15th International Conference on Electronics Computer and Computation (ICECCO 2019)

Deep Learning-based Mammogram Classification using Small Dataset

Adeyinka P. ADEDIGBA*, Steve A. ADESHINA†, and Abiodun M. AIBINU*,
*Department of Mechatronics Engineering, Federal University of Technology, Minna, Nigeria
Email: adeyinka.adedigba@futminna.edu.ng
† Comupter Vision Lab, Nile University of Nigeria, Abuja

***Abstract*—Breast Cancer is one of the most diagnosed cancer and the leading cause of death among women worldwide, second only to lung cancer. Mammographic screening has been the most successful screening technology capable of detecting up to 90% of all breast cancer even before a lump growth can be felt using breast exam. However, mammogram is a low intensity image and the heterogenous nature of breast can make healthy breast tissue appears as cancerous, this is most common among women with dense breast (aged 40 – 44). Thus, the sensitivity for early detection of breast cancer from mammogram has been estimated to 85 – 90%. This result can be improved by Deep CNN, however, to achieve good generalization, it must be train with high voluminous dataset whereas, mammographic dataset exists in smaller volume. In this paper, we present a method of training deep CNN with few datasets to achieve high training result and good generalization. An augmentation technique that increase both size and variance of the dataset is presented herewith, the augmented dataset was used to train five state of the art models. Highest training and validation accuracy (99.01% and 99.99% respectively) were achieved with DensNet. Meanwhile, SqueezeNet, a deep CNN model with fewer parameter also shows promising result, which means soon this model can be deployed into microcontroller and FPGAs for clinical applications.**

***Index Terms*—Breast cancer, Deep Convolution Neural Network, Mammogram, Transfer learning**

I. Introduction

Breast Cancer (BC) is a malignant tumour that begins its development from the breast cells, commonly the milk-producing glands called lobules or the milk passages called ducts. The term carcinoma refers to malignant tumour that develops from the lining epithelial cells of the breast duct (ductal carcinoma) or lobules (lobular carcinoma). Over time, these cancerous cells can invade adjacent breast tissues or metastasize to remote cells of the body; this is referred to as Invasive Breast Cancer [1]. About 85-90% of BC result from genetic and environmental factors while about 5-10% are due to hereditary [1].

BC is one of the most diagnosed cancer and the leading cause of death among women worldwide, second only to lung cancer. It is estimated that every 1 in 8 US women will develop invasive breast cancer in the course of her lifetime; in China, the incidence rate of breast cancer is estimated to be 22.1% in 100,000 people per year [2].

The symptoms of BC ranges from lumps swelling to skin changes while some may show no obvious symptoms at all. The full symptoms of BC were outlined by American cancer Society as follows: swelling of all or part of the breast, skin irritation or dimpling, breast pain, inward turning of nipple accompany with pain, thickening or redness of the nipple or breast skin, unusual nipple discharge and a lump in the underarm [2]. Early detection of BC is easier to treat, offers less risk and reduced mortality rate up to 25% [3]. Thus, breast self-examination has been recommended as part of monthly health routine for women, meanwhile other diagnosis such as biopsy, mammogram, breast histopathology, thermography and ultrasound imaging are highly recommended for women aged 40 years and above [4].

Mammographic screening is the most common and most effective breast cancer screening technology capable of detecting up to 90% of all breast cancer even before a lump growth can be felt using breast exam [5]. Mammogram uses a low-dose X-ray imaging of the breast where tissues in the breast as well as tumours appear as different shades of grey on the image whereas fats which are radiolucent appear black on the image, and masses such as calcifications appears as white (see Fig. 1). Breast screening and breast diagnosis are the two type of breast exams performed using mammography. Mammographic screening aims at detecting the presence of lesion in the breast while breast diagnosis is a follow-up test performed on the patients who have demonstrated abnormality in the clinical findings [6]. Mammographic screening usually involves two views of the breast pairs namely Cranio-Caudal (CC) and Medio-Latera oblique (MLO).

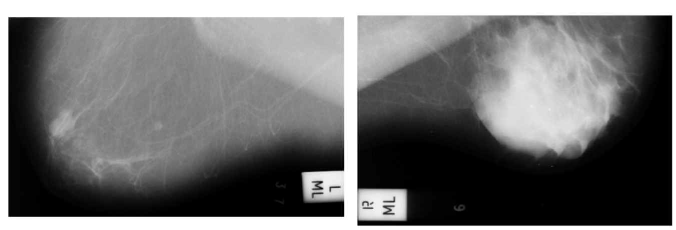

Fig. 1. Mammographic Image pair

Mammographic screening has increased early detection of Breast cancer resulting in early treatment of the tumour.

978-1-7281-5160-1/19/$31.00 © 2019 IEEE

However, mammogram is a low intensity image and the heterogenous nature of breast can make healthy breast tissue appears cancerous, this is most common among women with dense breast (aged 40 – 44) [5] [6] [7]. This results in difficulty of radiologists to interpret results of mammographic screening, the sensitivity of early detection of breast cancer from mammogram has been estimated to 85 – 90% [8]. To assuage this problem, double reading method has been employed where two expert radiologists are employed to interpret a single mammographic image [6]. This method was found to reduce false positive rate but at an extra cost and workload, more so, in some cases, this method is not feasible. Thus, Computer Aided Detection (CAD) algorithms has been proposed based on different machine learning techniques [9] [10] [11] [12] [13]. The most successful of these techniques has been the application of deep convolution networks to the detection of mammogram [14] [15] [16] [17] [18]. While this method is very successful with high accuracy, sensitivity and low false positive rates, it suffers from data availability. Deep Learning models requires huge training data to achieve high accuracy, meanwhile, medical image database is usually small, this has led to high training error and limited the clinical application of these models [16][18]. In this paper, we present a method of training deep learning models using smaller mammogram dataset. The dataset used contains only 322 images which was used to train five large and deep Convolution Neural Networks (CNN) models in a transfer learning approach.

The rest of this paper is organized as follows: theoretical framework and review of related works were presented in Section II, our dataset augmentation technique was presented in Section III along with architecture of models employed in transfer learning; finally the result was presented in Section IV.

II. Related Works

It has been shown that generalization error of deep CNN increases substantially when the training example is small [19], more so, all the state of the art models were trained on very large dataset (typically tens of millions of training dataset) to ensure their training, validation and test accuracy. However, medical images are available in limited number (less than hundreds of thousand), therefore there is need to develop an effective algorithm which adapt network trained on domains with voluminous training data to diminutive dataset available in the medical domain. Domain Adaptation techniques comes handy in this case, it provides mechanism of transferring knowledge from a source domain (for instance, domain with voluminous training examples) to a target domain (where training data is small) by exploring domain-invariant structures that underline distribution discrepancy in the two domains. Transfer learning, an example of domain adaptation technique, is a method of retraining a previously trained deep CNN (base model) in a way that facilitate the reuse of their learned features and applying them on a new task (target model) by fine-tuning their fully connected layers only [20]. Furthermore, Zeiler and Fergus [21] shows that regardless of dataset domain, deep CNN learns similar features in their early layers, owing to their presence in all natural images these features are called 'general' feature in [22]. Thus, deep CNN are able to disentangle underlying features in image distribution and group them together in an hierarchical way in accordance to their related invariance factors such as edges, curves and colour blobs [19] [21] [22].

Many researchers have exploited this mechanism to obtain good classification result from medical images, some also compare results of different base models. For instance, [6] trained four different CNN architectures namely: AlexNet, VGG, GoogLeNet and ResNet while [23] trained only two: AlexNet and GoogLeNet, their results shows that GoogLeNet performs better than the other networks models 95.06% and 93.4% respectively were obtained in [6] and [23]. Ten different (target) model configurations were trained with transferred weights from AlexNet in [24], similarly, with learned features from AlexNet [25] built multi-instance network to predict the presence of cancer in an image. More so, precision of localizing breast cancer tumour has been greatly by transfer learning approach [26] [27].

Although impressive results were obtained from transfer learning, yet, Yosinski et al. shows that model could overfit if the dataset in the target domain is small [22]. To prevent overfitting, data augmentation has been used in different forms. Data augmentation describes methods of increasing the size and variance of dataset utilized in training a machine learning model to achieve better generalization and to capture the underlying distribution of the training dataset. In mammogram classification, patches extraction is the common practise for data augmentation [24] [25] [28]. Although this increases size of the dataset, it fails to introduce variance to the training set besides lacking practical implication. The augmentation technique introduced in this paper not only increases the dataset but introduce variance to the dataset such as ones encountered in practical settings.

III. Methodology

This section discusses the methodology taken in this work. Starting with an introduction to the dataset used, the augmentation algorithm was presented, followed by a brief discussion on the (base) model architectures used in this work. This section concludes with the presentation of experimental setup for training the models.

A. Dataset

To demonstrate how small dataset can be utilize to achieve great training and validation accuracy, Mammographic Image Analysis Society (MIAS) database of mammogram version 1.21 was used. It contains 161 pairs (322) of mammogram taken at resolution of 50 micron which was further reduced to 200 micron pixel such that each image is 1024×1024 grayscale image. The dataset can be grouped into three classes: normal, benign and malignant. In the case of malignancy, the coordinate of centre and radii was provided to locate the calcification region. However, in the case of invasive (BC) where calcifications are widely scattered, centre and radii coordinates were omitted. The database label was clearly laid out to avoid mix-up as provided in Table I.

TABLE I
DESCRIPTION OF MIAS DATASET

Column	Description
1	MIAS database reference number
2	Character of the breast tissue: F – Fatty, G – Fatty-glandular and D – Dense-glandular
3	Class of Abnormality CALC – calcification CIRC – Well-defined masses SPIC – Spiculated Masses ARCH – Architectural distortion ASYM – Asymmetry NORM – Normal
4	Severity of Abnormality B – Benign M – Malignant
5,6	Coordinate of centre of abnormality
7	Approximate radius of a circle enclosing the abnormality

B. Data Augmentation

To train a deep convolution network with little training dataset as ours without overfitting, data augmentation is very imperative. Augmentation in this project was done with two goals: firstly, to increase the dataset and secondly, to increase variance within the dataset. To achieve these, additional images were synthesized by randomly performing gaussian blurring, horizontal flipping, internal refection and mild addition of white noise.

The Gaussian blurring applies two-dimensional Gaussian filters on input image with the aim of removing noise but in this case, Gaussian blurring was used to add within-class variance to the dataset. The filter is developed as an extension of one-dimensional Gaussian filter, given by:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \tag{1}$$

Thus, the two-dimensional Gaussian filter is given by:

$$G(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

Where σ^2 is variance of the Gaussian filter. It can be noticed that (2) is a product of two Gaussian filters. Applying (2) as image filter to pixel coordinate (r, c) according [29] to yields:

$$G(r,c) = e^{\frac{||r-c||^2}{v}} \tag{3}$$

Where r is the row and c is the column coordinate. From (3), it shows that Gaussian blurring works by adjusting the Euclidean distance between neighbouring pixel intensities. Blurring is common phenomenon encountered in medical images. It is usually introduced during the process of capturing the mammographic image. Thus, introducing Gaussian blurring to the training dataset not only increases the dataset, but also increases the variance within the dataset, making the model more robust.

The image horizontal reflection is formed by flipping the row coordinates of the original image to obtain new image, this is mathematically given by:

$$I(r', c) = I(-r, c) \tag{4}$$

Likewise, the image rotation is achieved by randomly adding or subtracting small angle ϕ from the coordinate of the original image, this is mathematically given by:

$$(r', c') = \pm rcos\phi \pm csin\phi \tag{5}$$

The overall augmentation algorithm is as follows: an image is randomly picked (with replacement) from the dataset, the three augmentation transformation to be performed are as defined, then a pipeline of these augmentation transformations is formed which is then randomly selected from and applied on the image. After the transformation is applied to the image, the new augmented image is then saved to disk. This algorithm is summarized in Fig. 2:

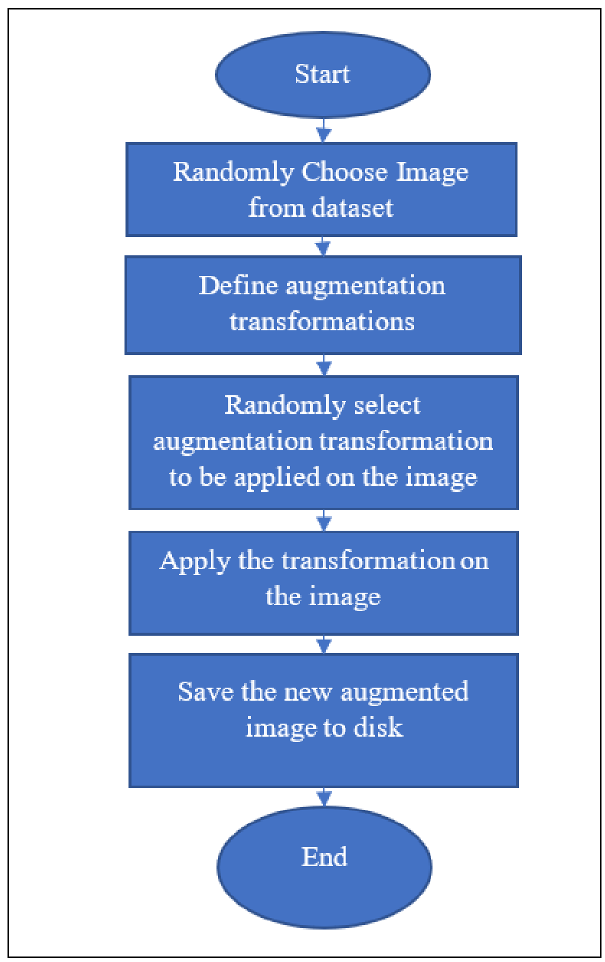

Fig. 2. Data Augmentation Algorithm

C. CNN Model Architecture

To test the hypothesis, transfer learning was performed using popular pre-trained CNN models. In this project, AlexNet, VGG, ResNet, DenseNet and SqueezeNet were used, the architectures as well as the central design idea of each of these networks is discussed herewith:

AlexNet [30]: is an eight-layer network consisting of five convolutional layers and three fully connected layers, pre-trained on the high-resolution ImageNet dataset. AlexNet, developed by Alex Krizhevsky, Geoffrey Hinton, and Ilya Sutskever, won the 2012 ImageNet competition with 15.3% top-5 error rate and has since become one of the baseline models in CNN history.

VGG [31]: is a 16-layer CNN developed by the Visual Geometry Group, Oxford University. The model was pre-trained on ImageNet dataset for ImageNet competition. VGG is the 1st runner-up of 2014 ImageNet classification task. VGG is desired for its uniform 3x3 convolution kernel used in building the architecture of the model and due to its simple kernel structure, it has become a favourable model for feature extraction tasks.

SqueezeNet [32]: achieves similar performance to AlexNet but with 50% fewer parameters. The parameter reduction was achieved by using 1×1 filter instead of larger filters and by decreasing the number of input channels to their 3×3 filters. This follows from [33] and [31] where smaller filters has been shown as an approximation of larger filters. Thus, instead of using larger filters, smaller filters are repeatedly used throughout the network which guarantee parameter reduction. The accuracy, on the other hand, was maintained by ensuring that each convolution layer receives large activation maps from the previous layer, that is, pooling (or downsampling) was not applied to earlier layers of the network. These key intuitions were implemented in the *fire* module of the network, which comprises of a *squeeze* module (1×1 filter) and an *expand* module (which has a 1×1 filter followed by 3×3 filter). The model was trained using similar parameters to AlexNet and its performance was benchmark against AlexNet. It was found that SqueezeNet perform as well as AlexNet, despite its fewer parameters [32].

ResNet [34]: generally, deeper convolutional network outperforms shallow counterpart [35], however training deeper model increases training error rate due to vanishing gradient problem [34]. To solve this, ResNet introduced Residual block (6) which creates a connection between the output of a convolutional layer and the original input to the layer using identity mapping [34]. Thus, the activation of a Residual block is given as:

$$\mathbf{a}_l = \mathcal{U}(\mathbf{a}_{l-1}) + \mathbf{a}_{l-1} \tag{6}$$

Where $\mathbf{a}_l$ is the activation of layer l , $\mathcal{U}(\cdot)$ is a nonlinear convolutional transformation of the layer and $\mathbf{a}_{l-1}$ is the activation of previous layer $l-1$. The skip connection of (6) enables more layers to be stacked on each other resulting in a deep network. The ResNet 152, a 152-layer convolutional network won the 2015 ImageNet competition with 3.57% top-5 error rate, higher than human-level performance. In this work, ResNet 101, a 100-layer convolutional network pretrained on ImageNet dataset was used.

DenseNet [36]: it is possible to train much deeper network with fewer parameters and better accuracy than ResNet, by implementing Dense block (7) instead of residual block (6). The dense block creates a form of connection that allows any layer within the network to be connected to all layers that follows it [36]. That is, layer l receives feature activations from all its preceding $l-1$ layers as follows:

$$\mathbf{a}_l = \mathcal{T}([\mathbf{a_0}, \mathbf{a_1}, \ldots \mathbf{a_{(l-1)}}]) \tag{7}$$

Where $\mathbf{a}$ is the activation of the l^{th} layer, $[a_0, a_1, \ldots a_{(l-1)}]$ is a concatenation of all the previous layer activations which can be seen as a form of collective information gathered by the network up to layer $l-1$. $\mathcal{T}(\cdot)$ is a nonlinear transformation that maps the concatenated activation to the activation of layer l. Comparing (6) and (7): the element-wise operation of the skipped connection in (6) resulted in parameter increase of $O(C{\times}C)$, whereas (7) resulted in fewer parameters of $O(l{\times}k{\times}k)$ where C is the number of channels, k is the growth order of the dense connection and l is the number of layers. For example, ResNet101, a 101 layers convolutional network has 10.2M parameters while DenseNet-BC (with k = 12), a 100 layers convolutional has 0.8M parameters [36].

D. Experimental Setup

Dataset: The image in the dataset was first resized to 400x400 from its original 1024x1024 resolution to reduce memory implication. Then augmentation algorithm presented in Fig. 2 was implemented on these resized images to generate 9,000 images of which 8,000 was used for training and 1,000 for validation.

Parameters: Transfer learning was used to initialize the weights of each model. The input to all the models is 224x224 grayscale image while the output layer was set to 3, according to the number of classes in this project. The base learning rate of 10^{-4} was used, which was gradually reduced by 0.1 at every 10 epochs. Adam optimization algorithm was used with momentum of 0.9, mini-batch size of 32 for 50 iteration. AlexNet, VGG and SqueezeNet were trained with dropout rate 0.4 to prevent overfitting.

Hardware/Software: The setup was implemented on PyTorch and trained on Assus with NVIDIA RTX 2070, Intel Core i7-8750H for three hours.

IV. Discussion of Result

The experimental results are presented in this section. A performance comparison of augmented and unaugmented training dataset in deep CNN has been presented in [23], in this work we focus on improving the performance of deep CNN with augmented dataset.

Results shown in Table II, Fig. 3 and Fig. 4 were obtained when the models were trained using augmented dataset.The best training and validation result were obtained

TABLE II
RESULT OF TRANSFER LEARNING WITH AUGMENTED DATASET

Model Name	Training Accuracy (%)	Validation Accuracy(%)
AlexNet	94.5	81.11
SqueezeNet	95.3	77.78
VGG	98.68	96.67
ResNet	98.39	99.99
DenseNet	99.01	99.99

with DenseNet while AlexNet performed least in training accuracy and SqueezeNet perform least in validation accuracy. From the result in Table II, the following can be deduced:

- Deeper network outperforms shallow counterparts. The top training and validation accuracies were obtained from DenseNet and ResNet which has depth of 100 and 101 respectively. The outstanding performance of DenseNet owes to their improved information (i.e availability of more activation maps) and gradient flow throughout the network, as mentioned earlier, this was made possible by the dense connection of all layers within the network.
- Smaller filter size is indeed an approximation of larger counterparts and can improve convergence as well as performance. For instance, VGG which uses filter size of *3×3* achieves slightly superior training result than deeper ResNet. This is further proved by SqueezeNet's better training performance compared to AlexNet. As earlier said, SqueezeNet used *1×1* and *3×3* filter size in its squeeze and expand module respectively, compared to AlexNet which uses *7×7* filter size in its first layer.
- Lastly, the performance of SqueezeNet raises hope of hardware realization of breast cancer detection using deep learning. Given its fewer parameter and low memory requirement, this can be easily deployed to FPGA and microcontrollers [32]. Although not shown in this work, SqueezeNet can be fine-tuned for better performance by adjusting its hyperparameters.

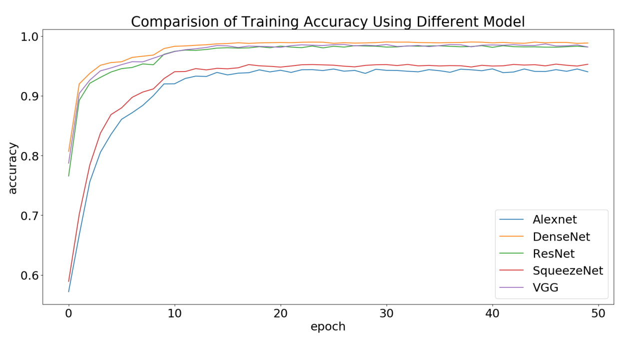

Fig. 3. Comparison of Training Accuracy using different models

Augmentation is a very germane to application of deep learning to medical dataset, a carefully augmented dataset can greatly improve performance. By carefully augmenting a small dataset, we achieve better performance compared to those reported in literature. Comparison of our result to those

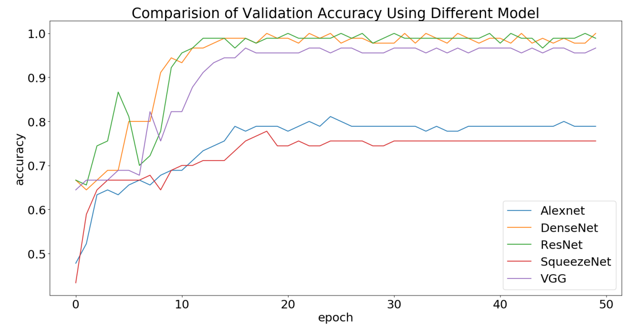

Fig. 4. Comparison of Validation Accuracy using models

TABLE III
COMPARISON OF OUR RESULT TO THOSE REPORTED IN LITERATURE

Reference	Highest Accuracy Reported (%)
[6]	95.06
[23]	93.4
[27]	88
[14]	**98.82**
[15]	92
[16]	91.5
Our method	**99.01**

reported in literature is presented in Table III. The highest accuracy was reported in [14] where 500 mammographic images were augmented to train a deep CNN in a transferred learning approach, Table II and Table III show that our approach yields better accuracy with different models.

V. CONCLUSION

A method of training deep learning models on smaller medical dataset has been presented in this paper. Medical data was passed through a pipeline of image transformations so as to increase size and introduce variance within the dataset. The augmented dataset is used to train popular deep CNN architectures. Comparing the result of different model, we observed that deeper models outperforms shallow ones and models that uses smaller filter (or kernel) size trains faster and obtained higher training accuracy than those that use larger filter size.

REFERENCES

[1] L. M. Mina and N. A. M. Isa, "A fully automated breast separation for mammographic images," in *2015 International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*. IEEE, 2015, pp. 37–41.

[2] G. Zhang, W. Wang, J. Moon, J. K. Pack, and S. I. Jeon, "A review of breast tissue classification in mammograms," in *Proceedings of the 2011 ACM Symposium on Research in Applied Computation*. ACM, 2011, pp. 232–237.

[3] B. Hela, M. Hela, H. Kamel, B. Sana, and M. Najla, "Breast cancer detection: A review on mammograms analysis techniques," in *10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13)*. IEEE, 2013, pp. 1–6.

[4] M. G. Marmot, D. Altman, D. Cameron, J. Dewar, S. Thompson, and M. Wilcox, "The benefits and harms of breast cancer screening: an independent review," *British journal of cancer*, vol. 108, no. 11, p. 2205, 2013.

[5] J. Dheeba, N. A. Singh, and S. T. Selvi, “Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach,” *Journal of biomedical informatics*, vol. 49, pp. 45–52, 2014.
[6] P. Xi, C. Shu, and R. Goubran, “Abnormality detection in mammography using deep convolutional neural networks,” in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 2018, pp. 1–6.
[7] T. Roberts, M. Newell, W. Auffermann, and B. Vidakovic, “Wavelet-based scaling indices for breast cancer diagnostics,” *Statistics in medicine*, vol. 36, no. 12, pp. 1989–2000, 2017.
[8] N. Dhungel, G. Carneiro, and A. P. Bradley, “Automated mass detection in mammograms using cascaded deep learning and random forests,” in *2015 international conference on digital image computing: techniques and applications (DICTA)*. IEEE, 2015, pp. 1–8.
[9] Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang, “Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7340–7351.
[10] C. Selvi and M. Suganthi, “A novel enhanced gray scale adaptive method for prediction of breast cancer,” *Journal of medical systems*, vol. 42, no. 11, p. 221, 2018.
[11] D. Saraswathi and E. Srinivasan, “An ensemble approach to diagnose breast cancer using fully complex-valued relaxation neural network classifier,” *International Journal of Biomedical Engineering and Technology*, vol. 15, no. 3, pp. 243–260, 2014.
[12] A. I. Mohd, A. Mohamed Alhaj, K. Shahreen, O. I. Ashraf, H. A. Anik, A. L. Saima, and A. Ali, “An enhancement of multi classifiers voting method for mammogram image based on image histogram equalization,” 2018.
[13] Y.-C. Zeng, “Mammogram density classification using double support vector machines,” in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*. IEEE, 2018, pp. 547–550.
[14] A. A. Mohamed, W. A. Berg, H. Peng, Y. Luo, R. C. Jankowitz, and S. Wu, “A deep learning method for classifying mammographic breast density categories,” *Medical physics*, vol. 45, no. 1, pp. 314–321, 2018.
[15] W. Lotter, G. Sorensen, and D. Cox, “A multi-scale cnn and curriculum learning strategy for mammogram classification,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 169–177.
[16] R. Zhang, H. Zhang, and A. C. Chung, “A unified mammogram analysis method via hybrid deep supervision,” in *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Springer, 2018, pp. 107–115.
[17] W. Zhu, X. Xiang, T. D. Tran, and X. Xie, “Adversarial deep structural networks for mammographic mass segmentation,” *arXiv preprint arXiv:1612.05970*, 2016.
[18] S. A. Adeshina, A. P. Adedigba, A. A. Adeniyi, and A. M. Aibinu, “Breast cancer histopathology image classification with deep convolutional neural networks,” in *2018 14th International Conference on Electronics Computer and Computation (ICECCO)*. IEEE, pp. 206–212.
[19] M. Long, Y. Cao, J. Wang, and M. I. Jordan, “Learning transferable features with deep adaptation networks,” *arXiv preprint arXiv:1502.02791*, 2015.
[20] S. Han, J. Pool, J. Tran, and W. Dally, “Learning both weights and connections for efficient neural network,” in *Advances in neural information processing systems*, 2015, pp. 1135–1143.
[21] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *European conference on computer vision*. Springer, 2014, pp. 818–833.
[22] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?” in *Advances in neural information processing systems*, 2014, pp. 3320–3328.
[23] D. Lévy and A. Jain, “Breast mass classification from mammograms using deep convolutional neural networks,” *arXiv preprint arXiv:1612.00542*, 2016.
[24] X. Zhang, Y. Zhang, E. Y. Han, N. Jacobs, Q. Han, X. Wang, and J. Liu, “Classification of whole mammogram and tomosynthesis images using deep convolutional neural networks,” *IEEE transactions on nanobioscience*, vol. 17, no. 3, pp. 237–242, 2018.
[25] W. Zhu, Q. Lou, Y. S. Vang, and X. Xie, “Deep multi-instance networks with sparse label assignment for whole mammogram classification,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 603–611.
[26] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, “Detecting and classifying lesions in mammograms with deep learning,” *Scientific reports*, vol. 8, no. 1, p. 4165, 2018.
[27] 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*. ACM, 2017, pp. 59–62.
[28] 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, no. 4, pp. 499–505, 2017.
[29] P. Milanfar, “A tour of modern image filtering: New insights and methods, both practical and theoretical,” *IEEE signal processing magazine*, vol. 30, no. 1, pp. 106–128, 2012.
[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[31] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
[32] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size,” *arXiv preprint arXiv:1602.07360*, 2016.
[33] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
[34] 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.
[35] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, “Multi-digit number recognition from street view imagery using deep convolutional neural networks,” *arXiv preprint arXiv:1312.6082*, 2013.
[36] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.