2023 26th International Conference on Computer and Information Technology (ICCIT)
13-15 December, Cox's Bazar, Bangladesh

Transfer Learning Based Approaches for Breast Cancer Classification using Mammogram Images

Md. Sharifujjaman
Computer Science & Engineering
Rajshahi University of Engineering & Technology,
Rajshahi-6204, Bangladesh
sharifjaman2499@gmail.com

Zannatul Ferdousee
Computer Science & Engineering
Rajshahi University of Engineering & Technology,
Rajshahi-6204, Bangladesh
methela1603060@gmail.com

Nahin Ul Sadad
Computer Science & Engineering
Rajshahi University of Engineering & Technology,
Rajshahi-6204, Bangladesh
nahinsd100@gmail.com

Boshir Ahmed
Computer Science & Engineering
Rajshahi University of Engineering & Technology,
Rajshahi-6204, Bangladesh
boshir78@gmail.com

***Abstract*—Breast cancer is a growing epidemic and a leading cause of death among women worldwide. Mammographic imaging has been found to be highly effective in detecting breast cancer at an early stage which leads to reduce mortality rates through prompt and appropriate treatment. In this research work, the proposed models have used two convolutional neural network (CNN) architecture known as VGG19 and ResNet50, which had been pre-trained with data from imageNet and then Mammographic Image Analysis Society (MIAS) database has been used to train and test. To identify potential cancer hotspots, the mammogram images from MIAS database have went through some image preprocessing steps such as image resizing and augmentation by rotation. The extracted features from the pre-trained architecture have been flattened into one dimension and then used as inputs to the trainable dense layers. The images have been classified either into benign or malignant type by using the sigmoid activation function. Measures of performance such as accuracy, recall, precision and F1-score have been calculated to evaluate the proposed models' efficacy. The experimental result depicts that pretrained VGG19 architecture performed 98.46% outperforming ResNet50 achieving a test accuracy of 97.94% .**

***Index Terms*—Breast Cancer, Mammographic Images, MIAS database, VGG19, ResNet50, Deep Learning, Transfer Learning, Convolutional Neural Network**

I. Introduction

According to the world health organization, breast cancer is the most prevalent type of cancer among women worldwide [1]. Due to rising life expectancy, urbanization and adoption of western lifestyles, breast cancer is rising in the developing globe. Although prevention can reduce risk to some extent, the cornerstone of breast cancer control remains early detection for bettering breast cancer outcome and survival [1]. In 2018, nearly 15% of all cancer fatalities were attributed to breast cancer among women, which resulted in 627,000 people losing their lives to the disease [2]. One of the most reliable methods for early diagnosis and detection of breast cancer lowering the mortality is mammography [3]. According to observational studies, mammography screening results in a mortality reduction of roughly 40% [4]. Other imaging procedures such as magnetic resonance imaging (MRI), ultrasound and X-rays are commonly used to check the breasts. Molecular breast imaging and digital breast tomosynthesis are two examples of the more recent imaging techniques that are now in the process of being developed (DBT) [1]. Mammograms are radiographic images of the breast which are used to identify breast cancer's early signs and symptoms. These radiographic images boost the accuracy of the diagnosis while decreasing the time spent diagnosing cysts and reducing human error. The breast is scanned using a device that produces low doses of two or three dimensional X-rays as part of this type of imaging [5].

The rapid advancement of deep learning and some machine learning architectures have raised a lot of interest in their potential applicability to clinical imaging problems where data preprocessing, feature extraction and classification are the three basic phases for evaluating the models' performance [3]. In this research study, two proposed model has been implemented on one dataset. At first the mammography images go through image preprocessing. Then the preprocessed data were given to the CNN architecture. For the binary classification a sigmoid function is utilized to separate the input sample into one of the two categories: benign or malignant. The rest of the paper is structured as follows: Section II discusses the Related Works in this research domain. Section III contains a detailed description of the dataset. Sections IV and V discuss the Methodology and the Results, respectively. Finally, Section VI contains the conclusion describing the overview of this research work.

II. Related Works

The field of mammography image anomaly automation began with notions in the 1960s [6]. Originally, the focus of development was on decreasing errors due to weariness in human execution. Since then, research and development have encompassed approaches ranging from basic image processing methods to the most recent deep learning techniques [7].

979-8-3503-5901-5/23/$31.00 ©2023 IEEE

Ismail et al. [8] demonstrated how well the VGG16 and ResNet50 architectures perform to classify the normal and aberrant tumors. The number of layers in their architecture affected the process time but with larger network it could also create overfitting. Jahangeer et al. [9] introduced another breast cancer diagnosis system by transfer learning where mammography images of the affected area were used to locate affected areas of the body and the use of a unique decision-based partial median filter eliminated salt and pepper noises in the mammograms which resulted in improved accuracy [9]. Later Isaza et al. [10] demonstrated the need to experiment with several networks for gaining better performance to find one that is well-suited to the task at hand, despite the fact that new innovations and methodologies have been evolved in the designs of deep learning models. To prevent the model from being overfit, data augmentation approaches were incorporated. Jayandhi et al. [11] introduced a method by using deep CNN architecture with transfer learning where due to use of reduced convolution and max-pooling layers VGG16 architecture performed better than VGG19. Alhussan et al. [12] proposed another transfer learning based approach with enhanced performance where the three stages of the proposed architecture were data augmentation, AlexNet based transfer learning feature extraction and optimum classification with a convolutional neural network (CNN).

III. Dataset Description

The database of digital mammograms maintained by the Mammographic Image Analysis Society (MIAS) which is publicly available [13]. It contains the original 322 grayscale images (161 pairs) having size of 1024 × 1024 pixel in the "Portable Gray Map" (PGM) format at a resolution of 50 microns as well as the true data is associated with them. The 322 different images are contained within an Exabyte tape of size of 2.3 GB 8 mm. The database consists of seven columns of which the first and fourth columns represent MIAS database reference number (from mdb001 to mdb322) and severity of abnormality respectively. Out of the 322 images, 115 images has severity of abnormality (64 benign and 51 malignant) and rest 207 are normal images [9].

IV. Technical work

One of the simplest and most common types of neural network architectures is the sequential model. It consist a linear stack of layers in a neural network where each layer has exactly one input tensor and one output tensor. Figure 1 depicts the workflow of the proposed methodology. In this research VGG19 and ResNet50 were used as pretrained (non-trainable) model and later various layers such as flatenning, dense, batch normalization, dropout were added sequentially in different amounts with different parameters to each of the pretrained architectures. Finally sigmoid activation function was used for classification in the last dense layers of the pretrained architectures. Before the training had started, the images of the MIAS database went through various preprocessing steps to prepare the dataset.

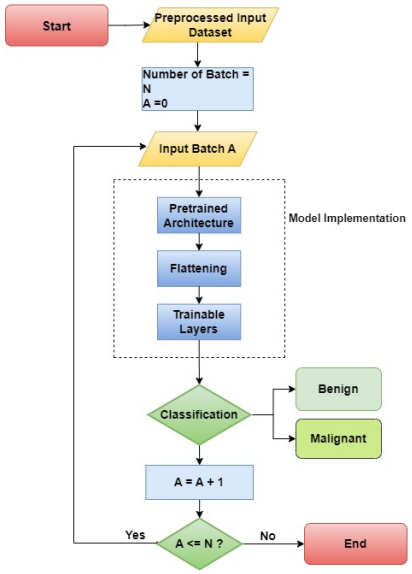

Fig. 1. Proposed methodology.

A. Data Preprocessing

The preprocessing of data is an extremely important step for each and every algorithm that uses deep learning. Image resizing, data augmentation and separating the dataset into train and test data are included in the preprocessing stage of our research work. The 115 images found with severity of abnormality by checking the presence of label 'B' for benign (64 images) and 'M' for malignant (51 images) through each of the rows of 322 images along severity of abnormality column were subjected for preprocessing.

1) Image Resizing: The VGG19 and ResNet50 pretrained architectures take input image of size 224 × 224 pixels. So the images were resized from 1024 × 1024 to 224 × 224 pixels when they act as input to the pretrained models.

2) Data Augmentation: It refers to the process of creating new data points from previously collected data. In this research from each of the labeled 115 images, 45 images were generated by a rotation scale of 8° starting from 0° to 360°. So total number of input images became 115 × 45 = 5175 which were saved in a nested dictionary where the first key was the labeled MIAS database reference number and the nested dictionary's key was the step of angle due to augmentation for each of the first key. The value set to 0 or 1 in the nested dictionary under each of the key (step of angle) within the sub-dictionary associated with the first key represented benign or malignant tumor respectively. Hence the new dataset with 5175 images' labeled value in the nested dictionary was prepared for our research work.

3) Train-Test Splitting: The training and testing data in the created dataset were divided randomly in the proportion of 85:15. The random state value was fixed and the shuffle parameter had been set as true such that the training and testing sets were both reflective of the whole dataset. So out of 5175 images, 4398 and 777 images were selected for training and

testing sets respectively. Out of 4398 training set images, 20% (880 images) were split for validation test.

B. Feature Extraction

Each of the proposed models contained two basic parts. These are feature extraction by the pretrained model and then binary classification replacing the classifier part of pretrained model by adding some sequential trainable layers. Pretrained weights were used from the ImageNet dataset. The final fully connected layers (classification layers) from the pretrained models were excluded by setting 'include top' to false to use VGG19 as a feature extractor and then to build our own unique layers on top for transfer learning purpose.

1) VGG19: A variety of the VGG model known as VGG19 contains sixteen convolution, three fully connected, 1 softmax and five maxpooling layers [14]. The first part is feature extraction. It is comprised of pretrained sequential layers using the VGG19 network. It consists 16 convolutional layers grouped in 5 convolutinal blocks with a ReLU activation function , with a total of 5 maxpooling layers. As input, it requires mammography images that have already been preprocessed and are of shape (224, 224, 3). This indicates that the image's input size is 224 × 224 pixels. All of these layers in these sections were made non-trainable so that they could ignore the weights that were provided by ImageNet. By freezing the layers, the model was able to discover appropriate weights for the target dataset and added the filter to VGG-19 in order to obtain feature information. It makes use of kernels that are 3 × 3 in size and has a stride size of one pixel, which enables it to cover the entire image. By the reduction of dimension maxpooling helps to use less computation. After maxpooling in the first convolutional block, the dimension of the input images was reduced by half which is equal to (112, 112, 128) from (224, 224, 64). In this context, 64 and 128 refer to the total number of (3 × 3) kernels that were utilized. As a consequence of this change, the dimensions of the input images were reduced by half for the subsequent four convolutional blocks. After conducting feature extraction with the assistance of the convolution, ReLU activation and the MaxPooling layers, the fifth block will eventually generate images with the shape (7, 7, 512). The rectified linear unit (ReLU) was constructed so that a non-linear component could be included to the model. The fully connected layer with softmax activation for classification was removed by setting the 'include top' parameter as false. From this pretrained model 45% nodes were dropped out using a dropout layer.

The second stage of this process was providing these images to the trainable layers as one-dimensional features. At first from the pretrained model 45% nodes were dropped out using a dropout layer. The overall convolutional layers can be viewed from Fig. 2. So its clear from the Fig. 2 that we added one flatten, four dense, three batch normalization and three dropout layers. Flattening was the immediate next to dropout step in the process, which consisted of converting each of the vectors into a single dimension. Flatten layer had been added between the freezing layers and the layer that is trainable. The output of the final convolutional layer of the pretrained model after applying the dropout was flattened into a one dimensional vector which acted as input into the dense layers of the trainable artificial neural network. For each image, this section generates a tensor with 7 x 7 x 512 values. With the first two dimensions matching those of the original image (7,7), this can be thought of as a very rough sketch with just 7 x 7 = 49 big pixels. In place of the traditional three color channels per pixel, our network made use of 512 characteristics to represent the image. The number of features collected by the network from the image are stored in this tensor, which comprised 7 x 7 x 512 = 25088 numbers. Table I demonstrates the overview of parameters of the proposed model. There are four

TABLE I
MODEL SUMMARY OF THE PROPOSED ARCHITECTURE WHEN VGG19 HAD BEEN USED AS PRETRAINED (BASE) MODEL.

Layer	Output shape	Parameters
vgg19 (Functional)	(None, 7, 7, 512)	20024384
dropout (Dropout)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 1024)	25691136
batch_normalization (BatchNormalization)	(None, 1024)	4096
activation (Activation)	(None, 1024)	0
dense_1 (Dense)	(None, 1024)	1049600
batch_normalization_1 (BatchNormalization)	(None, 1024)	4096
activation_1 (Activation)	(None, 1024)	0
dropout_1 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 1024)	1049600
batch_normalization_2 (BatchNormalization)	(None, 1024)	4096
activation_2 (Activation)	(None, 1024)	0
dropout_2 (Dropout)	(None, 1024)	0
dense_3 (Dense)	(None, 1)	1025
Total parameters:		47,828,033
Trainable parameters:		27,797,505
Non-trainable parameters:		20,030,528

dense or fully connected layers in our model. Each of the first three dense layers had 1024 units (neurons or nodes), their kernel initializer were set to 'he uniform' which uses the 'He' initialization method with a uniform distribution for initializing the weights that helps to mitigate the vanishing gradient problem and 'ReLU' as activation function. The last dense layer had 1 unit and 'sigmoid' as activation funtion for the binary classification. After each of the first three dense layers, batch normalization is carried out. So there were three batch normalization layers next to each of the first three dense layers. We applied 45% dropout on the pretrained model at first. After second batch normalization 45% of the input neurons were removed from the second dense layer. The overfitting problem was alleviated by the development of the dropout system. The model became more stable as a result of the further reduction of 30% of the neurons on the third dense layers where the dropout layer was added after third batch normalization layer. The dropout values were selected through a number of trials and errors to achieve the best performance.

2) ResNet50: A variation of the ResNet model known as ResNet50 contains 49 convolution layers organised into residualblocks and a fully connected layer for classification resulting 50 layers in total [15]. The left part of Fig. 3 shows the ResNet50 pretrained architecture. Each convolutional blocks have been separated by different colours where the first block has one kernel and each of the next four blocks has three kernels whose size is depicted immediate right to ResNet50 architecture. The size of the image reduces to half after it had

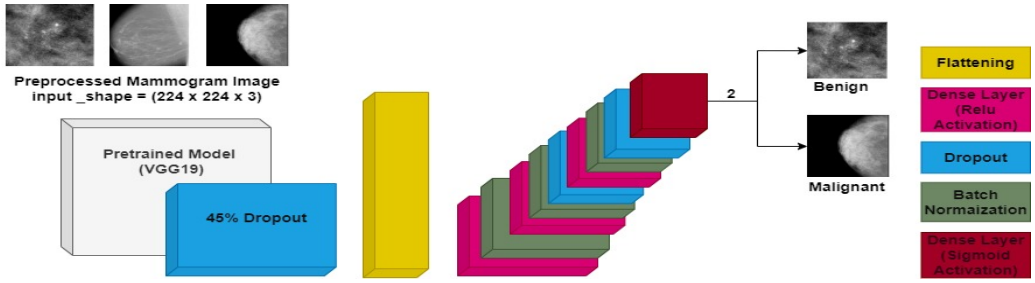

Fig. 2. Architecture of the Proposed Model when VGG19 had been used as pretrained (base) model.

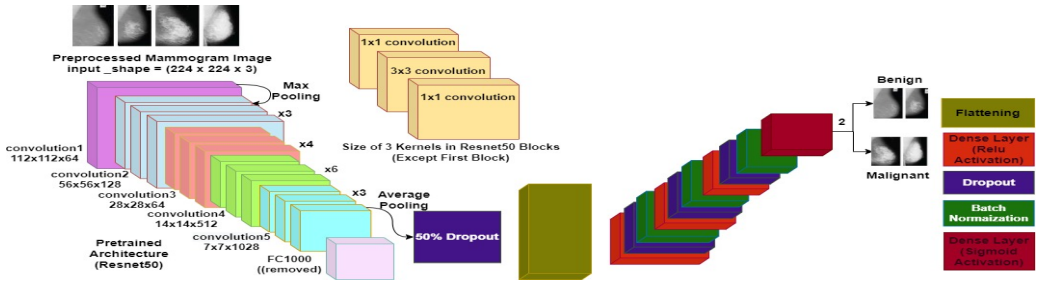

Fig. 3. Architecture of the Proposed Model when ResNet50 had been used as pretrained (base) model.

been processed through each block. All of the layers in the pretrained ResNet50 architecture have had their trainability disabled so that they can ignore the weights that have been provided by ImageNet. By freezing the layers, the model was able to discover appropriate weights for the target dataset. The 224 x 224 size preprocessed mammogram image acts as input to the first convolutional block. The first convolutional block has a kernel (7 x 7, 64) which indicates it has size of 7 x 7 and amount of distinct kernels is 64, each of which has a stride size of 2 giving us one layer. The 224 x 224 size input image gets converted to size 112 x 112. Next the architecture demonstrates maximum pooling with a stride size of two. In the second convolutional block, there is a (1 x 1, 64) kernel, then there is a (3 x 3, 64) kernel, and finally there is a (1 x 1, 256) kernel. These three layers are repeated in total 3 times, giving us a total of 9 layers in this step. The 112 x 112 size input image gets converted to size 56 x 56. The third block has a kernel of (1 x 1, 128), then one of (3 x 3, 128), and finally one of (1 x 1, 512); this phase was repeated four times, giving us a total of 12 layers in this step. The 56 x 56 size input image gets converted to size 28 * 28. After that there is a kernel with the value (1 x 1, 256), followed by two more kernels with the values (3 x 3, 256) and (1 x 1, 1024); this process is then repeated six times, giving us a total of 18 layers. The 28 x 28 size input image gets converted to size 14 x 14. And after that, a (1 x 1, 512) kernel, followed by two more of (3 x 3, 512) and one of (1 x 1, 2048); this process was repeated three times, giving us a total of nine layers. The 14 x 14 size input image gets converted to size 7 x 7. So the final image shape is (7 x 7, 2048). Then an average pooling is performed and the process finishes with a fully connected layer that has 1000 units and a softmax function. The final output image shape is (7, 7, 2048) meaning the image size is reduced to 7 x 7 pixels with 2048 features on each pixel. The last fully connected layer classifies object into 1000 categories. Table II demonstrates the overview of parameters of the proposed model.

TABLE II
MODEL SUMMARY OF THE PROPOSED ARCHITECTURE WHEN RESNET50 HAD BEEN USED AS PRETRAINED (BASE) MODEL.

Layer	Output shape	Parameters
ResNet50 (Functional)	(None, 7, 7, 2048)	23587712
dropout (Dropout)	(None, 7, 7, 2048)	0
flatten (Flatten)	(None, 100352)	0
dense (Dense)	(None, 256)	25690368
dropout_1 (Dropout)	(None, 256)	0
batch_normalization (BatchNormalization)	(None, 256)	1024
dense_1 (Dense)	(None, 256)	65792
dropout_2 (Dropout)	(None, 256)	0
batch_normalization_1 (BatchNormalization)	(None, 256)	1024
dense_2 (Dense)	(None, 256)	65792
dropout_3 (Dropout)	(None, 256)	0
batch_normalization_2 (BatchNormalization)	(None, 256)	1024
dense_3 (Dense)	(None, 1)	257
Total parameters: Trainable parameters: Non-trainable parameters:		49,412,993 25,823,745 23,589,248

In the second phase, the classifier part was skipped and trainable layers were added sequentially. Four dropout layers, one layer for flattening, four batch normalization layers and four dense layers (whose weights were initialized using the default initialization method) were added in total. The first three dense layers had ReLU activation function to introduce nonlinearity and the fourth one has sigmoid activation function to classify the images either into benign or malignant type tumor. Fig. 3 illustrates the proposed model. At first the pretrained model's 50% were dropped out using a dropout layer. The flatten layer was added after this dropout layer. The output of the final convolutional layer is flattened into

a one dimensional vector before acting as input to the dense layers. This part creates a tensor with 7x7x512 values for each image that is being processed. retaining the same proportions in the first two dimensions as the original image (7,7). This version contains merely 7x7 = 49 large pixels, so it can be looked of as more of a sketch than anything else. In order to accurately represent an image this network uses 2048 features as opposed to the more conventional three color channels for each pixel. This tensor, which is composed of 7x7x512 = 100352 numbers, is where the features that were extracted from the image by the network after it was processed. After that one dense layer using 256 units and ReLU as the activation function was added. Following each of this dense layer we applied a dropout layer and a batch normalization layer. In the similar manner there are two more layers of each of the three types. 25% were dropped out in each of the three dropout layers.

C. Classification

In this research, binary classification was used since our aim was to classify images either into benign or malignant tumor for both of the architectures, we used 'sigmoid' activation function in the last dense layer having one neuron. This activation function can be mathematically defined as Equation 1 where 'x' is the input to the sigmoid function, f(x) and 'e' represents Euler's number [16].

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

D. Experimental Setting

"Adam" was used as the optimizer to increase overall performance and efficiency where the hyperparameter learning rate was 0.001. Since the purpose of this research work was to classify the input images either into benign or malignant tumor, "Binary crossentropy" was used as the loss function. After the validation split, the 3518 training images were divided into 110 batches by setting the hyperparameter batch size as 32. The proposed model was iterated through 100 epochs on these training images.

E. Evaluation Metrices

From the confusion matrix measuring metrices such as Accuracy (AC), Recall (RC), Precision (PR), Specificity (SP) and F1-score (F1) were calculated for the evaluation of the proposed architectures. True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values found from the confusion matrix are necessary before proceeding with the calculation of any other evaluation metrics [17]. TP is the output where both the ground value and the result are positive whereas TN is the output where both of them are negative, FP is the positive outcome that corresponds to the negative ground truth whereas FN is the negative outcome that corresponds to the positive ground value [18].

1) Confusion Matrix: The confusion matrix (CM) provides a detailed summary of a classification algorithm's performance by displaying the number of correct and incorrect predictions generated by the model on a dataset. Fig. 4 and Fig. 5 represents the confusion matrices generated by the VGG19 and ResNet50 pretrained architectures respectively.

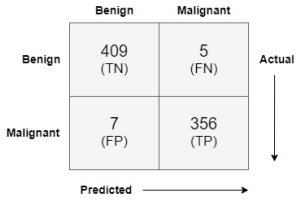

Fig. 4. CM for Pretrained VGG19.

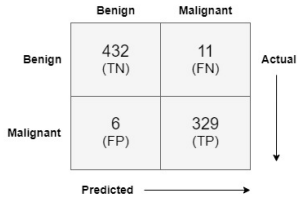

Fig. 5. CM for Pretrained ResNet50.

2) Measuring Matrices: The results for accuracy, precision, recall and f1-score were determined using the following equations.

$$\text{AC} = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

$$\text{PR} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{RC} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{F1} = \frac{2 \times RC \times PR}{RC + PR} \tag{5}$$

V. Results and Discussion

The results of the evaluation matrices evaluated on test set for the proposed models and their performance comparisons with other models has been displayed in Table III which shows that our models have outperformed the recently developed models. The gradual increase of accuracy and decrease of loss over 100 epochs on training and validation sets simultaneously shows the models' effectiveness to make small errors on few data and to mitigate overfitting depicted by Fig. 6 and Fig. 7, which also indicate the better performance of pretrained VGG19 architecture than ResNet50.

TABLE III
RESULTS OF THE EVALUATION MATRICES FOR THE PROPOSED MODELS AND THEIR PERFORMANCE COMPARISON WITH OTHER MODELS.

Model	AC (%)	PR (%)	RC (%)	F1 (%)
VGG16 [8]	94	89	99	93.73
Hybrid VGG16 [9]	96.45	-	-	-
ResNet50V2 [10]	97.73	-	-	-
VGG16 (reduced convolution) [11]	82.5	-	-	-
ABER-CNN (Dataset1: DDSM) [12]	96.2	96.8	97.1	96.9
Proposed Model (ResNet50)	97.94	97.94	97.94	97.94
Proposed Model (VGG19)	98.46	98.46	98.46	98.45

For the proposed pretrained ResNet50 and VGG19 architectures the obtained accuracy are 97.54% and 98.46% respectively suggesting that in the majority of circumstances they can make accurate predictions, obtained precision are 97.54% and 98.46% respectively showing the models' rates of false-positive predictions are comparatively low, obtained recall are 97.54% and 98.46% respectively demonstrating their ability to recognize true positive cases effectively and obtained F1-score are 97.54% and 98.45% respectively that offers a useful synopsis of the model's overall performance in recognizing positive and negative cases accurately.

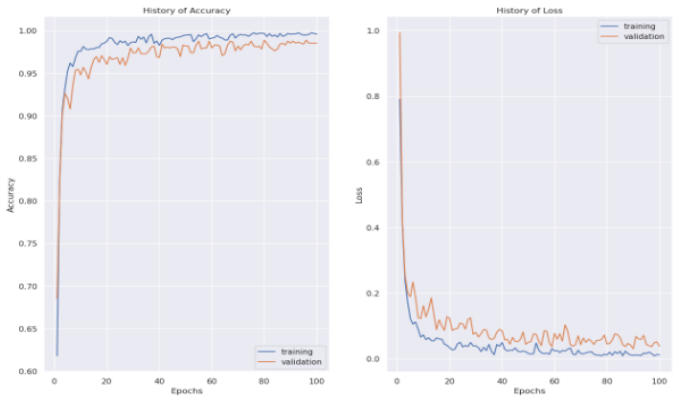

Fig. 6. Accuracy and Loss graph of pretrained VGG19 architecture on training and validation sets over 100 epochs.

Fig. 7. Accuracy and Loss graph of pretrained ResNet50 architecture on training and validation sets over 100 epochs.

VI. Conclusion

Worldwide an increasing number of women are succumbing to breast cancer making it the top cause of mortality among females. For it's early detection mammographic imaging has been demonstrated to be extremely useful, which can lead to a reduction in mortality if the disease is treated in a timely and suitable manner. The performance of the suggested model was dependent upon a number of important aspects such as the size and amount of the kernel, the learning rate, the amount of neurons used in the dense layers and the dropout rate. All of these has been carefully controlled and the best values were selected through many trials. In this research, the proposed models were trained on only one dataset. Besides the dataset is simplistic having only one object in each image and the total amount of images is small. It is possible that our model will run into difficulties when attempting to make a prediction using an image that contains more than one object. However, the proposed models outperform the most advanced models currently available. The enhanced performance results due to the fine tuning of different hyperparameters of the transfer learnings.

References

[1] 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.

[2] B. Arko-Boham, B. A. Owusu, N. A. Aryee, R. M. Blay, E. D. A. Owusu, E. A. Tagoe, A. R. Adams, R. K. Gyasi, N. A. Adu-Aryee, and S. Mahmood, "Prospecting for breast cancer blood biomarkers: death-associated protein kinase 1 (dapk1) as a potential candidate," *Disease markers*, vol. 2020, 2020.

[3] G. Meenalochini and S. Ramkumar, "Survey of machine learning algorithms for breast cancer detection using mammogram images," *Materials Today: Proceedings*, vol. 37, pp. 2738–2743, 2021.

[4] C. Nickson, K. E. Mason, D. R. English, and A. M. Kavanagh, "Mammographic screening and breast cancer mortality: a case–control study and meta-analysis," *Cancer epidemiology, biomarkers & prevention*, vol. 21, no. 9, pp. 1479–1488, 2012.

[5] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances," *IEEE transactions on information technology in biomedicine*, vol. 13, no. 2, pp. 236–251, 2009.

[6] F. Winsberg, M. Elkin, J. Macy Jr, V. Bordaz, and W. Weymouth, "Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis," *Radiology*, vol. 89, no. 2, pp. 211–215, 1967.

[7] Y. Gao, K. J. Geras, A. A. Lewin, and L. Moy, "New frontiers: an update on computer-aided diagnosis for breast imaging in the age of artificial intelligence," *AJR. American journal of roentgenology*, vol. 212, no. 2, p. 300, 2019.

[8] N. S. Ismail and C. Sovuthy, "Breast cancer detection based on deep learning technique," in *2019 International UNIMAS STEM 12th engineering conference (EnCon)*. IEEE, 2019, pp. 89–92.

[9] G. S. B. Jahangeer and T. D. Rajkumar, "Early detection of breast cancer using hybrid of series network and vgg-16," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7853–7886, 2021.

[10] A. Anaya-Isaza, L. Mera-Jiménez, J. M. Cabrera-Chavarro, L. Guachi-Guachi, D. Peluffo-Ordóñez, and J. I. Rios-Patiño, "Comparison of current deep convolutional neural networks for the segmentation of breast masses in mammograms," *IEEE Access*, vol. 9, pp. 152 206–152 225, 2021.

[11] G. Jayandhi, J. L. Jasmine, and S. M. Joans, "Mammogram image classification system using deep learning for breast cancer diagnosis," in *AIP Conference Proceedings*, vol. 2519, no. 1. AIP Publishing LLC, 2022, p. 030066.

[12] A. A. Alhussan, A. A. Abdelhamid, S. Towfek, A. Ibrahim, L. Abualigah, N. Khodadadi, D. S. Khafaga, S. Al-Otaibi, and A. E. Ahmed, "Classification of breast cancer using transfer learning and advanced al-biruni earth radius optimization," *Biomimetics*, vol. 8, no. 3, p. 270, 2023.

[13] J. Suckling, "The mammographic images analysis society digital mammogram database," in *Exerpta Medica. International Congress Series, 1994*, vol. 1069, 1994, pp. 375–378.

[14] Y. Zheng, C. Yang, and A. Merkulov, "Breast cancer screening using convolutional neural network and follow-up digital mammography," in *Computational Imaging III*, vol. 10669. SPIE, 2018, p. 1066905.

[15] G. D. Pearse, M. S. Watt, J. Soewarto, and A. Y. Tan, "Deep learning and phenology enhance large-scale tree species classification in aerial imagery during a biosecurity response," *Remote Sensing*, vol. 13, no. 9, p. 1789, 2021.

[16] F. E. Ahmed, "Artificial neural networks for diagnosis and survival prediction in colon cancer," *Molecular cancer*, vol. 4, no. 1, pp. 1–12, 2005.

[17] K. A. Hasan and M. A. M. Hasan, "Classification of parkinson's disease by analyzing multiple vocal features sets," in *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE, 2020, pp. 758–761.

[18] M. A. R. Ridoy and M. R. Islam, "A lightweight convolutional neural network for white blood cells classification," in *2020 23rd International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2020, pp. 1–5.