IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A REVIEW ON BREAST CANCER DETECTION APPROACHES FOR VARIOUS IMAGE DATA SETS

[1]Mr. Gurusiddayya Hiremath,[2]Dr. Jose Alex Mathew

[1]Assistant Professor, [2]Professor,

[1]Department of Electronics & Communication Engineering, [2]Department of Electrical & Electronics Engineering

[1]Sahyadri College of Engineering and Management, Mangaluru, India, [2]Srinivasa Institute of Technology Mangaluru, India

Abstract: According to global statistics, deaths among women due to breast cancer (BC) are the leading cause out of all the several types of cancers. Hence treating breast cancer as early as possible, and even relatively complicated to detect and analyse at the beginning stage. The conventional method is time-consuming and not efficient and very less accurate. Henceforth an efficient technique to diagnose the cancerous cell, not including human association close to accuracy. CT scan is a particular case of Mammography, which adopts the X-ray technique & uses superior-resolution pictures such that, it perceives fine tumors in the breast. The review paper elaborates on the recognition of the BC by employing various image processing (IP) application and techniques

***Index Terms* - Cancer, Breast Cancer, Machine Learning, Mammogram, Histopathology image, Pre-Processing, Segmentation**

I. INTRODUCTION

According to stats of the “World Health Organization (WHO)”, Breast cancer is the foremost cause of mortality in women of all cancers. [1], hence initial caveat and analysis helpful in preventing the death rate by making out the disease in its early stages. Patients who are diagnosed on time will have a better chance of avoiding the undesirable spread of malignant cancer cells. The anomalies in the breast are of different categories like masses, speculated lesions, micro calcifications plus architectural distortions. These anomalies take place in two means called begin as well as malignant. Non-cancerous abnormalities are begun and abnormalities are reported as malignant cancers. The breast masses usually occur in the intense region with dissimilar shapes and sizes, which include shapes such as stellate, circumscribed, and lobulated. Breast muscles, fibrous tissues, and breast parenchyma are difficult to identify due to low contrast, varying sizes and forms, and similarities to other blood vessels. At present, it is fairly simple to gather and store a bunch of data sets and create databases of patients in electronic documents with the latest computing tools [2], like the National Breast Cancer Foundation (NBCF), patient database [3]. With the help of the computer, it is possibl9e for health professionals to examine these various databases predominantly whenever required to study the various complex breast cancer data. figures 1 and figure 2 give the image of ultrasound and mammogram images of a breast cancer cell.

Fig1. Ultrasound Image [1]

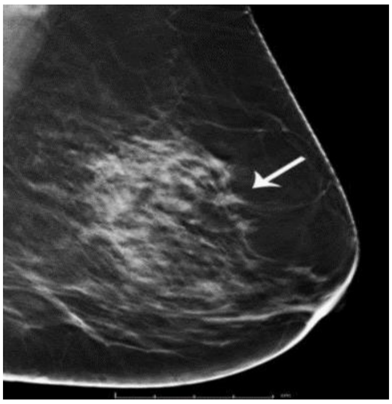

Fig 2. Mammogram Image [1]

In current scenario cancer is one of the deadliest disease as compared to other types of disease. If we enlist they type of cancers, and list as follows “Brain and nervous system cancer, Bladder cancer, Non-Hodgkin lymphoma, Leukaemia, Prostate cancer, Liver and intrahepatic bile duct cancer, Breast cancer, Pancreatic cancer, Colorectal cancer, Lung and bronchus cancer”. When you consider different sets of cancer and its death rate statistics as follows.

Table 1. Details for type of cancer and numbers in lacs

Cancer Type	Cases in Lacs
Lung	20.09
Breast	20.09
Colorectal	10.80
Prostate	10.28
Skin Cancer	10.04
Stomach	10.03

When we are dealing with cancer the initial diagnosis and prognosis is very essential to reduce the mortality rate. Based on the various category of the survey one of the deadliest cancer is BC in women [4]. In our paper, we are going to deal with different BC data sets and related algorithms and their implementation and conclusion based on the analysis.
Paper content starts with introduction of the breast cancer followed by different data sets, Comparative analysis of different dataset with different algorithm and conclusion.

II. DATA SETS

Any cancer can be diagnosed fundamentally by three main category of images first one is histopathology images, second one is x-ray or mammography image and third one is thermal image. Based on this main classification various data sets being curated/generated for the diagnosis and further understand condition of patient. when you enlist the different BC data sets and data sets are as follows Wisconsin data set [4], BreakHis data set [4], [5], [8], "BACH (BreAst Cancer Histology)- ICIAR 2018 Grand Challenge [8], [10], Patch-Camelyon [8], Bioimaging [8], The NYU breast cancer screening dataset V1.0[11], DDSM database [12], INbreast database [12], Mini-MIASdatase (very old one) [16]. Each one has certain unique feature to base on the condition and stage of cancer. Any research or doctor wants to identify different Metastasis Process; one can refer the different data sets based on the domain knowledge. Out of all the datasets few are taken for discussion and following are details.

Cancer may be properly diagnosed by using multiple Machine-Learning (ML) approaches to analyse the Wisconsin Breast Cancer Database (WBCD) database, and table1 offers the WBCD Attributes for breast cancer.

Table2. Attributes of Wisconsin Breast Cancer Database (WBCD)

Number	Attribute	Domain
01	Sample code number	id number
01	Clump Thickness	1-10
02	Uniformity of Cell Size	1-10
03	Uniformity of Cell Shape	1-10
04	Marginal Adhesion	1-10
05	Single Epithelial Cell Size	1-10
06	Base Nuclei	1-10
07	Bland Chromatin	1-10
08	Normal Nucleoli	1-10
09	Mitoses	1-10
10	Class	2 for benign 4 for malignant

Table3 give the description of the four sample mammograms like background tissue nature and irregularity present in the class. And figure 3 & figure 4 gives the double thresholding applied.

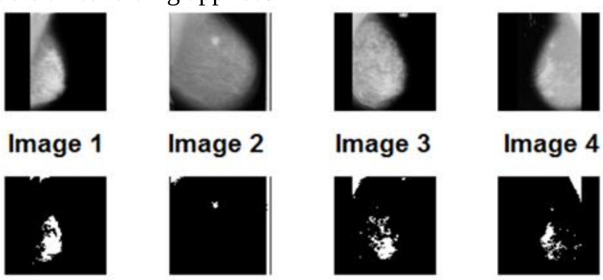

Fig3: Samples of four mammogram images results;
row1: the original set of mammogram images, row2: after performing double thresholding [9]

Figure 4 gives two different images captured from the fine needle biopsies of the breast for fine needle aspirate (FNA) of breast tumor [14]

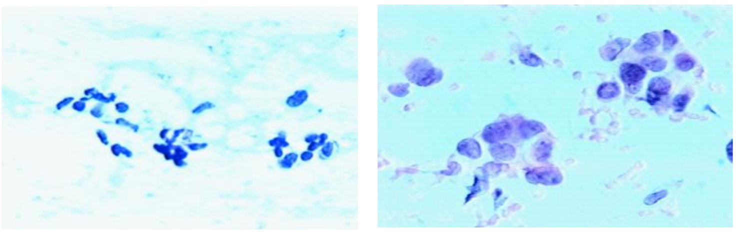

Fig4: Images from the FNA test: (a) Benign, (b) Non-Benign

Table3. Description of the sample mammograms [9]

Images	Background tissue nature	Abnormality present in Class	Abnormality present in Class
Image 01	Fatty-glandular	Asymmetry	Malicious
Image 02	Fatty	Other, ill-defined masses	Malicious
Image 03	Dense-glandular	Architectural distortion	Malicious
Image 04	Fatty-glandular	Well-defined/circumscribed masses	Malicious

Within the WPBC/WPBC dataset, Table 3 provides nuclei ten characteristic attributes such as radius, texture, area, smoothness, compactness, concavity, concave points, symmetry, Perimeter and fractal dimension for tumour size" and "lymph node status."

Table4. Cell nuclei of WDBC/WPBC characteristic attributes

1	Radius	Mean of distances from centre to points on the perimeter
2	Texture	Standard deviation of grey-scale values
3	Perimeter	--
4	Area	--
5	Smoothness	Local variation in radius lengths
6	Compactness	((perimeter)2 / area) – 1
7	Concavity	Severity of concave portions of the contour
8	Concave points	Number of concave portions of the contour
9	Symmetry	--
10	Fractal dimension	coastline approximation – 1

BreakHis (Breast Cancer Histopathological Image Classification) data set is a collection of "9,109 microscopic images of various breast tumor tissue and samples collected from 82 patients using different magnifying factors like 40X, 100X, 200X, and 400X" [8], [15], [16].

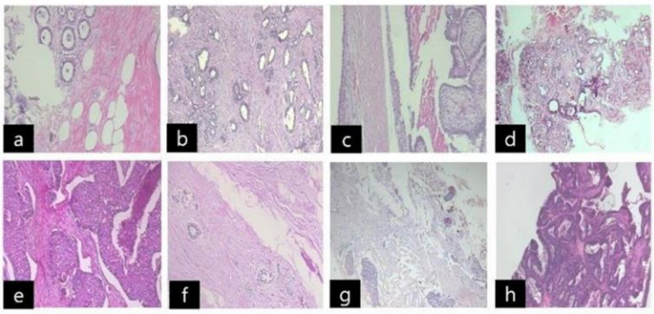

Fig4. BreakHis dataset Benign & Malignant images: [8]

Trials are prepared from breast tissue biopsy slides, marked with "hematoxylin and eosin (HE)" [17]. To diagnose with confidence, the presence of cancer by doing the biopsy. Out of all the biopsy techniques like "Fine needle aspiration, Core needle biopsy and Vacuum-Assisted, and surgical or open biopsy" are general one.

III. DIFFERENT COMPARATIVE APPROACHES FOR RESPECTIVE DATA SET

The subsequent table 5 describes the associated works for the recognition of breast cancer with IP and ML techniques

Author Name	Paper Reference Number	Approaches and Results
Omar Ibrahim Obaid et al.,	[1]	Comparison among several classifiers are support vector machine (SVM), K-Nearest Neighbour, Decision Tree (DT), and artificial neural network (ANN) to find the best classifier for better accuracy for Wisconsin Breast Cancer (Diagnostic) datasets According to the findings of this study, quadratic (SVM) support vector machines have the highest accuracy (98.1 percent) with the lowest false discovery rates. The area under the ROC curve in percentage for malignant tumours is 98.43% and 98.85% respectively, which is better than other classifiers.
WenbinYue et al.,	[2]	With benchmarking database WBCD different ML approaches like SVM, K-NN, and DT are being used. Each approach mentioned the accuracy of diagnosis and prognosis, for example, K-NN maximum classification precision by 98.70% as the Euclidean distance and the City block distance by k = 1, achieved 98.48%, and Cosine distance and Correlation by 95.67% and 94.69% respectively.
B.M. Gayathri et al.,	[6]	BC analysis using different machine learning algorithms and methods corresponding to Neural networks, SVM, RVM, and ELM. Paper explains the various method and their accuracy of detection and prediction from various authors like the Backpropagation algorithm used for training Multilayer Perceptron (MLP) -99.28%, SVM-KNN classifier -98.06%
Ahmed Hamza Osman et al.,	[7]	The paper has outlined two-step SVM algorithms; the hybrid method enhances precision by 99.1 percent in comparison to other classifiers on the UCI-WBC data set like SVM-RBF -96.84%, *Particle Swarm Optimization*(PSO)-SVM -96.99%, Evolutionary NN -95.06%, etc.
Samir M. Badawy et al.,	[9]	Double thresholding segmentation applied to mammogram images, Masking and Morphological Operations, The Applied Segmentation Approach for better detection and visualization.
Ilias Maglogiannis, et al.,	[14]	SVM algorithm does brilliantly, demonstrating high values of precision with 96.91%, specificity with 97, 67% and sensitivity with 97,84%
A. Marcano-Cedeño et al.,	[16]	Neural Network training for pattern classification with AMMLP classification with an accuracy of 99.26%. and Getting 100% accuracy is near to real, but getting a second opinion about a patient's condition will help further for the doctor to do diagnosis.
B. M. Abed et al.,	[18]	Hybrid classification algorithm based upon Genetic Algorithm (GA) and K-NN algorithm with an accuracy of 99% with Limited numbers of features being selected for only one data set out of two datasets.
C. Shahnaz et al.,	[22]	Different classifiers are like the "Naive Bayes, SVM, Logistic Regression, KNN, Random Forest Neural Network and deep learning" method CNN used to get better accuracy. For deep networks, the convergence time suggestively rises and it is very tough to improve the system performance.
D. Bardou et al.,	[23]	CNN (convolutional neural network) and "handcrafted features plus convolutional neural network and convolutional neural network features plus classifier" configurations achieved binary classification accuracy of 96.15 percent and 98.33 percent, respectively. The multi-class categorization received 83.31 percent and 88.23 percent, respectively. It replaces traditional classifiers with fully-connected layers to train the handmade features of DSIFT and SURF, which aids in improving their performance.
Y. Song et al.,	[26]	Histopathology image classification based on transfer learning. Feature extraction by Fisher Vector (FV), Convolutional Neural Network (CNN) model pertained on ImageNet. Patient and image-level classification to find out begin and malignant. Image-level classification high compared to patient-level classification.
P. J. Sudharshana et al.,	[27]	12 Multiple instance learning (MIL) used. Patient classification rates with 92.1% for magnification factor 40X. MIL for histopathological image segmentation further will lead to better results.

IV. Conclusion

Even now, there remains need for research in the subject of breast cancer, with many particular issues that must be addressed in future study. Papers cited in this review have been achieved using their private as well as a public dataset. Evaluation method and numerical comparisons of the studies various with different datasets with multiple performance metrics. The discussion gave insight into various methods of machine learning. SVM method is good classification accuracy and double thresholding for mammogram images. The hybrid method enhances the accuracy UCI-WBC data set and segmentation for better visualization. The paper discusses the prediction and diagnosis of whether a breast cancer cell is benign or malignant using various data sets such as mammography and histology pictures.

REFERENCES

[1] Omar Ibrahim Obaid, Mazing Abed Mohammed, Mohd Khanapi Abd Ghani, Salama Mostafa" Evaluating the Performance of Machine Learning Techniques in the Classification of Wisconsin Breast Cancer" International Journal of Enginnering & Technology 7(4.36) (2018) 160-166, December 2018.

[2] Wenbin Yue, Zidong Wang, Hongwei Chen, Annette Payne, Xiaohui Liu. "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis". 9 May 2018.

[3] National Breast Cancer Foundation (NBCF), Texas, ―Mammogram ‖ , URL: http://www.nationalbreastcancer.org/diagnostic-mammogram

[4] WHO, "World Health Organization: Breast cancer,"

[5] N. Wu et al., "Breast Cancer Wisconsin (Diagnostic) Data Set | Kaggle," Kaggle, vol. 4, no. November, pp. 1–9, 2019,

[6] B.M.Gayathri, C.P.Sumathi, and T.Santhanam. Breast Cancer Diagnosis Using Machine Learning Algorithms –A Survey, International Journal of Distributed and Parallel Systems (IJDPS) Vol.4, No.3, May 2013

[7] Osman, A.H. An enhanced breast cancer diagnosis scheme based on two-step-SVM technique. Int. J. Adv. Computer. Sci. Appl. 2017, 8, 158–165.

[8] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Classification of histopathological biopsy images using ensemble of deep learning networks," CASCON 2019 Proc. - Conf. Cent. Adv. Stud. Collab. Res. - Proc. 29th Annu. Int. Conf. Comput. Sci. Softw. Eng., pp. 92–99, 2020.

[9] Samir M. Badawy, Alaa A. Hefnawy, Hassan E. Zidan, Mohammed T. GadAllah "Breast Cancer Detection with Mammogram Segmentation: A Qualitative Study", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017

[10] G. Aresta and T. Ara, "BACH: grand challenge on breast cancer histology images," pp. 1–4, 2019.

[11] N. Wu et al., "The NYU Breast Cancer Screening Dataset v1.0," pp. 1–9, 2019

[12] "A curated mammography data set for use in computer-aided detection and diagnosis research _ Enhanced Reader.pdf."

[13] M. L. Huang and T. Y. Lin, "Dataset of breast mammography images with masses,"

[14] Ilias Maglogiannis, E Zafiropoulos "An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers" Applied Intelligence, 2009 – Springer.

[15] A. Marcano-Cedeño, J. Quintanilla-Domínguez, and D. Andina, "WBCD breast cancer database classification applying artificial metaplasticity neural network," Expert Syst. Appl., vol. 38, no. 8, pp. 9573–9579, 2011

[16] P. Kaur, G. Singh, and P. Kaur, "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification,

[17] B. M. Abed et al., "A hybrid classification algorithm approach for breast cancer diagnosis," IEACon 2016 - 2016 IEEE Ind. Electron. Appl. Conf., pp. 269–274, 2017

[18] Chandra PrasetyoUtomo, AanKardiana and Rika Yuliwulandari, "Breast Cancer Diagnosis using Artificial Neural Networks with Extreme Learning Techniques", YARSI University, Jakarta, Indonesia, IJARAI, Vol. 3, No. 7, 2014.

[19] LuqmanMahood Mina and Nor Ashidhi Mat Isa, "A Fully Automated Breast Separation for Mammographic Images", IEEE International Conference on Bio Signal Analysis, Processing and Systems ICBAPS, pp. 37-41, 2015.

[20] Monica Jenefer and Cyrilraj, "An Efficient Image Processing Methods for Mammogram Breast Cancer Detection", Journal of Theoretical and Applied Information Technology, Vol. 69, Nov. 2014.

[21] C. Shahnaz, J. Hossain, S. A. Fattah, S. Ghosh, and A. I. Khan, "Efficient approaches for accuracy improvement of breast cancer classification using wisconsin database," in 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dec. 2017, vol. 2018-Janua, no. December, pp. 792–797

[22] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks," IEEE Access, vol. 6, pp. 24680–24693, 2018,

[23] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks," IEEE Access, vol. 6, pp. 24680–24693, 2018,

[24] Computer-Aided Breast Cancer Detection Using Mammograms: A Review

[25] Shanmugavadivu and Sivakumar "Wavelet Transformation-Based Detection of Masses in Digital Mammograms", International Journal of Research in Engineering and Technology, Vol. 3, No. 2, Feb. 2014.

[26] Y. Song, J. J. Zou, H. Chang, and W. Cai, "ADAPTING FISHER VECTORS FOR HISTOPATHOLOGY IMAGE CLASSIFICATION School of Information Technologies, University of Sydney, Australia School of Computing, Engineering and Mathematics, Western Sydney University, Australia Biological Systems and Engineer," pp. 600–603, 2017.

[27] P. J. Sudharshana et al., "Multiple Instance Learning for Histopathological Breast Cancer Images to cite this version: HAL Id: hal-01965039 Multiple Instance Learning for Histopathological Breast Cancer Images," 2019.

[28] M. Kanchana and P. Varalakshmi, "Breast Cancer Diagnosis Using Wavelet Based Threshold Method", Middle-East Journal of Scientific Research, Vol. 23, No. 6, pp. 1030-1034, 2015.

[29] Lisa Hutchinson, ―Breast cancer: Challenges, controversies, breakthroughs‖, Nature Reviews Clinical Oncology, Vol. (7), pp. 669-670, December 2010. Doi:10.1038/nrclinonc.2010.192. URL: https://

[30] Overview: Breast Cancer—Patient Version‖, National Cancer Institute, National Institutes of Health (NIH), U.S. URL: https://www.cancer.gov/types/breast

[31] Radiologyinfo.org, ―Mammography‖,

URL: https://www.radiologyinfo.org/en/info.cfm?pg=mammo#overview

[32] H.D. Cheng, Juan Shan, Wen Ju, Yanhui Guo, Ling Zhang "Automated breast cancer detection and classification using ultrasound images: A survey", Pattern Recognition 43 (2010) 299 – 317.