International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

EARLY-STAGE BREAST CANCER DETECTION USING MACHINE LEARNING

DR. Vijaya R Pawar
Bharati Vidyapeeth's CoE for Women, Pune, India
Radhika Ajabe
Bharati Vidyapeeth's CoE for Women, Pune, India
Srushti Burhade
Bharati Vidyapeeth's CoE for Women, Pune, India
Aachal Jaiswal
Bharati Vidyapeeth's CoE for Women, Pune, India

Abstract –Segmentation plays significant role in detection of the breast tumor.Nowadays Artificial intelligence is used to perform the segmentation. Using different image processing tools combined and successively same result can be achieved . This tools are are playing very important role in identifying breast tumor by processing a image. In this proposed method to identify the desired regions of interest from the MRI images otsu's threshold segmentation is used with different image processing steps . Results obtained by this research have shown that accuracy rate has been increased compared to other existing approaches and and there is no need of human intervention .

Keywords: Breast Tumor**,** MRI image, median filter, Contrast adjustment, Edge detection, Morphological operations.

I.INTRODUCTION

Breast cancer is considered as the one of the common cause of cancer death among women's. [1] According to World Health Organization (WHO), it is given that 6,27,000 females died because of this cancer in 2018 which is about 15 % women in the world and this rate is rapidly growing. Because of this early detection of breast cancer is very important to considerably decrease the mortality rates because according to a statistical report, it was found that the survival rate of women diagnosed with breast cancer is almost 96 % over five years. This is because, their tumors were identified at the early stage. The diagnosis of breast cancer can be done by two ways, using imaging tests and biopsy.

Most common methods used for the diagnosis of breast cancer are mammography, ultrasound, and magnetic resonance imaging . mammography is the most standard breast tumor screening technique. [2] Mammography method is most conventional method screening method among this three. Mammography generally has less efficiency for patients under age of 40 and for the dense breasts, it also shows less sensitivity to tiny tumors as well as it done not give any possible outcome of the disease. Ultrasound is another tool used for detection and diagnosis of breast cancer.But it contributes very little to early detection of breast cancer. One more method of breast cancer detection and diagnosis is Magnetic resonance imaging (MRI).It is very efficient method than other two methods and is capable of detecting the breast cancer t early stage which cannot be achieved by other two methods. Nowadays various Computer Aided (CAD) breast imaging programs have been developed to overcome the various drawbacks of MRI. This drowbacks includes need of significant time to acquire, process, and interpret images.

II . ORGANIZATION OF THE PAPER

In this publication, section III discusses the Literature survey of earlier publications. Sections IV discuss the methodology and V discuss setup of the system. . Section VIII discusses the obtained result and section VI discusses the future scope of the system. The last section IX discusses the conclusion of this paper.

III. LITERATURE SURVEY

Breast Cancer is the most popular and growing disease in the world. Breast Cancer is mostly found in women.[1] Early detection is a way to control breast cancer. Many cases are handled by early detection and decrease the death rate. Much research works have been done on breast cancer. The Most common technique that is used in research is machine learning. Many previous types of research conducted through machine learning. Machine learning algorithms like decision tree, KNN, S VM, naive bays, etc. gives better performance in their field. But nowadays, a newly developed technique is used to classify breast cancer. The newly developed technique is deep learning. Deep learning is used to overcome the drawbacks of machine learning. Deep learning technique that is mostly used in data science is Convolution neural network, Recurrent neural network, deep belief network, etc. Deep learning algorithms give better results as compared to machine learning. It extracts the best features of the images. In this research, CNN is used to classify the images. [2] Our research is based on the images and CNN is the most popular technique to classify the images.[2] In the present paper, reviews of all authors are conducted. The research is about the prediction of breast cancer using machine learning techniques. Prediction of cancer type is one of the crucial aspects of statistical analysis. In the paper, we presented a hybrid approach for the development of a tool that can predict breast cancer with the help of Machine Learning techniques. [4] The main reason behind developing this tool is that the number of bio informatics tools for the prediction of the target class is scarce. Breast cancer is common and one of the most deadly diseases found in women. Prevention of breast cancer in its early stage is difficult as the cause of the disease is still not known. X-ray mammography is a very useful technique for screening and early-stage detection of cancer. Breast cancer is common and one of the most deadly diseases found in women.[5] Prevention of breast cancer in its early stage is difficult as the cause of the disease is still not known. X-ray mammography is a very useful technique for screening

IV . PRESENT METHODOLOGY

This model will use labels in the coder to label categorical data. Label encoder is the part of sci-kit learn library in Python and is used to convert categorical data, or text data, into numbers that our predictive model can better understand. The data we are using is generally split into training data and test data. The training set contains a known output and the model learns from this data to generalize to order data later on. To check the correct predictions we have to check the confusion matrix object and add the predicted result diagonally which will be the number of correct predictions and then divide by the total number of predictions. This model is built using android studio IDE. The proposed methodology will help to distinguish between malignant and benign tumors at a faster rate. Despite being a complex and complicated classifier, CNN can extract vital features from image automatically without depending on prepossessing. It is more efficient because it filters the important parameters of image and also is flexible as it is capable to work exceptionally well on image data. The main focus is to differentiate between malignant and benign tumors using Convolution Neural Network with Keras in the back-end and then analyze the result to see how the model can be useful in a practical scenario. The following are some steps performed for model building and evaluation: 1. Importing all the essential libraries. 2. Making a dictionary of images and labels 3.

Labels are based on image category. 4. Normalization of the image set. 5. Splitting data from dataset into training and testing sets. 6. Building Architecture of the model (CNN). 7. Cross Validating Model (comparing with different classifiers). 8. Testing Model.

1. Dataset -Well dataset is an essential requirement to produce a good method for the detection of Breast Cancer. The collection of a dataset is highly difficult due to the unavailability of samples and the confidentiality of the patient's demographic information. In this case, the Dataset is taken from the M. G Cancer hospital & research institute, Vishakhapatnam, India. The dataset comprises of 2556 image samples of over 683 patients of various levels. The summary of the taken dataset is given below. The given dataset consist a set of histopathology images. These images are then classified as benign and malignant tumors. The pre-processed images of the benign and malignant are applied to the proposed methodology to achieve the effective classification of Breast Cancer cases. Further, the proposed method has produced highly accurate and efficient results when compared to the existing methods data augmentation is employed to enlarge the dataset to lessen the problem of limited data size. Natural images are generally analyzed in a bottom-up approach. But majority of the Medical images are analysed by a top-down approach. It is perceived that the augmentation techniques applied to natural images will not hold well for medical images.

2. Image processing-Image processing is a method to perform some operations on an input image to get an enhanced image or to take some useful information from it which is required for application. Image processing is a type of signal processing for which the input is generally an image and the output may be an image and characteristics/features associated with that image. The input image went through a gray-scale conversion before the pre-processing phase. On the converted image, image filtering and enhancement operations were incorporated to reduce the noise and increase the contrast of the images.

A)Noise Reduction: To reduce unwanted regions, to improve smoothness, sharpness, and edge improvement; filter operation was conducted on the test images. In the place of other filters such as the Gaussian filter and mean filter, a median filter was implemented due to its less blurry impact and edge-preserving characteristics. In addition, the median filter can effectively reduce the salt and pepper noise incorporated with MRI images.

B)Contrast Adjustment: The image intensity values are adapted after noise reduction to remove regions close to bright white when the background is bright and dark. Contrast stretching was adopted to map the processed image intensity values into new intensity values so that low-intensity regions become lower and high-intensity regions become higher. To perform this step gray-scale images were converted to RGB images and then RGB images were divided into 3 channels (R, G, B). Instead of using contrast adjustment for all the channels only Green and Blue channels were considered as a fact that it prevailed under and over-segmentation issues during segmentation. Finally, images were converted back to grayscale images which are then used for segmentation

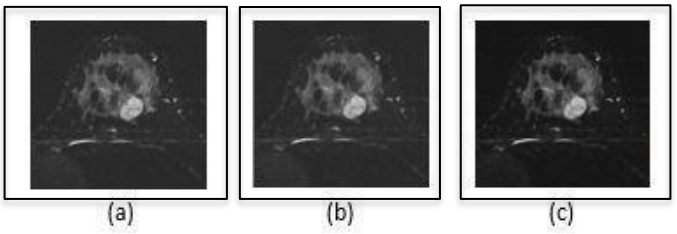


fig shows that during pre-processing stage, filter operation and contrast adjustment were done on MRI test images. A sample MRI test image of Fig.2 (a) was then used for pre-processing shown in Fig.3 (a), median filter was applied on preprocessed image (b), and Contrast adjusted results in (c).

c) Segmentation-Threshold-based image segmentation is a simplified but effective approach to segmenting images with bright objects on a dark surface or vice-versa by converting multilevel images into binary images. A threshold value, T divides the image pixels into multiple regions and distinguishes objects from the background. Binary images from gray-level images are acquired by converting all pixels below a certain threshold to zero and all pixels above that threshold to one. This study implemented Otsu's threshold because of its simplicity and the variance in gray-level intensities that requires distinct thresholds for distinct images. At the initial stage, the OTSU threshold approach was used to get a threshold value and then using coordinates that threshold value was fed again to the OTSU method to find the cut-off threshold value. This has helped this study images to pick the threshold cut-off value automatically rather than

setting a threshold value for each as a general value.

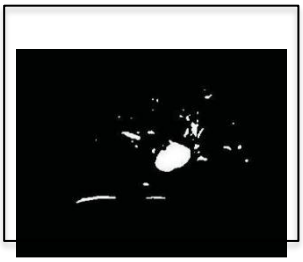


Figure 4. Result of Otsu's threshold-based segmentation

A)Detection of Tumor Region- The segmentation technique split the image into several regions including the breast tumor region. At this point, to differentiate the tumor region from the other regions several morphological operations were conducted on the segmented image. a hole-filling operation is used to fill holes in the segmented binary image. Then the system removes all objects connected to the image border as it is necessary to remove border-touching regions during the identification of ROI. During segmentation, the region of interests was successfully segmented

but the objects were not fully detected because some regions were still available which do not match the breast tumor detection feature criteria. Also, shrink and dilate morphological operations are implemented on the binary image to remove pixels to reduce objects without to a point and objects with holes shrink between each hole and exterior boundary to a connected ring halfway.

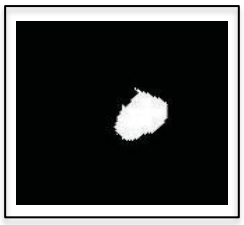


figure.5 Output image with detected tumor

3. Feature extraction- The entropy, geometrical and textural features are extracted from the pre-processed images. Entropy (E) is the estimation of haphazardness that is utilized to describe the texture of the input image. Shape features play an important role to distinguish the characteristics between normal and malignant cells. In textual features, each picture is partitioned into N sub-squares and quantized.

4. Feature Selection-The importance of feature selection in a machine learning model is inevitable. It turns the data to be free from ambiguity and reduces the complexity of the data. Also, it reduces the size of the data, so it is easy to train the model and reduces the training time. It avoids things of data. Selecting the best feature subset from all the features increases the accuracy. Some feature selection methods are wrapper methods, Iter methods, and embedded methods.

5. Classifier- A Convolutional neural network is a deep learning algorithm that can take in an input image, assign importance(learnable weights and biases to various aspects and objects in the image and be able to differentiate one from the other.

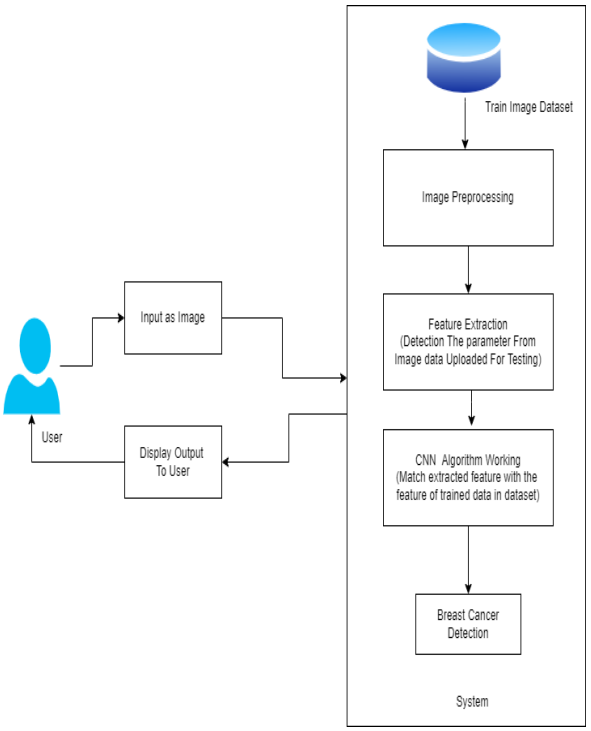


Fig 6. Block Diagram for Breast Cancer Detection

V. Experimental SETUP

The following diagram shows the setup utilized in this methodology. In this module admin is present. In this module, the Admin has to log in by using a valid username and password. After login was successful, he can do some operations such as View All Users and Authorize.

View All E-Commerce Websites and Authorize, View All Products and Reviews, View All Products Early Reviews, View All Keyword Search Details, View All Products Search Ratios, View All Keyword Search Results, and View All Product Review Rank Results. View and Authorize Users • In this module, the admin can view the list of users who are all registered. In this, the admin can view the user's details such as user name, email, and address and the admin authorizes the users. View Charts Results. View All Products Search Ratios, View All Keyword Search Results, and View All Product Review Rank Results. Ecommerce User.

In this module, there are n numbers of users present. Users should register before doing any operations. Once user registers, their details will be stored in the database. After registration is successful, he has to login

in using an authorized user name and password Once Login is successful user will do some operations like Add Products, View All Products with Reviews, View All Early Product Reviews, and View All Purchased Transactions. End User.

In this module, there are n numbers of users present. Users should register before doing any operations. Once user registers, their details will best or to the database. After registration is successful, he has to log in by using an authorized username and password. Once Login is successful user will do some operations like Manage Account, Search Products by keyword and Purchase, View Your search transactions, and view.

RESULT

The performance of the proposed method is measured based on

the following factors: i) Whether it can differentiate between the

images of a normal breast and a breast with the tumor, ii)

Whether it can correctly detect the tumor region or not. The

accuracy has been derived using the confusion matrix. Accuracy

is calculated as,

Accuracy = T P + T N /P + N

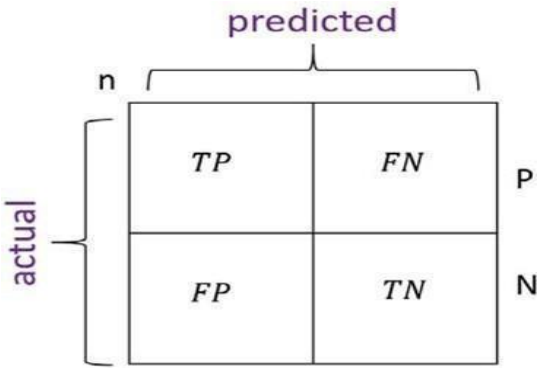


fig 7. Confusion Matrix

A.Result Analysis of the Proposed Method-

Based on the criteria of breast tumor size test images were divided into two batches. Table 1 demonstrates that it was possible to detect the tumor regions with an accuracy of 100% for batch 1 and 93.33% for batch 2. This study's results indicate that the overall accuracy rate for 150 test images is 97.33% which output armed many other existing techniques[14]. Table I gives the detection ratio of the proposed system.

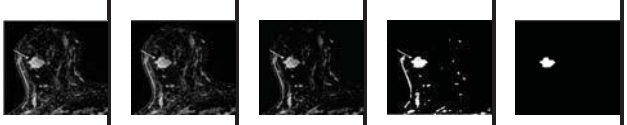
Table I
Detection ratio of proposed system

Dataset (No. of images)	true positive	true negative	false positive	false Negative	Accuracy (%)
Batch1 (90)	84	6	0	0	100%
Batch2 (60)	55	3	5	0	93.36 %
Total = 150	136	8	5	0	97.23 %

Experimental Outcome

The potential experimental results of this proposed method of research are shown in figure 7. It illustrates two breast tumor images from two different batches of the dataset. Fig. 7(a) shows the input images. The images after reducing the noise using median filter are shown in the in fig. 7(b).Image enhancement activity was introduced to improve image contrast shown in fig. 7(c).

Table 2

Input Image (a)	Noise Reduction using Median Filter(filtered image) (b)	Image Enhancement using Contrast Adjustment (c)	Image Segmentation Using OTSU Method (segmented image) (d)	Identified Region (e)
				

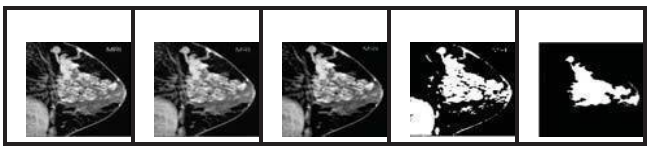


Figure 8 (a) Input Image (b) Noise Reduction using Median Filter(filtered image) (c) Image Enhancement using Contrast Adjustment (d) Image Segmentation Using OTSU Method(segmented image)
(e) Identified Region

VI. FUTURE SCOPE

In the future, various advances in Sensors, Contrast agents, Molecular methods, and Artificial Intelligence will help detect cancer-specific signals in real time at ver early stage. To reduce the burden of cancer on society, risk-based detection and prevention need to be cost-effective and widely accessible.

IX. CONCLUSION

In this work, a robust breast tumor detection system is designed to assist the pathologists to detect breast tumors. The proposed system automatically segments the image and identify the tumor region of the image from other regions. The precision can be still increased if image quality remains similar for different laboratories and experts' intra and inters-observation variability issues can be minimized.

Fig 9. Breast Cancer Detection

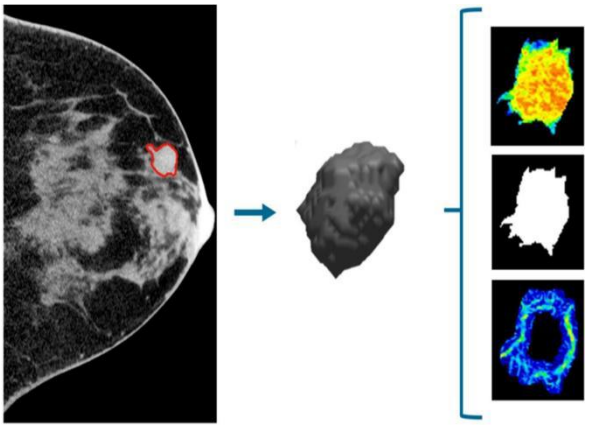


Fig 9. Breast Cancer Detection

VIII REFERENCES

1. F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L. A. Torre, and A. Jemal, "global cancer statistic 2018 : GLOBOCAN estimates of incidence and mortality rate worldwide for 36 cancers in the 185 countries," ca. Cancer J. Clin., vol. 68, no. 6, pp. 394–424, 2018.

2.H. Gelband, P. Jha, R. Sankaranarayanan, and S. Horton, Disease Control Priorities, (Volume 3): Cancer. The World Bank, 2015.

3.J. Rimal, A. Shrestha, I. K. Maharjan, S. Shrestha, and P. Shah, "risk assessment of smokeless tobacco among oral pre-cancer and cancer patients in the eastern developmental region of Nepal," asian pacific J. Cancer prev., vol. 20, no. 2, pp. 411–415, 2019.

4.J. G. Doss, W. M. Thomson, B. K. Drummond, and R. J. R. Latifah, "validity of the FACT-HN among Malaysian oral cancer patients," oral oncol., vol. 47, no. 7, pp. 648–652, 2011.

5. H. Amarasinghe et al., "economic burden of managing oral cancer patients in Sri Lanka: a cross-sectional hospital-based costing study," BMJ Open, vol. 9, no. 7, 2019.

6. Gayathri BM, Sumathi CP., "comparative study of relevance vector machine with various machine learning techniques used for detecting breast cancer," In Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on 2016 Dec 15 (pp. 1-5). IEEE.

7. forsyth AW, barzilay R, hughes KS, Lui D, lorenz KA, enzinger A, tulsky JA, lindvall C., "machine learning methods to extract documentation of breast cancer symptoms from electronic health records," journal of pain and symptom management. 2018 Jun 1;55(6):1492-9.

8.A. I. for Cancer Research. (2018) Breast cancer statistics. [Online].Available:https://www.wcrf.org/dietandca cancer/cancer trends/breast-cancer-statistics

9.V. S. Subbhuraam, E. Ng, U. R. Acharya, and O. Faust, "breast imaging: A survey," World Journal of clinical oncology, vol. 2, p. 171, 01 2014. [3] B. N. Hellquist, K. Czene, Hjalm, L. Nystr ¨ om, and H. Jonsson, effectiveness of population-based service screening with mammography for women ages 40 to 49 years with a high or low risk of breast cancer: Socioeconomic status, parity, an age at birth of a first child," cancer, vol. 121, no. 2, pp. 251–258, 2015.

10. W. A. Berg, Z. Zhang, D. Lehrer, R.Jong, E. D. Pisano, R. G. Barr, M. Bohm- V ¨ Velez, M. C. Mahoney, W. P. Evans, L. H. Larsen ´ et al., "Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk," Jama, vol. 307, no. 13, pp.1394–1404