DESIGN AND IMPLEMENTATION OFHYBRID SEGMENTATION AND CLASSIFICATION APPROACH USING MAMMOGRAM IMAGE FOR DETECTION AND PREVENTION OF CANCER

[1]Neeti Arora,[2]Dr. Gaurav Aggarwal

[1]Research Scholar, [2]Associate Professor & Head Department of Computer Science & Engineering.
Jagannath University, NCR, Haryana.

Abstract : Breast cancer is the commonest form of cancer in women, all around the world[1]. Early detection can help in complete treatment of breast cancer. But due to lack of awareness and proper diagnostic tools, many ignore the symptoms or suffer in silence. The advent of mammography has helped in extensive breast cancer detection in women. Digital mammography is the most widely used technique in breast cancer detection. Computer Aided Diagnosis (CAD) and technology used to decode images have an important role to play in advanced diagnosis of breast cancer. In this paper efficiency and accuracy of these techniques are studied and theses are directly related to the rate of early breast cancer detection

IndexTerms: Breast cancer, Mammography, Detection & Image.

I. INTRODUCTION

Breast cancer is the commonest form of cancer in women, all around the world[2]. Early detection can help in complete treatment of breast cancer. But due to lack of awareness and proper diagnostic tools, many ignore the symptoms or suffer in silence. The advent of mammography has helped in extensive breast cancer detection in women. With advancements in technology, new tools have enabled doctors to identify and classify cancerous mass without the need of exploratory sample collection[3]. Differentiation between benign and malignant tumours can decide the long term prognosis of the patient. Rising number of breast cancer cases can be detected early with the help of computer aided cancer diagnostic technology.New hybrid segmentation approach assists in the processing of mammography images for early detection of cancer. The tumour images obtained by mammography are compressed, restored, recognized and the edge of the tumour is analysed in hybrid segmentation method. Tumour detection and detailed analysis is done with a set of algorithms that are used to decipher the mammography image. This fully automatic process involves two stages and hence is known as hybrid segmentation.

Steps in Mammography for Breast Cancer Detection:

Mammography uses X-rays as diagnostic tools for tumour detection during human breast examination. The images generated are studies using various tools and techniques to locate and identify possible abnormalities[4].

- Pre-processing:

  Mammograms cannot be interpreted easily due to low contrast and background artefacts. By setting certain threshold values, images can be processed to remove unwanted areas and increasing contrast for prominence of Regions of Interest.

[1]https://www.sciencedirect.com/science/article/pii/S1470204500002540
[2]https://www.sciencedirect.com/science/article/pii/S1470204500002540
[3]https://academic.oup.com/jnci/article/96/3/185/2521133
[4]https://pdfs.semanticscholar.org/650f/075316bcb7ad720c7206b572139d2daa1110.pdf

- Post-processing:

This step includes the division of pre-processed image into pixel blocks and abnormal mass area is highlighted by techniques discussed in the next section. Then the resultant images are subjected to segmentation techniques for image optimization, feature extraction, tumour identification and classification.

Different Approaches of Hybrid Segmentation:

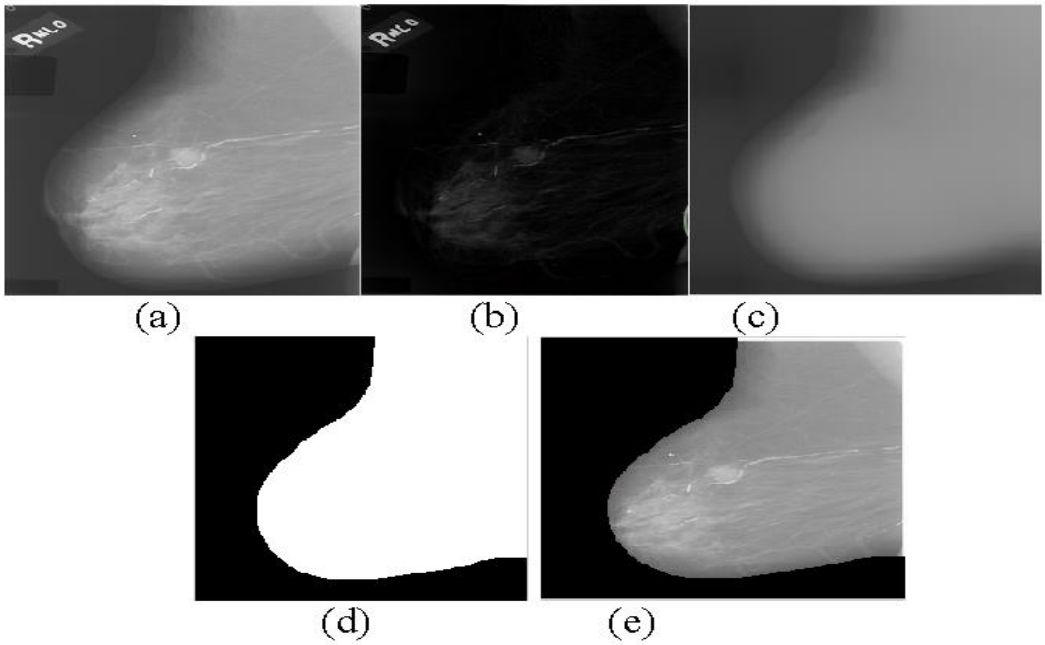
Though mammography is the gold standard in detecting breast cancer, poor contrast of the images and repetitive procedure makes tumor detection a difficult task. Advanced diagnostic segmentation results in high quality digital image processing[5].Expert radiologists use a variety of segmentation techniques to divide the mammography image into non-overlapping images. The aim of segmentation is to identify conceivable regions of interest in the image which could lead to early breast cancer detection. A clearer result is obtained when two or more segmentation techniques are combined, also known as hybrid segmentation.

Thresholding Technique and Watershed Segmentation:

Hefnawy, 2013[6], studied an improved approach of hybrid segmentation for semi-automated breast cancer detection. Mammography images collected from mammographic Image Analysis Society were processed using Median Filter, Global Threshold, Watershed Segmentation and Level Set Segmentation. Image preprocessing was done to filter out black background and unnecessary artifacts. Then global thresholding was applied to locate the breast tissue and crop out unwanted regions. Immersion based approach of Vincent and Soille's watershed segmentation was applied followed by smoothening of boundaries by level set method. The results provided significant improvement in accuracy of final segmentation. Further study remains be done on removal of false positives in level set segmentation technique.

A study by Hima Bindu and Swetha, 2015[7] assessed the usefulness of segmentation technique in breast cancer detection using edge based segmentation and Otsu's thresholding technique. The process employed segmentation algorithm to remove artifacts and patient details from the images. The two stage approach of hybrid segmentation was based on local region decision and final boundary localization. In the first approach used gradient information for morphological dilation of local region in an adaptive manner. The second approach used the dual front evolution to achieve final boundary of the image. Then Otsu's threshold was used to reduce intra class variance. The study proposed thecombined use of three techniques to generate greater accuracy in tumor edge detection. But the dependence of this method on clarity of mammography radiography required additional techniques.

[5]https://www.sciencedirect.com/science/article/pii/S0031320305002955
[6]https://pdfs.semanticscholar.org/225d/7648cd14d4273395828499073c8f391e8721.pdf
[7]https://www.researchgate.net/publication/300410139_Detection_of_Breast_cancer_with_Hybrid_image_segmentation_and_Otsu's_thresholding



a. Original input image.
b. Image with top-hat morphological operation.
c. a-b
d. Threshold image.
e. Output Image.

In a study conducted by Paschapur et al in 2016[8], the efficiency of a new hybrid technology including CLAHE histogram enhancement, Canny horizontal masking and k-means algorithm. The hybrid technique was applied to 30 hospital mammogram images. Images were preprocessed using CLAHE and subjected to canny edged masks and parallel k –means algorithm. After application of watershed segmentation, performance analysis parameters like PSNR, MSE and SSIM were used and the results were compared. Though the study evaluated the efficiency of proposed hybrid method over existing methods, it does not discuss the evaluation tools that can note the drawback of this method.

Multi-wavelet Technique

Accuracy in tumor mass detection in mammography image and differentiation using Local Seed Region Growing –Spherical Wavelet Transform hybrid method was demonstrated in a study conducted by Gorgel et al, 2013[9]. Mammograms obtained from Istanbul University and MIAS, were subjected to homomorphic filter for enhancement. ROI were identified using LSRG algorithm. After incorporating feature extraction and SWT, classification was done into mass or non- mass and benign or malignant using Support Vector machine with 96% and 93.59% accuracy respectively. More research remains to be done in exploration of LSRG for achieving greater accuracy with external validation.

Jabbari et al, 2016[10] proposed a new improved hybrid technology using combination of algorithms. The study aimed at development of technology to aid in the automatic tumor classification. 150 mammography images both of malignant and benign tumors were processed to remove artifacts. A combination of wavelet transform, genetic algorithm was used for image segmentation. The proposed method had greater accuracy when compared to ant colony optimization and PSO algorithms. It showed 91.4% accuracy in tumor edge detection without the need of unnecessary breast tissue biopsy. The study has scope in development of a near exact tumor detection technology with minimum errors.

[8]https://irjet.net/archives/V3/i8/IRJET-V3I8173.pdf
[9]https://www.sciencedirect.com/science/article/pii/S0010482513000796
[10]http://ijbd.ir/browse.php?a_id=556&sid=1&slc_lang=en&ppup=0

Machine Learning Techniques:

Early breast cancer can be detected early by locating microcalcifications in mammographic images[11]. But evaluation of microcalcifications is difficult due to their small size and low contrast in comparison to image background. Automated deep learning model can aid in more efficient identification and classification of these breast cancer indicators.

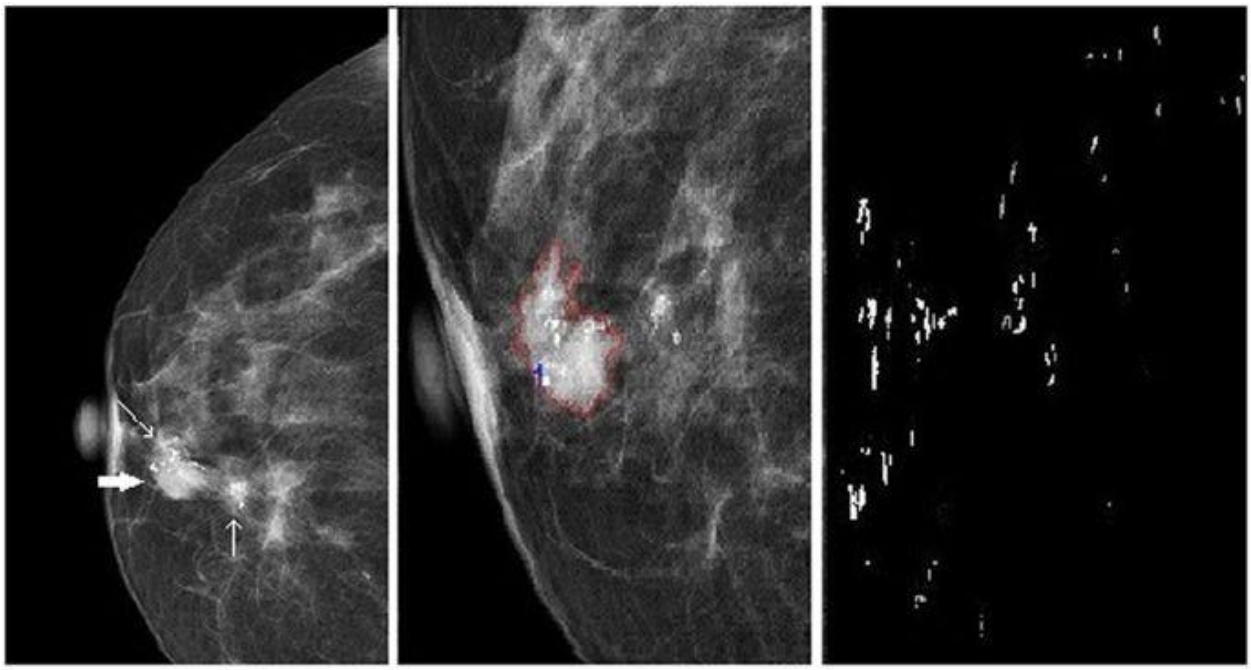
Saki, F. et al, 2012[12], studied three new machine learning rules which aimed at accelerating the training time of Multilayer Perceptron (MLP) classifier. Opposite Weight Back Propagation per Pattern (OWBPP), Opposite Weight Back Propagation per Epoch (OWBPE), Opposite Weight Back Propagation per Pattern in Initialization (OWBPI) and traditional Back Propagation algorithms were employed on MLP classifiers. The accuracy of the resultant classifications was compared and analysed quantitatively which showed that OWBPE had a greater convergence rate than usual Back Propagation. But the study doesn't address the problem of false positives and false negatives.

In another study on machine learning, Kozegar, E. et al, 2013[13], proposed a superior method for detecting masses on mammogram[14]and implementation of mass detection was done in machine learning. Adaptive thresholding technique was used to extract suspicious regions from mini-MIAS database. The false positives generated in the step were reduced by machine learning approach. The mass detection algorithm was analyzed by FROC and was found to have a good accuracy, although increasing the sensitivity in segmentation phase requires more study.

Localization of tumor mass is a daunting problem in Computer Aided Detection technique for diagnosis of suspicious mass on mammograms. Liu and Zeng, 2014[15] studied a new automatic breast cancer detection method. They used Multiple Layer Concentric (MLC) approach to locate suspicious regions in 219 images. Parameters were tuned with training data set and the regions of focus were refined further using Narrow Band Based Active Contour (NBAC). After extraction of texture and geometry features of ROI, computation was done by gray level co-occurrence matrix and Completed Local Binary Pattern (CLBP). Final classification was done using SVM and the entire method was evaluated on a data set of 429 images with 504 masses in craniocaudal view. The results showed improved detection sensitivity, though false positives were not completely removed.

Wang, J. et al, 2016[16] carried out a study to improve the diagnostic accuracy for discrimination of breast cancer with microcalcifications on mammography by employing deep learning. Mammograms from 1204 female patients with histopathological diagnosis of benign and malignant breast lesions were subjected to feature and data extraction using hybrid segmentation. Those features were used as input data in deep learning model for accurate classification. Their deep leaning model used stacked auto encoder that stacked multiple auto encoders in hierarchy to create deep network. Below is the segmentation of a mammogram of a 60 year old breast cancer patient.

[11]https://www.sciencedirect.com/science/article/pii/S0031320303001924
[12]https://www.sciencedirect.com/science/article/pii/S0010482512001709
[13]http://www.cancerjournal.net/text.asp?2013/9/4/592/126453
[14]http://mammoimage.org/interesting-papers/sampat_01.pdf
[15]https://www.sciencedirect.com/science/article/pii/S0925231214014003
[16]https://www.nature.com/articles/srep27327



a. Focal clustered microcalcifications and an irregular mass shown by thick arrow.
b. Automatic delineation of suspicious mass in red curve.
c. Segmented microcalcifications were used for feature characterization.

The study produced superior deep learning technique for large datasets, while it faced shortcomings like expansion of dataset and insufficiency of features for full characterization of microcalcifications.

A study on machine learning technique was conducted by Wang et al in 2018[17]. The study aimed at development of new hybrid learning approach for determining the likelihood of a detected breast tumor, being malignant. 301 mammographic images with suspicious lesions were collected as a dataset. A Convolutional Neural Network was pre trained to act as feature extractor and a pseudo color ROI method generated ROIs with RGB channels. Transmitted CNN features were collected and linear Support Vector Machine was trained to predict results. The results demonstrated the improved performance of CAD Scheme using new hybrid deep learning method. The study presents scope achievement of greater accuracy on different dataset sizes.

K-nn and Fuzzy Set Algorithm:

Rouhi, R and Jafari, M, 2015[18] conducted a study on classification of breast tumors based on hybrid level set segmentation, whose objective was development of an enriched mammography CADx system. Spatial Fuzzy Clustering, improved Region Growing or Cellular Neural Network was used for initial segmentation for the first algorithm. Parameters controlling level set were obtained from dynamic training technique by combining Genetic Algorithm and Artificial Neural network (ANN) or Memetic Algorithm and ANN, for the second algorithm. After segmentation using one of the proposed hybrid methods, feature extraction and selection were done by another Genetic Algorithm. Classification of tumors was done using ANN, random forest, SVM, k-nearest neighbor and naïve Bayes. Improved efficiency of the adaptive method in breast tumor classification was established while presenting scope for quicker classification of breast tumors.

Chowdhary and Acharjya, 2016[19], demonstrated the effectiveness of segmentation techniquein a study by hybridizing intuitionistic fuzzy set and rough set with statistical feature extraction method[20]. Processed images were subjected to zone

[17]https://www.researchgate.net/profile/Morteza_Heidari6/publication/323594047_A_hybrid_deep_learning_approach_to_predict_malignancy_of_breast_lesions_using_mammograms/links/5aa1a3b445851543e63a55a3/A-hybrid-deep-learning-approach-to-predict-malignancy-of-breast-lesions-using-mammograms.pdf
[18]https://www.sciencedirect.com/science/article/pii/S0957417415006983
[19]https://www.scientific.net/JBBBE.30.12

extraction and edge enhancement using intuitionistic fuzzy set. Feature extraction was done using gray-level-co-occurrence-matrix. Minimal reducts and rules were stimulated by rough set and fed to a classifier for zone identification and differentiation. This method offered 98.3% accuracy. The Possibilistic Exponential Fuzzy Clustering Means (PEFCM) segmentation can confer diagnostic imaging techniques with additional accuracy for early breast cancer detection.

Random Walks Algorithm and Active Contour:

Mass segmentation is crucial step in Computer Aided Diagnosis Systems. When suspicious mass lesions are surrounded by normal tissues, poor contrast and abstruse margins are major hindrances in lesion mass identification and differentiation.

Hao, X. et al, 2012[21] studied the accuracy of segmentation method involving Random Walks algorithm and active contour. A data set of 1095 ROIs were preprocessed and a set of seed points were produced for initial random walks segmentation. This produced initial mass contour and two probability matrices. Final segmentation was achieved by modified Chan-Vese contour model in which probability of matrices were updated by random walks round. Comparison with 4 other methods yielded higher accuracy in segmentation while applications of this segmentation technique can achieve greater finesse in segmentation of complex masses.

Feature Selection Techniques:

Feature selection is an important step in classification of breast cancer. It is the key preprocessing step to machine learning and is imperative in methods like pattern recognition, data mining methods, AI and medical image processing[22]. The increase in data dimensions hampers the efficiency of many feature selection methods. Rough set theory can reduce dimensionality and preserve features, but it cannot deal with real-valued data.

Jothi et al, 2013[23] studied the functionality of hybrid segmentation using Tolerance Rough Set (TRS) and Particle Swarm optimization (PSO). 19 features were extracted from segmented image using gray-level-co-occurrence-matrix. Hybrid Administered TRS, PSO grounded relative reduct and hybrid TRS, PSO based quick reduct were used for feature reduction and classification. Though this method fared well than rough set approach, minimal information loss does occur.

Tan, E. et al, 2014[24], conducted a study on maximizing breast mass distinction using two techniques namely SFFS (Sequential Forward Floating Selection) and SVM (Support Vector Machine). 181 mammogram image features were computed and relevant features were extracted using SFFS based feature selection technique. The performance was analyzed by SVM trained for classification. 600 malignant and 600 benign ROI were studied using ten- fold cross validation method. The characterization task was valued as 0.805+-0.012 AUC. It also indicated a reduction in efficiency of mass classification computation due to overlapping of tissues within benign masses.

Wajid and Hussain, 2015[25], proposed a novel feature extraction technique for detection of breast cancer, in their study. INbreast and MIAS datasets were used for the study. After extraction of Regions of Interest (ROI), their Local Energy Based Shape Histograms (LESH) features were calculated. The values were fed into Support Vector Machine (SVM) for classification of masses into benign and malignant. The additional classification of abnormality into different subtypes (masses and microcalcifications) was also accomplished with 100% accuracy by linear kernel. The study offers more prospects for the application of this technology in 3-D MRI images.

[20]https://www.sciencedirect.com/science/article/pii/S0262885606000746
[21]https://link.springer.com/article/10.1631/jzus.C1200052
https://www.sciencedirect.com/science/article/pii/S0957417413006659[22]
[23]https://www.igi-global.com/article/hybrid-tolerance-rough-set/101767
[24]https://link.springer.com/article/10.1007/s11548-014-0992-1
[25]https://www.sciencedirect.com/science/article/pii/S0957417415002997

Conclusion:

Digital mammography is the most widely used technique in breast cancer detection. Computer Aided Diagnosis (CAD) and technology used to decode images have an important role to play in advanced diagnosis of breast cancer. Efficiency and accuracy of these techniques/algorithms are directly related to the rate of early breast cancer detection. Cancer detected in its early stages has a greater chance of cure than those which are detected in advanced stages. With good prognosis, chances of successful cure and a better quality of life is higher.