

Deep Learning based Prediction of Breast Cancer in Histopathological images

V Sansya Vijayan

M. tech student

Department of Computer Science and Engineering,
LBS Institute of Technology for Women
Thiruvananthapuram, India

Lekshmy P L

Assistant Professor

Department of computer Science and Engineering,
LBS Institute of technology for Women
Thiruvananthapuram, India

Abstract— Cancer is a baneful affliction across the world .It mainly affect both public and personal health system. Particularly in females, breast cancer is the second mass type of cancer affecting largely and particularly mass dangerous types when not decently perceived and treated. Malignant breast cancer is one of the dangerous types in humans .Early diagnosis can be curable. Many trending imaging technologies are there for verdict, biopsy is the most common way to detect cancer when it is present. Histopathological images are mainly used in diagnosis purpose .This paper mainly explains the techniques for detection of breast cancer applying both image processing and deep learning techniques. This paper mainly help to predict cancer as malignant and benign. Advances in image processing and machine learning modes, in which CAD (Computer-Aided Diagnosis) systems are built, which helps pathologists to be more, objective and consistent in the diagnosis process .Effective methods are required for detection of cancerous cells .This paper focuses an Lloyd's algorithm for clustering and convolutional neural network for classification purpose .About 96% accuracy is obtained from the results.

Keywords— Breast Cancer, Computer-aided diagnosis, Convolutional Neural Network, Deep Learning, Histopathological images, Lloyd's algorithm

I. INTRODUCTION

Cancer is one of the significant health issues. According to the word cancer research fund there was an increase of 20% of disease in recent days. Unhealthy diet is the important factor causing cancer. When compared to other type of cancer BC have high mortality rate. Among all kinds of cancer BC is the inferior type of cancer occur commonly in females .Early detection of cancer is one of the major challenging task. Detection means classifying tumor two classes (i) non-cancerous tumor (benign) and (ii) cancerous tumor (malignant). Biopsy is the nearly common way to detect cancer when it is present. In biopsy first samples of cells are collected. Detection of cancer from a histopathology image persist the "gold standard" especially in BC. The core of this paper is detection of breast cancer in histopathological images using Lloyd's algorithm and CNN. First breast cancer dataset is selected .Image enhancement is done using local contrast stretching .This is followed by pre - processing which uses Gaussian filter which helps in removal of unwanted noises. Next the image is clustered using Lloyd's algorithm. This technique mainly help in grouping similar data's. Finally, classification is done using CNN.

Organization of paper is done in this sequence .Section II describe the fundamentals of breast cancer. In section III depict the review on previous works. Section IV describes the processing stages. Section V covers a brief description of experimental results. Section VI is the conclusion part

II. FUNDAMENTALS OF BREAST CANCER

A) Breast Cancer Diagnosis

The main exams for initial analysis of breast cancer are imaging test. Final stage diagnosis can be established through Anatomopathological exam.

a) Imaging exam

Imaging exam mainly including technologies mammography, MRI, sonography and thermography, mammography is the current method for breast cancer masking as well as investigation at the beginning stages. This is performed by using equipment which uses X ray to produce high resolution image.

Sonography is the second crucial imaging exam .It is due to the ejection of strong frequency sound wave.

MRI is also an imaging test which is parallel to mammography and sonography which mainly apply electromagnetic waves and radio frequency pulses to produce the image .MRI allows detailed examination of nodes because it reaches deeper regions of tissue.

Thermography otherwise known as infrared thermograph is a rapid, passive, non-invasive method which is benefit to diagnose breast cancer.

b) Anatomopathological Exams

In breast cancer ,if physical examination detects lumps or any other suspicious tissues .Anatomopathological exams are required .Pathological exam analyses cellular and tissue microscopic alteration present in samples collected from biopsies .Biopsy is the technique of removal of small amount of breast tissue for pathological evaluation to determine whether it is cancerous or not.

B) Characteristics of Breast Cancer

All tumors are not cancerous .Non-cancerous types of tumors are known as benign .They will not affect or spread additional parts of the body. .Cancer is mainly used to describe malignant tumors that spread deeper from original part to all sections of the body. This may even cause death. Breast cancer is not a single disease it is constituted of other

subtypes based on evolution .By histopathological analyses it is better to classify different tumors.

C) Breast Cancer Types

There are 2 broad categories 1) Non special type 2) Special types .Non special types are the types of cancer which cannot be differentiate under microscope. Special type are the types of tumors which can be classified.

D) Causes Of Breast Cancer

This is caused due to the changes or mutation in DNA of the cells. Some of the hazard factors are benign condition like hyper plasia increase risk of breast cancer. Having a earlier history of cancer increases the chance of causing cancer.

E) Symptoms of Breast Cancer

- a) Colic or difficulty in swallowing
- b) Nagging cough
- c) Persistent headache or fatigue
- d) Unexplained loss of weight
- e) Chronic pain in bone

III. RELATED WORKS

Shannon Agner et.al [2] proposed a unique method for instinctive discovery of breast cancer histopathological images and differentiate as high and low degree .They bare a dataset of 3400 images which include formal and nuclear based features. Spectral clustering is used to abate the magnitude of images. Support Vector Machine classifier is used to assort images as cancerous and non-cancerous image and to distinguish low and high degrees of cancer.

Arnav Bhavsar et.al [15] proposed an access which uses textural features and aggregate method for classifying histopathological images .The main design of the act is to classify images based on different augmentation levels. Large variability in tissue advent is used to get different augmentation levels

Y Li et.al [13] proposed a rule known as curvature scale space corner detection for nuclei detection in breast histopathological images. This system mainly splits the surrounding cells to get better fidelity. Here for finding the ROI wavelet disintegration and multi scale region growing are aggregate together. For classification chain parallel agent genetic algorithm is used.

Maruf Hossain Shuvo et.al [10] proposed a classifier known as Wavelet Neural Network. Wavelet Neural Network is a type of artificial neural network. This current method is established on wavelet transform and neural net .The proposed model admit how WNN classifies by using certain formulas.

Daniel Racocanu et.al [6] proposed a medicinal knowledge guided epitome for indexing of histopathological images. A rule based decision method is used to close the semantic gap sole crucial confronts in medical image evaluation and inventory. This method is a robustic tool for the visual location and semantic retrieval.

Luiz S. Oliveira et.al [14] proposed a method which depict a brief description of all accessible dataset for breast carcinoma histopathological image classification. Dataset mainly comprise of 7909 breast cancer images from 82 patients. The main intent of this paper is automatic classification of images into 2 classes. The classification accuracy grades with 80-85.

Jia Qu et.al [12] proposes a scheme for pathological image classification by applying higher order local auto association attribute. Here a novel image pre-processing and area scalable evaluation form is used. In the novel pre-processing arrangement a effect with no false negative and with 3% false positive rank is exposed. In the scalable evaluation method pathological images are partitioned into smaller regions and local area is estimated without any extra computational cost .Anomaly detection achievement is enhanced and the location of anomaly is judged.

IV. PROPOSED METHODOLOGY

A) Overview

This paper presents a method for recognition of breast cancer in histopathological images using deep learning .Histo means tissue pathos means suffering. Histopathology refers to the study of diseases in tissues .Deep learning is subset of machine learning. which use multiple layers to get more high level features .First stage the input image is enhanced then pre-processed .Then image is clustered and final stage is classification which predict whether the image is malignant or benign.

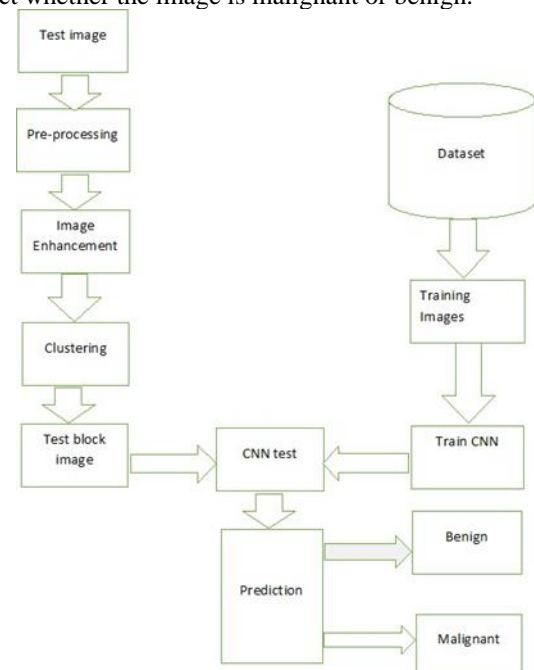


Fig 4.1: Block Diagram of the Proposed system

B) Dataset

Dataset here used is Break His which is obtained from Kaggle. The dataset Break His is dispense into two primary association: benign tumors and malignant tumors. Histologically benign is a term cite to a abrasion that does not match any benchmark of malignancy. Commonly, benign tumors are comparably “innocent”, present slow growing and

remain confined. Malignant tumor is a alternate for cancer. Break His contain 7909 images from 82 patients which contain benign and malignant tissue samples. In the Break His dataset, each image filename accumulate information nigh the image itself method of biopsy procedure, tumor class, tumor type, slide identification, and magnification factor.

C) Pre-Processing

Pre-processing is done mainly to remove all the unwanted noises from the image. This is an important step in machine learning .In data analysis quality of data is very much important. So pre-processing is an important step in computational biology. Data pre-processing comprise clean up, instance selection ,normalization. Data pre-processing mainly affect the way in which aftermath of final data processing can be construe. The main task are data cleansing, data editing, data reduction, data wrangling. First images are converted from RGB to grey level. Then noise reduction is done by using a Gaussian filter. Main advantage of Gaussian filter is it will un sharp all the edges in an image.Guassian filter is faster than median filter

Algorithm

For pre-processing breast cancer image

1. Read image
2. R,G,B are separated from colour image
3. Apply Gaussian filter to reduce noise

D) Image Enhancement

Image enhancement is a technique mainly used to action an image to get more suitable results than the original image. It is the alternate area of image processing. There are 2 main grade spatial domain methods and frequency domain methods. Enhancement at any point in an image dangle only on the gray level at that point techniques in this grades are often referred as point processing. Here contrast stretching is used for enhancement. Contrast stretching mainly stretches the range of intensity values to get high quality image.

E) Clustering

Clustering is done to group similar data points .Lloyd's algorithm is used for clustering .This is an important algorithm which extract all the given clusters from a training set .Initial methods used in clustering are Forgy and random partition .This clustering method mainly uses random partition .This algorithm is fast and easier to understand.

ALGORITHM

1. Number of clusters is specified as K
2. Place centroids at random location and select data points without replacing it.
3. Repeat step until no change in centroids
4. Calculate the distance between centroid and data points.
5. Place all data points to the nearby clusters.
6. Take average of all data points to compute the centroids.

F) Classification.

Classification is done by using Convolutional neural network. CNN is a deep neural network otherwise known as convNet.This deep neural network have the ability to capture

all spatial and temporal characteristics .Reducing the images into a suitable form without losing any features is one of the major advantage of CNN. CNN mainly consist of 4 layers such as Convolutional layer, RELU layer. Pooling layer and fully connected a normalized layer. CNN shares weights in the convolutional layer reducing the memory footprint and increases the performance of the network .High level features are extracted from convolutional layer. A feature map is produced by convolution layer through convolution of different sub regions of the input image with a learned kernel. Then, anon-linear activation function is applied through ReLu layer to improve the convergence properties when the error is low .Pooling layer is used to reduce the spatial size .This layer mainly extract prominent features. The pooling layer usually carries out two types of operations viz. max pooling and means pooling. In mean pooling, the average neighborhood is calculated within the feature points and in max pooling it is calculated within a maximum of feature points. Fully-Connected (FC) layer will be used in conjunction with the convolutional layers towards the output stage.

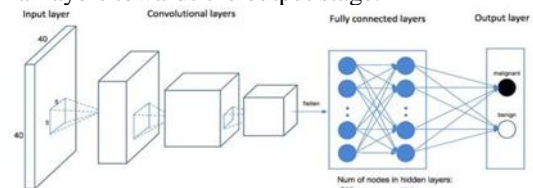


Fig1.Architecture of CNN

G) Evaluation Metrics

Performance and stability of system is calculated by using certain parameters. Few of these parameters are mentioned below. Sensitivity, Specificity, Accuracy are some of the parameter .ROC curve act as a tradeoff between sensitivity and specificity. When the area is statically greater than0.5 then will give better prediction.

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

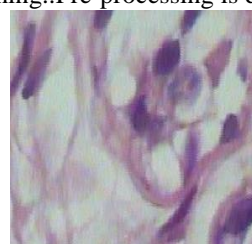
$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

V. RESULTS

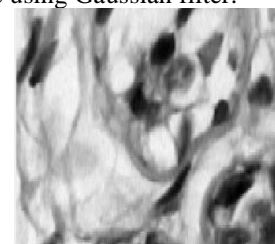
There are lots of mechanism for medical image diagnosis. These techniques plays an important milestone in diagnosis purpose. The present paper gives a brief description on breast cancer detection based on deep learning. A suitable Lloyd's algorithm is used for clustering. Prediction is done using CNN.CNN will detect pathological conditions for breast cancer accurately.

A) Pre-processing

Pre-processing is done mainly to remove all the unwanted noises from the image. This is an important step in machine learning..Pre-processing is done using Gaussian filter.



(a)



(b)

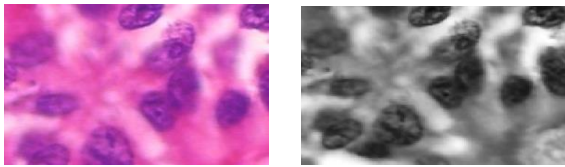


Fig 2. (a) Original image (b) pre-processed image (benign) (c) original image (d) pre-processed image (malignant).

B) Image enhancement.

Image enhancement is a technique mainly used to action an image to get more suitable results than the original image. Image enhancement is done using contrast stretching which is a point processing technique.

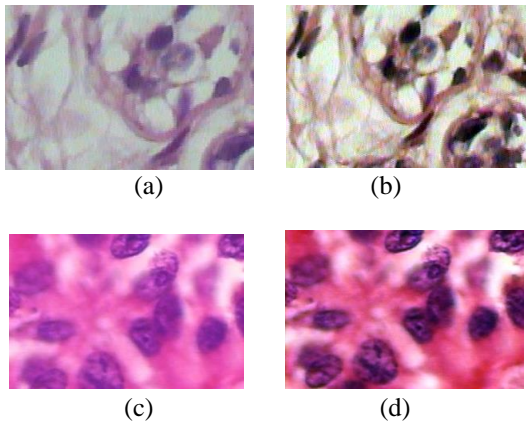


Fig 3 (a) original image (b) contrast stretched image (benign) (c) original image (d) contrast stretched image (malignant).

C) Clustering

Clustering is done to group similar data points. Clustering is done using Lloyd's algorithm.

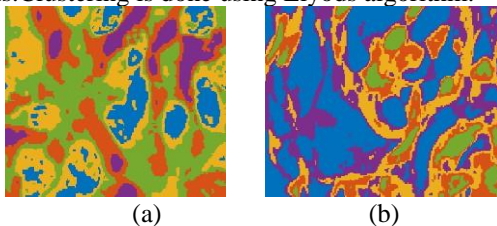


Fig 4. (a) clustered image (benign slice) (b) clustered image (malignant).

Table 1: Comparison of classification techniques.

S.NO	Classification techniques	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	SVM	86.30	90.30	93.00
2	Bayes Net	88.54	91.80	95.90
3	Decision Stumb	83.76	86.00	91.43
4	Adaboost	85.34	84.00	89.06
5	KNN	89.49	85.67	87.00
6	CNN	96.00	92.26	99.00

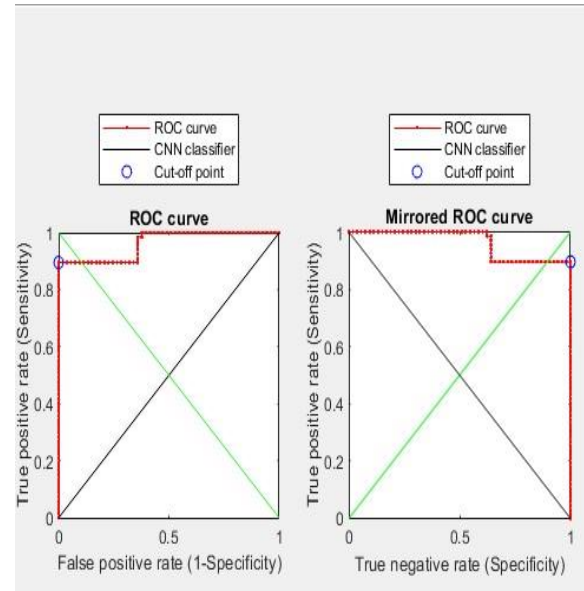


Fig 5: ROC Curve

Receiver operating characteristic (ROC) is a fundamental tool for diagnostic test evaluation. If the value of false positive is 0, that means there is no misclassification among the cancer cell images. In our work, there is no false positive detection. So we can make our classification as optimal

VI. CONCLUSION

A convolutional neural network-based system was implemented to detect the malignancy tissues present in the breast cancer histopathological image. Breast cancer histopathological image with different shape, size of the cancerous tissues has been fed at the input for training the system. The proposed system is able to detect the presence and absence of cancerous cells of breast cancer with accuracy of about 96%.

REFERENCES

- [1] Kamal H. Sager, Loay E. George ; , "Breast Cancer Diagnosis using Multi-Fractal Dimension Spectra", IEEE International Conference on Signal Processing and Communications ,NOV 2007
- [2] Shannon Agner ; Anant Madabhushi Scott Doyle ; , "Automated Grading of Breast Cancer Histopathology Using Spectral Clustering with textural and architectural image features", MAY 2008
- [3] Wiem Lassoued, Yousef Al-Kofahi ; , "Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images", IEEE Transactions on Biomedical Engineering APRIL 2010
- [4] AB Tobsun "Graph Run Length Matrices For Histopathological Images" IEEE Transactions on Medical Imaging Volume: 30 , Issue: 3 MARCH 2011
- [5] L. Rodney Long, Lei He , "Histology Image Analysis For Carcinoma Detection and Grading", from Computer Methods and Programs in Biomedicine 2012.
- [6] Daniel Racocanu, Adina Eunice Tutac , "Knowledge-Guided Semantic Indexing Of Breast Cancer Histopathology Images", International Conference On Biomedical Engineering And Informatics MAY 2018
- [7] Josien P.W, Mitko Veta, "Breast Cancer Histopathology Image Analysis: A Review", IEEE Transactions on Biomedical Engineering Volume: 61 , Issue: 5 , MAY 2014
- [8] Wei Lu , Zhang X, "Towards a large scale histopathological image analysis" IEEE Transactions on Medical Imaging Volume: 34 , Issue: 2 , FEB 2015
- [9] Smriti H. Bhandari , "A Bag-Of-Features Approach For Malignancy Detection In Breast Histopathology Images", IEEE International Conference on Image Processing SEP 2015

- [10] Md. Maruf Hossain Shuvo, Fatema-Tuz Johra ; , “ Detection Of Breast Cancer From histopathology image and classifying benign and malignant state using fuzzy logic”, 3rd International Conference on Electrical Engineering and Information Communication Technology , SEP 2016
- [11] Luiz S. Oliveira ,Fabio A. Spanhol , “A Dataset for Breast Cancer Histopathological Image Classification”, IEEE Transactions on Biomedical Engineering Volume: 63 , Issue: 7 , JULY 2016
- [12] Jia Qu ,Hirokazu Nosato “CANCER DETECTION FROM PATHOLOGICAL IMAGES USING HIGH ORDER LOCAL AUTO CORRELATION” ICSP OCT 2012
- [13] Y LI, P Wang X HU ,“AUTOMATIC CELL NUCLEI SEGMENTATION AND CLASSIFICATION OF BREAST CANCER HISTOPATHOLOGY IMAGES”, Signal Processing Volume 122, MAY 2016.
- [14] Luiz S. Oliveira,Fabio A. Spanhol , “Deep Features For Breast Cancer Histopathological Image Classification”, IEEE International Conference On Systems, Man, And Cybernetics (SMC) OCT 2017.
- [15] Arnav Bhavsar,Vibha Gupta, , “Breast Cancer Histopathological Image Classification ,Is Magnification Important” IEEE Conference On Computer Vision And Pattern Recognition,JULY 2017.