

A Framework for Breast Cancer Classification using Multi-DCNNs

Dina A. Ragab, Omneya Attallah, Maha Sharkas, Jinchang Ren, Stephen Marshall

Dina A. Ragab (Corresponding author): Electronics & Communications Engineering Department, Arab Academy for Science, Technology, and Maritime Transport (AASTMT), Alexandria 1029, Egypt, and the Electronic & Electrical Engineering Department, University of Strathclyde, Glasgow G1 1XW, United Kingdom (e-mail: dinaragab@aast.edu; dina.ragab@strath.ac.uk).

Omneya Attallah: Electronics & Communications Engineering Department Arab Academy for Science, Technology, and Maritime Transport (AASTMT), Alexandria 1029, Egypt, (e-mail: o.attallah@aast.edu).

Maha Sharkas: Electronics & Communications Engineering Department, Arab Academy for Science, Technology, and Maritime Transport (AASTMT), Alexandria 1029, Egypt, (e-mail: msharkas@aast.edu).

Jinchang Ren: Electronic & Electrical Engineering Department, University of Strathclyde, Glasgow G1 1XW, United Kingdom (e-mail: jinchang.ren@strath.ac.uk).

Stephen Marshall: Electronic & Electrical Engineering Department, University of Strathclyde, Glasgow G1 1XW, United Kingdom (e-mail: stephen.marshall@strath.ac.uk).

Abstract

Background: Deep learning (DL) is the fastest-growing field of machine learning (ML). Deep convolutional neural networks (DCNN) are currently the main tool used for image analysis and classification purposes. There are several DCNN architectures among them AlexNet, GoogleNet, and residual networks (ResNet). **Method:** This paper presents a new computer-aided diagnosis (CAD) system based on feature extraction and classification using DL techniques to help radiologists to classify breast cancer lesions in mammograms. This is performed by four different experiments to determine the optimum approach. The first one consists of end-to-end pre-trained fine-tuned DCNN networks. In the second one, the deep features of the DCNNs are extracted and fed to a support vector machine (SVM) classifier with different kernel functions. The third experiment performs deep features fusion to demonstrate that combining deep features will enhance the accuracy of the SVM classifiers. Finally, in the fourth experiment, principal component analysis (PCA) is introduced to reduce the large feature vector produced in feature fusion and to decrease the computational cost. The experiments are performed on two datasets (1) the curated breast imaging subset of the digital database for screening mammography (CBIS-DDSM) and (2) the mammographic image analysis society digital mammogram database (MIAS). **Results and Conclusions:** The accuracy achieved using deep features fusion for both datasets proved to be the highest compared to the state-of-the-art CAD systems. Conversely, when applying the PCA on the feature fusion sets, the accuracy did not improve; however, the computational cost decreased as the execution time decreased.

Keywords: Deep convolutional neural networks, machine learning, principal component analysis, and support vector machines

1. Introduction

Breast cancer is considered a severe danger threatening women's life and health. Breast cancer is observed to be one of the most prevalent types of cancer among women worldwide [1]. In Egypt, all types of cancers are increasing rapidly, especially in the breast. Early detection of breast cancer is crucial for successful treatment and reducing the mortality rate.

Medical image examination is the most effective method for the diagnosis of breast cancer. There are different imaging modalities used for diagnoses such as digital mammography, ultrasound (US), magnetic resonance imaging (MRI), and infrared thermography, although, mammography imaging is mostly recommended [2,3]. Mammography produces high quality images to visualize the internal anatomy of the breast. There are several indicators of breast cancer from mammograms. Among them are masses, macrocalcifications (MCs), and architectural distortions. The former two indicators are the crucial indicators of tumors in the primary stage, while the architectural distortions are found to be less significant compared to the masses and MCs [4].

Radiologists cannot easily provide accurate manual evaluation due to the increasing number of mammograms generated in widespread screening. Therefore, a computer-aided diagnosis (CAD) system has been developed to

detect the indicators of breast cancer and improve the accuracy of diagnosis. These systems will facilitate the diagnosis process and can be considered as a second opinion for radiologists [2].

Recently, several researchers proposed machine learning (ML) methods for classifying breast abnormality in mammogram images. Assiri et al. [5] proposed an ensemble classifier based on a majority voting mechanism. The performance of different state-of-the-art ML classification algorithms was evaluated for the Wisconsin breast cancer dataset (WBCD) achieving an accuracy of 99.42%. Ragab et al. [6] used image processing techniques to remove the pectoral muscle of the mammographic image analysis society digital mammogram database (MIAS) [7] and the digital mammography dream challenge dataset [8]. The authors extracted the features using some statistical metrics and classified them using single and multiple classifiers. The highest accuracy achieved was 99.7% [6]. Zhang et al. [9] use Fourier transforms and principal component analysis (PCA), followed by a support vector machine (SVM) to classify the samples of the MIAS dataset [7]. The accuracy achieved was 92.16%. Moreover, the classical CAD systems using ML methods were cited in some papers as in [10–16].

In the last few years, deep learning (DL) using convolutional neural networks (DCNN) has emerged as one of the most powerful ML tools in image classification [17]. It has surpassed the accuracy of traditional classification methods and human ability. The convolutional process simplifies an image that has millions of pixels to a small feature map, thus the dimension of the input data is reduced whilst retaining the most-important differential features [18–21].

Recently, several researchers studied and proposed methods for breast mammography abnormalities classification using DCNN. Zhang et al. [22] developed and evaluated DCNN models for whole mammography image classification introducing transfer learning and data augmentation techniques. The authors used the images from the Department of Radiology, University of Kentucky. The area under the receiver operating characteristics curve (AUC) scored 0.73 (73%) [22]. Hepsag et al. [23] used deep features to classify the MIAS dataset by using 8-fold cross-validation. The accuracy achieved was 68%. Tan et al. [24] proposed a DCNN architecture by using TensorFlow to classify normal and abnormal MIAS samples [7]. The accuracy achieved was 85.85%. Jiang et al. [25] used GoogleNet and AlexNet DCNNs architectures to classify breast lesions of a new dataset named breast cancer digital repository film mammography dataset number 3 (BCDR-F03). The AUC of the GoogleNet and AlexNet was 0.88 (88%) and 0.83 (83%), respectively. Jadoon et al. [26] proposed a model that resolves three classes; normal, benign, and malignant. The model proposed two methods, namely convolutional neural network discrete wavelet (CNN-DW) and convolutional neural network curvelet transform (CNN-CT). The authors classified the samples of the Image Retrieval in Medical Applications (IRMA) dataset using the SVM classifier.

They achieved accuracy rated from 81.83% to 83.74% [26]. Ragab et al. [27] extracted the features of the curated breast imaging subset of DDSM (CBIS-DDSM) [28] using a fine-tuned DCNN-SVM AlexNet architecture. The accuracy achieved was 87.2% with AUC equaling to 0.94 (94%). Mendel et al. [29] extracted and classified the features of 78 mammogram lesions using pre-trained DCNN VGG-19 and SVM, respectively. The AUC achieved was 0.81 (81%). Khan et al. [30] proposed a multi-view feature fusion (MVFF) based CAD system using a feature fusion technique for the classification of mammograms. The deep features were extracted and fused from four fine-tuned DCNN architectures: VGG-16, VGG-19, GoogleNet, and ResNet-50. The authors performed their experiments on CBIS-DDSM [28] and MIAS [7] datasets. The classification accuracy and AUC achieved were 96.66%, 0.934 (93.4%), respectively [30]. Khan et al. [31] extracted and fused the features using the fine-tuned DCNN architectures VGGNet, GoogleNet, and ResNet-50. They classified the fused features by the average pooling method achieving 97.67% accuracy. They used two breast microscopic image data sets: Breast cancer histopathology images [32] and a locally collected dataset from LRH hospital Peshawar, Pakistan to evaluate the performance of the proposed CAD system [31]. Song et al. [33] proposed a new CAD system to classify three classes, normal, benign, and malignant samples of the DDSM dataset [34]. The authors fused the deep features of GoogleNet, Inception-v2, and Inception with $n \times n$ convolution with handcrafted features. The handcrafted features included scoring features, GLCM, and histogram of oriented gradient (HOG) features. The features were classified by SVM and extreme gradient boosting (XGBoost) classifiers. The authors first classified the end-to-end DCNN features achieving an accuracy of 82.84%; however, when classifying the fused features, the accuracy increased reaching 92.8%. The results achieved by the XGBoost proved to be higher than those achieved by the SVM classifier alone [33]. Moreover, deep learning (DL) was used in a breast classification problem as in [35–41].

The novelty of this paper lies in the design of an efficient CAD system based on feature extraction and classification using DL techniques to classify benign and malignant, (or normal and abnormal) breast cancer lesions. This CAD system is evaluated using two datasets, CBIS-DDSM and MIAS. Several papers in the literature employed individual DCNNs to classify breast cancer in their CAD systems [22–25] and [27,29], where the classification accuracies between 68-94% were not sufficient for a reliable and powerful CAD system. Other papers proposed the use of feature fusion from several DCNNs of different architectures [30,31,33]. Although the fusion techniques could improve the accuracy to 92.80-97.67%, the combination of deep features, which contributed the most to the improved performance was not analyzed. Moreover, they did not investigate how to reduce the computational cost of the CAD system. To tackle these drawbacks, in this paper, a novel CAD system

is proposed to explore the fusion of various features extracted from different DCNNs for choosing the best combination of the features, which improves the accuracy of the CAD system. Moreover, the proposed CAD system used PCA to reduce the feature dimension as well as the associated computational cost. These procedures are made through the following four different experiments. The first is an end-to-end DCNN process, where features are extracted and classified using several DCNN architectures including, AlexNet, GoogleNet, ResNet-18, 50, and 101. These architectures are fine-tuned to distinguish between two classes instead of 1000 on which they were originally trained. In the second experiment, the features are extracted using the fine-tuned DCNN architectures. These features are used separately to construct SVM classifiers with different kernels. In the third experiment, a deep feature fusion process is investigated, which is performed by ranking the extracted deep features and using them to form four feature sets that include a different combination of deep features. This experiment is performed to determine if combining deep features from different DCNNs can enhance the classification performance of SVM. Finally, in the fourth experiment, the effect of using a feature reduction method such as PCA to reduce the feature space of the four sets of features is evaluated as well as the computational cost. The number of principal components is chosen in a sequential forward strategy.

This paper is organized as follows: Section 2 introduces the methodology of the paper, section 3 shows the experimental setup, and section 4 gives the computed results of the technique. A discussion of the suggested technique is presented in section 5 and finally, the work is concluded in section 6.

2. Methodology

The CAD system consists of four modules, which are (1) image preprocessing, (2) feature extraction, (3) feature classification and reduction, and (4) classifier evaluation. The proposed framework is shown in Fig. 1.

2.1 Image Preprocessing

Usually, in this step, the images are enhanced and the region of interest (ROI) is cropped. In this paper, the contrast-limited adaptive histogram equalization (CLAHE) is used as an enhancement technique [42]. CLAHE is a type of adaptive contrast enhancement method (AHE). AHE is a technique accomplished by improving local contrast and using all the details in the image. It proved to be an exceptional contrast enhancement method for all types of images [43,44]. CLAHE creates and clips the histogram of each contextual region at a predefined value. The clipped amount is reallocated among the histogram bins. The created histogram is a modified version of the original one. This method solves the edge-shadowing effect of AHE and reduces the problem of over-enhancement [45,46].

2.2 Feature Extraction

In traditional CAD systems, much effort is exerted by the human designer to extract and produce handcrafted features, including the shape and density information of the cancerous area in medical images [47]. This is actually a very challenging task, as this process is lengthy yet the extracted handcrafted features may still not have the discrimination power for classifying cancerous regions [48,49]. Consequently, the emerging DL techniques have attracted increased attention for their outstanding performance by skipping the design of handcrafted features and providing high classification accuracy consistency. The main advantage of the DL and more specific the DCNN is its ability to learn and extract the optimal features by itself through a training process [50]. Furthermore, DCNN consists of several layers of nonlinear or quasi-nonlinear processing to attain a high-level representation of features in images compared to conventional CAD systems using handcrafted features as an input. On the other hand, DL has the ability to deduce an optimal representation of the raw images without image preprocessing i.e. enhancement, segmentation, and feature extraction processes, leading to more effective classification and even lower complexity of design compared to conventional CAD systems [51]. To this end, in this paper, DL techniques were employed to extract significant features using several DCNNs.

DCNN can be used either as a classifier or a feature extractor [52,53]. In the first experiment, DCNNs are used as classifiers. However, in experiments (2) – (4) they are used as feature extractors. DCNN networks are built-up using three main types of layers; (1) convolutional layer, (2) pooling layer, and (3) fully connected (fc) layer [54,55]. The feature extraction is carried out by the convolution layers while the classification is performed using a fully connected layer. This layer classifies which category input belongs to the extracted features. The pooling layer is employed to minimize the dimensions of feature maps and network parameters [55]. In this paper, several DCNN are evaluated such as AlexNet [18], GoogleNet [56], and the ResNet [57] architectures.

2.2.1 AlexNet Architecture

AlexNet consists of five convolution layers, three pooling layers, and two fully connected layers with approximately 60 million free parameters [18]. The AlexNet DCNN architecture is shown in Fig. 2. In the convolution layers, a dot product operation is performed for each neuron between the weights and the local region that is connected to the input [18]. While the pooling layers perform a down-sampling operation on the previous layers to reduce the amount of computation [18]. Moreover, the neurons of the fully connected layers have full connections to all neurons in the previous layer [58,59].

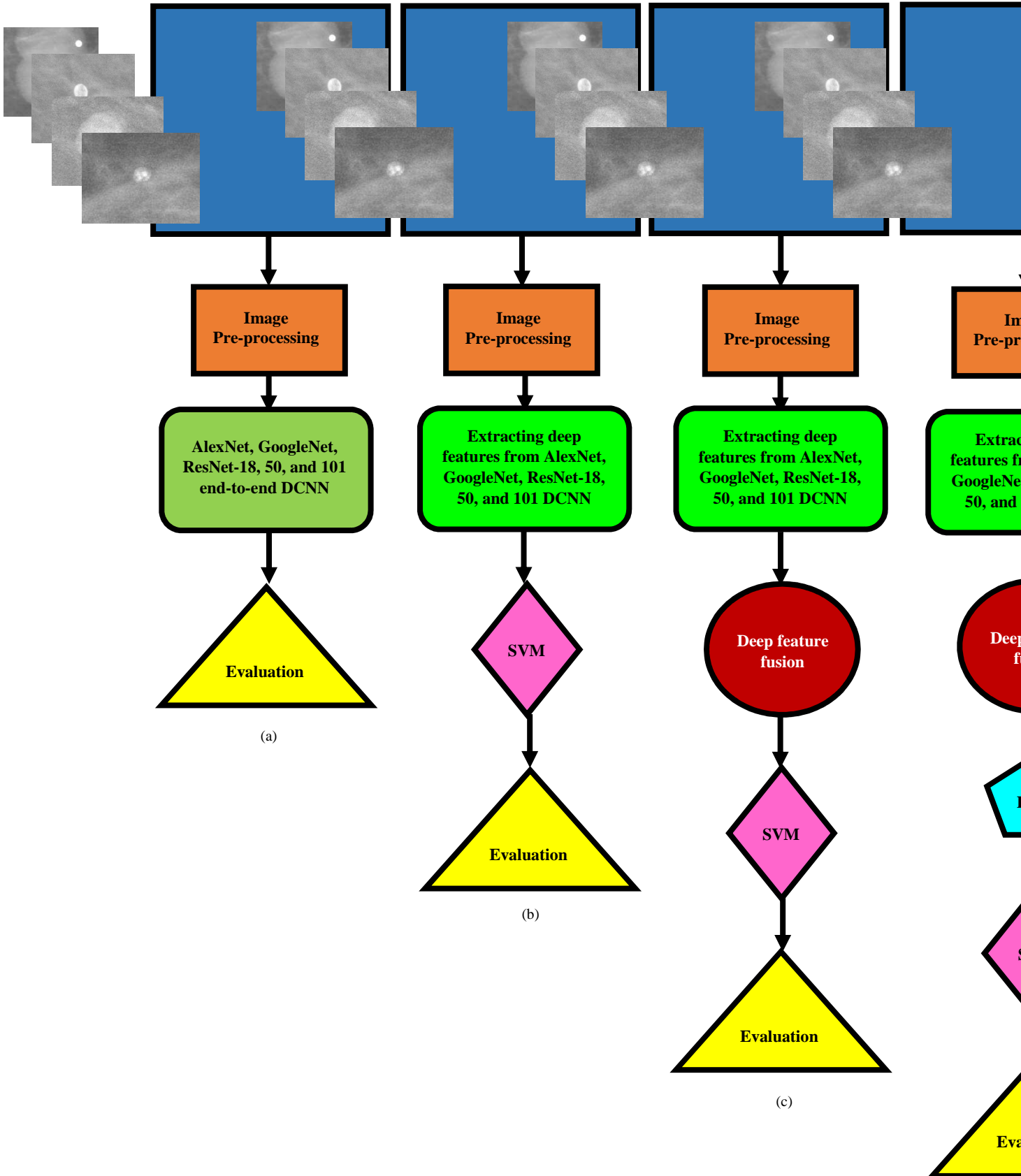


Fig. 1. The framework of the CAD system proposed.

(a) Experiment (1), (b) experiment (2), (c) experiment (3), and (d) experiment (4).

2.2.2 *GoogleNet Architecture*

GoogleNet is a DCNN architecture consisting of 22 layers. It was introduced by Szegedy C. et al. [56] who proposed a computationally efficient structure. This network structure is based on the Inception module; therefore, it is called Inception-v1. Each Layer of the GoogleNet has nine inception units and finally a fully connected layer before the output. GoogleNet has several Inception modules weighted upon each other, with a maximum pooling layer. Although GoogleNet is incredibly deep, it has twelve times fewer parameters than AlexNet, which makes it faster to train. The GoogleNet architecture is shown in Fig. 3.

2.2.3 *Residual Networks (ResNet) Architecture*

ResNet is one of the recent architectures that are commonly used for medical imaging applications. It received first place in ILSVRC and COCO 2015 competition in ImageNet Detection, ImageNet localization, Coco detection, and Coco segmentation [57,60]. The main building block in ResNet is the residual block introduced by He et al. [57]. This method adds shortcuts (called residuals) between layers of traditional DCNNs to bypass a few convolution layers at a time. It increases the number of deep layers as well to enhance its performance and employs the residual shortcuts to accelerate the convergence of these large numbers of deep layers. A ResNet has several stacks of residual blocks. Each block is made up of several stacked convolution layers. Every single convolution layer takes the output fields of the feature map of the previous layer as its input. The output of every residual block is added to its input through an associate identity mapping path [61,62]. ResNet has some common architectures such as ResNet-18, 50, and 101.

2.3 *Feature Reduction*

Principal Component Analysis (PCA) reduces the number of observed variables to a smaller number of principal components that still contain most of the information of the large set. PCA is performed using the variance-covariance structure of a set of variables through linear combinations. It is used when variables are highly correlated, and it is suitable for data sets in multiple dimensions. PCA provides a powerful tool for data analysis and pattern recognition. It is used frequently in signal and image processing [63].

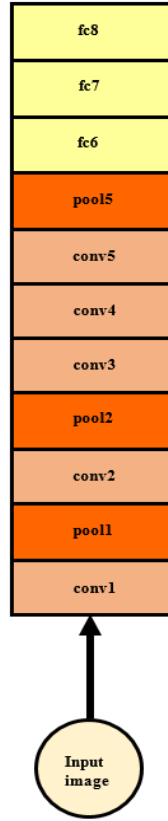


Fig. 2. The AlexNet DCNN Architecture.

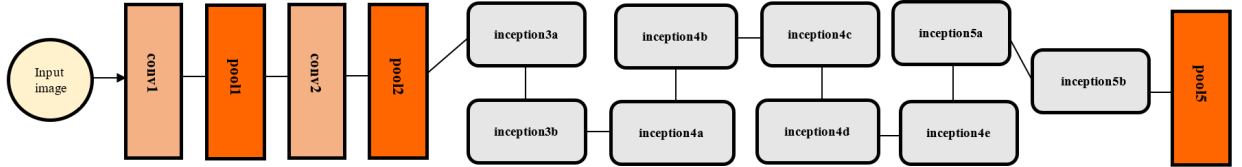


Fig. 3. The GoogleNet DCNN Architecture.

2.4 Feature Classification

In this step, the ROI is classified as either benign/malignant or normal/abnormal lesions according to the features. There are lots of classifier techniques among them decision trees (DT), artificial neural networks (ANN), and support vector machines (SVM) [64,65].

In this paper, specifically in the first experiment, five end-to-end DCNN are constructed using AlexNet, GoogleNet, and the ResNet-18, 50, 101 architectures to classify breast cancer lesions. In experiments (2), (3), and (4) the SVM classifier of different kernels is used. SVM is a supervised learning method that groups data into categories. The goal of the SVM classifier is to formulate an efficient way to separate the data into classes by creating hyperplanes [66]. Many hyperplanes could classify two data points, but the ideal is the one having the

maximum margin. The margin is defined as the width by which the boundary could increase before encountering a data point. The support vectors are the vectors that define the hyperplane [67].

2.5 Feature Evaluation

There are several tools to evaluate a classifier, amongst them the accuracy, the sensitivity or true positive rate (TPR), the specificity or true negative rate (TNR), and the Matthews correlation coefficient (MCC) as defined in equations (1) to (4).

$$accuracy = \frac{TP + TN}{TN + FP + FN + TP} \quad (1)$$

$$sensitivity (TPR) = \frac{TP}{TP + FN} \quad (2)$$

$$specificity (TNR) = \frac{TN}{TN + FP} \quad (3)$$

$$Matthews \text{ correlation coefficient (MCC)} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4)$$

In equations (1) to (4), TP, TN, FP, and FN are the abbreviations for the true positive, true negative, false positive, and false negative, respectively. Accuracy indicates to what extent the relevant classifier has classified the items correctly. In addition, FPR and FNR indicate the false positive rate and false-negative rate, respectively, as in equations (5) and (6).

$$FPR = 1 - TNR \quad (5)$$

$$FNR = 1 - TPR \quad (6)$$

FPR and FNR criteria represent the system error. FPR is a false alarm rate indicating the percentage of class (1) that has been incorrectly classified as class (2). Whereas FNR provides the percentage of class (2) that have been incorrectly classified as class (1).

The area under the receiver operating characteristic (ROC) curve (AUC) is one of the well-known criteria that is used by most medical diagnostic systems for assessing the efficiency of classification. ROC is a curve based on the TPR and FPR and it is generated by changing the threshold decision. AUC provides an approach for evaluating models based on the average of each point on the ROC curve. A classifier has a higher efficiency when the AUC value is approaching one.

3 Experimental Setup

The proposed CAD system was applied to the mammogram images to estimate the possibility of each image belonging to one of the two classes either benign or malignant (or normal and abnormal). All the experiments

were performed on the Intel® CORE™ I7 processor and NVIDIA GeForce 940MX, Windows 10, 64 bit with 8 GB of random-access memory (RAM). The software used to implement the experiments was MATLAB R2018b with an academic license provided by the University of Strathclyde.

3.1 Dataset Selection

In this study, two datasets were used to test the performance of the proposed CAD system. These datasets are (1) the curated breast imaging subset of DDSM (CBIS-DDSM) [28] and (2) the mammographic image analysis society digital mammogram database (MIAS) [7]. The format of the MIAS dataset is PGM; however, the CBIS-DDSM is DICOM, which is the format as obtained directly from the digital equipment. For simplicity, a MATLAB tool was used to convert all the samples to the JPG format.

CBIS-DDSM: Recently, Lee et al. [28] released an updated and standardized version of the DDSM dataset for the evaluation of CAD systems in mammography named CBIS-DDSM. It includes an easily accessible dataset and improved ROI segmented images. The dataset contains 753 and 891 microcalcification and mass cases, respectively. In this work, only the mass samples were used, which are categorized as benign and malignant mass tumors.

MIAS: An organization of the UK research groups called mammographic image analysis society (MIAS) created a database of digital mammograms [7]. The films have been digitized to a 50-micron pixel edge. All images are available in a size of 1024×1024 . Mammogram images are available via the pilot European image processing archive (PEIPA) at the University of Essex [7]. The MIAS dataset has 322 annotated images of left and right breasts classified as normal and abnormal lesions. The abnormal samples are divided into six categories, i.e. calcification, architectural distortion, asymmetry, well-defined, spiculated, and ill-defined masses. In addition, the severity of abnormal, benign or malignant, is stated beside each abnormal sample. However, the number of samples of normal, benign, and malignant is not normalized. Therefore, we differentiated between only two classes, i.e. normal and abnormal, which is consistent with most existing works on the MIAS dataset as reported in [23,68–72]. Moreover, the ROI was cropped using the co-ordinates of the center and the radius of the abnormality provided by the dataset.

3.2 Transfer Learning

The DCNN networks are trained using the ImageNet dataset, which has 1.2 million natural images in 1,000 labeled classes. The transfer learning technique is performed on these networks so that it can be used in any classification problem [73,74]. This is performed by replacing the last fully connected layer in any network with a

new layer for the classification of two classes: Benign/malignant for the CBIS-DDSM, normal/abnormal for the MIAS dataset.

To retrain the DCNN after fine-tuning the fully connected layer, some parameters must be set; the iteration number and the primary learning rate are set to 10^4 and 10^{-3} , respectively [75]. However, the momentum and weight decay are set to 0.9 and 5×10^{-4} , respectively [75]. The number of epochs and mini-batch size were set to 20 and 4. Additionally, the validation frequency was set to the maximum number of iterations per epoch this is done to validate the result at the end of each epoch. These configurations are to confirm that the parameters are fine-tuned for medical breast cancer analysis. Other parameters are set to default values. The optimization algorithm used is the stochastic gradient descent with momentum (SGDM) [75].

3.3 DCNN Architectures

As mentioned in section II, five DCNN architectures are used in this paper after being fine-tuned to differentiate between two classes instead of 1000. A detailed layers description for AlexNet, GoogleNet, and ResNet-18, 50, and 101 are illustrated in Tables 1, 2, and 3, respectively. The input layer of each of the five DCNN architectures constructed requires a specific image size. For example, the input layer of the GoogleNet architecture requires the image samples to be of dimension $224 \times 224 \times 3$. Thus, there was a preprocessing step to change all the image sizes to the size required for each DCNN architecture. However, the output layer of the first convolutional layer in the AlexNet architecture is calculated using equation (7). The output equals $55 \times 55 \times 96$, which demonstrates that the size of the feature map is 55×55 in width and height respectively. In addition, the number of feature maps is 96. On the other hand, the output size of the pooling layer is calculated using equation (8).

$$\text{The output size of the conv layer} = \left\lceil \left(\frac{\text{input} - \text{filter size} + (2 \times \text{Padding})}{\text{Stride}} \right) + 1 \right\rceil \quad (7)$$

$$\text{The output size of the pooling layer} = \left\lceil \left(\frac{\text{output of conv} - \text{pool size}}{\text{Stride}} \right) + 1 \right\rceil \quad (8)$$

3.4 Data Augmentation

In general, when training on a huge number of samples, the classifiers perform well and give high accuracy rates. On the other hand, the biomedical datasets contain a small number of samples due to limited patient volume. Consequently, data augmentation is a vital process. Data augmentation aims to increase the number of images; this is done by generating new images, which are variations on the original ones. Data augmentation has many forms such as rotation, flipping, and transformation; the one used in this paper is the rotation [76–78] Each

original image was rotated by four angles, which were 0, 90, 180, and 270 degrees. Accordingly, each original image was augmented to four images. The total number of samples used for each dataset is illustrated in Table 4.

Table 1. AlexNet architecture layers.

Layer Name	Description		Output Size
Input layer			$227 \times 227 \times 3$
conv1	Filter Size	11×11	$55 \times 55 \times 96$
	Stride	4	
	Padding	0	
pool1	Pooling Size	3×3	$27 \times 27 \times 96$
	Stride	2	
conv2	Filter Size	5×5	$27 \times 27 \times 256$
	Stride	1	
pool2	Pooling Size	3×3	$13 \times 13 \times 256$
	Stride	2	
conv3	Filter Size	3×3	$13 \times 13 \times 384$
	Stride	1	
conv4	Filter Size	3×3	$13 \times 13 \times 384$
	Stride	1	
conv5	Filter Size	3×3	$13 \times 13 \times 256$
	Stride	1	
pool5	Pooling Size	3×3	$6 \times 6 \times 256$
	Stride	2	
Fully connected (fc)			4096×2

Table 2. GoogleNet architecture layers.

Layer Name	Filter Size	Stride	Output Size
Input Layer			$224 \times 224 \times 3$
conv1	7×7	2	$112 \times 112 \times 64$
pool1	3×3	2	$56 \times 56 \times 64$
conv2	3×3	1	$56 \times 56 \times 192$
pool2	3×3	2	$28 \times 28 \times 192$
Inception (3a)	-	-	$28 \times 28 \times 256$
Inception (3b)	-	-	$28 \times 28 \times 480$
pool3	3×3	2	$14 \times 14 \times 480$
Inception (4a)	-	-	$14 \times 14 \times 512$
Inception (4b)	-	-	$14 \times 14 \times 512$
Inception (4c)	-	-	$14 \times 14 \times 512$
Inception (4d)	-	-	$14 \times 14 \times 528$
Inception (4e)	-	-	$14 \times 14 \times 832$
pool4	3×3	2	$7 \times 7 \times 832$
Inception (5a)	-	-	$7 \times 7 \times 832$
Inception (5b)	-	-	$7 \times 7 \times 1024$
average pooling	7×7	1	$1 \times 1 \times 1024$
fully connected (fc)			1024×2

Table 3. ResNet architecture layers.

Layer Name	Output Size	ResNet-18	ResNet-50	ResNet-101
Input Layer		224 × 224 × 3		
conv1	112 × 112 × 64	Filter size = 7 × 7 Number of filters = 64 Stride = 2 Padding = 3		
pool1	56 × 56 × 64	Pooling size = 3 × 3 Stride = 2		
conv2_x	56 × 56 × 64	$\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, & 64 \\ 3 \times 3, & 64 \\ 1 \times 1, & 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, & 64 \\ 3 \times 3, & 64 \\ 1 \times 1, & 256 \end{bmatrix} \times 3$
conv3_x	28 × 28 × 128	$\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3, & 128 \\ 1 \times 1, & 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3, & 128 \\ 1 \times 1, & 512 \end{bmatrix} \times 4$
conv4_x	14 × 14 × 256	$\begin{bmatrix} 3 \times 3, & 256 \\ 3 \times 3, & 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3, & 256 \\ 1 \times 1, & 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3, & 256 \\ 1 \times 1, & 1024 \end{bmatrix} \times 23$
conv5_x	7 × 7 × 512	$\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3, & 512 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3, & 512 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3$
Average pooling		Pool size = 7 × 7 Stride = 7		
		1 × 1 × 512	1 × 1 × 2048	1 × 1 × 2048
Fully connected (fc)		2 (512 × 2)	2 (2048 × 2)	2 (2048 × 2)

Table 4. The total number of samples for the datasets used in this paper.

			Training	Testing	Total
CBIS-DDSM	Benign	2728	3691	1581	5272
	Malignant	2544			
MIAS	Normal	836	901	387	1288
	Abnormal	452			

4 Results

The proposed CAD system performs four experiments. For the first experiment, 70% of the images were used for training and the rest for testing, as this is a common ratio used in the classification problem. Although, for the rest of the experiments, the ratio of the training and testing was 80%:20%. This was because these experiments were validated using five-fold cross-validation. Additionally, the SVM parameters were tuned using a Bayesian optimization technique [79]. In the following sub-sections, the results of the four experiments will be presented for the CBIS-DDSM and the MIAS datasets.

4.1 CBIS-DDSM dataset

For the CBIS-DDSM dataset, the mass samples from the two mammogram views: Craniocaudal (CC) and the mediolateral-oblique (MLO) views were extracted and used in the four experiments of the CAD system [30,80]. Images of this dataset were already segmented and the breast cancer lesion was shown. Therefore, they did not need to be segmented. The samples were only enhanced using the CLAHE method. Table 4 shows the numbers of

training and testing samples used for the CBIS-DDSM dataset. In the first experiment, five end-to-end DCNNs including AlexNet, GoogleNet, ResNet-18, 50, and 101 were constructed. The end-to-end GoogleNet ranked the highest accuracy of 76.01% as illustrated in Table 5. Moreover, the training time of each of these DCNNs is illustrated in Table 5 as well.

Furthermore, in the second experiment, the deep features of the five DCNNs architectures were extracted and used separately to train and test SVM classifiers with different kernels. Table 6 shows the accuracy, AUC, sensitivity, and specificity of the SVM classifiers with different kernels constructed with the five deep feature sets. The highest classification accuracy was 93.7% achieved by the medium Gaussian kernel SVM classifier constructed using ResNet-18 deep features. Fig. 4 displays the ROC curve and the AUC computed for the ResNet-18 deep features with medium Gaussian kernel function SVM.

In the third experiment, four sets of deep features were generated. These features include a different combination of deep features extracted from the five DCNNs. These sets of features were produced by ranking the accuracies of the different DCNN architectures shown in Table 5. It was clear that the AlexNet and the GoogleNet features achieved the highest accuracies compared to the others. Therefore, the AlexNet and the GoogleNet features were combined to produce a single feature vector with 5120 features named feature set (1). The linear kernel SVM achieved the highest accuracy, which was 94.4%. Moreover, the ResNet-18 features were added to the feature set (1). This set was named feature set (2) containing 5632 features. The highest accuracy achieved was 96.9% for the linear kernel SVM as well. Additionally, the features of ResNet-101 were added to feature set (2) producing a feature vector of 7680 features in length named feature set (3). The accuracy increased to 97.5% for the linear kernel SVM. Finally, all the deep features were combined to produce a feature vector with 9728 features in length named feature set (4). Fig. 5 shows a comparison for the accuracies of SVM classifiers of different kernels for the four feature sets. Table 7 shows the accuracy, AUC, sensitivity, and specificity for the SVM classifiers of different kernels for feature set (4) as it achieved the highest accuracy of 97.9% as shown in Fig. 5. Moreover, the ROC and the AUC of the cubic and quadratic SVM classifiers, which achieved the highest accuracy using feature set (4) is shown in Fig. 6.

In the fourth experiment, PCA was applied to reduce the feature space of each feature set and the complexity of the classification process. To choose the optimal number of principal components, a sequential forward selection was carried out. Fig. 7 represents the classification accuracy, using the PCA reduction for the four feature sets.

Table 5. The accuracy and the trained time of the DCNN architectures for the CBIS-DDSM dataset.

DCNN Architecture	DCNN Accuracy	Training Time
AlexNet	74.68%	6 hours, 30 min
GoogleNet	76.01%	12 hours
ResNet-18	72.23%	14 hours
ResNet-50	71.09%	33 hours
ResNet-101	71.47%	62 hours

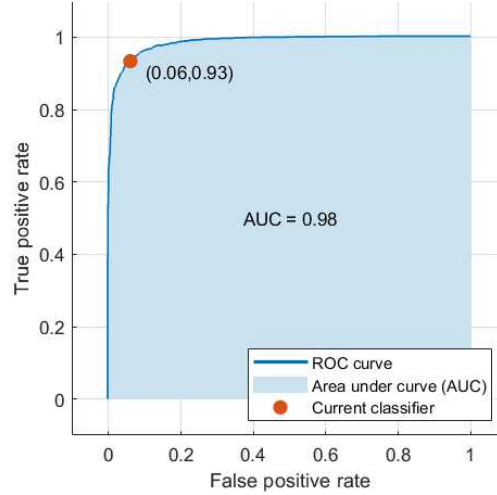


Fig. 4. The computed ROC for the ResNet-18 with Medium Gaussian kernel function SVM of the CBIS-DDSM dataset.

Table 6. The calculated scores of the different SVM kernel functions for the DCNN features of the CBIS-DDSM dataset

DCNN	Different Kernels	Accuracy (std)	AUC (std)	Sensitivity (std)	Specificity (std)	MCC (std)
AlexNet	Linear	91.3% (0.001)	0.97 (0)	0.918 (0.005)	0.911 (0.001)	0.829 (0.005)
	Quadratic	91.0% (0.002)	0.96 (0.004)	0.909 (0.006)	0.911 (0.001)	0.819 (0.007)
	Cubic	90.9% (0.002)	0.96 (0.001)	0.904 (0.004)	0.913 (0.005)	0.816 (0.007)
	Medium Gaussian	91.1% (0.001)	0.97 (0)	0.910(0.001)	0.910 (0.001)	0.820 (0.001)
	Coarse Gaussian	89.2% (0.001)	0.96 (0.001)	0.884 (0.003)	0.899 (0.001)	0.782 (0.004)
GoogleNet	Linear	90.1% (0.002)	0.97 (0)	0.900 (0.001)	0.900 (0.001)	0.800 (0.001)
	Quadratic	89.4% (0.003)	0.97 (0.004)	0.900 (0.001)	0.900 (0.001)	0.793 (0.004)
	Cubic	88.7% (0.004)	0.96 (0.004)	0.883 (0.004)	0.891 (0.004)	0.773 (0.005)
	Medium Gaussian	87.9% (0.002)	0.95 (0.004)	0.859 (0)	0.905 (0)	0.762 (0.001)
	Coarse Gaussian	88.6% (0.004)	0.95 (0.004)	0.898 (0.001)	0.876 (0.006)	0.773 (0.007)
ResNet-18	Linear	93.5% (0.002)	0.98 (0.001)	0.931 (0.001)	0.939 (0.004)	0.870 (0.004)
	Quadratic	93.1% (0.002)	0.98 (0.001)	0.930 (0.003)	0.931 (0.004)	0.861 (0.005)
	Cubic	93.0% (0.001)	0.98 (0.001)	0.930 (0.001)	0.930 (0.001)	0.860 (0.001)
	Medium Gaussian	93.7% (0)	0.98 (0.013)	0.940 (0.001)	0.931 (0.003)	0.872 (0.003)
	Coarse Gaussian	93.4% (0.001)	0.98 (0.001)	0.932 (0.003)	0.94 (0.001)	0.872 (0.003)
ResNet-50	Linear	87.2% (0.002)	0.95 (0.005)	0.864 (0.003)	0.879 (0.001)	0.742 (0.004)
	Quadratic	88.4% (0.003)	0.95 (0)	0.880 (0.008)	0.887 (0.005)	0.765 (0.011)
	Cubic	87.8% (0.001)	0.95 (0)	0.873 (0.001)	0.884 (0.005)	0.756 (0.006)
	Medium Gaussian	87.3% (0.01)	0.95 (0)	0.861 (0.006)	0.894 (0.004)	0.752 (0.004)
	Coarse Gaussian	85.2% (0.005)	0.93 (0.004)	0.833 (0.006)	0.874 (0.001)	0.704 (0.007)
ResNet-101	Linear	89.5% (0.001)	0.96 (0.001)	0.878 (0.021)	0.906 (0.007)	0.782 (0.028)
	Quadratic	89.3% (0.002)	0.96 (0.001)	0.893 (0.006)	0.900 (0.006)	0.790 (0.004)
	Cubic	89.1% (0.001)	0.95 (0.004)	0.890 (0.003)	0.893 (0.005)	0.783 (0.005)

	Medium Gaussian	89.3% (0.002)	0.96 (0.001)	0.877 (0.006)	0.908 (0.001)	0.783 (0.007)
	Coarse Gaussian	87.9% (0.004)	0.94 (0.004)	0.853 (0.006)	0.905 (0.001)	0.754 (0.007)

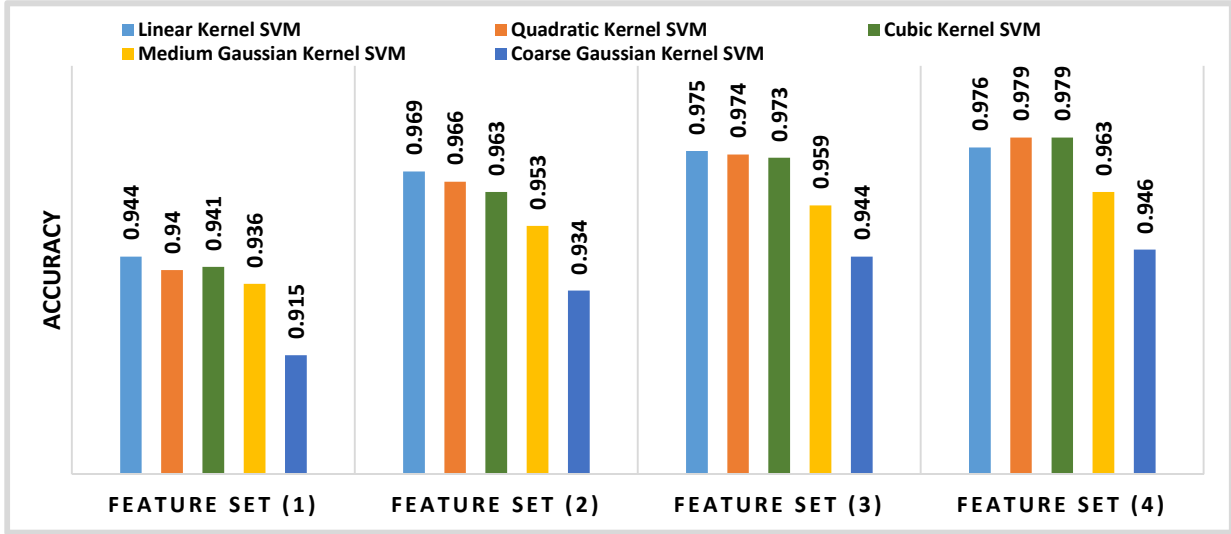


Fig. 5. The accuracies of different SVM kernels for the different DCNN features combination of the CBIS-DDSM dataset.

Feature Set (1) = AlexNet and GoogleNet.

Feature Set (2) = AlexNet, GoogleNet, and ResNet-18.

Feature Set (3) = AlexNet, GoogleNet, ResNet-18, and ResNet-101.

Feature Set (4) = AlexNet, GoogleNet, ResNet-18, ResNet-50, and ResNet-101.

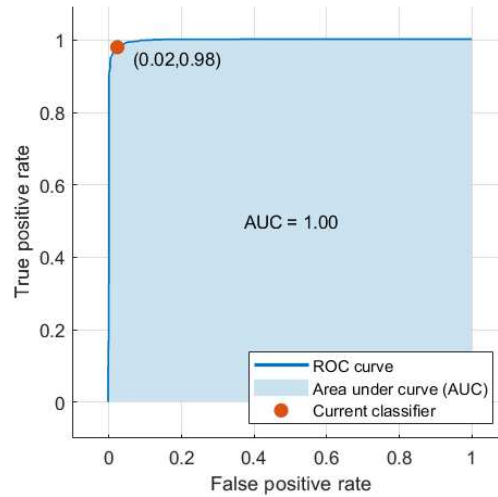


Fig. 6. The computed ROC for all combined DCNN features for the cubic and quadratic SVM kernel functions of the CBIS-DDSM dataset.

Table 7. The calculated scores of the different SVM kernel functions for feature set (4) for the CBIS-DDSM dataset.

AlexNet, GoogleNet, ResNet-18, 50, and 101 DCNN Features					
	Accuracy (std)	AUC (std)	Sensitivity (std)	Specificity (std)	MCC (std)
Linear	97.6% (0.001)	1.00 (0)	0.971 (0.001)	0.98 (0.001)	0.951 (0)
Quadratic	97.9% (0.001)	1.00 (0)	0.980 (0.003)	0.98 (0.001)	0.960 (0.003)
Cubic	97.9% (0.001)	1.00 (0)	0.980 (0.001)	0.98 (0.001)	0.960 (0.001)
Medium Gaussian	96.3% (0.001)	0.99 (0)	0.951 (0)	0.97 (0)	0.921 (0)
Coarse Gaussian	94.6% (0)	0.99 (0)	0.950 (0)	0.95 (0)	0.900(0.001)

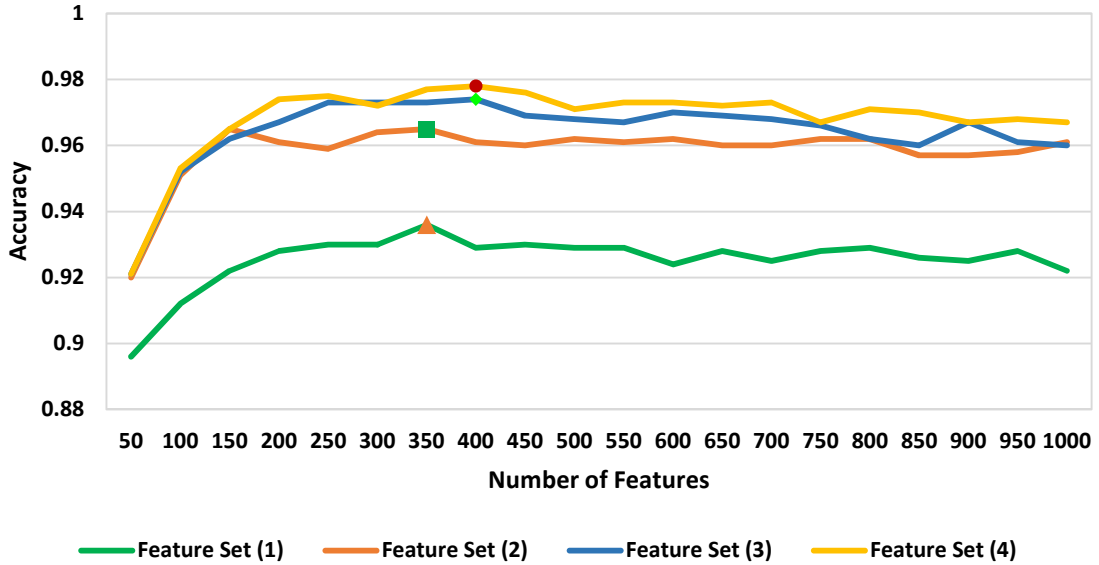


Fig. 7. A comparison of different combined DCNN features accuracy after PCA feature reduction for the CBIS-DDSM dataset.

4.2 MIAS dataset

For the MIAS dataset, the images were enhanced and segmented according to the information provided in the dataset. The center of the abnormal lesion was given in the description of the dataset. These images were used in the first experiment to construct the five DCNNs. Table 8 shows the classification accuracy for the different DCNN architecture to classify normal and abnormal lesions. Furthermore, in the second experiment, the deep features were extracted and used to train and test the SVM classifiers built-up with different kernel functions. The classification accuracies of these models are illustrated in Table 9. The features of the ResNet-50 with quadratic kernel function achieved the highest accuracy, which was 95.3% with AUC equaling to 0.99 (99.0%). The ROC and AUC of the quadratic SVM classifier constructed using deep features of ResNet-50 are shown in Fig. 8.

In the third experiment, four sets of different combinations of deep features were generated in the same manner as those produced in the CBIS-DDSM. Feature set (1) represents those extracted from the GoogleNet and ResNet-50 as these DCNNs achieved the highest accuracies compared to the others as in Table 8. The feature-length of feature set (1) was 3072 features. The highest accuracy achieved was 95.0% using the quadratic kernel SVM. On the other hand, feature set (2) consists of feature set (1) plus ResNet-18 producing 3584 feature length. The accuracy was 96.3% achieved using the quadratic kernel SVM as well. The features of the ResNet-101 were added to the feature set (2) and named feature set (3). The highest accuracy achieved was 97.4% for the quadratic kernel SVM. Finally, adding the AlexNet features to feature set (3) produces a feature set (4). This time the accuracy decreased to 96.6%. Fig. 9 shows the accuracies of the SVM classifiers constructed using different

kernels for the four sets of features. Table 10 shows different scores for the feature set (3) which achieved the highest accuracy using the quadratic SVM classifier. The ROC curve and the AUC for the quadratic kernel SVM function of feature set (3) are shown in Fig.10. In the fourth experiment, the PCA was used to reduce the large dimension of the feature sets generated in experiment (3). Fig. 11 shows a comparison between the classification accuracies of the four sets of features and the number of principal components.

Finally, the results of the proposed CAD system are compared to existing CAD systems described in the literature. Table 11 shows a comparison between our proposed CAD and the applicable state-of-the-art CAD systems.

Table 8. The accuracy of the DCNN architectures for the MIAS dataset.

DCNN Architecture	DCNN Accuracy
AlexNet	59.69%
GoogleNet	74.40%
ResNet-18	68.22%
ResNet-50	68.73%
ResNet-101	67.44%

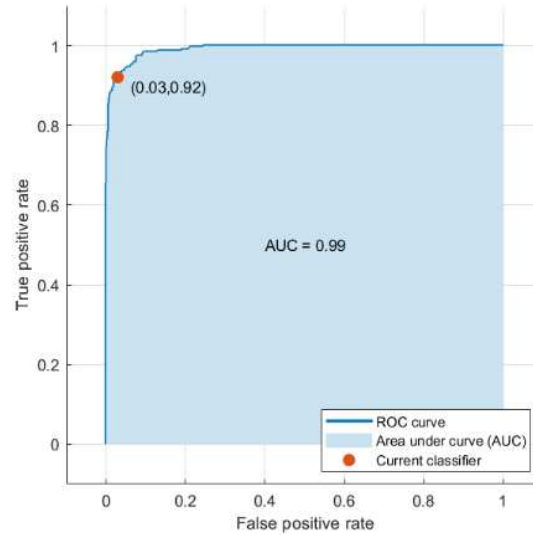


Fig. 8. The computed ROC for the ResNet-50 with quadratic kernel SVM function of the MIAS dataset.

Table 9. The calculated scores of the different SVM kernel functions for the DCNN features of the MIAS dataset.

DCNN	Different Kernels	Accuracy (std)	AUC (std)	Sensitivity (std)	Specificity (std)	MCC (std)
AlexNet	Linear	79.7% (0.002)	0.86 (0.001)	0.871 (0.007)	0.684 (0.003)	0.522 (0.006)
	Quadratic	80.8% (0.007)	0.88 (0.004)	0.841 (0.04)	0.723 (0.016)	0.561 (0.018)
	Cubic	80.0% (0.002)	0.86 (0.001)	0.840 (0.004)	0.725 (0.006)	0.552 (0.008)
	Medium Gaussian	78.8% (0.007)	0.86 (0.006)	0.884 (0.006)	0.662 (0.007)	0.498 (0.013)
	Coarse Gaussian	72.7% (0.004)	0.82 (0.001)	0.902 (0.003)	0.573 (0.002)	0.342 (0.005)
GoogleNet	Linear	76.6% (0.007)	0.83 (0.004)	0.841 (0.011)	0.650 (0.005)	0.451 (0.014)
	Quadratic	76.4% (0.007)	0.83 (0.004)	0.802 (0.044)	0.667 (0.009)	0.460 (0.014)
	Cubic	77.2% (0.007)	0.84 (0.005)	0.817 (0.011)	0.688 (0.002)	0.488 (0.01)
	Medium Gaussian	77.3% (0.004)	0.83 (0.007)	0.888 (0.002)	0.641 (0.003)	0.467 (0.006)
	Coarse Gaussian	71.0% (0.003)	0.79 (0)	0.916 (0.004)	0.557 (0.003)	0.305 (0.01)
ResNet-18	Linear	87.8% (0.002)	0.94 (0.001)	0.937 (0.004)	0.792 (0.001)	0.714 (0.004)
	Quadratic	89.1% (0.002)	0.94 (0.001)	0.921 (0.037)	0.821 (0.011)	0.744 (0.012)
	Cubic	87.9% (0.003)	0.94 (0.004)	0.921 (0.005)	0.810 (0.001)	0.722 (0.005)
	Medium Gaussian	85.3% (0.001)	0.93 (0.004)	0.944 (0.001)	0.740 (0.002)	0.651 (0.003)
	Coarse Gaussian	73.7% (0.002)	0.91 (0.004)	0.964 (0.001)	0.574 (0.002)	0.368 (0.006)
ResNet-50	Linear	94.4% (0.004)	0.99 (0)	0.968 (0.004)	0.894 (0.007)	0.867 (0.009)
	Quadratic	95.4% (0.001)	0.99 (0)	0.966 (0.015)	0.921 (0.011)	0.890 (0.009)
	Cubic	94.6% (0.002)	0.99 (0)	0.971 (0.005)	0.908 (0.001)	0.875 (0.005)
	Medium Gaussian	92.0% (0.001)	0.98 (0.001)	0.976 (0.001)	0.838 (0.001)	0.802 (0.001)
	Coarse Gaussian	79.1% (0.002)	0.95 (0.004)	0.977 (0.01)	0.627 (0.002)	0.492 (0.003)
ResNet-101	Linear	93.1% (0.003)	0.98 (0.001)	0.976 (0.004)	0.860 (0.001)	0.828 (0.004)
	Quadratic	93.6% (0.002)	0.98 (0.001)	0.961 (0.026)	0.887 (0.01)	0.851 (0.009)
	Cubic	93.6% (0.003)	0.98 (0.001)	0.966 (0.004)	0.882 (0.003)	0.843 (0.007)
	Medium Gaussian	91.0% (0.004)	0.97 (0.004)	0.974 (0.004)	0.817 (0.003)	0.774 (0.007)
	Coarse Gaussian	79.1% (0.001)	0.96 (0.001)	0.977 (0.001)	0.627 (0.001)	0.492 (0)

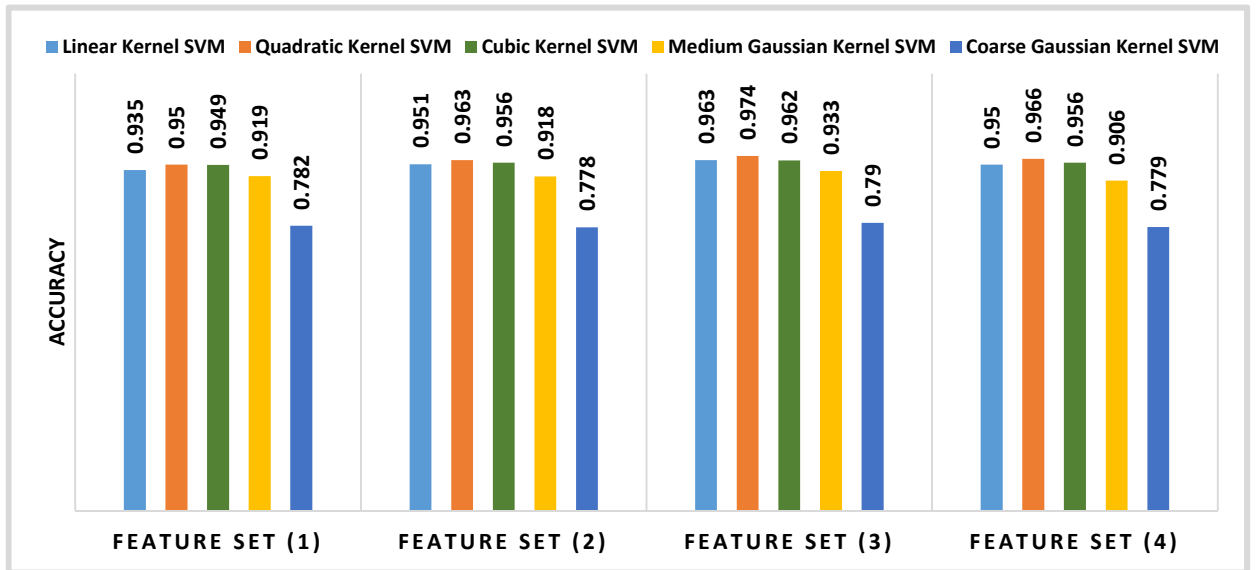


Fig. 9. The accuracies of different SVM kernels for the different DCNN features combination of the MIAS dataset.

Feature Set (1) = GoogleNet and ResNet-50.

Feature Set (2) = GoogleNet, ResNet-18, and ResNet-50.

Feature Set (3) = GoogleNet, ResNet-18, ResNet-50, and ResNet-101.

Feature Set (4) = AlexNet, GoogleNet, ResNet-18, ResNet-50, and ResNet-101.

Table 10. The calculated scores of the different SVM kernel functions for feature set (3) of the MIAS dataset.

	Accuracy (std)	AUC (std)	Sensitivity (std)	Specificity (std)	MCC (std)
Linear	96.3% (0.001)	0.99 (0)	0.99 (0)	0.917 (0)	0.903 (0.001)
Quadratic	97.4% (0)	1.00 (0)	0.99 (0.013)	0.952 (0.012)	0.938 (0.013)
Cubic	96.2% (0)	1.00 (0)	0.99 (0.001)	0.926 (0.001)	0.913 (0)
Medium Gaussian	93.3% (0)	0.99 (0)	0.99 (0.001)	0.854 (0.001)	0.831 (0.001)
Coarse Gaussian	79.0% (0)	0.96 (0.001)	1.00 (0.001)	0.629 (0.001)	0.508 (0)

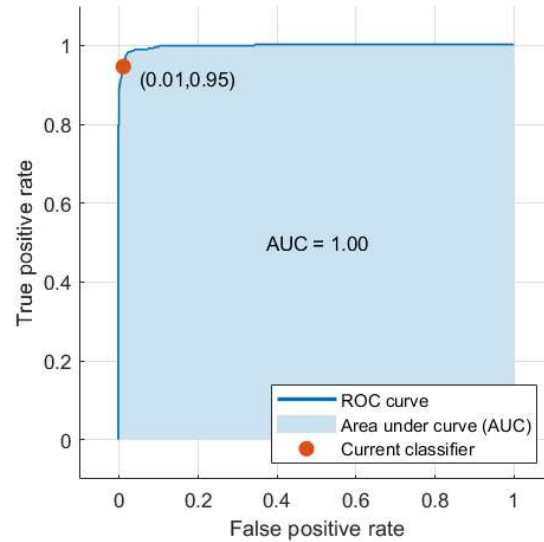


Fig. 10. The computed ROC for feature set (3) combined DCNN features for the quadratic SVM kernel function of the MIAS dataset.

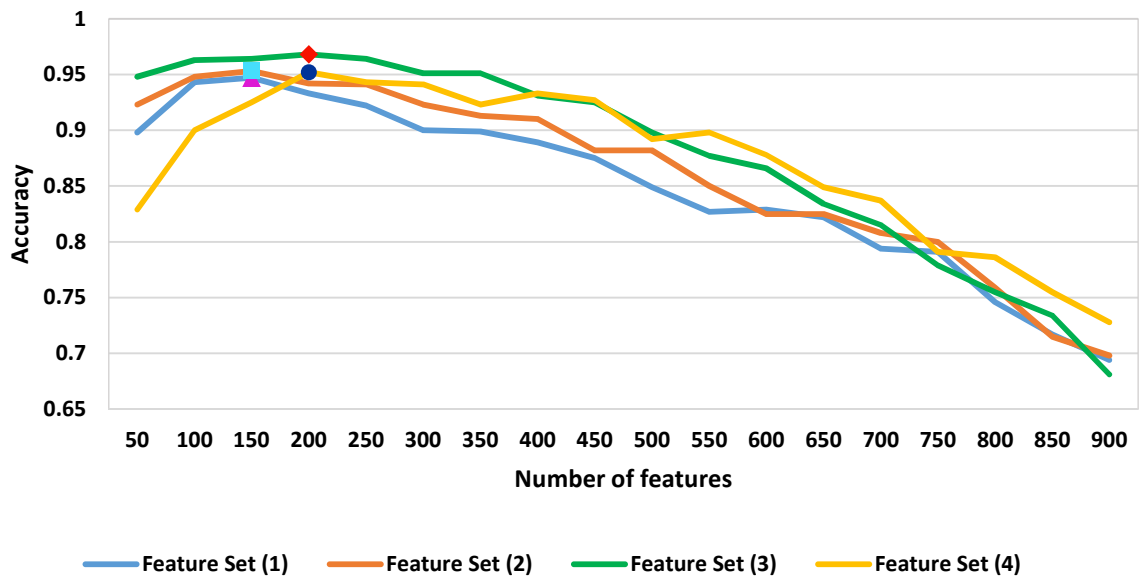


Fig. 11. A comparison of different combined DCNN features accuracy after PCA feature reduction for the MIAS dataset.

Table 11. Classification results for different breast classification methods.

Reference	Year	Feature Extraction	Classification	Dataset	Accuracy	AUC
Hepsag et al. [23]	2017	DCNN		MIAS	68.00%	-
Tan et al. [24]	2017	CNN using Tensorflow		MIAS	85.85%	-
Jiang et al. [25]	2017	GoogleNet and AlexNet		BCDR-F03	-	0.88 0.83
Jadoon et al. [26]	2017	CNN-DW CNN-CT	SVM	IRMA	81.83% 83.74%	-
Ragab et al. [27]	2019	DCNN-AlexNet	SVM	CBIS-DDSM	87.20%	0.94
Khan et. al [30]	2019	Deep features fusion of VGG-16, VGG-19, GoogleNet, and ResNet-50		CBIS-DDSM MIAS	96.6%	0.934
Khan et al. [31]	2019	Deep features fusion of VGG, GoogleNet, and ResNet		Microscopic images	97.67%	-
Song et al. [33]	2020	GoogleNet Inception-v2	XGBoost	DDSM	92.80%	-
Zhang et al. [81]	2020	The fusion of Gist, SIFT, HOG, LBP, VGG, ResNet, and DenseNet features	SVM, XGBoost, Naïve Bayes, k-NN, DT, AdaBoosting	CBIS-DDSM INbreast	90.91% 87.93%	-
Proposed CAD	2020	Deep features fusion of AlexNet GoogleNet ResNet-18 ResNet-50 ResNet-101	SVM	CBIS-DDSM MIAS	97.90% 97.40%	1.00 1.00

5 Discussions

This paper proposed a novel CAD system to classify breast cancer lesions by constructing four different experiments. Transfer learning was used so that the last fully connected layer (fc) of the pre-trained DCNNs architectures was replaced with a new one to classify two classes instead of 1000. In order to increase the number of training samples, data augmentation was used principally based on the rotation technique. The samples were rotated by four angles, which were 0, 90, 180, and 270 degrees. All the experiments were tested on the two datasets; CBIS-DDSM and MIAS. The two-dimensional scatter plot based on the feature vectors for benign and malignant samples of the CBIS-DDSM breast cancer dataset is shown in Fig.12. This figure represents the fifth feature versus the sixth feature as an example for the features of ResNet-18 DCNN architecture for the first 10 samples of CBIS-DDSM dataset images and their orientations with a total of 40 images for each class.

5.1 Experiment (1)

As stated before, in this experiment an end-to-end DCNN of different architectures was constructed. These networks include AlexNet, GoogleNet, ResNet-18, 50, and 101. For the CBIS-DDSM dataset, it was clear from Table 5 that the classification accuracy ranged from (71.09% – 76.01%) with the best accuracy achieved using GoogleNet. The training for the five networks varied between (6:30 – 62 hours) which is quite significant. On the

other hand, the accuracy of DCNN networks in the case of the MIAS dataset ranged from (59.69% – 74.40%) with the highest accuracy achieved using GoogleNet architecture as well, as shown in Table 8.

5.2 Experiment (2)

To improve the classification accuracy of DCNNs constructed in the first experiment, deep features were extracted from each network. These deep features were used separately to train and test SVM classifiers with different kernel functions. From Table 6, it was obvious that the classification accuracies for the CBIS-DDSM dataset increased and ranged between 85.2% and 93.7%. The scores obtained from the deep features of the ResNet-18 proved to be the highest compared to the other networks. Moreover, when comparing the different SVM kernels constructed using ResNet-18 deep features the best accuracy was for the medium Gaussian kernel function. The accuracy was 93.7% and the AUC scored 0.98 (98%) as shown in Fig. 4. Additionally, the sensitivity, specificity, and MCC were 0.94 (94.0%), 0.931 (93.1%), and 0.872 (87.2%), respectively. The deep features obtained for each DCNN for the CBIS-DDSM are visualized in Fig. 13 – 17. In these figures, the first and second convolutional layers were visualized for AlexNet, GoogleNet, ResNet-18, 50, and 101, respectively.

For the MIAS dataset, the accuracies of SVM classifiers constructed using each deep feature of the DCNN have also increased to reach a range of (71.0% – 95.4%). This time the highest accuracy was achieved using the deep features of the ResNet-50 architecture. Furthermore, the quadratic kernel SVM constructed using these deep features ranked the first accuracy which was 95.4% compared to the other kernels. The sensitivity, specificity, and MCC of the quadratic SVM in this case, were 0.966 (96.6%), 0.921 (92.1%), 0.89 (89.0%), respectively. Moreover, the AUC calculated from the ROC curve was 0.99 (99.0%) as shown in Fig. 8.

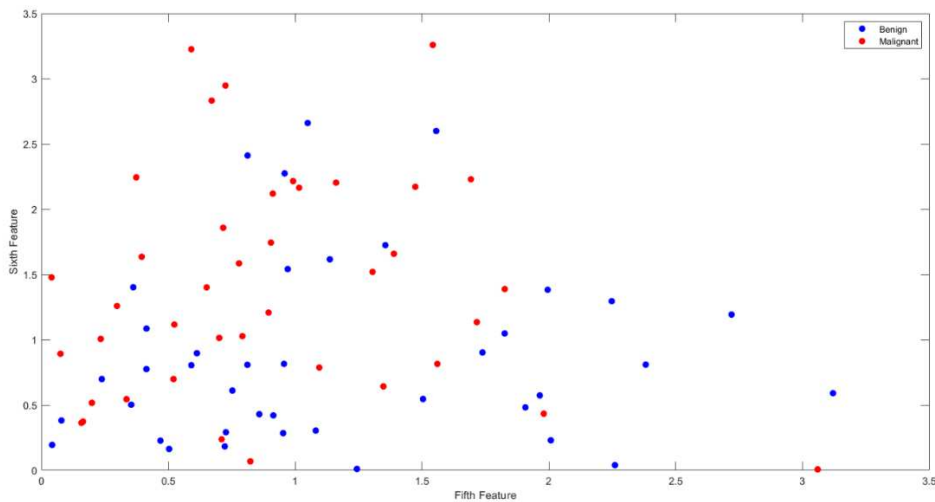


Fig. 12. The fifth feature values versus the sixth feature values for the first 10 samples of CBIS-DDSM images and their rotated versions for the ResNet-18 DCNN architecture features.

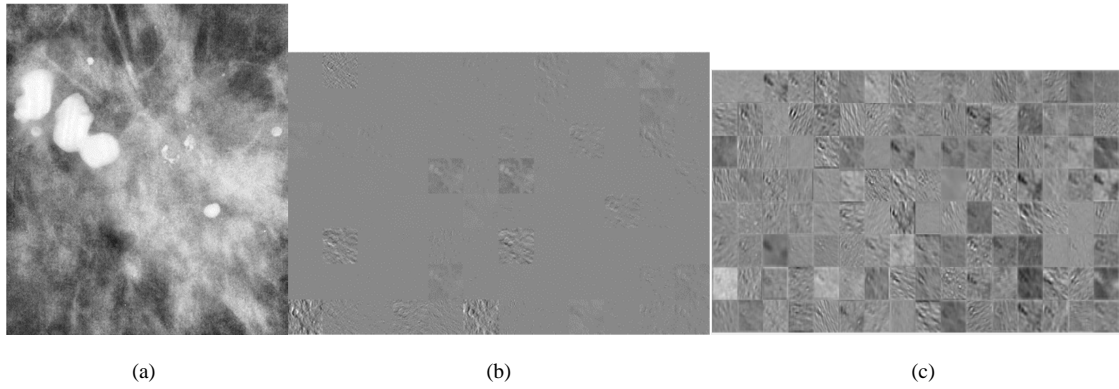


Fig. 13. Visualizing the deep features for the fine-tuned AlexNet DCNN architecture; (a) Malignant ROI from the CBIS-DDSM dataset, (b) the activation features from the first convolutional layer, and (c) the activation features from the second convolutional layer.

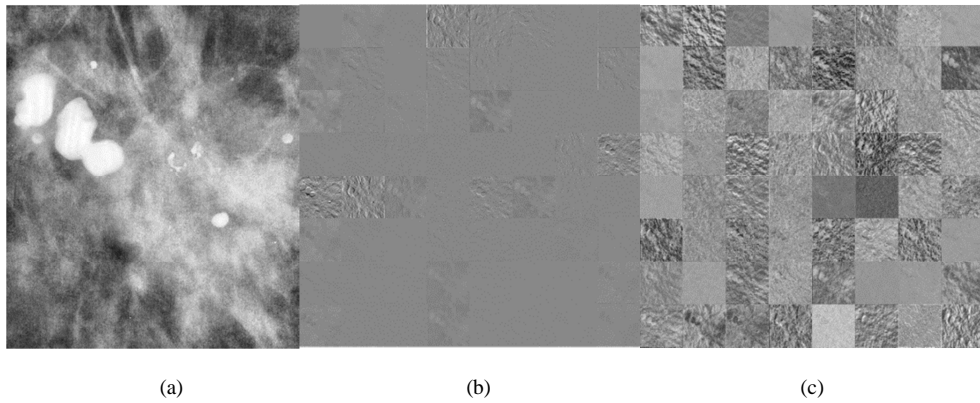


Fig.14. Visualizing the deep features for the fine-tuned GoogleNet DCNN architecture; (a) Malignant ROI from the CBIS-DDSM dataset, (b) the activation features from the first convolutional layer, and (c) the activation features from the second convolutional layer.

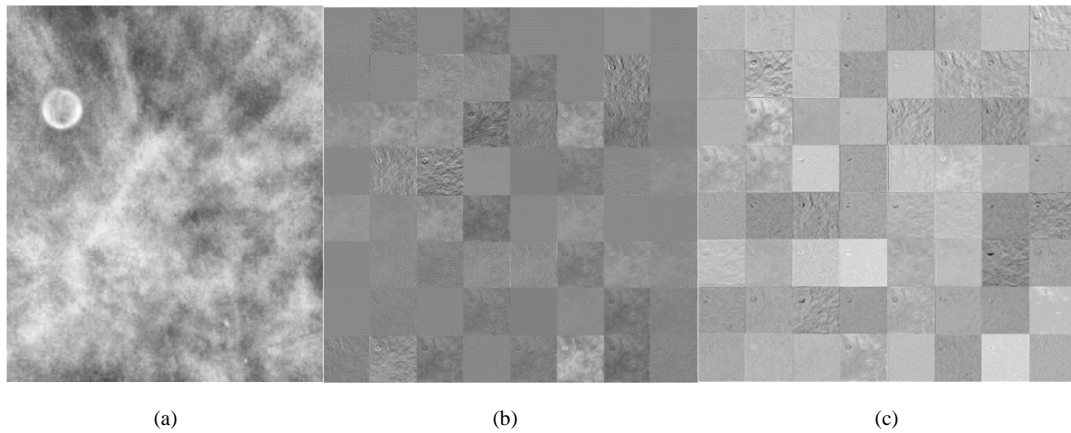


Fig. 15. Visualizing the deep features for the fine-tuned ResNet-18 DCNN architecture; (a) Benign ROI from the CBIS-DDSM dataset, (b) the activation features from the first convolutional layer, and (c) the activation features from the second convolutional layer.

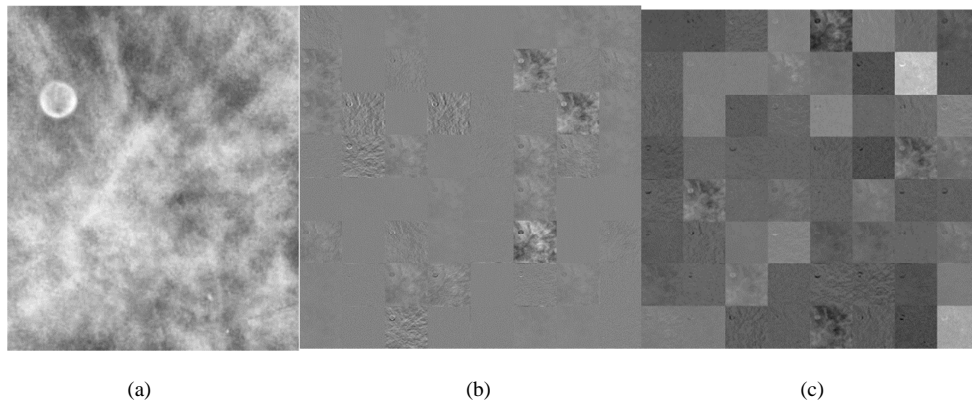


Fig. 16. Visualizing the deep features for the fine-tuned ResNet-50 DCNN architecture; (a) Benign ROI from the CBIS-DDSM dataset, (b) the activation features from the first convolutional layer, and (c) the activation features from the second convolutional layer.

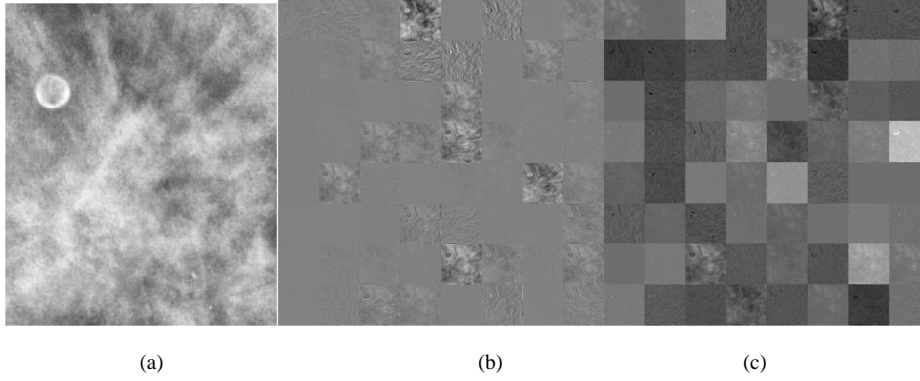


Fig. 17. Visualizing the deep features for the fine-tuned ResNet-101 DCNN architecture; (a) Benign ROI from the CBIS-DDSM dataset, (b) the activation features from the first convolutional layer, and (c) the activation features from the second convolutional layer.

5.3 Experiment (3)

This experiment was conducted to determine if combining deep features would enhance the performance of the SVM classifiers, four feature sets representing different combinations of deep features were produced. To generate these sets, the classification accuracies produced in experiment (1) were used as a ranking method to order the deep features extracted from each DCNN in descending order. Subsequently, this ranking was employed in sequential forward feature set selection to identify the best combination of deep features. For the CBIS-DDSM dataset, the first set includes the deep features of AlexNet and GoogleNet, which have the two highest accuracies compared to the other networks as in Table 5. Feature set (2) represents the feature set (1) plus deep features of ResNet-18. Feature set (3) is a combination of feature set (2) and deep features of ResNet-101. Additionally, feature set (4) consists of feature set (3) and features of ResNet-50. These feature sets were added sequentially to construct the model. Fig. 5 shows that increasing the number of deep features will increase the classification accuracy of the SVM classifiers. Feature set (4) had improved the accuracy to reach 97.9% using the quadratic and cubic kernels. This was higher than the 93.7% of the linear SVM classifier constructed using only the deep features of ResNet-18 in experiment (2). Table VII also indicated that the sensitivity and specificity were both equal to 0.98 (98.0%), which were higher than those achieved experiment (2). Additionally, the AUC increased to 1.00 (100.0%) as shown in the ROC curve in Fig. 6. Furthermore, the MCC yielded to 0.96 (96.0%) for both SVM kernels.

Conversely, for the MIAS dataset, the first set includes the deep features of GoogleNet and ResNet-50 which have the highest two accuracies compared to the other networks as in Table 8. Then, the deep features of ResNet-18 were added to the feature set (1) creating a feature set (2). Then feature set (3) was created as a fusion of feature set (2) and the deep features of ResNet-101. Finally, feature set (4) consisted of feature set (3) plus AlexNet deep features. Fig. 9 reveals that feature set (3) had the highest classification accuracy 97.4% using the quadratic kernel SVM classifier. This was higher than the 95.4% accuracy achieved using the deep features of

ResNet-50 only in the second experiment. Furthermore, the sensitivity, the specificity, and MCC increased to 0.987 (98.7%), 0.949 (94.9%), and 0.938 (93.8%), respectively, which were higher than 0.966 (96.6%), 0.921 (92.1%), and 0.89 (89.0%) achieved in experiment (2). Additionally, the AUC increased from 0.99 (99.0%) to 1.00 (100.0%) as clear in Fig. 10.

5.4 Experiment (4)

PCA was applied in this experiment in order to reduce the large feature space of the feature sets generated in the third experiment. In order to select the number of principal components that achieve the highest accuracy, a sequential forward strategy was used. It started with 50 principal components and added the components iteratively. For the CBIS-DDSM dataset, when reducing the features of sets (1) and (2) the accuracies reached 93.6% and 96.5% with 300 principal components only. However, for feature sets (3) and (4), the accuracies became 97.4% and 97.8% with 400 principal components only. Therefore, the highest classification accuracy in this experiment was achieved using feature set (4) as shown in Fig. 7. Moreover, the execution time for training decreased from 287.48 s to 57.85 s.

Whereas for the MIAS dataset, when the PCA method was applied to feature sets (1) and (2) the accuracies yielded to 94.7% and 95.3%, respectively. This was achieved using 150 principal components. Furthermore, the accuracies of feature sets (3) and (4) became 96.8% and 95.2% with 200 principal components. Thus, the highest accuracy achieved for the MIAS dataset in this experiment was for feature set (3) as shown in Fig. 11. Additionally, the operating time decreased from 40.77 s to 1.97 s when applying PCA on the feature set (3).

To validate the statistical significance of the results obtained in all experiments, an ANOVA test was performed on all the results by a five-fold cross-validation repetition method. The null hypothesis H_0 for all classification was that the mean accuracies of all SVM kernel classifiers were the same. Tables 12 – 16 and Tables 17 – 21 show the ANOVA test for the deep features of the five DCNN architectures constructed in the second experiment for the CBIS-DDSM and MIAS datasets, respectively. In the third experiment, since the highest accuracy was achieved using feature set (4) and feature set (3) for the CBIS-DDSM and the MIAS datasets, therefore, the ANOVA test was computed to these sets. Tables 22 and 23 show the ANOVA test for feature set (4) and feature set (3) performed in the third experiment for the CBIS-DDSM and the MIAS datasets, respectively. From these tables, it was revealed that the p-values achieved were lower than α , where $\alpha = 0.05$. Consequently, it can be concluded that there was a statistically significant difference between the accuracies of the classifiers.

Table 12. ANOVA test details for the different kernel functions of the SVM classifier for the deep features of AlexNet DCNN for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00298	4	0.00074	532.99	< 0.001
Error	0.00006	45	0		
Total	0.00304	49			

Table 13. ANOVA test details for the different kernel functions of the SVM classifier for the deep features of GoogleNet DCNN for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00267	4	0.00067	100.86	< 0.001
Error	0.0003	45	0.00001		
Total	0.00296	49			

Table 14. ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-18 DCNN for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00037	4	9.33×10^{-5}	68.29	< 0.001
Error	0.00006	45	1.3667×10^{-6}		
Total	0.00043	49			

Table 15. ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-50 DCNN for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00605	4	0.00151	60.45	< 0.001
Error	0.00113	45	0.00003		
Total	0.00717	49			

Table 16. ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-101 DCNN for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00173	4	0.00043	120.51	< 0.001
Error	0.00016	45	0		
Total	0.00189	49			

Table 17. ANOVA test details for the different kernel functions of the SVM classifier for the deep features of AlexNet DCNN for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.04284	4	0.01071	509.95	< 0.001
Error	0.00095	45	0.00002		
Total	0.04379	49			

Table 18. ANOVA test details for the different kernel functions of the SVM classifier for the deep features of GoogleNet DCNN for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.02842	4	0.00711	241.35	< 0.001
Error	0.00132	45	0.00003		
Total	0.02975	49			

Table 19. ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-18 DCNN for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.15896	4	0.03974	11332.57	< 0.001
Error	0.00016	45	0		
Total	0.15912	49			

Table 20. ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-50 DCNN for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.18644	4	0.04661	16208.77	< 0.001
Error	0.00013	45	0		
Total	0.18657	49			

Table 21. ANOVA test details for the different kernel functions of the SVM classifier for the deep features of ResNet-101 DCNN for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.15574	4	0.03894	8375.22	< 0.001
Error	0.00021	45	0		
Total	0.15595	49			

Table 22. ANOVA test details for the different kernel functions of the SVM classifier for the feature set (4) in experiment (3) for the CBIS-DDSM dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.00795	4	0.00199	99391	< 0.001
Error	0	45	0		
Total	0.00795	49			

Table 23. ANOVA test details for the different kernel functions of the SVM classifier for the feature set (3) in experiment (3) for the MIAS dataset.

Source of Variation	SS	df	MS	F	p-Value
Columns	0.23117	4	0.05779	9.3699×10^{16}	< 0.001
Error	0	45	0		
Total	0.23117	49			

Finally, the proposed CAD system has been compared with the applicable state-of-the-art CAD systems to prove the efficiency of the proposed method as shown in Table 11. The results reveal that the proposed CAD system has outperformed other CAD systems. Regarding the CBIS-DDSM dataset, the results have shown that the proposed CAD system recorded a slightly higher classification accuracy and AUC compared to Khan et. al [30]. This was obvious as Khan et. al [30] achieved 96.6% for accuracy and 0.934 (93.4%) for AUC. However, Khan et. al [30] fused the deep features of VGG-16, VGG-19, GoogleNet, and ResNet-50 DCNN. Moreover, it was found that the accuracy increased compared to the CAD system proposed by Ragab et al. [27]. Moreover, recently, in 2020, Zhang et al. [81] fused some handcrafted features with deep features and classified them using several classifiers. However, the accuracy was lower than that achieved by feature set (4) generated in the fourth experiment by 6.99%. Whereas for the MIAS dataset, the accuracy achieved was higher than that in Hepsag et al.

[23] and Tan et al. [24] as well. This was clear as the highest classification accuracy and AUC achieved were 97.4% and 0.99 (99%).

On the other hand, when comparing the usage of different DCNN architectures, it was obvious that the scores achieved by the proposed experiments were high as well. In 2017, Jiang et al. [25] and Zhang et al. [82] achieved an AUC of 0.83 (83%) and 0.8 (80%), respectively using the AlexNet DCNN. However, they evaluated the approaches on different datasets BCDR-03 and DDSM, respectively. Moreover, Jiang et al. [25] used the GoogleNet DCNN to achieve a better AUC compared to using AlexNet although, it was still lower than that of the proposed CAD system. Moreover, in 2020, Song et al. [33] extracted and classified the deep features of GoogleNet and Inception-v2 using XGBoost classifier achieving an accuracy of 92.8%. Besides, there was a slight difference between the accuracy achieved in this proposed CAD system and with the work offered by Khan et al. [31]. They fused the deep features of VGG, GoogleNet, and ResNet DCNN. However, Khan et al. [31] applied their experiments on microscopic samples.

6 Conclusions

The accurate and early diagnosis of breast cancer is essential to control the progression of tumors and reduce death rates. Therefore, in this paper, an efficient and accurate solution to diagnose breast cancer is proposed. Radiologists cannot easily provide accurate manual evaluation due to the huge number of mammograms generated in widespread screening. Therefore, a CAD system has been developed to detect the indicators of breast cancer and improve the accuracy of diagnosis.

In this paper, a novel CAD system is proposed to explore different deep feature combinations and choose the one, which best improves the classification accuracy. In addition, it studies the process of reducing the computational cost of classification. This was performed by constructing four different experiments and evaluated on two datasets. The first experiment was composed of constructing five end-to-end pre-trained fine-tuned DCNN networks of different architectures. These networks included AlexNet, GoogleNet, ResNet-18, 50, and 101. In order to increase and enhance the classification accuracy of the first experiment, a second experiment was created. The second experiment was constructed by extracting the deep features of the DCNNs constructed in experiment (1). These deep features were used separately to design SVM classifiers with different kernel functions. The results showed that the classification accuracies in experiment (2) were higher than those of experiment (1). The third experiment was devised to test if combining deep features would enhance the accuracy of the SVM classifiers. In this experiment, the accuracy achieved in experiment (1) was used to rank deep features in descending order. Consequently, four feature sets were generated using this ranking. These feature sets

were used to train the SVM classifiers. The results showed that combining more deep features increased the performance of the SVM classifiers in both datasets. This demonstrated that the feature fusion for one or two views using feature set (4) for the CBIS-DDSM dataset and feature set (3) for the MIAS dataset had improved the accuracy to reach 97.9% and 97.4%, respectively.

Finally, in the fourth experiment, PCA was used to reduce the large dimension of the feature space produced in experiment three. The results showed that PCA had reduced the feature space to 400 and 200 principal components for the CBIS-DDSM and the MIAS datasets, respectively. In addition, the classification accuracies were the same as in experiment (3). However, the computational cost decreased when applying PCA to the feature fusion. This was clear as the execution time for the classification process reduced from 287.48 s to 57.85 s and from 40.77 s to 1.9794 s for the CBIS-DDSM and MIAS datasets, respectively.

The results of the experiments indicated that the proposed CAD system is capable of successfully classifying breast cancer lesions. This is because the highest accuracy achieved was 97.9% and 97.4% using the fusion of deep features for CBIS-DDSM and MIAS datasets, respectively. However, the highest AUC for the CBIS-DDSM and MIAS datasets was 1.00 (100%) using the fusion of deep features as well. These results were higher than other techniques that appeared in the literature.

This study is a crucial trial compromising a simple construct, low cost, efficient, and automatic CAD system. It has been demonstrated that it can achieve a high accuracy by determining the optimal fusion of multiple DCNNs and PCA. DCNN methods. The results demonstrate that it is more capable of distinguishing between cancerous and non-cancerous cases than manual diagnosis by mammogram images. Radiologists may use this CAD system to assist them in accurately diagnosing breast cancer. It will also reduce the time and effort during the examination process and reduce human misdiagnosis that could occur due to human fatigue.

New DCNN architectures are emerging at regular intervals and will be investigated in future work. Moreover, other types of feature extraction techniques could be combined with deep features. In addition, alternative feature reduction techniques such as linear discriminant analysis (LDA) could be investigated.

References

- [1] A.D. Darwish, A.M. Helal, N.H. Aly El-din, L.L. Solaiman, A. Amin, Breast cancer in women aging 35 years old and younger: The Egyptian National Cancer Institute (NCI) experience, *Breast*. 31 (2017) 1–8. doi:10.1016/j.breast.2016.09.018.
- [2] J. Tang, S. Member, R.M. Rangayyan, J. Xu, I. El Naqa, Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances, *IEEE Trans. Inf. Technol. Biomed.* 13 (2009) 236–251.

- [3] P. Agrawal, M. Vatsa, R. Singh, Saliency based mass detection from screening mammograms, *Signal Processing*. 99 (2014) 29–47. doi:10.1016/j.sigpro.2013.12.010.
- [4] J. Bozek, M. Mustra, K. Delac, M. Grgic, A survey of image processing algorithms in digital mammography, in: M. Grgic, K. Delac, M. Ghanbari (Eds.), *Recent Adv. Multimed. Signal Process. Commun.*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009: pp. 631–657. doi:10.1007/978-3-642-02900-4_24.
- [5] A.S. Assiri, S. Nazir, S.A. Velastin, Breast tumor classification using an ensemble machine learning method, *J. Imaging*. 6 (2020). doi:10.3390/jimaging6060039.
- [6] D.A. Ragab, M. Sharkas, O. Attallah, Breast Cancer Diagnosis Using an Efficient CAD System Based on Multiple Classifiers, *Diagnostics*. 9 (2019) 165. doi:10.3390/DIAGNOSTICS9040165.
- [7] J. Suckling, J. Parker, D. Dance, S. Astley, E. Al., The mammographic image analysis society digital mammogram database, *Exerpta Medica, Int. Congr. Ser.* 1069 (1994) 375–378.
- [8] Digital Mammography Dream Challenge dataset, (n.d.). <https://www.synapse.org/#!Synapse:syn4224222/wiki/401743> (accessed November 3, 2019).
- [9] Y.D. Zhang, S.H. Wang, G. Liu, J. Yang, Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform, *Adv. Mech. Eng.* 8 (2016) 1–11. doi:10.1177/1687814016634243.
- [10] M. Sharkas, M. Al-Sharkawy, D.A. Ragab, Detection of microcalcifications in mammograms using support vector machine, in: *Proc. - UKSim 5th Eur. Model. Symp. Comput. Model. Simulation, EMS 2011, 2011*. doi:10.1109/EMS.2011.23.
- [11] M. Al Sharkawy, M. Sharkas, D.A. Ragab, Breast cancer detection using support vector machine technique applied on extracted electromagnetic waves, *Appl. Comput. Electromagn. Soc. J.* 27 (2012) 292–301.
- [12] D.A. Ragab, M. Sharkas, M. Al-Sharkawy, A comparison between support vector machine and artificial neural network for breast cancer detection, in: *12th Int. Conf. Signal Process. Robot. Autom. (ISPRA '13)*, Cambridge, UK, 2013., Cambridge, UK, 2013: pp. 171–176.
- [13] Z. Wang, G. Yu, Y. Kang, Y. Zhao, Q. Qu, Breast tumor detection in digital mammography based on extreme learning machine, *Neurocomputing*. 128 (2014) 175–184. doi:10.1016/j.neucom.2013.05.053.
- [14] N.I.R. Yassin, S. Omran, E.M.F. El Houbay, H. Allam, Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review, *Comput. Methods Programs Biomed.* 156 (2018) 25–45. doi:10.1016/j.cmpb.2017.12.012.
- [15] S.M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafiyan, T. Back, M. Chesus, G.C. Corrado, A. Darzi, M. Etemadi, F. Garcia-vicente, F.J. Gilbert, M. Halling-brown, D. Hassabis, S. Jansen, A. Karthikesalingam, International evaluation of an AI system for breast cancer screening, *Nature*. 577 (2020) 89–94. doi:10.1038/s41586-019-1799-6.
- [16] S. Kala, M. Ezhilarasi, Comparative analysis of serial and parallel fusion on texture features for improved breast cancer diagnosis, *Curr. Med. Imaging*. 14 (2018) 957–968. doi:10.2174/1573405613666170926164625.
- [17] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (2013) 1798–1828. doi:10.1109/TPAMI.2013.50.
- [18] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), *Adv. Neural Inf. Process. Syst.* 25, Curran Associates, Inc., 2012: pp. 1097–1105. doi:http://dx.doi.org/10.1016/j.protcy.2014.09.007.
- [19] D.A. Ragab, O. Attallah, FUSI-CAD: Coronavirus (COVID-19) diagnosis based on the fusion of CNNs and handcrafted features Dina, *Peer J Comput. Sci.* 6 (2020) e306. doi:10.7717/peerj-cs.306.
- [20] O. Attallah, D.A. Ragab, M. Sharkas, MULTI-DEEP : A novel CAD system for coronavirus (COVID-19) diagnosis from CT images using multiple convolution neural networks, *PeerJ*. 8 (2020) e10086.

- doi:10.7717/peerj.10086.
- [21] P. Meyer, V. Noblet, C. Mazzara, A. Lallement, Survey on deep learning for radiotherapy, *Comput. Biol. Med.* 98 (2018) 126–146. doi:10.1016/j.combiomed.2018.05.018.
- [22] X. Zhang, Y. Zhang, E.Y. Han, N. Jacobs, Q. Han, X. Wang, J. Liu, Whole mammogram image classification with convolutional neural networks, in: 2017 IEEE 11th Int. Conf. Bioinforma. Biomed., IEEE, Kansas City, MO, USA, 2017: pp. 700–704. doi:10.1109/BIBM.2017.8217738.
- [23] P.U. Hepsağ, S.A. Özel, A. Yazici, Using deep learning for mammography classification, in: 2nd Int. Conf. Comput. Sci. Eng. UBMK 2017, 2017: pp. 418–423. doi:10.1109/UBMK.2017.8093429.
- [24] Y.J. Tan, K.S. Sim, F.F. Ting, Breast cancer detection using convolutional neural networks for mammogram imaging system, in: 2017 2nd Int. Conf. Robot. Autom. Sci., 2017: pp. 1–5. doi:10.1109/ICORAS.2017.8308076.
- [25] F. Jiang, H. Liu, S. Yu, Y. Xie, Breast mass lesion classification in mammograms by transfer learning, in: Proc. 5th Int. Conf. Bioinforma. Comput. ICBCB '17, 2017: pp. 59–62. doi:http://dx.doi.org/10.1145/3035012.3035022.
- [26] J. M.M., Z. Q., H. I.U., B. S., J. A., Three-class mammogram classification based on descriptive CNN features, *Hindawi BioMed Res. Int.* 2017 (2017). doi:10.1155/2017/3640901.
- [27] D.A. Ragab, M. Sharkas, S. Marshall, J. Ren, Breast cancer detection using deep convolutional neural networks and support vector machines, *PeerJ.* 7 (2019) e6201. doi:10.7717/peerj.6201.
- [28] R.S. Lee, F. Gimenez, A. Hoogi, K.K. Miyake, M. Gorovoy, D.L. Rubin, A curated mammography data set for use in computer-aided detection and diagnosis research, *Sci. Data.* 4 (2017) 1–9. doi:10.1038/sdata.2017.177.
- [29] K. Mendel, H. Li, D. Sheth, M. Giger, Transfer learning from convolutional neural networks for computer-aided diagnosis: A comparison of digital breast tomosynthesis and full-field digital mammography, *Acad. Radiol.* 26 (2019) 735–743. doi:10.1016/j.acra.2018.06.019.
- [30] H.N. Khan, A.R. Shahid, A.H. Dar, H. Alquhayz, Multi-view feature fusion based four views model for mammogram classification using convolutional neural network, *IEEE Access.* 7 (2019) 165724–165733. doi:10.1109/ACCESS.2019.2953318.
- [31] S.U. Khan, N. Islam, Z. Jan, I. Ud Din, J.J.P.C. Rodrigues, A novel deep learning based framework for the detection and classification of breast cancer using transfer learning, *Pattern Recognit. Lett.* 125 (2019) 1–6. doi:10.1016/j.patrec.2019.03.022.
- [32] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, *IEEE Trans. Biomed. Eng.* 63 (2016) 1455–1462. doi:10.1109/TBME.2015.2496264.
- [33] R. Song, T. Li, Y.A.N. Wang, Mammographic classification based on XGBoost and DCNN with multi features, *IEEE Access.* 8 (2020) 75011–75021. doi:10.1109/ACCESS.2020.2986546.
- [34] R.M. and P.K.J. M. Heath, K. Bowyer, D. Kopans, The digital database for screening mammography, in: Fifth Int. Work. Digit. Mammography, M.J. Yaffe, Ed., Med. Phys. Publ. 2001., 2001: pp. 212–218. doi:ISBN 1-930524-00-5.
- [35] Z. Wang, M. Li, H. Wang, H. Jiang, Y. Yao, H. Zhang, J. Xin, Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features, *IEEE Access.* 7 (2019) 1–1. doi:10.1109/access.2019.2892795.
- [36] G. Murtaza, L. Shuib, A.W. Abdul Wahab, G. Mujtaba, G. Mujtaba, H.F. Nweke, M.A. Al-garadi, F. Zulfiqar, G. Raza, N.A. Azmi, Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges, *Artif. Intell. Rev.* (2019). doi:10.1007/s10462-019-09716-5.
- [37] S. Ekici, H. Jawzal, Breast cancer diagnosis using thermography and convolutional neural networks, *Med. Hypotheses.* 137 (2020). doi:10.1016/j.mehy.2019.109542.
- [38] A. Ibrahim, P. Gamble, R. Jaroensri, M.M. Abdelsamea, C.H. Mermel, P.H.C. Chen, E.A. Rakha, Artificial intelligence in digital breast pathology: Techniques and applications, *Breast.* 49 (2020) 267–

273. doi:10.1016/j.breast.2019.12.007.
- [39] I. Sechopoulos, R.M. Mann, Stand-alone artificial intelligence - The future of breast cancer screening?, *Breast*. 49 (2020) 254–260. doi:10.1016/j.breast.2019.12.014.
- [40] N. Wu, J. Phang, J. Park, Y. Shen, Z. Huang, M. Zorin, S. Jastrze, T. Févry, J. Katsnelson, E. Kim, S. Wolfson, U. Parikh, S. Gaddam, L. Leng, Y. Lin, K. Ho, J.D. Weinstein, B. Reig, Y. Gao, H. Toth, K. Pysarenko, A. Lewin, J. Lee, K. Airola, E. Mema, S. Chung, E. Hwang, N. Samreen, S.G. Kim, L. Heacock, L. Moy, K. Cho, K.J. Geras, Deep neural networks improve radiologists' performance in breast cancer screening, *IEEE Trans. Med. Imaging*. 39 (2020) 1184–1194. doi:10.1109/TMI.2019.2945514.
- [41] R. Agarwal, O. Díaz, M.H. Yap, X. Lladó, R. Martí, Deep learning for mass detection in full field digital mammograms, *Comput. Biol. Med.* 121 (2020) 103774. doi:10.1016/j.combiomed.2020.103774.
- [42] A. and H.S. Sahakyan, Segmentation of the breast region in digital mammograms and detection of masses, *Int. J. Adv. Comput. Sci. Appl.* 3 (2012) 102–105.
- [43] S. M.Pizer, E.P. Amburn, J. D.Austin, R. Cromartie, A. Geselowitz, T. Greer, B.T.H. Romeny, J. B.Zimmerman, K. Zuiderveld, Adaptive histogram equalization and its variations, *Comput. Vision, Graph. Image Process.* 39 (1987) 355–368.
- [44] E.D. Pisano, S. Zong, B.M. Hemminger, M. DeLuca, R.E. Johnston, K. Muller, M.P. Braeuning, S.M. Pizer, Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms, *J. Digit. Imaging*. 11 (1998) 193–200.
- [45] J. Lee, S.R. Pant, H.-S. Lee, An Adaptive Histogram Equalization Based Local Technique for Contrast Preserving Image Enhancement, *Int. J. Fuzzy Log. Intell. Syst.* 15 (2015) 35–44. doi:10.5391/ijfis.2015.15.1.35.
- [46] G. Yadav, S. Maheshwari, A. Agarwal, Contrast limited adaptive histogram equalization based enhancement for real time video system, *Proc. 2014 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2014*. (2014) 2392–2397. doi:10.1109/ICACCI.2014.6968381.
- [47] H. Fujita, AI-based computer-aided diagnosis (AI-CAD): The latest review to read first, *Radiol. Phys. Technol.* 13 (2020) 6–19. doi:10.1007/s12194-019-00552-4.
- [48] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks, *NeuroImage Clin.* 21 (2019) 101645. doi:10.1016/j.nicl.2018.101645.
- [49] M.D. Hssayeni, S. Saxena, R. Ptucha, A. Savakis, N.Y. Us, Distracted driver detection: Deep learning vs handcrafted features, *Electron. Imaging*. 10 (2017) 20–26. doi:10.2352/ISSN.2470-1173.2017.10.IMAWM-162.
- [50] K. Suzuki, Overview of deep learning in medical imaging, *Radiol. Phys. Technol.* 10 (2017) 257–273. doi:10.1007/s12194-017-0406-5.
- [51] S. Vieira, W.H.L. Pinaya, A. Mechelli, Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications, *Neurosci. Biobehav. Rev.* 74 (2017) 58–75. doi:10.1016/j.neubiorev.2017.01.002.
- [52] D. Abdelhafiz, C. Yang, R. Ammar, S. Nabavi, Deep convolutional neural networks for mammography: Advances, challenges and applications, *BMC Bioinformatics*. 20 (2019). doi:10.1186/s12859-019-2823-4.
- [53] S. Soffer, A. Ben-Cohen, O. Shimon, M.M. Amitai, H. Greenspan, E. Klang, Convolutional Neural Networks for Radiologic Images: A Radiologist's Guide, *Radiology*. 290 (2019) 590–606. doi:10.1148/radiol.2018180547.
- [54] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, Breast cancer histopathological image classification using convolution neural networks, in: *2016 Int. Jt. Conf. Neural Networks, IEEE, Vancouver, BC, Canada, 2016*: pp. 2560–2567. doi:10.1109/IJCNN.2016.7727519.

- [55] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, Deep learning for visual understanding : A review, *Neurocomputing*. 187 (2016) 27–48. doi:10.1016/j.neucom.2015.09.116.
- [56] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conf. Comput. Vis. Pattern Recognit., IEEE, Boston, MA, USA, 2015: pp. 1–9. doi:10.1109/CVPR.2015.7298594.
- [57] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conf. Comput. Vis. Pattern Recognit., IEEE, Las Vegas, NV, USA, 2016. doi:10.1109/CVPR.2016.90.
- [58] S. Suzuki, X. Zhang, N. Homma, K. Ichiji, N. Sugita, Y. Kawasumi, T. Ishibashi, M. Yoshizawa, Mass Detection Using Deep Convolutional Neural Network for Mammographic Computer-Aided Diagnosis, in: Proc. SICE Annu. Conf. 2016 Tsukuba, Japan, Tsukuba, Japan, 2016: pp. 1382–1386.
- [59] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, F.-F. Li, ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conf. Comput. Vis. Pattern Recognit., Miami, FL, USA, 2009: pp. 248–255. doi:10.1109/CVPRW.2009.5206848.
- [60] H. Lei, T. Han, F. Zhou, Z. Yu, J. Qin, A. Elazab, B. Lei, A deeply supervised residual network for HEp-2 cell classification via cross modal transfer learning, *Pattern Recognit.* 79 (2018) 290–302. doi:https://doi.org/10.1016/j.patcog.2018.02.006.
- [61] Z. Li, S. Yu, W. Qin, L. Liu, K. Li, T. Xiao, Comparison of transferred deep neural networks in ultrasonic breast masses discrimination, *Biomed Res. Int.* 2018 (2018) 1–9. doi:10.1155/2018/4605191.
- [62] S. Wu, J. Xu, S. Zhu, H. Guo, A deep residual convolutional neural network for facial keypoint detection with missing labels, *Signal Processing*. 144 (2018) 384–391. doi:10.1016/j.sigpro.2017.11.003.
- [63] L.I. Smith, A tutorial on Principal Components Analysis, Cornell Univ. USA. (2002).
- [64] J. Ren, ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging, *Knowledge-Based Syst.* 26 (2012) 144–153. doi:10.1016/j.knosys.2011.07.016.
- [65] J. Ren, D. Wang, J. Jiang, Effective recognition of MCCs in mammograms using an improved neural classifier, *Eng. Appl. Artif. Intell.* 24 (2011) 638–645. doi:10.1016/j.engappai.2011.02.011.
- [66] S.R. Gunn, Support vector machines for classification and regression, University of Southampton, 1998.
- [67] I. El-naqa, S. Member, Y. Yang, M.N. Wernick, S. Member, N.P. Galatsanos, S. Member, R.M. Nishikawa, A support vector machine approach for detection of microcalcifications, *IEEE Trans. Med. Imaging*. 21 (2002) 1552–1563.
- [68] S. Khan, M. Hussain, H. Aboalsamh, G. Bebis, A comparison of different Gabor feature extraction approaches for mass classification in mammography, *Multimed. Tools Appl.* 76 (2017) 33–57. doi:10.1007/s11042-015-3017-3.
- [69] F. Mohanty, S. Rup, B. Dash, B. Majhi, M.N.S. Swamy, Mammogram classification using contourlet features with forest optimization-based feature selection approach, *Multimed. Tools Appl.* (2018) 1–30. doi:10.1007/s11042-018-5804-0.
- [70] H. Chougrad, H. Zouaki, O. Alheyane, Multi-label transfer learning for the early diagnosis of breast cancer, *Neurocomputing*. 392 (2020) 168–180. doi:10.1016/j.neucom.2019.01.112.
- [71] F.F. Ting, Y.J. Tan, K.S. Sim, Convolutional neural network improvement for breast cancer classification, *Expert Syst. Appl.* 120 (2019) 103–115. doi:10.1016/j.eswa.2018.11.008.
- [72] S. Wessels, D. van der Haar, Applying deep learning for the detection of abnormalities in mammograms, in: K.H. (eds) Kim K. (Ed.), *Inf. Sci. Appl.*, Springer Singapore, Singapore, 2020: pp. 201–210.
- [73] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, in: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (Eds.), *Adv. Neural Inf. Process. Syst.* 27, Curran Associates, Inc., 2014: pp. 3320–3328. <http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf>.

- [74] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (2010) 1345–1359.
- [75] F. Gao, H. Yoon, T. Wu, X. Chu, A feature transfer enabled multi-task deep learning model on medical imaging, *Expert Syst. Appl.* 143 (2020) 112957. doi:10.1016/j.eswa.2019.112957.
- [76] Z. Hussain, F. Gimenez, D. Yi, D. Rubin, Differential data augmentation techniques for medical imaging classification tasks, in: *AMIA 2017, Am. Med. Informatics Assoc. Annu. Symp.*, Washington, DC, 2017: pp. 979–984.
- [77] D. Han, Q. Liu, W. Fan, A new image classification method using CNN transfer learning and web data augmentation, *Expert Syst. Appl.* 95 (2018) 43–56. doi:10.1016/j.eswa.2017.11.028.
- [78] J. Lemley, S. Bazrafkan, P. Corcoran, Smart augmentation learning an optimal data augmentation strategy, *IEEE Access.* 5 (2017) 5858–5869. doi:10.1109/ACCESS.2017.2696121.
- [79] T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning: Data mining, inference, and prediction*, 2009. doi:10.1007/978-1-4419-9863-7_941.
- [80] A. Jouirou, A. Baâzaoui, W. Barhoumi, Multi-view information fusion in mammograms: A comprehensive overview, *Inf. Fusion.* 52 (2019) 308–321. doi:10.1016/j.inffus.2019.05.001.
- [81] H. Zhang, R. Wu, T. Yuan, Z. Jiang, S. Huang, J. Wu, J. Hua, Z. Niu, D. Ji, DE-Ada*: A novel model for breast mass classification using cross-modal pathological semantic mining and organic integration of multi-feature fusions, *Inf. Sci. (Ny)*. 539 (2020) 461–486. doi:10.1016/j.ins.2020.05.080.
- [82] X. Zhang, T. Sasaki, S. Suzuki, Y. Takane, Y. Kawasumi, T. Ishibashiz, N. Homma, M. Yoshizawa, Classification of mammographic masses by deep learning, in: *2017 56th Annu. Conf. Soc. Instrum. Control Eng. Japan*, 2017: pp. 793–796. doi:10.23919/SICE.2017.8105545.

- Proposing a multi-DCNN based CAD system for breast cancer classification.
- Deep features fusion had an influence on the classification accuracy.
- Searching for the best combination of deep features.
- Selecting the number of principal components with a sequential forward method.
- The computer-aided diagnosis system proposed proved to be reliable.

We confirm that we do not have any conflicts of interest to declare.

Authors Names:

Dina A. Ragab (Corresponding Author)

Omneya Atallah

Maha Sharkas

Jinchang Ren

Stephen Marshall

On behalf of all authors

Dina A. Ragab

A handwritten signature in purple ink that reads "Dina Ragab". The signature is written in a cursive style with a long, sweeping underline that extends to the left and then curves back under the name.



Dina A. Ragab received her B.Sc. and M.Sc. in 2007 and 2011, respectively from the Electronics and Communications Engineering department in Arab Academy for Science, Technology, and Maritime Transport (AASTMT).

She is currently a Ph.D. researcher at the University of Strathclyde, Glasgow, United Kingdom. She is working as an assistant lecturer in Electronics and

Communications engineering department at AASTMT.

Eng. Ragab, research interests are biomedical engineering, medical image processing, and machine/deep learning.



Omneya Attallah received her B.Sc. and M.Sc. in 2006 and 2009, respectively from the Electronics and Communications Engineering department in the Arab Academy for Science, Technology, and Maritime Transport (AASTMT). She received her Ph.D. in 2016 from Aston University UK. Her M.Sc. and Ph.D. research focused on biomedical engineering specifically in medical image

processing and medical informatics using machine learning. She is currently an associate professor in the Electronics and Communications Engineering department at the AASTMT. Her current research interests include; biomedical engineering, biomedical informatics; signal/bio-signal processing, image/biomedical image processing, machine/deep learning, and data mining.



Maha Sharkas received her B.Sc., M.Sc., and Ph.D. in 1990, 1996, and 2002, respectively from the Electrical Engineering Department, Alexandria University. She is currently a Professor and Head of the Electronics and Communications Department at AASTMT. Her research interests are in the field of signal processing, image watermarking, biometrics, biomedical engineering, and pattern recognition.



Jinchang Ren received his B.E. degree in computer software, MEng in image processing, Deng in computer vision, all from Northwestern Polytechnical University (NWPU), Xi'an, China. He was also awarded a Ph.D. in Electronic Imaging and Media Communication from Bradford University, U.K. Currently he is with CeSIP, University of Strathclyde, Glasgow, U.K. His

research interests focus mainly on visual computing and multimedia signal processing, especially on semantic content extraction for video analysis and understanding and more recently, hyperspectral imaging. He has published over 150 peer-reviewed journal and conference papers and acts as an Associate Editor for two international journals including Multidimensional Systems and

Signal Processing (Springer) and Int. J. of Pattern Recognition and Artificial Intelligence.



Stephen Marshall received the B.Sc. degree in electrical and electronic engineering and the Ph.D. degree in image processing respectively from the University of Nottingham and the University of Strathclyde, UK. With over 150 papers published, his research activities focus in nonlinear image processing and hyperspectral

imaging. Currently he is a Professor with the Department of Electronic and Electrical Engineering in Strathclyde, and a Fellow of the IET.