

An Effective Classification Algorithm for Breast Cancer using Dyadic Projection



T Yasodha, Arun Balaji G

Abstract—Breast Cancer is a wide spread reason for the deaths of women in the world. Nowadays, CAD systems have become the most increasing interest in its detection. In this paper, a new computer-aided diagnosis method is introduced to help oncologists to classify it as benign or evil breast tumors in ultrasound. In the proposed model, bi clustering is done for feature acquisition and then dyadic transform is applied. Biclustering mining is used as a key to identify the regularity patterns in columns on the working out data, Biclustering mining is utilized as a key. At last, to identify the perfect combinations and put them into a strong classifier, AdaBoost learning is applied. Using a dataset the proposed method is evaluated validated and the results are compared with the results of existing methods. The results of the proposed model showed the best calculation, proving it to be effective in laboratory applications.

Keywords—Biclustering, computer-aided diagnosis, adaboost, feature scaling, ensemble learning.

I. INTRODUCTION

An abnormal extra cell growth in the breast tissue leads to Breast cancer. If proper treatment is not given, then the cancer extends to all parts of the body. Breast cancer is the second most common type of cancer affecting 33% of women in USA next to Skin cancer. Approximately 2,11,240 fresh persistent cases of breast cancer were expected to arise among women in USA. Further, disgracefully 1,690 fresh male cases likely to expect with breast cancer. The vulnerability is prevalent after the age of 40. The highest anomaly (approximately 80%) occurs in women over the age of 50. Besides invasion, 58,590 fresh cases of expressive breast cancer cases were anticipated among women during 2005[11]. From the above facts, it is classified that around 88% of cases as Ductal Carcinoma In-Situ (DCIS). The use of mammography screening for the diagnosis of DCIS cases increases its count.

To identify the cancer at the initial stage and to prevent its occurrence, the Mammography screening method is used. In 2005, the death due to breast cancer is found to be 40,400 women and 460 men, which come around 40,860 deaths. Breast cancer is second among all deaths caused due to cancer in women. Statistically, mortality rates declined significantly during the 90s, with the largest decrease in younger women from all descend[10].

Manuscript received on April 02, 2020.

Revised Manuscript received on April 15, 2020.

Manuscript published on May 30, 2020.

* Correspondence Author

Dr T Yasodha, Head of the Department, Electronics and Communication Engineering, Christian College of Engineering and Technology, Oddanchatram, Tamilnadu, India.

Arun Balaji G, PG Scholar, Department of Electronics and Communication Engineering, Christian College of Engineering and Technology, Oddanchatram, Tamilnadu, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

II. LITERATURE SURVEY

[17] Z. L chen et al. propose the cancer image reconstruction for detection of tumor, realization with only extracted signals from a detected waveform as it is, with no further modifications. [15] Y.

Liu et al. propose a innovative, automatic method for initialization of active contour model designed specifically for US-based imaging models. The method can estimate an initial contour by utilizing the conventional Color Doppler. [7] Qinghua Huang, et al propose a method to identify micro calcifications and embedded mass of cells and also classified them as benign or malignant. Here, Curvelet and texture analysis are used for extracting feature and a PNN classifier for classification.

[4] J. L. Koning, and J. E. Grabowski et al. present a novel scheme for detection from mammographic digital images based on Local Binary Patterns (LBP). The method successfully used LBP based features with a classifier and thresholding levels of readings. The proposed method is tested on a set of images extracted from MIAS and DDSM databases. [13] X. J. Shi, et al et al., utility of this cause has already been shown in an ex vivo setting. [12] T.ungi et al. present a UWB differential equivalent time sampler for breast cancer detection assembled on a PCB.

[1]B.K.Singh et al. investigated the probability of predicting PCR using only the nodal sizes of the first three treatment records. Feature combination for each breast cancer subtype was screened from the real nodal sizes of the first three treatments and the nodal sizes of the next three treatments estimated from those of the first three ones. [2] B. Sahiner et al., describe the testing of 3D computer aided radio logic system for use in research of early detection of breast cancer.

[3] H. D .Cheng et al., proposed the approach of pattern recognition for identification of masses in mammograms. [5] J. Y. Choi et al., proposed a multi classifier system aiming to classify the breast mass using mammogram. [6] M. Costantini, and P. Belli suggested the idea of sonography for identifying breast masses using lexicon system. [14]Y. L. Huang et al., presented the idea of support vector machines for breast tumor classification and the results proved to be effective in comparison with the existing techniques.[16] Y.Su suggested the idea of propagation clusters that would be assisted with computer aided methodology for the classification of breast tumors.

A. Types of breast cancer

Adenocarcinoma is the reason for 90 % of breast cancer, which starts from tissue glands. In wide category, there around 30 different subtypes of adenocarcinoma.DCIS is the least advanced stage of the growth, constituting about 20-25% of the breast cancers.

An Effective Classification Algorithm for Breast Cancer using Dyadic Projection

It forms in the milk ducts, fully. This expands through the duct walls and enters the breast tissue leading to the common form of cancer, known as invasive DCIS [8].

Lobular carcinoma or small cell carcinoma means the cancer formed from the lobes or lobules on both sides. 15-20% of invasive lobular carcinoma starts from the milk secretive glands. Either the carcinomas mentioned above are spread through the walls of the duct or lobe to nearby tissues and can be in situ, or self-contained.

The following comprise less common types of breast cancer:

- **Diffused infiltration:** red or inflammatory breast indicates quick spread.
- **Medullary carcinoma:** cancer initiating from central breast tissue.
- **Mucinous carcinoma:** It is an invasive cancer seen in women post menopause.
- **Paget disease of the nipple:** Starts from the milk ducts and expands to the skin of nipples or areola.
- **Phyllodes tumour :** This Tumours expands to the ducts with a leafy look. It occasionally goes to metastasize.
- **Tubular carcinoma:** It is small and could not be identified by palpation.
- **Sarcomas :** it gets formed in the connective tissue and
 - **Lymphomas:** this gets formed in the lymph tissue and occasionally forms in the breasts.

B. Risk factors

The factors that cause risk of breast cancer are unknown, but studies suggest that oestrogen hormone, produced by the ovaries, is involved. One reason for suspecting oestrogen is the association of breast cancer is certain reproductive changes in a woman's life. It is thought that about 10-15% of breast cancers are connected to heredity. For identifying BRCA1 and BRCA2 genes, molecular testing can be used. These genes are present in low percentage of the population. Mutation of these genes is frequently seen in Jewish ancestry women.

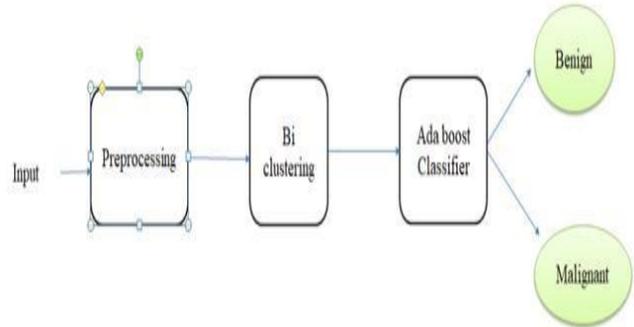
It is said that the greater risk of suffering is seen in women who are the first-degree relatives of the women with breast cancer. Such women with family history of breast cancer advised to have earlier screening tests, compared to women without the family history of cancer.

Age is a risk factor for breast cancer. The death risk is directly proportional to age. The death risk for 50-year-olds is 1/36, 60 years old is 1/25 and 70 years old is 1/23. Atypical hyperplasia, family history of breast cancer, increase of breast density, a late menstrual, abnormal weight increase after menopause, use of oral contraceptives or postmenopausal oestrogen and progesterone, women who have never gave birth to children or delivered first child after the age of 30, or who took oral contraceptives or consume alcohol are the categories of greater risk for breast cancer.

Globally, the rate of occurrence of breast cancer is related to diet intake, specifically fat intake. The impact of diet intake affecting breast cancer has not been conformed so far. Regular physical activity and healthy body weight maintenance are less likely case breast cancer. The data collected shows that tamoxifen and raloxifene decreases the risk of breast cancer.

III. EXISTING SYSTEM

[9] R.M Jales et al., described a novel framework to attain the aim of increasing the learning perfection with very minimum tagging for deep actively learning, classification and implementation of breast cancer histopathological images. This approach includes physical annotation of important unlabelled data, and is incorporated in the work set. Then the model is repeatedly reorganized for training with by increasing work dataset. Deep active learning framework uses two selection strategies namely a confidence-boosting and entropy based strategies. Existing approach is evaluated utilizing accessible breast cancer histopathological digital sample data from the open source. Every patches of the image are categorized as malignant or



benign.

Fig.1: Block Diagram of the Existing System

The figure shows the existing system for breast cancer classification. Here there is no transforms such as FFT, DCT are used. The novel concept this paper proposes includes a transform to frequency domain so as to easily process the image with more accuracy. The active learning concept is a basic and machine learning algorithm which provides better performance taking less work dataset given to it, when it is permitted to select data from where it learns.

Consider a sample data B includes N_B images from categories of Y . Sample data B is divided into a work sample pool B_{pool} containing N_{pool} images and testing sample B_{test} containing N_{test} images. For further subdivisions, B_{pool} is split up into the pre-training workset $B_{pre-train}$ containing $N_{pre-train}$ images, labelled data set B_l containing N_l images, and remaining unlabelled data set B_u with N_u images ($N_{pre-train} + N_l + N_u = N_{pool}$). $B_{pre-train}$ includes arbitrarily picked labeled data from B_{pool} , that are not connected with B_u and B_l and of fixed size.

Prior to initiating, let's assume that no samples are labeled, that is, $B_u = B_{pool} - B_{pre-train}$, $B_l = \emptyset$. Further to identify the precious data samples from B_u to B_l , N_l and N_u vary during the active learning process and are then provided into a fine tuning model. It is observed that the labeled sample data B_l need not be the work set since query strategies may include few special data that are not manually assigned labels. To differentiate between B_l and the work set, B_{train} is used as the work set. For Finishing the categorizing task, A deep network M is essential particularly CNN network.

Hence, in an active query t , we choose the best possible data from B_u using exact query strategy that directs Which data in B_u are supposed to be physically labeled, and which labeled data is supposed to be dropped in to B_{train} to take part in the working process of model M_t . The fine tuning of the model is recorded as $M_t, t \in (0, 1, ,T_a)$ during every query and the maximum of queries is placed at T_a .

IV. PROPOSED SYSTEM

In the proposed system of classification, acquiring features of the image is executed by the opinion of the doctors and Breast Imaging Reporting and Data System (BI-RADS) based subject-participated feature scoring scheme . For digging deeper to get data, to explore the column patterns consistently on the work set data, Biclustering mining is used. The regularly appearing pattern in tumors having the same label is considered as a proper diagnostic protocol . Consequently, the diagnostic procedure is used to construct components that classify data of the Adaboost algorithm with innovative rules of mixed strategy that can resolve the Problem of Classification in Different Feature Spaces (PC-DFS) . At last, the AdaBoost learning is observed to explore efficient combinations and incorporate them into a strong classifier. Subject-participated feature scoring scheme performed the feature extraction. To find the column consistency patterns on the workout data, Biclustering mining is utilized as a useful tool. Adaboost algorithm via new rules of mixed strategy that resolves the PC-DFS.

In par with the medical findings and BUS images, the benign tumor appears oval in shape, parallel in lesion orientation, that has embedded, distinct border, and speculated or bony free margin; and malignant tumor is normally uneven in shape, alignment being non-parallel, un constrained or imprecise border with speculated or angular margin. Malignant tumor normally has hypo echoic pattern, vascularity, micro calcification in mass and duct leeway. However, Benign tumor have these patterns in rare or slight degree.

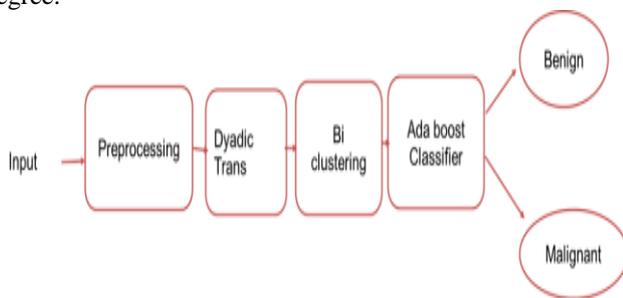


Fig.2: Block Diagram of Proposed System

A. Dyadic transformation

Dyadic transformation may be expressed as the piecewise linear function,

$$f(x) = \begin{cases} 2x & 0 \leq x < 0.5 \\ 2x - 1 & 0.5 \leq x < 1. \end{cases} \quad (1)$$

The map of bit shift comes into play because, if the value of an repetition is binary, the next repetition is obtained by shifting the binary point one bit to the right, and if the bit is

shifted to the left of the new binary point is a "one", replaced with a zero.

B. Biclustering algorithm

A bicluster can be considered as a sub matrix N obtained from matrix M . From the doctor’s point of view, bicluster some reveal some type of diagnostic regulation as a local lucid pattern. In breast tumor analysis, tumors in the bicluster have the same influence to the classification of scratch type that are sharing the same cut and subdivision. A recurrently looked pattern shows that it is one of the common medical directories of tumors and can be observed as a major diagnostic regulation. Here, in a bicluster, every feature should obtain the equal or comparable values in the subset of breast cancer cases. Therefore, this study focuses on mining the bi clusters with constant columns. In order to obtain a strong diagnostic rule, a lawful bicluster should contain at least 5 rows in the study. To calculate the quality of bi cluster, Mean- Square-Residue Score (MSRS) is used. In a bi cluster with R rows and C columns, the element u_{ij} represent the value at the i^{th} row and j^{th} column .

Subsequently in bi clustering mining, the diagnostic patterns or bi clusters are explored and transformed into diagnostic rules as given below. To find the category and reliability of a rule, a confidence-based metric is proposed in the process of adaptation of a bicluster in to the diagnostic rule. In this study, diagnostic rules for benign (B) and malignant (M) are different. Assume R_{benign} and $R_{malignant}$ to denote the number of rows with benign and malignant label, respectively, and R_{bic} denotes the total number of rows of bicluster. Sureness of benign category confidence (B) and malignant category confidence (M) can be computed based on the final diagnostic result shown below.

$$\begin{cases} confidence(B) = R_{benign} / R_{bic} \\ confidence(M) = R_{malignant} / R_{bic} \end{cases} \quad (2)$$

The confidence C of a bicluster is,

$$C = \max \{ confidence (B), confidence (M) \} \quad (3)$$

Hence, the consistent leeway of existence is indicated by the buoyancy of a bicluster for benign or malignant [8]. The category of the larger confidence finds the group of a bicluster. If the confidence C of the bi cluster is larger than the a predefined threshold T_c ., then it is selected as a diagnostic rule.

C. Adaboost

AdaBoost, short for Adaptive Boosting, is a machine learning algorithm using meta data. This can be combined with other types of learning algorithms to increase the performance. The output of the other learning algorithms ('weak learners') is combined and made as a weighted sum representing the final output of the boosted classifier. AdaBoost is adaptive in a logic that successive weak algorithms are modified in favor of those instances wrongly classified by previous classifiers. AdaBoost is fragile to noisy data and ripples. In certain examples, it can be less vulnerable to the incomplete problems than any other learning algorithms.

An Effective Classification Algorithm for Breast Cancer using Dyadic Projection

The learners are separately weak, but the performance of each are to some extent better than arbitrarily guessing the output, the final model can be established closely to a stronger learner.

Each learning algorithm inclines to handle some problem types better than that of others, and has many diverse parameters and conformations to regulate before it attains optimum performance over a dataset, AdaBoost is often referred to as the best classifier available. On combination used with decision tree learning, information congregated at every stage of the algorithm about the relative 'hardness' of each training trial is given to the growing tree algorithm such that the later trees tend to emphasis on examples that are hard to classify.

Machine learning problems endlessly suffer from the dimensionality problem; each sample may consist of a enormous number of potential features (for example, there can be millions of Haar features, in par with the Viola-Jones object detection framework, in a pixel image window), and evaluating every feature can reduce not only the speed of classifier training and execution, but in fact reduce predictive power. Unlike neural networks and SVMs, the AdaBoost training process chooses only those features known for improving the predictivity of the model, lessens dimensionality and potentially improves implementation time as inappropriate features are ignored.

V. RESULT

The breast images are provided as input to MATLAB tool and processed to get output with more accuracy than in the previously used methods, which process the image directly without applying any kind of transform. The below table I depicts that the proposed method outperforms the existing method in accuracy, sensitivity and specificity.

Table I: Comparison between the Proposed and Existing system

Parameter	With Dyadic Transform (Proposed)	Without Dyadic Transform (Existing)
Accuracy	98.1	96.3
Sensitivity	93.4	91.9
Specificity	97.6	96.5

The stage thresholds for different types of cancer are identified as below,

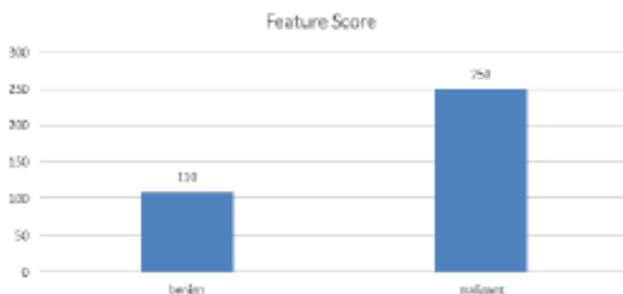


Fig.3: Cancer Stage Thresholds

This work aims to accurately classify the type of breast cancer present in women, in order to improve the survival rate by improving detection; there by the detection of the tumor cells before they could grow beyond the breast. So, the accuracy and the precision factors depend purely on:

The Classifier used.

Clustering method used.

VI. CONCLUSION

This work suggests a fresh human-in-the-loop CAD system for categorizing breast tumors with human judgment on the BI-RADS reference book-based features. It is a pioneering attempt to implement operator-based feature scoring scheme slightly measuring the image de-noising, image segmentation and feature extraction in traditional CAD systems. Separately, traditional dealings remain a challenging crisis in the computer vision exclusively in ultrasound images and image processing field, that unquestionably affect the output of final classification. In distinction, the doctor's experience is used in feature extraction, i.e easily acceptable by doctor's real-time applications and advances the validity of our proposed work. In order to extract diagnostic rules, bi clustering mining is used behind the huge data of breast tumor. This could help to explore possibly related medical indicators correlated with benign and malignant tumors.

ACKNOWLEDGMENT

Primarily, we submit our gratitude & sincere thanks to our college principal Dr. B JUSTUS RABI M.E., Ph.D., MBA, MISTE, MIE, LLB for his constant motivation. We would not be able to do justice to this project work, if we do not express our heartiest thanks to the Dept. of Electronics and Communication Engineering, Christian College of Engineering and Technology, Oddanchatram, Tamilnadu, India.

REFERENCES

1. B. K. Singh and A. S. Thoke, "Adaptive gradient descent backpropagation for classification of breast tumors in ultrasound imaging," Proceedings of the International Conference on Information and Communication Technologies, 2015, ICICT 2014, vol. 46, pp. 1601-1609.
2. B. Sahiner, et al., "Malignant and benign breast masses on 3D US volumetric images: Effect of computer-aided diagnosis on radiologist accuracy," Radiology, , Mar 2007, vol. 242, pp. 716-724.
3. H. D. Cheng, et al., "Approaches for automated detection and classification of masses in mammograms," Pattern Recognition, Apr 2006, vol. 39, pp. 646-668.
4. J. L. Koning, and J. E. Grabowski, "Breast Imaging-Reporting and Data System (BIRADS) Classification in 51 Excised Palpable Pediatric Breast Masses," Journal of the American College of Surgeons, Sep 2014, vol. 219, pp. S78-S79.
5. J. Y. Choi, "A generalized multiple classifier system for improving computer-aided classification of breast masses in mammography," Biomedical Engineering Letters, 2015, vol. 5, pp. 251-262.
6. M. Costantini, and P. Belli, "Characterization of solid breast masses - Use of the sonographic breast imaging reporting and data system lexicon," Journal of Ultrasound in Medicine, May 2006, vol. 25, pp. 649-659.
7. Qinghua Huang, et al., "On Combining Biclustering Mining and AdaBoost for Breast Tumor Classification, 2018"
8. R. L. Siegel, et al., "Cancer Statistics, 2015," CA: a Cancer Journal for Clinicians, Jan-Feb 2015, vol. 65, pp. 5-29.

9. R. M. Jales et al., "Simple rules for ultrasonographic sub-categorization of BI-RADS (R)-US 4 breast masses," European Journal of Radiology, Aug 2013, vol. 82, pp. 1231-1235.
10. S. Ackerman, "ACR BI-RADS for Breast Ultrasound," Ultrasound in Medicine & Biology, 2015, vol. 41, pp. S95-S95.
11. T. F. Yin and F. H. Ali, "A Robust and Artifact Resistant Algorithm of Ultrawideband Imaging System for Breast Cancer Detection," IEEE Transactions on Biomedical Engineering, Jun 2015, vol. 62, pp. 1514-1525.
12. T. Ungi, et al., "Navigated Breast Tumor Excision Using Electromagnetically Tracked Ultrasound and Surgical Instruments," IEEE Transactions on Biomedical Engineering, Mar 2016, vol. 63, pp. 600-606.
13. X. J. Shi, et al., "Detection and classification of masses in breast ultrasound images," Digital Signal Processing, May 2010, vol. 20, pp. 824-836.
14. Y. L. Huang, et al., "Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines," Neural Computing & Applications, Apr 2006, vol. 15, pp. 164-169.
15. Y. Liu, et al., "Computer Aided Diagnosis System for Breast Cancer Based on Color Doppler Flow Imaging," Journal of Medical Systems, Dec 2012, vol. 36, pp. 3975-3982.
16. Y. Su and Y. Wang, "Computer-Aided Classification of Breast Tumors Using the Affinity Propagation Clustering," in 2010 4th International Conference on IEEE in Bioinformatics and Biomedical Engineering, 2010, pp. 1-4.
17. Z. L. Chen and H. Strange, "Topological Modeling and Classification of Mammographic Microcalcification Clusters," IEEE Transactions on Biomedical Engineering, Apr 2015, vol. 62, pp. 1203-1214.

AUTHORS PROFILE



Dr. T. Yasodha, is currently working as Associate Professor and Head, of the Department of Electronics and Communication Engineering, in Christian college of engineering and technology, Oddanchatram. She has put in 20 years of teaching experience in Polytechnic College and 12 years of experience in Engineering. Her area of specialization is Communication in VLSI design.



Mr. G. Arun Balaji, completed B.E degree in ECE from Christian college of Engineering and Technology, Oddanchatram, Tamilnadu, India in 2018. At present, he is doing final year M.E applied electronics in the same institution.