A Multifarious Diagnosis of Breast Cancer Using Mammogram Images – Systematic Review

M.Swapna¹, Dr Nagaratna Hegde²
¹ Asst Professor, Dept of CSE, Vardhaman College of Engineering.
² Professors, Dept of CSE, Vasavi College of Engineering
1swapna0801@gmail.com
2nagaratnaph@gmail.com

Abstract. Breast Cancer is most common disease in worldwide leads to high rate in mortality. Detection of symptoms at early stage is difficult to identify the breast cancer. Diagnosis through General Medical Examination cannot detect the disease. This paper represents various common methods to detect the Breast Cancer and detection of tumors using mammogram images with Artificial Intelligence (AI). Algorithms helps Images in preprocessed using various feature extraction methods and classification algorithms to predict the class of tumors is benign or malignant. This article insight the recent approaches used in detection of tumors and stage of breast cancer.

Keywords Detection Breast Cancer, Volatile Organic Compounds, Urinary Exosomal, CT scan, Mammography, Ultra Scan.

1. Introduction

One of the leading diseases, which lead to more death in women, is cancer. When new patients are diagnosing with cancer, one of the four detected as breast cancer [1]. As per the ICMR (Indian Council of Medical Research), every year 70,000 people die from this disease and 1.5 lakhs are new cases [4]. Breast cancer is most common at age of above 40 years [5]. The p53 gene act as a tumor suppressor, which will regulates the normal cell proliferation and in p53 mutations obtains an output of nonfunctional protein, which accumulates in tumor cell nuclei leads to breast cancer [6]. Patients delay in regular screening and ignoring symptoms of the disease can cut the survival rate of current stage [7] and it is also observed and proved that tumor size also depending on the cause of the prognosis of the patient, as tumor size accumulate, less in survival rate [8]. Breast cancer occurs in Male and female, in few countries men has more incident rate but most countries female have more incident rate [9].

The radiologists identify the calcification (break down chips) and lesions (cotton balls) using mammograms [11]. Benign disorder breasts Asymmetric breast development and discharge of nipples, breast cysts, breast pain can because of breast cancer [12]. The presence of receptors like Estrogen, the progestin in cells lead to cancer, it is called as HER2 Positive [13]. When Estrogen and progestin receptor are tested negative and excess proteins are said to be Triple-negative Breast Cancer [14]. Lymph was built on the skin of the breast, symptom of breast cancer, swollen lymph are present under the armpits [15].

In this article, section 2 describes the extensive literature survey on various approaches used to predict the breast cancer, comprehensive study and performance is summarized in Table 1, section 3-detailed review on various preprocessing, segmentation, Feature extraction and section-4 reviews on classification methods.
2. Literature survey

2.1 Multifarious Methods for Breast Cancer Examination

Test for breast-screening method is very simple and attainable [18]. Breast Cancer SelfExamination has to perform at regular menstrual cycle from day three to five. Observations are Dimpling, Shape, and outline are deformed, nipple pin inside lumps near armpits to know the abnormality of breast [18]. A breath Analysis Test can be the primary screen for breast cancer disease, this test performed by analyzing the breath samples using the analytical technique is gas chromatography and measures the mass-to-charge ratio of ions in mass spectroscopy. The breath methylated alkenes contour values are determining breast cancer. Which determines a woman has breast cancer or not. The negative-positive values are more accurate than mammograms [37]. VOCs will help to identify the existence of breast cancer or not. Components like heptanes, 1-phenyl-ethanone, and isopropyl myristate, 2, 3-dihydro-1-phenyl-4(1H)-quinazolinone, 2-propanol. [34]. the novel approach for Breast Cancer diagnostics applies to premenopausal, younger, pregnant women who are Liquid biopsy-based on Biomarkers. Urinary miRs are used for breast cancer detection [22].

Ultrasound is an important approach for the detection of breast cancer, to avoid the needle biopsy and needle aspiration. Apart from lesion characteristics and reporting of lesions it also helps us to identify the lesion is benign or malignant. MRI Examination is used to detect and study the breast more accurately than any another method, even it is more costly and it is preferred for high-risk women’s and it does not expose the human body to heavy radiations and the duration of the test is for 15 to 30 minutes, the human body is passed into a narrow tube structure and create a magnetic field and radio frequency wave environment [16]. MRI Screening will result in effective Positive prediction values when we compare with mammography, ultrasound, and mammography with MRI. A patient lay on the machine. An X-Ray beam is projected only on the Breast portion of the body and it is a circular scan on different projection views and builds the 3D Images of the breast (300 to 500 projections in a 360-degree rotation of X-ray Tube). The disadvantage of CT scan is radiations observed are 10 times more than mammography [51]. CT scan helps to the detection of benign breast lesions and primary and secondary malignancies, it observed that the detection rate various from 24% to 70%, and it is an alternative examination. The radiologist should concentrate while performing regular chest CT scans carefully [2].

3. Related Work

3.1 Breast Cancer Detection using Mammogram

Mammogram images are used to detection of abnormal growth of cells. It uses a low dose of X-Ray and High quality with low radiations on Human breast [23]. In figure 1, the model diagram represents process flow to detect breast cancer. These Digital Images are meant to observe calcification or Lumps look like cotton balls. Mammogram images are preprocessed after reading an image, smoothing an image by removing noise data, identify the boundaries, Region of Interest, detection methods applied to the image [24].
Breast images were screened in two views one is the CC view another is the MLO view. The images are X-Ray Film records in either Digital or Film Screen Mammography. Digital Mammogram is viewed [26]. Images are stored on the storage devices with DICOM file format [27]. DICOM images are viewed with Open- Software like OSIRIX [28], ImageJ[29], Micro View, MIPAV[30]. Smoothing Image is a Images are preprocessing by Removal of noise like Salt & Pepper noise, Gaussian noise using Median Filter, Min and Max Filters to change the intensity of the image for clarity of the organ and denoising image is done by Winer filter [25].

Table 1: literature Survey on comparative Analysis of Detection of Breast Cancer

| S. No | Method of Detection | Examination Procedure | Factors for Observation | Size of the Dataset | Results |
|-------|---------------------|-----------------------|-------------------------|---------------------|---------|
| 1     | Breast Self Examinatio n | Self Examination of breast and observe of Armpits. | Lumps, Deformation of the shape, Dimpling, Nipple pins | None | AUC 81.40% |
|       |                     |                       |                         | None | AUC 89.70% |
| 2     | Biomarkers Analysis | Breath Analysis (VoC). analyze gas chromatography and mass spectroscopy[37] | VoC’s such as alkanes, ketenes, halogenated, hydrocarbon, aldehydes, esters [36]. | 101 | 93.8% sensitivity and 84.6% specificity |
|       |                     | Urinary Exosomal MicroR | miR-Factors | 109 | 98.6% sensitivity 100% specify |
| 3     | Magnetic Resonance Imaging | human body is passed into a narrow tube structure and create an magnetic field and radiofrequency wave enviourment | BRCA Gene Mutation Analysis | 529 | sensitivity is 91% Specificity of MRI is 97.2% |
| 4     | Ultrasound Images | sound waves are transmitted into body and bounce back ,they are collected by computer to create images[46] | Shape of lesion, margins, boundaries and orientation. | 167 | 71.3% Accuracy |
| 5     | Computer Tomograph y | Humandy body is passed an X-Ray Beam on 360-degree projection. | Abnormal Edges, shapes and RIM enhancement on breast | - | 24% to 70% Variance with radiologist |

The Edge detection method is alimages are divided into (Region of Interest) RoI using edge Detection techniques will able to identify the boundaries of the breast and analyze the shape of the breast, it also helps to observe both breasts are symmetry in shape or not. There are many proved
efficient methods for edge detection like Robert Edge Detection use for horizontal and vertical filters, Sobel Edge Detection used for edge detection, Prewitt Edge Detection for horizontal and vertical edges [31]. Zero-Cross, and Canny uses mathematical functions [32].

**Feature Extraction Methods** is a Detection of masses and calcification, identification of the region of interest. This phase is most important in breast cancer detection can be obtained by the classification of regions and clustering algorithms [17] [33].

4. Classification Techniques used for Breast Cancer Detection

Classification Techniques helps to classify the input data into a labeled class. Class features are defined based on the training dataset. This will help new facts to predict the class label. The broadly categorized algorithms are quadratic classifier, Random Forest Method, Neural Network, Support Vector Machine, Linear Classification and Logistic Regression.

4.1 Sequential Minimal Optimization (SMO) is methods used to train the Support Vector Machine (SVM), it is invented by John Carlton Platt. This approach is simple and fast, the concept used to solve the problem is dual quadratic optimization, which will include two elements are at each iteration. The large quadratic program divides into a small optimization problem. The smallest problem solved individually and parallel to decrease the time. The attributes like clump thickness, Cell Size, Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses are input variable X1,X2,…, Xn and Y1, Y2,…Yn are binary integer value has positive or negative values for classification. Support Vector Machine has trained by a quadratic optimization algorithm to create the margin. SMO method able to classify the instances correctly is 657 and incorrectly classify are 26, the overall accuracy achieved 96.19% [23].

4.2 Radial Basis Function Neural Network (RBFNN). This method is an artificial feed-forward neural network, which consists of three layers, Input Layer has four neurons Use to represent features like ASM, Correlation, Sum Entropy, and Sum Variance of the GLCM features extraction approach. The second layer is a hidden layer, which has a total number of neurons less than the no of epochs in the dataset; they depend upon the total no of datasets available. The third layer is the output layer, which is the linear sum of the radial basis function. The x value considers based on the threshold. Here ‘k’ is the upper threshold value and ‘j’ is the lower threshold value. If ‘x’ value is greater than ‘k’, then x=k or if ‘x’ is less than ‘j’ then x=j. This Work, There are two different classifications are performed, first time RBFN used to classify normal or abnormal and the second time RBFN used to classify the data is benign or malignant. ‘Y’ is the outputs of the network refer in equation (1).

\[ Y = \sum_{j=1}^{p} W_j \phi_j \]

\[ Y = \sum_{j=1}^{p} W_j \phi_j \]

\[ W_j \] are the weights from jth neuron to output layer.

\[ \phi_j \] is a kernel function, i.e. Gaussian Function refer in equation (2)

\[ \phi_j = \exp \left( -\frac{(x-c_j)^2}{2\sigma_j^2} \right) \]

\[ x = (x_1, x_2, x_3, x_d)^T \] is input vectors.
\[ C_j = (c_1, c_2, \ldots, c_d) \] is center vector, ‘\(\sigma\)’ is a spread.

### 4.3 Region Extraction by Clustering and Classification Algorithms (RECC)

In this work, mammogram images are giving as inputs and they are dividing into three steps. In step-1, removal of noise and enhancement of image using medium filter because it can efficiently remove noise for non-linear digital image and Gaussian Filter used to enhance the images. In step-2, images have segmented using a k-Means algorithm, every pixel compared with the centroid value of the cluster and the nearest cluster has mapped to the pixel. CART is a Classification and Regression Trees, which is a machine-learning algorithm, which helps to build a binary tree using the GINI Index to split the attributes. The recursive splitting of attributes stops at min-count at leaf node condition. The outcome is to predict the disease of class. This method has been achieved an accuracy of 93.75%. J48 algorithm has used for classification in the data-mining tool Weka. This algorithm will help to generate a binary tree based on the training data. It is an extension of the ID3 algorithm. It will generate a decision tree, which may be pruned or unpruned. the training set is generated based on information entropy. The attributes are divided into subset and entropy used to find the optimum attribute until the leaf nodes are pure. This algorithm is able to obtain an accuracy of 92.85%. JRIP is a repeated incremental pruning. The basic concept is to reduce the errors and it is one of the fast learning and efficient methods to generate a decision tree. It has different steps a) Growing Rule b) Pruning c) Optimization d) Selection.In Growing Rule will add the antecedents until it is 100% pure, the incremental inspecting the rule with previous rule of class, Optimization of rule, after generating the initial rule, and in the selection phase it will regulate the classes. This method is able to achieve an accuracy of 95.48%. Navi Bayesian Algorithm is a classification algorithm based on statistical attribute values, members of the class dependence on the attribute dependency of class. The training of the data has to perform and noisy data is to be replaced with the most frequent occurred value, a transformation of data into relational tables. Finally the predication of the class is obtained based on the attribute and class table values. This algorithm has achieved an accuracy of 93.61%. SVM (Support Vector Machine) in a machine-learning algorithm, it the most commonly used method used for binary classification, which separates the group of data into classes. The decision line drawn between the data objects falls. It will separate them into two classes. This approach has obtained an accuracy of 92.40%.
Fig 2 Comparative Analysis of Performance Measure of Accuracy

Table 2: Multifarious Algorithms used to classify the Classes

| S. No | Author | Feature Extraction | Attributes | Classification Algorithm | No. of Images | Accuracy (%) |
|-------|--------|-------------------|------------|--------------------------|---------------|--------------|
| 1     | Vikas et. al [23] | Numeric values as input | clump thickness, Cell Size, Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses are attributes numerical values | Sequential Minimal Optimization | 683 | 96.19% |
| 2     | Melissa et. al [33] | Gray-level co-occurrence matrix | ASM, Correlation, Sum Entropy and Sum Variance | Radial Basis Function Neural Network | 330 | 91.57% |
| 3     | Venkatesan et. al [37] | Gray-level co-occurrence matrix | Clustering images into K cluster using K-means algorithm. Each pixel is comparing with centric of cluster to consider in nearest cluster. | Classification and Regression Trees (CART) | 50 | 93.75 % |
|       |        |                   |            | Naive Bayes Algorithm     |               | 93.61 %      |
|       |        |                   |            | Support Vector Machines   |               | 92.40 %      |
|       |        |                   |            | J48 Algorithm             |               | 92.85 %      |
|       |        |                   |            | JRecursive Incremental Pruning |         | 95.45 %      |

5. Conclusion
In this paper, Different approaches to detection of breast cancer and brief knowledge of characteristics and symptoms of BC. Preliminary image profile screening is best option for detection of breast cancer. Mammogram image analysis proven to be more accurate compare to any another approach, even patient will undergoes low radiations, cheaper in cost and most important is highly availability of technology. Images are preprocessed and identify the ROI with feature extraction, In this work images are preprocessed and segmented using Gray-level co-occurrence matrix and one of the example numeric values are considered. This paper discuss on various method used to classify normal/abnormality of patients using different classification methods SMO, RBFN, JRIP, Naive Bayesian, SVM, J48 and CART. The Sequential Minimal Optimization is achieving maximum accuracy in comparison of other methods.

References
[1] American Institute for Cancer Research, Published an article on “world cancer statistics for the most common cancers” in 2018.

[2] O’Connell, A.M., Karellas, A., Vedantham, S. and Kawakyu-O’Connor, D.T., 2018, February. Newer technologies in breast cancer imaging: dedicated cone-beam breast computed tomography. In Seminars in Ultrasound, CT and MRI (Vol. 39, No. 1, pp. 106-113). WB Saunders.

[3] Kanimozhi, G., Shanmugavadivu, P. and Rani, M.M.S., 2020. Machine Learning-Based Recommender System for Breast Cancer Prognosis. Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries, p.121.

[4] Jemal, A., Ward, E. and Thun, M.J., 2007. Recent trends in breast cancer incidence rates by age and tumor characteristics among US women. Breast Cancer Research, 9(3), pp.1-6.

[5] Allred, D.C., Clark, G.M., Elledge, R., Fuqua, S.A., Brown, R.W., Chamness, G.C., Osborne, C.K. and McGuire, W.L., 1993. Association of p53 protein expression with tumor cell proliferation rate and clinical outcome in node-negative breast cancer. JNCI: Journal of the National Cancer Institute, 85(3), pp.200-206.

[6] Rauscher, G.H., Ferrans, C.E., Kaiser, K., Campbell, R.T., Calhoun, E.E. and Warnecke, R.B., 2010. Misconceptions about breast lumps and delayed medical presentation in urban breast cancer patients. Cancer Epidemiology and Prevention Biomarkers, 19(3), pp.640-647.

[7] Garg, M., Nagpal, N., Sidhu, D.S. and Singh, A., 2016. Effect of lump size and nodal status on prognosis in invasive breast cancer: Experience from Rural India. Journal of clinical and diagnostic research: JCDR, 10(6), p.PC08.

[8] Ly, D., Forman, D., Ferlay, J., Brinton, L.A. and Cook, M.B., 2013. An international comparison of male and female breast cancer incidence rates. International journal of cancer, 132(8), pp.1918-1926.

[9] Karmilasari, S.W., Hermita, M., Agustiyan, N.P., Hanum, Y. and Lussiana, E.T.P., 2014. Sample K-means clustering method for determining the stage of breast cancer malignancy based on cancer size on mammogram image basis. IJACSA) Int J Adv Comput Sci Appl, 5(3), pp.86-90.

[10] Nguyen, Q.D., Tenreiro, A., Roberts, J.T., Tavana, A. and Robinson, A.S., 2020. Hematoma Mimicking Breast Cancer on CT Scan and Breast Ultrasound. Cureus, 12(7)

[11] Santen, R.J. and Mansel, R., 2005. Benign breast disorders. New England Journal of Medicine, 353(3), pp.275-285.

[12] Harrell, J.C., Dye, W.W., Allred, D.C., Jedlicka, P., Spoelstra, N.S., Sartorius, C.A. and Horwitz, K.B., 2006. Estrogen receptor positive breast cancer metastasis: altered hormonal sensitivity and tumor aggressiveness in lymphatic vessels and lymph nodes. Cancer research, 66(18), pp.9308-9315.

[13] Anders, C. and Carey, L.A., 2008. Understanding and treating triple-negative breast cancer. Oncology (Williston Park, NY), 22(11), p.1233.

[14] Marino, M.A., Avendano, D., Zapata, P., Riedl, C.C. and Pinker, K., 2020. Lymph Node Imaging in Patients with Primary Breast Cancer: Concurrent Diagnostic Tools. The oncologist, 25(2), p.e231.

[15] Bakic, P.R., Barufaldi, B., Pokrajac, D.D., Weinstein, S.P. and Maidment, A.D., 2018, July. Optimized simulation of breast anatomy for virtual clinical trials. In 14th International Workshop on Breast Imaging (IWBI 2018) (Vol. 10718, p. 107181X). International Society for Optics and Photonics.

[16] Kuhl, C.K., Schrading, S., Leutner, C.C., Morakabati-Spitzen, N., Wardelmann, E., Fimmers, R., Kuhn, W. and Schild, H.H., 2005. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. Journal of clinical oncology, 23(33), pp.8469-8476.

[17] Prasad, G.V. and Raju, S.V., 2018. A survey on local textural patterns for facial feature extraction. International Journal of Computer Vision and Image Processing (IJCVIP), 8(2), pp.1-26.
[18] Semiglazov, V.F., Moiseyenko, V.M., Bavli, J.L., Migmanova, N.S., Seleznyov, N.K., Popova, R.T., Ivanova, O.A., Orlov, A.A., Chagunava, O.A., Barash, N.J. and Matitzin, A.N., 1992. The role of breast self-examination in early breast cancer detection (results of the 5-years USSR/WHO randomized study in Leningrad). European journal of epidemiology, 8(4), pp.498-502

[19] Costantini, M., Belli, P., Lombardi, R., Franceschini, G., Mulè, A. and Bonomo, L., 2006. Characterization of solid breast masses: use of the sonographic breast imaging reporting and data system lexicon. Journal of ultrasound in medicine, 25(5), pp.649-659.

[20] Sehgal, C.M., Weinstein, S.P., Arger, P.H. and Conant, E.F., 2006. A review of breast ultrasound. Journal of mammary gland biology and neoplasia, 11(2), pp.113-123.

[21] Thaineua, V., Ansusinha, T., Auamkul, N., Taneepanichskul, S., Urairoekkun, C., Jongvanich, J., Kannawat, C., Traisathit, P. and Chitapanarux, I., 2020. Impact of regular Breast Self-Examination on breast cancer size, stage, and mortality in Thailand. The Breast Journal, 26(4), p.822.

[22] Sood, R., Rositch, A.F., Shakoor, D., Ambinder, E., Pool, K.L., Pollack, E., Mollura, D.J., Mullen, L.A. and Harvey, S.C., 2019. Ultrasound for breast cancer detection globally: A systematic review and meta-analysis. Journal of global oncology, 5, pp.1-17.

[23] Singh, A.K. and Gupta, B., 2015. A novel approach for breast cancer detection and segmentation in a mammogram. Procedia Computer Science, 54, pp.676-682.

[24] Makandar, A. and Halallli, B., 2015. Breast cancer image enhancement using median filter and CLAHE. International Journal of Scientific & Engineering Research, 6(4), pp.462-465.

[25] Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J. and Cardoso, J.S., 2012. Inbreast: toward a full-field digital mammographic database. Academic radiology, 19(2), pp.236-248.

[26] Vachon, C.M., Fowler, E.E., Tiffenberg, G., Scott, C.G., Pankratz, V.S., Sellers, T.A. and Heine, J.J., 2013. Comparison of percent density from raw and processed full-field digital mammography data. Breast Cancer Research, 15(1), p.R1.

[27] Yakami, M., Ishizu, K., Kubo, T., Okada, T. and Togashi, K., 2011. Development and evaluation of a low-cost and high-capacity DICOM image data storage system for research. Journal of Digital Imaging, 24(2), pp.190-195.

[28] Hermena, J., Molla, C. and Jahnen, A., 2008. Managing and viewing DICOM images with ImageJ. Proc. ImagesJConf.

[29] Haak, D., Page, C.E. and Deserno, T.M., 2016. A survey of DICOM viewer software to integrate clinical research and medical imaging. Journal of digital imaging, 29(2), pp.206-215.

[30] Singh, N., Mohapatra, A.G. and Kanungo, G., 2011. Breast cancer mass detection in mammograms using K-means and fuzzy C-means clustering. International Journal of Computer Applica1.

[31] Pratiwi, M., Harefa, J. and Nanda, S., 2015. Mammograms classification using gray-level occurrence matrix and radial basis function neural network. Proceeding Computer Science, 59, pp.83-

[32] Hirschfeld, M., Rücker, G., Weiß, D., Berner, K., Ritter, A., Jäger, M. and Erbes, T., 2020. Urinary Exosomal MicroRNAs as Potential Non-invasive Biomarkers in Breast Cancer Detection. Molecular Diagnosis & Therapy, pp.1-18.

[33] Phillips, M., Cataneo, R.N., Ditkoff, B.A., Fisher, P., Greenberg, J., Gunawardena, R., Kwon, C.S., Tietje, O. and Wong, C., 2006. Prediction of breast cancer using volatile biomarkers in the breath. Breast cancer research and treatment, 99(1), pp.19-21.

[34] Speiser, D., Schneider, A., Staeck, O., Freitag, C., Lanowska, M. and Mangler, M., 2012.
Volatile organic compounds (VOCs) in exhaled breath of patients with breast cancer in a clinical setting. Ginekologia polska, 83(10).

[37] Venkatesan, E. and Velmurugan, T., 2016. Extraction of cancer affected regions in mammogram images by clustering and classification algorithms. Indian Journal of Science and Technology, 9(30), pp.1-9.