



# Detection of Breast Cancer from Mammogram Images Using Deep Transfer Learning

Akalpita Das<sup>1(✉)</sup>, Himanish Shekhar Das<sup>2</sup>, Utpal Barman<sup>1</sup>,  
Arijeet Choudhury<sup>3</sup>, Sourav Mazumdar<sup>3</sup>, and Anupal Neog<sup>3</sup>

<sup>1</sup> Girijananda Chowdhury Institute of Management and Technology,  
Guwahati 781017, Assam, India

akalpita\_cse@gimt-guwahati.ac.in,  
utpalbelsor@gmail.com

<sup>2</sup> Cotton University, Guwahati 781001, Assam, India  
hsdas0815044@gmail.com

<sup>3</sup> Jorhat Engineering College, Jorhat 785007, Assam, India  
arijeetchoudhury100@gmail.com, souravmzdr@gmail.com,  
neog.anupall01@gmail.com

**Abstract.** Among all types of cancers found in women, breast cancer is having the second highest mortality rate and it is also considered to be the prime cause of high death rate. Mammographic images are often investigated by the experienced and trained radiologists to recognize the breast abnormalities like masses and micro-calcifications. This paper focuses on computer aided diagnosis to help the radiologists so that breast cancer can be detected with better accuracy. In this paper, our aim is to process the mammogram images for breast cancer affected patients using deep learning architectures. In this paper, two approaches are considered. In first approach, mammogram images are feed into various pre-trained models which are trained from scratch. In second approach, exactly same techniques are being replicated with the exception that fine tuning has been performed using transfer learning for all the models. For this work, we have considered variants of convolutional neural networks. Experimental works have performed on two different datasets and for both datasets, the results of the fine-tuned network outperforms all other approaches.

**Keywords:** Breast cancer · Medical imaging · Deep learning · Convolutional Neural Networks · Transfer learning

## 1 Introduction

According to the report of International Agency for Research on Cancer (IARC) presented by the World Health Organization (WHO), in the year 2012, it has been found that women are mostly affected by the breast cancer and it is considered to be the second deadliest disease in women [1]. The abnormal growth of the tumor cells followed by invasion in to the surrounding tissues leads to cancer in the human body. In general, tumors are classified into benign and malignant. The cells of the benign tumor are non-cancerous in nature whereas the cells present in malignant tumor are considered as cancerous. The cells of benign tumor grow locally and they are unable to

spread to the surrounding tissues by invasion. On the other hand, the malignant tumor cells can grow uncontrollably and also invade to the surrounding tissues and later on to various parts of the body [2]. There are different screening methods used to diagnosis breast cancer and among all mammography is the most efficient screening technique used. In mammography, different kinds of views like Mediolateral oblique (MLO, craniocaudal (CC) views are used for better understanding of the abnormalities present in the breasts. To discriminate the abnormality found in breast among benign, normal or malignant classes, radiologists use either of these MLO or CC views to analyze the breast lesion markers such as masses and micro-calcifications [3]. The interpretation of the mammographic images is not only time consuming but also requires the expertise of a highly experienced and trained radiologist [4]. Due to these issues, now a day, Computer assisted diagnosis/detection (CAD) tools are in demand to automate the medical image analysis.

In last decade, deep learning being the major breakthrough technology used in various field of research also proves its emergence in healthcare research. Different deep architectures [5] are explored and efficiently used in the breast cancer research which showed promising results too. A good number of researchers are still carrying out research to diagnose breast cancer using computer aided diagnosis process. Arevalo et al. [6] proposed a hybrid Convolutional Neural Network (CNN) method in which supervised learning technique is used to learn image based handcrafted features. For the problem of mammographic tumor classification, Huynh et al. [7] used transfer learning on the pre-trained AlexNet model without fine-tuning it and for the classification at the back-end, Support Vector Machine (SVM) algorithm is used.

In this paper, an attempt has been made to develop an automated system that has been trained with variants of convolutional neural network architectures using mammographic images using Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) and INbreast datasets. In this study, the investigations are performed for the classification of breast cancer between benign or malignant classes. Pre-trained models are trained from scratch in first approach, thereafter fine tuning is performed using transfer learning in the second approach. For this work, we have considered variants of convolutional neural networks such as VGG19, ResNet50, MobileNet-v2, Inception-v3, Xception and Inception-ResNet-v2 respectively. The existing research works were mostly focused on various breast cancer databases with their accuracy achieved from various machine learning approaches as well as proposed techniques. But this work mainly focuses on how different deep learning models behave for full mammogram images and region of interest (ROI) images for a particular database. The main contribution of this paper is as follows:

- Utilization of transfer learning for fine tuning and regularization approaches in breast cancer detection.
- To analyze the usefulness of various pre-trained models which are trained from scratch without tweaking the deep architectures.
- Impact of transfer learning approach on the pre-trained convolutional neural networks using regularization technique.

The organization of the paper is as follows: Sect. 2 presents literature survey. Section 3 presents the dataset description. Section 4 elaborates proposed methodology. Experimental results and discussions are presented in Sect. 5 and Sect. 6 concludes the paper.

## 2 Related Work

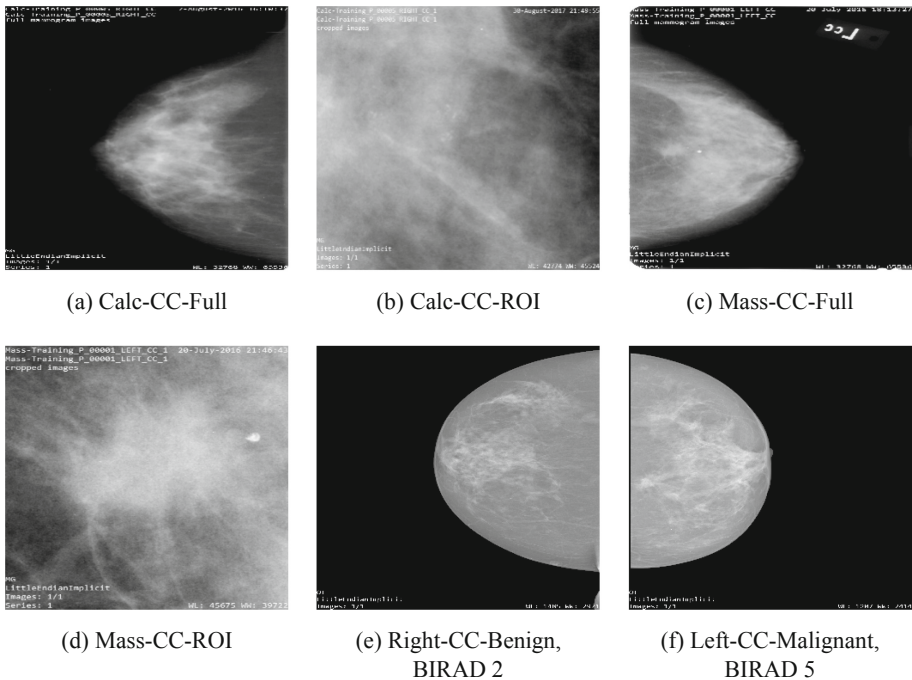
Currently in healthcare sector, the most recent advances of deep learning technologies play a potential role to improve the quality of medication. Carneiro et al. [8] used fine tuning on ImageNet, a pre-trained CNN to discriminate micro-calcification and masses. Breast Imaging-Reporting and Data System (BI-RADS) score is very helpful to classify the breast cancer type. In authors in their work also used BI-RADS score to discriminate the breast cancer classification. For the classification of mammographic images, Levy and Jain [9] used transfer learning on two pre-trained models such as AlexNet as well as GoogleNet. They compared both the networks where AlexNe was found to be the one with superior results. Ting et al. [10] proposed a new network named Convolutional Neural Network Improvement for Breast Cancer Classification (CNNI-BCC) network. For BC classification, the proposed model was trained from scratch. The experiment was conducted on MIAS database and the region of interests (ROIs) detected by a one-shot detector was fed to the model. After training with all CNN, Rampun et al. [11] used the three best performing models predictions in an ensemble and modified version of AlexNet model. Tsochatzidis et al. [12] carried out a detail survey where they compared different architectures such as AlexNet, VGG16, VGG19, ResNet50, ResNet101, ResNet152, GoogleNet, and Inception-BN (v2) from scratch as well as using fine-tuning for DDSM-400 and CBIS-DDSM datasets. Arora et al. [13] proposed a deeply ensembled transfer learning model for discrimination of benign and malignant tumors along with which a neural network classifier is used for auto feature extraction.

## 3 Database Description

In this research, two different datasets are used. It includes a publicly available dataset CBIS-DDSM [14], and another public dataset from INbreast, acquired with a license agreement. Figure 1 represents sample images from CBIS-DDSM and INbreast datasets.

**CBIS-DDSM:** It is a subpart of the DDSM [15] database which includes a total of 6775 studies. The mammographic images from DDSM are selected by the trained and experienced radiologists and represented in updated and standardized version in CBIS-DDSM. All the images are converted to DICOM format after lossless decompression. Along with pathologic diagnosis information, the segmented region of interest (ROI) for training data is also included in the database. Though the dataset is broadly classified based on the types of abnormalities, such as mass and calcification, it can also be classified based on the subclass such as benign and malignant.

**INbreast:** A total of 410 numbers of digital mammographic images are there in the INbreast dataset. Experienced radiologists were assigned to interpret the mammogram images and after the analysis, the lesions captured in the images are assigned a standard score known as BI-RADS [16]. The six BI-RADS scores indicates different stages of abnormalities found in breasts where score 0 recognizes non-conclusive examination; score 1 implies no findings; score 2 indicates benign; with score 3 it implies probably benign findings; score 4 indicates suspicious findings; score 5 ensures a high probability of malignancy; and score 6 predicts breast cancer. INbreast dataset is not available publicly but it can be obtained by a request from [17].



**Fig. 1.** Example of breast images (a), (b), (c), (d) from CBIS-DDSM for Full, ROI images of Mass, Calcification and (e), (f) from INbreast

### 4 Proposed Methodology

The main contribution of deep learning architectures is that it can extract low level to high level features by their own. CNNs are the best models to extract deep features from images. CNN can learn feature representation automatically as compared to handcrafted features.

In this work, various variants of pre-trained CNNs such as VGG19, ResNet50, MobileNet-v2, Inception-v3, Xception and Inception-ResNet-v2 are examined for experimental purposes. In pre-processing step, all mammographic images have been

converted from the DICOM images to portable network graphics (PNG) format. The reason of converting the DICOM to PNG format is to avoid the degradation of image quality. Thereafter, normalization of the pixel values in the range of (0–1) is performed so that the higher pixel values do not influence the result of the investigation. In this paper, two different approaches have been taken into consideration for this work. In the first approach, these pre-trained models are trained from scratch for both the datasets. For the second approach, we have utilized the concept known as transfer learning. Transfer learning can be used in two different ways with respect to the deep learning models. First method is transfer learning for feature extraction, which means chopping off all layers from fully connected layer to output layer. The extracted features are known as deep features. Another method is transfer learning using fine tuning, it means that we can freeze some layers, change the dimension of fully connected layer and make a new architecture utilizing pre-trained model. In this paper, we have utilized transfer learning using fine tuning only. The aim of transfer learning is to apply the knowledge, the network learned from one problem to another similar kind of problem. Transfer learning utilizes the pre-trained networks, where the network learned weights to obtain features and those weights, bias values would be later used for the new problem. In the first approach, network initialization is done with some pre-trained weights, and in the second approach though the network gets initialized with random weights but fine tuning of the parameter is induced to get better performance.

#### 4.1 Fine-Tuning Using Transfer Learning with CNNs

In Table 1, the different CNN architectures employed in the classification task along with the finely tuned parameters used for transfer learning are presented.

**Table 1.** The CNNs and their fine-tuned parameters for transfer learning

Network	Number of layers cut off	Number of layers added	Number of nodes added
VGG19 [18]	22	4	2048
ResNet50 [19]	175	2	1024
MobileNet-v2 [20]	155	2	1024
Inception-v3 [21]	249	2	512, Dropout(0.4)
Xception [22]	131	2	1024, Dropout(0.4)
Inception-ResNet-v2 [23]	779	0	0

## 5 Results and Discussion

To assess the performance of the proposed model in different scenarios, two different approaches have been taken into account for CBIS-DDSM dataset. Mass and calcification are two different abnormality types present in this dataset. In first approach, all

the calcification full and ROI mammographic images are trained and tested separately. In second approach, the training and testing performed on both mass full and ROI images separately to identify benign tumor and malignant tumor. For the evaluation of this work, various performance measures considered are given in the following equations:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall/Sensitivity} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

### 5.1 Analysis for CBIS-DDSM Dataset

Table 2, 3, 4 and 5 contains the result set for calcification full, ROI and mass full, ROI mammogram images.

**Table 2.** Result set for calc-full mammogram images (in %)

Network	From scratch				From fine-tuning			
	Accuracy	Precision	Recall	Specificity	Accuracy	Precision	Recall	Specificity
VGG19	62	25	11	85	71	39	42	80
ResNet50	66	41	27	82	74	33	12	93
MobileNetv2	67	44	22	87	68	33	35	78
Inceptionv3	77	100	19	100	74	0	0	96
Xception	71	0	0	100	66	11	6	85
InceptionResNetv2	76	75	19	98	66	23	18	82

**Table 3.** Result set for calc-ROI images (in %)

Network	From scratch				From fine-tuning			
	Accuracy	Precision	Recall	Specificity	Accuracy	Precision	Recall	Specificity
VGG19	62	37	42	70	66	38	24	84
ResNet50	69	44	19	90	72	55	24	92
MobileNetv2	20	36	16	12	69	33	5	96
Inceptionv3	77	71	52	89	65	44	43	76
Xception	42	24	39	46	77	71	36	94
InceptionResNetv2	76	75	39	94	70	53	32	87

**Table 4.** Result set for mass-full mammogram images (in %)

Network	From scratch				From fine-tuning			
	Accuracy	Precision	Recall	Specificity	Accuracy	Precision	Recall	Specificity
VGG19	56	47	28	76	59	36	50	63
ResNet50	61	54	42	73	51	31	54	50
MobileNetv2	59	66	7	97	77	59	50	85
Inceptionv3	59	42	59	59	49	38	58	44
Xception	66	0	0	100	71	64	55	82
InceptionResNetv2	70	100	9	100	56	44	67	49

**Table 5.** Result set mass-ROI images (in %)

Network	From scratch				From fine-tuning			
	Accuracy	Precision	Recall	Specificity	Accuracy	Precision	Recall	Specificity
VGG19	64	53	25	86	69	79	45	90
ResNet50	67	58	37	84	71	64	84	60
MobileNetv2	62	66	13	33	64	64	48	77
Inceptionv3	64	51	67	63	69	62	24	92
Xception	62	45	19	87	65	40	6	95
InceptionResNetv2	66	67	15	96	59	38	33	73

From the above results, it is found that the model achieved a highest accuracy up to 74% with both ResNet50 and Inception-v3 model for the calc-full mammogram images. For calc-ROI images, the model achieved the highest accuracy of 75% with Xception model. On the other hand, for mass-full mammogram images, the model achieved a highest accuracy of 75% with MobileNet-v2 model and for mass-ROI images; the accuracy obtained is 71% with ResNet50 model. From Table 2, 3, 4 and 5, it can also be concluded that fine-tuned models outperforms the performance of the same models when they are being trained from scratch.

## 5.2 Analysis for INbreast Dataset

The experimental results for the full mammogram images taken from the INbreast dataset are shown in Table 6 and Table 7. The experimental work on the state-of-the-art models are performed using train from scratch approach and fine-tuning approach. In the first approach, where the model is trained using INbreast full mammogram images from the scratch, the model achieved a highest accuracy of 82% and for the fine-tuning approach, the model achieved an accuracy of 84%. In this case too it can also be referenced that fine-tuned models overshadows the models which are trained from scratch.

**Table 6.** Result set for training from scratch (in %)

Network	Accuracy	Precision	Recall	Specificity
VGG19	80	0	0	100
ResNet50	70	27	36	78
MobileNetv2	80	0	0	100
Inceptionv3	78	0	0	100
Xception	82	0	0	100
InceptionResNetv2	82	0	0	100

**Table 7.** Result set after fine tuning models (in %)

Network	Accuracy	Precision	Recall	Specificity
VGG19	81	33	9	96
ResNet50	84	50	18	96
MobileNetv2	82	40	18	94
Inceptionv3	82	33	18	92
Xception	82	50	27	94
InceptionResNetv2	82	50	36	92

The state-of-the-art method [12], has also exploited the deep features for breast cancer diagnosis using from scratch and fine-tuning approaches for CBIS-DDSM dataset and achieved an accuracy of 66% for ResNet-101 model trained from scratch whereas for the same dataset, ResNet-152 achieved an accuracy of 76% when fine-tuning is used. In this paper, the extension of the same work has been performed on the pre trained networks and the performance comparison has been evaluated to verify which pre-trained network shows the improvement in terms of accuracy. It has been found out that for CBIS-DDSM dataset, InceptionResNet-v2 achieved an accuracy of 76% when the model is trained from scratch and using fine-tuning approach, MobileNet-v2 and Xception achieved 77% accuracy. From the analytical comparison of the obtained results with the state-of-the-art method, it can be also concluded that fine-tuning approach performs well than the trained from scratch approach.

## 6 Conclusion

For computer-aided diagnosis for breast cancer classification from mammographic images, this paper introduces deep convolutional neural networks using transfer-learning method. The performance of various networks is analyzed on two mammogram databases CBIS-DDSM and INbreast. In this work, two different scenarios of training are considered. In the first scenario, training is carried out from-scratch in which the network weights are initialized using a random distribution approach, and the other one is the fine-tuning approach, where the training is initiated with the already learned network weights. In the second scenario, fine-tuning is performed using



transfer learning. Above mentioned two approaches are performed on VGG19, ResNet50, MobileNet-v2, Inception-v3, Xception and Inception-ResNet-v2 models respectively. Experimental results show that fine-tuned models outperforms the models which are trained from scratch with an accuracy of 75% and 83% for CBIS-DDSM and INbreast respectively.

## References

1. Boyle, P., Levin, B.: World cancer report 2008. IARC Press, International Agency for Research on Cancer (2008)
2. Hamed, G., Marey, M.-R., Amin, S.-S., Tolba, M.F.: Deep learning in breast cancer detection and classification. In: Hassanien, A.-E., Azar, A.T., Gaber, T., Oliva, D., Tolba, F. M. (eds.) AICV 2020. AISC, vol. 1153, pp. 322–333. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-44289-7\\_30](https://doi.org/10.1007/978-3-030-44289-7_30)
3. Bick, U.: Mammography: how to interpret microcalcifications. In: Hodler, J., von Schulthess, G.K., Kubik-Huch, R.A., Zollikofer, C.L. (eds.) Diseases of the Abdomen and Pelvis 2014–2017, pp. 313–318. Springer, Milano (2014). [https://doi.org/10.1007/978-88-470-5659-6\\_40](https://doi.org/10.1007/978-88-470-5659-6_40)
4. Hubbard, R.A., Kerlikowske, K., Flowers, C.I., Yankaskas, B.C., Zhu, W., Miglioretti, D.L.: Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. *Ann. Intern. Med.* **155**(8), 481–492 (2011)
5. Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., Zwiggelaar, R.: Deep learning in mammography and breast histology, an overview and future trends. *Med. Image Anal.* **47**, 45–67 (2018)
6. Arevalo, J., González, F.A., Ramos-Pollán, R., Oliveira, J.L., Lopez, M.A.G.: Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput. Methods Programs Biomed.* **127**, 248–257 (2016)
7. Huynh, B.Q., Li, H., Giger, M.L.: Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J. Med. Imag.* **3**(3), 034501 (2016)
8. Carneiro, G., Nascimento, J., Bradley, A.P.: Unregistered multiview mammogram analysis with pre-trained deep learning models. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A. F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 652–660. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_78](https://doi.org/10.1007/978-3-319-24574-4_78)
9. Lévy, D., Jain, A.: Breast mass classification from mammograms using deep convolutional neural networks. arXiv preprint [arXiv:1612.00542](https://arxiv.org/abs/1612.00542) (2016)
10. Ting, F.F., Tan, Y.J., Sim, K.S.: Convolutional neural network improvement for breast cancer classification. *Expert Syst. Appl.* **120**, 103–115 (2019)
11. Rampun, A., Scotney, B. W., Morrow, P. J., Wang, H.: Breast mass classification in mammograms using ensemble convolutional neural networks. In: IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), pp. 1–6. IEEE (2018)
12. Tsochatzidis, L., Costaridou, L., Pratikakis, I.: Deep learning for breast cancer diagnosis from mammograms—a comparative study. *J. Imag.* **5**(3), 37 (2019)
13. Arora, R., Rai, P.K., Raman, B.: Deep feature-based automatic classification of mammograms. *Med. Biol. Eng. Compu.* **58**(6), 1199–1211 (2020). <https://doi.org/10.1007/s11517-020-02150-8>
14. Curated Breast Imaging Subset of DDSM. <https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM>

15. Lee, R.S., Gimenez, F., Hoogi, A., Rubin, D.: Curated breast imaging subset of DDSM. *Cancer Imaging Arch.* **8** (2016)
16. Orel, S.G., Kay, N., Reynolds, C., Sullivan, D.C.: BI-RADS categorization as a predictor of malignancy. *Radiol.* **211**(3), 845–850 (1999)
17. The INbreast Dataset. <https://medicalresearch.inescporto.pt/breastresearch/index.php/GetINbreastDatabase>
18. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014).
19. Akiba, T., Suzuki, S., Fukuda, K.: Extremely large minibatch sgd: Training resnet-50 on imagenet in 15 minutes. arXiv preprint [arXiv:1711.04325](https://arxiv.org/abs/1711.04325) (2017)
20. Howard, A.G., et al.: Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861) (2017)
21. Xia, X., Xu, C., Nan, B.: Inception-v3 for flower classification. In: 2nd International Conference on Image, Vision and Computing (ICIVC), pp. 783–787. IEEE (2017)
22. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)
23. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI Conference on Artificial Intelligence (2017)