Breast Cancer Detection and Localization using MobileNet based Transfer Learning for Mammograms

Wajeeha Ansar[1], Ahmad Raza Shahid [1], Basit Raza [1],Amir Hanif Dar [1] and Asad Ali Safi [2]

Wajeeha.ansar@yahoo.com, ahmadrshahid@comsats.edu.pk, basit.raza@comsats.edu.pk, daramir@comsats.edu.pk, asafi@hct.ac.ae

[1] Medical Imaging and Diagnostic (MID) Lab, National Centre of Artificial Intelligence (NCAI), Department of Computer Science,
COMSATS University Islamabad (CUI), Islamabad, 45550, Pakistan.
[2] Sharjah Women's Campus - Higher Colleges of Technology, Department of Computer Information Science, Sharja UAE.

Abstract. Breast cancer is the major cause of death among women. The best and most efficient approach for controlling cancer progression is early detection and diagnosis. As opposed to biopsy, mammography helps in early detection of cancer and hence saves lives. Mass classification in mammograms remains a major challenge and plays a vital role in helping radiologists in accurate diagnosis. In this work, we propose a MobileNet based architecture for early breast cancer detection and further classify mass into malignant and benign. It requires less memory space and provides faster computations with 86.8% and 74.5% accurate results for DDSM and CBIS-DDSM, respectively. We have achieved better results than other deep CNN models such as AlexNet, VGG16, GoogleNet, and ResNet.

Keywords: Breast Cancer, Mammography, Deep Learning, MobileNet, Localization.

1 Introduction

The second most prevalent cause of death from cancer among women is breast cancer. Based on the recent statistics from the American Cancer Society, it is estimated that 41,760 women in the United States are expected to die from breast cancer in 2019 [1]. Mammography is one of the most commonly used techniques of breast cancer screening which has made a significant contribution to reducing mortality rates through early cancer detection. Mammography involves exposing the breasts of a patient to low levels of X-ray radiation. The complexity of mammograms and the high volume of examinations per radiologist, however, can lead to a fake diagnosis[2] .

Computer-Aided Detection (CAD) systems are used to help radiologists for cancer detection and diagnosis. Studies have demonstrated the effectiveness of CAD systems;

however, accurate breast cancer detection remains a challenging task. Due to mammograms complexity and huge number of examinations by radiologist, results in false diagnosis. CAD system aims to decrease false diagnosis, which ultimately reduce unnecessary biopsies, patient anxiety, additional cost of health care and extra evaluation by radiologists [3].Current CAD constraints show the need for fresh and more accurate techniques of detection. Standard Machine Learning techniques (ML) used for CAD systems, include a hand-crafted feature extraction phase, which is challenging as it requires domain knowledge of expert radiologists which is costly and time consuming

Moving on from classical ML, more recent effort has been geared towards use of Deep Learning (DL) techniques They diagnose suspected lesions more accurately through quantitative analysis [4]. DL models are designed for large and diverse datasets (e.g imageNet dataset), whereas for mammograms we have small datasets publicly available. Consequently, the capability and complexity of such networks can lead to significant adverse effects on model performance when learning from scratch with small training samples.

To overcome afore mentioned issues, Transfer Learning (TL) techniques are widely used for mammograms. In TL we train deep models on large dataset and then test them on smaller dataset with fine tuning. In this work, we present our preliminary outcomes on the use of transfer learning to classify breast cancer. We use a MobileNet [5] based technique, which is efficient, require less memory and computational cost. This paper comprises of the following:

- We proposed a customized DL architecture that is better suited to mobile devices because of its compact size, reduced computational costs and competitive performance relative to existing models.
- We also perform methods such as pre-processing of images, data augmentation, choice of hyper parameters (e.g. batch size, learning rate), which are attractive alternatives for reducing overfitting and increasing the generalization of the TL technique for mammograms.
- Comparative assessment of DL models such as VGG [6], AlexNet [4], ResNet [7], GoogleNet[8] with MobileNet and MobileNet-v2[9] in terms of their efficiency and computational power.
- Our aim is to provide automatic CAD system to manage millions of routine imaging examinations, exposing prospective cancers to the radiologists who undertake follow-up operations. This can be used to help and serve as a second eye for radiologists.

The rest of the paper is organized as follows: literature review in Section 2 on the use of transfer learning in the classification of defects mammograms as benign or malignant in the current study. Section 3 introduces our proposed experimental technique and dataset. Section 4 provides the outcomes of experimental information. Finally, conclusions and future work are drawn in Section 5.

2 Previous Work

Current CNN models are intended to enhance the capacity of radiologists to locate even the smallest breast cancers at their earliest phases. Due to the accessibility of large information repositories, the power of parallel and distributed computing, use of DL techniques has resulted in breakthroughs in pattern recognition and classification tasks.

It is very difficult to train a deep CNN model with a small amount of medical information. Unfortunately, publicly available datasets for mammograms are not large enough for deep learning models. In recent years, several models of convolutional neural networks such as: AlexNet, VGG, GoogLeNet and Resnet, and others have been created to resolve such issues. Researchers have also begun to investigate the use of these models for transfer learning to classify mammograms [10]. For transfer learning, fine tuning is the most commonly used technique. In this case, with the new data, only a few of the model's last layers are retrained.

AlexNet was the first convolutional neural network (CNN) in the field of object detection and classification to exhibit performance beyond the state of the art. Jiang et al [11] and Huynh et al. [12] used AlexNet model for applying TL technique on mammograms . Both authors used pre-trained AlexNet for the issue of mass diagnosis without further fine-tuning. Using SVM for classification, they evaluated classification performance using characteristics from multiple intermediate network layers. They contrasted their outcomes with two methods: a classifier working on hand-crafted characteristics and a soft voting ensemble of both. Rampun et al.[13] used a mildly altered, pre-trained and fine-tuned version of AlexNet on CBIS-DDSM. They selected the three highest performing models during inference and merged their predictions.

The impact of network depth was explored in VGG while maintaining a smaller size for convolution filters. VGG-19 is an updated version of VGG-16, and they showed that by increasing the depth from 16 to19 layers, important improvement can be accomplished. The benefit of VGG is that the efficient receptive field of the network is enhanced by stacking various convolution layers with small kernels, while decreasing the number of parameters compared to using less convolutional layers for the same receptive field with bigger kernels. Hamidinekoo et al.[14] used VGG16 and GoogleNet. Similarly, Chougrad et al.[15], for instance, used VGG16, InceptionV3, and ResNet. They showed that the precision of the fine-tuned model decreases when the number of convolutional blocks exceeds 2.

Residual Networks (ResNet) consist of reformulated convolutional layers that learn residual features based on inputs. It is simpler to optimize this sort of networks by considerable reduction in depth such as the "residual block" implementation. Google has created the InceptionV3 model which is also known as GoogleNet. It's computational cost and memory requirements are much smaller than that of VGG and ResNet, making it a popular model being used for big data.

Morrell et al.[16] used InceptionV3 and deformable convolutional net for private mammographic dataset. This method is regarded to be computationally complicated, as TL is done with two different models. Cameiro et al.[17] worked on the whole mammogram image models in quest for an end-to-end design. Carneiro et al.[18] used a pre-trained CNN that was fine-tuned using unregistered mammograms and microcalcification and mass segmentation to estimate the risk of BIRADS-based breast cancer. They found that the pre-trained models are superior to those initialized at random.

Rather than training from scratch, more recent work has focused on use of pre-trained networks. However, pre-trained CNN architectures are designed, trained and tested on large datasets, which are diverse and different in nature than the available mammographic datasets. Consequently, the capability of such networks and their complexity can far exceed the requirements of larger datasets, resulting in significant negative effects when training from scratch and limit their use in mobile devices. Using such large networks on mobile devices is hard on memory and computational power which in turn decreases the performance. To resolve this issue, we propose a transfer learning-based technique using MobileNet based model with few parameters. As a result, they require less memory space and provide good and faster result as compared to other large VGG and inception models.

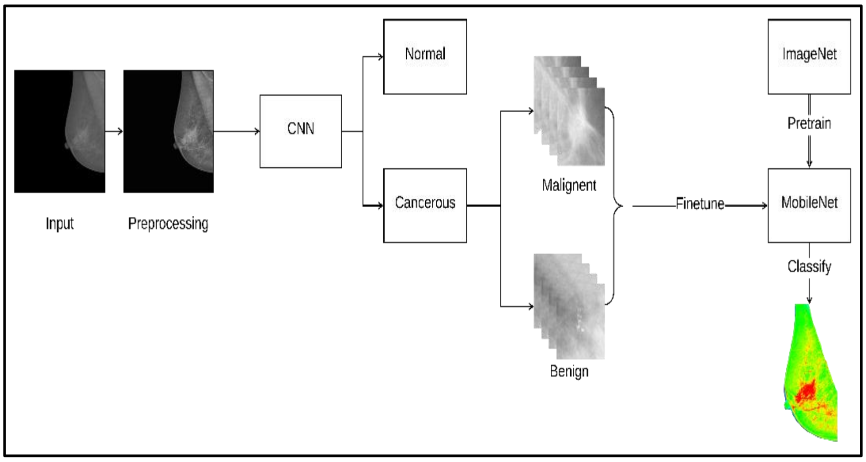

Fig. 1. The proposed framework for automated detection and localization of masses in full mammograms, where the first shallow CNN take pre-process full mammogram and performed binary classification between cancerous and non-cancerous. If cancer present in input then is further pass to MobileNet based model, which is pre-trained on ImageNet dataset. To reduce parameter, we used ROI for fine-tuning. At the end to localize cancer in full mammogram, we used class activation mapping technique.

3 Methodology

Mammograms cannot be considered as a simple image classification task, because there are defects in tiny areas within an entire image. For instance, a typical complete mammogram with 3000x4600 resolution (width and height in pixels) includes only about 200x200 (pixels) of abnormality. To use a pre-trained deep model for TL, it is necessary to resize and normalize the input to the same format that the network was initially trained on (e.g 224 x 224 for VGG models). Therefore, to address this challenge, we are proposing to train extensive CNN framework with fewer parameters on cropped image patches (labeled ROIs) and adapt them to complete mammogram images. Figure1 shows our approach's data flow. A binary shallow CNN classifier is trained as shown in Figure 2, followed by a MobileNet based architecture with training image patches from benign and malignant mass tissues.

A pre-trained MobileNet is altered to have two output classes in output layers. The output layers are then finely tuned while the network's initial layers are kept frozen. The performance of the model can be tracked by visualizing the learned features at different layers of MobileNet during the training process. In this manner, characteristics corresponding to distinct scales were acquired. In addition, it helps us to stop early training and reduce the size of learning parameters. The fine-tuned neural patch network is then used to locate mammographic defects in mammograms of full size by generating heat-map at the end of output layer.

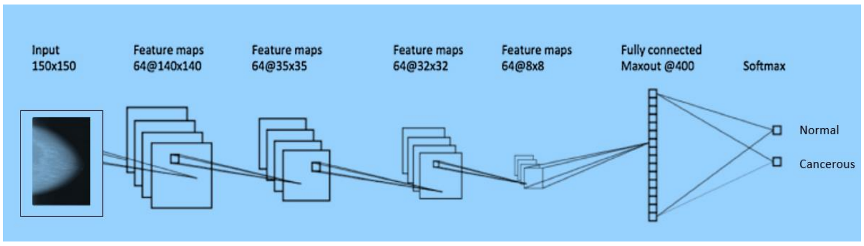

Fig. 2. CNN architecture for binary classification of mammogram into Normal or Cancerous.

3.1 Data set used for experiments

There is a shortage of datasets for standard assessment in mammography and are not publicly accessible, therefore most CAD algorithms are assessed on private datasets. It presents a challenge in comparing method efficiency or replicating previous results. The databases most frequently used are Digital Database for Screening Mammography (DDSM) [18] and Curated Breast Imaging Subset of DDSM (CBIS-DDSM) [19].

DDSM is the biggest publicly accessible mammography dataset. The database involves roughly 2,500 trials, each of which includes two views for each breast Cranial Caudal (CC) view captured from the top of breast and other is Medium Lateral Oblique (MLO) captured from side angle at 45 degree as shown in Figure 3. Text data about the patient and image is also available in dataset. Images comprising suspect areas connected "ground truth" pixel-level with the places and kinds of suspect regions. To test and train our automatic segmentation techniques, we only use images taken from the CC perspective. Both the characteristics of MLO and CC are used to test and train the model of mass lesion diagnosis.

An updated and standardized DDSM version for mammography is CBIS-DDSM. The dataset includes 753 cases of calcification and 891 cases of mass. We use CBIS-DDSM patch images to classify and test for localization in complete mammograms. We combine the two discussed datasets for training and testing and perform 80/20 split for training/testing sets respectively. In CBIS-DDSM dataset, experts cropped the ROIs of abnormality portion and also removed 339 questionable mass cases manually. Breast mass classification into malignant and benign is a challenging task due to the diverse characteristics of abnormal regions and overlapping with the dense breast tissues. Mass samples from CBIS_DDSM are shown in Figure 4.

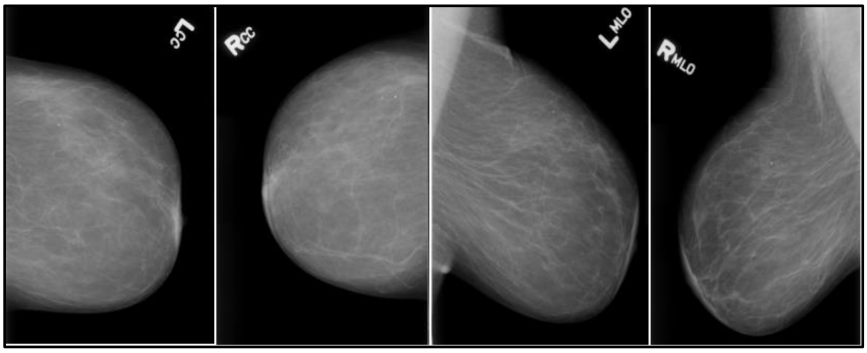

Fig. 3. CC and MLO views of a patient from DDSM dataset (From left to right: left CC view, right CC view, left MLO view, right MLO view).

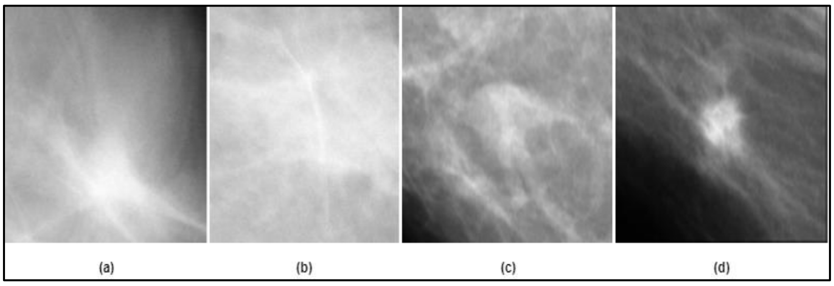

Fig. 4. Example of Mass benign cases (a) and (b); Mass malignant cases (c) and (d) from CBIS-DDSM dataset.

3.2 Image pre-processing and data Augmentation

Before training CNNs, mammogram's pre-processing is an important task. Mammograms images contain artifacts, labels, pectoral muscles and have low contrast, which affects training and hence reduces proper diagnosis. To remove noise and enhance mammograms contrast, we applied adaptive mean filter Contrast Limited Adaptive Histogram Equalization (CLAHE) and median filter [20, 21].

Since CBIS-DDSM's ROI images are of distinct dimensions, a shift in image size was essential. Considering the aspect ratio as defined in $r = \frac{w}{h}$ where *h*, *w* and *r* are the image's height, width and aspect ratio respectively. From dataset we removed images with an aspect ratio of less than 0.4 and greater than 1.5. We considered aspect ratio to resize the image as a parameter, which helps in both up-sampling and down-sampling procedures to preserve the best quality possible from the original images. Most pre-trained techniques use 224x224 (width, height) target sizes. Cubic interpolation was used for up-sampling, while area interpolation yields highest outcomes for down-sampling.

Due to relatively smaller size of the available dataset, CNN models memorized all the dataset rather than learning it, hence the model over fit the data and decreased its performance on future unseen data. To avoid over fitting, we have performed data augmentation, which generated more instances for training artificially by applying transformations to the actual data such as flipping and rotation. We flipped and rotated patches by 90, 180, and 270 degrees, sharing by 0.2, zooming and random scaled. Such data increase appropriate training samples because tumors can occur in different directions and sizes. Thus, methods of augmentation do not alter the masses fundamental pathology.

4 Experiments

In this section we discuss evaluation metrics, experimental setting and results in detail.

4.1 Evaluation metrics

In this paper, we have used traditional approaches to assess our model performance such as accuracy, recall and precision as indicated in equation 1, 2 and 3. Where *P* and *N* are the total of positive and negative class instances respectively. In our first simple classifier *P* is cancerous and *N* is non-cancerous, while in second MobileNet model *P* is Mass malignant and *N* is mass benign. TP, TN, FP and FN stand for True Positive, True Negative, False Positive and False Negative respectively.

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

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

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{3}$$

4.2 Experimental Setup

Hyper-parameters are variables that determine the structure of the network (e.g. number of hidden layers) and variables that determine how the network is trained (e.g. number of epochs, rate of learning, batch size, decay etc.). Before training the CNNs, hyper-parameters are selected manually.

Learning rate. Learning rate (LR) is one of the most significant hyper-parameters that affects the efficiency of the CNNs. Typically, stochastic gradient descent optimizer is used in deep learning models. Variants of stochastic gradient descent are RMS Prop, Adam, Adagrad, etc. All these optimizers allow the learning pace to be set by customers. To control and minimize losses of a model, all these optimizers use learning rate. Many iterations are required as learning rate is too small. However, if learning rate is too high, it can cause undesirable divergent behavior and destroy pre-trained weights of model. The popular strategies on learning rates are step decline, quadratic decline, square root decline, and linear decline[19]. For mammograms and patches we have used Adam with decay rate of 0.99 per epoch, and a regularization coefficient of $10^{-5}$for training their CNN.

Batch Normalization. A batch normalization layer standardizes input variables over a mini-batch (a subset of the training data set) in a CNN model. It reduces the network initialization sensitivity and speeds up the CNN training process. We chose the largest batch size our GPU memory would allow for batch size. This is a batch size of 64, although we had more computing energy. A greater batch size implies that the information set is better represented. Which implies that our model will converge more quickly and will have less variety when the model tries to discover the optimum minimum in our gradient.

Model Check Point. A model checkpoint was a significant technique we used. This saves the weights of the model for every epoch with the best score. We can use these weights that during practice were saved at a midpoint. This enables us to guarantee that the weights of the model we use do not fit too well on the training data and generalize well into fresh information. Maintaining the model weights with the highest outcomes is also significant. Only saving best weights saves our memory.

4.3 Transfer learning model comparison

Architectures like VGG and ResNet are either big in size or involve a lot of mathematical computations, although they have attained very high accuracies in the Imagenet dataset. Thus, these architectures may not be effective for embedded based vision applications and mobile devices. Deep learning-based architecture for mobile devices that is computationally effective, tiny in size and achieve higher accuracy is MobileNet, proposed by Google in 2017.

MobileNet's primary concept is that instead of using the standard 3*3 convolution filters, the procedure is divided into 3* 3 depth-wise separable convolution filters, followed by 1*1 point-wise convolutions. The new architecture needs fewer operations and parameters while achieving the same filtering and mixture method as a regular convolutional architecture.

Let input image size is $H * W$ with Nin number of channels (e.g 3 for RGB image), filter size $F * F$ and Number of output channels is $Nout$, then number of operations can be computed using equation 4 as follow (here we use F=3):

$$H * W * Nin * F * F * Nout \qquad (4)$$

$$H * W * Nin * (9 * Nout) \quad (5)$$

While in the case of depth-wise separable convolution with filters of size $F * F$ followed by point-wise 1*1 convolution, number of operations is given in equation 6 (here we use F=3):

$$(H * W * Nin * F * F) + (H * W * Nin * Nout) \quad (6)$$

$$H * W * Nin * (9 + Nout) \quad (7)$$

From equation 5 and 7, we can see that the conventional convolution layer needed $(9 * Nout)$ where only $(9 + Nout)$ operations were needed as a depth-separable convolution layer followed by point-wise convolutions. Multiplication is a costly operation with respect to addition for computers. We also used mobileNet-v2 for mass benign and mass malignant classification, which combine inverted ResNet architecture with depth-wise separable and point-wise convolutions.

In ResNet architecture, 3 x 3 convolution is performed on a reduced number of channels, whereas in MobileNet-v2 architecture the 3 x 3 convolution layer is replaced by a 3 x 3 convolution layer with an increased number of channels, but with a smaller number of parameters. The results of our experiments using pre-trained models are listed in Table 1. MobileNet-v1 gives us best accuracy for both DDSM and CBIS-DDSM datasets. Although AlexNet achieve same results as MobileNet-v1, but with huge number of parameters and thus required a lot of computation power.

Table 1. Result Comparison. In the table Acc, Rec, and Pre presents Accuracy, Recall and Precision respectively.

CNN	**Year**	**Number of parameters (million)**	**Mass malignant vs mass benign**					
			CBIS-DDSM			**DDSM**		
			Acc (%)	Rec	Pre	Acc (%)	Rec	Pre
AlexNet	2012	60	73.0	0.74	0.71	86.5	0.90	**0.86**
VGG-16	2014	138	70.2	0.72	0.69	83.0	0.84	0.85
VGG-19	2014	140	70.7	0.72	0.70	84.3	0.85	0.87
ResNet-50	2015	23	63.7	0.66	0.50	85.6	0.89	0.74
GoogLeNet	2014	10	60.0	0.70	0.58	83.0	0.86	0.81
MobileNet-v1	2017	3.2	**74.5**	**0.76**	**0.70**	**86.8**	**0.95**	0.84
MobileNet-v2	2018	2.5	**73.0**	**0.77**	0.65	**85.2**	**0.92**	**0.85**

5 Conclusion and Future Work

In this paper, we provide efficient and accurate CAD system that can help and serve as a second eye for radiologists. For mobile devices and vision-based tools, we provide MobileNet based CNN architecture which takes very less memory space and less computational power. We achieved good accuracy for DDSM, as it has more sample for training and testing. However, CBIS-DDSM is subset of DDSM dataset with high quality images, but less in numbers. Therefore, for model it is difficult to properly learn and generalize well by using less training samples. Accuracy can be achieved by using other preprocessing techniques available in literature or by increasing dataset size. Although MobileNet-v1 give us highest performance but latest version named as MobileNet-v2 is much batter as it performs only 1% less in results by using very less parameters.

In future we aim to provide efficient and accurate mobile and web based clinical-decision support system which is truly independent and deals with all kind of breast cancer screening modalities like mammograms, 3D breast tomosynthesis, MRI, ultrasound. Our system will not only detect cancer but also localize and segment cancer, prove information about cancer size and shape. In addition, generate report for risk assessment with solution treatments like suggest medicine and biopsy treatment. Also remind patient for regular check-up and to visit hospital for test. Exploring other characteristics not restricted to breast tissue would be interesting. Studies have shown powerful correlations with breast cancer in age, gender and family history Future research will also concentrate on domain adaptation and transfer learning techniques from one medical domain to other (e.g use large dataset for brain and other tumors to train a model from scratch and then fine-tune that model for breast cancer). Furthermore, we will explore cancer prediction techniques which can help and generate alarm before cancer generation. Additionally, developments in CNNs can not only assist radiologists, but eventually also allow independent reading of MGs by diagnostic instruments in the close future.

Acknowledgement

This work has been supported by Higher Education Commission under Grant # 2(1064) and is carried out at Medical Imaging and Diagnostics (MID) Lab at COMSATS University Islamabad, under the umbrella of National Center of Artificial Intelligence (NCAI), Pakistan.

References

1. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a cancer journal for clinicians,* vol. 69, pp. 7-34, 2019.
2. S. A. Feig, "Screening mammography benefit controversies: sorting the evidence," *Radiologic Clinics,* vol. 52, pp. 455-480, 2014.
3. H. G. Welch and H. J. Passow, "Quantifying the benefits and harms of screening mammography," *JAMA internal medicine,* vol. 174, pp. 448-454, 2014.
4. J.-G. Lee, S. Jun, Y.-W. Cho, H. Lee, G. B. Kim, J. B. Seo*, et al.*, "Deep learning in medical imaging: general overview," *Korean journal of radiology,* vol. 18, pp. 570-584, 2017.
5. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand*, et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861,* 2017.
6. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.
7. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
8. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov*, et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
9. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510-4520.
10. F. Jiang, H. Liu, S. Yu, and Y. Xie, "Breast mass lesion classification in mammograms by transfer learning," in *Proceedings of the 5th international conference on bioinformatics and computational biology*, 2017, pp. 59-62.
11. B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging,* vol. 3, p. 034501, 2016.
12. A. Rampun, B. W. Scotney, P. J. Morrow, and H. Wang, "Breast Mass Classification in Mammograms using Ensemble Convolutional Neural Networks," in *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2018, pp. 1-6.
13. A. Hamidinekoo, Z. Suhail, E. Denton, and R. Zwiggelaar, "Comparing the performance of various deep networks for binary classification of breast tumours," in *14th International Workshop on Breast Imaging (IWBI 2018)*, 2018, p. 1071807.
14. H. Chougrad, H. Zouaki, and O. Alheyane, "Deep convolutional neural networks for breast cancer screening," *Computer methods and programs in biomedicine,* vol. 157, pp. 19-30, 2018.
15. S. Morrell, Z. Wojna, C. S. Khoo, S. Ourselin, and J. E. Iglesias, "Large-Scale Mammography CAD with Deformable Conv-Nets," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*, ed: Springer, 2018, pp. 64-72.
16. G. Carneiro, J. Nascimento, and A. P. Bradley, "Deep learning models for classifying mammogram exams containing unregistered multi-view images and segmentation maps of lesions," *Deep learning for medical image analysis,* pp. 321-339, 2017.
17. G. Carneiro, J. Nascimento, and A. P. Bradley, "Unregistered multiview mammogram analysis with pre-trained deep learning models," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 652-660.

18. M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The digital database for screening mammography," in *Proceedings of the 5th international workshop on digital mammography*, 2000, pp. 212-218.
19. R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data,* vol. 4, p. 170177, 2017.
20. P. Teare, M. Fishman, O. Benzaquen, E. Toledano, and E. Elnekave, "Malignancy detection on mammography using dual deep convolutional neural networks and genetically discovered false color input enhancement," *Journal of digital imaging,* vol. 30, pp. 499-505, 2017.
21. D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, "Deep convolutional neural networks for mammography: advances, challenges and applications," *BMC bioinformatics,* vol. 20, p. 281, 2019.
22. D. Mishkin, N. Sergievskiy, and J. Matas, "Systematic evaluation of convolution neural network advances on the imagenet," *Computer Vision and Image Understanding,* vol. 161, pp. 11-19, 2017.