Analysis of Deep Learning Models to Detect Breast Cancer from Histopathology Images

Anjali Gautam[1] and Satish Kumar Singh[1]

Abstract— **Breast cancer is a disease where breast cells grow out of control and leads to cancer. Various methodologies have been developed to identify breast cancer. In this paper, we have developed an approach to classify breast cancer from histopathology images. The approach makes use of deep learning based architectures by setting same parameters for all while training and testing on them. Thereafter, all the architectures are compared to see which one is most suited for the classification of breast cancer. Previous works on AlexNet, VGG, ResNet have already been published, and here we have tried to see the performance of those models which have less number of trainable parameters, namely DenseNet121, DenseNet169, DenseNet201, EfficientNetB0, EfficientNetB5, EfficientNetV2B0 and EfficientNetV2S. Here, all the experiments are conducted on BreakHis histopathology dataset by utilizing all the images of resolutions 40X, 100X, 200X and 400X of benign and malignant cancer.**

I. Introduction

Breast cancer commonly found in the women and according to Global Cancer Observatory (GLOBOCAN) 2021 report, 19.3 million incident cancer cases were found. In India 2022, the breast cancer cases including men and women were 221,757 where a count of males were 5,649 and females were 216,108 according to [1]. Males were at a risk of 1 in 1021 and females were 1 in 29. According to Cancer Statistics 2022, in the United States, estimated new cases were 290,560, where 2,710 were of male and 287,850 were of female. The number of deaths estimated by them were 43,780 including both sexes where 530 were for male and 43,250 for female. In 2025, the predicted number of breast cancer cases would be 2,32,832 among males and females [1].

The commonly found reason of mortality among women is breast cancer. It may be caused due to inherited changes in genes like BRCA1 and BRCA2, however, there is no clear reason for its occurrence. This cancer can increase by smoking, obesity, lack of physical activity, alcohol, or other medical conditions. Clinicians found that its early detection can increase the rate of survival and therefore its accurate prediction is required. It can be diagnosed with various breast imaging modalities like mammography, magnetic resonance imaging (MRI), dynamic contrast-enhanced MRI (DCE MRI), ultrasound, nuclear, optical, microwave imaging and many more. Initially, apart from physical examination, mammography is used for detection than other imaging methods and if required doctors may prefer other modalities as well. If any sign of cancer is found, then doctors go for tissue biopsy.

[1]Anjali Gautam and Satish Kumar Singh are with Computer Vision and Biometrics Lab (CVBL), Department of Information Technology, Indian Institute of Information Technology Allahabad, Prayagraj, INDIA anjaligautam@iiita.ac.in, sk.singh@iiita.ac.in

Many machine learning (ML) and deep learning (DL) algorithms exist for disease prediction, however, due to false positives those systems are not in trend as they may cause worry in patients even if they don't have cancer. In this work, we have tried to predict the best suited and efficient DL based model for breast cancer detection from histopathology images. The image dataset which we have chosen for the experiment is Breast Cancer Histopathological Image Classification (BreakHis) and is available online for research work. The experiments have been conducted on the same setting for all the models so that the best model can be found. The models used for experiments are DenseNet121, DenseNet169, DenseNet201, EfficientNetB0, EfficientNetB5, EfficientNetV2B0 and EfficientNetV2S [9], [8], [11]. Here, we have not included any architecture such as LeNet, AlexNet [5], VGG [6] and ResNet [7] and their versions, as lots of work have already been done using them. Fig. 1 depicts the sample images of breast cancer tissues in benign and malignant cases.

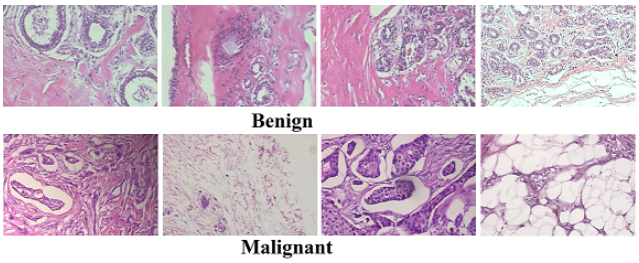


Fig. 1: Some sample images of benign and malignant breast cancer from BreakHis dataset.

This paper is written in 5 different sections: II section discusses about recent related works, III section gives the description of adopted methodology, fourth section discusses the results obtained after various experiments and finally the last section gives the overall conclusion of the paper.

II. Literature Review

In this section, previous works on breast cancer classification from medical images have been discussed.

Earlier computer-aided detection or diagnosis (CAD) systems designed for breast cancer detection, where pixel-based approaches were used to extract features of each pixel and classified images as normal or abnormal [2], [3], [4]. The classifiers used by them for classification tasks were binary decision tree, SVM, LDA, Bayesian, random forest, neural networks classifiers, etc.

In almost many works, general step which they followed for cancer detection were image preprocessing, image seg-

979-8-3503-0219-6/23/$31.00 ©2023 IEEE

mentation, features extraction and selection, and classification. After the popularity of deep learning based architectures like AlexNet [5], VGG [6], ResNet [7], DenseNet [8] and many others, they have started being used in developing CAD systems for health centers.

Now some recent works on classifying breast cancer into different categories is briefly discussed. Some papers focused on extracting discriminant feature from images so that cancer images can be categorized easily [10], [17], [13]. In [10], image preprocessing and enhancement techniques were used before deep learning, where image resizing, data augmentation, CLAHE enhancement and histogram matching were used. Thereafter, Resnet50 pre-trained model was used for the classification task. In [17], phylogenetic diversity indexes as the feature extractor had been used to characterize images in 4 categories invasive carcinoma, in situ carcinoma, normal, and benign lesion. This has been followed by classification using multilayer perceptron, XGBoost, random forest, and SVM. Chattopadhyay et al. developed multi-scale dual residual recurrent network (MTRRE-Net) to detect breast cancer [13]. It was based on using multi-scale feature fusion where two-fold residual recurrent operation was utilized for overcoming vanishing gradient problem.

The works on ensembling various CNN architectures has been done by many researchers [21], [16], [18]. Majumdar et al. [21] in their work developed the ensembled method where they combined the output of three well known architectures VGG11, GoogleNet and MobileNetV3_Small. The combined results obtained by getting the confidence scores of those three models using Gamma function and helped in getting the ensemble to get final prediction. EMS-Net: ensemble of multiscale CNNs had been developed by Yang et al. [16] where each image firstly converted to multiple scales, and then further utilized the cropped and augmented training patches at each scale to fine-tune DenseNet-161, ResNet-152, and ResNet-101 models. Thereafter, they used the models as the ensemble model. Senousy et al. [18] developed multi-level context and uncertainty aware (MCUa) dynamic deep learning ensemble model. As the name indicates this model extract multi-level context information from several patches of multiple image scales for learning spatial dependency in image patches. They have used ResNet-152 and DenseNet-161 as the pre-trained backbone architecture for extracting features, and then those features were sent to multi-level context-aware models for final prediction. Bhowal et al. [19] on the other hand, fused many CNN architectures VGG16, VGG19, Xception, Inception V3, and InceptionResnet V2 to classify images by using a fuzzy ensemble method which made use of Choquet integral, Coalition game theory, and Information theory.

In Li et al.'s work, DenseNet was used as the backbone architecture which was interleaved with the squeeze-and-excitation (SENet) module [14]. Feng et al. [15] proposed deep manifold preserving autoencoder as the new feature extractor. Thereafter, it was integrated with a softmax classifier for classification. Hameed et al. [22] developed a model which exploited six intermediate layers of the Xception (Extreme Inception) network. Therefore, for each images six different features were taken out then performed global average pooling, which were then concatenated by forming a single vector for the classification task. They have used 5-fold cross-validation approach to optimize their model.

In the recent works, still already existing CNN architectures are being used. In the method given by Taheri and Golrizkhatami [23], they developed two different models where the first one was based on pre-trained DenseNet201 architecture and fine-tuned the dataset for specific magnification factor to get classification results. The second model contains four submodels of DenseNet201 for each magnification factor where the obtained results from each model were then fused to get final prediction. Atban et al. [25] used ResNet18 as the feature extractor followed by classification using Atom Search Optimization (ASO), Particle Swarm Optimization (PSO) and Equilibrium Optimizer (EO) algorithms. In [26], VGG and ResNet were used in two teacher models for training two student models having less number of layers then teacher models.

Some literature review work on breast cancer detection can be also be found from [24], [12].

III. Methodology

A. Dataset

In this paper, we have worked on BreakHis dataset [1] [20]. The dataset contains total 9,109 histopathology microscopic images of 82 patients suffering from breast cancer. However, there are 2,480 images in benign class and 5,429 images in malignant class, which are of four different magnification factors i.e. 40X, 100X, 200X and 400X. This total number comes out to 7,909. All the images are of 700×460 pixels in RGB and are of *.png* format.

B. Cancer Classification Approach

In this section, the methodology adopted to analyze breast cancer classification performance by various versions of DenseNet and EfficientNet architectures using same experimental settings.

The methodology starts with downloading the BreakHis dataset, and then it is split into train, validation and test sets of 7:1:2. Thereafter, we have chosen some models of DenseNet and EfficientNet for training on 70% dataset and validation on 10% of the dataset. Here, we have used the standard models from Keras Applications and have not made any changes in any parameters except hyper-parameters. Here, no pre-trained models with ImageNet weights have been used, as histopathological images are different from the ImageNet images. Therefore, we have trained, validated and tested all the models on the BreakHis dataset. The methodology is depicted in Fig. 2 and its algorithm is given Algorithm 1.

The dataset which we have used is briefly described in III-A. For the experiments, we have used total 5536 images in training set, 1583 in testing set and 790 images are present in validation set. Every set contains two classes benign and malignant of all the four resolutions 40X, 100X, 200X and 400X. Please see first part of Fig. 2.

[1] https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-BreakHis/

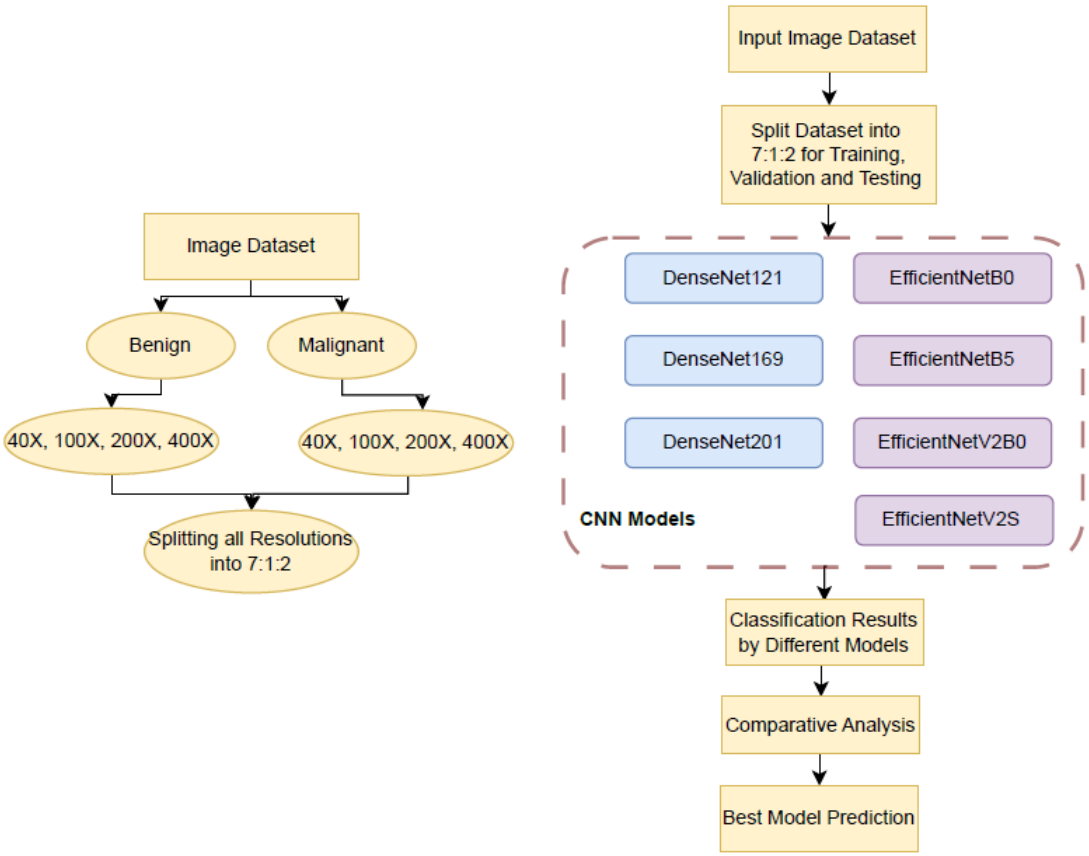


Fig. 2: Methodology for the Analysis of Deep Learning Models for BreakHis dataset.

Algorithm 1: Classification of histopathology breast cancer images

Data: Histopathology image dataset
Result: Classification accuracy in terms of accuracy, precision, recall

```
Batch Size=64, Image Size= 224 × 224 × 3;
Steps per Epoch=87, Epochs=100;
validation_steps=13,validation_freq=1;
Split dataset into training, validation and testing sets;
ImageDataGenerator() used for creating batches of training,
  testing and validation images;
while For_Each_Model do
    Model training and validation;
    Evaluating Model Accuracy on Testing Data;
end
Comparing each model on test results;
Model with best evaluation metrics
```

The main motive to perform this analysis task is that, many papers have been published on using various CNN models, however, the comparative analysis between different versions of same models is missing for BreakHis dataset. Therefore, we have tried to find out which model is best suited for this dataset. The second motive is to select only those models whose total training parameters are less in numbers and thus in this paper we have not considered all the versions of DenseNet and EfficientNet, the details about those models can be found in their respective papers. The number of parameters for the models which we have considered is shown for ImageNet and BreakHis datasets in Table I, the reason to show for ImageNet is to remove any confusion why parameters are less for BreakHis as compared to ImageNet, which is due to 1000 classes in ImageNet and only 2 classes (benign and malignant) for BreakHis.

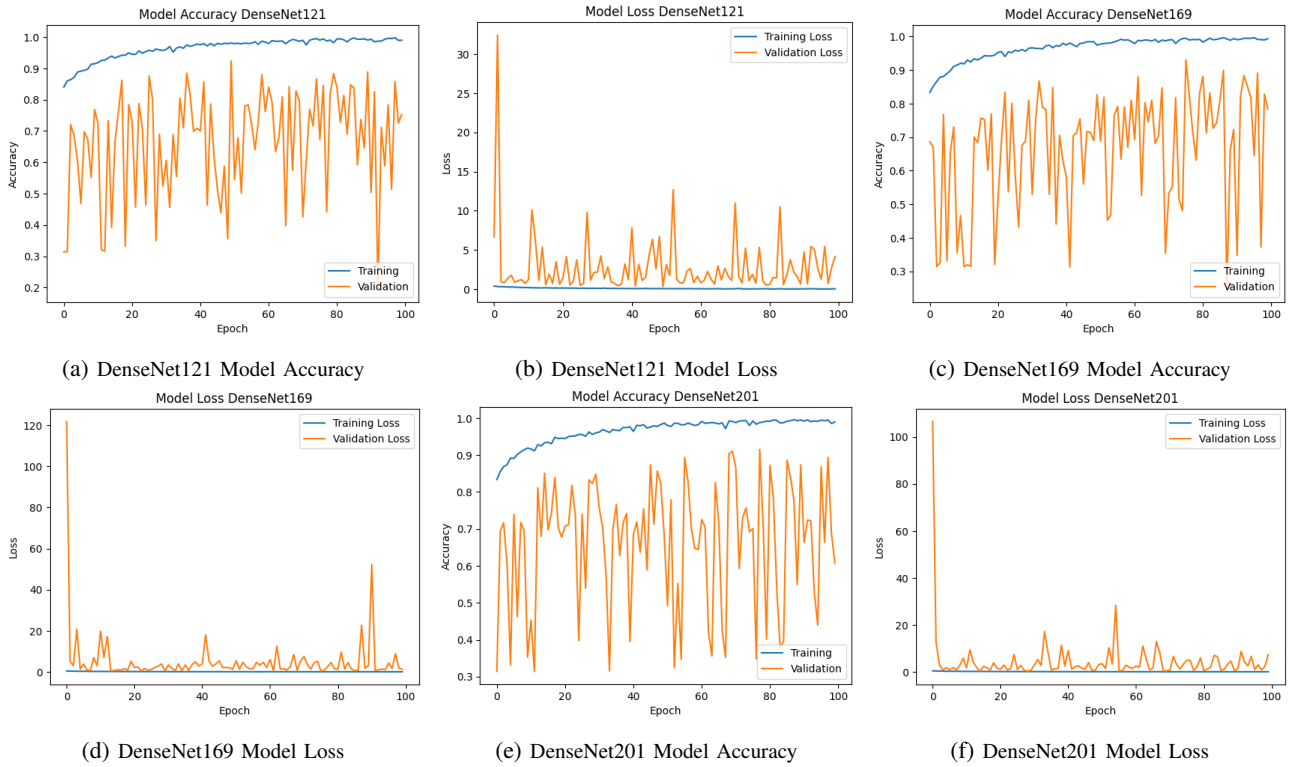
IV. Results

This section discussed about the achieved experimental results by using DenseNet and EfficientNet models on BreakHis dataset with same hyper-parameters.

All the experiments have been conducted using NVIDIA GeForce RTX 3090 with CUDA Version 11.7, Tensorflow 2.7.0 and Python 3.9.16. The hyper-parameters used for all the experiments in training, validation and testing are batch size of 64, image size 224×224, steps per epoch=87, epochs=100, validation steps=13 and validation freq=1. Total seven experimental results are mentioned here which were conducted for seven different CNN models

TABLE I: Total and Trainable Parameters of CNN Models on ImageNet and BreakHis Dataset.

Model	ImageNet (Total/Trainable Parameters)	BreakHis (Total/Trainable Parameters)
DenseNet121	8,062,504/7,978,856	7,039,554/6,955,906
DenseNet169	14,307,880/14,149,480	12,646,210/12,487,810
DenseNet201	20,242,984/20,013,928	18,325,826/18,096,770
EfficientNetB0	5,330,571/5,288,548	4,052,133/4,010,110
EfficientNetB5	30,562,527/30,389,784	28,517,625/28,344,882
EfficientNetV2B0	7,200,312/7,139,704	5,921,874/5,861,266
EfficientNetV2S	21,612,360/21,458,488	20,333,922/20,180,050



(a) DenseNet121 Model Accuracy (b) DenseNet121 Model Loss (c) DenseNet169 Model Accuracy

(d) DenseNet169 Model Loss (e) DenseNet201 Model Accuracy (f) DenseNet201 Model Loss

Fig. 3: Models accuracy and loss graphs of DenseNet121, DenseNet169 and DenseNet201 while training and validation on BreakHis dataset.

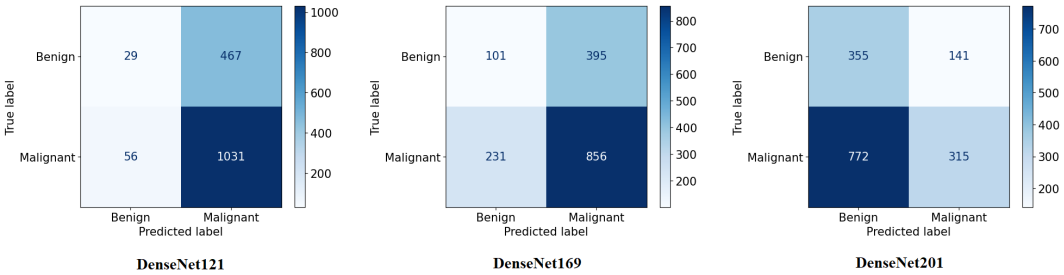


Fig. 4: Confusion matrices of DenseNet models on testing set.

DenseNet121, DenseNet169, DenseNet201, EfficientNetB0, EfficientNetB5, EfficientNetV2B0 and EfficientNetV2S.

Table I shows the number of total and trainable parameters of CNN models on ImageNet dataset on which the models are pre-trained. It also shows the number of parameters for BreakHis dataset on which we have trained the models. The lowest and highest number of parameters is for EfficientNetB0 and EfficientNetB5 respectively.

All the models trained and validated on 70% and 10% of the dataset respectively. Fig. 3 and 5, shows the accuracy and loss graph of all the models while their training and validation. Fig. 4 and 6 show the confusion matrices of the testing dataset on the trained models, which help to identify the number of correctly and wrongly classified images. From the figures, we can see that models are not able to perform well on validation and testing datasets, both the accuracy and loss values are very less and high respectively. Table II, shows the values of evaluation metrics namely accuracy, weighted average precision and weighted average recall on testing dataset. It shows that overall DenseNet121 achieved highest accuracy of 66.96% followed by EfficientNetB5 with 66.83% accuracy. From these two we can say that DenseNet121 is better as it has only 6,955,906 numbers of trainable parameters whereas EfficientNetB5 has 28,344,882 numbers of trainable parameters which are very huge.

Hence, from these experiments we can conclude that DenseNet121 model is best suited for breast cancer prediction from histopathology images and if implemented with improved methodology then it may be able to give satisfactory results.

V. Conclusion

This paper is about analyzing the performance of various models of DenseNet and some models of EfficientNet on breast cancer BreakHis dataset. The task of this paper is to classify histopathology images of various resolution into benign and malignant. Both the cancer categories contains images of all resolutions in training, validation and testing sets, and no separate resolution-wise experiments conducted for the classification task. Here, 3 models of DenseNet

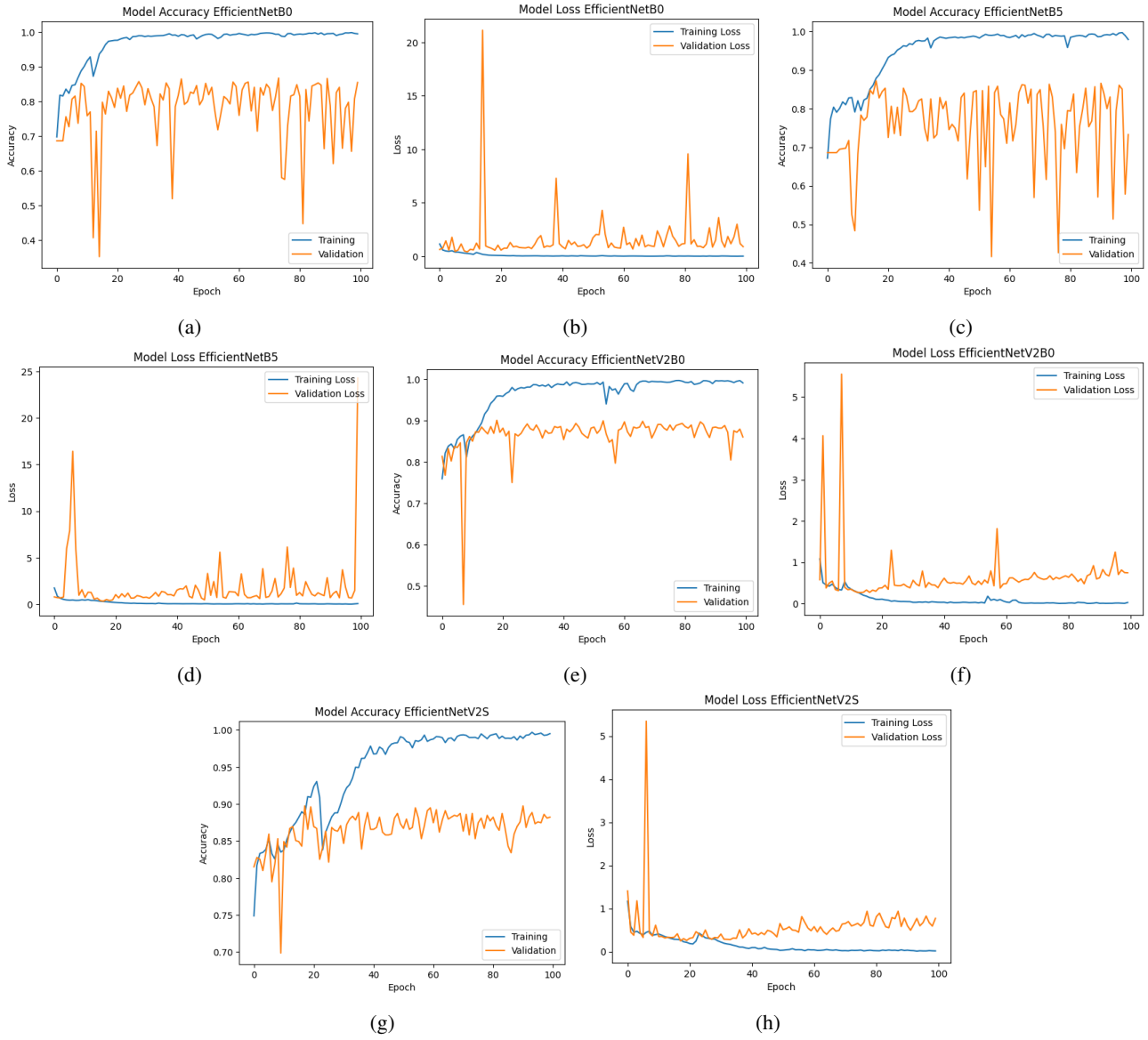


Fig. 5: Models accuracy and loss graphs of EfficientNetB0, EfficientNetB5, EfficientNetV2B0 and EfficientNetV2S while training and validation on BreakHis dataset.

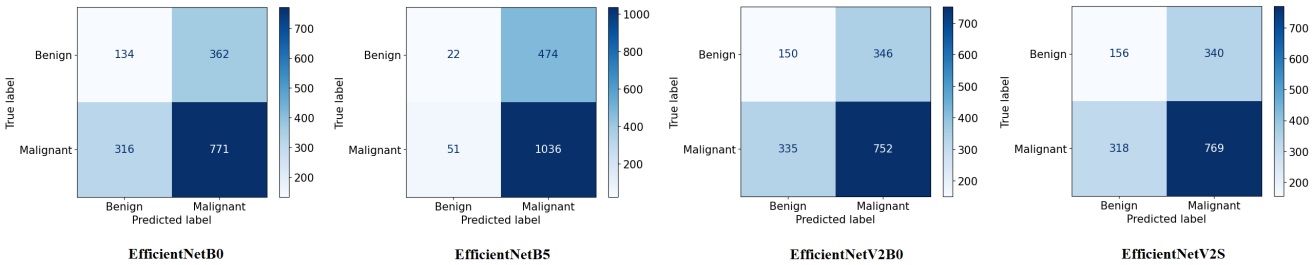


Fig. 6: Confusion matrices of EfficientNet models on testing set.

TABLE II: Evaluation metrics obtained on testing set of the dataset (in %).

Model	Accuracy	Precision	Recall
DenseNet121	66.96	57.95	66.96
DenseNet169	60.45	56.52	60.45
DenseNet201	42.32	58.91	42.32
EfficientNetB0	57.17	56.06	57.17
EfficientNetB5	66.83	56.55	66.83
EfficientNetV2B0	56.98	56.72	56.98
EfficientNetV2S	58.43	57.93	58.43

i.e. DenseNet121, DenseNet169 and DenseNet201, and 4 models of EfficientNet namely EfficientNetB0, EfficientNetB5, EfficientNetV2B0 and EfficientNetV2S have been considered for the analysis. All these selected models have less number of parameters as compared to other EfficientNet models. Here, we have tried to analyze the effect of model size on the classification performance and we saw that model of less size was able to perform well, however, models of smaller size EfficientNetB0 and EfficientNetV2B0 could'nt performed well. Overall, DenseNet121 gave the best performance on testing set followed by EfficientNetB5.

ACKNOWLEDGMENT

This paper is supported under the seed money research grant (F. No. IIITA/RO/2022/407) by Indian Institute of Information Technology Allahabad. The Central Computing Facility (CCF) of IIIT Allahabad has been used to conduct all experiments.

REFERENCES

[1] Sathishkumar, Krishnan, Meesha Chaturvedi, Priyanka Das, S. Stephen, and Prashant Mathur. "Cancer incidence estimates for 2022 & projection for 2025: Result from national cancer Registry Programme, India." Indian Journal of Medical Research (2023).

[2] Sampat, Mehul P., Mia K. Markey, and Alan C. Bovik. "Computer-aided detection and diagnosis in mammography." Handbook of image and video processing 2, no. 1 (2005): 1195-1217.

[3] Kegelmeyer Jr, W. Philip, Joe M. Pruneda, Philip D. Bourland, Argye Hillis, Mark W. Riggs, and Michael L. Nipper. "Computer-aided mammographic screening for spiculated lesions." Radiology 191, no. 2 (1994): 331-337.

[4] Tang, Jinshan, Rangaraj M. Rangayyan, Jun Xu, Issam El Naqa, and Yongyi Yang. "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances." IEEE transactions on information technology in biomedicine 13, no. 2 (2009): 236-251.

[5] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6), pp.84-90.

[6] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[7] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[8] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[9] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[10] Zewdie, E.T., Tessema, A.W. and Simegn, G.L., 2021. Classification of breast cancer types, sub-types and grade from histopathological images using deep learning technique. Health and Technology, 11, pp.1277-1290.

[11] Tan, M. and Le, Q., 2019, May. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

[12] Gautam, A., 2023. Recent advancements of deep learning in detecting breast cancer: a survey. Multimedia Systems, 29, pp.917-943.

[13] Chattopadhyay, S., Dey, A., Singh, P.K., Oliva, D., Cuevas, E. and Sarkar, R., 2022. MTRRE-Net: A deep learning model for detection of breast cancer from histopathological images. Computers in Biology and Medicine, 150, p.106155.

[14] Li, X., Shen, X., Zhou, Y., Wang, X. and Li, T.Q., 2020. Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). PloS one, 15(5), p.e0232127.

[15] Feng, Y., Zhang, L. and Mo, J., 2018. Deep manifold preserving autoencoder for classifying breast cancer histopathological images. IEEE/ACM transactions on computational biology and bioinformatics, 17(1), pp.91-101.

[16] Yang, Z., Ran, L., Zhang, S., Xia, Y. and Zhang, Y., 2019. EMS-Net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images. Neurocomputing, 366, pp.46-53.

[17] Carvalho, E.D., Antonio Filho, O.C., Silva, R.R., Araujo, F.H., Diniz, J.O., Silva, A.C., Paiva, A.C. and Gattass, M., 2020. Breast cancer diagnosis from histopathological images using textural features and CBIR. Artificial intelligence in medicine, 105, p.101845.

[18] Senousy, Z., Abdelsamea, M.M., Gaber, M.M., Abdar, M., Acharya, U.R., Khosravi, A. and Nahavandi, S., 2021. MCUa: Multi-level context and uncertainty aware dynamic deep ensemble for breast cancer histology image classification. IEEE Transactions on Biomedical Engineering, 69(2), pp.818-829.

[19] Bhowal, P., Sen, S., Velasquez, J.D. and Sarkar, R., 2022. Fuzzy ensemble of deep learning models using choquet fuzzy integral, coalition game and information theory for breast cancer histology classification. Expert Systems with Applications, 190, p.116167.

[20] Spanhol, F.A., Oliveira, L.S., Petitjean, C. and Heutte, L., 2015. A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering, 63(7), pp.1455-1462.

[21] Majumdar, S., Pramanik, P. and Sarkar, R., 2023. Gamma function based ensemble of CNN models for breast cancer detection in histopathology images. Expert Systems with Applications, 213, p.119022.

[22] Hameed, Z., Garcia-Zapirain, B., Aguirre, J.J. and Isaza-Ruget, M.A., 2022. Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network. Scientific Reports, 12(1), p.15600.

[23] Taheri, S. and Golrizkhatami, Z., 2023. Magnification-specific and magnification-independent classification of breast cancer histopathological image using deep learning approaches. Signal, Image and Video Processing, 17(2), pp.583-591.

[24] Nassif, A.B., Talib, M.A., Nasir, Q., Afadar, Y. and Elgendy, O., 2022. Breast cancer detection using artificial intelligence techniques: A systematic literature review. Artificial Intelligence in Medicine, p.102276.

[25] Atban, F., Ekinci, E. and Garip, Z., 2023. Traditional machine learning algorithms for breast cancer image classification with optimized deep features. Biomedical Signal Processing and Control, 81, p.104534.

[26] Sepahvand, M. and Abdali-Mohammadi, F., 2023. Joint learning method with teacherstudent knowledge distillation for on-device breast cancer image classification. Computers in Biology and Medicine, 155, p.106476.