

# Detection of Breast Cancer and Segmentation of Abnormalities Using Deep Learning and Image Processing Techniques

<sup>1</sup>Swapna Yenishetti, <sup>2</sup>Lakshmi Panat, <sup>3</sup>Aishwarya Bharambe

<sup>1</sup>Principal Technical Officer, <sup>2</sup>Associate Director, <sup>3</sup>Student  
<sup>1&2</sup> HPC-MBA Group,

<sup>1&2</sup> Centre for Development of Advanced Computing, Pune, India

**Abstract:** Among all cancers detected breast cancer rates are alarming. Ministry of Health and Family Welfare under Government of India has ranked breast cancer as commonly occurring cancer among women, both in terms of incidence as well as mortality. Mortality rate can be reduced if it is diagnosed early, using techniques like Mammography, X-Ray, MRI (Magnetic Resonance Image), Computed Tomography (CT) etc. In this paper, we propose a method for detection of breast cancer and segmentation of abnormalities for MRI scans. We first classify the MRI scans as malignant or benign using deep learning algorithm and then pass the malignant images for segmentation and sizing. Segmentation of abnormalities (e.g. tumor) is done using image processing and the size (length and breadth) of tumor is calculated. For classification we used Resnet whose performance is measured on both CPU and GPU systems. Resnet has advantage of tolerance to vanishing gradient compared to other deep neural networks. Instead of widening the network, Resnet increases the depth of network which results in less trainable parameters. In only about 5 epochs Resnet gets trained, with good accuracy. During training phase, on CPU system, Resnet demonstrates 98.07% accuracy in 48 min whereas on GPU system it shows 96.96% accuracy in 24 min. During inference phase, Resnet exhibits 99.56% and 99.07% accuracy on CPU and GPU systems respectively.

**Keywords:** Breast Cancer, Deep Learning, Magnetic Resonance Imaging (MRI), Segmentation, Image processing

## I. INTRODUCTION

Breast Cancer is a disease in which cells in the breast tissue grow in an uncontrolled way. Observing the growth of cells helps in staging of breast cancer. Most breast cancers begin in ducts or lobules. Breast cancer not only in India but across the world is increasing at alarming rate. Earlier Breast cancer occurred primarily in the age group of 50-64 years, but now-a-days it is being diagnosed in younger women too. Breast cancer is not only increasing in terms of incidence but also in terms of mortality. The incidence can be reduced through early diagnosis, preventing spread to other parts of body through blood/ lymph vessels (metastasis). Commonly occurring breast cancer types include Invasive ductal carcinoma and Invasive lobular carcinoma.

There are various screening modalities for early detection of cancer like Mammography, Ultrasound, Computed Tomography (CT), X-Ray, Magnetic Resonance Imaging (MRI) etc. Every screening has its own advantages and disadvantages. Advantage of mammography is its suitability in diagnosing any abnormality in mammograms easily, whereas disadvantage is its non-suitability for mammogram with dense tissue, resulting in incorrect diagnosis of 10-30% of breast cancer cases [1]. Ultrasound is easily available everywhere at low cost, but abnormality detection is dependent on operator's experience and requires high resolution imaging. Computed Tomography based screening for breast cancer involves radiation and includes contrasting agent which cannot be recommended patients who are allergic to it. Magnetic Resonance Imaging (MRI) also uses contrasting agent and can easily detect abnormality or tumor, and is safe as the patient is not exposed to radiation. MRI is not recommended for patients with cardiac pacemaker, surgical and vascular clips, etc.

These imaging modalities can find the area where the abnormality is present but does not give indications about abnormality being benign or malignant. For this reason, additional tests like biopsy need to be done, which is a painful process, and results in monetary loss. Hence there is a need to have a computer aided system to help doctors make precise decisions and reduce biopsies.

In this paper we are proposing a model which will first detect malignancy in breast MRI scan using Deep learning model (Resnet) and then characterize the abnormal part through Image processing, and segmentation methods. The length and breadth of each abnormality is calculated on the basis of pixel values. Segmentation of abnormality uses region growing algorithm.

This paper is organized in sections as follows - in section II, Related work on detection of breast cancer is discussed, in section III, Objective, information about dataset used, system configuration and the software used in our research are explained. Section IV discusses the proposed method, architecture, model and algorithm. Finally, in section V results and direction of future work are outlined.

## II. RELATED WORK

Several studies have been done on computer aided systems for detection of the abnormalities for imaging modalities helping doctors make accurate decision. Segmentation is also an important characterization process which helps to know the exact area where abnormality is present, and aids in sizing.

In [2], Mixture Ensemble of Convolutional neural network (ME-CNN) is used which can be used as effective tool by radiologists for analyzing breast dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) images. Two steps are used in this system: i) Tumor Segmentation based on the morphological information of masses in image and intensity, ii) Tumor is classified into malignant or benign using ensemble of ME-CNN model. In [3], You Only Look Once (YOLO) CNN model for this CAD system. The system consists of four steps: pre-processing of images, feature extraction, mass detection and mass classification using fully connected network (FCN). Trained YOLO model is used to classify mass as benign or malignant. In this study the overall accuracy of detection of mass was 96.33% and of classification was 85.52%. There are many CAD systems based on CNN model such as VGGNet[4,5], Alexnet[6], UNET[6], etc.,

We reviewed the techniques for classification of breast cancer using Deep Learning like - BNN, VGG-16, YOLO, etc., some of the limitation of these techniques are: BNN has disadvantage of getting in premature convergence, VGG-16 if applied very deeply then it faces vanishing gradient problem. We also reviewed the techniques for Segmentation using image processing which helps in detecting the abnormal areas and can provide guidance for further decisions such as biopsy and surgery by knowing the size of tumor, from segmented images. Segmentation can be done using various methods such as Otsu in combination with region growing [7], morphological operations [8] etc., where in morphological operations' threshold values need to be given explicitly.

## III. OBJECTIVE AND SYSTEM SPECIFICATION

### *Objective of the research:*

The main objective of this work is to classify breast cancer as malignant or benign, and if found to be malignant, segment the abnormality along with sizing (length and breadth), to guide optimal referrals for biopsy.

### *Dataset:*

We have used dataset from TCIA library which contains MRI scans for 88 patients with 105,050 images of around 60.8 GB, out of which 84 patients are examined using MRI scan, 17 patients through mammography, 3 through CT and 3 through PET. We used MRI scans of patients for whom clear results are given in spreadsheet (metadata with the dataset) - 1080 scans of post-contrast images were separated into benign and malignant folders. From the data, 80% (864 images) were used for model training, and 20% (216 images) were used to test the model.

The Breast scans collection contains cases that are high-risk normal, DCIS (Ductal carcinoma in situ), Fibroids, Lobular carcinomas. Each case has 3 or more distinct magnetic resonance (MR) pulse sequences from a Phillips 1.5 T (e.g. T2, STIR and Bliss). We have focused on Axial Images. As the size of images varies from patient to patient, images are resized to 448\*448 before training the model. Metadata includes a spreadsheet with BI-RADS score, MRI impression (e.g. position/ size if a mass was found), observations from pathology report (e.g. which breast, ER, PR, HER2).

<https://wiki.cancerimagingarchive.net/display/Public/QIN+Breast+DCE-MRI>

### *CPU & GPU Configuration:*

- CPU: 8 GB RAM, Quad core processor.
- GPU: 16GB RAM, NVIDIA graphics card. Run on three GPUs in parallel.

### *Software:*

- Mango: It is non-commercial software for visualizing, editing and analyzing medical images. It is a tool for creation and editing Region of Interest (ROI) within the images. Headers in medical images can be easily read using Mango but while exporting images this software does not retain DICOM format of images requiring usage of RadiAnt.
- RadiAnt DICOM Viewer: It is used for processing and displaying medical images in DICOM format (Digital Imaging and Communications in Medicine).
- Anaconda: It is a package management software with free and open-source distribution of the Python and R programming language for scientific computations (data science, machine learning, large-scale data processing, predictive analytics, etc.), that aims to simplify deployment.

## IV. PROPOSED METHOD

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads—the template will do that for you.

Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar.

The model development, consists of the following steps.

- First, a Neural Network is trained to classify MRI scans into suspected malignant or benign cases.
- The model is saved as h5 file in JSON format.
- For purpose of testing, unseen image is provided to the model for classification.
- If the input image is found to be “benign”, program is terminated, else if image is “malignant” then it is passed on to image segmentation module.
- In image segmentation phase, the abnormality is segmented using morphology methods like cannyedge, watershed, erode, dilate and remove small objects. The segmented image is then passed for measurement of abnormality.
- As part of sizing of abnormality/ tumor, the length and width is determined and saved to text file in output folder.

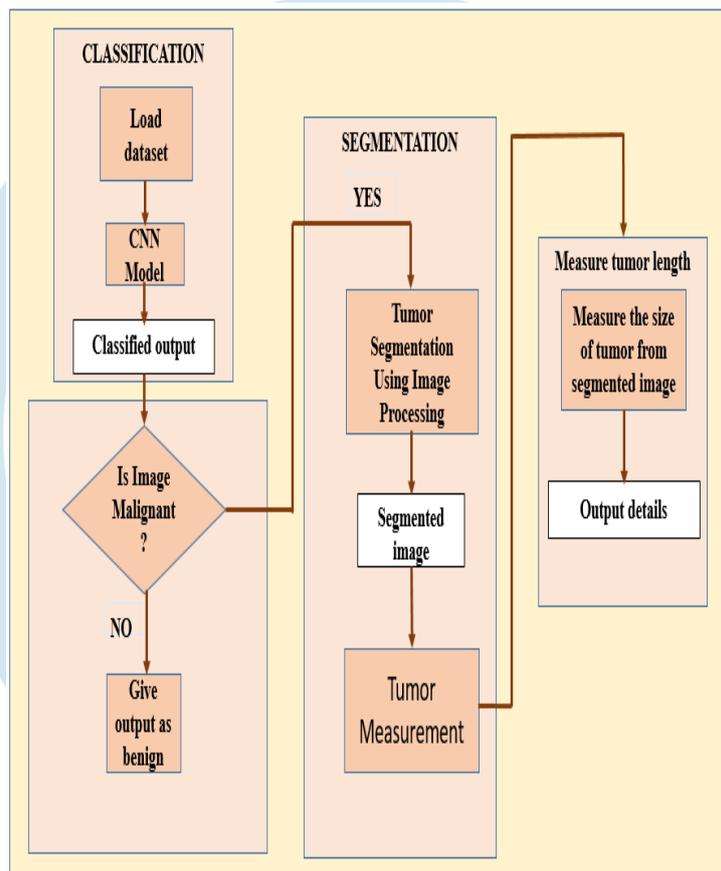


Figure 1 Architecture of the system

#### Data pre-processing:

We have used Post contrast scans i.e. the scans done after inserting the intravenous solution in patient’s body. The reason being ease of extraction of outline of abnormality, using segmentation method.

In post-contrast images, while visualizing through RadiANT, we observed that abnormality is clearly visible in scans from sequence number 40 to 68 during each scan.

Then we converted scans into an integer array and expanded the dimensions of scans as it should be three channels while passing to the Neural Network namely Image height, Image Width, Channel.

- Then, we added the image into a list, along with a label (“0” if the image is benign and “1” if the image is malignant).
- Then, the dataset is split into training and testing, normalized, labels are encoded using one hot encoding, so when the element in label list has value “0” it gets converted to [1.0 0.0] and if element in label list has value “1” it gets converted to [0.0 1.0].

The flow diagram for data pre-processing is shown in Figure 2.

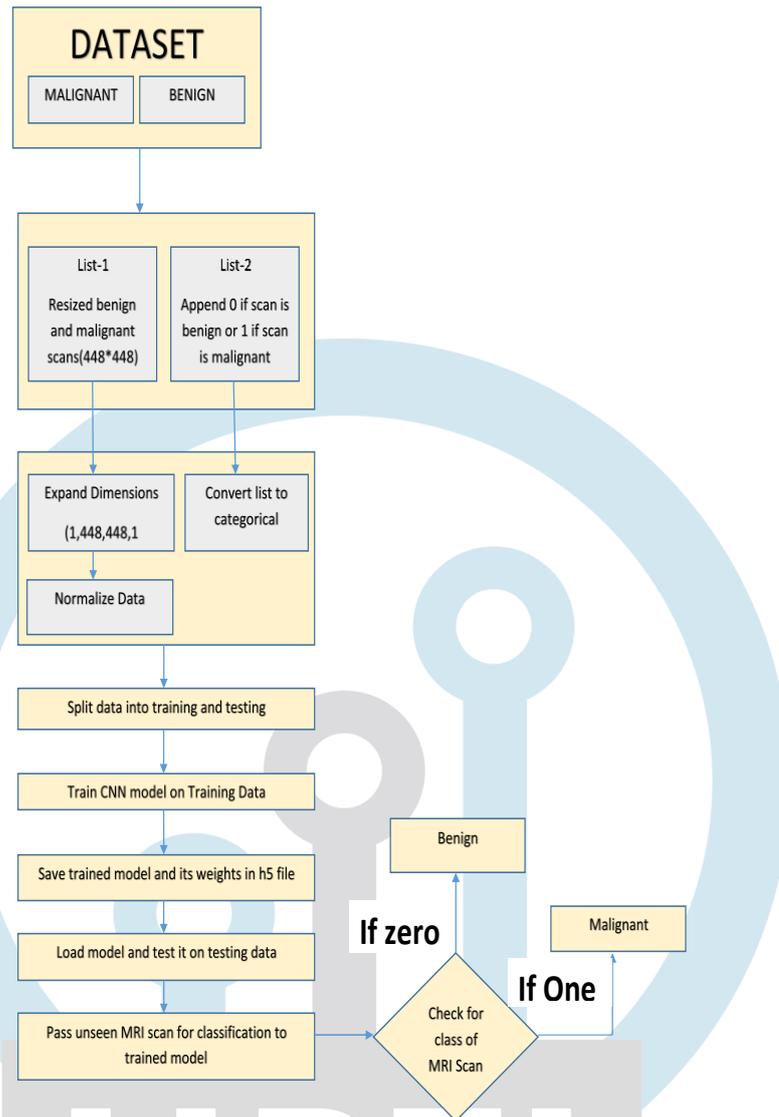


Figure 2. Flow for Data pre-processing

**Classification:**

For classification we used Resnet model as it increases the depth of network instead of widening it which results in fewer parameters for training, and tolerates vanishing gradient. For classification task we are using 50 layer deep residual network.

The classification is done to predict if the image is likely to be Benign or Malignant. The model for classification is trained on CPU and GPU system separately to understand the performance on both systems.

**Segmentation:**

In region growing segmentation method, we first plot histogram to analyze the intensity of pixels using which threshold values are calculated for thresholding. Then two seed values are initialized one for background and one for foreground, and checked every pixel's intensity distance from two seeds. Then it is labeled as the distance from seeds. Then the minimum distance label is appended to a list converted to array and this array is then reshaped into image size. Then the image is eroded to remove false positives and then it is dilated to get back original resolution. Flow for segmentation is shown in fig.3

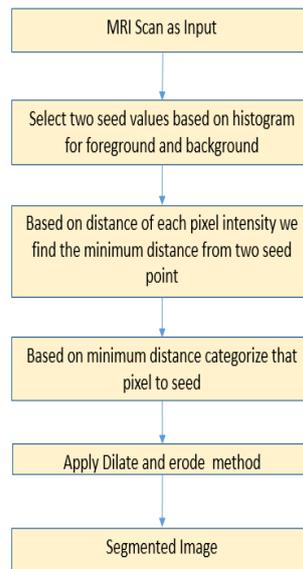


Figure 3. Flow for segmentation using Region growing

**Abnormality/ Tumor Sizing:**

From the segmented image we can see that there can be more than two discontinuous objects that are tumors. We have to identify number of objects according to connected components.

For each component we find pixel position of first and last position horizontally for measuring the length of tumor. Flow for tumor size measurement is as shown in fig 4.

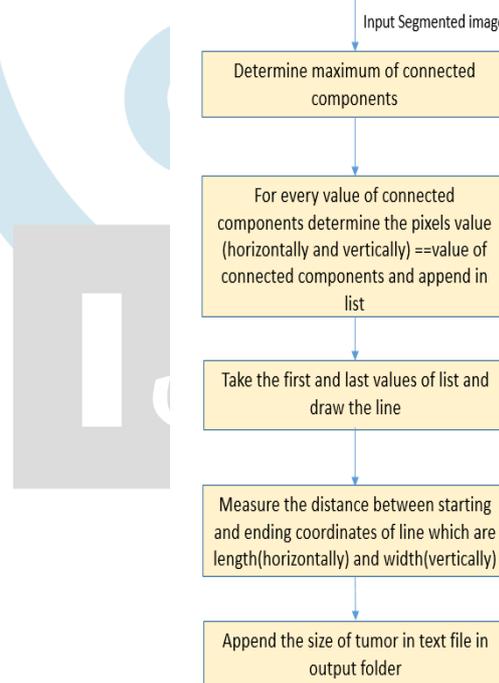


Figure 4. Flow for measuring tumour size

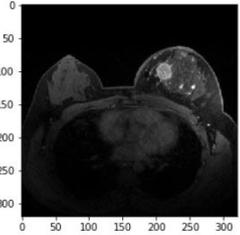
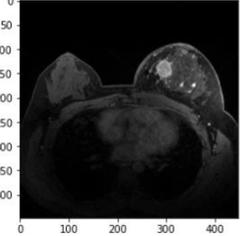
**V. RESULTS, CONCLUSION AND FUTURE SCOPE**

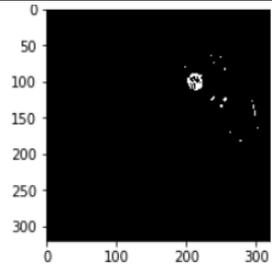
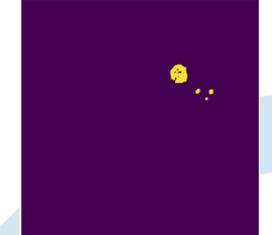
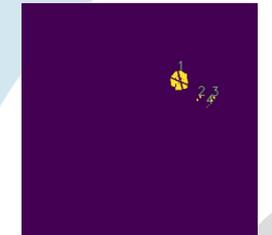
**Results:**

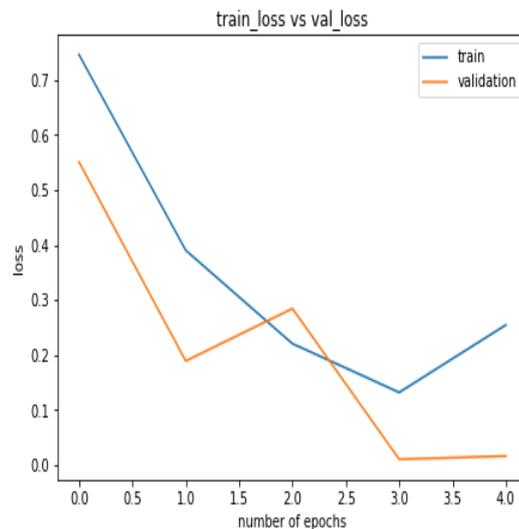
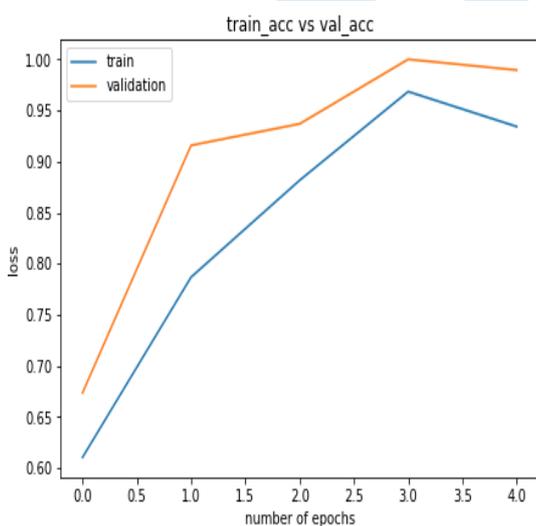
<i>Parameter</i>	<i>Resnet-CPU</i>	<i>Resnet-GPU</i>
Number of Convolutional layers	40	40
Number of epochs	5	5
Batch size	16	32
Training Time (Approx)	0:48:26	0:24:58
Training Accuracy	98.07%	96.96%
Validation accuracy	99.45%	99.42%
Testing accuracy	99.56%	99.07%
Trainable parameters	5,296,898	5,296,898

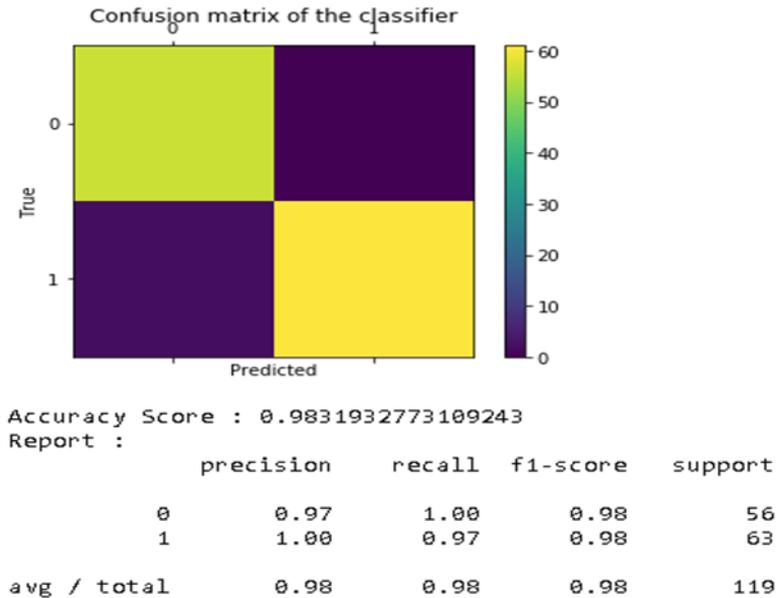
Table 1.

Table 1 shows the comparison between Resnet performance on GPU and CPU. From the observations obtained we observed that Resnet on GPU get trained in nearly half the time and still gives fair accuracy.

<p><b>Original Image</b></p>		<p><b>Original image size is varied for different patients</b></p>
<p><b>After resizing</b></p>		<p><b>Due to difference in sizes of scans for different patients we resized images into 448*448.</b></p>
<p><b>Classification Results</b></p>	<pre>Enter path of image which need to classify and segment=C:\Users\sqwqz\OneDrive\Documents\breast1.jpg C:\Users\sqwqz\OneDrive\Documents\breast1.jpg (0,0) breast-CX-HE-BCD x.shape= (3, 448, 448, 3) t1[ ]-----] is 2x step result= [[0.0000000 0.0000000]] prediction = benign result= benign</pre>	<p><b>This shows the results of classification. If image found to be malignant our model will pass this scan for segmentation otherwise terminate the code showing scan is benign</b></p>

<p><b>Segmentation results</b></p>		<p>After performing segmentation, we got this result but it contains false positives which are resolved by dilation and erosion</p>
<p><b>Results after dilation and erosion</b></p>		<p>Tumors are found through dilation and erosion</p>
<p><b>Measurement of tumors</b></p>		<p>Tumor measurements are found in this section and saved to text file in output folder</p>
<p><b>Length and width of tumor saved to output folder</b></p>	<pre> Patient's Name= QIN-Breast-DCE-MRI-BC01 Total number of tumor found =4 length of tumor 1=37.36919896452157 mm width of tumor 1=35.71643966718967 mm Total number of tumor found =4 length of tumor 2=8.566148857567208 mm width of tumor 2=7.735116757360551 mm Total number of tumor found =4 length of tumor 3=9.078003979399076 mm width of tumor 3=8.566148857567208 mm Total number of tumor found =4 length of tumor 4=6.0104076400856545 mm width of tumor 4=6.0104076400856545 mm                     </pre>	<p>In output folder we saves the segmented image, marked length and width on segmented image and details regarding tumors in text file.</p>



**Confusion Matrix:**

Sr.no	Virtual Environment Details
1	Tensorflow 1.9.0
2	Keras 2.2.4
3	Python 3.6.4
4	Anaconda3 5.0.0
5	Opencv-python
6	pydicom
6	PIL
7	Numpy 1.16.0
8	Scikit-Image 0.15.0

**Conclusion:**

The main objective of this work is to develop a methodology for Detection of Breast cancer and segmentation of abnormalities using MRI scans. If Breast cancer is detected as Malignant, region in MRI scan containing abnormality/ tumor is segmented using image processing and sized (length and width). For Cancer detection Resnet deep neural network is used On CPU system, it provide 98.07% accuracy in approx. 48 min whereas on GPU system it give 96.96% accuracy in approx. 24 min. Testing accuracy of Resnet on CPU and GPU system provide 99.56% and 99.07% accuracy respectively. Resnet gets trained in about 5 epochs. Image segmentation is done using region growing algorithm, and length and width of tumor is determined.

**VI. ACKNOWLEDGMENT**

The authors are indebted to C-DAC for providing the GPU system for the project.

**REFERENCES**

- [1] Jalalian, A., Mashohor, S., Mahmud, R., Karasfi, B., Saripan, M.I.B. and Ramli, A.R.B., 2017. Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection. EXCLI journal, 16, p.113.
- [2] Rasti, R., Teshnehlab, M. and Phung, S.L., 2017. Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks. Pattern Recognition, 72, pp.381-390
- [3] Al-masni, M.A., Al-antari, M.A., Park, J.M., Gi, G., Kim, T.Y., Rivera, P., Valarezo, E., Han, S.M. and Kim, T.S., 2017, July. Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network. In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)(pp. 1230-1233). IEEE.
- [4] Guan, S. and Loew, M., 2017, October. Breast Cancer Detection Using Transfer Learning in Convolutional Neural Networks. In 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) (pp. 1-8). IEEE.
- [5] Garcia, E., Hermoza, R., Castanon, C.B., Cano, L., Castillo, M. and Castañeda, C., 2017, June. Automatic Lymphocyte Detection on Gastric Cancer IHC Images Using Deep Learning. In Computer-Based Medical Systems (CBMS), 2017 IEEE 30th International Symposium on (pp. 200-204).

- [6] Goceri, N. and Goceri, E., 2015, December. A neural network based kidney segmentation from MR images. In Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference on (pp. 1195-1198). IEEE.
- [7] Yin, D. and Lu, R.W., 2015, November. A Method of Breast Tumour MRI Segmentation and 3D Reconstruction. In Information Technology in Medicine and Education (ITME), 2015 7th International Conference on (pp. 23-26). IEEE.
- [8] Tan, Y., Liu, L., Liu, Q., Wang, J., Ma, X. and Ni, H., 2011, October. Automatic breast DCE-MRI segmentation using compound morphological operations. In Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on (Vol. 1, pp. 147-150). IE

