Considering this fact, we have applied both algorithms distinctly and respective performance has been assessed. A brief of these techniques is given, as follows:

Principle Component Analysis (PCA)

It mainly projects feature space of a high dimension to the lower dimension while ensuring that the obtained feature space contains most significant features. In function, PCA rotates the p-dimensional feature space's axes to a new position called principle axes in such way that the principle axis 1 has the highest variance followed by axis 2, and so on. In practice among extracted breast cancer features there can be the highly correlated feature attributes of the elements. These highly correlated features (sometimes called homogeneous feature) are known as the most expressive features (MEF). In this case, PCA intends to transfer the feature attributes or the elements to a new feature set or feature vector having no correlation among the elements. Thus, applying this approach PCA reduces a significant amount of feature elements that makes computation more efficient. The finally retrieved feature vector is then projected for classification.

Linear Discriminant Analysis (LDA)

Unlike PCA that uses MEF to perform feature selection or the dimensional reduction, LDA employs the most discriminating features (MDF) to perform dimensional reduction. LDA performs dimensional reduction automatically. Executing PCA, the LDA dimensional reduction approach projects extracted features to a single principle component regardless of its class label. To perform feature selection and the dimensional reduction, LDA applied two distinct matrix called intra-class scatter matrix D_{ISW} and inter-class scatter matrix D_{IOS}. These matrix values are estimated, as follows:

$$D_{ISW} = \sum_{i=1}^{T} \sum_{j=1}^{S_{c_i}} (y_j - v_{ci})(y_j - v_{ci})^T \qquad (11)$$

$$D_{IOS} = \sum_{i=1}^{T} (v_{ci} - v_c)(v_{ci} - v_c)^T, \qquad (12)$$

In (12), T refers the total number of classes, v_{ci} signifies average vector of class i, and S_{c_i} presents the overall samples in class i. We estimate the v_{ci} as

$$v_c = \frac{1}{T} \sum_{i=1}^{T} v_{ci}, \qquad (13)$$

In (13), T refers the total number of classes.

LDA intends to increase the inter-class scatter and decrease the intra-class scatter. In our model, it was achieved by increasing a factor $\mathcal{F}_i$, given in equation (14).

$$\mathcal{F}_i = \frac{\Delta|S_B|}{\Delta|S_w|} \qquad (14)$$

For a non-singular D_{IOS}, F_i is increased provided the column vectors of the projection matrix P are the eigenvectors of $D_{ISW}{}^{-1}D_{IOS}$. Here, the projection matrix P with $T-1$ dimension projects the training data to form Fisher vector (FV). In this way, the final Fischer Vector obtained is given as input to the classifier. Noticeably, the overall Fischer Vectors projected for classification be $\mathcal{F}_{VC} = (\mathcal{f}_{1C}, \mathcal{f}_{2C}, \mathcal{f}_{3C}, \dots, \mathcal{f}_{4096C})$.

Once obtaining the final feature vectors, it is processed for breast cancer classification. In our proposed work, a two class classification has been done: BENIGN and MALIGNANT. To further enhance performance, 10-fold cross validation has been performed. A brief of the classification method used is presented as follows:

Classification

Considering ease of implementation and better non-linear pattern learning efficiency, in our model a Radial Basis Function (RBF) kernel based SVM classifier is applied to perform two-class BC-CAD classification. In our model, SVM has been trained over extracted feature vectors that intend to achieve the maximum margin by obtaining low m as given in equation (15).

$$\frac{1}{2} m^T m + e_r \sum E_i, \qquad (15)$$

Where, $E_i \geq 0$ and e_r signify the extent of error resiliency.

In our model, to perform two-class classification, FV is grouped in the labeled pairs, say K_i (s_i, t_i) where s_i presents the training vector. The class label of the training vector s_i is given by $t_i \in \{-1, 1\}$. In classification, the hyper plane classifies the maximum possible points of the same class on the same side of hyperplane. The use of 10-fold cross-validation ensures higher accuracy and reliability of the results. The overall performance of the proposed BC-CAD model is examined in terms of classification accuracy and the obtained results are compared with other state-of-art techniques. The results obtained in this research work are discussed in the next section.

5. Results and Discussion

To assess the performance of our proposed BC-CAD solution, we consider BreakHis dataset [39] that contains 9,109 Histopathological (say, microscopic) images of breast cancer tissue with different tumor types or associated traits, such as fibroadenoma, phyllodes tumor, tubular adenona, adenosis, carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma. For ease of implementation, we consider random 2000 Histopathological images for our study. The original Histopathological images are in the size of 700X460, which is further reduced using resampling approach (i.e., pixel area relation) and finally the images with 350X230 dimensions are obtained. It is then followed by the extraction of patches from the considered Histopathological images. In our work, we apply a sliding window approach with 50% overlapping. In addition, the patches can be obtained randomly too without incorporating any overlapping control between patches. In our simulation test, the patches with 32X32 dimensions are obtained. As pre-processing, we incorporate image resizing and noise filtering. Unlike generic GMM based background subtraction, we develop a new adaptive learning rate based GMM method for ROI subtraction, which is followed by CCA to ensure optimal ROI localization. The obtained ROI are then processed for CNN based feature extraction, where AlexNet-DNN in conjunction with Caffe-enet is used to extract and learn breast cancer features. The transferability feature of AlexNet DNN is used to feed Histopathological images as input. The multilayered architecture of AlexNet DNN (with 5 convolutional layers and 2 fully connected layers) enabled us to retrieve high dimensional features are the fully connected layers; FC6 and FC7. In our developed CNN architecture, we retrieved 4096-dimensional features at the FC6 and FC7 layers. In addition to the AlexNet-DNN based features extraction, we apply SIFT feature extraction model to obtain equivalent features for performance comparison. Unlike AlexNet-DNN, SIFT at first generate 128-dimensional features, which was then projected onto PCA, where we use only half of the initial principle components i.e., 64 dimensions connected to 32 Gaussian components distribution. Finally, we obtain a 4096-dimensional feature vector for SIFT feature extraction model. Once retrieving 4096-dimensional features, it is processed for PCA and LDA for dimensional reduction. It is then followed by SVM classification with 10-fold cross-validation. The overall simulation model is developed using MATLAB 2015a with VLFeat-0.9.20 image processing toolbox. The overall performance of the proposed breast cancer CAD (BC-CAD) solution has been examined in terms of classification accuracy and area under Region Operating Characteristics (ROCAUC) parameters. Results are illustrated in Table 4.

Table 4: AlexNet-DNN and Sift based BC-CAD classification accuracy

	Features	Classification Accuracy (%)
AlexNet-FC6	FC6-PCA	96.15
	FC6-LDA	96.85
AlexNet-FC7	FC7-PCA	96.20
	FC7-LDA	96.70
SIFT	SIFT-PCA	95.70
	SIFT-LDA	95.75

Table 4 presents the performance of the proposed AlexNet-DNN based feature with sub-sequent feature selection and SVM based classification. As depicted above, FC6 features with LDA feature selection and SVM classification exhibits better classification accuracy (96.85%) than the PCA based features (96.15%). On the other hand, our proposed BC-CAD model with FC7 features and LDA feature selection or dimensional reduction has exhibited accuracy of 96.70%. On the contrary, with the same FC7 feature, followed by PCA based feature selection and SVM classification has yield classification accuracy of 96.20%. On the other hand, SIFT based features with PCA feature selection has shown accuracy of 95.70%, while the same with LDA has exhibited 95.75%. Observing these all results, it can be found that the proposed AlexNet-DNN based FC6 feature with LDA dimensional reduction and 10-fold cross validation assisted SVM outperforms other state-of-art possibilities. To examine performance, confusion matrix is obtained for each simulation model. The feature based accuracy comparison too reveals that AlexNet-DNN features give better classification results than the SIFT. A graphical snippet of the performance observed is given as follows, as illustrated in Fig. 8.

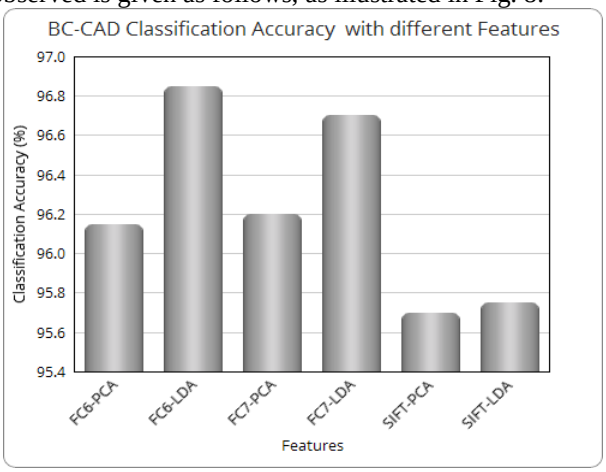


Fig. 8 Classification accuracy with different features

In addition to the developed BC-CAD solutions, we have compared it with different other approaches. Table V presents the comparison of our proposed BC-CAD solution with other state-of-the-art techniques for breast cancer detection. Table V presents comparative results of the different breast cancer detection systems, including CNN,

wavelet transform and other approaches. A wavelet transform technique was used in [8] to perform breast cancer detection, where authors achieved the maximum classification accuracy of 90.48%. In [9], authors applied wavelet and counterlet transform technique for breast cancer detection; however, their proposed system could achieve the maximum classification accuracy of 82.66%.

Table 5. Performance comparison

Techniques	Classification Accuracy (%)
[8]	90.48
[9]	82.66
[10]	96.00
[49]	90.00
[50]	88.03
[51]	95.03
[52]	88.00
[53]	80.00
[54]	91.02
[55]	(Precision) 65.03
[56]	92.40
[57]	97.34
[58]	87.00
[59]	94.00
FC7-PCA	96.20
FC7-LDA	96.70
SIFT-PCA	95.70
SIFT-LDA	95.75

A similar approach with wavelet transform based feature extraction developed in [20] for breast cancer detection exhibited breast cancer classification accuracy of 96%. However, authors could not examine the generality of their approach for large scale datasets. Authors in [49] applied CNN for breast cancer classification, where they obtained the maximum classification accuracy of 90%, which is lower than our proposed AlexNet-DNN based BC-CAD performance (96.85%). Interestingly authors [49] applied the same breast cancer Histopathological image; however, the generic approach of CNN confined its performance till 90%. Comparing this result with our developed model, it can be found that the use of GMM background subtraction, CCA based localization and feature selection along with n-fold cross validation strengthened our proposed model to exhibit better. In [50], a transferable learning approach was developed for CNN which was used to perform nuclie detection and classification over Histopathological images. In their approach, even applying AlexNet, authors [50] could achieve the maximum classification accuracy of 85.68%. Similarly, in [51] mammogram classification was performed using CNN. Wish a small dataset, authors managed to achieve the maximum accuracy of 92.30%. In [52], authors performed mitosis detection for breast cancer classification using DNN concept. Table 5 presents that the maximum classification accuracy obtained by [52] is 88%, which is lower than our proposed system. CNN was applied in [53] to classify cytological specimens in breast cancer; however, they could achieve the maximum accuracy of 80%. A recently developed deep multiple instance learning model was developed in [54]. Authors enhanced their approach with sparse label assignment of the mammogram. This as a result provided maximum accuracy of 91.02%. However, its efficacy with large scale real-world data remained unexplored. In [55], authors developed a joint multiple instance multiple label learning model for breast cancer diagnosis, where they exploited the concept of MIL to perform feature extraction followed by K-NN based clustering to perform breast cancer detection and classification. Authors could manage the average precision of 65.03%. Similar to our proposed approach was developed in [56], where authors applied AlexNet, Googlenet and Baseline approach to perform breast cancer detection. Authors could achieve the maximum classification accuracy of 89.0%, 92.9% and 60.4% with AlexNet, Googlenet and Baseline techniques, respectively. Recently, a hybrid segmentation model based DNN concept was derived in [57], where authors amalgamated Hough transform and DNN for tumor segmentation and feature extraction. This robustness could manage to achieve an appreciable performance (accuracy of 97.34%). However, they tested their approach with merely 40 images, and hence its efficacy for a large scale data remains questionable. With their simulated conditions, the generalization of their approach can't be done for major test cases. In [58] authors developed a deep CNN model to predict semantic characteristics of the breast cancer lesions in mammogram. Their approach could achieve the maximum classification accuracy of 87%. Similarly, in [59] authors developed a deep learning approach to perform breast cancer feature learning from mammogram. The prime novelty of their approach was the two-step training that comprised regression learning assisted pre-training to calculate large scale features. The other step dealt with the affine tuning model to perform better classification.

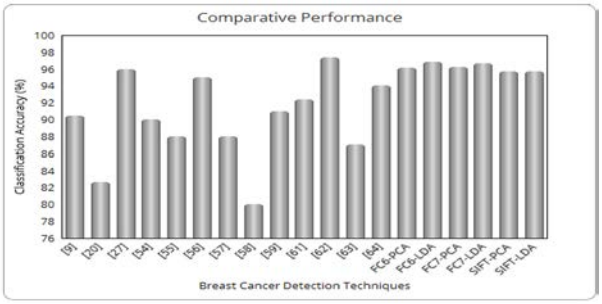


Fig. 9 Comparison of the results obtained

Thus, observing the performance of the various related works done so far, it can be found that the proposed research work exhibits better performance than other available stat-of art techniques. Therefore, the proposed work can be recommended for similar application. As shown in Fig. 9.

6. Conclusion and future work

In this paper, a CNN based breast cancer detection model was developed. The overall proposed model can be stated a multi-level optimization measure, where effort is made on enhancing each functional components of the BC-CAD model. At first, to develop a BC-CAD solution, BreakHis data, a standard breast cancer Histopathological data was taken into consideration. The proposed BC-CAD model has applied key enhancement, such as enhanced adaptive learning based Gaussian Mixture Model for region of interest, ROI (i.e., tumor) segmentation, connected component analysis based ROI localization, AlexNet-DNN based high dimensional feature extraction, feature selection using PCA and LDA, and 10-fold cross validation based classification. The use of enhanced GMM has strengthened proposed model to ensure optimal tumor region identification and localization, while the transferable deep neural network model, Caffe-enet-AlexNet DNN has enabled flexible implementation over general purpose computer. The implementation of feature selection approaches like PCA and LDA has played vital role in enhancing computational efficiency of the proposed BC-CAD solution. Further, SVM classification with 10-fold cross validation too has played significant role in enhancing classification accuracy of the proposed system with 96.15 for AlexNet-FC6 and 96.20 for AlexNet-FC7. The overall enhanced functional components as a cumulative BC-CAD solution has exhibited better than other state-of-art techniques, including spatial invariant feature transformation, wavelet transform, generic CNN …etc. In future work, the proposed approach can be assessed with more sophisticated GPUs such as NVidia …etc. to get more time efficient process to yield an optimal BC-CAD solution.

DECLERATION

Ethics approval and consent to participate No ethics approval or consent to participate was needed for this work, as it did not involve direct human material, human subjects or human data. Consent for publication No consent for publication was needed for the data used in this work. Competing interests the authors declare that they have no competing interests.

Availability of data materials: Not applicable

Funding: This study was funded by northern Border University project (SCI -2018-3-9-F-7782)

Conflict of Interest: No conflict of interest exists

Informed consent: Informed consent was obtained from all individual participants included in the study.

Authors' contributions: The present study provides a systematic review of accuracy to use Raman spectroscopy as a method for caries detection at an early stage. Absence of sample penetration makes the method simple and hence is widely used. Significant information related to Raman spectroscopy has been extracted and utilized in the presentation of systematic review paper. Various parameters have been taken into account, such as type of Raman spectroscopy, central wavelength, optical power, description of the system, scan rate, description of Raman micro-spectroscopy, Raman imaging, Raman peak, Peak intensities of laser polarization direction, Depolarization ratio and polarization anisotropy.

All authors read and approved of the final manuscript.

Acknowledgements: Acknowledgements to NORTHERN BORDERS UNIVERSITY and Deanship of Scientific Research

References

[1] World Health Organization, “Breast cancer: prevention and control,” Jan 2016.

[2] Australian Institute of Health and Welfare & Cancer, Australia, “Breast cancer in Australia: an overview,” Cancer series, no. 71, Cat. No. CAN 67, Canberra: AIHW, 2012.

[3] P. Boyle and B. Levin, Eds., World Cancer Report 2008. Lyon: IARC, 2008. [Online]. Available: http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/wcr_2008.pdf

[4] S. R. Lakhani, E. I.O., S. Schnitt, P. Tan, and M. van de Vijver, WHO classification of tumours of the breast, 4th ed. Lyon: WHO Press, 2012.

[5] Tabar L., Dean P.B., “Mammography and breast cancer: the new era,” Int J Gynaecol Obstet (PubMed: 14499978), 82, no. 3, pp. 319–326, 2003.

[6] Rankin, S.C., “MRI of the breast,” Br J Radiol, 73, pp. 806–818, 2000.

[7] P.L. Davis, M.J. Staiger, K.B. Harris , M.A. Gannot , J. Klementaviciene , K.S. MacCarty , Jr., H. Tobon. “Breast cancer measurements with magnetic resonance imaging, ultra sonography, and mammography,” Breast Cancer Res. Treatment, vol 37, pp. 1–9, 1996.

[8] S. Pramanik, D. Bhattacharjee and M. Nasipuri, "Wavelet based thermogram analysis for breast cancer detection," 2015 International Symposium on Advanced Computing and Communication (ISACC), Silchar, pp. 205-212, 2015.

[9] Y. Nasiri, M. Hariri and M. Afzali, "Breast cancer detection in mammograms using Wavelet and contourlet transformations," 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, pp. 923-926, 2015.

[10] B. Sanae, E. M. Samira, A. K. Mounir and F. Youssef, "Statistical block-based DWT features for digital mammograms classification," 2014 9th International Conference on Intelligent Systems: Theories and Applications (SITA-14), Rabat, pp. 1-7, 2014.

[11] B. Yousefi, Hua-Nong Ting, S. M. Mirhassani and M. Hosseini, "Development of computer-aided detection of breast lesion using gabor-wavelet BASED features in mammographic images," 2013 IEEE International Conference on Control System, Computing and Engineering, Mindeb, pp. 127-131, 2013.

[12] A. U. Rehman, N. Chouhan and A. Khan, "Diverse and Discrimintative Features Based Breast Cancer Detection Using Digital Mammography," 2015 13th International Conference on Frontiers of Information Technology (FIT), Islamabad, pp. 234-239, 2015.

[13] M. Naeemabadi, M. A. Saleh, M. Zabihi, G. Mohseni and N. A. Chomachar, "Malignant tumor detection using linear support vector machine in breast cancer based on new optimization algorithms," 2012 International Symposium on Instrumentation & Measurement, Sensor Network and Automation (IMSNA), Sanya, pp. 80-84, 2012.

[14] S. Caorsi and C. Lenzi, "Can an ANN based radar data processing approach be an aid in breast cancer detection?," 2015 International Conference on Electromagnetics in Advanced Applications (ICEAA), Turin, pp. 1464-1467, 2015.

[15] J. Moll, J. B. Harley and V. Krozer, "Data-driven matched field processing for radar-based microwave breast cancer detection," 2015 9th European Conference on Antennas and Propagation (EuCAP), Lisbon, pp. 1-4, 2015.

[16] Y. M. George, H. H. Zayed, M. I. Roushdy and B. M. Elbagoury, "Remote Computer-Aided Breast Cancer Detection and Diagnosis System Based on Cytological Images," in IEEE Systems Journal, vol. 8, no. 3, pp. 949-964, Sept. 2014.

[17] A. Tashk, M. S. Helfroush, H. Danyali and M. Akbarzadeh, "An automatic mitosis detection method for breast cancer histopathology slide images based on objective and pixel-wise textural features classification," The 5th Conference on Information and Knowledge Technology, Shiraz, pp. 406-410, 2013.

[18] A. Sanagavarapu Mohan, M. D. Hossain and M. J. Abedin, "Beamspace based time reversal processing for breast cancer detection," Proceedings of the 2012 IEEE International Symposium on Antennas and Propagation, Chicago, IL, pp. 1-2, 2012.

[19] Y. l. Li, J. Feng, Y. Ren, Q. p. Wang and B. y. Chen, "Breast cancer detection based on mixture membership function with MFSVM-FKNN ensemble classifier," 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, Sichuan, pp. 297-301, 2012.

[20] K. Ohri, H. Singh and A. Sharma, "Fuzzy expert system for diagnosis of Breast Cancer," 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, pp. 2487-2492, 2016.

[21] V. Gaike, R. Mhaske, S. Sonawane, N. Akhter and P. D. Deshmukh, "Clustering of breast cancer tumor using third order GLCM feature," 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Noida, pp. 318-322, 2015.

[22] M. A. S. Ali, G. I. Sayed, T. Gaber, A. E. Hassanien, V. Snasel and L. F. Silva, "Detection of breast abnormalities of thermograms based on a new segmentation method," 2015 Federated Conference on Computer Science and Information Systems (FedCSIS), Lodz, pp. 255-261, 2015.

[23] T. Gaber et al., "Thermogram breast cancer prediction approach based on Neutrosophic sets and fuzzy c-means algorithm," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, pp. 4254-4257, 2015.

[24] H. Ismahan, F. Amel and B. Abdelhafid, "Mass segmentation in mammograms for computer-aided diagnosis of breast cancer," 2015 3rd International Conference on Control, Engineering & Information Technology (CEIT), Tlemcen, pp. 1-5, 2015.

[25] T. M. Mejía, M. G. Pérez, V. H. Andaluz and A. Conci, "Automatic Segmentation and Analysis of Thermograms Using Texture Descriptors for Breast Cancer Detection," 2015 Asia-Pacific Conference on Computer Aided System Engineering, Quito, pp. 24-29, 2015.

[26] F. A. Maken and A. P. Bradley, "Multiple instance learning for breast MRI based on generic spatio-temporal features," 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, pp. 902-906, 2015.

[27] R. Sánchez de la Rosa, M. Lamard, G. Cazuguel, G. Coatrieux, M. Cozic and G. Quellec, "Multiple-instance learning for breast cancer detection in mammograms," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, pp. 7055-7058, 2015.

[28] Hatipoglu N., and Bilgin G, "Classification of histopathological images using convolutional neural network," 2014 Proc. 4th Int. Conf. on Image Processing Theory, Tools and Applications (IPTA) (Paris, France) pp 1-6, 2014.

[29] Spanhol F A, Oliveira L, Petitjean C and Heutte L 2016 Proc. Int. Joint Conf. on Neural Networks (IJCNN 2016) (Vancouver, Canada) pp 123-128

[30] Wang, D., Khosla, A., Gargeya, R., Irshad, H. & Beck, A. H. “Deep Learning for Identifying Metastatic Breast Cancer,”. https://arxiv.org/abs/1606.05718 ,pp. 1-6, 2016.

[31] Cruz-Roa A, Basavanhally A, Gonzalez F, Gilmore H, Feldman M, Ganesan S, Shih N, Tomaszewski J and Madabhushi A, “Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks” Proc. SPIE 9041, Medical Imaging 2014: Digital Pathology, 904103, pp. 1-15, 20 March 2014.

[32] M. G. Ertosun and D. L. Rubin, “Probabilistic visual search for masses within mammography images using deep learning,” in Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on, pp. 1310–1315, Nov 2015.

[33] J. Arevalo, F. A. Gonz´alez, R. Ramos-Poll´an, J. L. Oliveira, and 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), pp. 797–800, Aug 2015.

[34] Z. Jiao, X. Gao, Y. Wang, and J. Li, “A deep feature based framework for breast masses classification,” Neurocomputing, Vol. 197, pp. 221 – 231, 2016.

[35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” International Journal of Computer Vision (IJCV), vol. 115, no. 3, pp. 211–252, 2015.

[36] A. M. Abdel-Zaher and A. M. Eldeib, “Breast cancer classification using deep belief networks,” Expert Systems with Applications, vol. 46, pp. 139 – 144, 2016.

[37] E. Olfati, H. Zarabadipour and M. A. Shoorehdeli, "Feature subset selection and parameters optimization for support vector machine in breast cancer diagnosis," 2014 Iranian Conference on Intelligent Systems (ICIS), Bam, pp. 1-6, 2014.
[38] R. Nateghi, H. Danyali, M. SadeghHelfroush and F. P. Pour, "Automatic detection of mitosis cell in breast cancer histopathology images using genetic algorithm," 2014 21th Iranian Conference on Biomedical Engineering (ICBME), Tehran, pp. 1-6, 2014.
[39] http://web.inf.ufpr.br/vri/breast-cancer-database
[40] Mansour, Romany F. "Real-Time Iris Tracking-based on a Generalized Probabilistic Particle Filter." Indian Journal of Science and Technology 10.32 (2017).
[41] Mansour, Romany F. "Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy." Biomedical Engineering Letters: 1-17.
[42] Mansour, Romany F. "Evolutionary computing enriched ridge regression model for craniofacial reconstruction." Multimedia Tools and Applications (2017): 1-18.
[43] Mansour, Romany. "Evolutionary Computing Enriched Computer Aided Diagnosis System for Diabetic Retinopathy: A Survey." IEEE Reviews in Biomedical Engineering (2017).
[44] D. S. Lee, "Effective gaussian mixture learning for video background subtraction," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 827–832, 2005.
[45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, 2012.
[46] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in Advances in Neural Information Processing Systems, pp. 3320–3328, 2014.
[47] Y. Jia, et al., "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the ACM International Conference on Multimedia. ACM, pp. 675–678, 2014.
[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Proceedings of 26th Annual Conference on Neural Information Processing Systems 2012 (NIPS), P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1106–1114, Dec. 2012.
[49] Fabio A. Spanhol, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte "Breast Cancer Histopathological Image Classification using Convolutional Neural Networks" 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, pp. 2560-2567, 2016.
[50] Neslihan Bayramoglu and Janne Heikkil "Transfer Learning for Cell Nuclei Classi_cation in Histopathology Images" The 14th European Conference on Computer Vision, pp.1-8, 2016.
[51] Henry Zhou ,Yuki Zaninovich, "Mammogram Classification Using Convolutional Neural Networks", pp. 1-8. http://chrisgregory.me/projects/papers/cnn-mammogram.pdf
[52] Dan C. Cires¸an, Alessandro Giusti, Luca M. Gambardella, J¨urgen Schmidhuber "Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks" Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013, pp 411-418, 2013.
[53] Micha l Z_ ejmo, Marek Kowal, Jo_zef Korbicz and Roman Monczak "Classiffication of breast cancer cytological specimen using convolutional neural network" IOP Conf. Series: Journal of Physics: Conf. Series 783, pp. 1-11, 2017.
[54] Wentao Zhu, Qi Lou, Yeeleng Scott Vang, and Xiaohui Xie "Deep Multi-instance Networks with Sparse Label Assignment for Whole Mammogram Classification" Deep MIL with Sparse Label Assignment for Whole Mamm Class, pp. 1-12, 2016.
[55] Baris Gecer, Ozge Yalcinkaya, Onur Tasar and Selim Aksoy "Evaluation of Joint Multi-Instance Multi-Label Learning For Breast Cancer Diagnosis", Computer Vision and Pattern Recognition, pp. 1-4, 2015.
[56] Daniel Lévy, Arzav Jain "Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks" 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, pp. 1-6, 2016.
[57] Rekha Chakravarthi, NM Nandhitha, S Emalda Roslin, N Selvarasu, "Tumour extraction from breast mammographs through hough transform and DNN hybrid segmentation technique" Biomedical Research 2016; 27 (4): pp. 1188-1193, 2016.
[58] Vibhu Agarwal, Clayton Carson, "Using Deep Convolutional Neural Networks to Predict Semantic Feature of Lesions in Mammograms", semanticscholar, pp. 1-6, 2015.
[59] Neeraj Dhungel, Gustavo Carneiro, Andrew P. Bradley, "The Automated Learning of Deep Features for Breast Mass Classification from Mammograms" Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, pp.106-114, 2016.
[60] S. J. Nass, I. C. Henderson, J. C. Lashof, "Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer," Institute of Medicine (US) and National Research Council (US) Committee on Technologies for the Early Detection of Breast Cancer, pp. 1-267, 2001.
[61] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," pp. 1-9, 2012.

Appendix

Nomenclature	
CAD	Computer-Aided Diagnosis
BC	Breast Cancer
DNN	Deep Neural Network
CNN	Convolutional Neural Network
GMM	Gaussian Mixture Model
PCA	Principle Component Analysis
LDA	Linear Discriminant Analysis
IARC	International Agency for Research on Cancer
WHO	World Health Organization
CAD	Computer Aided Diagnosis
MRI	Magnetic Resonance Imagining
ROI	Region of Interest
SVM	Support Vector Machine
SIFT	Spatial Invariant Feature Transform

IFI	Initial Feature Point Image
ANN	Artificial Neural Network
PNN	Probabilistic Neural Network
CLBP	Completed Local Binary Pattern
NS	Neutrosophic Sets
F-FCM	Fast Fuzzy C-Mean
MIL	Multiple Instance Learning
ROC	Receiver Operating Characteristics
DBN	Deep Belief Network
GA	Genetic Algorithm
CCA	Connected Component Analysis
GPU	Graphical Processing Unit
PDF	Probability Density Function
CONV	Convlutional (Layer)
FC	Fully Connected
MEF	Most Expressive Features
MDF	Most Discriminating Features
RBF	Radial Basis Function
w_n	Weight Factor
y	Value of a Pixel at an Instant
$g(y)$	Gaussian Mixture
β	Rate of Learning
T_h	Threshold
ς_n	Lowest Standard Deviation
α	Fixed Sized Learning Rate
μ_n	Average Density
$\delta_n(t)$	Adaptive Learning Rate
F_V	Feature Vector

$\sum_n = \varsigma_n J$	Covariance Matrix
D_{ISW}	Intra-Class Scatter Matrix
D_{IOS}	Inter-Class Scatter Matrix
T	Total Number of Classes
v_{ci}	Mean of the Average Vector of a Class I
S_{c_i}	Total Number of Samples in Class I
P	Eigenvectors of $I_{ICW}{}^{-1}I_{IOS}$
$\mathcal{F}_i$	Increasing a Factor
q(y)	Output of the Neuron
e_r	Level of Error Tolerance
s_i	Training Vector