MAMMSIT: A Database For The diagnosis and detection of Breast Cancer in Mammography images

M.Parisa Beham
Department of ECE
Sethu Institute of
Technology
kariapatti,India
parisaphd2011@gmail.
com

N.Kayalvizhi
Department of ECE
Sethu Institute of
Technology
kariapatti,India
kayalvizhi@sethu.ac.in

R.Tamilselvi
Department of ECE
Sethu Institute of
Technology
kariapatti,India
tamilselvi@sethu.ac.in

A.Nagaraj
Department of ECE
Sethu Institute of
Technology
kariapatti,India
nagaraj.sa@gmail.com

Abstract

In recent days Mammography is most widely used techniques to detect the Breast cancer. Masses and calcifications are the abnormalities indicate the breast cancer. . Breast cancer is the leading cause of cancer deaths in women. Advanced technologies with existing imaging modalities improve breast cancer screening. Most of the Mammographic image data sets are not easily available. The Mammographic Image Analysis Society (MIAS) and Digital Database for Screening Mammography (DDSM) are the most easily available data set for Mammographic images. MAMMSIT is another data set for the mammographic image analysis. The mammogram data set is named as MAMMSIT and it contains Normal and Malignant group images of mammography. Our proposed data set provided with appropriate annotations details such as age, background tissue, ultra sound screen position (i.e., left and right) and BIRADS level are included in the database.

Keywords: Mammogram, Database, Malignant, Mammographic images.

I. INTRODUCTION

In the medical image processing filed, specialized methods such as Mammography is used to visualize the low X-ray system of breasts. Breast cancer is the leading cause of cancer deaths in women. In Recent medical technology field Mammograms takes an very important role and also supports for early detection and diagnosis of breast cancer. Mammography is adapted for screening or for diagnostic purposes. Screening mammography is one of methods used to detect breast changes in women who have no signs or symptoms or new breast abnormalities. At the age of 40, every women must undergo mammogram test atleast once in a year [2]. Image compression method of mammogram imaging, breasts is compressed between two firm surfaces to spread out the breast tissue. Reviewed by a doctor who looks for symptoms of cancer using an X-ray captures black-and-white images of breasts that are displayed on a computer screen. The objective is to detect cancer before clinical signs are markable. In diagnostic mammography method is analyzing suspicious breast changes such as breast pain, an unusual skin appearance, nipple thickening or nipple discharge. It's also used to find out abnormal images on a screening mammogram. In additional mammogram images are appending to the diagnostic mammogram. Breast cancer detection and diagnosis is currently questioning method of digital mammography. Separation of image acquisition and image display method using the development of imaging modality and it's allowing examination of both.

A mammogram of white area occurs on lump or tumor. Tumors can be cancerous or benign. If a tumor is benign, it is not a health risk and is unlikely to grow or change shape. Most tumors found in the breasts are non-cancerous. Small white specks are usually harmless. Sometimes occurring symptoms of cancer; the radiologist will check their finding areas of white, high density tissue, shape and pattern. A mammogram can also give person information about character of back ground tissue such as Fatty, Fatty - glandular, dense -glandular. The higher risk of breast cancer has a dense breast. Abnormalities of classes in dense breast on mammogram finding are very difficult. If a person has had breast cancer, Mammograms are still possible surgery or implants. However, it may be necessary to take more images of each breast, and it might take longer to check the images [1].

In imaging science, image processing is any form of signal processing for which an image as an input, such as a photograph or video and the output of image processing may be either an image or a set of characteristics or parameters related to the image [3]. In medical image processing system mammographic images are processed and it's used to convert an image signal, either digital or analog, converted into a physical image. As a result, very suitable differences between abnormal and normal but dense tissue can be made more obvious.

There are few Mammogram databases are publically available such as Digital Database for Screening Mammography (DDSM) ,DDSM data base contains 10480 image files in LJPEG format ,more than available all other databases[5][8] . LJPEG format is a non-standard version of the free format DDSM and finding the regions of interest of each image. Mammographic Image Analysis Society Digital Mammogram Database (MIAS) database consists of low quality 8 bit 320 mammogram images and compare among those databases required a greater

variety of images with different spatial and contrast resolutions. Computer Assisted Library for Mammography (CALMa)[6] and LLNL/UCSF database [7] consists of 3000 and 198 respectively.

Inspired by the above issues, in this work we proposed to build a database of mammography scan images which we named it as 'MAMMSIT Database', for the benefit of biomedical engineering research community. The important pitfall in the research and development of this community is that, unavailability of clinical interpretation. In the creation of database our key contributions in this work are:

- Create a new MAMMSIT database, involves 398 (normal and malignant) mammogram scan images.
- Present the annotation for the entire 398 subject's biological data.
- Provide clinical report as per WHO standard with suggestions of the health care provider.

II. MAMMSIT DATABASE

The MAMMSIT database is deliberated through subsequent stages:

1) Structure Details
2) Labeling the Images
3) Classifications

TABLE1. BIRADS SCORE

BIRADS (in Lesion)	Status
0	Incomplete
I	Negative
II	benign findings
III	probably benign
IV	suspicious abnormality
V	highly suspicious of malignancy
VI	known biopsy with proven malignancy

A. Structure Details

MAMMSIT dataset involves 398 mammogram scan images collected from various subjects. Each subject includes reference number, age, background tissue, abnormality classes and abnormality of severity in scan images. The samples of mammogram normal scan images of two subjects are shown in Figure. 1. Similarly, the samples mammogram and malignant images of 5 subjects are shown in Figure. 2. Table 1 shows the interpretation of malignant images.

B. Labeling the Images

In this proposed database of all the images are labelled perfectly for the ease understanding and interpretation for researchers. From the label itself one can easily identify the reference number, background tissue. For each subject, two samples are obtained viz., left screen image and right screen image. For example, labeling of left and right mammography scan image is given as MAMMSIT-L-001.jpg for a subject ID 001 and MAMMSIT-R-004.jpg for a subject ID 004 respectively. Here, in the first case MAMM refers to mammography scan image, 001L is the subject reference number of left screen image. In the next case 004R is the subject reference number of right screen image.

The list of annotations provided in the MAMMSIT database is as follows:
1. Database reference number
2. Age
3. Character of background tissue:
   F: Fatty
   G: Fatty-glandular
   D: Dense-glandular
4. Classes of abnormality present as
   CALC: Calcification
   CIRC : Well-defined/circumscribed masses
   SPIC : Speculated masses
   MISC : Other, ill-defined masses
   ARCH : Architectural distortion
   ASYM : Asymmetry
   NORM : Normal
5. Severity of abnormality
6. BIRADS (Breast Imaging Reporting and Data System) score

C. Classifications

The MAMMSIT dataset is classified into two types: Normal and Malignant. The classification performance was analyzed by using the complete set of abnormality presents report classified the following character presents in background tissue are calcification, well-defined circumscribed masses, speculated masses, ill-defined masses, architectural distortion, and asymmetry.
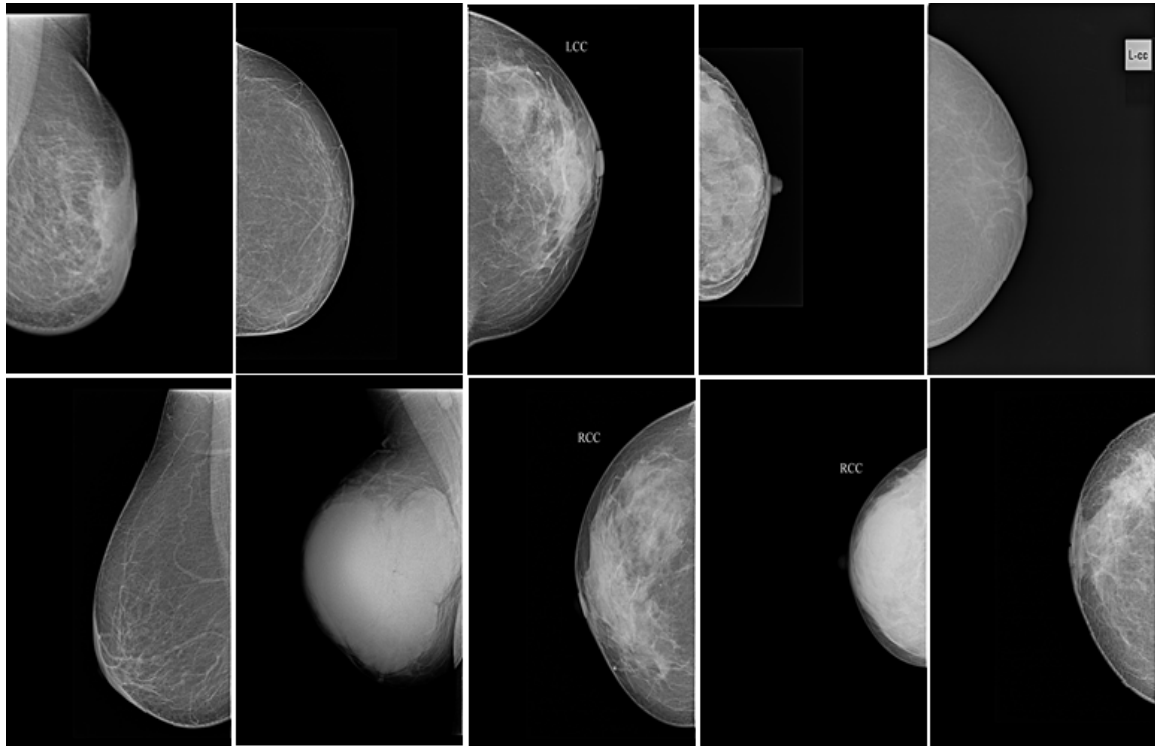
FIGURE.1. SAMPLES OF NORMAL AND ABNORMAL IMAGES

| MAMMSIT-L-001 | MAMMSIT-R-002 | MAMMSIT-L-003 | MAMMSIT-R-004 | MAMMSIT-R-005 |

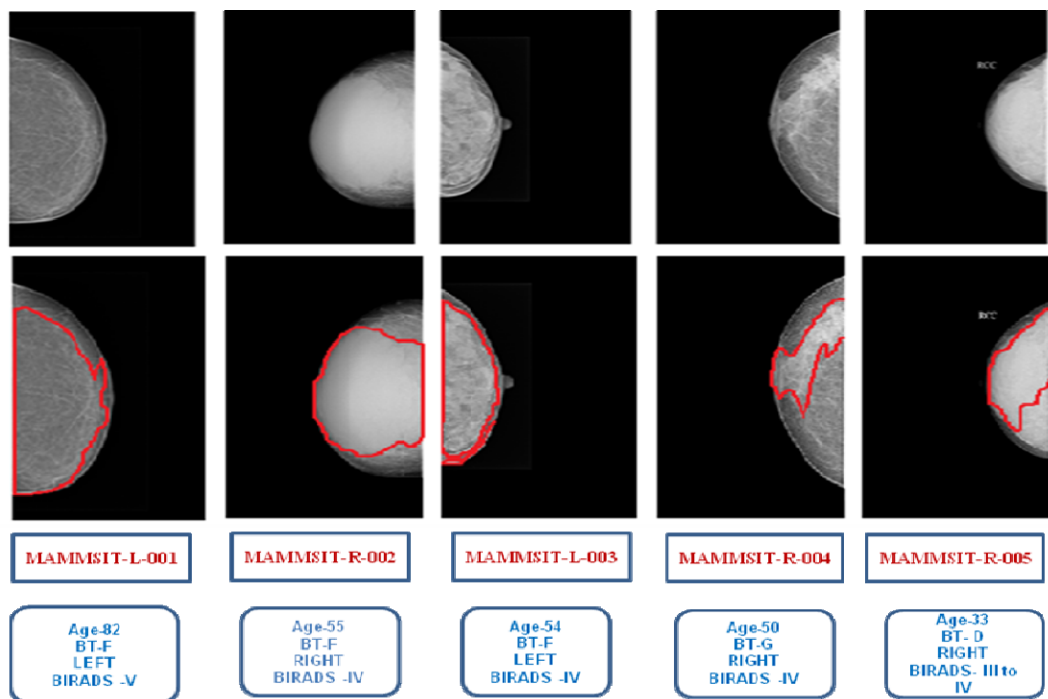| Age-82 BT-F LEFT BIRADS -V | Age-55 BT-F RIGHT BIRADS -IV | Age-54 BT-F LEFT BIRADS -IV | Age-50 BT-G RIGHT BIRADS -IV | Age-33 BT- D RIGHT BIRADS- III to IV |

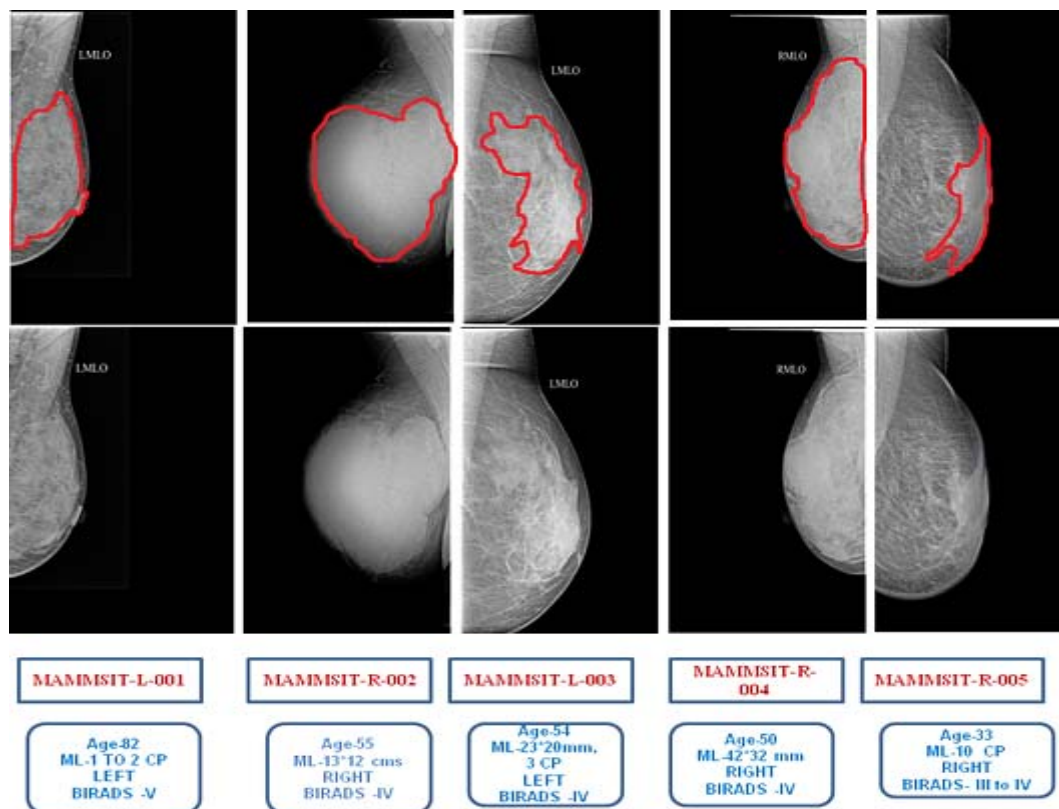FIGURE. 2. SAMPLES OF MALIGNANT IMAGES WITH GROUND TRUTH AND ANNOTATION

FIGURE. 3. SAMPLES OF MALIGNANT IMAGES WITH GROUND TRUTH AND ANNOTATION BASED ON CLOCK POSITION

TABLE.2. CLASSIFICATION DETAILS OF MALIGNANT IMAGES

Ref. No	BT	Classes of abnormality	AGE	Malignant	BIRADS
MAMMSIT-L-001	F	NORM	82	L	V
MAMMSIT-R-002	F	NORM	55	R	IV
MAMMSIT-L-003	F	ASYM	54	L	IV
MAMMSIT-R-004	D	CALC	50	R	IV
MAMMSIT-R-005	G	ARCH	33	R	III to IV

BIRADS classifications method is very useful in monitoring breast cancer treatment and supporting breast cancer research again by making statistics easier to calculate using BIRADS score as shown in Table 1. Figure 1 shows the sample normal and abnormal mammography images of 10 subjects. Figure 2 shows samples of malignant images with ground truth and annotation. Figure 3 depicts the samples of malignant images with ground truth and annotation based on clock position/mass lesions. Table 2 represents the examples of five subjects with the respective classification details.

III. CONCLUSION

The paper describes recent database for Screening Mammography. It acts as a resource of mammogram image analysis researchers. The database now contains substantial numbers of "normal" and "malignant" cases. The main characteristics of this MAMMSIT database are a) 398 mammogram images b) Marking the entire subject's biological data. By providing this database available to the biomedical research community, we hope to promote the analysis of many uncertain problems.

ETHICAL APPROVAL

The mammogram database used in this paper is provided by Pixel scans, Trichy. The ethical committee of Pixel scans has reviewed and approved to conduct research using this mammogram database and publish papers based on the results using those biomedical images.

REFERENCES

[1]Syed Jamal Safdar Gardezi Ahmed Elazab , "Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review", Journal of medical internet research , Vol.21,No 7,2019

[2]Giger "ML: Machine Learning in Medical Imaging." J Am Coll Radiol 15(3 Pt B),pp:512–520, 2018

[3]Karin Dembrower , PeterLindholm , Fredrik Strand "A Multi-million Mammography Image Dataset and Population-Based Screening Cohort for the Training and Evaluation of Deep Neural Networks the Cohort of Screen-Aged Women", Journal of Digital Imaging ,ISSN: 0897-1889

[4]Bruno Roberto ,Nepomuceno Matheus, and Homero Schiabel "Online Mammographic Images Database for Development and Comparison of CAD Schemes" Journal of Digital Imaging, Vol 24, No 3 (June), 2011: pp 500-506.

[5] Heath M, et al: The digital database for screening mammography. Proceedings of the Fifth International Workshop on Digital Mammography, 2001, pp 212–218

[6] Amendolia SR, et al: The CALMA project. Nucl Instrum Methods Phys Res, Sect A, Accel Spectrom Detect Assoc Equip 461(1-3):428–429, 2001

[7]Lawrence Livermore National Library/UCSF Digital Mammogram Database. Centre for Health Care Technologies Livermore. Livermore, CA, USA

8) Kallergi M: Computer-aided diagnosis of mammographic microcalcification clusters. Med Phys 31(2):314–326, 2004