Digital Image Processing For Breast Cancer Detection

Bhawna[1], Dr. N.H Simon[2]

[1]Research Scholar, Singhania University, Jhunjhunu, Rajasthan
[2]Research Supervisor, Singhania University, Jhunjhunu, Rajasthan

Abstract

Breast cancer is the most common cause of death in women and the second leading cause of cancer deaths worldwide. Primary prevention in the early stages of the disease becomes complex as the causes remain almost unknown. However, some typical signatures of this disease, such as masses and micro calcifications appearing on mammograms, can be used to improve early diagnostic techniques, which is critical for women's quality of life. X-ray mammography is the main test used for screening and early diagnosis, and its analysis and processing are the keys to improving breast cancer prognosis. As masses and benign glandular tissue typically appear with low contrast and often very blurred, several computer-aided diagnosis schemes have been developed to support radiologists and internists in their diagnosis. In this article, an approach is proposed to effectively analyze digital mammograms based on texture segmentation for the detection of early stage tumors. The proposed algorithm was tested over several images taken from the digital database for screening mammography for cancer research and diagnosis, and it was found to be absolutely suitable to distinguish masses and microcalcifications from the background tissue using morphological operators and then extract them through machine learning techniques and a clustering algorithm for intensity-based segmentation.

Keywords: *Breast Cancer, X –Ray.*

1. INTRODUCTION

Cancer cells are termed as abnormal tumors. Breast cancer is any form of abnormal tumor that develops from breast cells. Breast cancer is the second biggest cause of deaths in women mostly aged between 30-55 years. In order to differentiate and identify the tumor region, first of all mammograms are taken and Region of Interests (ROI) is taken into consideration. Image enhancement is performed on the mammograms and then various algorithms are applied. In the existing literature different algorithms for clustering the image data has been presented. Recently studies show that one in 10 women will contract breast cancer in their lifetime, and that breast cancer is the leading cause of death of women between the ages of 35 and 54. Every year 27% of the new cancer cases in women are breast cancers . Although X-ray mammogram detection is best way of screening the breast cancer and ultrasound method is more popular because of its non-invasiveness and low cost . Due to high noise, low contrast radiologists cannot detect and classify the tumor or dense in breast cancer. Image enhancement is a best way for the diagnostic reliability by reducing noise effects in mammogram and filtering is a challenging process in ultrasound image processing since the noise is of unknown source with non specific form and trend. Several algorithms have been proposed to enhance the signal-tonoise ratio and to eliminate noise speckles.

2. LITERATURE REVIEW

Zahra abdolali kazemi, et.al (2021) Breast cancer accounts for 19% of deaths caused by cancers and 24% of all cases afflicted to cancers in European countries. Approximately 25% of breast cancer deaths happen in women between the ages of 40 and 49. Anomaly detection is done by the separation of the initial steps in computer detection systems. In this chapter, while reviewing the different techniques, a qualitative comparison between them will be provided. In this study, two approaches for the presentation of mammography (comparison of previous and current mammography images) are evaluated: together (simultaneously) and alternately on the same screen. In this study, MATLAB software is used.In this study, image processing algorithms of support vector machine (SVM), genetic algorithm (GA), convolutional neural networks (CNN), and K-nearest neighbours (KNN) are exploited. In this regard, the performance of these algorithms will be explained in this section. In this method, it is first essential to conduct training. Training means that a number of features related to class one and class two are given to the function, and the algorithm updates its parameters based on the labeling done. Then, the unlabeled data are given to the algorithm for the classification, and it automatically specifies the corresponding class. the Segmentation is the simplification or modification of image view for more meaningful and easier analysis. this is the process of labeling each pixel in each image, which results in a set of segments that together cover the whole image. By analyzing the resulting images, the physicians can identify cancer cells and offer their diagnostic results. It is possible to expand the MATLAB environment by adding a toolbox for various purposes.For simulation, training and classification need to be done with the classification method.

Saif Ali et.al (2020) Cancer, also called malignancy, is an abnormal growth of cells. There are more than 100 types of cancer, including breast cancer, skin cancer, lung cancer, colon cancer, prostate cancer, and lymphoma. Symptoms vary depending on the type. Cancer treatment may include chemotherapy, radiation, and/or surgery. According to American Cancer Society America will be encountering 1,806,950 new cases of cancer in the year 2020 causing 606,520 deaths. Cancer is the leading cause of death in the world. Cancer can be classified into two main categories malignant and benign. Early detection of cancer is the key to the successful treatment of cancer. There are various methodologies for the detection of cancer some include manual procedures, Manual identification is time-consuming and unreliable therefore computer-aided detection came into the research. Computer-aided detection involves image processing for feature extraction and classification techniques for the recognition of cancer type and stages. In this paper, several different algorithms have been discus sed such as SVM, KNN, DT, etc. for the classification of the different cancers. This paper also presents a comparative analysis of the researches done in the past.

Mutiullah et.al (2019) From last decade, lung cancer become sign of fear among the people all over the world. As a result, many countries generate funds and give invitation to many scholars to overcome on this disease. Many researchers proposed many solutions and challenges of different phases of computer aided system to detect the lung cancer in early stages and give the facts about the lung cancer. CV (Computer Vision) play vital role to prevent lung cancer. Since image processing is necessary for computer vision, further in medical image processing there are many technical steps which are necessary to improve the performance of medical diagnostic machines. Without such steps programmer is unable to achieve accuracy given by another author using specific algorithm or technique. In this paper we highlight such steps which are used by many author in pre-processing, segmentation and classification methods of lung cancer area detection. If pre-processing and segmentation process have some ambiguity than ultimately it effects on classification process. We discuss such factors briefly so that new researchers can easily understand the situation to work further in which direction.

Yousif M.Y Abdallah et.al (2018) Enhancement of mammography images considers as powerful methods in categorization of breast normal tissues and pathologies. The digital image software gives chance to improve the mammographs and increasing their illustration value. The image processing methods in this paper were using contrast improvement, noise lessening, texture scrutiny and portioning algorithm. The mammography images kept in high quality to conserve the quality. Those methods aim to augment and hone the image intensity and eliminate noise from the images. The assortment factor of augmentation depends on the backdrop tissues and type of the breast lesions; hence, some lesions gave better improvement than the rest due to their density. The computation speed examined used correspondence and matching ratio. The results were 96.3 ± 8.5 ($p>0.05$). The results showed that the breast lesions could be improved by using the proposed image improvement and segmentation methods.

Prannoy Giri et.al (2017) Breast Cancer is one of the significant reasons for death among ladies. Many research has been done on the diagnosis and detection of breast cancer using various image processing and classification techniques. Nonetheless, the disease remains as one of the deadliest disease. Having conceive one out of six women in her lifetime. Since the cause of breast cancer stays obscure, prevention becomes impossible. Thus, early detection of tumour in breast is the only way to cure breast cancer. Using CAD (Computer Aided Diagnosis) on mammographic image is the most efficient and easiest way to diagnosis for breast cancer. Accurate discovery can effectively reduce the mortality rate brought about by using mamma cancer. Masses and microcalcifications clusters are an important early symptoms of possible breast cancers. They can help predict breast cancer at it's infant state. The image for this work is being used from the DDSM Database (Digital Database for Screening

Mammography) which contains approximately 3000 cases and is being used worldwide for cancer research. This paper quantitatively depicts the analysis methods used for texture features for detection of cancer. These texture featuresare extracted from the ROI of the mammogram to characterize the microcalcifications into harmless, ordinary or threatening. These features are further decreased using Principle Component Analysis(PCA) for better identification of Masses. These features are further compared and passed through Back Propagation algorithm (Neural Network) for better understanding of the cancer pattern in the mammography image.

3. METHODOLOGY

The search strategy was performed by searching the databases such as Medical Literature Analysis and Retrieval System Online (MEDLINE) via PubMed, Springer, IEEE, ScienceDi-rect, and Gray Literature (including Google Scholar, articles published in conferences, government technical reports, and other materials not controlled by scientific publishers) for relevant publications from 2007 to 2017.

Keyword searches based on Mesh included “breast cancer”, “breast cancer screening techniques”, “artificial intelligence techniques”, and “medical image processing”. The symbol “*” was also used to allow retrieving all variations with suffixes of the source words. The above terms were combined using the logical connectives “AND”, “OR”, and “NOT”.

Among the articles searched, non-English articles as well as articles that used similar techniques and reported similar results were excluded. Moreover, articles whose full texts were not available were also excluded. Two experts independently reviewed all potentially relevant studies. Disagreements were solved with discussion and by using the viewpoint of a third expert.

Finally, 18,651 articles were extracted; however, majority of the extracted articles were deleted due to being repetitive and not having access to their full texts, and, 40 articles remained for review (Figure 1 shows the process.). The selected articles were examined in terms of the name of the method used, type of the image, advantages and limitations of each method, features used, and application results of each method, and all the results were presented in the form of tables

Image Annotation

Study images along with the corresponding radiology and pathology reports for each biopsied case were shown to 2 radiologists at our institution (R.W. and S.G.) for annotation. We asked the radiologists to identify masses and architectural distortions that were biopsied and to put a rectangular box enclosing them in the central slice using a custom software developed by a researcher (N.L.) in our laboratory. Each case was annotated by 1 of 2 experienced radiologists. The first radiologist, with 25 years of experience in breast imaging (R.W.), annotated 124 cases, whereas the second radiologist, with 18 years of experience in breast imaging (S.G.), annotated 77 cases. This way we obtained 190 bounding boxes for cancerous lesions in 173 reconstruction views and 245 bounding boxes for benign lesions in 223 reconstruction views. There were 336 and 99 bounding boxes for masses and architectural distortions, respectively, across cancerous and benign lesions.

4. RESULT

Training, Validation, and Test Sets

In total, our data set contained 22 032 reconstructed volumes that belonged to 5610 studies from 5060 patients. It was randomly split into training, validation, and test sets in a way that ensured no overlap of patients between the subsets. The test set included 460 studies from 418 patients. For the validation set, we selected 312 studies from 280 patients, and the remaining 4838 studies from 4362 patients were in the training set. The selection of cases from the benign and cancer groups into the test and validation sets was performed to assure a similar proportion of masses and architectural distortions. Descriptive statistics for all the subsets are provided in Table 1.

Table 1. Descriptive Statistics of the Data Set Used for Training, Validation, and Testing

Characteristics	No. Training set	Validation set	Test set
Patients			
Total	4362	280	418
Normal group, No. (%)	4109 (94.2)	200 (71.4)	300 (71.8)
Actionable group, No. (%)	178 (4.1)	40 (14.2)	60 (18.9)
Benign group, No. (%)	62 (1.4)	20 (7.1)	30 (7.2)
Cancer group, No. (%)	39 (0.9)	20 (7.1)	30 (7.2)
Studies	4838	312	460
Reconstruction volumes	19 148	1163	1721
Bounding boxes for cancerous lesions	87	37	66
Bounding boxes for benign lesions	137	38	70
Bounding box diagonal, mean (SD), pixels	344 (195)	307 (157)	317 (166)

If we discard skin cancer, breast cancer is the largest cause of cancer among women, accounting for one-third of all the cancer types. Obtaining the best outcomes in breast cancer depends on early diagnosis. Therefore, imaging techniques have been developed to increase the likelihood of early diagnosis of breast cancer and reduce unnecessary

Figure 1 Stages of systematic review

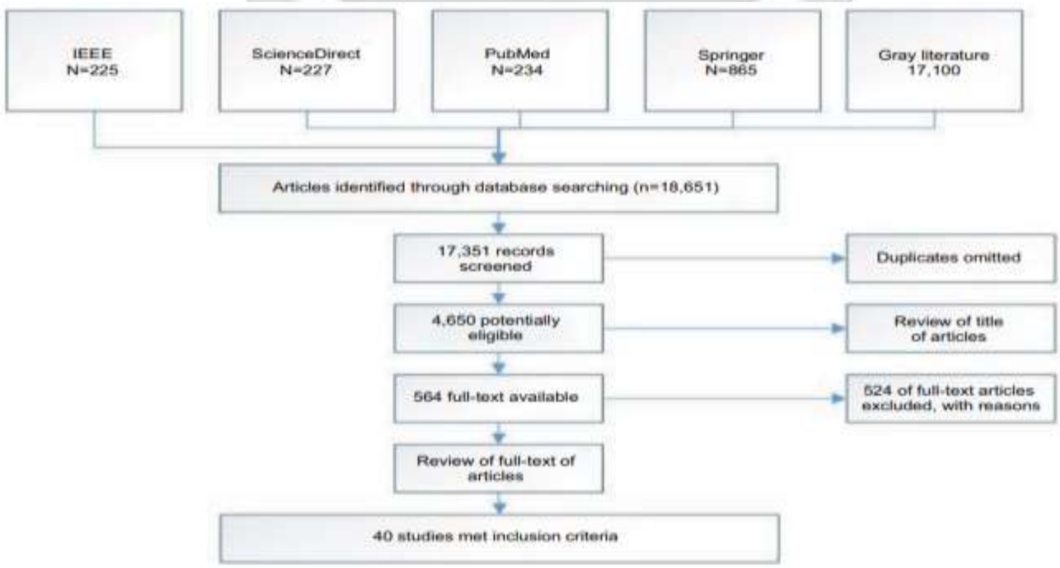

Table 2 Advantages and disadvantages of various imaging techniques in breast cancer

Imaging method	Application	Advantage	Disadvantage
Mammography	Golden standard imaging and diagnosis of breast cancer early stages	• It uses low levels of X-rays for imaging • This method is good for detecting DCIS and calcifications • Mammography is the gold standard method to detect early-stage breast cancer before the lesions become clinically palpable	• Radiation risk and other risks • Risk of false alarm • It is difficult for the radiologist to interpret the results from mammograms as mammograms generally have low contrast • Double reading of mammogram leads to increase in the cost of detection • Mammography alone misses many cancers in dense-breasted women
Ultrasound	Suitable for dense and soft tissues	• Widely available and accessible • Noninvasive • Quick • Highly sensitive • Suitable for women with dense breasts	• Quality and interpretation of the image depends highly on the skill of the person doing the scan
Thermography	Suitable for muscle tissue	• Noninvasive	• Physicians can have difficulty interpreting the images because of the low quality and low resolution of the images taken by the first generation of the medical infrared imaging cameras

Biopsy. Table 2 shows a summary of advantages and disadvantages of each method. Currently, digital image processing techniques are often used in solving machine visual problems and have provided good results. The importance and necessity of processing digital images are examined in the following two directions: 1) to improve images for human interpretation and 2) to process images for automatic understanding and interpretation by the machine.

In the field of diagnosis, first, medical images are collected and, then, preprocessing, segmentation, extraction of features, and eventually categorization are performed. Figure 2 shows the steps involved.

Image acquisition

The first step in processing images is to capture an image. In this step, data are collected in the form of digital images. The image format is usually a portable gray map, which is fixed format, and does not erase the image data when compressing images.

Image preprocessing

The next step is to preprocess input images to improve their quality by eliminating noise. Preprocessing of images is done through the middle filter. Preprocessing, in addition to deleting or reducing noise, improves image quality through increased contrast. Some of the most important preprocessing techniques are presented in Table 2.

Figure 2 Stages of cancer detection by image processing.

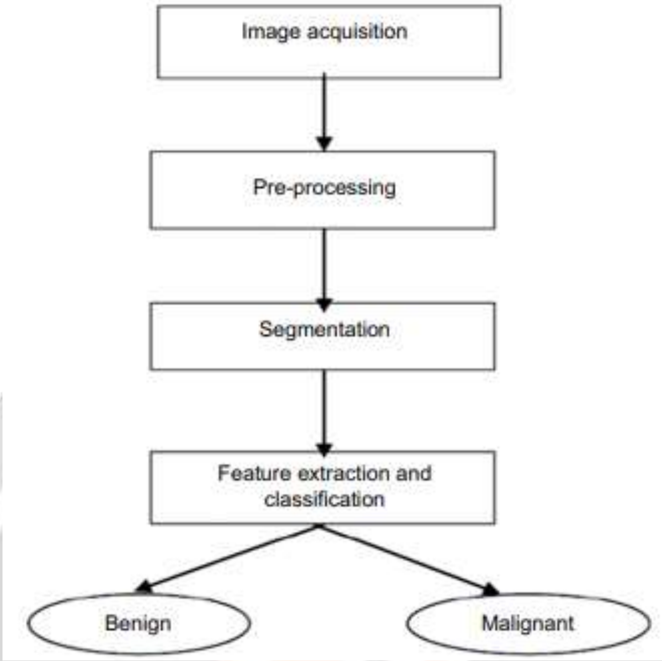

Table 3 Preprocessing techniques

FPN	FPN is the result of differences in responsivity of the detectors to incoming irradiance. It is a common problem when working with FPA. FPN for a particular configuration can be recovered from a blackbody image for later subtraction from the thermogram sequence.
Badpixels	A badpixel can be defined as an anomalous pixel behaving differently from the rest of the array. For instance, a dead pixel remains unlit (black), while a hot pixel is permanently lit (white). In any case, badpixels do not provide any useful information and only contribute to deteriorate the image contrast. A map of badpixels is generally known from the FPA manufacturer or they can be detected manually or automatically; the value at badpixel locations is then replaced by the average value of neighboring pixels.
Vignetting	Vignetting is another source of noise on thermograms that cause a darkening of the image corners with respect to the image center due to limited exposure. It depends on both pixel location and temperature difference with respect to the ambient. A correction procedure has been proposed.3
Temperature calibration	A transformation function is used to convert the grayscale values *g* provided by the infrared camera into a linearly incrementing physical quantity, eg, temperature. The procedure3 is used to position the IR camera in front of a reference temperature source (such as a blackbody and a thick copper plate) brought to various known temperatures. As the reference temperature source is varied, the IR

	images are recorded. Average of the central pixels in the field of view allows getting the calibration curve through a polynomial fit
Noise smoothing	One of the most useful preprocessing (and postprocessing) techniques is noise smoothing. For instance, neighbor processing can be performed bypassing a mask or kernel through the image. More elaborate noise removal techniques are available.

5. CONCLUSION

Breast cancer is the second leading disease for women in the world. In recent years, different types of methods developed to segment the mammogram images and it helped the radiologist to make an accurate decision about breast cancer cells. This research analyzed different stages of CAD and diagnosis methods for breast cancer using mammography. The mammogram image processed through various stages to detect whether a person is suffering from benign or malignant breast cancer. This paper analyzed segmentation and classification methods for breast cancer detection, which are much helpful for other researchers to enhance the traditional techniques in order to get better and accurate results. Also, this research paper is much useful for software developers to develop algorithms to improve the existing methods of fulfilling future requirements. The unsupervised classification of the cancer nuclei obtained 98.8% of accuracy by a combination of four CCS techniques in comparative analysis. In future work, the performance of the BCD can be enhanced by efficient classification and segmentation with optimization algorithms such as Ant colony algorithm, cuckoo search algorithm, particle swarm optimization and so on.

6. REFERENCE

[1] Yousif M.Y Abdallah et.al "Breast cancer detection using image enhancement and segmentation algorithms" 2018 Volume 29, Issue 20

[2] Prannoy Giri et.al "Breast Cancer Detection using Image Processing Techniques" DOI : http://dx.doi.org/10.13005/ojcst/10.02.19

[3] Zahra abdolali kazemi, "Diagnosis of breast cancer using image processing with SVM and KNN" Vol 13 No 1 Mar-Apr 2021

[4] Saif Ali et.al "Identification of Cancer Disease Using Image Processing Approahes" Volume 9, Issue 2, April 2020, Pages: 6-15

[5] Mutiullah et.al "Lung Cancer Detection Using Digital Image Processing Techniques: A Review" Vol. 38, No. 2, 351-360 April 2019

[6] Alhadidi B, Zu'bi MH, Suleiman HN. Mammogram breast cancer image detection using image processing functions. Information Technology Journal. 2007; 6 (2): 217-21.

[7] Vikrant Bhateja, Swapna Devi, An Improved Non-Linear Transformation Function for Enhancement of Mammographic Breast Masses, 978-1-4244-8679-3/11/$26.00 ©2011 IEEE,pp341-346.

[8] D. Coradini and M. G. Daidone, Bio-molecular prognostic factors in breast cancer, in Current Opinion Obstetrics Gynaecology, vol. 16, pp. 4955, 2004.

[9] Choudhari ,G, Swain ,D, Thakur, D, Somase, K . (2012) ."An adaptive approach to classify and detect the beast cancer using image processing", International Journal of Computer Applications (0975-8887) 45(17).

[10] Manohar S, Dantuma M. Current and future trends in photoacoustic breast imaging.Photoacoustics.(2019); 100134. https://doi.org/10.1016/j.pacs.2019.04.004 pmid:31871887

[11] Gupta AK, Chauhan SS, Shrivastava M. Low Contrast Image Enhancement Technique by Using Fuzzy Method. International Journal of Engineering Research and General Science. 2016 March-April; Vol. 4, Issue 2

[12] Chaira T, Ray AK. Fuzzy Image PreProcessing. Ch. 4 in: Fuzzy Image Processing and Applications with MATLAB. CRC Press, Taylor & Francis Group, LLC. 2009; pp. 45–68.

[13] American Cancer Society. Breast Cancer Facts & Figures 2019–2020. Atlanta: American Cancer Society, Inc. 2019.

[14] Bezdek, J. C., Hall, L. 0., Clark, M., Goldof, D. and Clarke, L. P. (1997). Segmenting medical images with fuzzy models: An update. In Fuzzy Information Engineering, ed. Dubois, D., Prade, H. and Yager, R., Wiley, NY, 69-92.
[15] Basset L.W., Manjikian V. III, Gold R.H. (1990). Mammography and breast cancer screening (Review). Surg Clinics of North America, 775-800.