

# Forecast Breast Cancer Cells from Microscopic Biopsy Images using Big Transfer (BiT): A Deep Learning Approach

Md. Ashiqul Islam<sup>1,\*,#</sup>, Dhonita Tripura<sup>2#</sup>, Mithun Dutta<sup>3</sup>, Md. Nymur Rahman Shuvo<sup>4</sup>, Wasik Ahmmed Fahim<sup>5</sup>  
Puza Rani Sarkar<sup>6</sup>, Tania Khatun<sup>7</sup>

Dept. of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh<sup>1, 4, 5, 6, 7</sup>

Dept. of Computer Science and Engineering, Rangamati Science and Technology University, Rangamati, Bangladesh<sup>2, 3</sup>

**Abstract**—Now-a-days, breast cancer is the most crucial problem amongst men and women. A massive number of people are invaded with breast cancer all over the world. An early diagnosis can help to save lives with proper treatment. Recently, computer-aided diagnosis is becoming more popular in medical science as well as in cancer cell identification. Deep learning models achieve excessive attention because of their performance in identifying cancer cells. Mammography is a significant creation for detecting breast cancer. However, due to its complex structure, it is challenging for doctors to identify. This study provides a convolutional neural network (CNN) approach to detecting cancer cells early. Dividing benign and malignant mammography images can significantly improve detection and accuracy levels. The BreakHis 400X dataset is collected from Kaggle and DenseNet-201, NasNet-Large, Inception ResNet-V3, Big Transfer (M-r101x1x1); these architectures show impressive performance. Among them, M-r101x1x1 provides the highest accuracy of 90%. The main priority for this research work is to classify breast cancer with the highest accuracy with selected neural networks. This study can improve the systematic way of early-stage breast cancer detection and help physicians' decision-making.

**Keywords**—Convolutional neural network (CNN); breast cancer; Big Transfer (BiT); densenet-201; NasNet-Large; Inception-Resnet-v3; mammography

## I. INTRODUCTION

Breast cancer is the second crucial illness worldwide [1]. In 184 significant countries, breast cancer is most common in 140 (70%) countries and a frequent cause of cancer mortality in 101 (55%) countries. Compared to other diseases, the ratio of breast cancer in women is higher in more developed countries and increases rapidly. The swift advancement of new technology is helping the doctor to identify the breast cancer cells in the inflammation stage with the help of Artificial Intelligence. Deep learning models aided in the prognosis of cancer cells and took the necessary steps.

Almost every woman is fearful if they feel something abnormal in their breasts. Moreover, they become much seared when they lump their breast [2]. However, most of them are not aware of the thing that all lump is not cancer. Furthermore, there is a thing if a women lumps in her breast she cannot even ignore this. Because according to DRHC (STYLE YOUR HEALTH), one out of four women has a complaint about their

breast at one time, and one-quarter of those patients who complain about their breast has cancer. It is considered one of the most severe and petrified diseases in women.

There are many issues affecting breast cancer. A woman can be affected by breast cancer if she gets it from an inherited genetic mutation. The main risk factor for breast cancer is being women and getting older. According to the CDC (Centers for Disease Control and Prevention), most breast cancer patients are found in a woman whose age is 50 years or more than 50 years [3]. These factors cause breast cancer, but they also have other factors that can because breast cancer. Reproducing history is another serious risk factor. In this factor, a woman who has menstrual periods before age 12 and menopause after age 55. This risk factor raises the risk of getting breast cancer. Some other factors like having dense breasts, previous treatment using radiation therapy, women who took the drug diethylstilbestrol (DES) can cause breast cancer.

Breast cancer disease has two advanced primary detections: early detection and screening [4]. Detecting early age mammography is the best technique for doctors. For being a most sensitive technique, a significant fraction of patients are referred for biopsy. As a reason, mammography findings do not have a fatality. However, a specific biopsy is an expensive, presumptuous, emotionally disturbing procedure for women.

At present, researchers are mainly working on deep learning, which performs well in image processing. Out of all deep learning, artificial neural network is the most popular, whereas Convolutional neural network is an extended version of it [5]. CNN is a robust algorithm that can extract features from raw input. CNN performs better than any other in image segmentation and achieves better accuracy.

Cancer tumors generally happen when the cells of breast ducts are overgrown from normal cells. A study shows that when diagnosing breast cancer, microscopic images are ubiquitous. Breast tissue can provide important microscopic level images to the pathologist for access. From the analysis, a pathologist can identify the tissue as normal tissue, benign or malignant tissue. If a patient has a benign tumor, cells form a lump and grow abnormally, but it does not spread to other body parts [6].

\*Corresponding author.

# Both authors contributed equally.

On the other hand, an untreated malignant tumor can spear any part of the body. Here, it is crucial to identify types of tumors from the beginning. It can easily save a women's life as well as the precious time of the doctors.

This decade, machine learning algorithms have been proposed for cancer diagnosis from microscopic biopsy images [14]. Neural network algorithm-based intelligent systems are proposed to be a part of the systematic diagnosis process. Artificial intelligence (AI) can impact the ratio of early-stage cancer cell identification with its immense power and growing improvements.

## II. LITERATURE REVIEW

Cancer is currently a dire thread for women, especially for elders. Breast cancer is the second most cancerous disease in the world right now. As a result, many works have been done on breast cancer. Deep learning plays an essential role in medical science. With the help of deep learning algorithms, people can predict different diseases from very early stages, significantly reducing the suffering of patients and doctors.

Yadavendra discussed the detection of benign and malignant tumors in the breast using machine learning and deep learning algorithms [3]. They used almost 2lakhs color patches sized 50×50. For the library, the author implemented sci-kit, Keras, and tensor flow with CNN based classifier. For training, testing and validation, sigmoid activation provided output in malignant and benign classes. CNN-based classifier outperformed all other machine learning algorithms.

Preprocessing is very important for removing noises, artifacts, and muscle regions as they increase the probability of false-positive values. As a result, Luqman Ahmed proposed a method of fine-tuning and preprocessing to decrease the probability of false-positive rates [4]. The datasets were collected from mammographic image analysis society digital mammogram database (MIAS) and Curated Breast Imaging (Digital Database for Screening Mammography). The author said that for handling big data vast amount of memory resources is required; here, the author used a method of converting data in small patches for batch training. ResNet is 152 layer CNN. The limitation of this method is that the cancerous region is irregular in shape and border, affecting calculation and resulting in a drop in precision. Authors [7] apply machine learning-based classification to determine the risk factor and prognosis of the CKD disease.

ZHIWEN HUANG said a hybrid neural network is better than other CNN classifiers in accuracy, sensitivity, specificity [8]. The author here presents a hybrid model based on DenseNet and PCANet. For learning about the dataset, he used a kernel of modified PCANet. However, the DenseNet is used for high-level image classification and constructing dense blocks or transition blocks. The method uses the same number of feature map outputs in transition block input. The global pooling technique was implemented for blocking a large number of weights in the network.

Data augmentation increases the image dataset ten times with original quality with random geometric image transformations, flipping, rotating, scaling, and shifting. The

proposed models by Li-Qiang Zhou were trained with Keras 2.2.0 and Tensor Flow [2] and the weight of the pre-trained model on ImageNet. Feature visualization methods can increase the predictive ability of deep learning networks. The author also shows that his model performs better than the radiologist's inaccuracy.

Deep Convolutional Neural Network is a robust deep learning algorithm [9]. It is pretty used in breast cancer segmentation and detection. Author Md Zahangir Alom compared the existing algorithms concerning image, patch, image-level, and patient-level classification. For the implementation, he used Keras and Tensorflow frameworks. Here, the author mentioned the eight types of breast cancer with image level and patient-level performance classification. The image-wise and patch-wise classification was discussed for better understanding. Researchers [10] evaluated the papaya disease classification using a Convolutional Neural Network.

Convolutional Neural Network outperformed any other algorithms in image recognition [6]. The author gives an example of ResNet outperforming human participants with an error rate of 3.6%. Different dataset types were used in the proposed model, and it seemed like the performance was varying with different datasets. The model successfully predicts 12 types of skin tumors. For improving the performance, it is crucial to add standardized diagnostics images.

Vikas Chaurasia described the 10-fold cross-validation for measuring the performance of the models [11]. The author used prediction models for malignant and benign parameters, represented by "1" for malignant and "0" for benign.

The author presents a framework of unsupervised feature learning by integrating principal components [12]. DEJUN ZHANG aimed to merge feature selection and feature extraction in a deep learning algorithm. For tackling overfitting, he implemented some scarcity penalties in hidden layers. Elu was introduced for speeding up the training process in the deep neural network.

The study [13] proposed a deep learning model with Convolutional layers for breast cancer classification to extract visual features. First, the author implies a preprocessing method to transform images into common space or variances for improving detection performance. They discussed the effectiveness of data augmentation for increasing training data. Also, multiple deep Convolutional networks can improve the performance significantly with a tensor flow machine learning system. Extracted features were used to boost the framework.

The paper proposed an approach of DenseNet deep learning based on CNN for multiclass classification [14]. DenseNet uses a transition layer to reduce the size of the dataset. The author applied dropout and batch operation for optimization—transfer learning, fine-tuning pre-trained CNN models from natural to medical images. The advantage of using DenseNet is feature concatenation that helps to learn features without compression from any stages. Furthermore, the DenseNet model shows that the deep learning model can obtain good performance from natural images.

Author Angel Cruz-Roa said the main objective of their proposed work is to identify tumors from digital images with deep learning [15]. They show the experimental result of the method that can detect breast cancer regions. Five cohorts were used to train, test and validation. With the help of the dice coefficient, the performed evaluation shows positive predictive value, negative predictive value, true positive rate, true negative rate, and false positive and false negative rate all over the test slides.

The workflow of computer vision-based breast cancer cell detection primarily focused on the supervised learning-based system rather than the semi-supervised or unsupervised system. For the most part, the authors depend on hybrid neural network architectures like ResNet, DenseNet, and Inception. During the neural network-based models training and validation of the trained models, many different sources of public datasets took part as raw input data. Recent researches contain pre-trained weights from Keras deep learning models. Using pre-trained weights initiated by the ImageNet dataset has validated the models [26] [31]. To enrich the good prediction ability of models, researchers mainly focused on data preprocessing stages. Data augmentation has taken part for a smaller dataset, and a different cross-validation process follows through some research.

### III. PROPOSED SYSTEM

This study presents quantitative research on deep learning models based on Convolutional Neural Network (CNN). CNN is applying in the breast mammography dataset to prognosis the breast cancer disease in the human breast cell [16][17]. In this study, four deep learning algorithms (NasNet-Large, Inception ResNet-V3, DenseNet-201, and Big Transfer) are applied to achieve better accuracy from the prediction model [18][9][19]. In the beginning, data are acquisition and preprocessed with different parameters to split the dataset. Various preprocessing techniques such as zoom, rotation, rescaling, flip, shuffle are used to improve the data quality [20][18]. The preprocessed images go through the deep convolutional neural network. Convolutional blocks of the models extract the main features from the input images [21]. For training and testing the splits, the pre-trained model is customized with fully connected layers. The parameter of resizing is (224 x 224), (331 x 331), and the rescaling technique (0-1) is applied to split the data. In fine-tuning, selected convolutional layers are responsible for extracting feature extraction with fully connected layers [11]. This process changed the pre-trained model weights and evaluated the testing data for comparative analysis. Fig.1 visualizes the entire workflow of the proposed system.

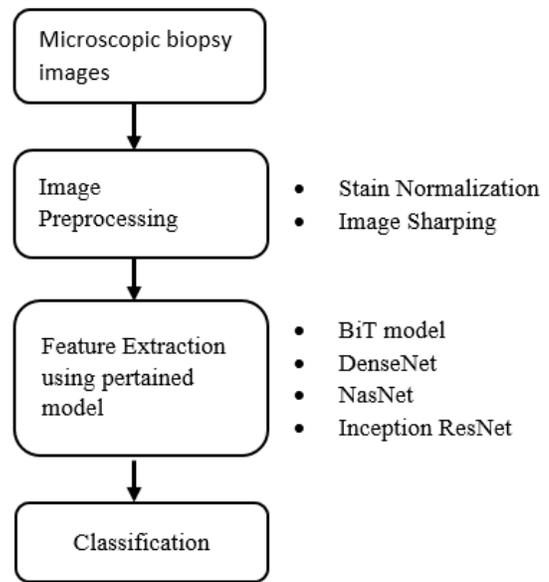


Fig. 1. Process Diagram of the Entire System.

### IV. DATASET DESCRIPTION

Benign and malignant breast cancer is microscopically imaged. Benign and malignant breast cancer is microscopically imaged. Table 1(I, II) provide general information about the dataset. The dataset is divided into two sections, Data for training and testing. In training data, benign and malignant directories contain 436 and 918 image data, respectively. On the other hand, test data is also divided into benign and malignant directories [22]. In testing, the benign directory has 111 files, and the malignant directory has 228 files.

TABLE I. (I): DATASET DESCRIPTION

Dataset features	Parameters
Total instance	1693
Total training data	1354
Total testing data	339
Number of classes	2

(II): DATASET DESCRIPTION

Breast Cancer classification	Number of Images	Number of Images Used for Train and Validation	Number of Images used for Test
Benign	547	436	111
Malignant	1146	918	228
<b>Total</b>	1693	1354	339

This breast cell microscopic biopsy patient's images were gathered from Kaggle [23]. Fig.2 visualize the microscopic images of a cluster of cells. The cluster of cells is called a tumor in medical science, and it is an abnormal growth of the cell. General tumors without invading the nearby area are categorized as benign, Fig.2 (a). A tumor with vast and uncontrollable speed is mentioned as malignant, as shown in Fig.2 (b).

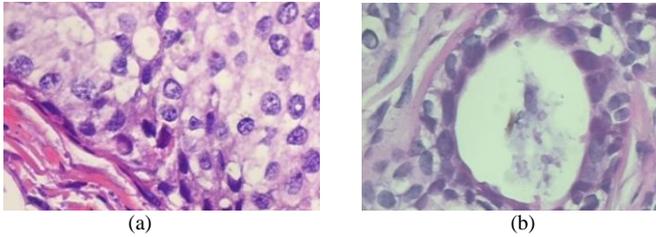


Fig. 2. Microscopic Biopsy Images of a Cluster of Cells from the Selected Dataset [23]. Noncancerous and Cancerous Tumors are Mentioned as Benign (a) and Malignant (b), Respectively.

### A. Data Preprocessing

The dataset contains a variety of image types. The dimensions are as follows: 1500x750, 1254x836, 1024x768, 800x533, 220x230, 100x100, and 61x159. Due to the different image sizes, classification becomes difficult without a fixed image size. As a result, a proposed method used the 200x200 shape as an input shape function. This dataset must contain RGB images. Stain normalization and sharpening the edges of image contents taken part.

## V. MODEL DESCRIPTION

### A. Big Transfer

A BiT, also known as Big Transfer, is a recipe for an image classification model used for pre-training on large supervised datasets [19]. This algorithm has excellent fine-tuning efficiency on the given task. It can provide fantabulous performance on different tasks. R101x1 architecture is implemented in this model; as a result, it can execute multi-label classification on an ImageNet-21k dataset which can contain 14 million images. The output can detect the existence and lack of multiple classes of objects. There are two features in Big Transfer, one is the fine-tuning, and the second one is the BiT collection. Fine-tuning refers to the exact adjustment of parameters in a model. BiT uses a recipe, aka BiT Hyper Rule, for fine-tuning the parameters. This fine-tuning protocol is applied on many down-streaming tasks. BiT collections are used for performing multi-label classification on legit ImageNet-21k datasets. For classifying, it needs an imagenet21k\_classification model. For up streaming pre-trained data scale is essential because it can transfer to tasks with few data points. Residual block showed in Fig. 3 used in M r-101 model with minor changes in the normalization process.

**Residual Block:** Generally, the residual block is a layer stack where the preceding output feeds to the block's deeper layer [24]. For addressing the degradation problem of a complex function, a simpler function should be a subset which is the main idea of the residual block. I consider an input as  $x$  and desired mapping from input as  $g(x)$  then the simpler

function will be  $f(x) = g(x) - x$ . On the other hand, optimization of residual can compromise the dreaded identity mapping in a deep network. When the identity mapping is optimal, the optimization will carry out the weights of residual function to zero.

### B. NasNet Large

NasNet search space is known as Neural Architecture Search, which is used to find convolutional architecture from a dataset. In this search space, a Recurrent Neural Network controller is used to sample different architectures of a child network. Fig. 4 shows the neural architecture generation and selection process of the NasNet model. The child network is trained on a validation set to achieve some accuracy. These accuracies update the controller so that it can generate improved architecture with time. Policy gradient updates the weights of the controller. There are two types of cells named Normal cell and Reduction cell. The average cell returns the feature map's exact dimension, and the Reduction cell returns the reduced feature map (reduced by height and width by factor two). All convolutional cells have two striding for reducing the height width. NasNet can be forced as a hierarchical structure for asserting a well-designed network. For finding optimal architecture:

$$J(\theta_c) = E_{p(a_{1:T}; \theta_c)}[R] \quad (1)$$

Where  $\theta_c$  controller RNN,  $a_{1:T}$  list of actions.

For updating  $\theta_c$  policy gradient method of REINFORCE rule is:

$$\nabla J(\theta_c) = \sum_{t=1}^T E_{P(a_{1:T}; \theta_c)} [\nabla \theta \log P(a_t[a_{(t-1):1}; \theta_c]) R] \quad (2)$$

Where R is non-differentiable.

To overcome unbiased estimation and for reducing variance here need baseline function:

$$\frac{1}{m} \sum_{k=1}^m \sum_{t=1}^T \nabla \theta \log P(a_t[a_{(t-1):1}; \theta_c]) (R_k - b) \quad (3)$$

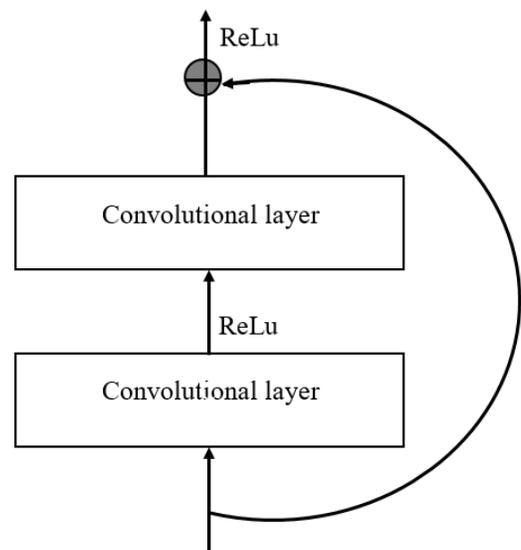


Fig. 3. Basic Residual Blocks of ResNet Architecture [27]. ResNet Architecture is at the Core of BiT Models.

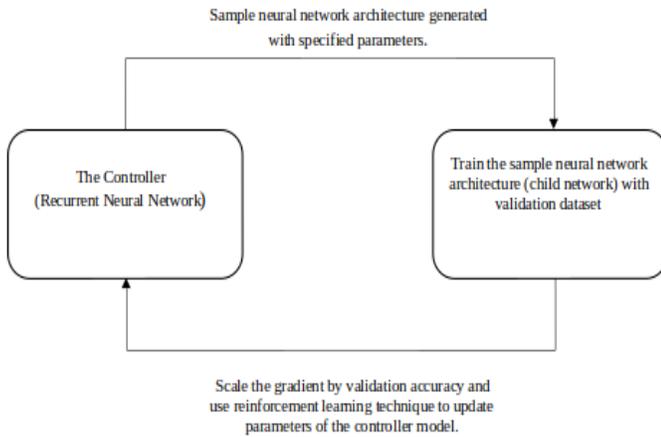


Fig. 4. NasNet Models Process Diagram [28]. The Controller Network is taken Part in Selecting the most efficient Neural Parameters for a Specific Dataset.

Here  $m$  is the number of various architectures,  $T$  is the number of hyper parameters. The validation accuracy of the  $k$ -th neural network after training a dataset is  $R_k$ . If  $b$  doesn't depend on current action then the function is unbiased gradient estimation.

**NasNet Controller Architecture:** On Neural Architecture Search, a controller is being used by us for creating architectural hyperparameters for neural networks [25]. The controller is mainly prepared as a recurrent neural network. Fig. 5 visualize the internal architecture of the NasNet controller. For predicting feed-forward neural networks with convolutional layers, the controller provides hyperparameters as tokens. The generating process of architecture is halted when the layer exceeds a particular value. For sampling convolutional network RNN anticipates filter height-width, stride height-width, and filter numbers for one or more layers. A Softmax classifier makes each prediction, and then the output is fed into the next step.

C. DenseNet-201

Considering CNN, DenseNet contains fewer perimeters than most conventional models [9]. A DenseNet architecture visualizes in Fig. 7. It contains sense built-in blocks with steam and transactional neural blocks. This architecture is not needed necessary to learn redundant feature maps. DenseNet can add a small set of new features because the layers it is narrow. To train intense networks, it has to face problems due to alluded flow of information and gradients. It can solve the problems by DenseNet. Cause each layer of this algorithm has direct access to gradients from the loss function.

Equations of DenseNet would be:

$$X_L = H_L(X_{L-1}) \tag{4}$$

For including skip connection ResNet extends the behavior by reformulating the equation:

$$X_L = H_L(X_{L-1}) + X_{L-1} \tag{5}$$

In this algorithm, the incoming feature maps don't sum up the output of the feature maps layer. So, the final equation is:

$$X_L = H_L([X_0, X_1, \dots, X_{L-1}]) \tag{6}$$

This thing makes the main difference between ResNet and DenseNet.

Using convolutional neural networks in a systematic flow for a specific task is known as blocks of CNN. Dense Block (showed in Fig. 6) is a benchmark model for specific feature extraction purposes. This module connects all matched feature maps (layers) with other available layers. For retraining from all feed forward quality each layer gets extra inputs from all foregoing and leaves them to subsequent layers. On DenseNet features are concatenated and the  $l^{th}$  layer has 1 input which consists of feature maps of previous convolutional blocks and its feature maps are left for  $L-l$  subsequent layers. For traditional dense connectivity, this layer implemented  $\frac{L(L+1)}{2}$  instead of doing  $L$ . The networks defined  $L$  as layers, and  $H_l(\cdot)$  is non-linear transformation,  $x_l$  is the output of the  $l^{th}$  layer.

The convolutional block is consists of two or more convolutional layers with essential activation functions. Fig. shows that every convolutional layer receives direct input from preceding layers. The transition layer mainly does the work of pooling and convolution, which is used in batch normalization with 1x1 convolution and 2x2 pooling. Input feature maps first go to batch normalization and standardize input data. After completing each convolution, the number of channels shows the growth rate of a block by remaining the same as before. After mapping out the feature, the output sends to all of the layer blocks. For extracting features from input data, activation function and convolutional layer are used. VGG-16 networks hold 16 convolutional layers with the same kernel.

D. Inception ResNet-V3

Inception networks interrupt the same filter size concept in a block [16][30]. A single built-in block searches for a different convolutional layer with different shapes shown in Fig. 8. Multiple layers subsist in parallel in a block and concatenate each sequence of layers. 1x1 convolutional layer at the beginning reduces the dimensionality of a sequence. Multilevel feature extraction is expedited by concatenating various filter-sized sequences. Res-Net passes the output of a convolutional layer typically.

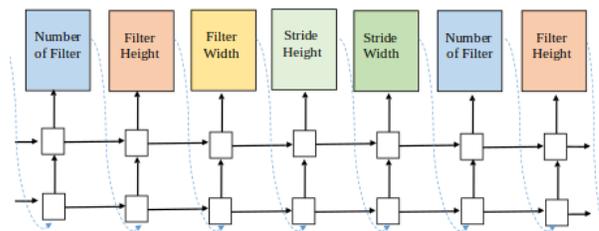


Fig. 5. NasNet Model Controller Architecture [28] based on Recurrent Neural Network (RNN).

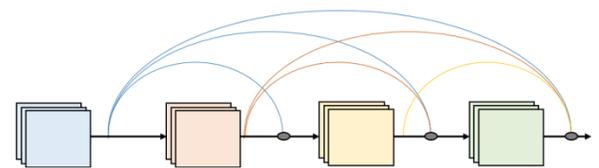


Fig. 6. 4-Layer Dense Block with Growth Rate (Number of Channel in Layer) of 3.

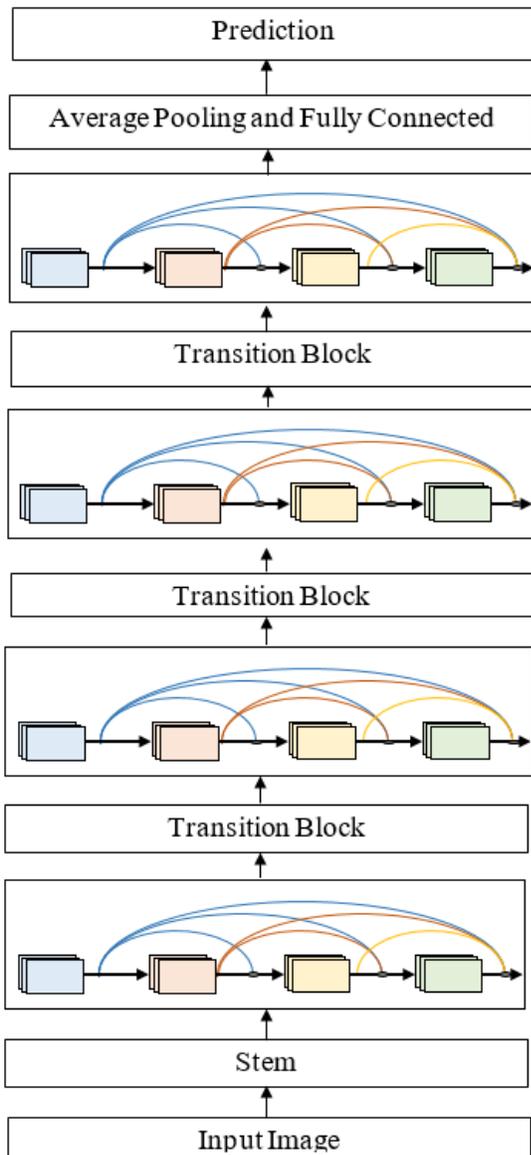


Fig. 7. DenseNet201 Model Architecture [29].

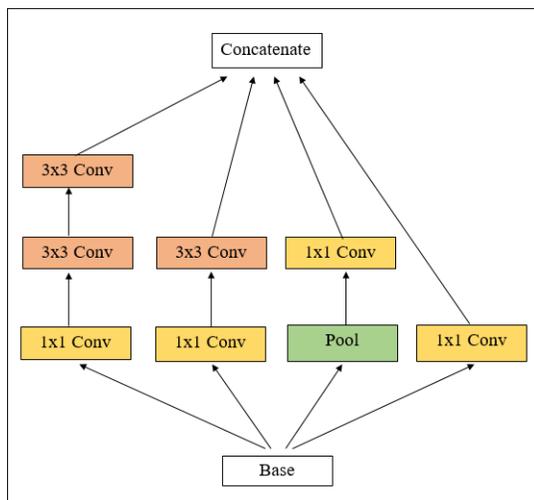


Fig. 8. Basic Inception Built-in Block [30].

## VI. SYSTEM OPERATION

Much work has been done on breast cancer using different types of machine learning and deep learning algorithms. This paper made a complete guideline to classify breast cancer-affected patients by following a flowchart given in Fig. 1. In this flowchart, after acquiring affected people's images, it needs to process those images. In this step, the machine takes the images from the dataset as input. In processing, the system split a dataset into two parts train data and test data. The system uses 80% of data for training and 20% of data for testing. This train data has to go through different processing techniques (rotates, room, flip and shuffle). This process improves the data quality by enhancing image features essential for the next training part of the process [22]. On the other side, in training data, the system applied resizing (224×224) and rescaling (0-1). Only the standard processes are applied between training and testing data.

Then these processed images have to go through the deep convolutional neural network. It is a feed-forward artificial neural network and has taken the role of feature extraction [12]. It is a fully connected layer containing multiple nodes, and every node is connected to the subsequent nodes with the next layer. Pre-trained weights can be changed by training procedure [15]. For comparative analysis, trained models are evaluated by testing data split.

## VII. RESULT ANALYSIS

This analysis was performed on four different types of deep learning algorithms. Table 2 contains the performance summary of selected models. The selected architecture's performance was promising. This study includes NasNet-Large, DenseNet-201, Inception-ResNet-V3, and Big Transfer, with a total of 1693 instances in the dataset. From here, 1354 were used in training and 339 instances in testing. The dataset contains two classes: Malignant and Benign. Out of 1693 selected datasets has 547 images in Benign and 1146 images in Malignant. Again 434 Benign images were used in training, and 111 were in testing.

On the other hand, 918 malignant images were used in the training phase and 228 in the testing phase. After analyzing the training and testing phase, Big Transfer achieved the highest accuracy of 90%, whereas DenseNet-201 has 89%, Inception-ResNet-V3 has 86%, and NasNet Large 81%. One of the great features of this study is that the applied algorithm will take the highest accuracy from 200 epochs intelligently, whereas other typical algorithms take values from predefined epochs [17] (Fig. 9). Table 3(I, II, III, IV) contains the performance measurement of the models with different classes of data. Here support is 111 for Benign and 228 for Malignant for each algorithm.

TABLE II. PERFORMANCES OF SELECTED MODELS

Models	Accuracy
NasNet Large	0.8183
DenseNet-201	0.8917
Inception-ResNet-V3	0.8642
Big Transfer	0.9000

NasNet Large: For Benign class NasNet has 0.31 precision with 0.29 Recall, 0.30 F1 Score. For Malignant class algorithm has 0.67 precision with 0.70 Recall, 0.68 F1 Score.

TABLE III (I): THE PERFORMANCE SCORE FOR EACH CLASS OF NASNET LARGE MODEL

Classes	Precision	Recall	F1 Score	Support
Benign	0.31	0.29	0.30	111
Malignant	0.67	0.70	0.68	228

DenseNet-201: For Benign class DenseNet has 0.28 precision, 0.27 Recall, 0.27 F1 Score and for Malignant class the precision is 0.66, Recall & F1 Score 0.67.

TABLE III (II): THE PERFORMANCE SCORE FOR EACH CLASS OF DENSENET-201 MODEL

Classes	Precision	Recall	F1 Score	Support
Benign	0.28	0.27	0.27	111
Malignant	0.66	0.67	0.67	230

Big Transfer: For the Benign class Big Transfer has 0.29 precision with 0.27 Recall, 0.28 F1 Score. For the Malignant class, the algorithm has 0.66 precision with 0.68 Recall, 0.67 F1 Score.

TABLE III (III): THE PERFORMANCE SCORE FOR EACH CLASS OF THE BIG TRANSFER MODEL

Classes	Precision	Recall	F1 Score	Support
Benign	0.29	0.27	0.28	111
Malignant	0.66	0.68	0.67	228

Inception-ResNet-V3: For Benign class Inception-ResNet-V3 has 0.31 precision, 0.31 Recall, 0.31 F1 Score and for Malignant class the precision, Recall & F1 Score is 0.67.

TABLE III (IV): THE PERFORMANCE SCORE FOR EACH CLASS OF THE INCEPTION-RESNET-V3 MODEL

Classes	Precision	Recall	F1 Score	Support
Benign	0.31	0.31	0.31	111
Malignant	0.67	0.67	0.67	228

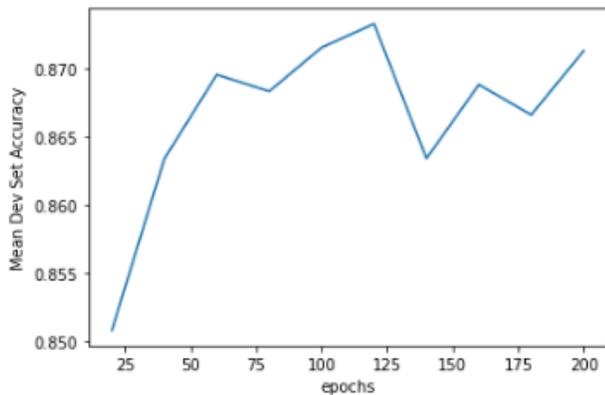


Fig. 9. Best Epochs Selection between 20 and 200.

Training and validation are a significant part of deep learning. If the algorithm is well trained and validated, it can perform better in higher accuracy output. Fig. 10 shows that the dataset was well trained. The training accuracy was almost 0.98 with a standard deviation of 0.0258021, and the highest epoch generated was 120.

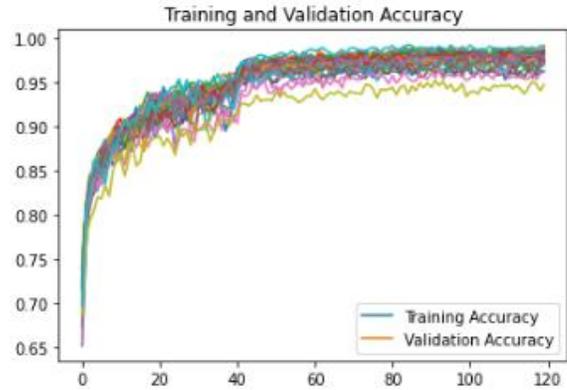


Fig. 10. Training & Validation Accuracy.

Fig. 11 shows that the number of accuracies repeated while performing accuracy with algorithms. Here 0.90 and 0.89 were achieved once. 0.86 accuracy was achieved two times, accuracy between 0.86 and 0.87 was achieved three times. 0.87 And 0.88 were achieved six times. The highest repeated number is 8. It was between 0.87 and 0.88.

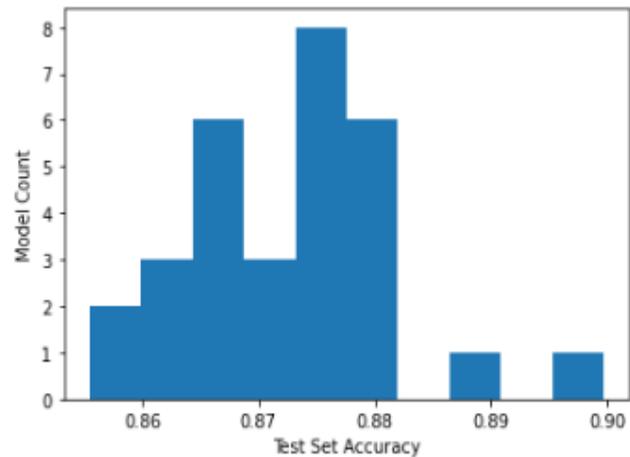


Fig. 11. Multiple Test Accuracy Bar Chart.

Dropout is the method of removing random neurons from the network while training. As a result, the activity of the downstream neurons is temporarily removed from the forward pass, and weight changes do not apply to the backward pass [14]. L2 rate means that it will reduce overfitting with the small size of weight and biases. After analyzing Fig. 12, the best L2 rate was 1.47, and the best dropout was 0.45.

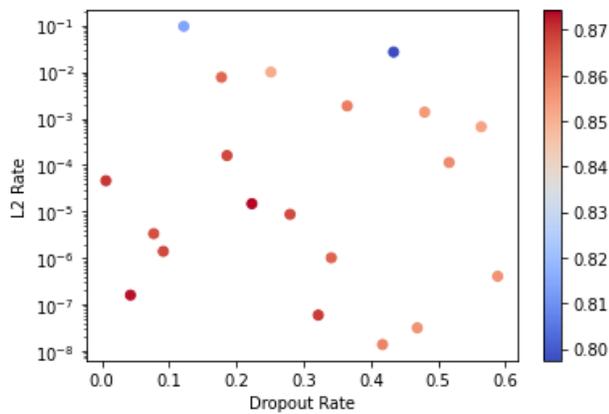


Fig. 12. Networks Dropout Rate Selection.

### VIII. DISCUSSION

Breast cancer is very deadly for women, especially older. Scientists are working very hard to find easy and practical solutions. With the help of modern technology, people are different types of deep learning or machine learning algorithms in medical science. This technology dramatically improves medical treatments as it is very effective, low on cost, and saves much time for doctors and patients.

This paper aims to identify breast cancer from images by using four deep learning algorithms. Those algorithms are NasNet Large, DenseNet-201, Inception-ResNet-V3, and Big Transfer. All of the experimented algorithms provide promising results. The dataset was collected from Kaggle. We worked with total 1693 instance divided into two classes named Malignant and Benign, from here 1354 was used in training, and 339 instances were used in testing. For all the algorithms, the total support instance was 339.

NasNet-Large is a convolutional neural network with more than 1 million images trained from the ImageNet database [28]. Input image size was 331\*331. This analysis shows it obtained 81% of accuracy, which is the lowest out of 4 algorithms.

DenseNet-201[29] is a convolutional neural network with 201 layers. It can add a pre-trained network trained on the ImageNet database. The input image size was 224\*224. This algorithm achieved 89% of accuracy, which is the 2nd highest.

Inception-ResNet-V3 is the descendant from the Inception family [30]. It has some improvements in Label Smoothing, 7\*7 factorized convolutions, and the use of an auxiliary classifier for propagating label information. Here input size was 299\*299. The algorithm provides the third-highest accuracy of 86%.

Significant Transfer is one of the newest in deep learning algorithms [19]. Although it is new, the performance of this algorithm is awe-inspiring. On current analysis, Big Transfer owned first place with 90% accuracy.

### IX. CONCLUSION

This study worked with deep learning methods to predict breast cancer. The proposed method goes through image data selection and preprocessing the data for focusing on the prime

features. DenseNet, NasNet Large, Inception ResNet (v3), m-r101x1x1 (BiT) neural architecture are selected for this study. The architecture selection process was based on their architectural variety and their efficiency in the computer vision sector. The selected public microscopic biopsy dataset is split into two parts for training and testing purposes, where 20% of the data is reserved for evaluating the trained neural network models. This study use pre-trained weights for the benchmark architectures from Keras. This study applies necessary modification in the tail order layers of benchmark models, customizes the neural models for forecasting breast cancer from histological images. Thirty evolution cases occurred for each trained model to finalize the reliable performance rate. The best accuracy of 90% was found once in the thirty evaluations from BiT based m-r101x1x1 architecture which contains 101 layers of neural network. The average accuracy of the Bit based model was found at 87.51%, which is higher than the average accuracy of comparative models. This inaugural work exhibit the probability and promise of working with large data set with more accuracy and larger multicenter studies for ulterior appraise the methods and findings.

### REFERENCES

- [1] M. H. Yap et al., "Breast ultrasound region of interest detection and lesion localisation," *Artif. Intell. Med.*, vol. 107, p. 101880, 2020.
- [2] L.-Q. Zhou et al., "Lymph node metastasis prediction from primary breast cancer US images using deep learning," *Radiology*, vol. 294, no. 1, pp. 19–28, 2020.
- [3] S. Chand, "A comparative study of breast cancer tumour classification by classical machine learning methods and deep learning method," *Mach. Vis. Appl.*, vol. 31, no. 6, pp. 1–10, 2020.
- [4] L. Ahmed, M. M. Iqbal, H. Aldabbas, S. Khalid, Y. Saleem, and S. Saeed, "Images data practices for semantic segmentation of breast cancer using deep neural network," *J. Ambient Intell. Humaniz. Comput.*, pp. 1–17, 2020.
- [5] K. N. Ramanto and A. A. Parikesit, "The usage of deep learning algorithm in medical diagnostic of breast cancer," *Malaysian J. Fundam Appl Sci*, vol. 15, no. 2, pp. 274–281, 2019.
- [6] V. Chaurasia, S. Pal, and B. B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," *J. Algorithm. Comput. Technol.*, vol. 12, no. 2, pp. 119–126, 2018.
- [7] M. A. Islam, S. Akter, M. S. Hossen, S. A. Keya, S. A. Tisha, and S. Hossain, "Risk Factor Prediction of Chronic Kidney Disease based on Machine Learning Algorithms," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 2020, pp. 952–957.
- [8] Z. Huang, X. Zhu, M. Ding, and X. Zhang, "Medical image classification using a light-weighted hybrid neural network based on PCANet and DenseNet," *IEEE Access*, vol. 8, pp. 24697–24712, 2020.
- [9] J. Hai et al., "Fully convolutional densenet with multiscale context for automated breast tumour segmentation," *J. Healthc. Eng.*, vol. 2019, 2019.
- [10] M. S. Hossen, I. Haque, M. S. Islam, M. T. Ahmed, M. J. Nime, and M. A. Islam, "Deep Learning based Classification of Papaya Disease Recognition," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 2020, pp. 945–951.
- [11] D. Zhang, L. Zou, X. Zhou, and F. He, "Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer," *IEEE Access*, vol. 6, pp. 28936–28944, 2018.
- [12] F. M. Alakwaa, K. Chaudhary, and L. X. Garmire, "Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data," *J. Proteome Res.*, vol. 17, no. 1, pp. 337–347, 2018.
- [13] M. Nawaz, A. A. Sewissy, and T. H. A. Soliman, "Multi-class breast cancer classification using deep learning convolutional neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 6, pp. 316–332, 2018.

- [14] R. Platania, S. Shams, S. Yang, J. Zhang, K. Lee, and S.-J. Park, "Automated breast cancer diagnosis using deep learning and region of interest detection (bc-droid)," in Proceedings of the 8th ACM international conference on bioinformatics, computational biology, and health informatics, 2017, pp. 536–543.
- [15] A. Cruz-Roa et al., "Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumour extent," *Sci. Rep.*, vol. 7, no. 1, pp. 1–14, 2017.
- [16] L. Tsochatzidis, L. Costaridou, and I. Pratikakis, "Deep learning for breast cancer diagnosis from mammograms—a comparative study," *J. Imaging*, vol. 5, no. 3, p. 37, 2019.
- [17] S. Ray, A. AlGhamdi, K. Alshouli, and D. P. Agrawal, "Selecting Features for Breast Cancer Analysis and Prediction," in 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), 2020, pp. 1–6.
- [18] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *J. Digit. Imaging*, vol. 32, no. 4, pp. 605–617, 2019.
- [19] A. Kolesnikov et al., "Big transfer (bit): General visual representation learning," in Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16, 2020, pp. 491–507.
- [20] M. N. R. Shuvo, S. Akter, M. A. Islam, S. Hasan, M. Shamsojjaman, and T. Khatun, "Recognizing Human Emotions from Eyes and Surrounding Features: A Deep Learning Approach."
- [21] A. Islam, N. Rahman Shuvo, M. Shamsojjaman, S. Hasan, S. Hossain, and T. Khatun, "An Automated Convolutional Neural Network Based Approach for Paddy Leaf Disease Detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 1, pp. 280–288, 2021, doi: 10.14569/IJACSA.2021.0120134.
- [22] D. M. Vo, N.-Q. Nguyen, and S.-W. Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," *Inf. Sci. (Ny)*, vol. 482, pp. 123–138, 2019.
- [23] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *Ieee Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, 2015.
- [24] S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park, and S. E. Chang, "Classification of the clinical images for benign and malignant cutaneous tumours using a deep learning algorithm," *J. Invest. Dermatol.*, vol. 138, no. 7, pp. 1529–1538, 2018.
- [25] M. Chung, J. Lee, M. Lee, J. Lee, and Y.-G. Shin, "Deeply self-supervised contour embedded neural network applied to liver segmentation," *Comput. Methods Programs Biomed.*, vol. 192, p. 105447, 2020.
- [26] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, "A deep learning mammography-based model for improved breast cancer risk prediction," *Radiology*, vol. 292, no. 1, pp. 60–66, 2019.
- [27] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [28] B. Zoph, V. Vasudevan, J. Shlens and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8697–8710, doi: 10.1109/CVPR.2018.00907.
- [29] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
- [30] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (pp. 4278–4284). AAAI Press.
- [31] Tsochatzidis, L.; Costaridou, L.; Pratikakis, I. Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study. *J. Imaging* 2019, 5, 37. <https://doi.org/10.3390/jimaging5030037>.