Fully Automated Segmentation and Classification of Breast Cancer using Machine Learning Techniques

N.Shivaanivarsha[1], N.Premalatha[2], A.Ramya[3]

1, 2, 3 Department of Electronics and Communication Engineering, Sri Sairam Engineering College, Chennai-600044

Abstract— In the field of radiology, mammographic screened images (i.e. X-rays image sensing) square measure terribly difficult and difficult to interpret. The skilled radiotherapist visually hunts the mammograms for any specific abnormality. However, human factor causes an occasional degree of preciseness which frequently ends up in biopsy and anxiety for the patient concerned. It proposes a novel Computer-Aided Detection (CAD) system to scale back the human issue involvement and to assist the radiotherapist in automatic diagnosing of benign/malignant breast tissues by utilizing the basic morphological operations. The input Region of Interest (ROI) is extracted manually and subjected to additional variety of preprocessing stages. The geometrical and texture features are used for feature extraction of suspicious region. After that a KNN classifier is introduced to classify the required class of the breast cancer.

Keywords—Breast cancer, Computed tomography, Breast CT image, Machine Learning, K-Nearest Neighbour

I. INTRODUCTION

Breast cancer is one of the most dreadful diseases in the developing countries and its mortality rate is 19.4%. Early detection of breast cancer is done by using many imaging techniques such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Detection means classifying tumor two classes (i) non-cancerous tumor (benign) and (ii) cancerous tumor (malignant). The chance of survival at the advanced stage is less when compared to the treatment and lifestyle to survive cancer therapy when diagnosed at the early stage of the cancer. Manual analysis and diagnosis system can be greatly improved with the implementation of image processing techniques. A number of researches on the image processing techniques to detect the early stage cancer detection are available in the literature. But the hit ratio of early stage detection of cancer is not greatly improved. With the advancement in the machine learning techniques, the early diagnosis of the cancer is attempted by lot of researchers.

The term digital image refers to processing of a two dimensional picture by a digital computer. In a broader context, it implies digital processing of any two dimensional data. A digital image is an array of real or complex numbers represented by a finite number of bits. An image given in the form of a transparency, slide, photograph or an X-ray is first digitized and stored as a matrix of binary digits in computer memory.

II. LITERATURE SURVEY

An Improved Object Detection Method for Mitosis Detection Haijun Lei1, Shaomin Liu1, Hai Xie2, Jong Yih Kuo3, and Baiying Lei4*-2019.Proposes an improved object detection method for automatic mitosis detection from histological images. First, we use a convolutional neural network (CNN) to automatically extract mitosis features. Then, we use the region proposed network (RPN) to locate a set of class-agnostic mitosis proposals. Finally, we use the improved R-CNN subnet to screen for mitosis from these proposals.

Weakly supervised mitosis detection in breast histopathology images using concentric loss Chao Li a, Xinggang Wang a Wenyu Liu a , ∗, Longin Jan Latecki b, Bo Wang c , Junzhou Huang d , e -2019.It proposes an automatic method for detecting mitosis. We regard the mitosis detection task as a semantic segmentation problem and use a deep fully convolutional network to address it. Different from conventional training data used in semantic segmentation system, the training label of mitosis data is usually in the format of centroid pixel, rather than all the pixels belonging to a mitosis. The centroid label is a kind of weak label, which is much easier to annotate and can save the effort of pathologists a lot. However, technically this weak label is not sufficient for training a mitosis segmentation model.

Multi-Level Wavelet Convolutional Neural Networks PENGJU LIU 1,HONGZHI ZHANG1,(Member, IEEE),WEILIAN 2,ANDWANGMENGZUO1,(Senior Member, IEEE) -2019. Proposes a novel multi-level wavelet CNN (MWCNN) model to achieve a better tradeoff between receptive eld size and computational efciency. The core idea is to embed wavelet transform into CNN architecture to reduce the resolution of feature maps while at the same time, increasing receptive eld. Specially, MWCNN for image restoration is based on U-Net architecture, and inverse wavelet transform (IWT) is deployed to reconstruct the high resolution (HR) feature maps. The proposed MWCNN can also be viewed as an improvement of dilated lter and a generalization of average pooling and can be applied to not only image restoration tasks, but also any CNNs requiring a pooling operation.

III. PROPOSED WORK

It proposes a novel Computer-Aided Detection (CAD) system to reduce the human factor involvement and to help the radiologist in automatic diagnosis of benign/malignant breast tissues by utilizing the Basic morphological operations. The input Region of Interest (ROI) is extracted manually and subjected to further number of preprocessing stages. The geometrical and texture features are extracted for feature extraction of suspicious region. After that a KNN classifier is introduced to classify the required class of the breast cancer.

A. Methodology Diagram

Figure 1 shows the block diagram of Image processing using segmentation and classification of Breast cancer.

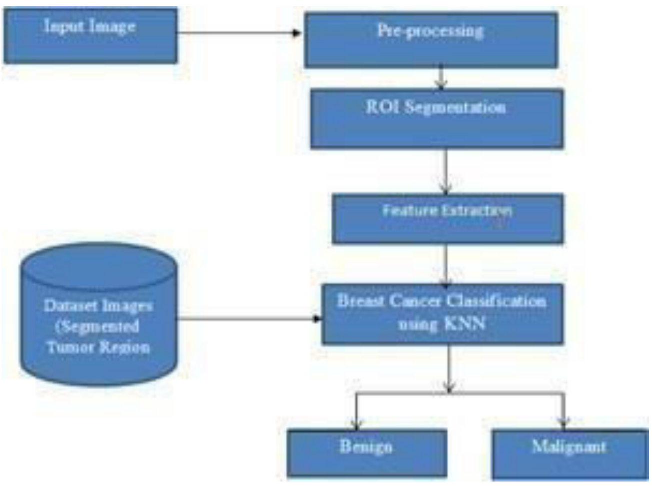


Fig (1).Overall Block Diagram

B. Block Diagram Description:

Breast cancer is a serious threat to women's life and health, and the morbidity and mortality of breast cancer are ranked first and second out of all female diseases. Early detection of lumps can effectively reduce the mortality rate of breast cancer. The mammogram is widely used in early screening of breast cancer due to its relatively low expense and high sensitivity to minor lesions. In the actual diagnosis process, however, the accuracy can be negatively affected by many factors, such as radiologist fatigue and distraction, the complexity of the breast structure, and the subtle characteristics of the early-stage disease. The computer- aided diagnosis (CAD) for breast cancer can help address this issue. The classical CAD for breast cancer contains three steps: (a) finding the Region of Interest (ROI) in the preprocessed mammogram, and hence locating the region of the tumor. (b) Then, extracting features of the tumor based on expert knowledge, such as shape, texture, and density, to manually generate feature vectors. (c) Finally, diagnosing benign and malignant tumors by classifying these feature vectors.

IV. HARDWARE AND SOFTWARECOMPONENTS

The hardware tools used are:

Processor : Pentium Dual Core 2.00GHZ

Hard Disk : 500GB.

Floppy Drive : 1.44Mb.

Keyboard : 110keys

Enhanced Mouse : Logitech.

RAM : 4GB

The software tools used are:

Coding Language : MATLAB

Tool : MATLAB R2013a

V. MODULE DESCRIPTION

A. Input:

Read and Display an input Image. Read an image into the workspace, using the image read command. In image processing, it is defined as the action of retrieving an image from some source, usually a hardware-based source for processing.

B. Preprocessing:

Data sets can require preprocessing techniques to ensure accurate, efficient, or meaningful analysis. This technique consists of resize the input image and converting the input image into gray scale image and black and white image using filters. Data cleaning refers to methods for finding, removing, and replacing bad or missing data.

C. Segmentation:

The technique of partitioning the image into segment can be defined as image segmentation. Considering the similar property, segmentation is implemented. This similar property is our propounded approach implements the Morphology based segmentation techniques. This technique aids in the extraction of important image characteristics, based on which information can be easily perceived. Then we use different morphological operations like DILATION, EROSION, AREA OPENING, CLOSING, BORDER CLEARING and etc.

 ISSN: 2582-3930

D. Feature Extraction:

Transforming the input data into the set of features is called feature extraction. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. While still describing the data with sufficient accuracy. Here we use geometrical based feature extraction method like Area, Diameter, Perimeter and Texture based feature extraction method like GLCM (Grey level co- occurrence matrix) for feature extraction. The GLCM gives the texture features of the test image like contrast, correlation, energy and etc. Then the region based features gives the various different features of the input image like area, diameter etc. From the all above extracted features we have to identify the best features that are related to differentiate the benign and malignant cancers.

E. KNN CLASSIFICATION:

The KNN (K-Nearest Neighbor) binary (as two class) is given more accurate data classification which beneficial to select k as an odd number which avoids the irregular data. The KNN procedure is the technique in ML procedures: It is an object which classified through a mainstream selection of its neighbors, with the determination assigned occurrence for most mutual class amongst its k nearest neighbors (k is a positive integer, classically small). Classically Euclidean distance is used as the distance metric; however, this is only suitable for endless variables. KNN is a new process that deliveries all available cases and categorizes novel cases built on an evaluation quantity (e.g., distance functions). KNN procedure is identical simple. It works built on a minimum distance from the interrogation instance to the training samples to regulate the K-nearest neighbors. The information for KNN procedure contains numerous attribute which will be used to categorize. The information of KNN can be any dimension scale from insignificant, to measurable scale.

VI. BREAST CANCER DATASET

Dataset for training is obtained from Breast Image Database. CT scans of both large and small tumors saved in Digital Imaging and Communications in Medicine

In photography and computing, a gray scale or gray scale digital image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. Images of this sort, also known as black-and-white, are composed exclusively of shades of gray, varying from black at the weakest intensity to white at the strongest.

Gray scale images are distinct from one-bit bi-tonal black-and- white images, which in the context of computer imaging are images with only the two colors, black, and white (also called bi-level or binary images). Gray scale images have many shades of gray in between. Gray scale images are also called monochromatic, denoting the presence of only one (mono) color (chrome).

Figure 2 shows the gray scale image:

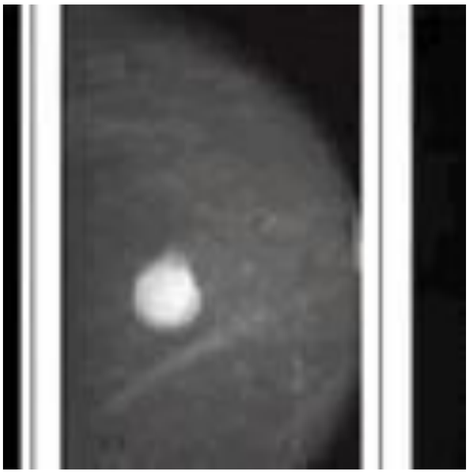


Figure (2) gray scale image

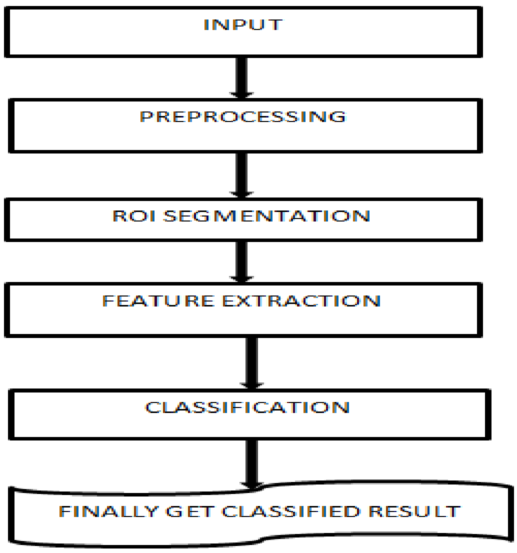
VII. K-Nearest Neighbour (K-NN)

In K-NN, an object is classified by the majority of its neighbors. The value of K determines the number of neighbours to be considered for the classification. Figure 1 depicts the flow chart of K-NN. In the test data, we select the distance metric and k value to be used. Here k value represents the number of neighbors to be considered for prediction. After based on distance selected, difference of test data from each training sample is computed. Then k minimum distances are selected, based on the majority of their category the class is assigned to the test data.

.

VIII. BLOCK DIAGRAM OF WORKFLOW

Figure (3) Methodology Work flow



ISSN: 2582-3930

IX. ADVANTAGE OF PROPOSED SYSTEM.

- In this paper instead of histopathological images we process with mammogram images.
- Here we also do the classification task like benign or malignant cancer.
- The machine learning processes needs only the normal hardware requirements.
- The process is easy to understand.

X. RESULTS ANDDISCUSSIONS

In our proposed system we used 241 data set images extracted from patients. This images are used to training, testing and detect the cancerous part. Each image in our datasets has resolution of 1024x1024.In our data sets consists of 169 normal images, 72 effected images.

OUPUT IMAGES:

Figure (4) Benign Input Image

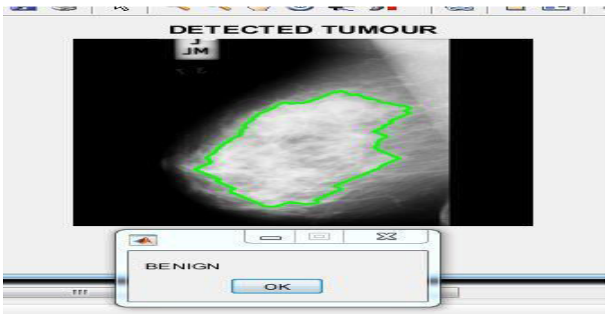


Figure (5) Benign Detected Image

Figure (6) Malignant Input Image

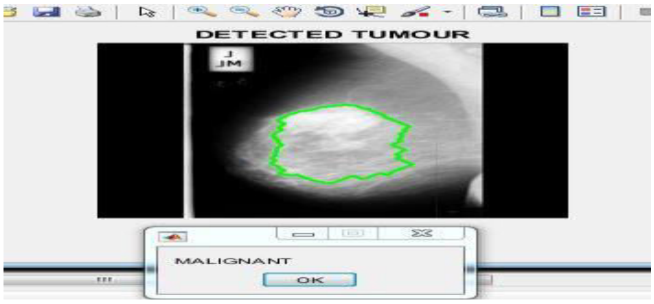


Figure (7) Malignant Detected Image

XI. CONCLUSION

It proposes novel breast cancer detection and classification method which uses region based and texture based features for breast cancer representation and classification. The proposed method was evaluated on a set created from mammogram database images. We have showed that the proposed method is efficient and effective for the detection and classification of benign and malignant breast cancers effectively. In this paper we combine region features and texture features, taking the doctor's experience and the essential attributes of the mammogram into account at the same time.

XII. REFERENCES

[1] "Breast-cancer early detection in low-income and middle-income countries: do what you can versus one size fits all," Lancet Oncology, vol. 12, no. 3, pp. 306– 312, 2011.

[2] C.Lerman,M.Daly,C.Sands,A.Balshem,E.Lustbader,T.Hegg an, L. Goldstein, J. James, and P. Engstrom, "Mammography adherence and psychological distress among women at risk for breast cancer," Journal of the National Cancer Institute, vol. 85, no. 13, pp. 1074–1080, 1993.

[3] P. T. Huynh, A. M. Jarolimek, and S. Daye, "The false-negative mammogram," Radiographics, vol. 18, no. 5, pp. 1137–54,1998.

[4] M. G. Ertosun and D. L. Rubin, "Probabilistic visual search for masses within mammography images using deep learning," in IEEE International Conference on Bioinformatics and Biomedicine, 2015, pp.1310–1315.

[5] S. D. Tzikopoulos, M. E. Mavroforakis, and H. V. Georgiou, "A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry," Computer Methods and Programs in Biomedicine, vol. 102, no. 1, pp. 47–63,2011.

[6] D. C. Pereira, R. P. Ramos, and M. Z. do Nascimento, "Segmentation and detection of breast cancer inmammograms combining wavelet analysis andgenetic algorithm," Computer Methodsand Programs in Biomedicine, vol. 114, no. 1, pp. 88–101,2014.

[7] S. A. Taghanaki, J. Kawahara, B. Miles, and G. Hamarneh "Paretooptimal multi-objective dimensionality reduction deep auto-encoder for mammography classification," Comput Methods Programs Biomed, vol. 145, pp. 85–93, 2017.

[8] X. W. Chen andX. Lin, "Big data deep learning:Challenges and perspectives," IEEE Access, vol. 2, pp. 514–525, 2014.

[9] K. Ganesan, U. R. Acharya, C. K. Chua, L. C. Min, K. T. Abraham, and K. H. Ng, "Computer-aided breast cancer detection using mammograms: A review," IEEE Reviews in Biomedical Engineering, vol. 6, pp. 77–98,2013.

[10] X. Sun, W. Qian, and D. Song, "Ipsilateral-mammogram computer-aided detection of breascancer," Computerized Medical Imaging and Graphics the Official Journal of the Computerized Medical Imaging Society, vol. 28, no. 3, pp. 151–158,2004.

[11] N. Saidin, U. K. Ngah, H. Sakim, and N. S. Ding, Density based breast segmentation for mammograms using graph cut and seed based region growing techniques. IEEE Computer Society,2010.

[12] S. Xu, H. Liu, and E. Song, "Marker-controlled watershed for lesion segmentation in mammograms," Journal of Digital Imaging, vol. 24, no. 5, pp. 754–763,2011.

[13] K. Hu, X. Gao, andF. Li, "Detection of suspicious lesions by Adaptive thresholding based on multiresolution analysis in mammograms," IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 2, pp. 462– 472,2011.

[14] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A. K. Davison, and R. Marti,"Automated breast ultrasound lesions detection using convolutional neural networks," IEEE J Biomed Health Inform, vol.22, no. 4, pp. 1218–1226,2017.

[15] K. C. Jr, L. M. Roberts, K. A. Shaffer, and P. Haddawy, "Construction of a bayesian network for mammographic diagnosis of breast cancer," Computers in Biology and Medicine, vol. 27, no. 1, pp. 19–29,1997.

[16] Z. Wang, G. Yu, Y. Kang, Y. Zhao, and Q. Qu,"Breast tumor detection in digital mammography based on extreme learning machine," Neurocomputing, vol. 128, no. 5, pp. 175–184,2014.

[17] Y.Qiu,Y.Wang,S.Yan,M.Tan,S.Cheng,H.Liu,andB. Zheng, "An initial investigation on developing a new method to predict short-term breast cancer risk based on deep learning technology," in Medical Imaging 2016: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2016.

[18] W. Sun, T. L. Tseng, B. Zheng, and W. Qian, A Preliminary Study on Breast Cancer Risk Analysis Using Deep Neural Network. Springer International Publishing, 2016.