ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

Breast Cancer Prediction using ResNet50

M .Prasanna[1], Dr.M.Thamarai[2] and Dr.S.P.Malarvizhi[3]

Department of ECE, Sri Vasavi Engineering College, Andhrapradesh, India[1]

Department of ECE, Sri Vasavi Engineering College, Andhra Pradesh India[2]

Department of CSE, Sri Vasavi Engineering College, Andhra Pradesh India[3]

Email:thamarai.muthusamy@srivasaviengg.ac.in

Abstract: Breast cancer is one of the most feared and frequently occurring cancers in the society especially for women. But the prediction of the cancer in early stages is still challenging. Early diagnosis significantly increases the chances of correct treatment and survival, but this process is tedious and often leads to a disagreement between pathologists. Deep learning (DL) plays a vital role in predicting cancers at earlier stage. This paper uses convolution neural network - ResNet50 to predict breast cancer at the initial stage itself with a few hundreds of sample images. This proposed work is to be carried out in the environment of Google Co-lab. The performance of the proposed method is analyzed by using different optimizers with different dropout rates. The maximum accuracy and AUC obtained are 0.986 and 0.98 respectively for the Adam optimizer with the dropout rate of 0.4

Keywords: Breast cancer diagnosis, Deep Learning, Convolution Neural Network, ResNet50, Transfer learning, Histological analysis, Fine Needle Aspiration.

I. INTRODUCTION

Cancers have become one of the major public health issues. According to statistics by the IARC (International Agency for Research on Cancer) from the WHO (World Health Organization), and GBD (Global Burden of Disease Cancer Collaboration), cancer cases increased by 28% between 2006 and 2016, and there will be 2.7 million new cancer cases emerging in 2030 [1]. Among the various types of cancer, breast cancer is one of the most common and deadly in women. Approximately 1 in 8 women (13%) will be diagnosed with invasive breast cancer in their lifetime and 1 in 39 women (3%) will die from breast cancer [2].

Therefore, the diagnosis of breast cancer has become very important. Although the diagnosis of breast cancers have been performed for more than 40 years using X-ray, MRI (Magnetic Resonance Imaging), ultrasound etc. now a days breast tissue biopsies allow the pathologists to histologically assess the microscopic structure and elements of the tissue using Computer aided diagnosis for cancer prediction.

The Breast Cancer cells mainly are of two types Benign and Malignant. The non cancerous cells are called benign cells and they do not disturb nearby cells or spread to other cells. They look like hard (tumor). A benign tumor is not making problem unless it is pressing on nearby tissues, nerves or blood vessels and causing damage. The cancerous cells are called malignant cells. These cells spread the cancer to nearby cells easily. Some cancer cells can move into the bloodstream or lymph nodes, where they can spread to other tissues within the body.

First, we give an overview of the tissue preparation and staining processes of histological slides after taking the biopsy from the patient. First the suspected breast tumor excisions or biopsies are performed in the operating room as shown in Fig.1. using Fine needle aspiration (FNA).

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 7, Issue 9, September 2020

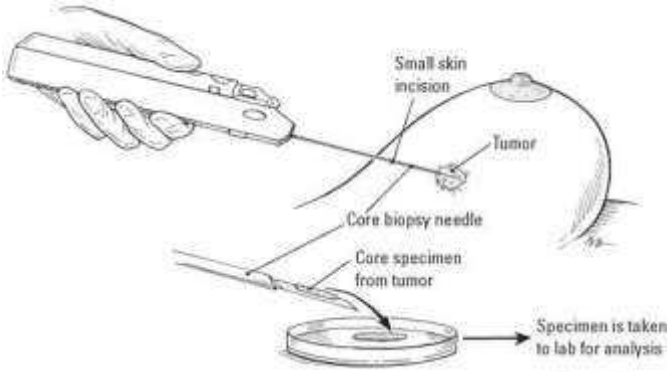


Fig.1. Fine needle aspiration biopsy process[3].

FNA Biopsy is a simple procedure that involves injecting a thin needle through the skin to collect sample fluid or tissue from the cyst or solid mass, as shown in Fig.1. The sample of cellular material taken during an FNA is then sent to a pathology laboratory for analysis. FNA biopsies are performed with less painful and quicker method when compared to tissue sampling such as surgical biopsy for cancer tests.

In pathology lab [3], the first step of the tissue preparation process is formalin fixation and embedding in paraffin. From the paraffin blocks, sections with a thickness of 3-5 µm are cut using a microtome (a high precision cutting instrument) and mounted on glass slides. The structure of interest in the tissue, the nuclei and cytoplasm would be clearly visible after adding dye with the strains. The standard staining protocol uses Hematoxylin and Eosin. Hematoxylin binds to DNA and thereby dyes the nuclei blue/purple, and Eosin binds to proteins and dyes other structures (cytoplasm, stroma, etc.) pink.

Fig. 2A shows the H&E stained breast biopsy image. Immuno Histo Chemistry (IHC) is a more advanced staining technique, which makes use of antibodies to highlight specific antigens in the tissue shown in Fig.2.B and 2.C. In breast cancer, IHC is commonly used to highlight the presence of estrogen (ER), progesterone (PR) and human epidermal growth factor 2 (HER2) receptors, as well as to assess the proliferation of the tumor.

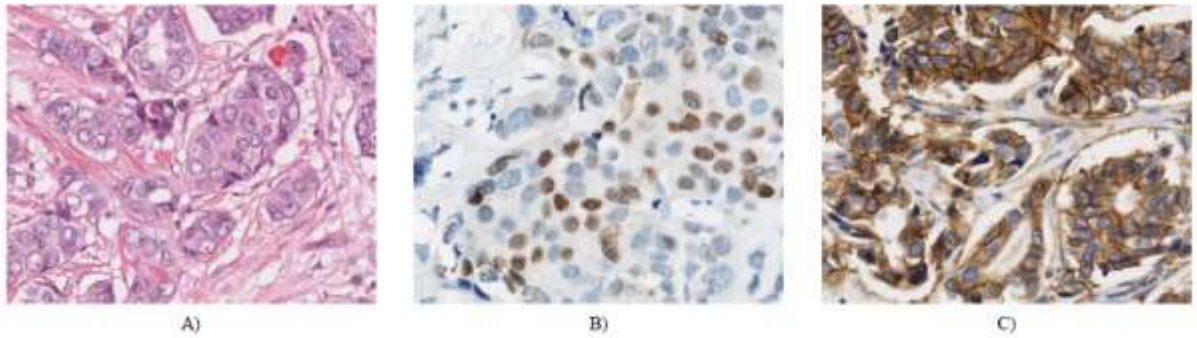


Fig.2 Example histological stains. A)Hematoxylin and Eosin. B)IHC staining for estrogen. C)IHC straining for HER2[3].

The samples taken are examined by a pathologist under a microscope. A detailed report will then be provided about the type of cells that were seen, including any suggestion that the cells might be cancer.

The Surveillance, Epidemiology, and End Results (SEER) summary staging system is used for descriptive and statistical analysis of tumor registry data. According to the SEER summary staging system [2] the stages are classified into 4.

• Situ stage refers to the presence of abnormal cells that are confined to the layer of cells where they originated.

• Local stage refers to invasive cancer that is confined to the breast.

• Regional stage refers to cancer that has spread to surrounding tissue and/or nearby lymph nodes.

• Distant stage refers to cancer that has spread to distant organs and/or lymph nodes, including nodes above the collarbone.

CNNs are multilayered neural networks (Deep learning network) that have more number of hidden layers. It is mainly used for image classification and image recognition applications. Many Researchers have proved the ability of CNN in cancer histology image classification. Deep learning is the advanced

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 7, Issue 9, September 2020

version of neural networks. It has at the maximum of 128 hidden layers between input and output layers.

For Machine learning algorithms, features for classification have to be provided, but the deep learning networks take image as input and extract features automatically and perform classification. Deep learning algorithms reduce the classification error below the human error rate (5%). Its classification error rate is less than 5%. CNNs are multilayered neural networks (Deep learning networks) with more number of hidden layers. It is mainly used for image classification and image recognition applications. It extracts the image features in various layers like human brain processing.

ResNet50 is a subclass of convolution neural networks most popularly used for image classification. Skip connections/Identity are a short cut connection that skips one or more layers in the forward path as shown in Fig.3. Skip Connections are used to explicitly copy features from earlier layers (lower layers) into later layers (Higher layers). This helps the higher layer to perform at least as good as the lower layer, and not worse. This ResNet structure minimizes vanishing gradient problem in CNNs.

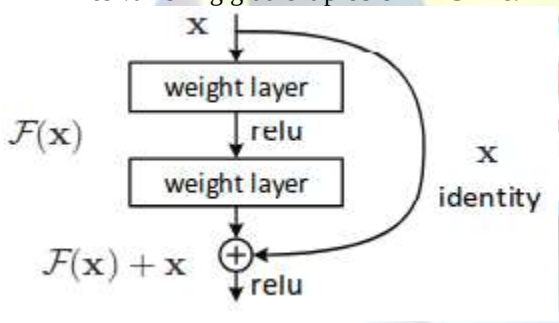


Fig.3. ResNet structure

Kaiming He et al.[4] proved that residual networks (ResNet) are easy to optimize and can gain considerable accuracy when compared to plain networks for image recognition.
The proposed method uses, ResNet50 for image breast cancer prediction using histologist image stains of cancer tissues.

II RELATED WORK

There are many methods developed for the digital pathology image analysis, from rule-based to machine learning approaches. Recently, deep learning based approaches were shown to outperform conventional machine learning methods in many image analysis tasks, automating end-to-end processing. The CNN has performed well in many image classification tasks [5]. CNN learns features from whole images directly [6]. The usage of pre trained CNN model and transfer learning concept is explained in [7,8].

Sara Hosseinzadeh et.al [9] proposed an ensemble deep learning based approach for binary classification of breast histology images. They used three pertained CNNs namely, VGG-19, Mobilenet, and Desnet in their system for prediction and proved that the multimodel ensemble method provides better predictions than single classifiers and Machine Learning algorithms. A.Kalinin 2017 [10], invented biopsy techniques that are still the main methods relied on to diagnose breast cancer correctly. Pierre Baldi 2011 [11] proposed a new auto encoder network structure to apply non-linear transformations to features in histopathological images of breast cancer extracted by the Inception_ResNet_V2 network. This effectively maps the extracted features to a lower dimensional space. The newly obtained features are then used as input for the classical clustering algorithm known as K-means to perform clustering analysis on histopathological images of breast cancer.

Juanying Xie et.al [12] analyzed InceptionV3 and Inception-ResNet-V2 –CNN structures trained with transfer learning for H&E biopsy images for breast cancer prediction. They trained the output layer of the model for both binary and multi classification of histopathological images of breast cancer and proved that ResNet-V2 is more suitable for the analysis of histopathological images than the Inception V3 network. Alexander Rakhlin et.al [13] used Gradient Boosted Tree classifier along with several deep NN architectures (CNN) for H&E strained breast histology microscopy image dataset. They trained the networks with the strong data augmentation with Ligh GBM .Their proposed model performs single and multi classification. The two classification model identifies benign and malignant cancer cells and four classification the model identifies normal, benign, in situ carcinoma and invasive carcinoma

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 7, Issue 9, September 2020

cells. For two classification task the accuracy was 93.8% and AUC was 97.3%.

Teresa Araujo et.al [14] proposed a method using CNN and SVM classifier. They have obtained classification accuracies of 77.8% for four class prediction and 83.3% for carcinoma and non carcinoma Histology image classification. Shuyue Guan,Murray Loew [15] used mammographic images from the two databases MIAS and DDSM and applied transfer learning in CNN for breast cancer prediction. Firstly, they tested three training methods on MIAS: 1) trained a CNN from scratch, 2) applied the pre-trained VGG-16 model to extract features from input images and used these features to train a Neural Network (NN)-classifier, 3) updated the weights in several last layers of VGG-16 model by back-propagation (fine-tuning) to detect abnormal regions.

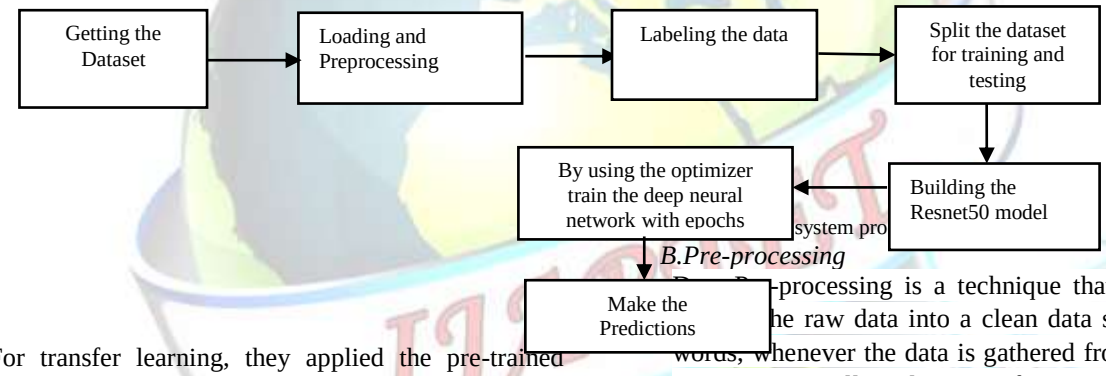
For transfer learning, they applied the pre-trained VGG-16 model in their study. They proved that method 1 and 2 have slight variation in classification accuracy but the method 1 needs more training time than the method 2.Shallu & Rajesh mehra [16] used VGG16, VGG19 and ResNet50 for their H&E image analysis with different training and testing data splitting and proved that VGG16 with logistic regression yielded best performance (accuracy) for 90%-10% training testing data split ratio.

III PROPOSED METHOD

The process flow of the proposed method is shown in Fig.4. The process is explained as follows**.** The H&E stain biopsy images are taken from dataset [17]. The BreaKHis database is introduced in this work. The database is composed of 7,909 image samples generated from breast tissue biopsy slides, which are stained with H&E. The images are divided into benign and malignant tumors based on the aspect of the tumor cells under the microscope.

A.Loading

Load the required libraries and dataset to the working directory in google Co-Lab. We have used standard Scikit learn functions for parameter metrics and training-test data split up. tensorflow keras, Matlibplot functions are also loaded. We divided the BreaKHis dataset into training (80%) and testing (20%) sets. BreaKHist is mainly used to analyze the classification performance and evaluate the compression strategy of the model.



B.Pre-processing

-processing is a technique that is used to he raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

There are 3 different pre-processing techniques:

- **Rescale**
  It is used to make uniform/same scale value for all the raw data obtained using rescaling. We can rescale the data using SciKit-learn using the 'MinMaxScaler' class.
- **Binarize**
  This converts the data into binary form. The image pixels/data values are compared with the given threshold and all the values above the threshold are marked 1 and all equal to or below are marked as 0. We can create new binary attributes in Python using Scikit-learn with the 'Binarize' class.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 7, Issue 9, September 2020

- **Standardize**
  Standardization is a useful technique to transform attributes with a Gaussian distribution and differing means and standard deviations to a standard Gaussian distribution with a mean of 0 and a standard deviation of 1 We can standardize data using scikit-learn with the 'Standard Scaler' class. In this process, data is converted to binary form.

C.Labelling

The labeling process assigns tag values (labels) to the samples as binary -1 (malignant) binary -0 (benign) H&E stains. Next split the dataset using Sckitlearn as training data and testing data.
Using [4] standard ResNet50 model avaialbe in ImageNet the CNN ResNet50 structure is formed. It has 50 hidden layers between input and output layers.

D.Data augmentation

Deep learning neural networks are data hungry networks and require standard hardware for its implementation. In the proposed work only 1000 image samples are used. The augmentation increases the size of the training data and one of the ways to prevent over fitting. Data augmentation is a way to create different images from one image while keeping the context same. Rotating flipping scaling, shifting, varying the contrast, enhancement and brightness control are the different ways for data augmentation.

E.Transfer learning

In transfer learning a pre-trained CNN is used as a fixed feature extractor. The following are the steps in transfer learning

- Take a pre-trained CNN architecture trained on a large data set (like ImageNet).
- Remove the last fully connected layer of this pertained network.

Thus the remaining CNN acts as a fixed feature extractor for the new dataset.
First one or two layers in the CNN are more generic layers and gives general information. These layers transfer Knowledge very well. Higher layers in this network are not task specific.

With this pre-trained CNN, connect a last layer (fully connected layer) according to the task such as classification, segmentation etc., and train the last layer according to the task. Transfer learning and fine tuning of higher layers often lead to better performance than the training from scratch on the target dataset. In the proposed work ImageNet CNN resNet 50 is used for transfer learning.

Next the network is trained for the given input images and the test images are applied for prediction. The network performance is analyzed by varying the optimizers and different dropout rates. Optimizers are used to fine tune the model into its most accurate form and avoids getting trapped in sub optimal local minima points in the solution space. It modifies the learning rate and makes the training in the right direction based on the loss function guidance .Three optimizers are used in this proposed work for performance analysis.
They are

- Adagrad
- RMS Prop
- Adam.

Adagrad: It adaptively scales the learning rate for different dimensions. The scale factor of a parameter is inversely proportional to the square root of sum of historical squared values of the gradient. It reduces learning rate faster for parameters showing large slope and slower for parameters giving smaller slope. It works well for sparse input datasets. It is not suitable for non convex functions training. The adaptive learning rate tends to get small value over a period of time. This is the negative side of ad grad.

RMS Prop: It is a special version of ada grad and it uses exponentially decaying average of squared errors in a fixed window and discards history from the extreme past gradient errors as in Ada grad.

Adam: stands for adaptive moment estimation. It is the variant of RMS prop and momentum. Momentum optimizer is incorporated in RMS PROP by adding momentum (current gradient) to the rescaled gradients (fractions of the previous gradients).This oprimizer is generally used in training neural nets.
During training randomly selected neurons are dropped from the network with some probability temporarily. Their activation is not passed to the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 7, Issue 9, September 2020

downstream neurons. The dropout in neural nets connections after dropout is shown in Fig 5.There is no dropout during testing.

Normally increase in dropout increases the number of iterations for the network to convergence but at the same time helps to avoid over fitting. In the over fitting condition, the gap between training and testing error is large and the network tries to memorize the training data. In the proposed work the dropout rates are varied with the probability of 0.4, 0.5 and 0.6, and the dropout regularization is performed to avoid over fitting.

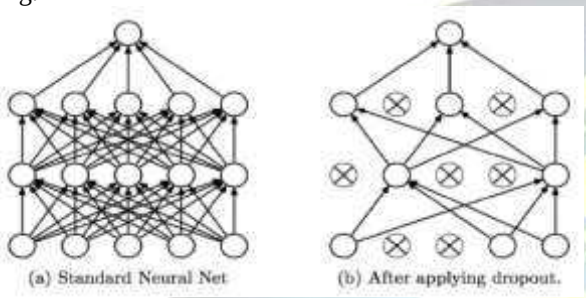


Fig. 5. Neural network structure before and after dropout

IV RESULTS AND DISCUSSION

In this proposed work, we used an input dataset consisting of 1000 Hematoxyline and Eosin (H&E) strained breast biopsy images form "BreakHist _Dataset" [16]. In that data set, for training 800 images & for testing 200 images are used. Each image size is 224*224*3. Here data augmentation is done by using ImageDataGenerator. In that we have applied zoom range-2, rotation range-90, horizontal and vertical flips condition true. Next the base ResNet 50 model is created and the model is fine tuned with the augmented data for breast H&E biopsy image classification.

In Preprocessing, we have done binorizing and labeling as per two classifiers(0 and 1 labels). Batch normalization is done with the batch size of 32. Model is initialized with ImageNet pretrained weights. Binary Cross entropy loss function is used for training. Global average pooling and output activation function is softmax.

The system with processor intel core-i3 -7020u is used to run the algorithm. It takes computation time of 21.22 min for classification. The data loading itself it takes around 14 minutes. Training time 6 min 21 seconds. Number of epochs are 20.

The confusion matrix obtained is shown in Fig.6.The benign image predicted as benign is high a (True positive) and the benign image predicted as malignant is less (i.e. 2 only).The malignant image predicted as benign is 0 and the malignant image predicted as malignant is high (67-True positive).

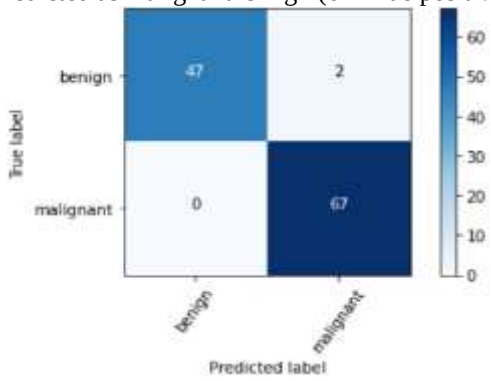


Fig. 6.Confusion matrix

Precision

It is the ratio of correctly predicted positive items to the total predicted items

Precision=TP/(TP+FP)

The precision value obtained for our applied model is 0.97.

Recall

It calculates how many actual true positives the model has captured, labelling them as positive.

Recall=TP/(TP+FN)

The value of Recall 1.00 is predicted by the applied model.

F-Score

F1-score is the function of precision and recall. It is calculated when a balance between precision and recall is needed.

F-Score= (2*Recall*Precision)/(Recall +Precision)

The obtained value of precision is 0.99.

E.Classification Accuracy

Accuracy in percentage= (number of correct predictions made)/total number of predictions made)*100

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology
(IJARTET) Vol. 7, Issue 9, September 2020

It should be as high as possible. Fig.7. shows the accuracy plot obtained for the model.

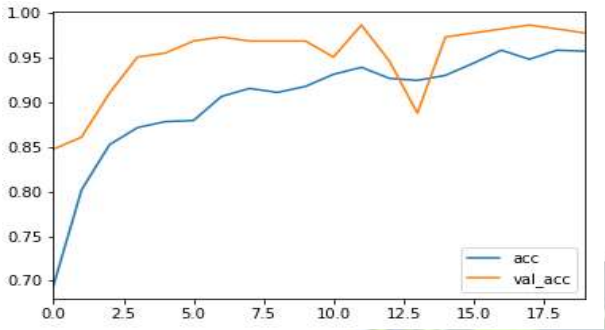


Fig.7. Accuracy plot

Area under the ROC Curve (AUC):
AUC provides the area under the ROC-curve integrated from (0, 0) to (1, 1). It gives the aggregate measure of all possible classification thresholds. AUC has a range from 0 to1. If a model classifies 100% correctly then the AUC value is 1.0 and it will be 0.0 if there is a 100% wrong classification. That is for all the test data the model produces wrong output. AUC is attractive for two reasons, first, it is scale-invariant, which means it checks how well the model is predicted rather than checking the absolute values, and second, it is classification threshold invariant as it will check the model's performance irrespective of the threshold being chosen. Fig.8. shows the ROC curve obtained for our obtained model and the AUC value can be obtained from this.

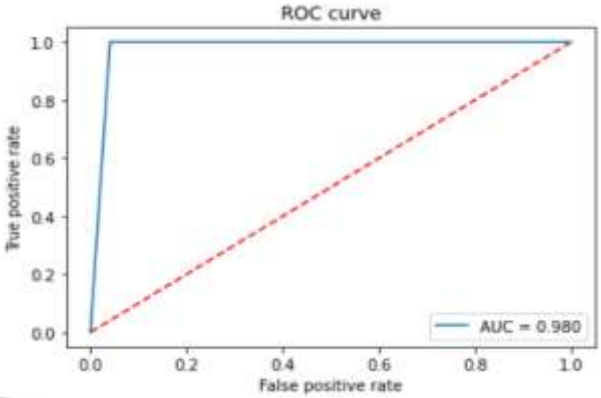


Fig.8.ROC curve.

The Stochastic Gradient Descent (SGD) method is carried out based on back propagation and the batch size of 32 is used to update the network parameters, including all the convolution layers and SEP blocks. Our CNN model is trained (fine tuned) for 20 iterations. The higher the AUC value the better the model and the maximum value of AUC is 1.0. From the above graph it is clear that the AUC value is 0.98 which means our model is the better one and it is covering the maximum number of input images dataset, predicting and giving the better results. Fig. 9. Shows the sample input H&E strain biopsy input imagesand fig.10 shows the obtained results after testing the images. It shows the predicted result (classification labels) and actual result (true labels) are same.

The below table 1 shows the comparison of accuracy value by changing the optimizers and dropout rate. Among the three optimizers, Adam optimizer produces best accuracy of 0.986 with the dropout rate of 0.4.In this optimizer, the increase in dropout rate decreases the prediction accuracy.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET) Vol. 7, Issue 9, September 2020

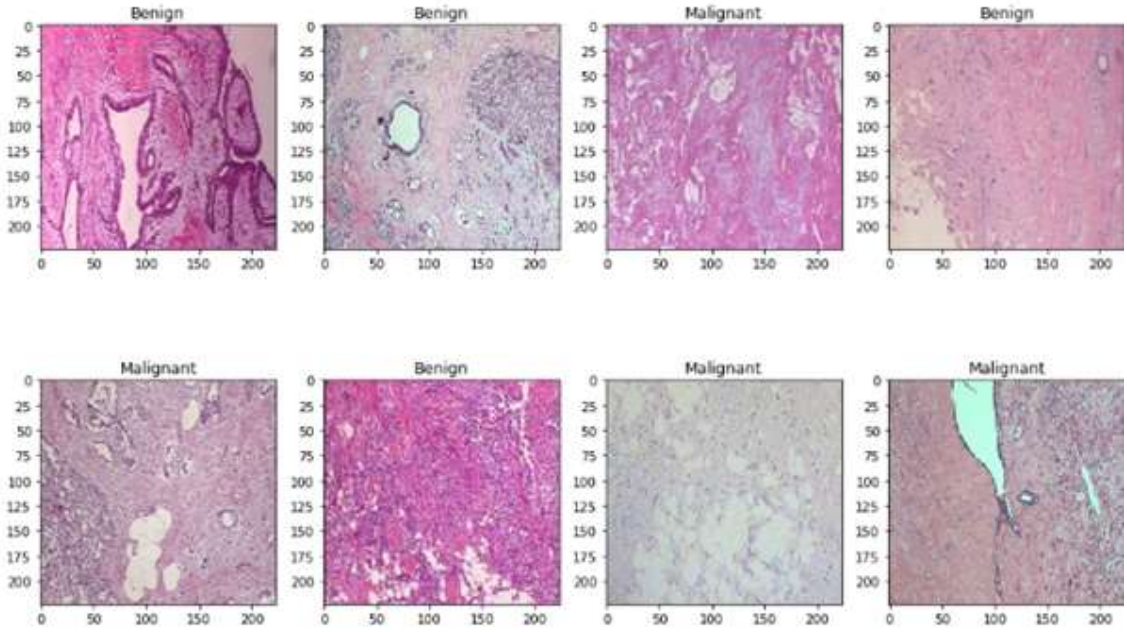


Fig.9 Input H&E strain biopsy images

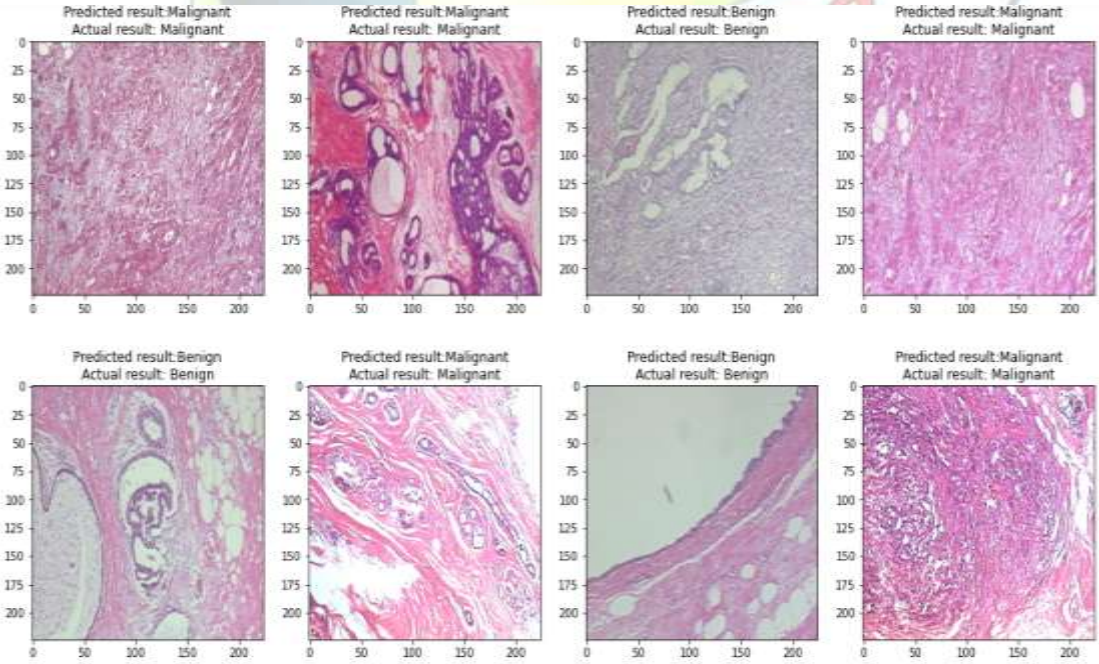


Fig.10. Output images with predictions

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 7, Issue 9, September 2020

For the Ada Grad optimizer, the best result is obtained for the dropout rate of 0.5 with an accuracy of 0.829. In case of RMS prop best prediction accuracy is obtained for the dropout rate of 0.4 only. Thus the model performance is analyzed using the same input data (Images).

Table 1 Accuracy value for different optimizers and dropout rate

SL .No.	Optimizer	Dropout Rate	Accuracy in %
1	Adam	0.4	98.6
2		0.5	97.7
3		0.6	96.8
4	Ada grad	0.4	79.3
5		0.5	82.9
6		0.6	77.5
7	RMS prop	0.4	96.7
8		0.5	93.7
9		0.6	94.6

V. CONCLUSION

In this paper, we proposed a simple and effective method for the classification of histological breast cancer images in the situation of very small training data (few hundred samples). To increase the robustness of the classifier we used strong data augmentation and deep convolution features extracted from publicly available CNNs ResNet50 pre trained on ImageNet. We analyzed the accuracy value for different optimizers like Adam, Ada grad, RMS prop by applying different value of dropout rate. For the Adam optimizer and 0.4 dropout rate the proposed method provides maximum accuracy as 0.986 and AUC as 0.98. In future ,various pretrained models will be analysed for H&E biopsy images for improved cancer prediction.

REFERENCES

.

[1]. Boyle P.Levin ,Advances in Breast cancer Report ,International Agency for Research on Cancer vol-4 ,sept 2015

[2]. Breast Cancer Facts & Figures 2019-2020 American cancer society.

[3]. Mitko Veta, Josien P.W. Pluim, Paul J. van Diest, and Max A. Vierg,."Breast Cancer Histopathology Image Analysis: A Review", *IEEE transactions on bio-medical engineering* · May 2014 DOI: 10.1109/TBME.2014.2303852 ·

[4]. Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun,"Deep Residual Learning for Image Recognition" arXiv:1512.03385v1 [cs.CV] 10 Dec 2015

[5]. V. Sze, Y.-H. Chen, T.-J. Yang, and J. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," *ArXiv170309039 Cs*, Mar. 2017.

[6]. A. R. Jamieson, K. Drukker, and M. L. Giger, "Breast image feature learning with adaptive deconvolutional networks," presented at the Medical Imaging 2012:Computer-Aided Diagnosis, 2012, vol. 8315, p. 831506.

[7]. H. C. Shin *et al.*, "Deep Convolutional Neural Networksfor Computer-Aided Detection: CNN Architectures,Dataset Characteristics and Transfer Learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, May2016.

[8]. N. Tajbakhsh *et al.*, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?,"*IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[9]. Sara Hosseinzadeh Kassani Peyman Hosseinzadeh Kassani and Michal J. Wesolowski ,"Classification of Histopathological Biopsy Images Using Ensemble of Deep Learning Networks" arXiv:1909.11870v1 [eess.IV] 26 Sep 2019.

[10]. Alexander A.Kalinin, 3D cell nuclear morphology :microscopy imaging dataset and voxel-based morphometry classification results, bioRxiv, 208207, 2017

[11]. Pierre Baldi ,Autoencoders unsupervised Learning and deep architectures ,proceedings of 2011international conference volume 27, 37-50 ,July 2011.

[12]. Juanying Xie, Ran Liu,Joseph Luttrel IV and Chaoyang Zhang "Deep Learning Based Analysis of Histopathological Images of Breast Cancer" Frontiers in Genetics vol.10 article 80,p.np.1-19,Feb 2019 DOI:10.3389/2019.00080

[13]. Alexander Rakhlin , Alexey Shvets , Vladimir Iglovikov , and Alexandr A. Kalinin," Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis" arXiv;1802.00752,DOI 10.1007/978-3-319-93000-8_83.

[14]. Teresa Araujo, Guilherme Aresta, Eduardo Castro, Jose´ Rouco, Paulo Aguiar, Catarina Eloy, Anto´nio Polo´nia, Aure´lio Campilho, Classification of breast cancer histology images using Convolutional Neural Networks, PLoS ONE 12(6): e0177544. https://doi.org/10.1371/journal. pone.0177544

[15]. Shuyue Guan and Murray Loew ,Breast cancer detection using transfer learning in Convolutional Neural Networks ,2017 IEEE Applied Imagery Pattern Recognition Workshop AIPR ,2332-5615 ,10 sept 2018.

[16]. Shallu Sharma Dr. Rajesh Mehra,"cancer histology images classification: Training from scratch or transfer learning?" RESEARCH GATE/ ICT Express 4 (2018) p.no.247–254 Article · November 2018 DOI: 10.1016/j.icte.2018.10.007

[17]. https://www.kaggle.com/ankur1809/breakhist-dataset.