

Breast Cancer Classification Using Transfer Learning



Animesh Seemendra , Rahul Singh , and Sukhendra Singh 

Abstract Cancer is one of the most lethal forms of the disease. And in females, breast cancer is the most common cancer which could even lead to death if not properly diagnosed. Over the years, a lot of advancement can be seen in the field of medical technology but when it comes to detecting breast cancer biopsy is the only way. Pathologists detect cancer by using histological images under the microscope. Inspecting cancer visually is a critical task; it requires a lot of attention, skill and is time-consuming. Therefore, there is a need for a faster and efficient system for detecting breast cancer. Advancements in the field of machine learning and image processing lead to multiple types of research for creating an efficient partially or fully computer monitored diagnosis system. In this paper, we have used histological images to detect and classify invasive ductal carcinoma. Our approach involves convolutional neural networks which are a very advanced and efficient technique when dealing with images in machine learning. We compared various famous deep learning models, and we used these pre-trained CNN architectures with fine-tuning to provide an efficient solution. We also used image augmentation to further improve the efficiency of the solution. In this study, we used VGG, ResNet, DenseNet, MobileNet, EfficientNet. The best result we got was using fine-tuned VGG19 and with proper image augmentation. We achieved a sensitivity of 93.05% and a precision of 94.46 with the mentioned architecture. We improved the F-Score of the latest researches by 10.2%. We have achieved an accuracy of 86.97% using a pre-trained DenseNet model which is greater than the latest researches that achieved 85.41% [30] accuracy.

Keywords Breast cancer · Deep learning · Invasive ductal carcinoma (IDC)

A. Seemendra (✉) · R. Singh · S. Singh
JSS Academy of Technical Education, Noida, India
e-mail: animeshseemendra1997@gmail.com

R. Singh
e-mail: rahulsingh2240@gmail.com

S. Singh
e-mail: sukhendrasingh@gmail.com

1 Introduction

Breast cancer in women is the most common form of cancer [1]. The American Cancer Society [2] reports that breast cancer leads to deaths of 40,610 women and 460 men in 2017 in the USA. Another report by the American Cancer Society reveals in every 8 women 1 which is diagnosed with invasive breast cancer and 1 in 39 women die from breast cancer [3]. In the developed countries like USA and UK, breast cancer survival rate is relatively high but in poor and developing countries like India, survival rates are much less major reasons being lack of awareness, delay in diagnosis, and costly screenings. Though there has been an increase in survival rates due to the advancement of technology, a number of cases have also increased over the years. Invasive ductal carcinoma (IDC) constitutes about 80% of all breast cancer cases [4]. Doctors use various techniques to detect IDC such as physical examination, mammography, ultrasound, breast MRI, and biopsy. The biopsy is often done to check the suspicious mammogram. It involves examining the abnormal-looking tissue under a microscope by a pathologist. Such an examination requires a lot of time, skill, and precision. Therefore over the years, many computer-aided systems are being developed to make the work of a pathologist easier and efficient.

Since the advancements in machine learning and image processing a lot of research is oriented towards creating a better and efficient model to detect breast cancer. Machine learning is a field of computer science that uses data or past experiences to generate or predict outcomes for unseen data [5–8, 37]. Multiple methods have been applied to create an efficient machine learning model for detecting breast cancer. Classification using standard machine learning techniques often requires feature selections. Some of them used image segmentation such as thresholding along with the SVM classifier to create a classification model [9–12]. Preprocessing is an important task in machine learning which depends upon the type of data. Some studies used mammography images; therefore, a common preprocessing step was cropping and resizing [9–12]. Different studies used different image preprocessing techniques such as adaptive histogram equalization followed by segmentation, image masking, thresholding, feature extraction, and normalization before training the classifier [9], morphological operations for image preprocessing followed by an SVM classifier [13] and high-pass filtering followed by a clustering algorithm [14]. Some other preprocessing and classification steps involved were median filtering and fuzzy-C clustering with thresholding [15], region-based image segmentation followed by SVM classifier [16] and ostu-based global thresholding method followed by radial basis neural network for classification [17].

Later advancements in deep learning and availability of high computational GPU lead to multiple studies in this sub-domain of machine learning. Deep neural networks do not require any feature selection or extensive image preprocessing techniques. These networks mimic human brain neurons and automatically does feature selections [18]. The use of neural network architectures in breast cancer classification leads to the state-of-the-art results. Convolutional neural networks proved to be one

of the most efficient models when working with images and videos. CNN maintains the structural architecture of the data while learning; this makes it achieve a higher accuracy over traditional machine learning methods [19]. Therefore, multiple studies for breast cancer detection revolve around convolutional neural networks [20–24]. Some of the studies combined deep neural architectures for feature selection and standard machine learning classifiers like SVM for the classification task [22]. Other approaches involve using transfer learning as a key part. Transfer learning is a technique to using knowledge of another network trained on a different dataset and use it in with your data. Various studies have used ResNet50, VGG16, VGG19, etc., models for breast cancer detection and have shown impressive results [25–28]. Studies mentioned in [29] have also used CNN and deep learning architectures for transfer learning to classify invasive ductal carcinoma. With the proposed methods, they have achieved an accuracy of 85.41%. We worked to improve these results.

Many of these approaches have used a multi-classification dataset that contains more than one class of breast cancer. Also, most of the approaches were applied to a mammographic dataset. The mammographic dataset can detect any abnormalities in the tissues which have to further go with biopsy. Our approach is on a binary classification dataset which addresses the problem of Invasive ductal carcinoma. The effective classification on the dataset will help in easing the task of the pathologist. In this paper, we have approached the problem by using various deep learning architectures through transfer learning. We applied image undersampling and image augmentation to handle imbalance dataset and to increase the accuracy of the model by providing necessary image transformation parameters. The objective of this paper is (a) to compare the performances of various famous convolutional deep learning architectures through transfer learning and achieve high efficiency in less computational time on the given dataset (b) to increase the efficiency in the IDC classification task over previous works.

2 Approach

In our approach, we used transfer learning and image augmentations on the given dataset. The following sections include steps applied to the dataset to detect invasive ductal carcinoma.

2.1 *Undersampling and Data Preparation*

In medical imaging, dataset class imbalance is a common problem; there are generally more images of negative results than positive results of disease. Similarly, in our dataset, there is a huge imbalance of the data between IDC(–) and IDC(+) classes. Such imbalance data could mislead the model in learning one type of class more than the other. Therefore, we used resampling techniques to create balanced data.

2.2 *Image Augmentation*

Image augmentation is a technique that is used to generate more data by applying certain processing to the images. It helps to create unseen yet valuable data which could help in further increase the accuracy. The common image augmentation techniques are zoom, flips, rotation, etc. These techniques help the model to learn variation in the dataset and not to get constraint in one particular type of format. It is also a powerful technique to solve the lack of data problem. We applied image augmentation techniques to add variation to the dataset which could further increase the efficiency of the model.

2.3 *Transfer Learning*

Transfer learning is a technique of using models that were trained on different datasets on your data with some tuning. We used various state-of-the-art CNN architectures that were trained on the ImageNet dataset. The idea was that the starting layers of CNN are used to capture the high-level features such as texture and shape which are not dependent on the data; therefore, utilizing those trained weights and fine-tuning it according to our model could help us achieve good results. The various deep learning architectures we used are present in Sect. 3.

3 **Deep Learning Architectures**

The various deep learning architectures used are mentioned below. While training, each of these architectures was fine-tuned so that they give the most efficient solution. While training validation set was used to prevent overfitting of the data. The weights were stored when the minimum loss was found, and the same weights were used to test the model efficiency on the test set. All the following architectures were trained on the ImageNet dataset [30] and achieved remarkable results.

3.1 *VGG16 and VGG19*

The VGG16 and VGG19 models were sequential CNN models with 3×3 convolutional layers stacked upon one another. The architecture contained max-pooling layers to reduce the volume as the layer increases finally the fully connected network layer with 4096 nodes followed by a 1000 node layer with a softmax activation function [31]. VGG16 and VGG19 are slow to train and have large weights themselves.

3.2 *ResNet50*

ResNet50, unlike VGG models, is a non-sequential model. This is a collection of CNN stacked together with residual networks added to each layer. The output of each layer of CNN layer is added with the actual input of that layer [32]. ResNet50 contains 50 weight layers and is faster to train with VGG networks.

3.3 *DenseNet*

DenseNet involves some advancements over ResNet50 networks, instead of adding feature maps DenseNet network concatenates output feature maps with input feature maps. Each layer output is concatenated with the outputs of all the previous layers, thus creating a dense architecture [38].

3.4 *EfficientNet*

This model performed so well on the ImageNet dataset that it was able to achieve 84.4% top 1 ranking. It crossed the state-of-the-art accuracy with 10 times better efficiency. The model was smaller and faster. Width, depth, and image resolution were scaled, and the best result was observed [33]. EfficientNet has shortcuts that directly connect between the bottlenecks and a fewer number of channels than expansion layers.

3.5 *MobileNet*

MobileNet is a lightweight model. Instead of performing combined convolutions on the three channels of colors and flattening, it applies convolutions on each color channel [34]. They are ideal for mobile devices.

4 Experiment and Results

4.1 *Datasets*

The dataset from which our data was derived was a 162 whole mount slide images of the breast cancer specimens scanned at $40\times$ [35, 36]. The derived dataset contained 25,633 positive samples and 64,634 negative samples. Each image is of size $(50 \times$

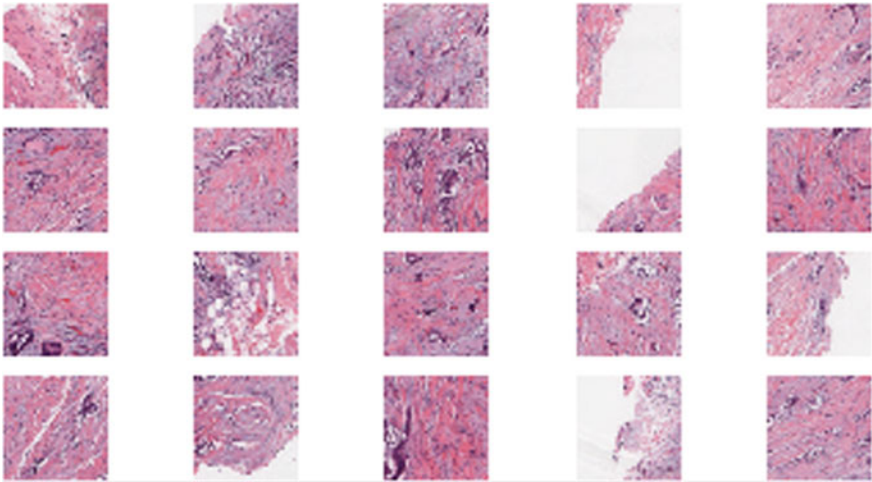


Fig. 1 Sample images from the dataset

50) extracted from the original dataset (Fig. 1). The image filename is in the form of $z_xX_yY_classD.png$. For example, $55634_idx8_x1551_y1000_class0.png$ where z is the patient ID (55634_idx8), X represents the x-coordinate and Y represents y-coordinate from which the patch is cropped and D indicates the class of the data. Class0 being IDC(-) and Class 1 being IDC(+).

4.2 Data Preparation

We performed undersampling on our dataset, i.e., removing samples from classes to make it more balanced (Fig. 2). The undersampling was done at random without replacement to create a subset of data for the target classes.

The final distribution contains 25,367 positive samples and 25,366 negative samples of data. Final data split into the train, test, and valid sets containing 31,393, 7849, 11,490 files, respectively. The class names were extracted from the names of the files and were one-hot encoded, i.e., binary class was represented by two features instead of one; for example, class0 was represented by an array of [1, 0] and class1 was represented by [0, 1] so that for deep learning model will give probabilities for the two indexes of the array and whichever is maximum that index will be the class of our array after which image augmentation techniques were applied such as zoom, rotation, width shift, height shift, height shift, shear, flip.

Now many deep learning model architectures were used to provide transfer learning. They were fine-tuned to give change their domain to our dataset.

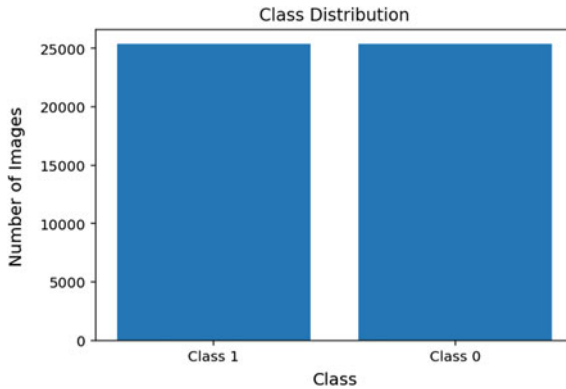


Fig. 2 Final distribution of class after undersampling

4.3 Results

The experiments were performed on Kaggle Notebook with its GPU with Keras API of Tensorflow. NVidia K80 GPU was used to perform experiments on the dataset. We used various famous deep learning architectures for transfer learning (VGG16, VGG19, ResNet50, DenseNet169, DenseNet121, DenseNet201, MobileNet, and EfficientNet). All these models were trained on the ImageNet dataset, and over the years, they have achieved high accuracy of the dataset. We have used only CNN architectures of these models and then added layers according to the requirement of the complexity of the dataset. The first set of experiments were conducted without any data augmentation, and then data argumentation was added to see the change. The evaluation metrics used to compare results were F1-score, recall, precision, accuracy, specificity, sensitivity. They are used to evaluate our model using true positives, true negatives, false positives, and false negatives.

DenseNet169 was trained for 60 epochs with a global pooling layer and two dense layers each followed by a dropout layer of 0.15 and 0.25, respectively, with ReLU activation added after the frozen CNN layers. The last layer was a dense layer with two classes and a softmax function. DenseNet 121, ResNet50, VGG19, and VGG16 were trained for 60, 60, 100, 100 epochs, respectively. The architecture of the last layers was made different than DenseNet169. We added batch normalization having $1e-05$ epsilon value and 0.1 momentum followed by a dense layer with 512 nodes and a dropout of 0.45. After that, a dense layer with 2 nodes representing each class was added with softmax activation. The experiment results without data augmentation can be seen in Table 1, the highest value in each column is highlighted (in bold).

Data augmentation was further added to push the accuracy even further with more different types of images in the dataset. Augmentation was all general, i.e., zoom range of 0.3, rotation range of 20, width shift range, sheer range, and height shift range was all set to 20 and also horizontal flip was set as true. With the mentioned augmentation, all previous models were used with SGD optimizer, global pooling

Table 1 Results without data augmentation

Models	Sensitivity	Specificity	Accuracy	Precision	Recall	F-Score
DenseNet169	83.37	88.14	84.43	89.29	81.37	85.15
DenseNet121	83.42	86.61	84.94	87.22	83.42	85.28
ResNet50	81.21	89.38	84.82	90.61	81.21	85.65
MobileNet	85.83	85.26	85.54	85.13	85.83	85.48
VGG16	85.16	87.49	86.29	87.83	85.16	86.54
VGG19	85.40	87.71	86.52	88.10	85.40	86.73
EfficientNet b0	87.60	79.92	83.32	77.63	87.60	82.31
EfficientNet b4	85.36	84.55	84.95	84.38	85.25	84.87
EfficientNet b5	86.78	81.40	83.88	79.94	86.78	83.22

Bold signifies highest value in respective column

layer followed by 32-node and 64-node layer with 0.15 and 0.25 dropout layer, respectively, and ReLU was used as an activation function. Results can be seen in Figs. 3 and 4.

Result of the experiments can be seen in Table 2.

By using different deep architectures for transfer learning, we were able to achieve higher accuracy and recall with our models over previous claimed methods. Table 3 compares our best model’s efficiency with respect to the efficiency claimed by previous works.

We presented F-Score as evaluation metrics for comparison with previous methods. We can see that our approach and method improved the F-Score by 10.2% compared to the latest researches.

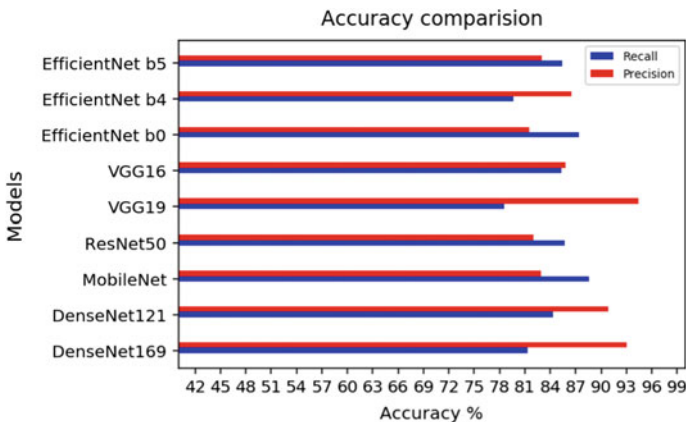


Fig. 3 Precision and recall versus fine-tuned models

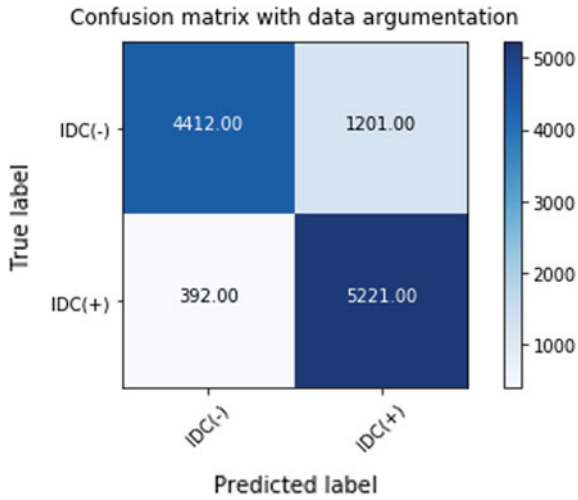


Fig. 4 Confusion matrix of DenseNet 169 with data augmentation

Table 2 Results with data augmentation

Models	Sensitivity	Specificity	Accuracy	Precision	Recall	F-Score
DenseNet169	81.29	91.84	85.80	93.01	81.29	86.76
DenseNet121	84.30	90.10	86.97	90.87	84.30	87.46
MobileNet	88.57	83.92	86.10	82.89	88.57	85.64
ResNet50	85.71	82.79	84.19	82.00	85.71	83.84
VGG19	78.51	93.05	84.30	94.46	78.51	80.82
VGG16	85.32	85.68	85.50	85.75	85.32	85.53
EfficientNet b0	87.36	82.71	84.88	81.56	87.36	84.36
EfficientNet b4	79.64	85.17	82.16	86.44	79.62	82.89
EfficientNet b5	85.35	83.44	84.37	82.98	85.35	84.14

Table 3 Our score versus existing deep learning approaches

Studies	Method	F1-Score
Paper [29] (2019)	Convolution Neural Networks trained from scratch with AdaDelta optimizer and image augmentation	0.8528
Our Result with DenseNet121	Transfer Learning with deep learning architectures (DenseNet121) with SGD optimizer and image augmentation	0.8746

5 Conclusion

We classified invasive ductal carcinoma (IDC) using deep learning. In our study, we took advantage of various pre-trained models and used their knowledge and added some fine-tuning to get an efficient model. We first tried undersampling techniques to balance classes then image argumentation followed by transfer learning. We were able to get a high precision and moderate recall model with VGG19 with image argumentation. The value of precision was 94.46, and recall was 78.51. We got an accuracy of 86.97 with DenseNet121 which when compared with other latest research that got 85.41% [29] accuracy.

Hence, we can conclude that transfer learning which is the simplest technique available in deep learning frameworks can be used to detect dangerous diseases such as cancer. We have also shown that deep learning can perform well while detecting IDC, and therefore, it is an improvement over manual segmentation. Deep learning has given us the freedom to detect the disease on smaller datasets with small size images from which only deep learning models can infer useful information. An increase in the dataset will improve the results. Hence, in the future, we are planning to work on the full dataset with full-size slide images and will work with more advanced techniques such as GAN networks to obtain better results.

References

1. American Institute of Cancer Research. <https://www.wcrf.org/sites/default/files/Breast-Cancer-2010-Report.pdf>
2. American Cancer Society. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf>
3. American Cancer Society. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf>
4. Breast Cancer Website. <https://www.breastcancer.org/symptoms/types/idc>
5. Dhall D, Kaur R, Juneja M (2020) Machine learning: a review of the algorithms and its applications. In: Singh P, Kar A, Singh Y, Kolekar M, Tanwar S (eds) Proceedings of ICRIC 2019. Lecture Notes in Electrical Engineering, vol 597. Springer, Cham. https://doi.org/10.1007/978-3-030-29407-6_5
6. Pillai R, Oza P, Sharma P (2020) Review of machine learning techniques in health care. In: Singh P, Kar A, Singh Y, Kolekar M, Tanwar S (eds) Proceedings of ICRIC 2019. Lecture Notes in Electrical Engineering, vol 597. Springer, Cham. https://doi.org/10.1007/978-3-030-29407-6_9
7. Jondhale SR, Shubair R, Labade RP, Lloret J, Gunjal PR (2020) Application of supervised learning approach for target localization in wireless sensor network. In: Singh P, Bhargava B, Paprzycki M, Kaushal N, Hong WC (eds) Handbook of wireless sensor networks, issues and challenges in current scenario's, advances in intelligent systems and computing, vol 1132. Springer, Cham. https://doi.org/10.1007/978-3-030-40305-8_24
8. Singh YV, Kumar B, Chand S, Sharma D (2019) A hybrid approach for requirements prioritization using logarithmic fuzzy trapezoidal approach (LFTA) and artificial neural network (ANN). In: Singh P, Paprzycki M, Bhargava B, Chhabra J, Kaushal N, Kumar Y (eds) Futuristic trends

- in network and communication technologies. FTNCT 2018, communications in computer and information science, vol 958. Springer, Singapore. https://doi.org/10.1007/978-981-13-3804-5_26
9. Breast Cancer Classification using Image Processing and Support Vector Machine. <https://pdfs.semanticscholar.org/d414/5b40d6a65b84e320a092220dc8e6cc54a7dc.pdf>
 10. Sudharshan PJ et al (2019) Multiple instance learning for histopathological breast cancer image classification. *Expert Syst Appl* 117:103–111. <https://doi.org/10.1016/j.eswa.2018.09.049>
 11. Rejani YA, Selvi ST (2009) Early detection of breast cancer using SVM classifier technique. *Int J Comput Sci Eng*
 12. Pathak R (2020) Support vector machines: introduction and the dual formulation. In: *Advances in cybernetics, cognition, and machine learning for communication technologies. Lecture Notes in Electrical Engineering*, vol 643. Springer, Singapore. https://doi.org/10.1007/978-981-15-3125-5_57
 13. Naresh S, Kumari SV (2015) Breast cancer detection using local binary patterns. *Int J Comput Appl* 123(16):6–9
 14. Guzman-Cabrera R, Guzman-Supulveda JR, Torres-Cisneros M, May-Arrioja DA, Ruiz-Pinales J, Ibarra-Manzano OG, AvinaCervantes G, Parada GA (2013) Digital image processing technique for breast cancer detection. *Int J Thermophys* 34:1519–1531
 15. Kashyap KL, Bajpai MK, Khanna P (2015) Breast cancer detection in digital mammograms. In: *IEEE international conference in imaging systems and techniques*, pp 1–6
 16. Oliver A, Marti J, Marti R, Bosch A, Freixenet J (2006) A new approach to the classification of mammographic masses and normal breast tissue. In: *International conference on pattern recognition*, pp 1–4
 17. Kanojia MG, Abraham S (2016) Breast cancer detection using RBF neural network. In: *IEEE conference on contemporary computing and informatics*, pp 363–368
 18. Goodfellow I, Bengio Y, Courville A, *Deep learning book*
 19. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang L, Wang G, Cai J, Chen T (2015) Recent advances in convolutional neural networks. arxiv: 1502.07108
 20. Selvathi D, Poornila AA (2017) Deep learning techniques for breast cancer detection using medical image analysis. In: *Biologically rationalized computing techniques for image processing applications*, pp 159–186
 21. Ragab DA, Sharkas M, Marshall S, Ren J (2019) Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ* 7:e6201. <https://doi.org/10.7717/peerj.6201>
 22. Shen L, Margolies RL, Rothstein JH, Fluder E, McBride R, Sieh W (2017) Learning to improve breast cancer detection on screening mammography. arxiv: 1708.09427
 23. Zou L, Yu S, Meng T, Zhang Z, Liang X, Xie Y (2019) A technical review of convolutional neural network-based mammographic breast cancer diagnosis. In: *Computational and mathematical methods in medicine*. <https://doi.org/10.1155/2019/6509357>
 24. Ciresan CD, Giusti A, Gambardella ML, Schmidhuber J (2013) Mitosis detection in breast cancer histology images with deep neural networks. In: *International conference on medical image computing and computer-assisted intervention*
 25. Le H, Gupta R, Hou L, Abousamra S, Fassler D, Kurc T, Samaras D, Batiste R, Zhao T, Dyke AL, Sharma A, Bremer E, Almeida SJ, Saltz J (2019) Utilizing automated breast cancer detection to identify spatial distributions of tumor infiltrating lymphocytes in invasive breast cancer
 26. Wu N et al (2019) Deep neural networks improve radiologists performance in breast cancer screening. In: *Medical imaging with deep learning conference*
 27. Rakhlin A, Shvets A, Iglovikov V, Kalinin AA (2018) Deep convolutional neural networks for breast cancer histology image analysis. In: *International conference on image analysis and recognition*
 28. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride RB, Sieh W (2017) Deep learning to improve breast cancer early detection on screening mammography. arxiv: 1708.09427

29. Romano AM, Hernandez AA (2019) Enhanced deep learning approach for predicting invasive ductal carcinoma from histopathology images. In: International conference on artificial intelligence and big data
30. Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: IEEE conference on computer vision and pattern recognition
31. Simonyan K, Zisserman A (2015) Very deep convolutional network for large-scale image recognition, arxiv: 1409.1556
32. He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition, arxiv: 1512.03385
33. Tan M, Le QV (2019) EfficientNet: rethinking model scaling for convolutional neural networks, arxiv: 1905.11946
34. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) MobileNets: efficient convolutional neural networks for mobile vision applications, arxiv: 1704.04861
35. Janowczyk A, Madabhushi A (2016) Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases
36. Cruz-Roa A et al (2014) Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: Proceedings of SPIE—the international society for optical engineering, vol 9041
37. Kral P, Lenc L (2016) LBP features for breast cancer detection. In: IEEE international conference on image processing, pp 2643–2647
38. Huang G, Liu Z, Maaten LD, Weinberger QK (2016) Densely connected convolutional networks, arxiv: 1608.06993