BREAST CANCER HISTOPATHOLOGICAL IMAGE CLASSIFICATION USING EFFICIENTNET ARCHITECTURE

Maheshvar Chandrasekar[a], Mukkesh Ganesh[b], Saleena B[c], Prakash Balasubramanian[d]
c.maheshvar2017@vitstudent.ac.in
g.mukkesh2017@vitstudent..ac.in
saleena.b@vit.ac.in
prakash.bala@vit.ac.in
a, b, c & d Vellore Institute of Technology, Chennai -600127, India

Abstract— Breast cancer is the most common type of cancer affecting women. The formation of lumps in the breast is one of the first signs of the presence of this disease. These tumors can either be cancerous or benign and hence a breast tissue biopsy is conducted to determine their nature. Advancements in the field of vision-based Deep Learning have facilitated the wide adoption of automated diagnostic systems in hospitals, for tasks such as cancer and COVID detection from lung X-ray scans, diabetic retinopathy detection from retinal fundus images, brain MRI segmentation, etc. Moving forward, reduction in training, validation and development times, and efficient usage of training resources for these models will be more in focus. The EfficientNet architecture proposed by Google has recently outperformed prior state-of-the-art architectures such as DenseNet and ResNet on the ImageNet classification task while using fewer parameters and epochs to converge faster. In this paper, we compare the performance of the EfficientNetB3 architecture with the above-mentioned architectures for the tasks of binary and multinomial tumor classification on the benchmark BreakHis dataset, which consists of around 8000 breast histopathology images of varying magnification. Our results show that under similar training conditions, the EfficientNetB3 can converge faster and outperform the previous benchmark models by a significant margin. Our best models achieved 100% sensitivity and accuracy on certain binary classification tasks and a sensitivity of 95.45% and precision of 95.15% on 8-ary classification tasks.

Keywords— Breast Cancer Detection; Artificial Neural Network; Deep Learning; EfficientNet; Transfer Learning;

I. INTRODUCTION

Breast cancer is cancer that arises in breast tissue. According to the WCRF survey in 2018, breast cancer is the most common cancer that affects women. Most women discover the presence of this disease when tangible lumps/tumors form in the breast. The tumors that develop in breasts can either be cancerous or benign in nature. To diagnose the nature of the tumors, tissue from the suspected area is collected through a process known as biopsy and is studied under a microscope and tested for the presence of cancerous cells. The survival rate for this disease varies with the severity stage at which it was diagnosed, and hence, faster results and examinations can greatly reduce the time between the initial checkup and the start of treatment.

Deep Learning based systems have shown tremendous results in the field of computer vision and have even surpassed human-level performance in certain niche tasks. With high sensitivities and specificities in medical image analysis tasks such as COVID detection from X-Rays, Brain MRI Segmentation, cancer detection from mammograms, diabetic retinopathy detection, etc. automated diagnostic systems are becoming widely adopted in the field of medicine. Newer architectures are being proposed by computer vision researchers, which are able to perform similar or even better at these tasks while making use of lesser or similar computing power. One such breakthrough architecture is the EfficientNet architecture, which outperforms previous state-of-the-art architectures such as DenseNet and ResNet, while having fewer parameters.

In this paper, we have explored the use of EfficientNetB3 model as an alternative to the above-mentioned architectures on the task of Breast Cancer Detection from Histopathological images, since it can greatly reduce training and inference time. We have made use of the BreakHis dataset (Breast Cancer Histopathological Database) since it provides us the flexibility to perform both binary and multinomial tumor classifications.

II. LITERATURE SURVEY

The authors of [1] and [2] have proven that the usage in Deep Learning (DL) in cancer studies are feasible and also efficient. In fact, from the study conducted with respect to detecting Lymph Node Metastases, 7 DL models showed better discrimination than a panel of 11 pathologists in a time-constrained setting [1]. In [3], a method of using Transfer Learning on a pre-trained Inception Resnet V2 is explored. It works in a mechanism of firstly training the dataset with the network initialized with the ImageNet weights and then further fine-tuning it by a smaller learning rate to obtain a best accuracy of 90%. The usage of Deep Convolutional Neural Networks such as ResNet-50, InceptionV3 and VGG-16 networks for feature extraction from images was done in [4]. The features were then trained using Gradient Boosting and obtained a classification accuracy of 93.8%. The authors in [5] used improved Resnet and Inception networks to obtain accuracies as high as 99.6% using the BreakHis dataset for the binary classification. In [6], in addition to Resnet, transfer learning was applied to GoogLeNet and Alexnet. The maximum accuracy of 86% was however obtained by the Resnet model. The authors in [7] have used a Generative Adversarial Network to augment the dataset with more data to improve the performance of the model. They then applied a CNN to classify the dataset and obtained a sensitivity level of 93.6%, with 90.8% specificity. There are also Machine Learning approaches used by authors such as the ones in [8]. ML Algorithms such as Logistic Regression, Random Forest, Support Vector Classifier (SVC), AdaBoost Classifier, Bagging classifier, voting classifier, and a DL Xception model to classify the BreakHis dataset. A maximum F1 score of 0.90 was obtained by the Xception model. Convolution

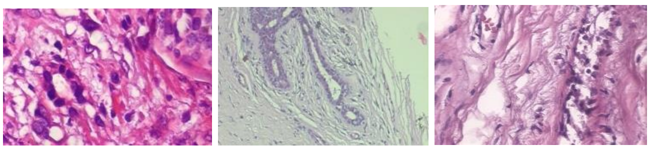
Neural Networks have reached new milestones since the development of Efficient Net Architecture. According to [9], if the width, depth and resolution of a Deep Convolution network were increased by scaling them with a constant, it improves the performance of the network. The usage of the algorithm has been proven successful in many places. For instance, it was used in a study [10] to detect COVID-19 infected patients and obtained a state of the art F1 Score of 97.11%, and 99.62%.

III. Dataset

Breast Cancer Histopathological Database (BreakHis) [11] consists of around 8000 microscopic images of breast tissue biopsy slides. These samples were collected from 82 patients, and the generated images are of 4 different orders of magnification, namely 40x, 100x, 200x and 400x. Apart from containing just the binary malignant and benign labels, the images are also sorted into 8 different breast tumor classes and can hence also be used for multilabel classification. The benign classes are Adenosis (A), Fibroadenoma (F), Phyllodes Tumor (PT) and Tubular Adenoma (TA), while the Malignant tumor classes are Ductal Carcinoma (DC), Lobular Carcinoma (LC), Mucinous Carcinoma (MC) and Papillary Carcinoma (PC). These samples are RGB images with a resolution of 700x460 pixels.

Table 1: Class-wise distribution of images

Magnification	Benign	Malignant	Total
40X	625	1370	1995
100X	644	1437	2081
200X	623	1390	2013
400X	588	1232	1820
Total Images	2480	5429	7909

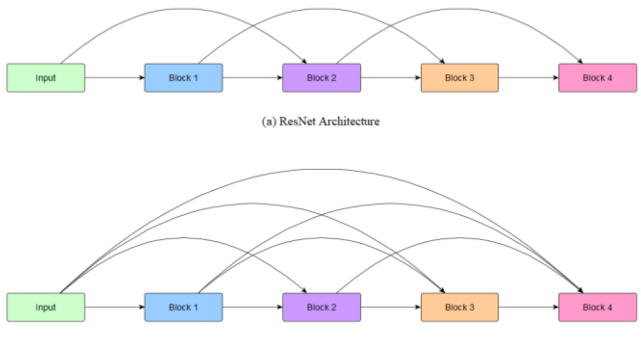


(a) 40x DC *(b) 100x PT* *(c) 400x PC*

Figure 1: *Microscopic breast tissue images of varying magnification*

IV. Prediction Model

The usage of Convolutional Neural Networks in Computer Vision has been popular ever since 2012 when AlexNet obtained the then state-of-the-art accuracy on the ImageNet challenge. However, the applications of such deep networks were not without obstacles. There were a few problems related to computational power which were solved by the advancements in GPU and CPU technologies which started to handle many more instructions per second. There were also a few theoretical problems that stalled the development of deeper networks. For instance, the Vanishing Gradient problem was one such limitation that inhibited the convergence of models. This was overcome by the introduction of the ResNet architecture which has residual/skip connections between every block for the gradient to flowthrough. An advancement over this particular architecture is the DenseNet architecture. Each block has residual connections with all blocks that came prior to it. By doing so, the collective knowledge of the neural network until the ith block is present for the (i+1)th block to work with.



(b) DenseNet Architecture

Figure 2: *ResNet vs DenseNet Architectures*

Another such problem which had challenged scientists was the expansion of architectures. Networks are firstly developed with limited resources to obtain good results and are then expanded to improve the performance. The major drawback faced by researchers was the absence of an empirical method to scale up this network. Networks can be expanded across their depth, width, or image resolution. Initially, the scaling-up process was not much understood and required tedious tuning and often did not result in substantial improvements. And even when they did, the increase in accuracy was exponentially less when compared to the increase in FLOPS (Floating Point Operations per Second). The recently introduced EfficientNet architecture addresses this particular issue. It scales up the network in a principal manner to avoid diminishing returns. The invention of this network is a huge breakthrough in Computer Vision as it has now achieved the highest accuracy of 84.3% in the ImageNet dataset while being 8.4 times smaller, and converging 6.1 times faster than the previous best network.

The major finding from [9] was to establish a relationship between the depth, width and resolution of a network and to increase it uniformly. This is called the Compound Scaling Method. The method states that the parameters of the network should be assigned as follows:

$$depth: d = \alpha^{\phi} \tag{1}$$
$$width: w = \beta^{\phi} \tag{2}$$
$$resolution: r = \gamma^{\phi} \tag{3}$$
$$st\ \alpha.\beta^2.\gamma^2 \approx 2 \tag{4}$$
$$\alpha, \beta, \gamma \geq 1 \tag{5}$$

where α, β, γ are constants and can be found using a grid search and Ø is a user defined coefficient that relates to the computing resources available. This is also intuitive since scaling up the width of the network increases the number of parameters (directly proportional to FLOPS) at a higher rate than an increase in depth would. Similarly, the authors hypothesized that an increase in resolution should also see an increase in width, since more spatial data is available to the network.

Different Ø values produce different EfficientNet architectures. In this paper, EfficientNetB3 architecture was used. It consists of 5 different models and its main building block is a Mobile Inverted Bottleneck Convolution (MBConv), an Inverted Residual Block. The architecture diagram for the network segregated by the models is shown in fig. 3.

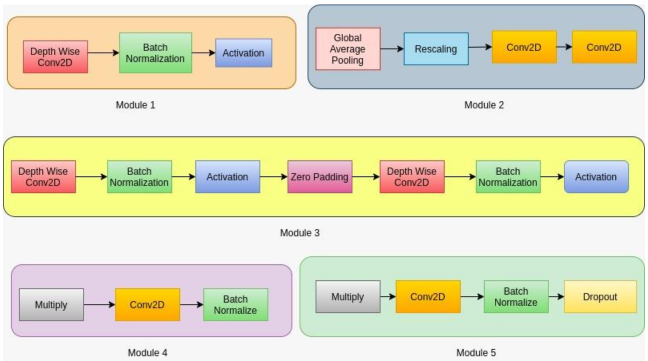


Figure 3: *Modules in EfficientNet Architecture*

V. TRAINING PROCEDURE

A. *Dataset split and Augmentation:*

The dataset was split in a ratio of 8:1:1 for training, validation and testing respectively. This was done for both binary and multinomial classification tasks. Since we have a limited number of images in each class, we used TensorFlow's Image Data Generator to augment our dataset. While creating batches for training and validation, augmentation techniques such as shearing, zooming mirroring and channel shifts were applied randomly to the data to diversify the batches. The images were also resized to (230,350) to preserve the spatial ratio of the rectangular images.

B. *Model Architecture:*

Since there are 4 different magnifications of images available for 2 classification tasks, 8 separate models were created to correspond to these different tasks and architectures. EfficientNetB3, DenseNet169 and ResNet101V2 each had 8 different models trained on data which was split class-wise and magnification-wise. The pretrained models with weights initialized to the ImageNet challenge, were loaded into the environment using Tensorflow.

Due to the images being resized to a non-standard resolution, we proceeded to use a Global Average Pooling 2D layer to restrict the volume of the output to that of the number of channels in the neural network. A Dense layer of 256 neurons with ReLU activation was added, followed by a Dropout layer with 70% keep probability. The output layers of these models were decided based on the classification task at hand (2-ary or 8-ary). Softmax function was used as the activator for the last layer.

Table 2: Model parameters

Model Architecture (imagenet)	No. of Parameters
ResNet101V22	44,675,560
DenseNet169	14,307,880
EfficientNetB3	12,320,535

C. *Loss function and Optimizer:*

The Categorical Cross Entropy was the loss function used in the multiclass classification problem. This function penalizes every wrongly classified output from the model using (6).

$$Loss = -1 . \sum_{i=1}^{output\ size} -y_i . log\ \hat{y}_i \qquad (6)$$

where y_i is the target value of the i^{th} data point in the dataset and $\hat{y}_i$ is the predicted value for the same data point from the model. This loss value is calculated for the batch of data passed through the network in each iteration.

The Nadam Optimizer [12] was used for making the model converge better. Nadam is a combination of Nesterov accelerated gradient (NAG) descent and Adam, which is a combination of Adagrad and RMSProp. NAG works on a 'look before you leap' basis. Every iteration step is measured to see whether the gradient decreases before execution in this method. This gives the models a chance to converge faster.

D. *Standardising Resources:*

All the models were trained for 50 epochs and the weights of the model which produced the highest validation accuracy were saved to be used for evaluation on the test set. By standardizing the number of epochs and keeping the weights initialized to the same ImageNet challenge without any further alterations, such as freezing intermediate layers, we can level the playing ground for the models to use a similar amount of computing resources while trying to converge.

VI. RESULTS AND DISCUSSION

The 2-ary and 8-ary models were tested on 10% of the remaining data after being left to run for 50 epochs. If the best validation accuracy was achieved after the 45^{th} epoch, an extra buffer of 10 epochs were provided for the model to reduce the validation error. However, none of the models' validation accuracies increased beyond the 45^{th} epoch, even when let to run for a total of 60 epochs. Hence, the best weights for these models were achieved during the initial 50 epoch runtime. Evaluation metrics for the 2-ary and 8-ary results are tabulated below.

Table 3: Binary Classification results

Zoom	Model	Prec.	Sensit.	F1	Acc.
40x	EffNB3	100.0	100.0	100.0	100.0
	Res101V2	92.86	92.68	92.50	92.68
	Dense169	93.43	93.17	93.24	93.17
100x	EffNB3	100.0	100.0	100.0	100.0
	Res101V2	94.26	94.29	94.26	94.29
	Dense169	91.35	91.43	91.35	91.43
200x	EffNB3	98.06	98.02	98.03	98.02
	Res101V2	89.56	89.60	89.37	89.60
	Dense169	93.60	93.56	93.58	93.56
400x	EffNB3	97.37	97.28	97.30	97.28
	Res101V2	86.81	86.96	84.60	86.96

	Dense169	95.53	95.11	95.17	95.11

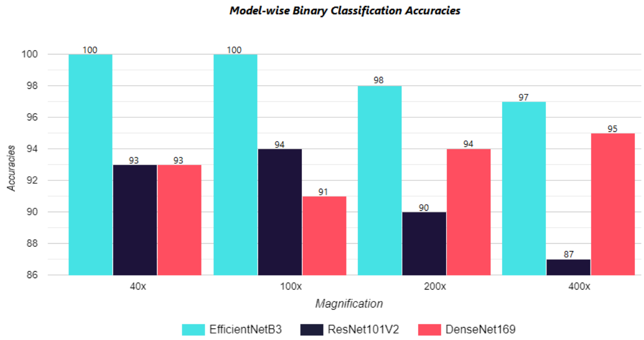


Figure 4: *Comparison of 2-ary classification accuracies*

Table 4: 8-ary Classification results

Zoom	Model	Prec.	Sensit.	F1	Acc.
40x	EffNB3	95.46	95.15	95.20	95.15
	Res101V2	73.23	68.93	65.75	68.93
	Dense169	85.69	84.47	83.97	84.47
100x	EffNB3	94.39	94.34	94.24	94.34
	Res101V2	68.38	68.87	66.71	68.87
	Dense169	85.82	82.55	83.10	82.55
200x	EffNB3	91.86	91.83	91.76	91.83
	Res101V2	69.82	69.71	66.81	69.71
	Dense169	79.04	76.92	76.68	76.92
400x	EffNB3	88.04	87.89	87.38	87.89
	Res101V2	70.04	65.79	64.27	65.79
	Dense169	83.28	80.53	81.18	80.53

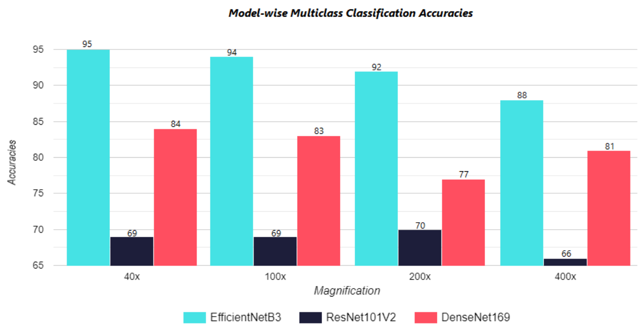


Figure 5: *Comparison of 8-ary classification accuracies*

From table 3, we can infer that EfficientNetB3 provides perfect or near-perfect binary classification accuracies across all the magnifications. High sensitivities and precision are also reported for this model. The DenseNet binary classification model converged similarly to the EfficientNetB3 model when trained on 400x magnification images. While ResNet performed relatively better when using lower magnification images, its drop in performance in noticeable while using 400x images.

For the task of 8-ary classification, the EfficientNetB3 model was consistently reaching precision and sensitivity values of over 90% on a per-class basis, on 40x, 100x and 200x images, as shown in table 4. Though there is a loss in performance when trained on 400x images, it can hold the performance lead over the other two models by a significant margin. This is shown in fig. 5. While ResNet and DenseNet struggle with the prediction of LC and PT tumors with accuracies around 60% and 70% for the models respectively, EfficientNet is able to achieve high accuracies such as 95% and 100% on the same when trained on 40x images.

Almost all EfficientNetB3 models converged within the first 25 epochs, barring the 400x 8-ary classification model, while the DenseNet and ResNet models' best validation accuracies were generally achieved during the 40-45th epoch. DenseNet 8-ary 200x and ResNet 8-ary 100x were the two models that converged only after the 45th epoch, and as alluded to before, didn't achieve a better score when run till the 60th epoch.

VII. Conclusion and Future Enhancement

Using the BreakHis dataset, binary and multi-class classification of tumors using various neural network architectures were explored. When DenseNet169, Resnet101V2 and EfficientnetB3 were trained using the pretrained weights of the ImageNet problem for 50 epochs, the EfficientnetB3 model produced outstanding results of up to 100% accuracy in 40x and 100x magnified images. When the models were trained to classify the type of tumor present in the biopsy slides which is a multiclass classification problem, EfficientnetB3 consistently outperformed the other two networks by a significant margin. The best EfficientNet model was able to achieve ~95% sensitivity and precision on both 40x and 100x images. These results are roughly 10% higher than the other two architectures. These EfficientNet models also converged at a faster rate when compared to the others.

Further developments can be made by using a Generative Adversarial Model (GAN) to augment the dataset with more samples of images to improve performance. It was found that the "DC" class in the 8-way classification contained a lot more samples than the other classes. Hence, an undersampling method such as Near Miss Undersampling can be used to achieve a balanced dataset.

VIII. References

[1] Bejnordi, Babak Ehteshami, et al. "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer." Jama 318.22 (2017): 2199-2210.

[2] Zhou, Li-Qiang, et al. "Lymph node metastasis prediction from primary breast cancer US images using deep learning." Radiology 294.1 (2020): 19-28.

[3] Ferreira, C. A., Melo, T., Sousa, P., Meyer, M. I., Shakibapour, E., Costa, P., & Campilho, A. (2018, June). Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2. In

International Conference Image Analysis and Recognition (pp. 763-770). Springer, Cham.

[4] Rakhlin, A., Shvets, A., Iglovikov, V., & Kalinin, A. A. (2018, June). Deep convolutional neural networks for breast cancer histology image analysis. In International Conference Image Analysis and Recognition (pp. 737-744). Springer, Cham.

[5] Motlagh, M. H., Jannesari, M., Aboulkheyr, H., Khosravi, P., Elemento, O., Totonchi, M., & Hajirasouliha, I. (2018). Breast cancer histopathological image classification: A deep learning approach. BioRxiv, 242818.

[6] Ahmad, H. M., Ghuffar, S., & Khurshid, K. (2019, January). Classification of breast cancer histology images using transfer learning. In 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 328-332). IEEE.

[7] Kansal, S., Goel, S., Bhattacharya, J., & Srivastava, V. (2020). Generative adversarial network–convolution neural network based breast cancer classification using optical coherence tomographic images. Laser Physics, 30(11), 115601.

[8] Chand, S. (2020). A comparative study of breast cancer tumor classification by classical machine learning methods and deep learning method. Machine Vision and Applications, 31(6), 1-10.

[9] Tan, M., & Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946.

[10] Marques, G., Agarwal, D., & de la Torre Díez, I. (2020). Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. Applied Soft Computing, 96, 106691.

[11] Spanhol, F., Oliveira, L. S., Petitjean, C., Heutte, L., A Dataset for Breast Cancer Histopathological Image Classification, IEEE Transactions on Biomedical Engineering (TBME), 63(7):1455-1462, 2016.

[12] Dozat, Timothy. "Incorporating nesterov momentum into adam." (2016). Accessed on: Oct. 11, 2020. [Online]. Available:
http://cs229.stanford.edu/proj2015/054_report.pdf

[13] “Breast Cancer Statistics”, 2018. Accessed on: Oct. 14, 2020. [Online]. Available:
https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics