



# Convolutional neural network-based models for diagnosis of breast cancer

Mehedi Masud<sup>1</sup> · Amr E. Eldin Rashed<sup>1</sup> · M. Shamim Hossain<sup>2,3</sup>

Received: 14 July 2020 / Accepted: 24 September 2020 / Published online: 9 October 2020  
© Springer-Verlag London Ltd., part of Springer Nature 2020

## Abstract

Breast cancer is the most prevailing cancer in the world and each year affecting millions of women. It is also the cause of largest number of deaths in women dying in cancers. During the last few years, researchers are proposing different convolutional neural network models in order to facilitate diagnostic process of breast cancer. Convolutional neural networks are showing promising results to classify cancers using image datasets. There is still a lack of standard models which can claim the best model because of unavailability of large datasets that can be used for models' training and validation. Hence, researchers are now focusing on leveraging the transfer learning approach using pre-trained models as feature extractors that are trained over millions of different images. With this motivation, this paper considers eight different fine-tuned pre-trained models to observe how these models classify breast cancers applying on ultrasound images. We also propose a shallow custom convolutional neural network that outperforms the pre-trained models with respect to different performance metrics. The proposed model shows 100% accuracy and achieves 1.0 AUC score, whereas the best pre-trained model shows 92% accuracy and 0.972 AUC score. In order to avoid biasness, the model is trained using the fivefold cross validation technique. Moreover, the model is faster in training than the pre-trained models and requires a small number of trainable parameters. The Grad-CAM heat map visualization technique also shows how perfectly the proposed model extracts important features to classify breast cancers.

**Keywords** Deep learning · Convolutional neural network · Transfer learning · Breast cancer

## 1 Introduction

Breast cancer is affecting millions of women every year in the world and is the reason of highest cause of deaths by cancers among women [1]. The survival rates of breast

cancer largely vary in the countries. In North America, it is greater than 80%, in Sweden and Japan it is around 60%, while in low-income countries it is below 40% [1]. The main reason of low survival rate in low-income countries is the lack of programs for early detection and the shortage of enough diagnosis and healthcare facilities. Therefore, it is vital to detect breast cancer in the pre-mature stage to minimize the rate of mortality. Mammography and ultrasound images are the common tools to identify cancers by the experts and requires expert radiologists. Manual process may cause to generate high false positive and false negative numbers. Therefore, nowadays computer aided diagnosis systems (CADs) are vastly used to aid radiologists during the process of decision making in identifying cancers. The CAD systems now potentially reducing the efforts of radiologists and minimizing the false positive and negative numbers in diagnosis.

Machine learning techniques have showed significant performance in various healthcare applications over the

---

✉ M. Shamim Hossain  
mshossain@ksu.edu.sa

Mehedi Masud  
mmasud@tu.edu.sa

Amr E. Eldin Rashed  
a.rashed@tu.edu.sa

<sup>1</sup> College of Computers and Information Technology, Taif University, Taif, Saudi Arabia

<sup>2</sup> Chair of Pervasive and Mobile Computing, King Saud University, Riyadh 11543, Saudi Arabia

<sup>3</sup> Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

traditional computer aided systems for disease diagnosis and patient monitoring [2]. However, the traditional machine learning techniques involve a hand-created step for extraction of features which is very difficult sometimes. It also requires domain knowledge and an expert radiologist. Meanwhile, deep learning (DL) models automatically develop a learning process adaptively and can extract features from the input dataset considering the target output [3, 4]. The DL methods tremendously reduce the exhaustive process of data engineering and feature extraction while enabling the reusability of the methods. Numerous researches [5] have been conducted to study breast cancer images from various perceptions. Machine learning (ML), convolutional neural networks (CNNs), and deep learning methods are now widely used to classify breast cancers from the breast images.

CNN models have been effectively used in the wide-ranging computer vision fields for years [6, 7]. Since the last few years numerous researches have been conducted applying CNN-based deep learning deep architectures for disease diagnosis. A CNN-based image recognition and classification model perhaps first applied in the competition of ImageNet [8]. After then CNN-based models are currently considered in various applications, for example, image segmentation in medical image processing, feature extraction from images, finding region of interests, object detection, natural language processing, etc.

CNN has incredible trainable parameters at the various layers that are applied for extracting important features at various abstraction levels [9]. Meanwhile, a CNN model needs a huge dataset to train. Particularly in the medical field medical dataset may not be always possible to obtain. Moreover, a CNN model requires high speed computing resources to train and tune its hyper parameters. To overcome the data unavailability transfer learning techniques at present vastly applied in classification of medical images. Applying transfer learning techniques, a model can use knowledge from other the pre-trained models (e.g., VGG16 [10], AlexNet [11], DenseNet [12], etc.) that are trained over a huge dataset to classify images. This lessens the requirement of data linked to the problem we are tackling with. The pre-trained models often are used as feature extractors in images from abstract level to more detailed levels. Transfer learning techniques using pre-trained models have shown promising results in different medical diagnosis, such as chest X-ray image analysis for pneumonia and COVID-19 patients' identification [13], retina image analysis for blind person classification, MRI image analysis for brain tumor classification, etc. Deep learning models leveraging CNNs are used widely to classify breast cancers. We now discuss some of the promising researches that have been proposed using CNN.

Authors in [14] proposed a learning framework leveraging deep learning architecture that can learn features automatically from mammography images in order to identify cancer. The framework was tested on the BCDR-FM dataset. Although they showed improved results, however, did not compare employing pre-trained models. Authors in [15] considered AlexNet as a feature extractor for mass diagnosis in mammography images. Support Vector Machine (SVM) is applied as a classification model after AlexNet generates features. The outcome of the proposed model is higher compared to the analytically feature extractor method. In our approach, we considered eight different pre-trained models and show their performances using ultrasound images. Authors in [16] considered transfer learning approach using GoogleNet [17] and AlexNet pre-trained models and some preprocessing techniques. The model is applied on mammograms images, where cancers are already segmented. The authors claim that the model achieves improved performance than the human involved methods. Authors in [18] proposed a convolutional neural network leveraging Inception-v3 pre-trained model to classify breast cancer using breast ultrasound images. The model supports facility for extracting Multiview features. The model is trained only on 316 images and achieved 0.9468 AUC, 0.886 sensitivity, and 0.876 specificity.

Authors in [19] developed an ensemble CNN model leveraging VGG192 and ResNet1523 pre-trained models with fine tuning. The authors considered a dataset managed by JABTS. There are 1536 breast masses that include 897 malignant and 639 benign cases. The model achieves 0.951 AUC, 90.9% sensitivity, and 87.0 specificity. Authors in [20] developed another ensemble-based computer aided diagnosis (CAD) system combining VGGNet, ResNet, and DenseNet pre-trained models. They considered a private database that consists of 1687 images that includes 953 benign and 734 malignant cases. The model achieved 91.0% accuracy and 0.9697 AUC score. The model is also tested on the BUSI dataset. In this dataset, the model achieved 94.62% accuracy and 0.9711 AUC score. Authors in [21] implemented two different approaches (1) a CNN and (2) a transfer learning to classify breast cancer from combining two sets of datasets, one containing 780 images and another containing 163 images. The model showed better performance results combining traditional and generative adversarial network augmentation techniques. In the transfer learning approach, the authors compared the performance of four pre-trained models, mainly, VGG16, Inception [22], ResNet, and NASNet [23]. In the combined dataset, the NASNet achieved highest accuracy value 99%. Authors in [24] compared three CNN-based transfer learning models ResNet50, Xception, and InceptionV3, and proposed a base model that consists of three convolutional layers to classify breast cancers from the breast

ultrasound images dataset. The dataset comprised of 2058 images that includes 1370 benign and 688 malignant cases. According to their analysis, InceptionV3 showed best accuracy of 85.13% with AUC score 0.91. Authors in [25] analyzed four pre-trained models VGG16, VGG19, InceptionV3, and ResNet50 on a dataset that consists of 5000 breast images comprised of 2500 benign and 2500 malignant cases. InceptionV3 model achieved the highest AUC of 0.905.

Authors in [26] proposed a CNN model for breast cancer classification considering the local and frequency domain information using Histopathological images. The objective is to utilize the important information of images that are carried by the local and frequency domain information which sometime shows better accuracy for the model. The proposed model is applied on the BreakHis dataset. However, the model obtained 94.94% accuracy.

Authors in [27] proposed a novel deep neural network consisting of clustering method and CNN model for breast cancer classification using Histopathological images. The model is based on CNN, a Long-Short-Term-Memory (LSTM), and a mixture of the CNN and LSTM models. In the model, both Softmax and SVM are applied at the classifier layer. However, the model achieved 91% accuracy.

From the above discussion, it is evident researchers still on the search for a better model to classify breast cancers. In order to overcome the scarcity of datasets, this research combines two publicly available ultrasound image datasets. Then eight different pre-trained models after fine tuning are applied on the combined dataset to observe the performance results of breast cancer classification. However, the pre-trained models did not show expected outcome. Therefore, we also develop a shallow CNN-based model. The model outperforms all the fine-tuned pre-trained models in all the performance metrics. The proposed model is also faster in training. We also employed different evaluation techniques to prove the better outcome of the proposed model. The details of the methods study, evaluation results and discussion are presented in Sect. 3.

The paper is organized as follows: Sect. 2 discusses materials and methods that are used for the purpose of breast cancer classification. Section 3 proposes the custom CNN model. Section 4 discusses evaluation results of the pre-trained models and the proposed custom. Finally, the paper concludes in Sect. 5.

## 2 Materials and methods

In this research, we consider two publicly available breast ultrasound image datasets [28, 29]. The two datasets are considered mainly for two reasons: (1) to increase the size

of the dataset for the training purpose in order to avoid overfitting and biasness and (2) to consider three classes (benign, malignant and normal). Combining the datasets also will improve the reliability of the model. Dataset in [28] contains 250 images in which there are two categories: malignant and benign cases. The size of the images is different. The minimum and the maximum size of the images are  $57 \times 75$  and  $61 \times 199$  pixels with gray and RGB colors, respectively. Therefore, all the images are transformed into gray color to fit into the model. The dataset in [29] contains 780 images, in which there are three categories: malignant, benign, and normal cases. The average image size of the images is  $500 \times 500$  pixels. The breast ultrasound images are collected from 600 women in 2018, and the age range of the women is between 25 and 75 years. Table 1 shows the class distribution of the images in the two datasets. Figure 1 demonstrates examples of ultrasound images of different cases in the two datasets.

Data normalization is an important pre-processing phase before feeding the data into a model for training. With pre-processing the data features become easily interpretable by the model. Lack of correct pre-processing makes the model slow in training and unstable. Generally, standardization and normalization techniques are used in scaling data. Normalization technique rescales the data values between 0 and 1. Since the datasets that are considered In this research, are both gray and color images, hence the values of the pixels lie between 0 and 255. We consider zero-centering approach that shifts the distribution data values in such a way that its mean becomes equal to zero. Assume a dataset  $\mathbf{D}$ , that consists of  $\mathbf{N}$  samples and  $\mathbf{M}$  features. Therefore,  $\mathbf{D}[:, i]$  denotes  $i$ th feature and  $\mathbf{D}[\mathbf{j}, :]$  denotes sample  $\mathbf{j}$ . The equation below defines zero-centering.

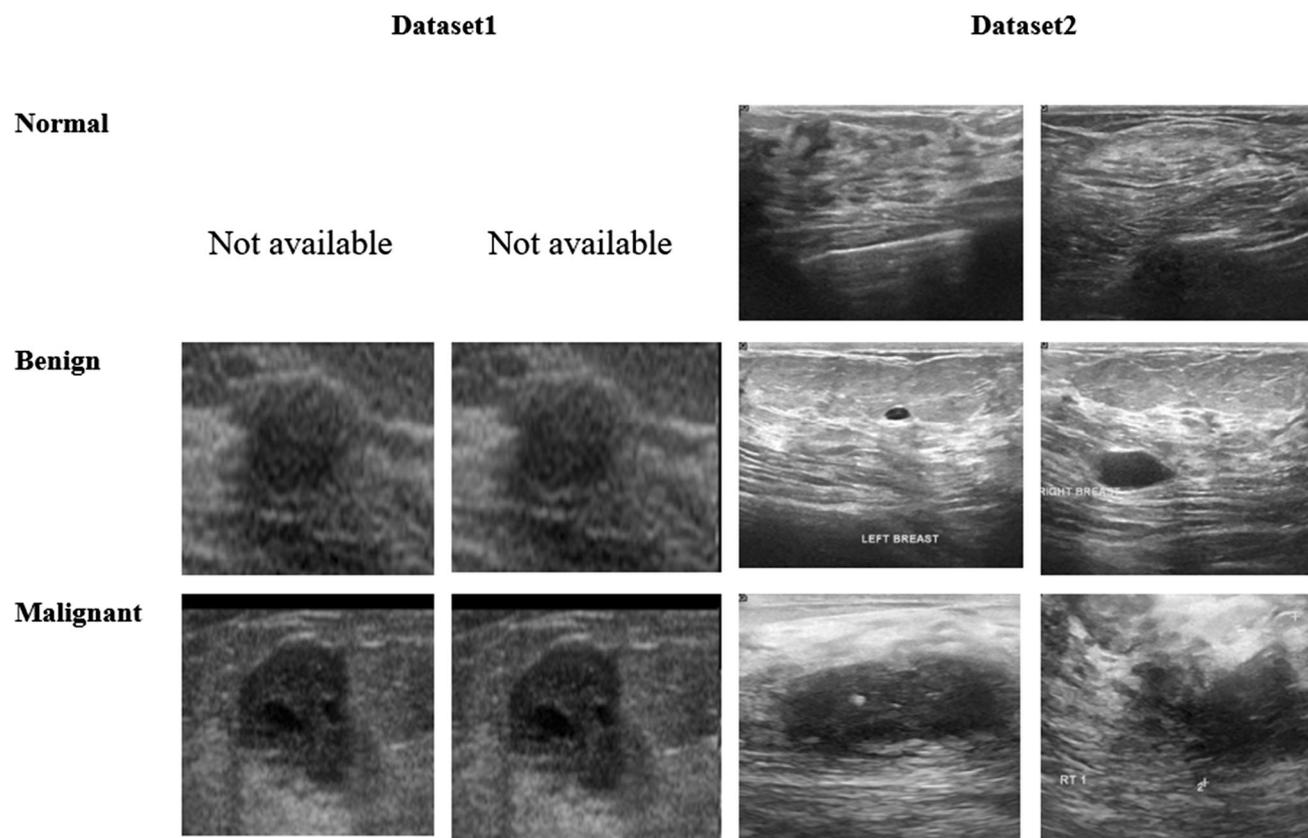
$$\hat{D}[:, i] = \frac{D[:, i] - \mu_i}{\sigma_i}, \quad \text{where } \mu_i = \frac{1}{N} * \sum_{k=1}^N D[k, i] \text{ and}$$

$$\sigma_i = \sqrt{\frac{1}{N-1} * \sum_{k=1}^N (D[k, i] - \mu_i)^2}$$

In this research, we employed  $k$ -fold ( $k = 5$ ) cross-validation on the dataset to overcome overfitting problem during model training. In  $k$ -fold cross validation method,  $K$  different datasets of same size is generated, where each

**Table 1** Datasets and distribution of classes

Datasets	Benign	Malignant	Normal	Total samples
Dataset 1	100	150	–	250
Dataset 2	437	210	133	780
Total	537	360	133	1030



**Fig. 1** Sample images of breast ultrasound images of different cases in the two datasets

fold is used to validate the model, and  $k-1$  folds are considered for the purpose of model training. This ensures that the model produces reliable accuracy. Cross-validation is a widely used mechanism to resample data for evaluating machine learning models when the dataset sample size is small. Cross-validation is mainly considered to approximate the learning skill of a machine learning model using data which the model has not seen previously. The result of a model obtained using cross-validation is normally less biased or gives optimistic estimation skill of the model compared to train/test split method. Table 2 shows how the fivefold cross-validation generates five different datasets of ultrasound images from the two datasets.

## 2.1 Fine-tuned pre-trained CNN models and the proposed custom model

During the last few years, transfer learning algorithms are widely used in many research problems in machine learning which concentrate on preserving knowledge acquired during unraveling one problem and employing the knowledge into another but a relevant problem. For example, an algorithm that is trained to learn in recognizing dogs can be applied to recognize horses. Authors in [30] formally

define the transfer learning in terms of domain and task as follows:

Let an arbitrary domain  $D = \{X, P(X)\}$ . Here  $X$  denotes a feature vector  $\{x_1, x_2, \dots, x_n\}$  and the probability distribution in  $X$  is denoted by  $P(X)$ . A task  $T$  on  $D$  is defined as  $T = \{Y, F(\cdot)\}$ , where  $Y = \{y_1, y_2, \dots, y_n\}$  denotes the domain of labels of  $X$  and  $F(\cdot)$  denotes a predictive function which learns from  $\{x_i, y_i\} \in \{X, Y\}$ . Then  $F(\cdot)$  is applied to predict the corresponding label  $F(x')$  in an unknown instance  $x'$ .

Hence, the transfer learning is defined as: consider two pairs of domains  $(D_A, T_A)$  and  $(D_B, T_B)$ , where,  $D_A$  is a source domain and  $D_B$  is a target domain.  $T_A$  and  $T_B$  are learning tasks of  $D_A$  and  $D_B$ , respectively. The goal of a transfer learning technique is to enhance the learning of the predictive function  $F_B(\cdot)$  in  $D_B$  by applying the knowledge learned from  $D_A$  and  $T_A$ , such that  $D_A \neq D_B, T_A \neq T_B$ .

One of the reasons that the transfer learning algorithms being used when small size dataset is available to train a custom model, but the goal is to produce an accurate model. A custom model employing transfer learning, applies the knowledge of the pre-trained models that are trained over a huge dataset for a long duration. There are mainly two approaches to apply transfer learning: (i) model

**Table 2** Data distribution in fivefolds after applying fivefold cross validation

	Train			Test		
	Benign	Malignant	Normal	Benign	Malignant	Normal
Fold 1	429	288	107	108	72	26
Fold 2	429	288	107	108	72	26
Fold 3	430	288	106	107	72	27
Fold 4	430	288	106	107	72	27
Fold 5	430	288	106	107	72	27

developing and (ii) using pre-trained models. The pre-trained model approach is widely used in deep learning domain. Considering the importance of the pre-trained models as feature extractors this research implements eight pre-trained models using the weights of the convolutional layers of the pre-trained models. These weights act as feature extractors for classifying breast cancers applying on the ultrasound images. Table 3 shows the pre-trained models that are considered in this research. All the models are built on convolutional neural network and were trained on the ImageNet database [31] that consists of a million images. The models can classify 1000 objects (mouse, keyboard, pencil, and many animals) from different images. Therefore, all the models have learned huge feature representations from a large number of images.

From the Table 3, we see that different models use different input size. Therefore, the images in the dataset are transformed accordingly to feed into the models. In the fine-tuning process of the pre-trained models, the final layer is substituted with a classifier that can classify three objects since the dataset consists of images with three classes (normal, malignant, and benign). Hence the models are fine tuned at the top layers. In the fine-tuning process the last three layers of the models are substituted with (i) a fully connected layer (ii) Softmax activation layer, and (iii) a custom classifier.

We considered three different optimizers to train the models and to determine which model produces the best

**Table 3** Pre-trained models and their image input size

Pre-Trained models	Number of convolutional layers	Input size
AlexNet	8	227 × 227
DarkNet19	19	256 × 256
GoogleNet	22	224 × 224
MobileNet-v2	53	224 × 224
ResNet18	18	224 × 224
ResNet50	50	224 × 224
VGG16	16	224 × 224
Xception	71	229 × 229

results. The brief description of the optimizers is given below:

Stochastic Gradient Descent with Momentum (SGDM) is the fundamental optimizer in neural network that is used for the convergence of neural networks, i.e., moving in the direction of the optimum cost function. The following equation is used to update neural network parameters to calculate the gradient  $\nabla$ .

$$\theta = \theta - \mu * \nabla_{\theta} J(\theta; x, y)$$

here  $\theta$ : a parameter (weights, biases and activations),  $\mu$ : learning rate,  $\nabla$ : is the gradient, and  $J$  is cost function. Root Mean Square Propagation (RMSprop) is formulated by Geoffrey Hinton. RMSprop attempts to reduce the oscillations. It also adjusts learning rate automatically. In addition, for each parameter RMSprop selects a different learning rate. In RMSprop, an update is done according to the equations described below.

For each parameter  $w_j$

$$v_t = \rho v_{t-1} + (1 - \rho) * g_t^2, \quad \Delta w_t = -\frac{\mu}{\sqrt{v_t + \epsilon}} * g_t, \\ w_{t+1} = w_t + \Delta w_t$$

here  $\mu$ : initial learning,  $v_t$ : exponential average of squares of gradients, and  $g_t$ : gradient at time along  $w_j$ . Adam optimizer associates the heuristics of momentum and RMSprop. The equation is given below.

For each parameter  $w_j$

$$v_t = \beta_1 * v_{t-1} - (1 - \beta) * g_t, \quad s_t = \beta_2 * s_{t-1} - (1 - \beta_2) * g_t^2, \\ \Delta w_t = -\tau \frac{v_t}{\sqrt{s_t + \epsilon}} * g_t, \quad w_{t+1} = w_t + \Delta w_t$$

here  $\mu$ : initial learning,  $v_t$ : exponential average of gradients  $w_j$  and  $g_t$ : gradient at time t along  $w_j$   $s_t$ : exponential average of squares of gradients along  $w_j$ ,  $\beta_1, \beta_2$  are hyperparameters.

The fine-tuned pre-trained models used Softmax activation function to generate the probability between the range 0 and 1 of the class outcomes from the input images. Using Softmax activation function at the end of a CNN model to convert its outcome scores into a normalized

probability distribution is a very well-known practice. Softmax function is defined with following equation:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}}$$

where  $\vec{z}$  is a input vector,  $z_i$  are the elements in  $\vec{z}$ ,  $e^{z_i}$  is the exponential function, and  $\sum_{j=1}^k e^{z_j}$  is the normalization term.

### 3 Proposed custom model

Figure 2 shows the proposed custom model. The model consists of one convolutional layer consists of 20 filters of  $5 \times 3 \times 3$  with stride [1 1] and padding [0 0 0]. The input image size for the model is  $227 \times 227 \times 3$ , and the images are normalized with the zero-centering approach. The model applies batch normalization with 20 channels. It also consists of one max pooling layer, one fully connected layer. Dropout regularization is also added after the fully connected layer. Finally, SoftMax activation function is applied since the model needs to classify three classes. The initial learning rate  $1.0000e-04$  is considered during training. The model also considers mini-batch size 8. The model is trained using three optimizers as the pre-trained models are trained. The model is trained and validated using the configuration of Table 4.

### 4 Performance measure

The performance of the fine-tuned pre-trained models are evaluated with various standard performance of metrics. The metrics are accuracy (ACC), Area Under Curve (AUC), precision, recall, sensitivity, specificity, and F1-score. Confusion matrix for each model is also generated to observe the scores of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) of normal, malignant, and benign cases. The TP (e.g.,

**Table 4** Training parameters of the pre-trained models

Training options	Value
Initial learn rate	$1.0000e-04$ or $10e-5$
Learn rate schedule	Constant or piece wise
Mini-batch size	8
Shuffle	Every epoch
Optimizer	Adam, RMSprop, SGDM
Cross validation	5
Max epochs	20
ExecutionEnvironment	gpu

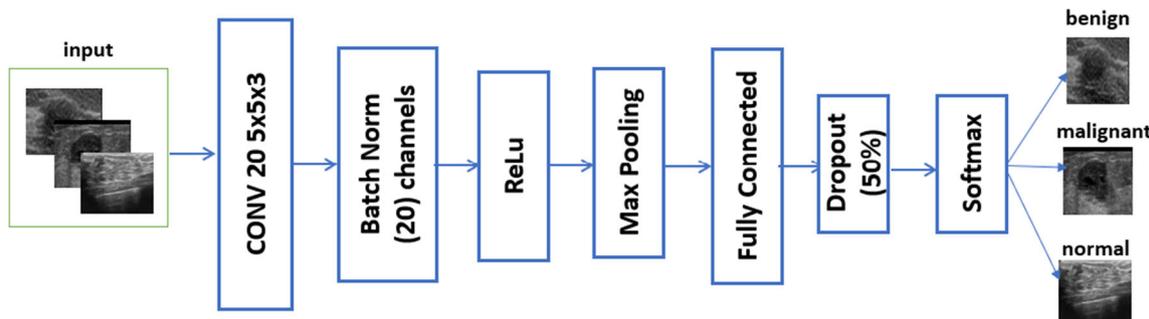
malignant) score represents how the model correctly classify real malignant cases as malignant. The FP (e.g., malignant) represents how the model wrongly classifies benign cases as malignant. Similarly, TN (e.g., benign) score represents how the model correctly classifies benign cases as benign, and FN (e.g., benign) score represents how the model wrongly classifies malignant cases as benign.

Another important metric is the precision that demonstrates the performance of a model in terms of proportion of the truly classified patients as malignant, benign, and normal cases. Meanwhile, sensitivity or recall value shows the proportion of a case (e.g., malignant) a model truly classifies as malignant cases. Specificity demonstrates the percentage of a case (e.g., benign) that a model classifies correctly. Through the F1-score we achieve a single score from precision and recall through evaluating their harmonic mean. In the below, we show the formula of different metrics.

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}, \text{ recall} = \frac{TP}{TP + FN},$$

$$\text{precision} = \frac{TP}{FP + TP}$$

$$\text{specificity} = \frac{TN}{FP + TN}, F1 = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})}$$



**Fig. 2** The architecture of the custom model

**Table 5** Evaluation results of the pre-trained models and the custom model

Models	Optimizer	AUC	Accuracy	Sensitivity	Specificity	Precision	Recall
AlexNet	SGDM	0.967	89.71	90.88	88.44	89.54	90.88
	Adam	0.943	84.85	85.85	83.77	85.21	85.85
	RMSprop	0.919	82.62	81.38	83.98	84.69	81.38
Darknet19	SGDM	0.948	87.18	88.27	86	87.29	88.27
	Adam	0.958	88.35	88.27	88.44	89.27	88.27
	RMSprop	0.959	88.93	89.57	88.24	89.24	89.57
GoogleNet	SGDM	0.959	88.93	91.25	86.41	87.97	91.25
	Adam	0.966	90.49	93.11	87.63	89.13	93.11
	RMSprop	0.96	91.26	93.85	88.44	89.84	93.85
MobileNet	SGDM	0.944	85.73	85.1	86.41	87.21	85.1
	Adam	0.969	90.19	89.01	91.48	91.92	89.01
	RMSprop	0.961	90.68	90.32	91.08	91.68	90.32
ResNet18	SGDM	0.948	87.67	89.94	85.19	86.87	89.94
	Adam	0.967	90	90.13	89.86	90.64	90.13
	RMSprop	0.97	91.26	90.32	92.29	92.73	90.32
ResNet50	SGDM	0.952	87.77	83.46	88.41	51.63	83.46
	Adam	0.971	92.04	93.23	91.86	62.94	93.23
	RMSprop	0.971	91.84	90.23	92.08	62.83	90.23
VGG16	SGDM	0.965	89.9	91.99	87.63	89.01	91.99
	Adam	0.972	89.9	92.92	86.61	88.32	92.92
	RMSprop	0.955	90.29	92.92	87.42	88.95	92.92
Xception	SGDM	0.924	83.11	83.8	82.35	83.8	83.8
	Adam	0.968	90.68	89.94	91.48	92	89.94
	RMSprop	0.968	91.17	91.43	90.87	91.6	91.43

**Table 6** Comparison results of the best pre-trained models and the custom model

Pre-trained models				Custom model	
Performance metric	Score	Models	Optimizers	Score	Optimizer
AUC	0.972	VGG16	Adam	1.0	Adam, RMSprop
Accuracy	92.04	ResNet50	Adam	100	Adam
Sensitivity	93.85	GoogleNet	RMSprop	100	SGDM, Adam, RMSprop
Specificity	92.29	ResNet18	RMSprop	100	Adam
Precision	92.73	ResNet50	RMSprop P	100	Adam
Recall	93.85	GoogleNet	RMSprop	100	Adam

### 5 Results and discussion

The scores of performance evaluation of the fine-tuned pre-trained models as well as the custom model are shown in Table 5. Among the pre-trained models ResNet50 demonstrates the highest accuracy score 92.04 with Adam optimizer and the highest precision score 92.73 with RMSprop optimizer. VGG16 achieves the highest AUC score 0.972 with Adam optimizer. GoogleNet achieved the highest score 93.85 in sensitivity and recall using RMSprop optimizer. ResNet18 achieved the highest score 92.29 in

specificity metric. Compared to all the pre-models the proposed custom model outperforms in all the evaluation metrics. The proposed model obtained 100% accuracy and 1.0 AUC score using Adam optimizer. In other metrics, the proposed model also shows better results. Table 6 summarizes the models’ performance with the best scores with different evaluation metrics and compares the results with the proposed CNN model.

Figure 3 shows the confusion matrix generated from the different pre-trained models as well as the proposed custom model. The figure only shows the confusion matrix of the

### ResNet50 with Adam optimizer

**Confusion Matrix: ResNetADAM**

Output Class	benign	494 48.0%	26 2.5%	7 0.7%	93.7% 6.3%
	malignant	28 2.7%	330 32.0%	2 0.2%	91.7% 8.3%
	normal	15 1.5%	4 0.4%	124 12.0%	86.7% 13.3%
		92.0% 8.0%	91.7% 8.3%	93.2% 6.8%	92.0% 8.0%
	benign	malignant	normal		
					<b>Target Class</b>

### VGG16 with Adam optimizer

Output Class	benign	499 48.4%	44 4.3%	11 1.1%	90.1% 9.9%
	malignant	19 1.8%	310 30.1%	1 0.1%	93.9% 6.1%
	normal	19 1.8%	6 0.6%	121 11.7%	82.9% 17.1%
		92.9% 7.1%	86.1% 13.9%	91.0% 9.0%	90.3% 9.7%
	benign	malignant	normal		
					<b>Target Class</b>

### ResNet18 with RMSprop optimizer

**Confusion Matrix: ResNet18**

Output Class	benign	485 47.1%	25 2.4%	7 0.7%	93.8% 6.2%
	malignant	32 3.1%	331 32.1%	2 0.2%	90.7% 9.3%
	normal	20 1.9%	4 0.4%	124 12.0%	83.8% 16.2%
		90.3% 9.7%	91.9% 8.1%	93.2% 6.8%	91.3% 8.7%
	benign	malignant	normal		
					<b>Target Class</b>

### GoogleNet with RMSprop optimizer

Output Class	benign	504 48.9%	37 3.6%	14 1.4%	90.8% 9.2%
	malignant	20 1.9%	319 31.0%	2 0.2%	93.5% 6.5%
	normal	13 1.3%	4 0.4%	117 11.4%	87.3% 12.7%
		93.9% 6.1%	88.6% 11.4%	88.0% 12.0%	91.3% 8.7%
	benign	malignant	normal		
					<b>Target Class</b>

### Custom Model with Adam

Output Class	benign	537 52.1%	0 0.0%	0 0.0%	100% 0.0%
	malignant	0 0.0%	360 35.0%	0 0.0%	100% 0.0%
	normal	0 0.0%	0 0.0%	133 12.9%	100% 0.0%
		100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
	benign	malignant	normal		
					<b>Target Class</b>

### Custom Model with SGDM

Output Class	benign	537 52.1%	1 0.1%	0 0.0%	99.8% 0.2%
	malignant	0 0.0%	359 34.9%	0 0.0%	100% 0.0%
	normal	0 0.0%	0 0.0%	133 12.9%	100% 0.0%
		100% 0.0%	99.7% 0.3%	100% 0.0%	99.9% 0.1%
	benign	malignant	normal		
					<b>Target Class</b>

Fig. 3 Confusion matrix of different pre-trained models and the custom model

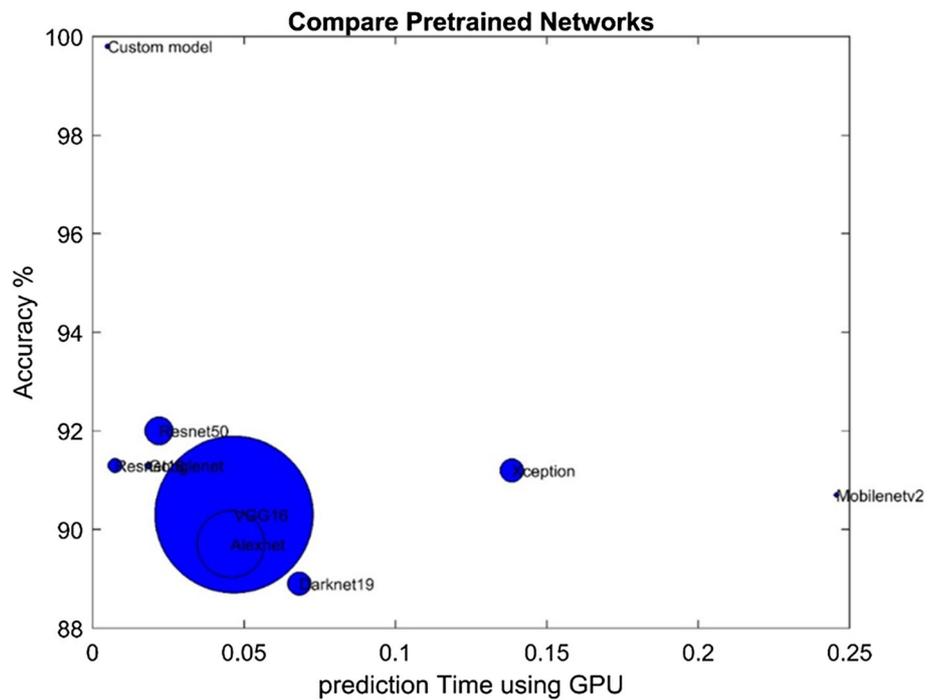
**Table 7** Comparison of classification results (the custom model (Adam optimizer) and best pre-trained models)

Models	Benign (537 original case)	Malignant (360 original case)	Normal (133 original case)
Custom model	537	360	133
GoogleNet	504	319	117
ResNet18	485	331	124
ResNet50	494	330	124
VGG16	499	310	121

**Table 8** Comparison of the pre-trained models and the custom model (accuracy, prediction time, and parameters)

Performance comparison									
Model	Custom model	AlexNet	DarkNet19	GoogleNet	MobileNet	ResNet50	ResNet18	VGG16	Xception
Accuracy	100	89.7	88.9	91.3	90.7	92	91.3	90.3	91.2
Prediction time (second)	0.00495	0.04564	0.06824	0.018441	0.245589	0.02194	0.007419	0.04667	0.13845
Number of parameters	2985263	56880515	20838179	5976627	4233133	23487619	11173251	1.36E + 08	20808883

**Fig. 4** Performance comparison of different pre-trained models and the custom model



best pre-trained models as mentioned in the Table 6. From the confusion matrix of the custom model, we observe that in all the classification of breast cancers the score is high. For example, the model classifies 100% benign class, 100% malignant class, and 100% normal classes using Adam optimizer. The results also outperform the results of the

pre-trained models. Table 7 shows the classification results of the models.

Table 8 shows performance comparison results between the custom and the pre-trained models. The custom model outperforms all the pre-trained models with respect to accuracy, prediction time and number of parameters. The custom model is also very fast in training than all the fine-

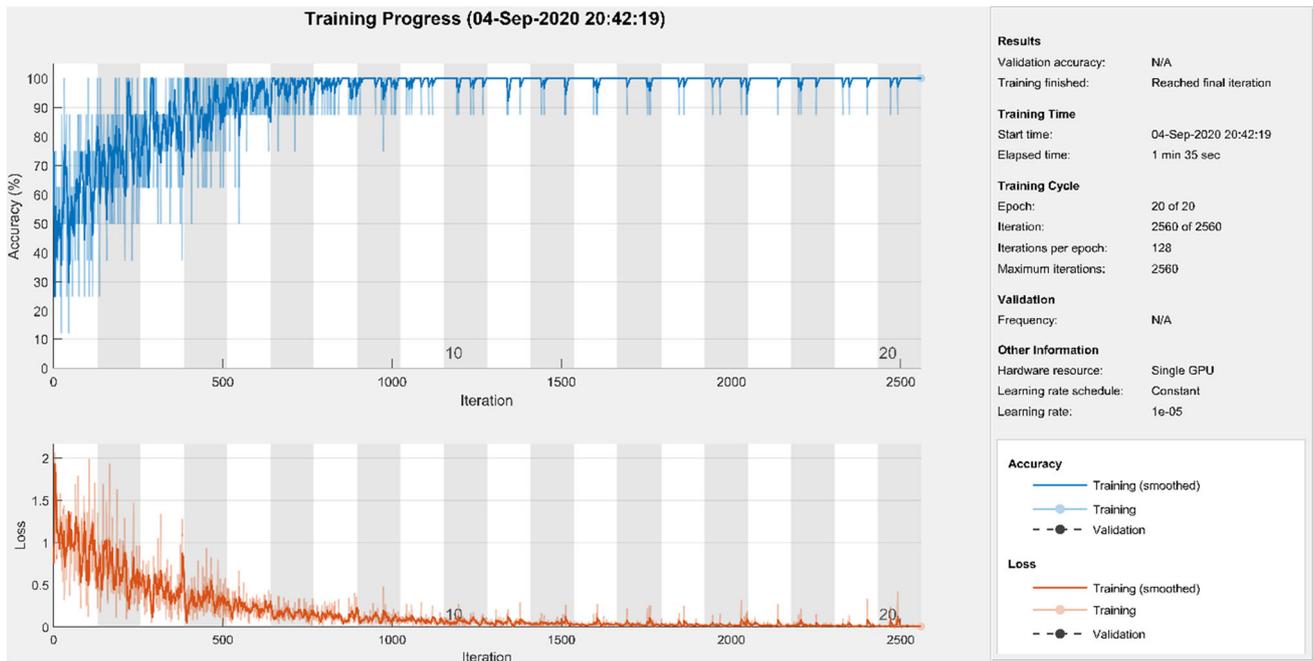


Fig. 5 Accuracy and loss of the custom model during training phase

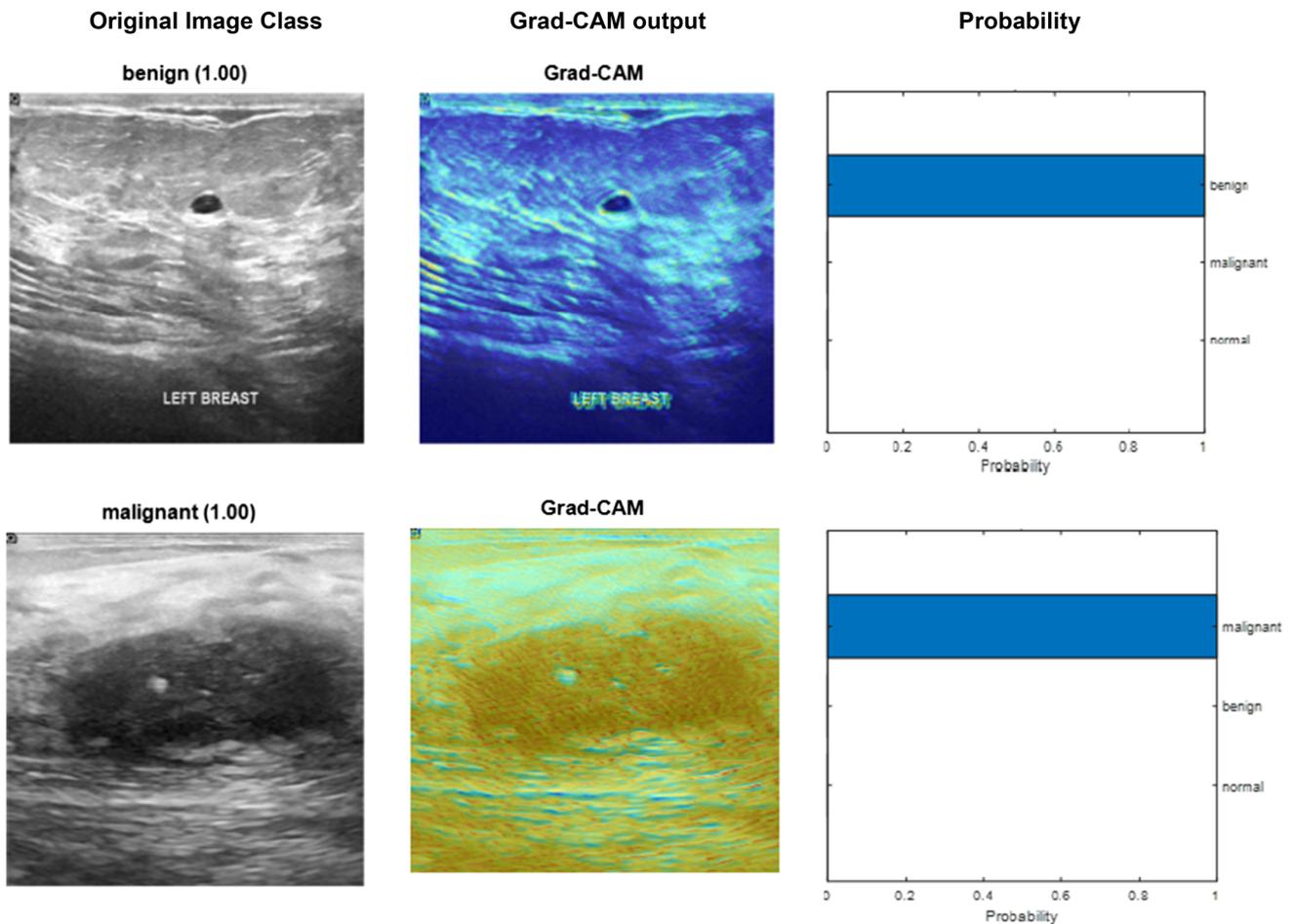


Fig. 6 Heat map visualization of same images using the custom model

tuned pre-trained models. The reason is that the custom model has only one fully connected layer. In addition, the custom model requires a very small number of trainable parameters compared to the other models. All the models are trained in a GPU (NVIDIA® GeForce GTX 1660 Ti with Max-Q design and 6 GB RAM) considering a mini-batch size of 8. Figure 4 shows execution time and the accuracy score of each model. To calculate accurate time, we run the code four times. The area of each marker in the Fig. 4 shows the size of the number of parameters in the networks. The time of models' prediction is calculated with respect to the fastest network. From the plot, it is quite evident that that custom model is fast and training and produces higher accuracy than the other pre-trained models. Figure 5 shows the accuracy and loss values when the custom model is trained and validated. From the graph in Fig. 5, it is evident that the custom model generates very high accuracy result as claimed in Table 8.

### 5.1 Heat map visualization

The custom model's performance is also evaluated by generating heat map visualization using Grad-CAM tool [32] to see how the model identifies the region of interest and how well the model distinguishes cancer classes. Grad-CAM is used to judge whether a model identifies the key areas in the images for prediction. Grad-CAM visualizes the portion of an image through heatmap of a class label that the model focuses for prediction. Figure 6 shows a sample Grad-CAM output of benign and malignant classes and prediction probability. From the output, we observe that the model perfectly focuses on the key areas of images to classify cancers.

## 6 Conclusion

This study implemented eight pre-trained CNN models with fine tuning leveraging transfer learning to observe the classification performance of breast cancer from ultrasound images. The images are combined from two different datasets. We evaluated the fine-tuned pre-trained models applying the Adam, RMSprop, and SGDM optimizers. The highest accuracy 92.4% is achieved by the ResNet50 with Adam optimizer and the highest AUC 0.97 score is achieved by VGG16. We also proposed a shallow custom model since the pre-trained models have not shown expected results and all the pre-trained models have many convolutional layers and need long duration in the training phase. The proposed custom model consists of only one convolutional layer as feature extractors. The custom model achieved 100% accuracy and 1.0AUC value. With respect to training time, the custom model is faster than any

other model and needs small size of trainable parameters. The future plan is to validate the model with other datasets that include new ultrasound images.

**Acknowledgements** The authors are grateful to the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia for funding this work through the Vice Deanship of Scientific Research Chairs: Chair of Pervasive and Mobile Computing.

**Funding** Not applicable.

**Availability of data and material** Datasets are collected from public repository [28, 29].

### Compliance with ethical standards

**Conflicts of interest** Not applicable.

**Code availability** Not applicable.

## References

1. Breast cancer: prevention and control, WHO. <https://www.who.int/cancer/detection/breastcancer/en/index1.html#:~:text=Breast%20cancer%20survival%20rates%20vary,et%20al.%2C%202008>. Accessed 15 Aug 2020
2. Hossain MS (2017) Cloud-supported cyber-physical localization framework for patients monitoring. *IEEE Syst J* 11(1):118–127
3. Hossain MS, Amin SU, Muhammad G, Sulaiman M (2019) Applying deep learning for epilepsy seizure detection and brain mapping visualization. *ACM Trans Multimed Comput Commun* 15(1):1–17
4. Garg S, Kaur K, Kumar N, Rodrigues JJPC (2019) Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN: a social multimedia perspective. *IEEE Trans Multimed* 21(3):566–578. <https://doi.org/10.1109/tmm.2019.2893549>
5. Ghoneim A, Muhammad G, Hossain MS (2020) Cervical cancer classification using convolutional neural networks and extreme learning machines. *Future Gener Comput Syst* 102:643–649
6. Hossain MS, Al-Hammadi M, Muhammad G (2019) Automatic fruit classification using deep learning for industrial applications. *IEEE Trans Ind Inf* 15(2):1027–1034
7. Hossain MS, Muhammad G (2019) Emotion recognition using secure edge and cloud computing. *Inf Sci* 504:589–601
8. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. *Int J Comput Vis* 115:211–252. <https://doi.org/10.1007/s11263-015-0816-y>
9. Yang X et al (2016) Deep relative attributes. *IEEE Trans Multimed* 18(9):1832–1842
10. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. arXiv reprint <https://arxiv.org/abs/1409.1556>
11. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Proceedings of the 26th Neural Information Processing Systems (NIPS' 12), December 3–8, Harrahs and Harveys, Lake Tahoe. <https://doi.org/10.1145/3065386>
12. Huang G, Liu Z, Maaten L, Weinberger KQ (2018) Densely connected convolutional network. arXiv:1608.06993. <https://arxiv.org/abs/1608.06993>

13. Hossain MS, Muhammad G, Guizani N (2020) Explainable AI and Mass Surveillance System-Based Healthcare Framework to Combat COVID-19 Like Pandemics. *IEEE Netw* 34(4):126–132
14. Arevalo J, González FA, Ramos-Pollán R, Oliveira JL, Lopez MAG (2016) Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput Methods Programs Biomed* 127:248–257. <https://doi.org/10.1016/j.cmpb.2015.12.014>
15. Huynh BQ, Li H, Giger ML (2016) Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J Med Imaging*. <https://doi.org/10.1117/1.jmi.3.3.034501>
16. Yuan-Pin L, Tzyy-Ping J (2017) Improving EEG-based emotion classification using conditional transfer learning. *Front Hum Neurosci*. <https://doi.org/10.3389/fnhum.2017.00334>
17. Szegedy C, Wei L, Yangqing J, Pierre S, Scott R, Dragomir A, Dumitru E, Vincent V, Rabinovich A (2015) Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. <https://arxiv.org/abs/1409.4842>
18. Wang Y, Choi EJ, Choi Y, Zhang H, Jin GY, Ko SB (2020) Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. *Ultrasound Med Biol* 46(5):1119–1132. <https://doi.org/10.1016/j.ultrasmedbio.2020.01.001>
19. Tanaka H, Chiu SW, Watanabe T, Kaoku S, Yamaguchi T (2019) Computer-aided diagnosis system for breast ultrasound images using deep learning. *Ultrasound Med Biol*. <https://doi.org/10.1016/j.ultrasmedbio.2019.07.426>
20. Moon WK, Lee YW, Ke HH, Lee SH, Huang CS, Chang RF (2020) Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput Methods Programs Biomed*. <https://doi.org/10.1016/j.cmpb.2020.105361>
21. Walid AD, Gomaa M, Khaled H, Fahmy A (2019) Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. *Int J Adv Comput Sci Appl*. <https://doi.org/10.14569/IJACSA.2019.0100579>
22. Szegedy C, Vanhoucke V, Ioffe, Shlens J, Wojna Z (2015) Rethinking the Inception Architecture for Computer Vision. arXiv reprint [arXiv:1512.00567](https://arxiv.org/abs/1512.00567). <https://arxiv.org/abs/1512.00567>
23. Barret Z, Vasudevan V, Shlens J, Quoc VL (2017) Learning transferable architectures for scalable image recognition. arXiv preprint [arXiv:1707.07012](https://arxiv.org/abs/1707.07012). <https://arxiv.org/abs/1707.07012>
24. Xiao T, Liu L, Li K, Qin W, Yu S, Li Z (2018) Comparison of transferred deep neural networks in ultrasonic breast masses discrimination. *Biomed Res Int*. <https://doi.org/10.1155/2018/4605191>
25. Zhang H, Han L, Chen K et al (2020) Diagnostic efficiency of the breast ultrasound computer-aided prediction model based on convolutional neural network in breast cancer. *J Digit Imaging*. <https://doi.org/10.1007/s10278-020-00357-7>
26. Nahid AA, Kong Y (2018) Histopathological breast-image classification using local and frequency domains by convolutional neural network. *Information* 2018(9):19. <https://doi.org/10.3390/info9010019>
27. Nahid AA, Mehrabi MA, Kong Y (2018) Histopathological breast cancer image classification by deep neural network techniques guided by local clustering. *Biomed Res Int*. <https://doi.org/10.1155/2018/2362108>
28. Paulo SR (2017) Breast ultrasound image. Mendeley data. <https://doi.org/10.17632/wmy84gzngw.1>
29. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A (2019) Dataset of breast ultrasound images. *Data Brief*. <https://doi.org/10.1016/j.dib.2019.104863>
30. Kathuria A (2018) Intro to optimization in deep learning: momentum, RMSProp and Adam. <https://blog.paperspace.com/intro-to-optimization-momentum-rmsprop-adam/>
31. ImageNet. <http://www.image-net.org/> Accessed 15 Aug 2020
32. Ramprasaath RS, Michael C, Abhishek D, Ramakrishna V, Devi P, Dhruv B (2019) Grad-CAM: Visual explanations from deep networks via gradient-based localization. arXiv preprint [arXiv:1610.02391](https://arxiv.org/abs/1610.02391) <https://arxiv.org/abs/1610.02391>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.