Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
 Volume 23, Issues 2, pp. 51–58, DOI 10.12694/scpe.v23i2.1975

A TRANSFER REPRESENTATION LEARNING APPROACH FOR BREAST CANCER DIAGNOSIS FROM MAMMOGRAMS USING EFFICIENTNET MODELS

PARITA OZA,* PAAWAN SHARMA, AND SAMIR PATEL †

Abstract. Breast cancer is a deadly disease that affects the lives of millions of women throughout the world. Over time, the number of cases of breast cancer has increased. Preventing this disease is difficult and remains unidentified, but the survival percentage can be improved if diagnosed early. The progress of computer-assisted diagnosis (CAD) of breast cancer has seen a lot of improvements thanks to advances in deep learning. With the notable advancement of deep neural networks, diagnostic capabilities are nearing a human expert's. In this paper, we used EfficientNet to classify mammograms. This model is introduced with the new concept of model scaling called compound scaling. Compound scaling is the strategy which scales the model by adding more layers to extend the receptive field along with more channels to catch the detailed features of larger input. We also compare the performance of various variants of EfficientNet over CBIS-DDSM mammogram datasets. We used the optimum fine-tuning procedure to represent the importance of transfer learning (TL) during training.

Key words: Convolutional Neural Networks, EfficientNet, Breast Cancer, Transfer Learning

AMS subject classifications.

1. Introduction. Breast cancer is the most frequent type of cancer worldwide, especially among women, and it is also the leading cause of death. Breast cancer can be detected early, allowing for better treatment planning and a higher survival rate. The most effective techniques for early detection of breast cancer are several imaging modalities such as mammography, Breast MRI, Breast Ultrasound, and PET CT [1]. Computer-aided diagnosis (CAD) systems are being developed for the automated diagnosis of breast cancer. This system enhances the accuracy of findings and the ability to distinguish between abnormalities such as mass, micro-calcification, architectural distortion, etc. CAD systems can act as a double reader solely meant to assist a radiologist; only expert clinicians make final choices.

Deep convolutional neural networks are commonly used in various medical imaging tasks such as cancer detection, classification, and segmentation [18]. Unfortunately, training a network from the ground up can take days or weeks and necessitates a lot of computational power. The research community, on the other hand, already has an access to pre-trained networks like as AlexNet [2], VGGNet [3], ResNet [4], Google Inception Family [5], EfficientNet [6], and so on. Rather than beginning from scratch, most current research suggests leveraging pre-trained networks. On the other hand, state-of-the-art networks are built and tested on datasets that are substantially more diverse [7]. As a result, such networks' capacity and complexity may exceed the needs of smaller datasets, resulting in severe drawbacks when learning from scratch. As a result, several papers have appeared in which the authors call for comprehensive training [7]. In light of the aforementioned, we examine the performance of EfficientNet [6] using the transfer learning approach. Furthermore, we compare the performance of various variants of the EfficientNet family by commencing the training with pre-trained weights.

The rest of the paper is organized as follows: Section 2 deals with the related work in the domain. Then, in section 3, we discuss the EfficientNet model in brief. Then, section 4 presents the proposed methodology used in this work. Experimental results are discussed in section 5. We finally end with the conclusion in section 6.

*Pandit Deendayal Energy University, Nirma University (parita.ophd19@sot.pdpu.ac.in, parita.prajapati@nirmauni.ac.in).

†Pandit Deendayal Energy University

2. Related Work in the Domain. Rahman et al. [8] proposed modified versions of InceptionV3 and ResNet50. The authors have altered the output layer and added two fully connected layers. During the experiment, the first seven layers of the InceptionV3 model were frozen, and two fully linked layers were added, with the last layer being replaced with a Softmax layer for binary classification. A similar logic was applied for ResNet50. Chougrad et al. [9] investigated the significance of transfer learning and tested various deep CNN models to find the optimum fine-tuning technique. With the Swish activation function, a modified VGG16 model was proposed in [10]. Authors have shown that the modified VGG16 model with the Swish activation function delivers better accuracy than Relu activation. A comprehensive study of mammogram classification techniques of various deep learning and machine learning approaches is presented in [11]. Apart from these, Support vector machine (SVM), naive bayes , artificial neural network (ANN), and set classifiers [12] are some of the machine learning algorithms that have proven popular for the development of computer-aided diagnosis systems for breast cancer [13, 14]. Another work by Ikechukwu et al. [19] presented a comparative study of two pre-trained models, such as ResNet-50 and VGG-19, against training a model from scratch (Iyke-Net). Data augmentation and dropout regularization were employed to reduce overfitting. Authors concluded that the pre-trained models with sufficient fine-tuning were comparable to Iyke-Net, a CNN developed from scratch, with a recall of 92.03 percent.

We found that transfer learning plays a substantial role in various deep learning algorithms based on our literature review. With a modest number of datasets, this method is useful in the medical arena [15, 21]. Different existing models based on a short dataset with the CNN architecture and the transfer learning method have not been completely investigated till now. As a result, using a modified state-of-the-art CNN architecture, there is potential for additional advancement in deep learning approaches.

3. EfficientNet Model Scaling. Convolutional Neural Networks have become common in the realm of Computer Vision since Alexnet won the 2012 ImageNet Challenge. However, one of the most challenging aspects of developing CNNs is model scaling so as to improve model accuracy. This process is time-consuming and also necessitates manual trial and error until a sufficiently accurate model is generated while meeting the resource constraints [6]. The procedure consumes a lot of resources and time, and it often results in models that aren't as accurate or efficient as they could be. In response to this issue, Google published a study in 2019 that discussed a new family of CNNs called EfficientNet [6]. The authors of this paper contributed two things:

- Development of mobile-friendly baseline architecture.
- The concept of compound scaling introduces a strategy for expanding model size and maximizing accuracy improvements.

The concept of compound scaling strategy for expanding model size and maximizing accuracy improvements. Depth, breadth, and resolution are three parameters to scale the convolutional neural network. The number of layers in a network refers to the network depth. The number of neurons in a layer, or the number of kernels or filters in a convolutional layer, is related to the width. The input image's height and width are used to determine the resolution. Figure 3.1 shows pictorial representation of compound scaling. An EfficientNet introduces two rules.

- The scaled models' layers/stages will all use the same convolution techniques as the baseline network.
- All layers must be scaled in the same way, with the same ratio.

All layers must be scaled in the same way, with the same ratio. Equation 3.1 mathematically presents the definition of EfficientNet imparting these two rules.

$$N(d, w, r) = \sum_{1..s} F_i^{d, L_i}(X_(r.H_i, r.W_i, w.C_i)) \tag{3.1}$$

where w, d, r are scaling coefficients to scale width, depth, and resolution of the network; F_i, L_i, H_i, W_i, C_i are predefined parameters in baseline network. The authors offer a simple but successful scaling strategy that employs a compound coefficient to uniformly scale network breadth, depth, and resolution in a principled manner (see equations 3.2 to 3.4):

$$depth : d = \alpha^{\phi} \tag{3.2}$$

$$width : w = \beta^{\phi} \tag{3.3}$$

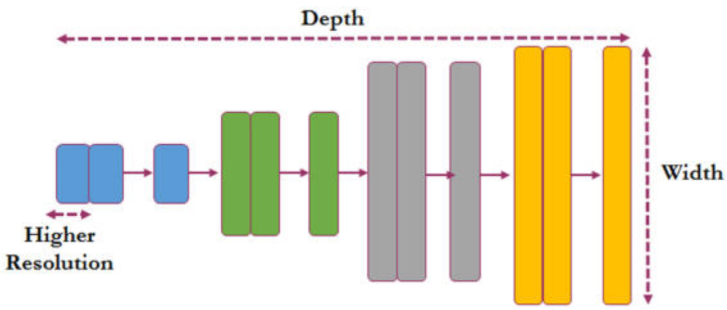

Fig. 3.1: Compound Scaling

Table 4.1: Image input shape expected for each model

Basemodel	**Resolution**
EfficientNetB0	224
EfficientNetB1	240
EfficientNetB2	260
EfficientNetB3	300
EfficientNetB4	380
EfficientNetB5	456
EfficientNetB6	528
EfficientNetB7	600

$$resolution : r = \gamma^{\phi} \tag{3.4}$$

With the help of grid search and by setting ϕ=1, the parameters α, β, and γ can be determined. Once these parameters have been identified, they can be fixed, and the compound coefficient ϕ increases to produce larger but more accurate models. EfficientNet B1-B7 are built in this manner, with the integer at the end of the name denoting the value of the compound coefficient.

4. Proposed Methodology. In this work, we used different versions of EfficientNet. We used TL [16, 17] approach to combat the effect of overfitting. Figure 4.1 shows the complete methodology adopted for the work carried out in this study. The TL is applied to the EfficientNet model (all versions) to classify mammograms in our work. As shown in figure 4.1, we used the recent implementation of deep neural networks that incorporates TL by using parameters of a pre-trained model for a particular task to initialize the new model with certain modifications. First, we created a base model and populated it with pre-trained weights. All the layers in the base model are then frozen by setting "trainable" as a "False". A new model is then created on top of the output of one (or several) layers from the base model. Finally, we train the new model on CBIS-DDSM [20] dataset. The classic oscillating problem is handled by varying the learning rate from 0.001 to 0.0005. The EfficientNet family has eight models, B0 to B7, out of which we used EfficientNetB0 to B5 and EfficientNetB7 in our work. Many factors control the choice of depth, resolution, and width. Therefore, the input shapes for B0 through B7 basic models differ. Table 4.1 shows the input shapes that are predicted for each model. To improve the model's performance and mitigate the effect of overfitting, data augmentation methods are also used in the proposed work. Table 4.2 shows the hyperparameters used to train all the variants of EfficientNet as well as parameters for the augmented strategies.

About Dataset: CBIS-DDSM (Curated Breast Imaging Subset of DDSM) [20] is a standardised and improved version of DDSM. The 10,239 mammographic images, with normal, benign, and malignant cases, were chosen and curated by a skilled mammographer. The images are converted to DICOM format, and the

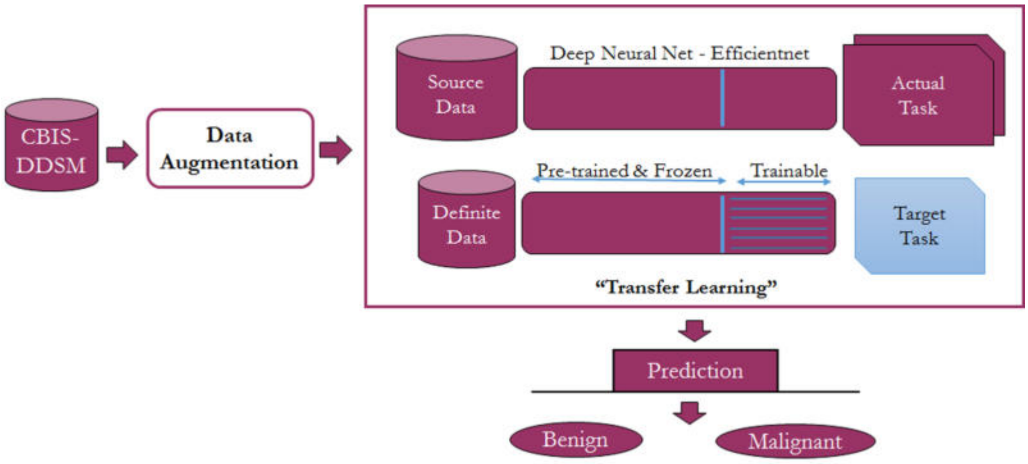

Fig. 4.1: Methodology

Table 4.2: Model Parameters for Training and Data Augmentation

Hyper Parameters for training	
Batch Size	64
Validation split	0.2
Epochs	100
Learning Rate	0.0005
Loss Function	Binary Crossentropy
Optimizer	Rmsdrop
Data Augmentation Parameters	
Rotation Range	180
Shear Range	10
Zoom Range	0.2
Fill Mode	reflect
Horizontal and Vertical Flip	True

ROI segmentation for each lesion is updated. The dataset is separated into training and testing subgroups to directly compare performance between different methodologies. Due to extensive memory usage during training time, we used 6700 images for our work.

5. Result Analysis and Discussion. We carried out experiments of the proposed model on "The PARAM Shavak system ." The system has two multi-core x86_64 CPUs, each having 12 or more cores. The GPU card used for this work is Intel Xeon Phi or Nvidia Tesla GPGPU. Moreover, the system has 64 GB RAM and 8 TB RAID-5 storage. With the default ratio of 80:20, we split the dataset into train and test random splits. The improved models are trained with 100 epochs and 64 instances each batch. In our research, we employed accuracy and loss as performance metrics. We measured accuracy and loss for all types of cases, including train, test, and validation. Figures 5.1, 5.2, and 5.3 show the training and validation performance of the improved EfficientNet models, respectively.

The validation and test accuracy, as well as the validation and test loss, for all EfficientNet models, are summarized in Table 5.1. We found that the performance of EfficientNetB2 and EfficientNetB3 are nearly identical. We set up the models so that the accuracy and loss are optimum. Early stopping was also employed to keep track of the validation loss. The best findings, as well as the results collected to the final epoch, have

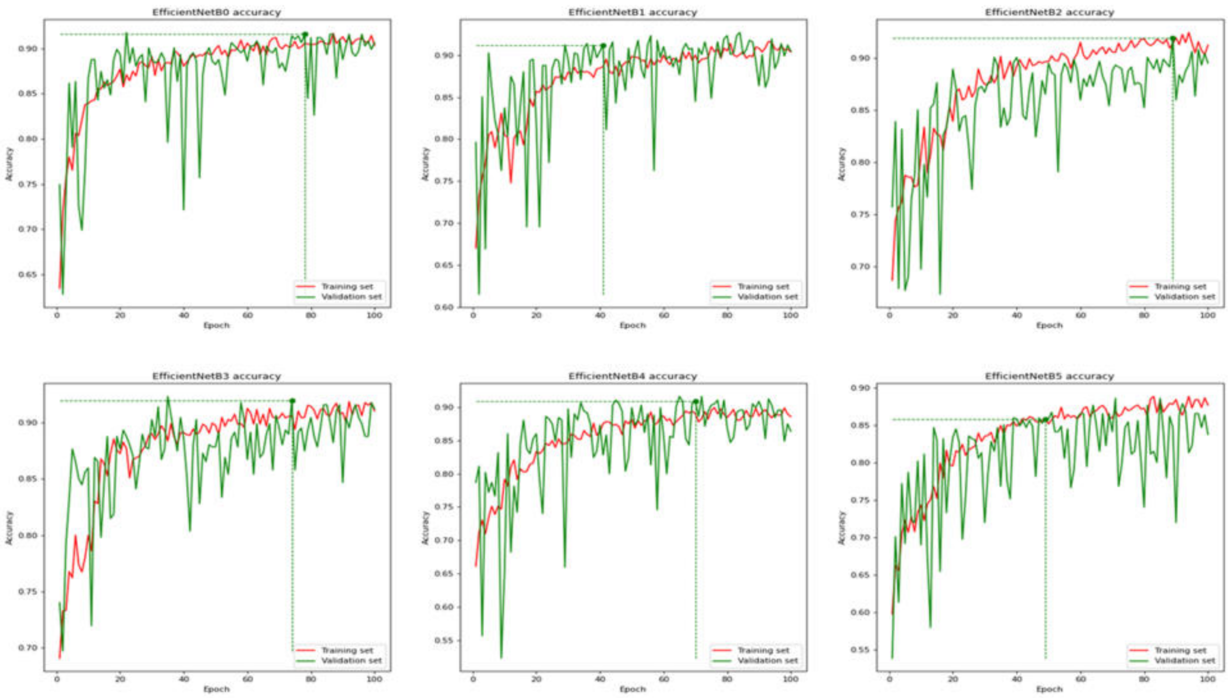

Fig. 5.1: Accuracy for EfficientNet B0 to B5

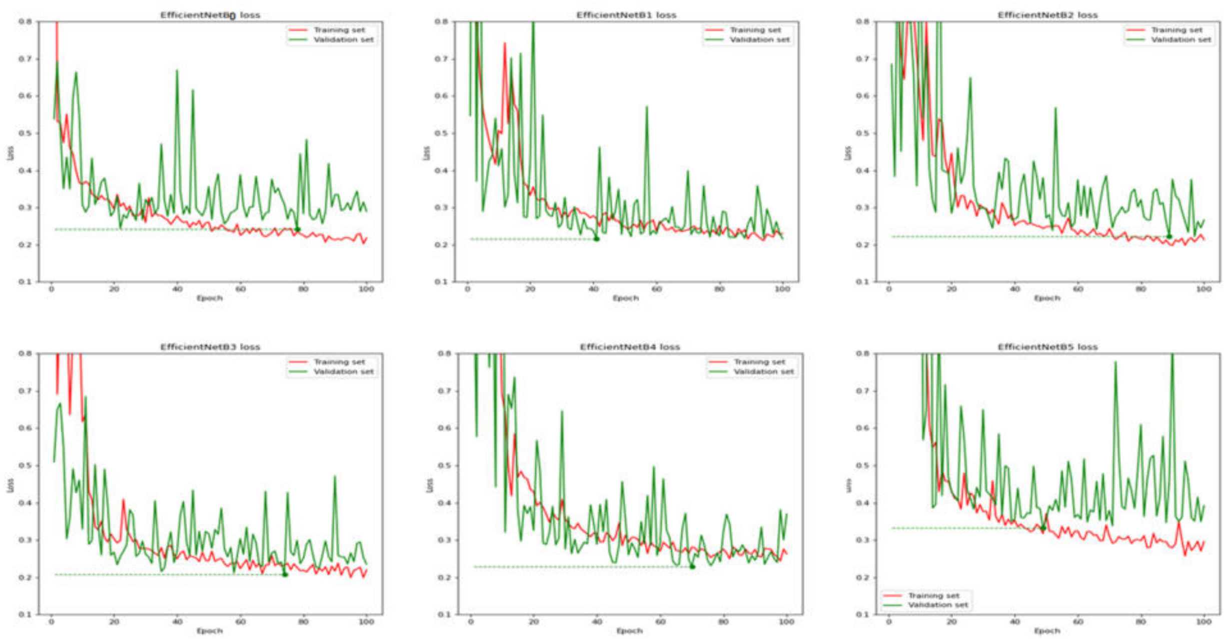

Fig. 5.2: Loss for EfficientNet B0 to B5

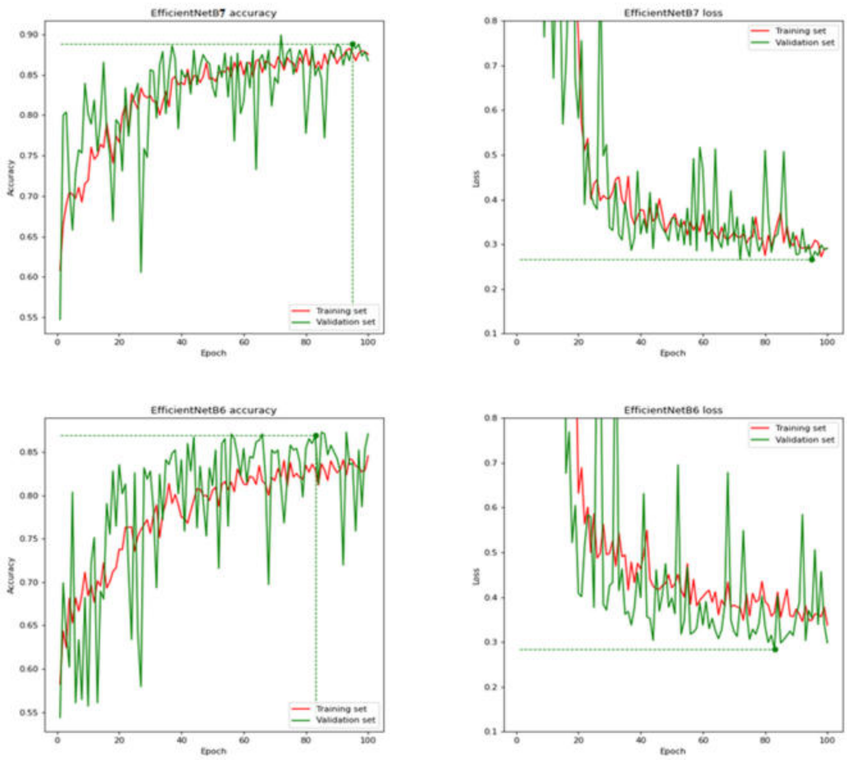

Fig. 5.3: Accuracy and Loss for EfficientNet B6 and B7

Table 5.1: Performance Measures by Various EfficientNet Models

	Validation Accuracy	**Validation Loss**	**Test Accuracy**	**Test Loss**
EfficientNetB0	0.9159	0.2413	0.8512	0.3928
EfficientNetB1	0.9121	0.2154	0.8810	0.2793
EfficientNetB2	0.9196	0.2223	0.8155	0.2734
EfficientNetB3	0.9196	0.2070	0.9018	0.2471
EfficientNetB4	0.9084	0.3699	0.8452	0.3940
EfficientNetB5	0.8579	0.3326	0.8482	0.3521
EfficientNetB6	0.8692	0.2839	0.8125	0.3888
EfficientNetB7	0.8879	0.2655	0.9018	0.2389

been reported. Figure 5.3 shows that EfficientNetB7 has the best accuracy and loss at almost the last epoch (100 for our example), while EfficientNetB6 could result into its best performance prior to the final epoch. EfficientNetB6 almost falls between EfficientNetB5 and EfficientNetB7 in terms of validation accuracy and loss. In terms of test accuracy, EfficientNetB2 and B6 are nearly equal. The optimized outcomes for the other models can be observed before the final epoch (See figure 5.1 and 5.2).

EfficientNet gets very high accuracy while using fewer parameters. A baseline network called EfficientNet-B0 was created first, and then scaled it up to create Efficient-B1 through B7. Comparing EfficientNetB7 to all other versions of the EfficienNet family, we can observe that it offers the best test accuracy and the lowest test loss. The idea behind this neural network is that larger input images necessitate additional layers, which expand the receptive field, and more channels, which enable the network to catch more fine-grained patterns on the larger images. With 600×600 resolution, EfficientNetB7 is the largest EfficientNet model that has obtained state-of-the-art performance on the datasets like CIFAR-100 and ImageNet. The outcome demonstrates that the model performs just as well on medical datasets, including the one utilized in this study.

Compound scaling is a better way to scale up neural networks. The main idea behind the compound scaling approach is the notion of balancing width, depth, and resolution dimensions by scaling with a fixed ratio. In the table 4.1, we present resolution parameters for each EfficientNet model that we employed in our research. The remaining parameters such as depth and width are predefined in the baseline network. Section 3 presents a brief discussion of the selection process utilized by EfficientNet models for all of these parameters.

6. Conclusion. It is preferable to scale up neural networks using compound scaling. The primary principle of the compound scaling method adopted by EfficientNet model family is to scale the model with a constant ratio in order to balance the width, depth, and resolution parameters. On several versions of EfficientNet, we present a transfer representation learning approach in this study. The deep neural model's classification accuracy improves when the fine-tuning approach is used. We discovered that the performance of EfficientNetB2 and EfficientNetB3 are practically equal in our tests. Furthermore, in comparison to other models, EfficientNetB3 is relatively stable in terms of validation and test accuracy. The presented approach is used for binary classification, but it can be modified to work with multi-class classification as well.

Acknowledgment. The authors express their gratitude to the Department of Computer Science and Engineering, Nirma University, Ahmedabad, for providing computing facilities for the studies.

REFERENCES

[1] Jalalian, Afsaneh, et al., *Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection.*, EXCLI journal 16 (2017): 113.

[2] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton., *Imagenet classification with deep convolutional neural networks.*, Advances in neural information processing systems 25 (2012): 1097-1105.

[3] Simonyan, Karen, and Andrew Zisserman. *Very deep convolutional networks for large-scale image recognition.*, arXiv preprint arXiv:1409.1556 (2014).

[4] He, Kaiming, et al., *Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016).

[5] Szegedy, Christian, et al., *Going deeper with convolutions.*, Proceedings of the IEEE conference on computer vision and pattern recognition.(2015).

[6] Tan, Mingxing, and Quoc Le., *Efficientnet: Rethinking model scaling for convolutional neural networks.*, International Conference on Machine Learning. PMLR,(2019).

[7] Tsochatzidis, Lazaros, Lena Costaridou, and Ioannis Pratikakis., *Deep learning for breast cancer diagnosis from mammograms—a comparative study.*, Journal of Imaging 5.3 (2019): 37.

[8] Rahman, Anas S. Abdel, et al., *Breast Mass Tumor Classification using Deep Learning.*, 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT). IEEE, (2020).

[9] Chougrad, Hiba, Hamid Zouaki, and Omar Alheyane., *Deep convolutional neural networks for breast cancer screening.*, Computer methods and programs in biomedicine 157 (2018): 19-30.

[10] Muduli, Debendra, Ratnakar Dash, and Banshidhar Majhi., *Enhancement of Deep Learning in Image Classification Performance Using VGG16 with Swish Activation Function for Breast Cancer Detection.*, International Conference on Computer Vision and Image Processing. Springer, Singapore, (2020).

[11] Oza, P.; Shah, Y.; Vegda, M., *A Comprehensive Study of Mammogram Classification Techniques.*, In Tracking and Preventing Diseases with Artificial Intelligence; Springer: Berlin/Heidelberg, Germany, 2021; pp. 217–238.

[12] Saxena, S.; Gyanchandani, M., *Machine learning methods for computer-aided breast cancer diagnosis using histopathology: A narrative review.* J. Med. Imaging Radiat. Sci. 2020, 51, 182–193.

[13] Pillai, R.; Oza, P.; Sharma, P., *Review of machine learning techniques in health care.* In Proceedings of the ICRIC 2019, Jammu, India, 8–9 March 2019; Springer: Cham, Switzerland, 2020; pp. 103–111.

[14] Oza, P.; Sharma, P.; Patel, S., *Machine Learning Applications for Computer-Aided Medical Diagnostics.*, In Proceedings of the Second International Conference on Computing, Communications, and Cyber-Security, Ghaziabad, India, 3–4 October; Springer: Singapore, 2021; pp. 377–392.

[15] Oza, Parita, et al., *A Bottom-Up Review of Image Analysis Methods for Suspicious Region Detection in Mammograms.*, Journal of Imaging 7.9 (2021): 190.

[16] Elshafey, Mohamed Abdelmoneim, and Tarek Elsaid Ghoniemy., *A hybrid ensemble deep learning approach for reliable breast cancer detection.*, International Journal of Advances in Intelligent Informatics 7.2 (2021): 112-124.

[17] Saber, Abeer, et al., *A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique.* IEEE Access 9 (2021): 71194-71209.

[18] Oza, Parita, et al., *Deep convolutional neural networks for computer-aided breast cancer diagnostic: a survey.*, Neural Computing and Applications. (2022) https://doi.org/10.1007/s00521-021-06804-y

[19] Ikechukwu, A. Victor, et al., *ResNet-50 vs VGG-19 vs training from scratch: A comparative analysis of the segmentation and classification of Pneumonia from chest X-ray images.*, Global Transitions Proceedings 2.2 (2021): 375-381.

[20] Rebecca Sawyer Lee, Francisco Gimenez, Assaf Hoogi , Daniel Rubin, *Curated Breast Imaging Subset of DDSM [Dataset].*, The Cancer Imaging Archive. DOI: https://doi.org/10.7937/K9/TCIA.2016.7O02S9CY

[21] Oza, P.; Sharma, P.; Patel, S.; Adedoyin, F.; Bruno, A., *Image Augmentation Techniques for Mammogram Analysis.*, J. Imaging 2022, 8, 141. https://doi.org/10.3390/jimaging8050141

Edited by: Gabriele Mencagli
Received: Jan 31, 2022
Accepted: Aug 20, 2022