EMBEDDED SYSTEM DESIGN FOR BIOMETRIC IMAGE PROCESSING

[DETECTION OF BREAST CANCER]

REPORT OF PROJECT SUBMITTED FOR PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

SUBMITTED BY

NAME	REG. NO.	ROLL NO.
Abhinaba Habishyashi	111150110002	11500111002
Swapnadip Chakraborty	111150110112	11500111112
Tuhin Kundu	111150110115	11500111115

UNDER THE SUPERVISION OF

Mr. Bikromadittya Mondal
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,
B. P. PODDAR INSTITUTE OF MANAGEMENT AND TECHNOLOGY

AT

B. P. PODDAR INSTITUTE OF MANAGEMENT AND TECHNOLOGY
[Affiliated to West Bengal University of Technology]
137, V.I.P. ROAD, PODDAR VIHAR, KOLKATA – 700052

<u>Certificate</u>

The report of the Project titled : <u>EMBEDDED SYSTEM DESIGN FOR BIOMETRIC IMAGE PROCESSING</u> submitted by Abhinaba Habishyashi (Roll No.: 11500111002), Swapnadip Chakraborty (Roll No.: 11500111112) and Tuhin Kundu (Roll No.: 11500111115) has been prepared under my supervision for the partial fulfillment of the requirements for B Tech degree in Computer Science & Engineering under the West Bengal University of Technology

The report is hereby forwarded.

…………………………………….

Mr. Bikromadittya Mondal

[Designation], Dept. of Computer Science & Engineering

B. P. PODDAR INSTITUTE OF MANAGEMENT & TECHNOLOGY

Countersigned by

……………………………………

Dr. Ananya Kanjilal

Head, Department of CSE

B. P. PODDAR INSTITUTE OF MANAGEMENT & TECHNOLOGY

Acknowledgement

I express my sincere gratitude to Mr. Bikromadittya Mondal of Department of Computer Science and Engineering, BPPIMT and for extending their valuable times for me to take up this problem as a Project.

I am also indebted to parents, friends and mentors for their unconditional help and inspiration.

Last but not the least I would like to express my gratitude to Dr. Ananya Kanjilal, Mr. Suprabhat Maity and Mr. Suvadeep Bhattacharjee of our department who helped me in their own way whenever needed.

Date: 19/05/2015

……………………………………
Abhinaba Habishyashi
Reg. No.: **111150110002**
Roll No.: **11500111002**

……………………………………
Swapnadip Chakraborty
Reg. No.: **111150110112**
Roll No.: **11500111112**

……………………………………
Tuhin Kundu
Reg. No.: **111150110115**
Roll No.: 11500111115

Abstract

Biometrics is the science of measuring and statistically analyzing biometric data. Breast cancer is one of the most infamously famous type of cancer present and the most prevalent these days. The presence of micro calcification clusters in mammograms contributes evidence for the detection of early stages of cancer. Detection requires a series of tests and lot of visits to the doctor. The objective of this project is to create an embedded system which uses a low-cost and high-speed neural network based breast cancer detection algorithm that helps to detect the cancer.

Table of contents

Topics **Page No.**

1. Introduction

1.1 Breast cancer is the development of cancer from breast tissue.
Worldwide, breast cancer is the most common invasive cancer in women. The incidence of breast cancer varies greatly around the world: it is lowest in less-developed countries and greatest in the more-developed countries.

1.2 Mammography is the process of using low-energy X-rays to examine the human breast, which is used as a diagnostic and screening tool. The goal of mammography is the early detection of breast cancer.

In information technology, biometrics refers to technologies that measure and analyze human body characteristics through DNA, fingerprints, mammograms, eye retinas, voice patterns, etc. for authentication and diagnostic purposes.

So, the objective is to build a low cost embedded system that would combine these three ideas and then by the application of an algorithm on the image of the mammogram, helps in the detection of breast cancer.

1.3 Now what is an embedded system?

It's a computer that has been built to solve only a few very specific problems and is not easily changed. It has a processor and software, input and output.

Embedded systems generally use a Real Time Operating System, or RTOS.
The program instructions written for embedded systems are referred to as firmware, and are stored in read-only memory or flash memory chips.

2. Review of Literature

The system of using machines, typically artificial intelligence for the study of and analyzing biometric images has been a field of interest among various researchers for a long time, but not without incurring a huge cost to the investors. So the recent general trend has been to search or to make a low cost effective solution which is within the grasp of common people. This trend has led researchers all over the world to generate various algorithms which work but are not entirely suitable when it comes to the low cost part. And so the search is on to create that near perfect embedded system design which fulfills the dream of being able to analyze a biometric system without incurring a heavy cost.

Many works have been formulated to develop a segmentation algorithm and diagnosis tools for detecting breast tumour and classify them whether it as benign and malignant. Samual H. Lewis and Aijuan Dong developed a Marker-Controlled Watershed segmentation algorithm to locate breast mass tumour candidates. In this method first selected foreground and background markers from the mammogram images and then applied Watershed segmentation algorithm to isolate a tumour region from its surrounding tissue. The Marker-Controlled Watershed segmentation was fairly successful in locating tumour under all conditions. Watershed segmentation is used in various image processing and computer vision tasks. But the major drawback is it produces false positive results. Arianna Mencattini, Marcello Salmaeri and Simona Salicone describes a CAD (Computer-Aided Detection) system and diagnosis (CADx) systems that are widely used in mammography. Both systems are involve the use of computer algorithms to detect patterns in images associated with signs of the disease and to assign them a malignancy index. The result should attract the clinicians' attention to potentially abnormal regions in mammograms. Jawad Nagi, Sameem Kareem and Farrukh Nagi proposed a Breast profile segmentation method .Breast profile segmentation is an automated technique for mammogram segmentation. It uses morphological preprocessing and Seeded Region Growing (SRG) to remove digitization noises, suppress radiopaque artefacts and remove the pectoral muscle to accentuate the breast profile region for use in CAD algorithms. SRG performs segmentation of an image with respect to a set of points, known as seeds. The algorithm has been tested using mammogram images of differing densities from multiple databases and has shown results with accuracy. Kekre.H.B, Saylee M. Gharge and Tanuja K. Sarode proposed an algorithm. The algorithm uses probability of mammographic image as input for vector quantization. Kekre's Proportionate Error (KPE) algorithm is used for region forming, and codebook of size 128 is created. Further the 128 clusters were utilized for region merging using KPE algorithm for reclustering. The tumour sectional area is calculated and centre point is compared with Linde-Buzo-Gray (LBG) algorithm for segmentation of mammographic images. The probability of original image is used for grouping pixels into regions and then the image of probability is formed. For image segmentation Equalized probability image is used as an input image for further segmentation. Roshan Dharshana, Yapa and Koichi Harada proposed Breast skin-line estimation and breast segmentation techniques. An image would increase the accuracy and efficiency of processing algorithms. Leonardo de Oliveira Martins and Geraldo Braz Junior described a Clustering algorithms that can be applied to solve the segmentation problem. It consists in choosing an initial pixel or region that belongs to one object of interest, followed by an interactive process of neighbourhood analysis, deciding if whether each neighbouring pixel belongs or not to the same object. K-means algorithm is used to resolve the mass detection task on mammograms using texture information obtained from Haralick's descriptors. The K-means algorithm is one of the simplest non-supervised learning algorithms classes that solve the clustering segmentation problem. The method follows the usual steps to satisfy the primary objective that is clustering all the image objects into K distinct groups. First, K centroids are defined, one for each group. Their initial position is very important to the result. After that, it is determined a property region for each centroid, which groups a set of similar objects. Maanasa N A S, V Gowri proposed a tumor cut segmentation method and also using Support Vector Machine Classifier to classify whether the segmented image as benign and malignant.

3. Preliminaries

The basic Embedded System design consists mainly of basic three parts.
The first part being the external input. The mammograms (images of the breasts) are scanned by the external connected scanner and relative information regarding age; breast density rating, subtlety record etc. are also entered. The scanner related information like spatial resolution etc. are pre-defined in certain subroutines that are invoked by the main algorithm. These images are then converted into greyscale images and the subsequent information is entered into the algorithm.
The second part is the processing unit. This unit is programmable and will have the algorithm that will be used for detection purposes. The processing unit in this case is a microcomputer/microcontroller. The architecture is built to support complex mathematical algorithms that involve a significant amount of multiplication and addition. It executes the multiply/add feature in a single cycle (compared to multiple cycles for RISC processors) with the help of the multiply/accumulate (MAC) hardware inside the arithmetic logic unit (ALU).
Developers of biometrics systems can take advantage of this architecture to enhance the resolution of the captured image with the use of two-dimensional fast Fourier transforms (FFT) and finite IR filters. Because the accuracy of a system is as much dependent on the input image as it is on the processing algorithm, this helps in improving the overall accuracy and error rate of the biometrics system - a key performance metric. The last part is the output device. With the advancement of technology any type of output devices can be used.
Finally, there is the output interface element, which will communicate the decision of the biometric system to the interfaced asset to enable access to the user. This can be a simple serial communication protocol like RS232, or the higher bandwidth USB protocol. It could also be the TCP/IP protocol via a wired medium like 10/100 Ethernet or through a wireless medium using either the

802.11b protocol, ISM RF band, RFID, Bluetooth, or one of the many cellular protocols.

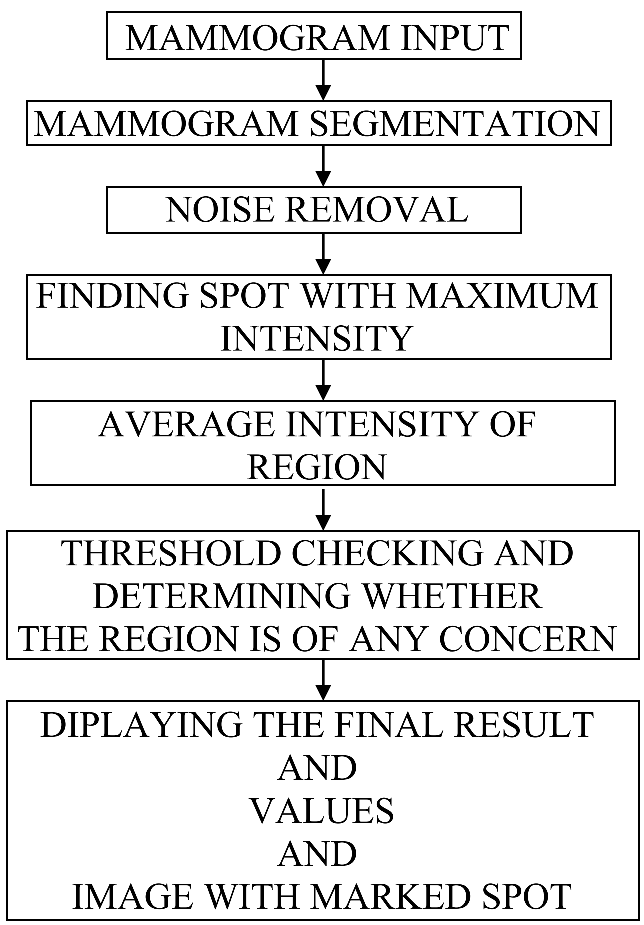
We use Raspberry Pi B+ for our required objective.

3.1 The Model B+ is the higher-spec variant of the Raspberry Pi. It replaced the original Model B in July 2014. Compared to the Model B it has:

- More GPIO. The GPIO header has grown to 40 pins, while retaining the same pin-out for the first 26 pins as the Model B.
- More USB. We now have 4 USB 2.0 ports, compared to 2 on the Model B, and better hot plug and overcurrent behaviour.
- Micro SD. The old friction-fit SD card socket has been replaced with a much nicer push-push micro SD version.
- Lower power consumption. By replacing linear regulators with switching ones we've reduced power consumption by between 0.5W and 1W.
- Better audio. The audio circuit incorporates a dedicated low-noise power supply.
- Neater form factor. We've aligned the USB connectors with the board edge, moved composite video onto the 3.5mm jack, and added four squarely-placed mounting holes.

4. Proposed Work

Digital Mammograms are medical images that are difficult to interpret, thus a preparation phase is needed in order to improve the image quality and make the segmentation results more accurate. Our objective during this process is to improve the quality of the image to make it ready for further processing by removing the irrelevant and unwanted parts in the background of the mammogram. After obtaining the processed mammogram image, we propose the following steps:



4.1 MAMMOGRAM INPUT :

The mammograms used during the course of developing this project was supplied in the form of images of JPEG format from hospital. Each mammogram has the resolution 600X1066 pixels. The image sections are lateral and are either the left or the right breast.

The mammograms are widely classified into each of these three types based on the breast tissue :

- ➔ **Fatty –** In this case the amount of fat in the breast region is very high. The fat appears black on the mammogram. So, the mammograms with high black intensity are grouped under this category. Suspicious spots are easiest to detect in this type of mammograms.

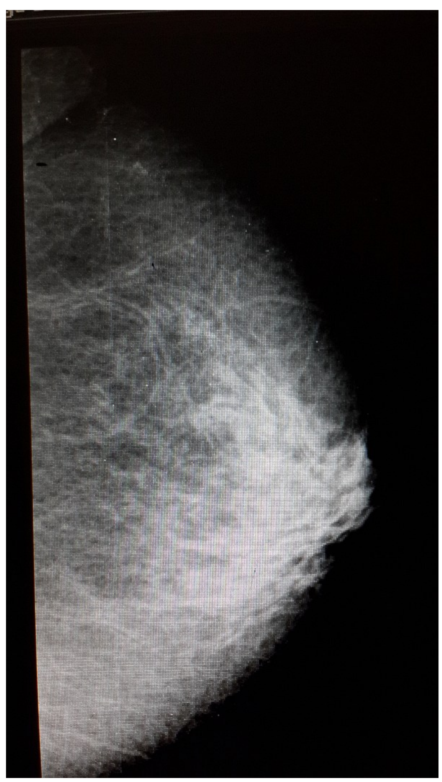


Fig : 1- Fatty mammogram sample

➔ **<u>Fatty Glandular –</u>** In this case there are both fat and blood vessels visible and dense tissue in the mammogram. The dense tissue region appear white on the mammogram.

Fig : 2 – Fatty Glandular mammogram sample

➔ **<u>Dense Glandular –</u>** In this case fat is almost absent and only the tissue part is present or rather visible in the mammogram. The detection of a suspicious region is visible in these type of mammogram is extremely difficult.

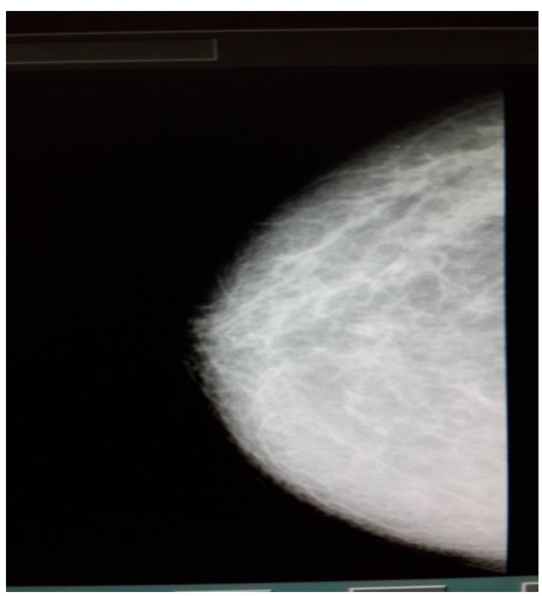


Fig : 3 – Dense Glandular mammogram sample

This project uses Python integrated with OpenCV for writing the algorithm in the embedded system and thus functions from these two modules have been used in the project.

4.2 MAMMOGRAM SEGMENTATION :

This refers to the grayscale conversion of the image so as for further reading. The mammograms appear to be black and white on viewing but still they have to be converted to grayscale for better experimentation and more accurate results.

In OpenCV the feature of the function **imread** has been used for this purpose.

This function takes as input the location of the image and also the parameter as to how to read it. The image can be read as Gray Scale image or in the same condition as it appears on the disk.

Python: `cv2.imread(filename[, flags])`

Parameters:

→ filename – Name of file to be loaded.
→ flags –

Flags specifying the color type of a loaded image:

CV_LOAD_IMAGE_ANYDEPTH - If set, return 16-bit/32-bit image when the input has the corresponding depth, otherwise convert it to 8-bit.

CV_LOAD_IMAGE_COLOR - If set, always convert image to the color one

CV_LOAD_IMAGE_GRAYSCALE - If set, always convert image to the grayscale one

>0 Return a 3-channel color image.
=0 Return a grayscale image.
<0 Return the loaded image as is (with alpha channel).

So If we include a 0 in the imread function along with the image is imported and converted to its Grey Scale version.

4.3 NOISE REMOVAL :

In this section the noise from the image is removed using Gaussian Blur function of OpenCV.

Noise removal is an essential part of the image transformation in order to get the correct results as required. Mammogram reading and detection of suspicious masses is a process that must be done with caution and thus it must be done carefully and thus noise removal is an essential portion of the process. Noise refers to fluctuations in the images that are wanted and may give false readings of those intensities.

Python: `cv2.`GaussianBlur(src, ksize, sigmaX[, dst[, sigmaY[, borderType]]]) → dst

Parameters:

src – input image; the image can have any number of channels, which are processed independently, but the depth should be CV_8U, CV_16U, CV_16S, CV_32F or CV_64F.

dst – output image of the same size and type as `src`.

ksize – Gaussian kernel size. ksize.width and ksize.height can differ but they both must be positive and odd. Or, they can be zeroes and then they are computed from `sigma`

sigmaX – Gaussian kernel standard deviation in X direction.

sigmaY – Gaussian kernel standard deviation in Y direction; if sigmaY is zero, it is set to be equal to sigmaX, if both sigmas are zeros, they are computed from ksize.width and ksize.height ,respectively to fully control the result regardless of possible future modifications of all this semantics, it is recommended to specify all of ksize, sigmaX, and sigmaY.

borderType – pixel extrapolation method .

This removes the noise from the images. The ksize taken in the procedure is 9x9 matrix as it gives the best result for the given resolution of the images.

4.4 FINDING SPOT WITH MAXIMUM INTENSITY :

This is the central part of the algorithm and this is the part which leads to the detection of the suspicious region if present within the mammogram.

Cancer starts with the growth of abnormal tissue which means that in certain part of the body where cancer occurs there is an abnormal growth of tissues which is visible physically.

How does this affect the appearance of a mammogram?

In a mammogram the tissues appear as white. The region which has the most growth of tissue will also appear white to the naked eye. But in case of an image it will have a higher intensity than its neighbouring regions. Thus one can conclude that the region with the highest intensity will be the suspicious region on the mammogram and thus should be examined.

In order to find the region with the highest intensity the function minMaxLox is used.

Python: `cv2.minMaxLoc`(src[, mask]) → minVal, maxVal, minLoc, maxLoc

Parameters:

src – input single-channel array.

minVal – pointer to the returned minimum value; `NULL` is used if not required.

maxVal – pointer to the returned maximum value; `NULL` is used if not required.

minLoc – pointer to the returned minimum location (in 2D case); `NULL` is used if not required.

maxLoc – pointer to the returned maximum location (in 2D case); `NULL` is used if not required.

mask – optional mask used to select a sub-array.

The functions `minMaxLoc` find the minimum and maximum element values and their positions. The extremums are searched across the whole array or, if `mask` is not an empty array, in the specified array region.

Thus, we get the location and the intensity of the maximum intensity location in the mammogram image.

Now the analysis of the image is required with respect to its neighbouring regions.

In order to do that we must crop the image. The cropped image is different for fatty, fatty glandular and dense glandular mammograms.

After cropping the image we have to find the average intensity of the region.

4.5 AVERAGE INTENSITY OF REGION :

To find the average intensity of the cropped region the mean function is used.

Python: `cv2.mean`(src[, mask]) → retval

Parameters:

src – input array that should have from 1 to 4 channels so that the result can be stored in `Scalar_` .

mask – optional operation mask.

The function `mean` calculates the mean value `M` of array elements, independently for each channel, and return it:

$$N = \sum_{I:\ \mathtt{mask}(I) \neq 0} 1$$
$$M_c = \left(\sum_{I:\ \mathtt{mask}(I) \neq 0} \mathtt{mtx}(I)_c \right) / N$$

4.6 THRESHOLD CHECKING AND DETERMINING WHETHER THE REGION IS OF ANY CONCERN :

After the mean has been calculated then it is necessary that we compare the value of the mean intensity with the threshold values and determine whether the region is indeed suspicious or not.

The threshold values for the three types of tissues are :

Fatty → Values within 180 and 195 are considered risky and must be forwarded for further testing and biopsy. Anything lower than 180 must be ignored and anything greater than 195 may be ignored as a calculation error.

Fatty Glandular → Values within 195 and 215 are considered risky and must be forwarded for further testing and biopsy. Anything lower than 195 must be ignored and anything greater than 215 may be ignored as a calculation error.

Dense Glandular → Values greater than 215 are considered risky and must be forwarded for further testing and biopsy. Anything lower than 215 must be ignored. This type of tissue is the most difficult to read and hence no upper bound is possible for it.

5. Experimental Results

```
enter tissue type:
f
enter the file name:
1
Max intensity pixel value : 238.0
mean intensity:  187.6
```

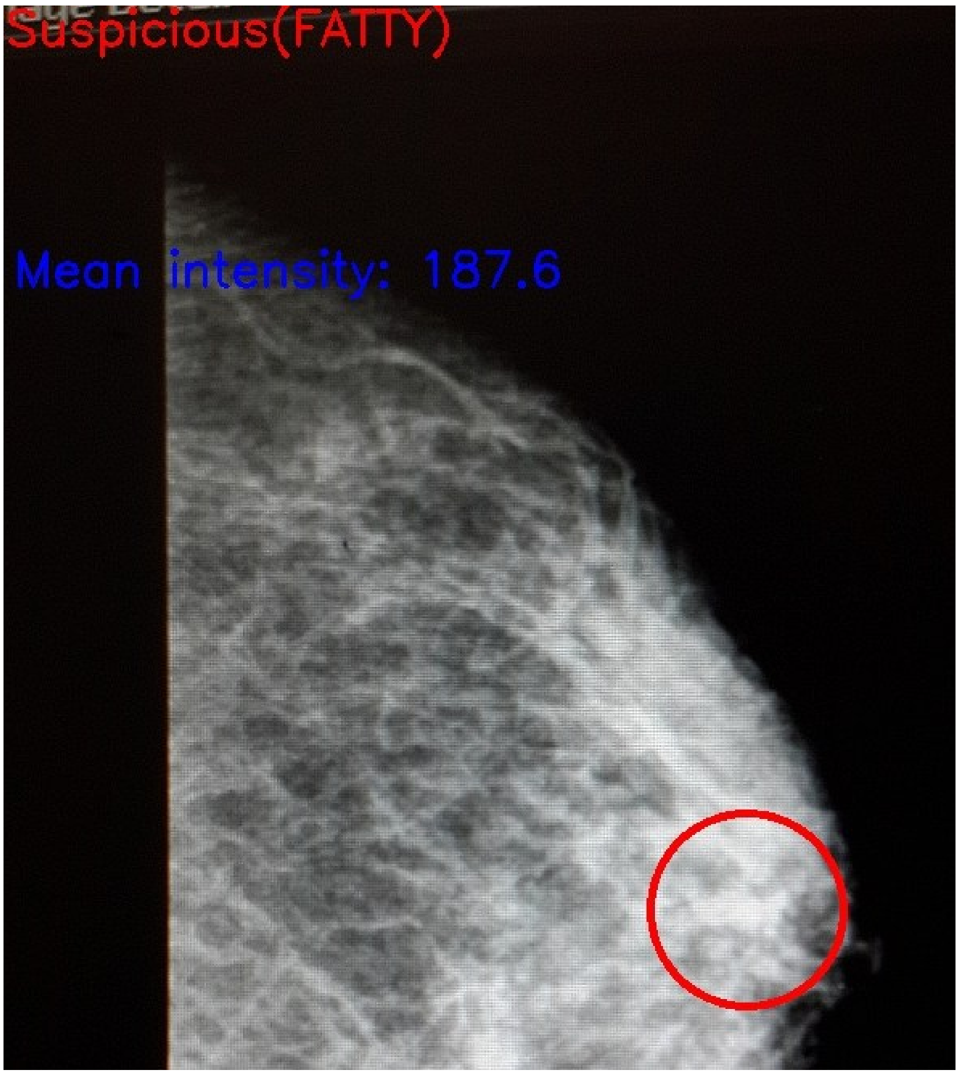


Fig : 4 Fatty mammogram result

```
enter tissue type:
fg
enter the file name:
2
Max intensity pixel value : 231.0
mean intensity:  187.6907
```

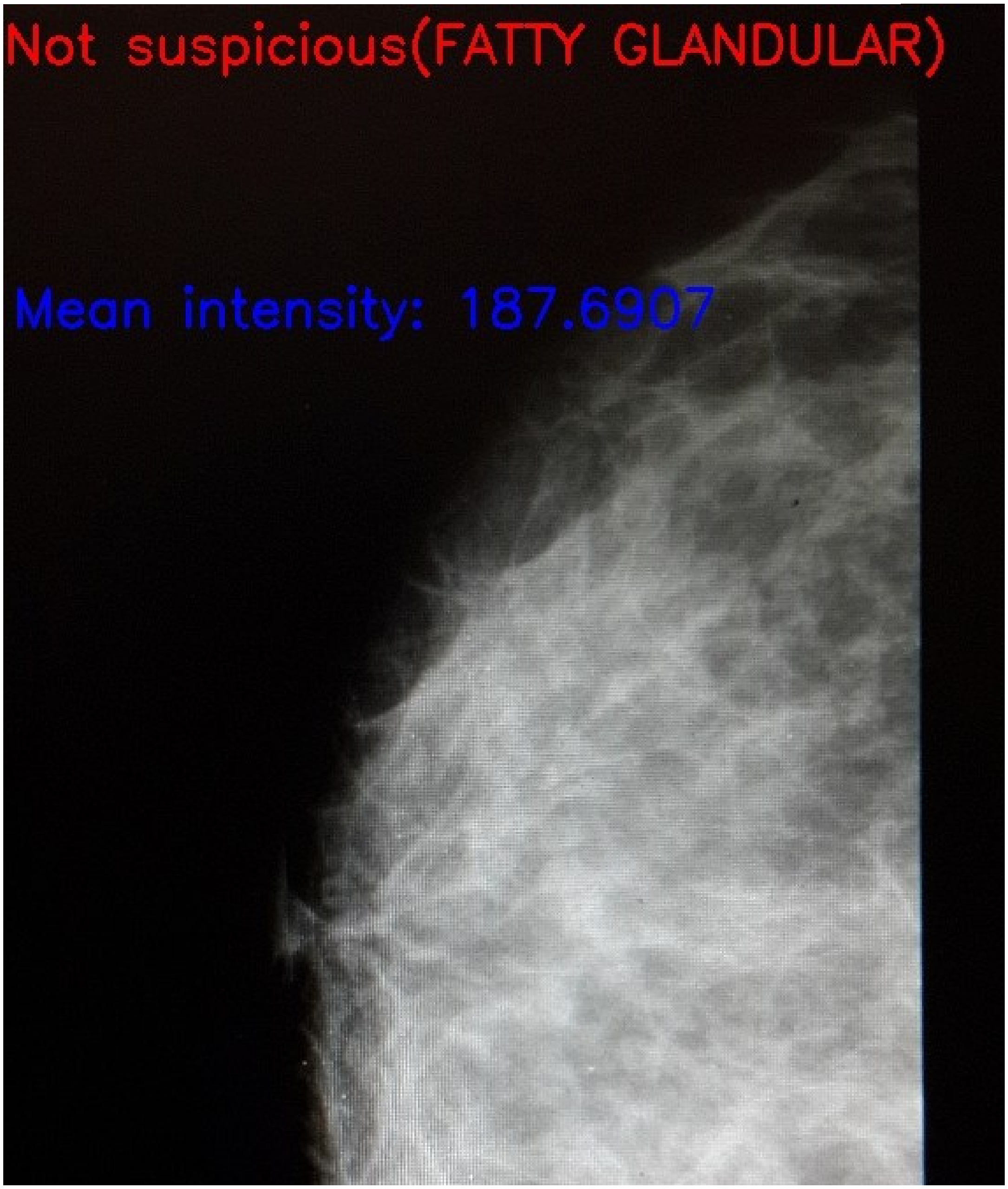


Fig : 5 – Fatty Glandular mammogram result

```
enter tissue type:
dg
enter the file name:
1
Max intensity pixel value : 255.0
mean intensity:  223.625246914
```

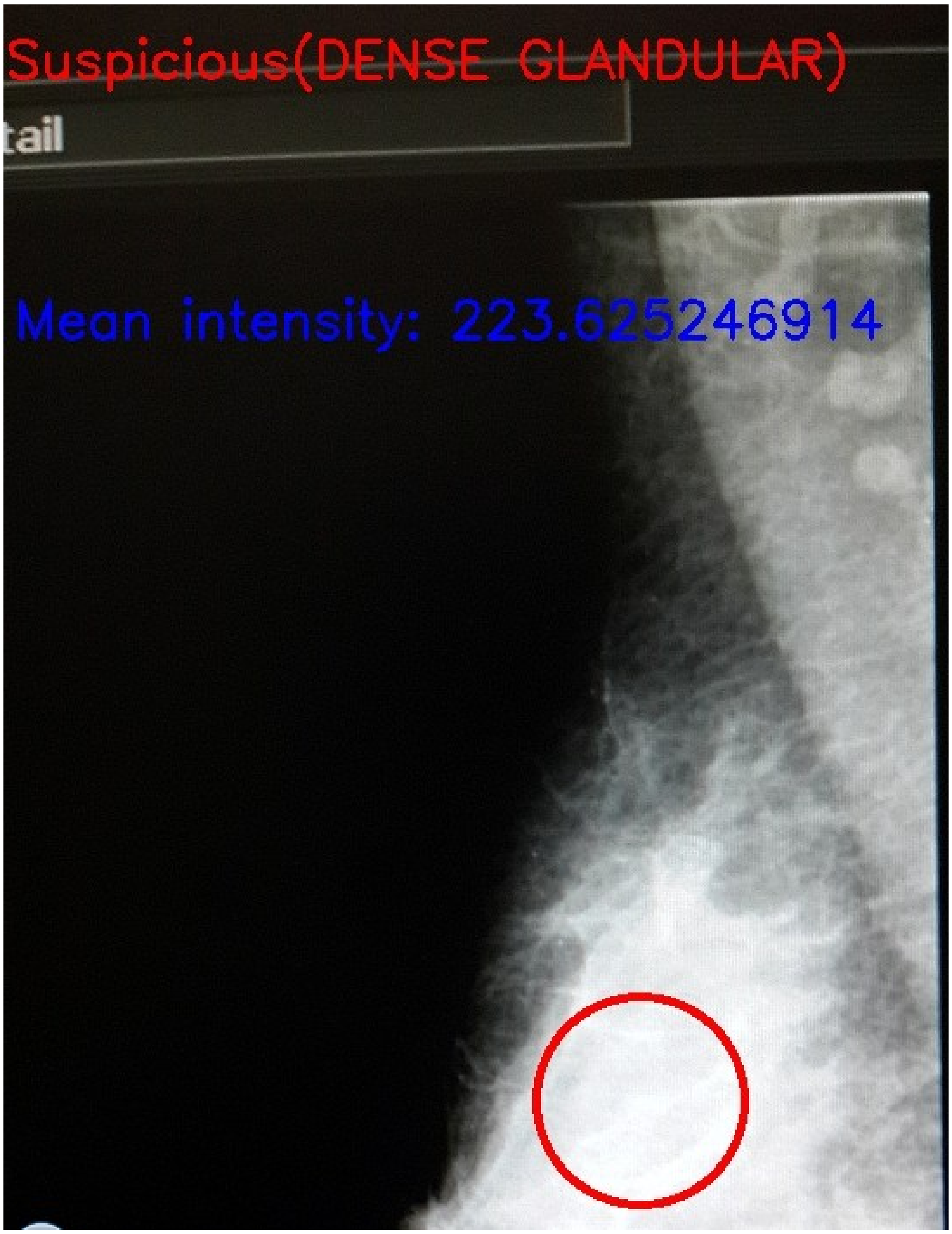


Fig : 6 – Dense Glandular mammogram result

Table 1a

FATTY

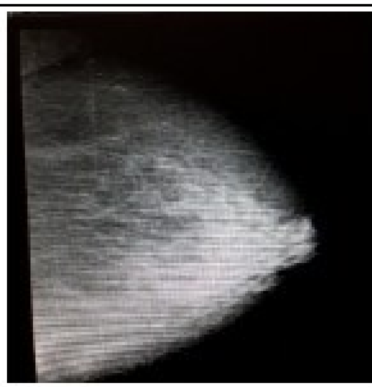
IMAGES	TYPE	MAXIMUM INTENSITY
	FATTY	238
	FATTY	223
	FATTY	221

Table : 1b

AVERAGE INTENSITY	**SUSPICIOUS?**
186.2694	YES
187.7683	YES
193.0251	YES

Table : 2a

FATTY GLANDULAR

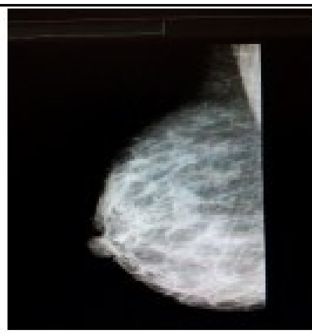
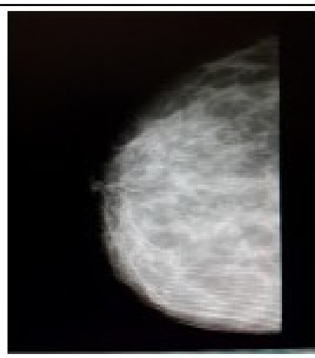
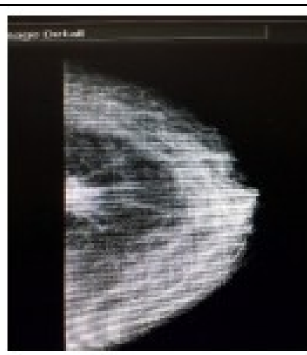
IMAGES	TYPE	MAXIMUM INTENSITY
	FATTY GLANDULAR	232
	FATTY GLANDULAR	231
	FATTY GLANDULAR	243

Table : 2b

AVERAGE INTENSITY	SUSPICIOUS?
193.7945	NO
186.266	NO
209.8271	YES

Table : 3a

FATTY GLANDULAR

IMAGES	TYPE	MAXIMUM INTENSITY
	FATTY GLANDULAR	219
	FATTY GLANDULAR	255
	FATTY GLANDULAR	251

Table : 3b

AVERAGE INTENSITY	SUSPICIOUS?
189.5467	NO
213.1477	YES
208.8557	YES

Table : 4a

FATTY GLANDULAR

IMAGES	TYPE	MAXIMUM INTENSITY
	FATTY GLANDULAR	239

Table : 4b

AVERAGE INTENSITY	SUSPICIOUS?
199.7579	YES

Table : 5a

DENSE GLANDULAR

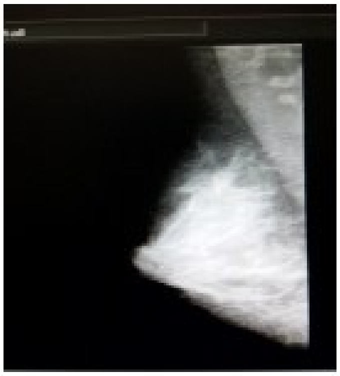
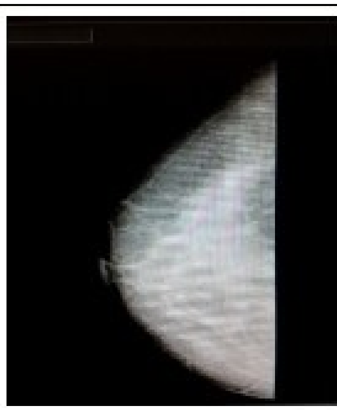
IMAGES	TYPE	MAXIMUM INTENSITY
	DENSE GLANDULAR	255
	DENSE GLANDULAR	223
	DENSE GLANDULAR	220

Table : 5b

AVERAGE INTENSITY	SUSPICIOUS?
221.9158	YES
196.4898	NO
197.2416	NO

Table: 6a

DENSE GLANDULAR

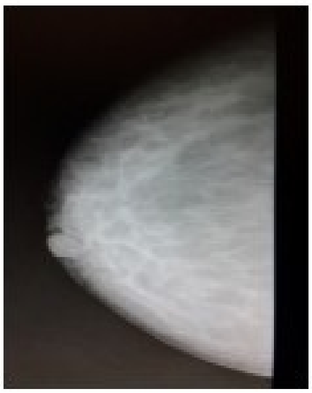
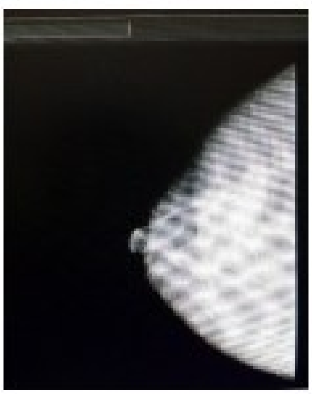
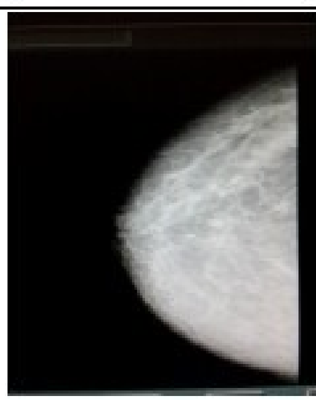
IMAGES	TYPE	MAXIMUM INTENSITY
	DENSE GLANDULAR	212
	DENSE GLANDULAR	254
	DENSE GLANDULAR	226

Table : 6b

AVERAGE INTENSITY	SUSPICIOUS?
191.7011	NO
227.5498	YES
189.9882	NO

6. Conclusion

Today's biometric systems are based mainly on interfacing the sensing element with a personal computer. The sensors are generally networked to a computer server to service unlimited users and multiple access points. The cost of using PCs is prohibitive and the communication link between the sensor and the PC/server could be a major cause for concern with regards to security and privacy. A biometrics solution based on DSPs can function both as a secure standalone device for recognition (1:1 or 1: few) and as a trusted network device for identification (1: many). A secure standalone device is one where all the functions of authentication are carried out within the confines of the embedded processor and the result is communicated or displayed along with control signals to deny or grant access to the secured asset. The original enrolled template or pattern is either stored in the memory within the product or on a smart card which is carried on the user's person.

Biometrics is a truly emerging market with great potential for success. Its roots may be in science fiction, but it is part of today's science and technology fact. In the near future, we will come to rely on biometric technology to protect our property, assets, and the people we love. We will see this technology become a secure and trusted form of authentication with uses varying from controlling access to personal information devices, to securing buildings and enabling e-Commerce. In conclusion, using an embedded processor of choice for enabling smart biometric systems can provide the following advantages:

- Fast, accurate, secure and trusted authentication

- Enable new applications with one scalable design

- Reduce overall cost of development

References

- S.L. Lou, H.D. Lin, K.P. Lin, and D. Hoogstrate, "Automatic Breast Region Extraction from Digital Mammograms for PACS and Telemammography applications", Computerized Medical Imaging and Graphics, 24:205-220. 2000.

- Gonzalez, R.C., Woods, R.E. (1992), "Digital image processing".

- M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models", International Journal of Computer Vision, 1:321-331. 1998.

- Highnam R and Brady M, "Mammographic Image Analysis", Kluwer Academic Publishers, 1999. ISBN: 0-7923- 5620-9.

- National Cancer Institute (NCI) Web site, http://www.cancernet.gov

- http://www.opencv-python-tutroals.readthedocs.org

Appendix 1 (list of figures)

Appendix 2 (list of tables)