© May 2019 | IJIRT | Volume 5 Issue 12 | ISSN: 2349-6002

Breast Cancer Classification Based on Histology Images Using CNN

Prachi Shingvi[1], Saurabh Waikule[2], Shashank Thigale[3], Prof. M. R. Mahajan[4]
[1,2,3]Department of Information Technology, PVG's College of Engineering & Technology, Pune – 411009
[4]Department of Information Technology, PVG's Information Technology, Pune – 411009

***Abstract-* Breast cancer is one of the most significant reasons for death among women. Early diagnosis significantly increase the chance of correct treatment and survival but this process is tedious and often leads to disagreement between pathologists. Computer based analysis showed potential for improving diagnostics accuracy. Many research has been done on the detection and classification of breast cancer using various image processing and classification techniques. The performance of most conventional classification system relies on feature extraction and appropriate data representation. On the other hand deep learning can organize discriminative information from data. Using CNN we propose a method for classification of histology images into benign and malignant and also subclasses. For 2 class classification task, we report 88% accuracy and for 4 class classification task, we report 85% to 89% accuracy.**

Index Terms- Breast cancer, Classification, Convolutional Neural Network (CNN), Deep learning, Histopathological image.

I. INTRODUCTION

Now-a-days, cancer is massive public health problem around the world. Cancer is a disease in which the cells in the body begin to divide at a faster rate than the body requires. According to International Agency for Research and Cancer (IARC) part of World Health Organization (WHO), there were 8.2 million deaths caused by cancer in 2012 and 26 million new cases of this disease are expected to occur until 2030 [1]. Among the cancer types, Breast cancer is the most common invasive cancer in women and the second cause of death in women, after lung cancer. The development of massive breast cancer screening has led to early diagnosis and rapid management with significant improvement in survival rate. Predicting a breast tumor is a challenging task for histopathologist. The tissue is taken and studied under a microscope to see if the area is cancerous. Detection is based on the qualification of the histopathologist, who will examine the tissue looking for cancerous cells. If histopathologist does not have much experience, this may lead to an incorrect diagnosis. Hence, there is a need for a system in order to help histopathologist for classification of breast cancer.

With the onset of pattern recognition and machine learning, many handcrafted (engineered) features-based studies are proposed for classifying breast cancer histology images. Some studies have focused on nuclei segmentation as in [2]. After selecting the region of interest, a set of features are extracted and fed into traditional classifiers to classify the breast histology images into either benign or malignant. CNNs have been applied to address the task of breast cancer histology images classification such as in [3], where the authors divided the histology images into small batches and then used to train CNN. To get the final classification result, the patches results are combined for the whole image. In [4], the authors also used CNN to classify breast histology images into four classes - normal tissue, benign lesion, in situ carcinoma, and invasive carcinoma. The authors also extracted a set of features from the CNN and fed them into support vector machines.

With the recent advancement in image processing and machine learning allows building computer aided detection systems that help histopathologist to be more productive and accurate in diagnosis. The proposed approach aims to classify breast cancer using histopathological images into benign and malignant and their subtypes. We perform image processing post which the processed image is given for training to Convolutional Neural Network. The model is then trained on the training dataset. After the model has been trained well, it can be used for classification of a new input image. Physicians can then use this output to make further decision regarding diagnosis.

II. METHODOLOGY

This paper presents a system for breast cancer classification using CNN. The system will classify histopathological images into benign and malignant and further into their sub types. The dataset [5] used is BreaKHis dataset. Dataset will be divided into training image set and test image set. The images from training set are first preprocessed which includes image conversion, augmentation and resizing. The deep learning model is built based on the input images by using Convolutional Neural Network. A new test image is given to the trained model to classify it first into benign or malignant and further into their subtypes.

A. PREPROCESSING MODULE

To bring the microscopic images into a common space to enable improved quantitative analysis, we perform image preprocessing. The various steps involved in image processing are as follows:

1. RGB to Gray Scale conversion:

An RGB image consists of 3 layers R, G, B i.e. Red, Green and Blue. It is a 3 dimensional matrix. Whereas grayscale image is of only 2 dimensions, and the values ranges between 0–255. Image is made up of number of pixels and various important factors like color and monochrome. Using different image processing techniques, image is processed and executed. Gray scale conversion is an important part of image processing. RGB or color information has a 3 dimensional characteristics which makes signal processing so much difficult to process. Therefore, to remove this disadvantage, Gray scale conversion is necessary. RGB to Grayscale conversion involves eliminating the hue and saturation information while retaining the luminance. Image conversion is shown in Fig 1.2

2. Augmentation:

Data augmentation is an important step to increase number of samples and learn deep network from the images. In order to achieve good performance, deep networks need large amount of training data. Image augmentation is helpful to boost the performance of a powerful image classifier which uses very little training data. Image augmentation artificially creates training images through different ways of processing or combination of multiple processing, such as random rotation, shifts, shear and flips, etc. Augmentation is used to utilize the full power of Convolutional Neural Network. The data augmentation in our project contains rotation of images by 900, 1800, 2700 and horizontal flipping of images

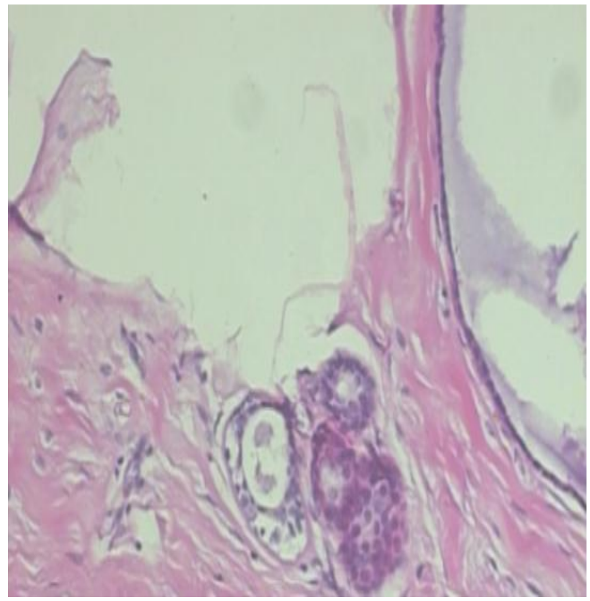

Figure 1.1: Normal image

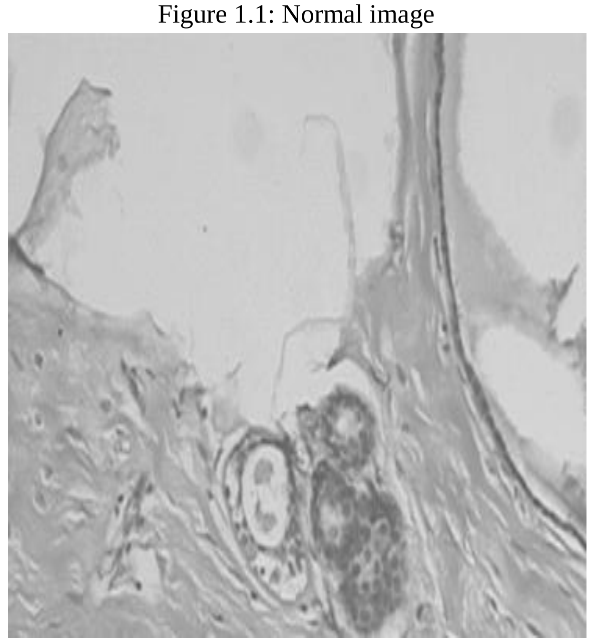

Figure 1.2: Gray scale image

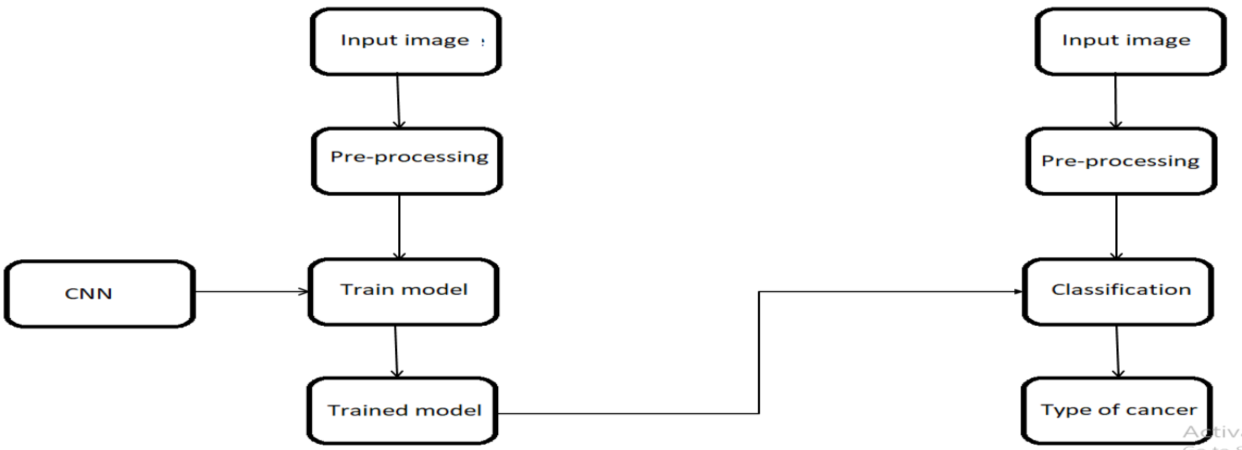

Figure 2: Architecture of proposed system

B. TRAINING MODULE

The preprocessed images from training image set will be given as input to the neural network. In the proposed system 2 models were trained. One for two-class classification and other for four-class classification.

1. Convolutional Neural Network

Convolutional Neural Network consists of convolutional layers present in their architecture. Convolutional layers use to identify features of an image. In each and every movement of the kernel on image, CNN learns the feature of an image. A deep Convolutional Neural Network is trained by feeding image as input to first layer and letting it compute and extract feature and generate the output. Training the CNN is an iterative process that involves multiple layers and the input is passed to these layers where parameter is updated in each layer until the model converges.

a. The architecture of two class model is composed of five convolutional layers and two fully connected layers.
   - 1st convolutional layer with filter size 3×3 and 32 feature maps.
   - 2nd convolutional layer with filter size 3×3 and 64 feature maps.
   - 3rd convolutional layer with filter size 3×3 and 96 feature maps.
   - 4th convolutional layer with filter size 3×3 and 128 feature maps.
   - 5th convolutional layer with filter size 3×3 and 128 feature maps.
   - Fully connected layer with 512 hidden units.
   - Fully connected layer with number of hidden units equal to the number of classes.
   - Softmax layer.

b. The architecture of four class model is composed of five convolutional layers and two fully connected layers.
   - 1st convolutional layer with filter size 3×3 and 64 feature maps.
   - 2nd convolutional layer with filter size 3×3 and 96 feature maps.
   - 3rd convolutional layer with filter size 3×3 and 128 feature maps.
   - 4th convolutional layer with filter size 3×3 and 256 feature maps.
   - 5th convolutional layer with filter size 3×3 and 256 feature maps.
   - Fully connected layer with 512 hidden units.
   - Fully connected layer with number of hidden units equal to the number of classes.
   - Softmax layer.

We have applied the RELU layer to all convolutional and fully connected layers to fasten the convergence learning and to introduce the non-linearity to the proposed system. The RELU layer changes all negative values to zero for the given input by applying the function $f(x) = \max(0, x)$. To train the model we have used the adam optimizer to minimize the loss function [7-41]. Adam optimizer is a gradient

descent algorithm with an adapter momentum that computes adaptive learning rate for each parameter.
Dropout layer was applied after second the fully connected layer with keep probability of p=0.5 and after first with p=0.15. The dataset was randomly shuffle to avoid any negative impact on the learning by using ordered training data.

C. CLASSIFICATION MODULE

The unseen test image will be preprocessed where RGB to Grayscale conversion of the image will be performed. This preprocessed image will be given as input to the trained model. The trained model will classify the image first into benign or malignant and further into its subtypes.

III. DATASET

The BreaKHis dataset contains microscopic biopsy images of benign and malignant breast tumors. Images were collected through a clinical study in year 2014. Dataset contains 7909 pathological breast cancer images with different magnification of 40X, 100X, 200X, 400X. Each slide of breast tumors is stained with hematoxylin and eosin (HE). The samples are collected by surgical (open) biopsy (SOB), prepared for histological study and labelled by pathologists of the P&D Lab. The preparation used in this work is the standard paraffin process, which is widely used in clinical routine. The complete preparation process includes steps such as fixation, dehydration, clearing, infiltration, embedding and trimming. Final diagnosis of each case is produced by experienced pathologists and confirmed by complementary exams such as IHC analysis [6].
The dataset contains four distinct histological sub-types of benign breast tumors: Adenosis, Fibroadenoma, Phyllodes tumor, Tubular adenoma; as well four malignant tumors: Ductile carcinoma, Lobular carcinoma, Mucinous carcinoma, Papillary carcinoma. The dataset has been randomly split into 70% training and 30% testing.

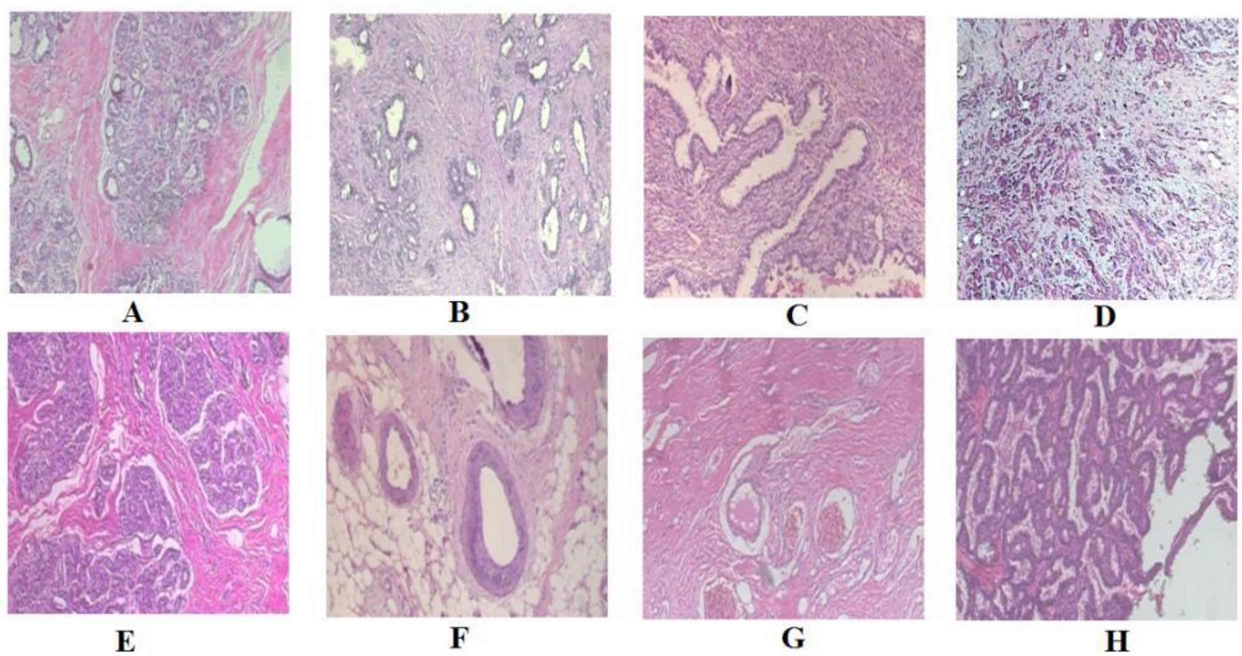

Figure 3: Samples of each type: (A): Adenosis, (B): Fibroadenoma, (C): Phyllodes tumor, (D): Tubular adenoma, (E): Ductal carcinoma, (F): Lobular carcinoma, (G): Mucinous carcinoma, (H): Papillary carcinoma

IV.CONCLUSION

This paper proposes a method for classification of breast cancer using convolutional neural network. CNN is used to in order to increase the accuracy of the current system. This system can be used to assist doctors for diagnosis of breast cancer which helps them to take more informed decisions.

REFERENCES

[1] P. Boyle and B. Lewin, Eds. World Cancer Report 2008. Lyon: IARC, 2008. [Online].Available:http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/wcr_2008.pdf .

[2] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak, "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images," Computers in Biology and Medicine, vol. 43, no. 10, pp. 1563-1572, 2013.

[3] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using Convolutional Neural Networks," 2016 International Joint Conference on Neural Networks (IJCNN), 2016.

[4] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using Convolutional Neural Networks," Plos One, vol. 12, no. 6, Jan. 2017.

[5] F. Spanhol, L. S. Oliviera, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," IEEE Transactions of Biomedical Engineering, 2016.

[6] A. L. Mescher, Junquiera's basic histology: text and atlas. New York: McGraw-Hill Lange, 2013.

[7] D. Kingma, and J. L. Ba, (2015), "Adam: a Method for Stochastic Optimization". International Conference on Learning Representations, 1-13.

[8] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropogation applied to handwritten zip code recognition," Neural Computation, vol. 1, no. 4, pp. 541-551, 1989.

[9] D. Bardou, K. Zhang, and S. Mohammad Ahmad, "Classification of breast cancer based on histology images using Convolutional Neural Networks," IEEE, 2017.

[10] Y. Zhang, B. Zhang, F. Coenen, and W. Lu, "Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles," Machine Vision and Applications, vol. 24, no. 7, pp. 1405-1420, 2013.