BREAST CANCER DIAGNOSIS USING NEURAL NETWORK BASED ON HISTOLOGY IMAGES

[1] Shruti Lohokare, [2,] Priyanka Nikam [3] Vrashali Talke, [4] Bhagyashri Wadkar ,

Abstract: Breast cancer is foremost cause of death in women and many efforts have been needed to obstruct it. Early diagnosis crucially increases the chances of meticulous treatment and survival, but this process is very tedious and often assists to a controversy between pathologists. To conquer with is problem we develop breast cancer detection system based on histological images. Diagnosis of breast cancer mainly depends on perusal of biomedical images such as Mammograms, Histological images, etc. In this work, Convolutional Neural Network (CNN) is used which is a fruitful technique to detect breast cancer and classify the histological images into benign and malignant and its sub-classes like adenosis, fibroadenoma, etc. Each classification method has predefined work such as features extraction, selection of dataset and constructing model to perform the classification. This system can be considered as a second opinion for doctors.

IndexTerms -Classification, Image processing, Neural network.

I. INTRODUCTION

Recent global cancer statistics reported that breast cancer is nevertheless the most common cancer type and the leading cause of malignant neoplastic disease-caused mortality among women, worldwide, with 2.4 million new cases and 523,000 deaths per year. Worldwide studies in 2012 reported 522,000 deaths from breast cancer, an increment of 14% over 2008 [2].

A biopsy can be used to identify breast cancer, in this tissue is detached and observed under microscope. The detection is depends on the qualification of the pathologist, who will study for abnormal cells After studying a large number of tumor tissue slides, pathologists can maximize their knowledge of cancer detection and interpretation. If images have low magnification, then manual inspection methods introduce different types of faults, like statistical, distributional and human errors.

By the extinct improvement in machine learning and image processing there is an engrossment in trying to bring a dependable pattern identification-based schema to enhance the quality of diagnosis. Histological images permit to differentiate the patterns according to cell nuclei types and their architecture. To determine cancerous areas and malignancy degree, histologists visually examine the invariability of tissue diffusion and cell shapes. The convolutional neural networks (CNN) can be considered as alternative of the standard neural networks.

II. RELATED WORK

R. Guzman Caberrera et al. [3], Proposed a Computer Aided Detection scheme to disparate sceptical regions containing the lump from background parenchyma using local entropy. The proposed algorithm composed of background and objects, texture segmentation and extraction. In the proposed work, window size of 9*9 was avail to find out local entropy, then the characteristics of the local textures are extracted depends on the adjoining pixel of the present pixel. After identifying regions of interest, it is removed using both the background texture and original input image as a disparity mask. However, the disadvantage of the proposed system is that the segmentation and extraction of the sceptical area is fully dependent on an appropriate value of the mention gray level.

Leonardo de Oliveira et al, [4] proposes a method for detection of lumps in mammographic images using support vector machine k-means Image segmentation is used k-means algorithm and texture characteristics of the segmented area. Established along the texture characteristics of the segmented area, support vector machine classifies the normal and cancerous tissues. There are certain limitations over the use of k-means algorithm used in the study, which is unable to detect the masses due to lack of ability to differentiate between mass region and sensitivity. But it is limited only for small tumors.

E. Mahmoudzadeh et al [5], proposed a method to improved separation of breast thermography images by applying extended hidden Markov model. They proposed a novel extended HMM for real image clarification and simple analysis of the TR thermal pattern. Re-estimation of model patterns can handle random sampling of breast IR image. But it is limited only for small tumors.

Anuj Kumar et al. [6], Proposes research which composed of detection and segmentation of region of malignant tissue. The method identifies the cancerous region and segment the identified region. At a detection stage, cancerous region is detected by applying a thresholding operation and an averaging filter to an actual input image. As per authors claim, because of the basic image processing technique and successful segmentation of the malignant tissues the proposed method is easy and rapid for the histological images. However, the drawback is, it is highly depending on bulk of the averaging filter and manual selection of threshold parameter.

Woten, D. A. et al [7], used a planar broadband antenna and a three-region breast approach for numerical examination into the betterment of artificial neural network and diagnosis of breast cancer. In this method, for producing various wave polarizations Modified Four-point antennas are used. Spectroscopic analysis-based proteome offers a capable technique for the approximate identification of various diseases. But the analysis of the proteomic pattern becomes very difficult due to the mass spectral data, such as huge mass, data entanglement and the existence of noise.

Kashyap, et al [8] proposed FCM Algorithm is used. The advantages of this algorithm are Sharp edges are obtained and Increases accuracy of segmented images. The disadvantage is, this method does not classify the benign and malignant masses.
Kontos, K., &Maragoudakis, M. [9] proposed a method for extraction of image segmentation done by The Statistical Region Merging (SMR) algorithm. Accuracy can be improved by identifying tissues of the tumor. Disadvantages are Classimbalance; Genetic selection feature method needs precision improvement.

Singh, N. et al [10] suggest fuzzy c-means and k-means methods for automatic breast cancer lump and calcification prediction in histological images. Clustering Algorithm i.e. means. Advantages are Low implementation cost and Detect early breast cancer. Disadvantages give unsatisfactory and noisy images. K-Means tends to run faster than FCM.

Guzmán-Cabrera et al [11] proposed microcalcification are used to distinguish it from background tissue and Morphological operators are used to distinguish masses. Advantages are Gray level value segmentation and extraction of regions is successful. It is not the best option to recognize suspicious areas along with irrelevant area.

Hairong Qi et al. [12] use asymmetry analysis to detect abnormalities. They have proposed a method which includes automatic segmentation and pattern classification. Here left and right breast is segmented to extract feature curves using Hough transform. Methods include steps like edge detection, feature curve extraction, segmentation and unsupervised learning. The drawback is of the proposed method is the experiments are done only on a specific set of images.

III. BACKGROUND

i. Classification

To classify histological images into benign and malignant and their subclasses we used Convolutional Neural Network (CNN) classification algorithm:

ii. Convolutional Neural Network

The CNN model contains input layer and output layer also contain three hidden layers which are convolutional layered, pooling layer and fully connected layer. This model has capability to extricate features in a hierarchical way that assure local connectivity and weight sharing property.

a. Input Layer

In input layer, the whole image is taken as input image and images are loaded from testing dataset and it produces output for convolutional layer. The given histological image is in the form of 2D array and in RGB form.

b. Convolutional Layer

The main aim of this layer is to reduce the size of image. In this layer feature of the input layer is matched with the filters. The match of features of the input image and the filter is calculated by taking dot product of each pixel of filter with each corresponding pixel value of input image.

c. Pooling Layer

Output from Convolutional layer is feed as input to pooling layer. The main aim of pooling layer is to reduce the large image generated in convolutional layer by preserving its important characteristics. The most common technique of pooling is max pooling.

d. Fully Connected Layer or Output Layer

The fully connected layer takes a highly-filtered image from pooling layer and transfers it into votes. In fully connected layer instead of taking two dimensional arrays it is converted into the single list and it is treated as identical. Every value gets in own vote based on whether the image belonging to given classes.

iii. Dataset

The dataset used in this work is BreakHis, which consist of histological images of benign and malignant and their sub-classes of breast tumor. It consists of 7909 images. The dataset contains 8 types of sub-classes which include four benign tumor types: Adenosis, Phyllodes tumors, Fibroadenoma and tabular adenoma and malignant tumor types: Ductal carcinoma, Mucinous carcinoma, lobular carcinoma, and papillary carcinoma.

a. Training and Testing Dataset

We used 80% of images from the dataset for training and validation and tested their performance on the remaining 20% of images.

IV. MATHEMATICAL MODEL

Mathematical model set theory S= {s, e, X, Y, Φ}
s= Start of the program
1. Register/Login into the system
2. Provide Symptoms and User Detail.
e= End of the program
Identify the Disease related to the symptoms
X= input of the program= {P, R, Q}
P = Symptoms of disease
R= Time required in queue for that disease
Q = Using Algorithm predicting the disease
First, users provide the symptoms and predict the particular disease
Let R be the set of Symptoms
R = {R1, R2, R3... Rn} //Number of Symptoms
Let A be the set of categories
Therefore,
A = {A1, A2, A3..., Am} // Number of Disease
Overall Disease is evaluated with the help of this Symptoms Y= E1+E2+... +Em / m
Where m is number of overall Disease

V. SYSTEM ARCHITECTURE

We developed a system for identification breast cancer depends on histological images. In this work, convolutional neural network is used for categorization of breast cancer within malignant, benign and its subclasses.

Classification of malignant and benign tissue images is always a complicated job. The complication of the image classification rises in case of histological images. Each classification method has predefined work such as features extraction, selection of dataset and constructing model to perform the classification. In this study, we developed method to classify the histological images into cancerous or healthy and then categorize into sub-types like Adenosis,Papillary Carcinoma, etc.

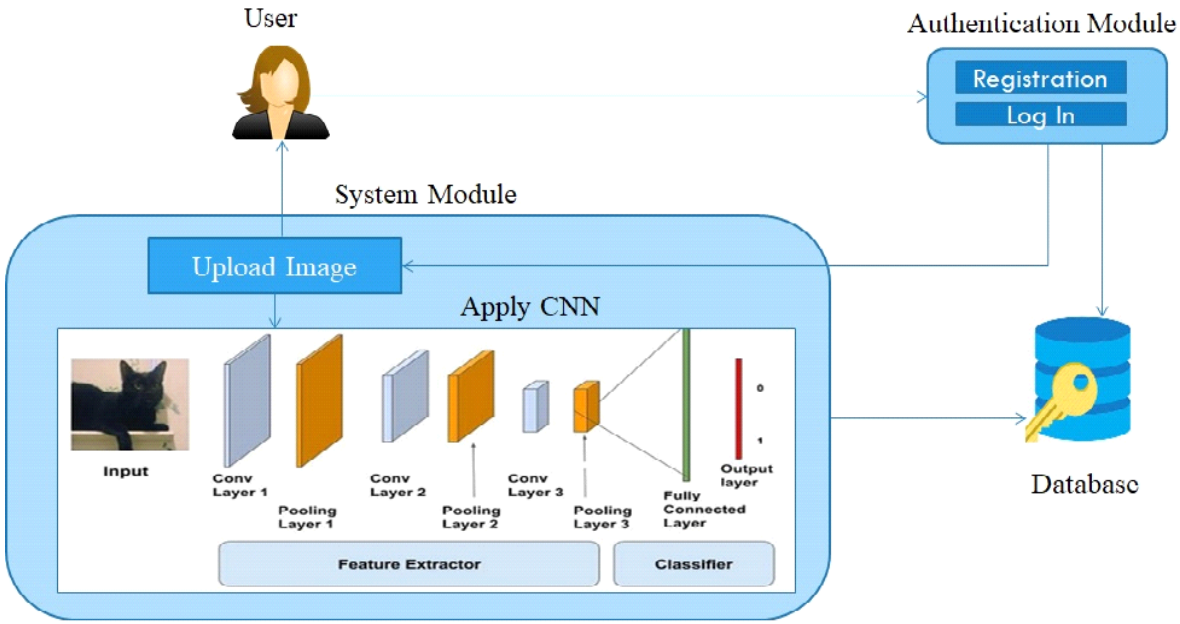

Figure: System Architecture

To predict the breast cancer first step is to upload the histological image on the web. The uploaded image is in the form of RGB hence RGB image is converted into a binary image. Then apply the CNN algorithm which has mainly three layers convolutional layer, pooling layer and fully connected layer. To extract the feature of image convolutional and pooling layer are used and fully connected layer is used for classification. The histological images are getting classified into cancerous or healthy and sub-types. After applying CNN, we get the output result, i.e. whether the cancer is present or not and if present, at which level it is.

i. Authentication Module

Firstly, users need to do registration by providing details like name, mobile number, email address, etc. After registration, user will be able to do login into the system by using username and password. To predict the breast cancer user needs to upload the histological image on the web and only authenticated users can upload the image.

ii. Database

All the details of user get stored into database after registration. When users upload images into system that image also get stored in a database with ImageId in the form of BLOB.

iii. System Module

To predict the breast cancer first step is to upload the histological image on the web. The uploaded image is in the form of RGB hence RGB image is converted into a binary image. Then apply the CNN algorithm which has mainly three layers convolutional layer, pooling layer and fully connected layer. To extract the feature of image convolutional and pooling layer are used and fully connected layer is used for classification. The histological images are getting classified into cancerous or healthy and sub-types. After applying CNN, we get the output result, i.e. whether the cancer is present or not and if present, at which level it is.

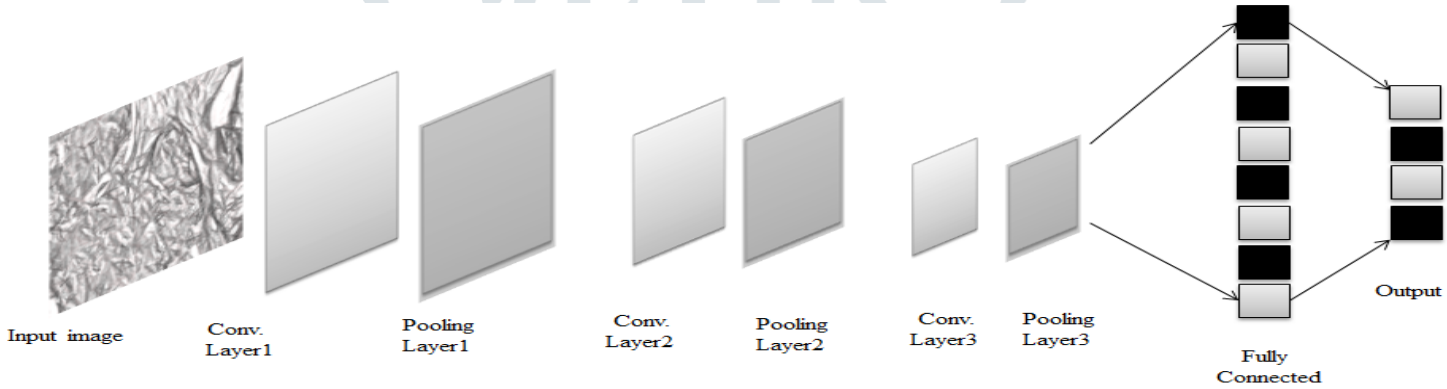

Figure: Convolution Neural Network

VI. CONCLUSION

The goal of this work is to detect the affected portion and to classify the histological images into benign and malignant. For the classification, we used Break His dataset which contain 7750 histological images. In this work, we used CNN classifier which extract the features from images as well as classify into benign and malignant and its sub-classes accurately. We used 80% of images from the dataset for training and validation and tested their performance on the remaining 20% of images

VII. ACKNOWLEDGMENT

It gives us great pleasure in presenting the preliminary project report on '**EXPERT SYSTEM FOR DIAGNOSIS BREAST CANCER ON HISTOLOGY IMAGES USING NEURAL NETWORK '**

I would like to take this opportunity to thank my internal guide for giving me all the help and guidance I needed I am really grateful to them for their kind support. Their valuable suggestions were very helpful.

I am also grateful to HOD for her in dispensable support and suggestions.

Name of Students

[1] Shruti Lohokare, [2,] Priyanka Nikam [3] Vrashali Talke, [4] Bhagyashri Wadkar,

REFERENCES

[1] DalalBardouet. al., "Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks", IEEE, ISSN: 2169-3536, May 2018.

[2] C. Fitzmaurice, C. Allen, R. M. Barber, L. Barregard, Z. A. Bhutta, H. Brenner, D. J. Dicker, O. Chimed-Orchir, R. Dandona, L. Dandona, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. JAMA oncology, 3(4):524–548, 2017.

[3] R. Guzman-Cabrera et. al., "Digital image processing technique for breast cancer detection", Int J Thermophys(2013)34:1519-1531.

[4] Leonardo de Oliveira Martins et. al., "Detection of Masses in Digital Mammograms using K-means and Support Vector Machine", Electronic Letters on Computer Vision and Image Analysis 8(2):39-50, 2009.

[5] E. Mahmoudzadeh, M. Montazeri, M. Zekri, and S. Sadri, "Extended hidden Markov model for optimized segmentation of breast thermography images," Infrared Physics & Technology, vol. 72, pp. 19–28, 2015.

[6] Anuj Kumar Singh, Bhupendra Gupta, "A Novel Approach for Breast Cancer Detection and Segmentation in a Mammogram" vol. 54, 2015.

[7] Woten, D. A., & El-Shenawee, M. "Improvement of artificial neural network detection of breast cancer using broadband dual polarized antenna". 2007 IEEE Antennas and Propagation International Symposium. doi:10.1109/aps.2007.439548.

[8] KanchanLataKashyapet. al., "Fractional Order Filter based Enhancement of Digital Mammograms", vol I, WCECS 2017.

[9] KonstantinosKontos, ManolisMaragoudakis, "Breast Cancer Detection in Mammogram Medical Images with Data Mining Techniques",DOI 10.1007/978-3-642-41142-7_35, ISSN 1868-4238, September 2013.

[10] Nalini Singh et. al., "GUI Based Automatic Breast Cancer Mass and Calcification Detection in Mammogram Images using K-means and Fuzzy C-means Methods", International Journal of Machine Learning and Computing, vol. 2, February 2012.

[11] R. Guzmán-Cabrera et. al., "Digital Image Processing Technique for Breast Cancer Detection", vol. 34, October 2012.

[12] Hairong Qi, P. T. Kuruganti, "Asymmetry analysis in breast cancer detection using thermal infrared images", IEEE, ISBN: 0-7803-7612-9, October 2002.