LITERATURE SURVEY OF AUTOMATED DETECTION OF BREAST CANCER

P.Thirumurugan[1], P.Shanthakumar[2], N. Lavanya Devi[3], Jothilakshmi[4]

[1,3]Department of ECE, PSNA College of Engineering and Technology, Dindigul, Tamil Nadu, India

[2]Department of CSE, Srisubramanya College of Engineering and Technology, Palani, Tamil Nadu, India

[4]Department of Mathematics, Thiagarajar College of Engineering, Madurai, India.

ABSTRACT:

Breast cancer has become the major cause of the increase in mortality rate among women. The causes of breast cancer can be abnormal growth in the breast cell, dividing the cell faster than usual, and gathering of cells to form a mass. This abnormal growth will affect the cells in the nearby tissue also. The cancer cells in the breast frequently start in the milk- producing ducts. The causes of breast cancer can be a family history of breast cancer, age factor, exposure to radiation, beginning period at an earlier age, menopause at an older age, having first pregnancy at an older age, drinking alcohol etc. The symptoms of breast cancer can be the formation of lumps, discoloration, pain, change in size or shape, etc. This paper discusses the various algorithms and methods available in the literature survey for the detection of breast cancer in detail.

I. INTRODUCTION

Breast cancer is the most common type of cancer among women which leads to death if not identified at an earlier stage and treated [1,2]. The cancer cell spreads to other parts of the body and advances to the next stage if proper treatment is not given. A mammogram is used as one of the screening methods of breast cancer. The indication of the cancer is the presence of masses, presence of calcium deposit in the breast tissue which is seen as a bright spot in the mammogram, change in the shape of the breast [3]. At present detection of breast cancer is done by examining the affected cell under the microscope by an expert and trained pathologist. The efficiency of this method will depend on the expertise of the pathologist. Since the human intervention is present this method is highly prone to error. To overcome this limitation of the manual method automatic method of breast cancer classification came into existence. The feature-extract based automatic breast cancer classification involves preprocessing, segmentation, feature extraction, classification, and validation of the proposed algorithm.

Section II deals with different preprocessing techniques available in the literature, section III deals with the segmentation process, section IV discusses feature extraction and classification. Section V briefs about the non-feature extraction based methods, section VI conveys the concluding remarks of this paper.

II. PREPROCESSING

Preprocessing is a mandatory step in medical image processing. As the output obtained from the preprocessing is used for further processing, the utmost care has to be taken to select the most appropriate filter. Filters can improve the information content of the image, remove the noise, smoothen the image, enhance the edges, etc.

Median filter shows promising results for salt and pepper noise, a mean filter can be used for the removal of Gaussian noise, adaptive mean filter adjusts themselves according to the input image and are efficient in the removal of speckle noise, histogram equalization shows successful results for the images with both foreground and background dark or bright [4].

Vasantha et al. [5] proposed filtering techniques to remove the noise. Low pass filter can be used to remove the noise present in the image but LPF causes distortion in the edges. High pass filters can also be used to improve the image information. However, usage of low pass and high pass filters partially can result in a better quality of the image.

Roselin et al. [6] put forward histogram equalization as the preprocessing step. This method can improve the contrast illumination of the input image.

III. SEGMENTATION

Segmentation is the process of dividing the image into meaningful regions. The two types of segmentation are supervised and unsupervised segmentation. In supervised learning, the user should have knowledge about the number of classes and in unsupervised learning, the user may not have knowledge about the number of classes [7].

Pegah Faridi et al. [8] proposed an algorithm for the automated detection of the abnormal masses present in the breast tissues. Bilateral filtering is used as the preprocessing step and edges are preserved while applying the filter. The algorithm segments the image to the centroid of the nucleus and boundary of the cell nucleus. Steps like Gamma correction and thresholding, a morphological operation like dilation and erosion, DoG filtering, and thresholding are applied on the bilateral filtered image to detect the centroid of the nucleus. The cell nucleus boundary is identified by finding the points of change in the brightness level. The average accuracy in finding the cell nucleus and cell boundaries are 86% and 87% respectively. The distinguishing feature of this algorithm is no need for training which is considered to be the most time-consuming process.

Samir Kumar Bandyopadhyay et al. [9] proposed two types of segmentation. One is edge-based segmentation and another is segmentation based on region. Laplacian kernel method uses the edge-based segmentation. In region-based segmentation, each pixel in the image is compared to each group of pixel and the pixel is assigned to the group whichever is most similar. Compression of images can be done in a lossless or lossy format. But the medical images should be stored in a lossless format. JPEG format stores the image in a lossless format. The region of interest is detected and they are separated from the rest of the image.

G.Bharatha Sreeja et al. [10] initiated an algorithm for the detection of tumors in the breast tissue. The author has used linear contrast stretching as the preprocessing step. In this paper, the author has done coarse and fine segmentation of the mammogram. For doing the coarse segmentation apply one-dimensional wavelet form of the PDF of wavelet applied images of various channels and compute the local minima. The local minimum is fixed as the threshold value. For performing the fine segmentation form small and large windows. Compute the mean, maximum, and minima of the window for calculating the threshold value. Also, compute the difference between the minimum and the maximum. If the gray level value of the pixel is greater than the threshold and if the difference is greater than the mean value then the particular pixel belongs to the abnormal region, otherwise the pixel belongs to the normal region.

Albert Gubern et al. [11] developed an algorithm for the automated segmentation of the breast and estimating its density using MRI. As an initial step for segmenting the breast it is separated from the other organs of the body like the heart, lungs, thorax, etc. Bias field correction and image normalization are carried out as the preprocessing step. From the MRI image sternum forms the main identification for separating the breast from the body. From the segmented breast, the dense parts are segmented. The EM algorithm is used to segment the fibroglandular tissue. Different threshold methods were used for the segmentation of the fibroglandular tissue. The use of a single threshold showed heavy asymmetry between the left and right breast. This asymmetry is reduced by N3 bias field correction and segmenting the two breasts separately.

Hadjidj Ismahan et al. [12] initiated an algorithm for segmenting the mass present in the breast using mammogram. The mammogram image is passed through a morphological filter to remove unwanted radiopaque artifacts present in the image. The Image enhancement process is carried out to enhance the features present in the image. Then the image is converted to a binary image. The breast alone is obtained by convolving the mask obtained by the morphological process and the filtered image. The masses present in the image are identified by the watershed algorithm. For evaluating the performance of the suggested algorithm area overlap parameter is found.

Amresh Nikam et al. [13] proposed an algorithm for the segmenting of the nucleus of the breast cells from histology images. The input images undergo histogram equalization for making the darker images darker and lighter images lighter. The image is converted to grayscale images. Average of Potent and applied on the gray image to remove the pixels which are more close to the background. Difference method, Minimum error method, Otsu method, Max entropy method are computed. Thus the pixels which are whiter are identified and they are superimposed on the histology images. After this, graph-based segmentation is applied to the superimposed image. The next step is eliminating the large region by overlapping the original histology image and white pixel obtained from the segmentation. Hard thresholding, is performed on the image to neglect the false region and exactly identify the nuclei. In the next step morphological operations like dilation and erosion are performed on the image.

IV. FEATURE EXTRACTION AND CLASSIFICATION

Rafael Llobet et al. [14] proposed five different feature extraction techniques and compared the performance of these methods. The feature extraction methods proposed are

1) gray map: The gray map values of a pixel along with its neighborhood followed by Principal Component Analysis are computed for every pixel in the image.

2) Sobel: This filter consists of two kernels that can be used for detecting the magnitude and direction of the image. However, the author has used Sobel filters to detect the magnitude only.

3) Spatial Gray Level Dependence Matrices (SGLDM): The features considered are Difference Entropy, Sum Variance, Entropy, Difference Variance, Angular Second Moment, Contrast, Correlation, Variance, Inverse Difference Moment, Sum Average, and sum entropy. For the calculation of SGLDM, four angles -0^0,45^0, 90^0, 270^0 , and 1 to 4-pixel distance is considered.

4) Average Fraction Under the Minimum (AFUM): The pixels with gray level values less than the gray level value of the pixel p_{ij} within a radius r is calculated. The average of all these pixels is computed. The outstanding feature of this method is avoiding the time-consuming training process**.**

5) SFUM: This method is similar to the AFUM. The only difference is that the average of all the minimum pixels is not calculated. Instead, all the minimum value pixels are maintained as features. PCA is computed to reduce the dimension of this huge database. For these five feature extraction methods, K nearest neighbor is used as the classifier. Gray map and SFUM showed promising results compared to the other methods. P.Shanthakumar et al [19] proposed to review various MR Images to classify and segment the different image sets.

Said Pertuz et al. [15] put forward a method for the detection of breast cancer. For Full Field Digital Mammography and Digitized Screen-Film Mammography segmentation techniques used are thresholding and Hough based line detector respectively. From the input image, the breast is segmented and the Region of Interest is identified. Gradient features, spatial frequency, gray level run length, gray-level co-occurrence, statistical features are extracted. Gradient feature extraction considers the difference in gray level value in vertical and horizontal directions. The Spatial frequency feature considers the texture properties in the frequency domain, spatial domain, or in both domains. Gray level co-occurrence considers the occurrence of a pair of gray level value within a particular length and direction. Gray level run length considers the occurrence of a gray level values in consecutive lengths. Statistical property deals with the computation of the statistical parameters. Finally, risk scoring is calculated to find the risk ($r \epsilon$ [0 1]). 0 represents low risk and 1 represents the highest risk. For evaluating the performance of the system, AUC is the computed and obtained AUC as 0.876.

Chaitanya Varma et al. [16] suggested an alternative approach to examine breast cancer efficiently using mammograms. As a preprocessing step pass the input mammogram image to the high pass filter. Compute local texture and local entropy of the image. Select a proper threshold value for converting the image into a binary image. Apply morphological operation opening for edge smoothening and for filling holes. Finally, the background texture is being extracted and the background is being removed from the foreground image. Then calculate the local entropy, select the threshold to convert the image into binary, apply a morphological opening operation to remove the small objects in the image. Finally extract the texture and outline the suspected region. In the end, the author concluded that their future work will be to evaluate their work with some benchmark.

V. OTHER METHODS OF BREAST CANCER DETECTION

A. D. Belsare et al. [17] proposed Atanassov's Intuitionistic Fuzzy Set (AIFS) uses a multi-texture image to segment the breast duct. This method uses clustering methods with a multi-texture image to prepare the affinity matrix. Normalized cuts (NCUT) segment the breast duct using the calculated affinity matrix. IFCM algorithm creates AIFS image, objective, and membership function are calculated for clustering, updates the cluster centers, repeat the steps until the algorithm converges. Finally, the author has proved the efficiency of the proposed algorithm quantitatively and qualitatively.

Tom Botterill et al. [18] developed an algorithm for the detection of breast tissues by reconstructing the motion of the skin surface in three dimensions for the Digital image-based elasto-tomography (DIET). DIET utilizes elastography which identifies the disease by the presence of hard or soft tissue. The breast is vibrated and this motion is used to find the stiffness of the tissue. This paper deals with accurately measuring the motion in 3 dimensions. The different steps involved in the process are segmentation of the breast based on the model, reconstructing the 3-D surface from the profile, computation of the optical flow, motion reconstruction of the 3-D surface. The authors have checked the system on the breast with silicon phantom, simulated images on human data, on humans. This proposed system can detect tumors up to 10mm in a silicon phantom breast.

VI. CONCLUSION

A detailed survey of breast cancer detection is presented in this paper involving both feature extraction based method and non-feature extraction based method. The feature extraction based method involves preprocessing, segmentation, feature extraction, and classification. All these methods aid in the automated detection of breast cancer. This will help in the early diagnosis of the disease and reduce the mortality rate.

REFERENCE

[1]S. Kowsalya et al., "A survey on dignosis methods of breast cancer using Mammography," International Journal of Engineering Technology Science and Research,Vol. 2, Issue 4, pp: 2394-3386, 2015.

[2]Sathesh Raaj et al.,"Breast Cancer Dtetection Method- A Survey using Digital Mammographic techniques", international journal of Advanced science and technology, Vol. 29, no.3, pp. 4615-4619

[3] B.K.Gayathri, et al.,"A Survey of Breast Cancer Detection Based on Image Segmentation Techniques", International Conference on Computing Technologies and Intelligent Data Engineering, 2016.

[4]Dr.K.Meenakshi Sundaram et al.,"A Study On Preprocessing A Mammogram Image Using Adaptive Median Filter", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Issue 3,pp.10333-10337, March 2014.
[5] M. Vasantha et al.,Classification of Mammogram images using hybrid feature," European Journal of Scientififc Research ISSN 1450-216, Vol.57 No.1, pp: 87-96, 2011.
[6]K. Thangavel et al., "Mammogram mining with genetic Optimization of Anti-Miner parameters," International Journal Of Recent Trends In Engineering, Vol. 2, No. 3, 2009.
[7] P.Thirumurugan et al., "Automated Detection of Cervical Cancer", International Journal of Innovative Technology and Exploring Engineering, Vol 8. No.10, 2278-3075, 2019.
[8] Pegah Faridi et al., "An automatic system for cell nuclei pleomorphism segmentation in histopathological images of breast cancer", IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2016.
[9] Samir Kumar Bandyopadhyay et al.," Pre-processing of Mammogram Images", Vol. 2, pp.6753-6758, 2010,
[10] G.Bharatha Sreeja et al., "Detection of Tumours in Digital Mammograms Using Wavelet Based Adaptive Windowing Method", I.J.Modern Education and Computer Science, vol.1, 1-8, 2010.
[11] Albert Gubern-M´erida et al.," Breast Segmentation and Density Estimation in Breast MRI: A fully automatic framework", IEEE journal of biomedical and health informatics, vol. 19, no. 1, January 2015.
[12] Hadjidj Ismahan et al.," Mass Segmentation in Mammograms For Computer-Aided Diagnosis Of Breast Cancer", 3rd International Conference on Control, Engineering & Information Technology (CEIT), 2015
[13] Amresh Nikam et al. , "Nuclei Segmentation from Breast Cancer Histology Images", 1st International Conference on Emerging Trends and Applications in Computer Science, 2013.
[14] Rafael Llobet et al.," Comparison of Feature Extraction Methods for Breast Cancer Detection", pp. 495–502, 2005.
[15] Said Pertuz et al.," Open Framework for Mammography-based Breast Cancer Risk Assessment", IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2019.
[16] Chaitanya Varma et. al, "An Alternative Approach to Detect Breast Cancer using Digital Image Processing Techniques", International Conference on Communication and Signal Processing,pp.134-137, 2018.
[17] A. D. Belsare, "Breast Epithelial Duct Region Segmentation Using Intuitionistic Fuzzy Based Multi-Texture Image Map", IEEE India Council International Conference (INDICON), 2017.
[18] Tom Botterill, "Reconstructing 3-D Skin Surface Motion for the DIET Breast Cancer Screening System", IEEE transactions on medical imaging, Vol. 33, no. 5, May 2014.
[19]. P.Shanthakumar et al "A Survey on Soft Computing Based brain Tumot Detection Techniques using MR Images",Journal of Seybold Report,Vol.15(9), pp.2951-2962, 2020.