International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Harnessing Convolutional Neural Networks for Histopathological Breast Cancer Classification.

Thyagaraj T[1], Keshava Prasanna[2], Hariprasad S A[3]
[1]Department of Electronics and Communication,
BMS Institute of Technology and Management, Visvesvaraya Technological University, Belagavi, India
thyagaraj_tanjavur@bmsit.in

[2]College of Horticulture
Keladi Shivappa Nayaka University of Agricultural and Horticultural Sciences, Shivamogga, India
keshavaprasanna2013@gmail.com

[3]Faculty of Engineering and Technology
JAIN deemed to be university, Bengaluru, India
sa.hariprasad@jainuniversity.ac.in

Abstract—Recent advancements in Convolutional Neural Networks (CNNs) have significantly supported the field of breast cancer discovery using medical imaging. An improvised DenseNet architecture for the classification of histo-pathological breast cancer images is explored in this work. Leveraging the effectiveness of DenseNet in capturing intricate patterns through dense connectivity, our improvised architecture aims achieve high accuracy and efficiency of classification. The model integrates novel features such as optimized bottleneck layers and attention mechanisms, contributing to improved feature extraction and classification capabilities. The improvised DenseNet produced a accuracy of 93.39% on the breakhis dataset. A summary of key findings and future research directions, emphasizing the need of custom CNN models in breast cancer detection is provided.

Keywords- convolutional neural networks, breast cancer, histopathological images

I. INTRODUCTION

One of the biggest worldwide public health concerns is still breast cancer, particularly affecting women. Its early detection and accurate classification are vital for effective treatment strategies and ultimately improving survival rates. The advent of Convolutional Neural Networks (CNNs) in medical imaging, especially in analyzing histopathological images, has revolutionized the field of oncology diagnostics, creating new pathways for the accurate and effective detection of breast cancer The breast cancer diagnosis of is a critical aspect of early intervention and improved patient outcomes. Convolutional Neural Networks (CNNs) have emerged as powerful tools in the arena of medical imaging in recent years, revolutionizing the way breast cancer is detected and classified. The significance of CNNs in the diagnosis of breast cancer lies in their ability to autonomously analyze complex patterns and subtle features within medical images, particularly mammograms and histo-pathological slides.

Traditional methods of diagnosis often relied on human expertise, introducing the potential for subjective interpretation and human error. CNNs, as a subset of deep learning, offer a data-driven and automated approach that enhances accuracy and efficiency in identifying abnormal tissue patterns indicative of cancer. Their capacity to learn hierarchical representations allows CNNs to discern intricate details in medical images, aiding for early detection of lesions, tumors, and abnormalities. CNNs capability to process vast amounts of imaging data efficiently accelerates the diagnostic process, enabling timely interventions and tailored treatment plans.

An architecture that translates the findings into a straightforward pattern of connectivity is proposed. All levels relate to corresponding feature-map sizes directly with each other to guarantee the maximum information flow among connecting layers in the network. Each layer receives extra inputs from all former layers and transmits its individual feature maps to all following layers to maintain the feed-forward nature.

The features are concatenated rather than summated before they are input to the layer. This ensures that the nth layer has ‘n’ inputs consisting of the attribute map of all the preceding convolution layers. This causes (n(n-1))/2 connections in the network.

The DenseNet layers being thin as they have few filters per layer, add small set of feature maps to the collective knowledge of the system and keep the remaining features unchanged.

II. RELATED WORK

A DenseNet, is indicated by its dense connectivity outline, where each layer receives a direct input from all preceding layers. This thick connection facilitates feature reuse, encourages parameter efficiency, and enhances gradient flow through the network is projected by Huang et al. [1].

DenseNet were employed for the histopathological categorization of breast cancer images. The study demonstrated superior performance in difference to traditional methods, showcasing DenseNet's ability to capture complex patterns within tissue samples [2].

The extended use of DenseNet to mammography images, achieving high accuracy in recognizing between benign and malignant lesions [3]. The dense connectivity in DenseNet proved advantageous in acquiring hierarchical representations from diverse image features.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Transfer learning with DenseNet for breast tumor classification via mammograms was presented. The pre-trained model significantly improved classification accuracy, emphasizing the magnitude of leveraging existing knowledge for medical image analysis [4].

Integrated radiomic features with DenseNet-based models, achieving improved performance in distinguishing between different breast cancer subtypes was demonstrated [5]. This integration showcases the versatility of DenseNet in handling multimodal data for comprehensive diagnosis.

Zhang [6] proposed a hybrid architecture combining DenseNet with attention mechanisms for breast cancer classification. The study demonstrated improved interpretability and highlighted the potential of hybrid models for better capturing relevant features, etc.

III. ARCHITECTURE

A. DenseNet

The design of DenseNet which differentiates from other CNN models is shown in Fig 1.

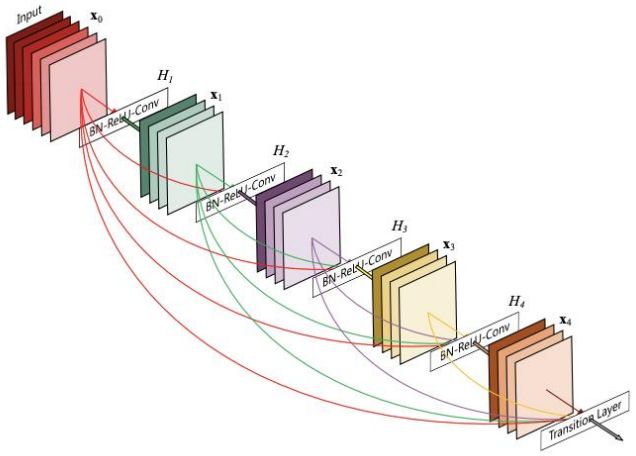

Fig 1: DenseNet architecture
(Courtesy: https://www.researchgate.net/figure/Five-layers-of-a-DenseNet-block-with-a-growth-rate-of-4-feature-maps-per-layer-source_fig1_320170916)

Densely Connected Convolutional Networks, is a profound learning architecture introduced by Huang et al. [8]. It addresses challenges associated with information flow and vanishing gradient problems in very deep networks by introducing dense connectivity between layers.

Unlike traditional architectures where each layer connects only to its successive layer, DenseNet connects each layer to all subsequent layers in a dense and direct manner. This dense connectivity is accomplished by concatenating the feature maps from all earlier layers, allowing each layer to gather direct input from all its predecessors. Mathematically, for a known layer *l*, the output is calculated as (1):

$$xl = Hl([x0, x1, \ldots, xl-1]) \quad (1)$$

Here, [*x*0,*x*1,…….,*xl*−1] denotes the chain of characteristic maps from all prior layers, and *H*l represents the composite function of convolution, batch standardization, and non-linearity for layer *l*.

Feature mining is a fundamental technique that maximizes classification accuracy[7-10]. DenseNet is provided, which makes use of dense connections between layers, lowers the quantity of parameters, improves propagation, and promotes feature reuse [11].

[18-20] discusses the application of Convolutional Neural Networks (CNNs) for breast cancer histopathological image classification

DenseNet introduces bottleneck layers to enhance computational efficiency. These bottleneck layers consist of a 1x1 convolution followed by a 3x3 convolution. The 1x1 convolution lowers the number of input channels before the 3x3 convolution, reducing the computational burden while maintaining expressive power. Mathematically, the operation in a bottleneck layer can be expressed as (2):

$$yl = Hl([x0, x1, \ldots, xl-1]) = \mathrm{Fl}(Wl \cdot [x0, x1, \ldots, xl-1]) \quad (2)$$

Where Fl represents the bottleneck layer function, Wl denotes the weights, and yl is the output of layer l.

Multiple dense blocks, each made up of numerous layers with dense connections, make up DenseNet. The dense blocks make it easier to understand complex patterns and encourage the reuse of features. Mathematically, a dense block with L layers can be expressed as (3):

$$xl + 1 = Hl([x0, x1, \ldots, xl]) \quad (3)$$

For each layer in the dense block, the yield is achieved by concatenating the feature maps beginning from all former layers.

To control the growth parameters and computational cost, DenseNet incorporates transition layers between dense blocks. These transition layers consist of a 1x1 convolution pursued by 2x2 average pooling. The 1x1 convolution reduces the number of passages, and pooling reduces the spatial dimensions. Mathematically, the transition layer operation can be represented as (4):

$$Zl = \Theta l([x0, x1, \ldots, xl]) \quad (4)$$

Where Zl is the yield of the transition layer, and Θl represents the composite function of 1x1 convolution and average pooling.

The transition layers are meant to aggregate the characteristic maps from a dense block and reduce its dimensions.

DenseNet employs global average pooling before the decisive fully connected layer for classification. GAP reduces the spatial dimensions to 1x1, and the ensuing feature maps are fed into a SoftMax tier for classification. The classification can be expressed as (5):

$$pred = \mathrm{Softmax}\big(Wfc * \mathrm{Pool}([x0, x1, \ldots, xL])\big) \quad (5)$$

Here, pred is the predicted probability distribution, Wfc represents the weights of the fully connected layer, and Pool denotes global average pooling.

B. ResNet

The Residual Network addressing the disappearing gradient problem linked with increasing network depth.

Traditional networks learn the underlying mapping functions directly, but residual learning focuses on learning

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

residuals—differences between the input and the desired output. The yield of each residual block is computed as the sum of the input and the learned residual as (6):

$$y = Q(x) + x \tag{6}$$

Here, x is the input to the block, Q(x) is the residual function to be learned, and y is the output.

Two convolutional layers with batch normalization and Rectified Linear Unit (ReLU) activation algorithms make up the residual cell block. By enabling the gradient to pass through the network directly, the skip connection helps to mitigate the vanishing gradient issue. Mathematically, the residual block operation is given by (7):

$$y = Q2(Q1(x)) + x \tag{7}$$

Here Q1, Q2 represents the two consecutive convolutional layers.

In ResNet, the skip network can take two forms: identity shortcut or projection shortcut. The identity shortcut is used when the input and output proportions are the same, and the projection shortcut is employed when dimensions differ. Mathematically, the identity shortcut is (8):

$$y = Q2(Q1(x)) + x \tag{8}$$

The projection shortcut introduces a linear transformation Wsx to match dimensions:

$$y = Q2(Q1(x)) + Wsx \tag{9}$$

The layer depth is achieved through the repetition of residual extensions and the down sampling of feature representations operating convolutional layers and max pooling.

ResNet employs global average pooling (GAP) before the last completely connected layer for classification. GAP reduces the spatial dimensions to 1x1, and the subsequent feature maps are flattened for input into the SoftMax layer (10):

$$\mathrm{Y} = \mathrm{Softmax}(Wfc \cdot \mathrm{GAP}(x)) \tag{10}$$

Here, Y is the predicted probability distribution, Wfc represents the weights of the fully connected layer, and GAP denotes global average pooling.

Resnet makes information preservation explicit through additive identity transformations.

IV. Methodology

The methodology followed is displayed in Fig 2. The processes include data collection and exploration. The preprocessing includes standardizing the images size and normalize pixel values to a common scale and augmentation to improve the dataset variability.

The normalization also includes the process of stain normalization. Since the images obtained are stained with H&E, and the process of imagery being different, there exists a need for stain normalization to bring all the test and validation samples to the common scale.

The quantity of samples available in the model are low when compared to the number of samples required to train a CNN model for optimal accuracy, augmentation techniques are employed which includes rotation, flip, picture-in-picture and other common techniques to surge the sample size.

The following step is to derive the fitting model for the classifier. The purported model has a modified DenseNet-201 architecture. The model employs four dense blocks and three transition layers.

The model is trained and verified on the data for its categorization performance. The hyper parameter tuning is the ensuing step to upgrade the performance metrics. The learning rate and the batch size alongside the total epochs are critical parameters that dictate the training time and precision of the model.

Selection of the optimizer is of critical importance as they decide the evaluation results. Adam and SGD optimizers were used. Metrics used with their corresponding formulae:

Accuracy measures the overall correctness of the model predictions (11).

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions} \tag{11}$$

A Confusion Matrix is a table that summarizes the model's performance. It shows the counts of True Positives, True Negatives, False Positives, and False Negatives.

Accuracy is the metric considered for assessment of the model.

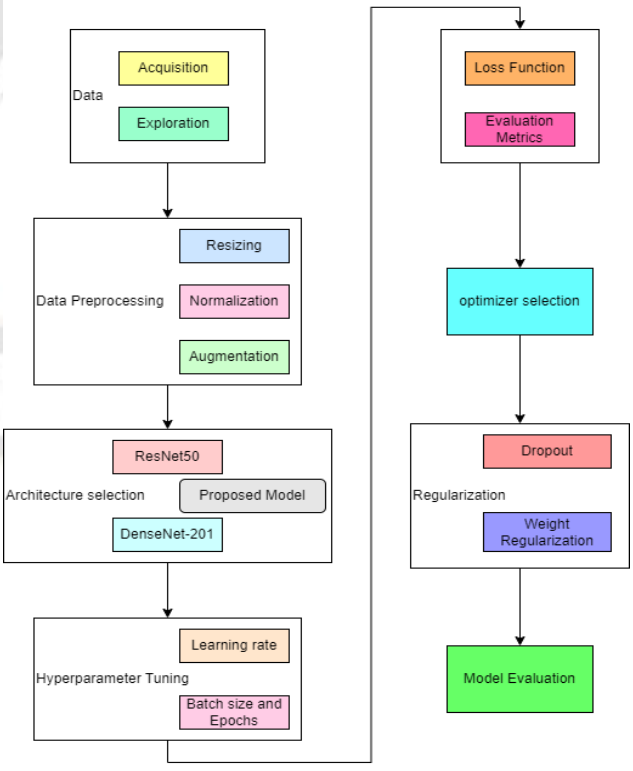

Fig 2: Methodology followed.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Fig 3 shows the layers in the proposed model.

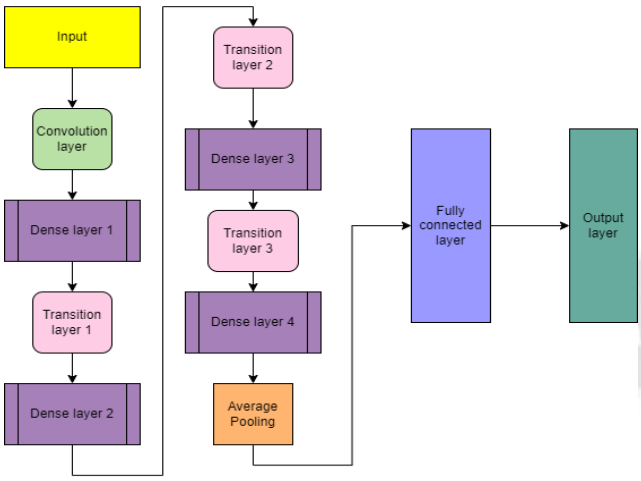

Fig 3: Block diagram of the proposed improvised model.

V. DATASET

BreakHis: The BreakHis dataset is a histopathological image dataset designed for the development and evaluation of machine learning algorithms for breast cancer classification. The key characteristics, composition, and significance of the BreakHis dataset. The Breast Cancer Histopathological Image Classification (BreakHis) is composed of 9,109 high-resolution microscopic images of breast tissue specimens, providing a diverse and comprehensive representation of various breast cancer subtypes of breast lump tissue assembled from 82 patients.

The images are amassed using different magnifying considerations (40X, 100X, 200X, and 400X).

They contain 2,480 benign and 5,429 malignant samples (700X460 pixels, 3-channel RGB, 8-bit depth in each channel, PNG format).

These images were put in using the method of procedure biopsy and the class of the tumor, type of tumor with patient identification details along with magnification factor are furnished in the naming nomenclature.

BreakHis images are accompanied by expert annotations indicating regions of interest (ROIs) within the tissue samples. These annotations identify areas relevant to the presence of malignant cells, facilitating the training and validation of machine learning models.

Sample images from BreakHis dataset is presented in fig 4.

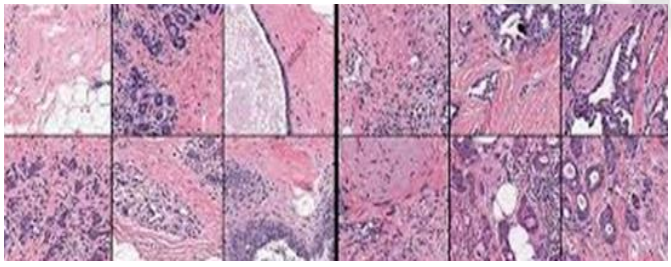

Fig 4: Sample BreakHis dataset, benign and malignant images.

VI. RESULTS:

1) *MODEL 1: RESNET WITHOUT AUGMENTATION*

The Resnet50 model was trained and verified on BreakHis dataset. The model configuration is shown in figure 5.

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 7, 7, 2048)	23587712
global_average_pooling2d (Gl	(None, 2048)	0
dropout (Dropout)	(None, 2048)	0
batch_normalization (BatchNo	(None, 2048)	8192
dense (Dense)	(None, 3)	6147

Total params: 23,602,051
Trainable params: 23,544,835
Non-trainable params: 57,216

Fig 5: Outline of the model1

The total parameters managed by the model are 23,602,051. Of the total parameters, 23,544,835 parameters were trainable parameters while 57,216 were non-trainable parameters.

The model was tutored on the Breakhis dataset without augmentation. Fig 6 projects the results of the model1's performance.

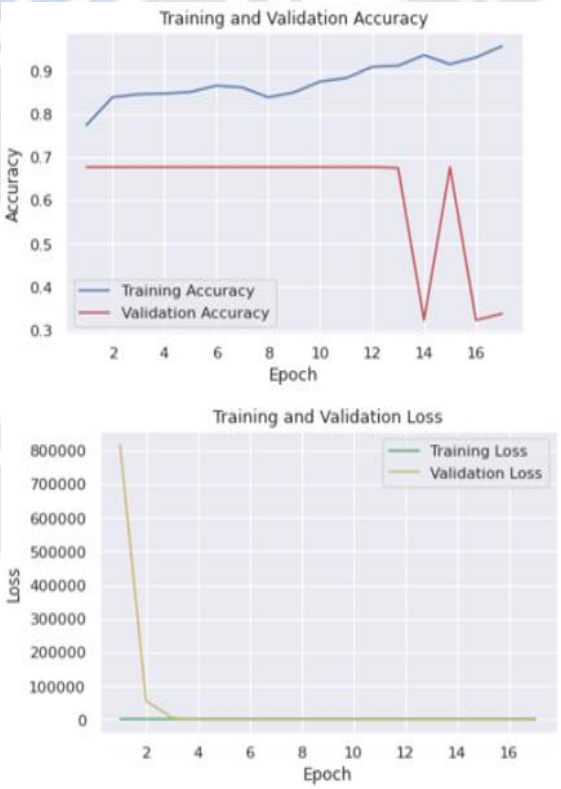

Fig 6: graph of Accuracy vs Epoch followed by Loss vs Epoch
The training accuracy reached 96% for 16 epochs, the validation accuracy was 67.5%.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

2) MODEL 1: RESNET WITH AUGMENTATION

The Resnet50 model with the identical configuration was trained on the augmented BreakHis dataset.

Fig 7 projects the results of the performance of the model1 with information augmentation.

A 90-degree rotation range , random horizontal and vertical flipping of images in the training dataset with a range of 2 set for random zoom and the range for brightness set between 0.5 and 1.5 combined with scaling of images of both training and test dataset was performed.

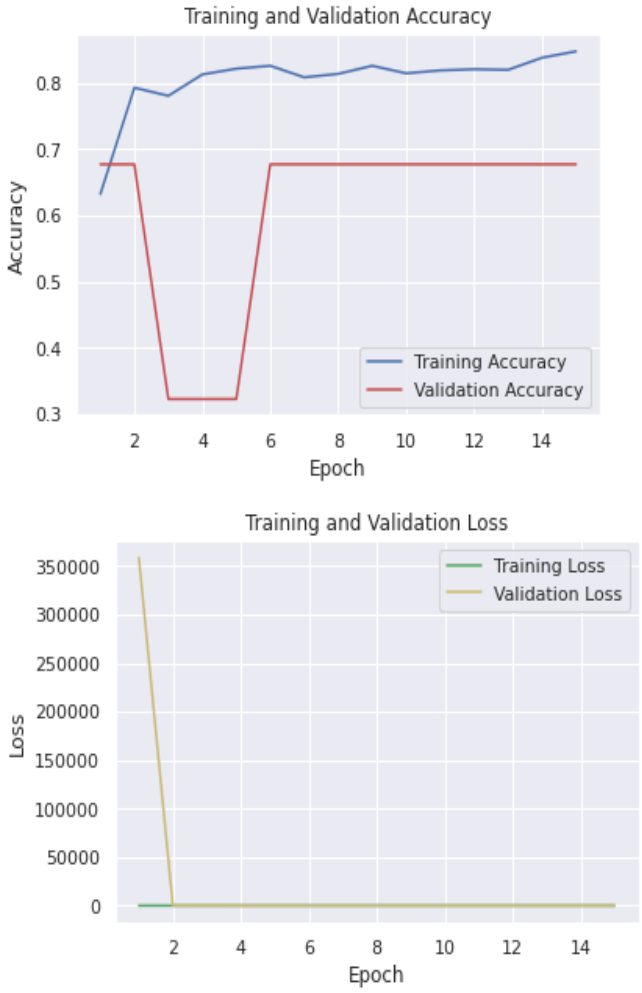

Fig 7: graph of Accuracy vs Epoch followed by Loss vs Epoch

The training accuracy reached 85% for 14 epochs, the validation accuracy was 67.5%.

B. MODEL 2: DENSENET

The DenseNet model was trained and tested on the BreakHis dataset. The model configuration is shown in Fig 8 .

The DenseNet has pre-trained weights with image data augmentation and early stopping as callback:

Layer (type)	Output Shape	Param #
densenet201 (Functional)	(None, 7, 7, 1920)	18321984
global_average_pooling2d (Gl	(None, 1920)	0
dropout (Dropout)	(None, 1920)	0
batch_normalization (BatchNo	(None, 1920)	7680
dense (Dense)	(None, 3)	5763

Total params: 18,335,427
Trainable params: 18,102,531
Non-trainable params: 232,896

Fig 8: Summary of the DenseNet201 model.

The total parameters used by the model is 18,335,427. Of the total parameters, 18,102,531 parameters were trainable parameters while 232,896 were non-trainable parameters.

One hyper-parameter that regulates how much the model is altered in response to the predicted error after each update of the model weights is the learning rate. Selecting the learning rate can be difficult since too small of a value could lead to an extended training process that could become stuck, while too large of a value could cause the training process to become unstable or acquire a suboptimal set of weights too quickly. A constant 0.0001 was used as the learning rate.

A GlobalAveragePooling layer is used. It directs average pooling on the spatial proportions till every spatial dimension is one and allows other dimensions unchanged.

It is ensued by a dropout layer set to 0.5 which means that there is a 50% chance that the output of a given neuron is forced to zero , this layer is used to prevent over-fitting.

Batch normalization is followed by a dense layer with 2 neurons for 2 output classes i.e. benign and malignant.

SoftMax is applied as the activation function. An activation function called SoftMax is used to convert numbers or logits into probabilities. A vector (let's say v) containing the probabilities of every possible result is the result of a SoftMax. Vector v's probabilities add up to one for every possible outcome or class. It is mathematically defined as (12):

$$S(yi) = \frac{\exp(yi)}{\sum_{j=1}^{n} \exp(yj)} \quad (12)$$

Where: y is an input vector to the SoftMax function S and consists of n elements for the n classes, y_i is the i^{th} element of the input vector.

$\text{Exp}(y_i)$ is the standard exponential function applied on y_i for values of $y_i<0$, the exponential value is very close to zero but not zero and when y_i is large, the exponential value is also large.

$\sum_{j=1}^{n} \exp(yi)$ value ensures the value of the output vector S(yi) adds to 1 for the i^{th} class and in the range of 0 to 1 for valid probability distribution.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Fig 9 depicts the outcome of the performance of model 2.

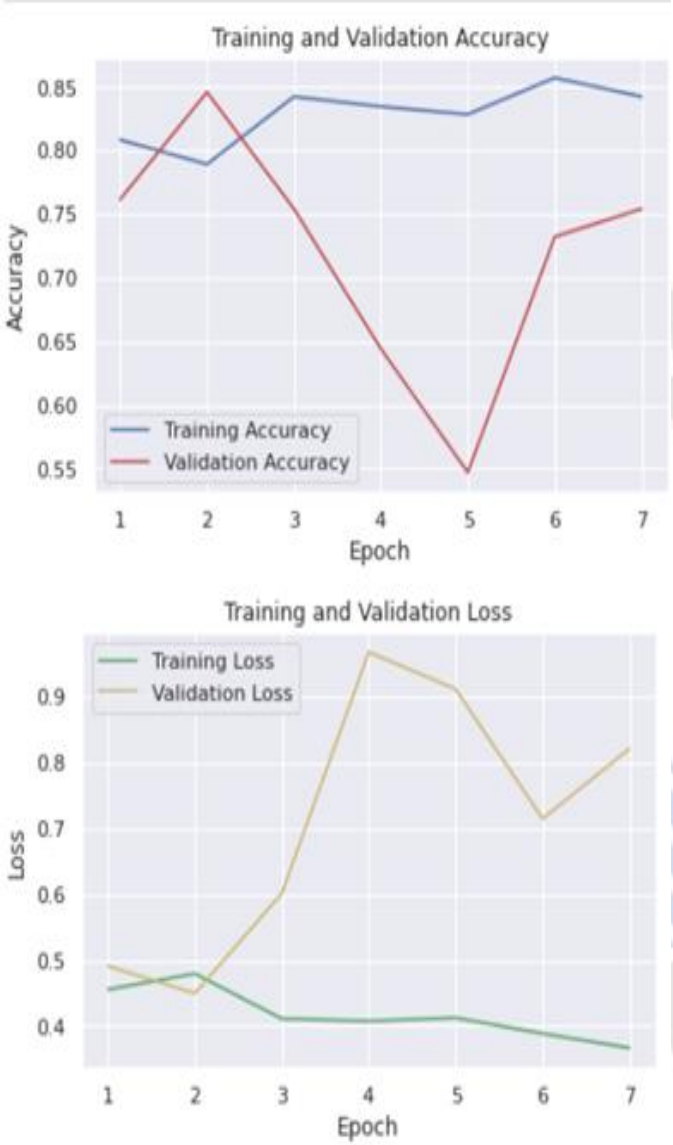

Fig 9: graph of Accuracy vs Epoch followed by Loss vs Epoch

The accuracy of this model is around 75.41% .

1) MODEL 2.1: DENSENET WITH EARLY STOPPING AS CALLBACK

Training will stop when the chosen performance measure stops improving. In the case of this model the chosen performance parameter is validation accuracy. Often, the first indication of no further advancement may not be the best time to halt training. This is because the model may coast into a plateau of no progress or even get faintly worse before progressing much better. We can account for this by adding a pause to the trigger in terms of the number of epochs on which we would like to see no improvement. This can be done by setting the "patience" argument.

The modification to the model generated the following results.

Fig 10, 11 display the results of the performance of model 2.

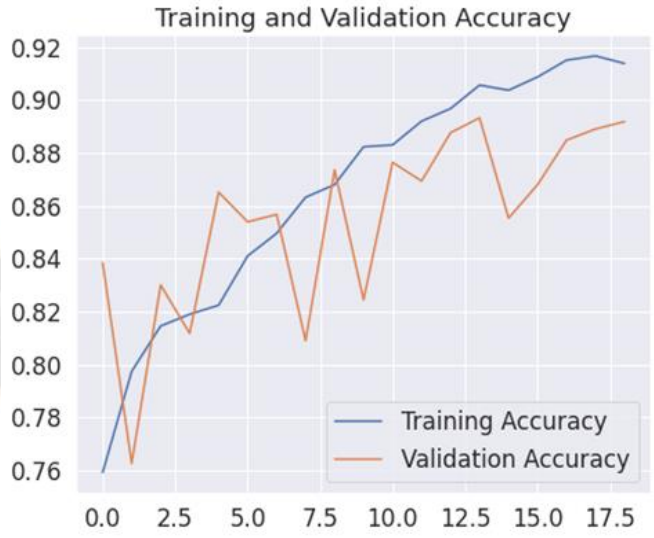

Fig 10: graph of training and validation accuracy

Fig 11: graph of training and validation loss.

The CNN model-2 is 88.5% accurate in predicting the cancer

2) MODEL 2.2: DENSENET WITH REDUCELRONPLATEAU AND MODELCHECKPOINT AS CALLBACKS

ReduceLROnPlateau : Lower the learning rate if a metric no longer improves. Once learning stalls, models frequently gain from lowering the learning rate by a factor of 2–10. This callback keeps an eye on a quantity and lowers the learning rate if a certain number of epochs (referred to as "patience") show no improvement.

Here, the validation accuracy is monitored and after 5 epochs with no improvement the learning rate will be reduced with a factor of 0.2 to a minimum of 0.0000001.

ModelCheckpoint: callback is used in combination with training using model.fit() to save a model or weights (in a checkpoint file) at some interval, so the model or weights can be loaded later to continue the training from the state saved. The

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

validation accuracy is monitored and only the best performance is saved.

The performance of the model is shown in Fig 12.

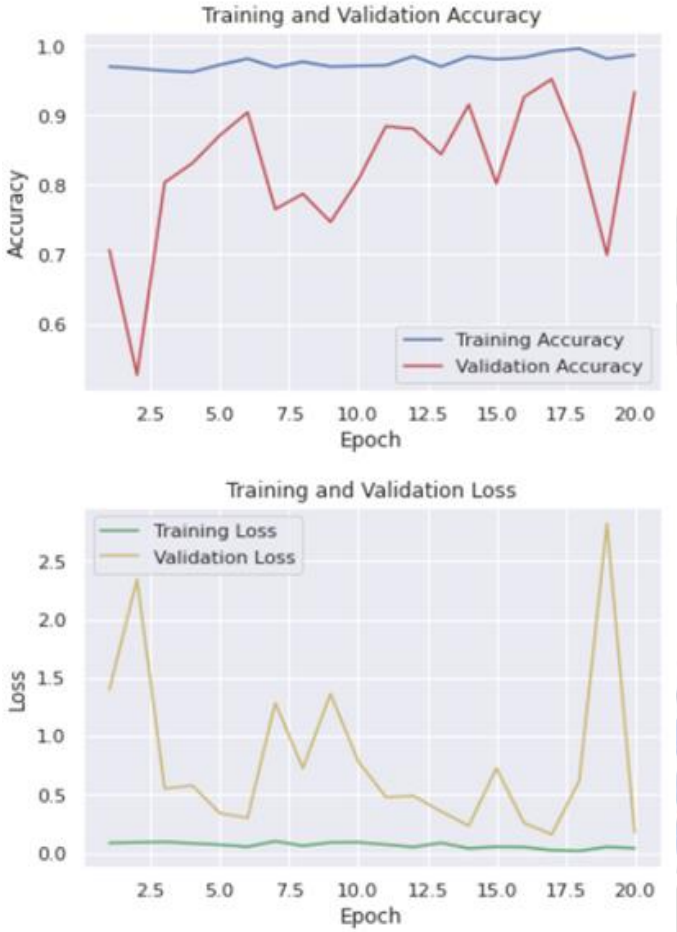

Fig 12: graph of Accuracy vs Epoch followed by Loss vs Epoch
The accuracy of this model is around 93.39% for 20 epochs.

VII. CONCLUSIONS:

Table 1: Training and validation accuracy of the proposed models.

BreakHis dataset	**Training Accuracy(%)**	**Validation Accuracy(%)**
Model 1 without data augmentation.	96	67.5
Model 1 with data augmentation.	85	67.5
Model 2 base version	85	75.41
Model 2.1 with early stopping call back	90.77	88.5
Model 2.2 with Reduce on plateau and model checkpoint call back	95	**93.39**

Table 1 shows the accuracy of the model with different modifications. The model 2.2 delivered the highest validation accuracy of 93.39%.

Table 2: comparison of training and validation accuracy of the proposed models with the models from literature.

BreakHis dataset	**Validation Accuracy(%)**
Alexnet[10]	92.68
ResNet[11]	98.57
VGG16[12]	97.85
Inception V3[13]	97.84
SqueezeNet[14]	97.56
DenTnet[15]	99.28
Proposed Model -2	**93.39**

Table 2 compares the performances of different models used for classifying the BreakHis dataset.

- AlexNet: Achieved a validation accuracy of 92.68%.
- ResNet: Outperformed other models with a validation accuracy of 98.57%.
- VGG16: Demonstrated a strong performance with a validation accuracy of 97.85%.
- Inception V3: Achieved a competitive validation accuracy of 97.84%.
- SqueezeNet: Showed good performance with a validation accuracy of 97.56%.
- DenTnet: Outperformed all other models with the highest validation accuracy of 99.28%.
- Proposed Model -2: Achieved a validation accuracy of 93.39%.

These results highlight the effectiveness of deep learning models in classifying breast cancer histopathological images, with DenTnet standing out as the top-performing model on the BreakHis dataset.

The split of the dataset for training and test/ validation plays a valuable role in establishing the accuracy of the model.

The ResNet, VGG16, Inception V3, SqueezeNet, and DenTnet models perform a overfit to the data and hence produce high accuracy values.

The dense connectivity of DenseNet and the residual learning mechanism of ResNet undeniably contribute to their robust feature extraction capabilities, allowing for nuanced discrimination crucial for accurate classification.

DenseNet's emphasis on feature reuse and parameter efficiency is particularly advantageous for tasks requiring intricate pattern recognition.

ResNet's residual learning approach mitigates vanishing gradient issues, providing stability in training and facilitating the construction of deeper networks.

REFERENCES

[1] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).
[2] Li, H., et al. (2018). Deep learning for predicting progression of microvascular disease in patients with diabetes. Diabetes Research and Clinical Practice, 142, 1-9.
[3] Ren, J., et al. (2019). Breast Cancer Diagnosis via a Robust and Fully Automated Deep Learning Approach. Nature Communications, 10(1), 4875.
[4] Cheng, J., et al. (2020). Transfer Learning with DenseNet for Mammographic Mass Classification. Computational and Mathematical Methods in Medicine, 2020.
[5] Wang, Y., et al. (2021). Integrating Radiomic Features with a DenseNet-Based Convolutional Neural Network for Breast Cancer Subtype Classification. Frontiers in Oncology, 11, 674186.
[6] Zhang, S., et al. (2022). A Hybrid Architecture of DenseNet and Attention Mechanism for Breast Cancer Classification. Journal of Healthcare Engineering, 2022.
[7] Ewees, L. Abualigah, D.Yousri et al., "Improved Slime Mould Algorithm Based on Firefly Algorithm for Feature Selection: A Case Study on QSAR Model." Engineering with Computers, vol. 38, pp. 1–15, 2021.
[8] Ewees, Z. Y. Algamal, L. Abualigah et al., "A cox proportional-hazards model based on an improved aquila optimizer with whale optimization algorithm operators." Mathematics, vol. 10, no. 8, p. 1273, 2022.
[9] M. Abd Elaziz, A. A. Ewees, D. Yousri, L. M. Abualigah, and M. A. A. Al-qaness, "Modified marine predators algorithm for feature selection: case study metabolomics." Knowledge and Information Systems, vol. 64, no. 1, pp. 261–287, 2022.
[10] L. Abualigah, K. H. Almotairi, M. A. Al-qaness et al., "Efficient text document clustering approach using multi-search Arithmetic Optimization Algorithm," Knowledge-Based Systems", vol. 248, Article ID 108833, 2022.
[11] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks." in Proceedings of the Conference on Computer Vision and Pattern Recognition. (CVPR), pp. 2261–2269, Honolulu, HI, USA, July 2017.
[12] Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25" in Proceedings of the 26th Annual Conference on Neural Information Processing Systems 2012, pp. 1106–1114, Lake Tahoe, Nevada, USA, October 2012.
[13] He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks" in Proceedings of the European Conference on Computer Vision (ECCV), vol. 9908, pp. 630–645, Netherlands, Europe, 2016.
[14] Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition." in Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, July 2015.
[15] Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions." in Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9, Boston, MA, USA, July 2015.
[16] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size." CoRR, 2016
[17] M.A.Wakili, H.A. Shehu et al.," Classification of Breast Cancer Histopathological Images Using DenseNet and Transfer Learning." Advanced Deep Learning and Neuro-Evolution Metaheuristic Techniques in Medical Applications, Volume 2022, Article ID 8904768
[18] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770-778).
[19] Spanhol, F. A., Oliveira, L. S., Petitjean, C., Heutte, L. (2016). Breast cancer histopathological image classification using Convolutional Neural Networks. In Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) (pp. 568-576).
[20] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2097-2106).