International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Secondary Screening Algorithm for Breast Cancer Detection Using Matlab

Umar Alqasemi[1*], Jawad Fikri[2], Fadi Tarmeen[3]

Submitted: 06/05/2024 **Revised**: 19/06/2024 **Accepted**: 26/06/2024

Abstract Physicians and radiologists utilize computer-aided detection (CAD) systems to detect breast cancer. In this study, through the use of CAD we are going to detect abnormal tumors in X-Ray images using statistical and histogram-based features along with 9 different SVM and KNN classifiers. DDSM from the University of South Florida is the source of the digital X-Ray images. The specificity, sensitivity, and accuracy is compared with previous similar studies.

Keywords: *Breast Cancer, Computer Aided Detection, CAD, Support Vector Machine, SVM, Gray-level Co-occurance Matrix, GLCM, k-nearest Neighbor, KNN.*

1. Introduction

The term breast cancer refers to a type of fatal disease that mainly affects female breast. It is caused by uncontrolled cell growth in the breast that can disperse throughout the body. Cells originating from cancer usually appear as a lump that can be found on an X-ray or by touch. The majority of breast cancer cases occur in women, but men are at risk as well. Typically, lumps in the breast are benign and not actually cancerous. The lumps which are not cancerous are abnormal growths that do not usually extend beyond the breast. In some instances, however, breast lumps may increase the possibility of developing cancer.

In the course of a biopsy, cancer is tested for HER2, estrogen receptors, and progesterone receptors. It is also necessary to carefully examine the tumor cells in the laboratory in order to determine the grade of the tumor. Identifying specific proteins within the tumor and determining its grade are crucial to determining treatment options. An individual's lymphatic system consists of a network of vessels connecting lymph nodes throughout the body. Known as lymphatic vessels, they are responsible for removing lymph fluid from the breast. Breast cancer cells may diffuse through lymph vessels and multiply within lymph nodes during the development of the disease.

Upon diffusion into lymph nodes, it is very likely that the cancerous cells may have passed through the lymphatic system and metastasized. In general, the larger the number of lymph nodes containing breast cancer cells, the greater the likelihood that other organs will be affected by the disease. As a result, accurate detection of cancer in one or more lymph nodes can have significant implications for treatment plan. One or more lymph nodes are usually removed during surgery in order to determine whether the cancer has spread. It is possible, however, for women without cancer cells in their lymph nodes to develop metastasis later in life [1] – [3].

In order to detect breast cancer, three methods may be used: film or digital mammography, ultrasound, and magnetic resonance imaging (MRI). All three methods have their pros and cons. A number of companies have created digital mammography devices in order to improve the quality of x-ray mammography. Unlike older film mammography machines, digital mammography machines capture the x-ray image digitally. A computer screen is used to display and manipulate the digitized image generated by an array of detectors. As a result, in theoretical terms, one could detect tumors that are obscured by dense breast tissue commonly found in young women more easily through computer software.

Having the capability of enlarging or adjusting the contrast of questionable areas without having to expose the patient to a new set of X-rays may assist in the cancer detection. Additionally, this technology could improve the screening of breast cancer through the ability to electronically store, retrieve, and transmit mammogram images. According to data available to date, digital mammography has not been found to be more effective or accurate in reducing breast cancer deaths.

Mammograms are can usually be followed by an ultrasound examination to determine how solid is the

[1] Department of Electrical and Computer Engineering, King Abdulaziz University, Saudi Arabia
ualqasemi@kau.edu.sa
[2] Department of Electrical and Computer Engineering, King Abdulaziz University, Saudi Arabia
jfksapphire@gmail.com
[3] Department of Electrical and Computer Engineering, King Abdulaziz University, Saudi Arabia
fadi8806@gmail.com

suspicious area. The ultrasound imaging device emits acoustic waves, then a computer uses the echoes generated when these waves bounce between tissues to create a sonogram. A fluid-filled cyst differs from a solid mass in that radiologists are able to detect them using ultrasound.

Women with dense breasts may benefit from ultrasound imaging of the breast to determine whether they have any lumps that they can feel (palpable lesions) but cannot see on a mammogram. Ultrasound technology had been shown to be very effective at diagnosing non-malignant abnormalities in women with palpable lesions and could have avoided over half of biopsies. In other studies, ultrasound has been shown to differentiate between benign and malignant non-palpable solid lesions. The utility of ultrasound screening in combination with mammography needs further investigation.

Despite its utility as a complement to mammography, ultrasound has limitations when used alone. Ultrasounds are not always able to detect small tumors up to 5 mm or about 14 inch in diameter or abnormalities related to certain types of breast cancer.

Since MRI was FDA-approved in 1985 for use in body imaging, doctors have been using it for scanning internal body structures and suspicious diseases. MRI is considered safer non-invasive imaging modality. A specially designed MRI system for breast imaging, which has been approved by the FDA, may prove useful as a method of detection when used in conjunction with mammography, particularly with dense breast tissue. Research suggests that although MRI can detect tissue changes indicative of cancer, it cannot always differentiate malignancies from harmless tissue changes.

MRI can also be used to detect breast cancer recurrence after lumpectomy, as, unlike mammography, MRI is not restricted by scarring associated with surgery. Breast implants and dense breasts can interfere with the x-ray interpretation of a mammogram and therefore make MRIs more effective. Therefore, MRI may provide useful information to screen for breast cancer among high-risk young women. These women tend to have dense breasts [4].

2. Literature Review

It is well known that previous studies have shown that CAD system can be used for breast cancer screening as secondary method to help radiologist in faster screening the regions of interest (ROI).

According to Arai et al. [5] , the data was divided into two: training and testing. There was a 74% and 26% data proportion. The following features have been utilized by the author: Wavelet, Variance, STD, CV, Mean, Centroid, Max, and 7 Hu. Each detail is decomposed using Wavelet decomposition, horizontally, diagonally, and vertically, respectively. The SVM classifiers resulted in 91.43 percent sensitivity and 90 percent specificity. It is more computationally expensive to train the classifier when using features obtained after image transformation.

Soulami et al. [6] identified abnormal areas in digital mammograms using Mini-MIAS as the database, based solely on the dense breast categories, and categorized the areas as abnormal (benign and malignant). In order to identify suspicious structures, an Electro-Magnetism Like (EML) optimization and edge-based detection were used. The accuracy of SVM classifier performed by this method is 86.36. It is computationally expensive to use these features, yet the accuracy attained is low.

RBFNN has higher sensitivity and specificity than computer-assisted diagnosis (CADx), as well as sensitivity and specificity in NMR. Pratiwi et al. [7] have found that RBFNN is more accurate in CADe (detection) than CADx (diagnosis). The author recommends that texture features, such as wavelets and curvelets, be used to enhance breast cancer classification accuracy.

Laws' Texture Energy Measure (LAWS) features were used to classify mammogram images in Setiawan et al. [8]. The experiment showed LAWS to be more accurate than GLCM at classifying mammograms. Based on CADx, the accuracy of each degree is less than 55% using the GLCM feature. The author suggests that improving the ANN model by changing the architecture and number of hidden layers can be done in this study.

Using Otsu's method, Saad et al. [9] developed an algorithm to detect Microcalcifications (MCs) and to diagnose breast cancer automatically. Features included CII, EPI, and PSNR, the enhancement algorithm significantly improved contrast between MCs and backgrounds, which ultimately improved MC detection. Adaptive boosting (Adaboost) is also shown in the algorithm to be more sensitive and accurate in detecting single and clustered MCs in comparison to ANN [9]. DDSM, MIAS, and local databases have been tested, and the algorithm has demonstrated high levels of accuracy (98.68%) and sensitivity (80.15%).

Based on Local Binary Patterns (LBP) features, Pavel, et al. [10] proposed a method for detecting breast cancer. A set of data compiled from MIAS and DDSM was used to evaluate the proposed method. A SVM classifier was used to evaluate the method and the accuracy was approximately 84%. In this study, LBP features were the only ones analyzed, which produced a highly accurate classification. Specifying the ROI in this study may improve the performance of the classifier.

Yahia Osman paper [11], which this study is based on, used SVM and KNN classifiers on 32x32 ROI and MIAS database. The method showed a high accuracy of 94%, sensitivity 92%, and specificity of 100%. Our study used DDSM database instead with 64x64 ROI size and showed similar best results using similar features and classifiers.

3. Database

DDSM database offers an image analysis resource for mammography researchers. MIT, Sandia Nat. Lab. and the

Univ. of S. Florida Comp. Sci. and Eng. Dept. collaborated on this project. There are about 2,500 studies in the database. Two images are included for each breast, along with patient data, and image info. They also provide software used to access the mammograms and truth images. [12]

4. Methodology

Preprocessing: On every image, the radiologist crops a 64x64 pixel area around the suspected tumor or lump. It is done in order to prevent the computational process from becoming overburdened, thereby allowing it to focus on the Region of Interest.

Features Extraction: Our initial analysis consisted of the computation of 94 features, starting with 14 statistical parametric features and 64 Histogram features (16 GLCM features). Initially, a T-test (significant p-value of 5%) is calculated to determine the viability of features along with the final classification performance metrics, such as accuracy, sensitivity, and specificity. As a result of trial and error, the final most significant statistical features that contributed to the success of this study included mode, median, mean, and quantile (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9), those proved the greatest significance in the t-test along with the highest classification accuracy.

Classifiers: In order to achieve this task we used Support Vector Machine (SVM) and the KNN (K-Nearest Neighbor) classifiers. Several kernel-functions were used for SVM, these include linear, polynomial, and radial basis functions. Also, different neighbor number were used with KNN. Eventually, we showed the results of the best of these classifiers.

5. Results

There are several T-Test results that are useful (P-Value 0.05) = 61, as shown by the final test results. Essentially, all of the first-order statistical features used in this study can be considered useful. As a result, the final result can be summarized as follows in Table 1 in comparison with previous studies:

6. Conclusion

In conclusion, we tried Yahia Osman’s algorithm on a different database of mammographic images, similar results is produced though the region of interest (ROI) size is bigger (64x64) pixels. Best results were obtained from SVM classifier with linear kernel. This shows that first-order statistical features along with histogram-based and GLCM-based features perform well in screening of breast tumors when used with linear SVM classifier. Future studies can include microcalcifications (MCs), note that, MCs were excluded in this study. Also, other classifiers can be tried and compared, such as ANN and RBFNN and CNN.

Table 1 Previous Studies Comparison

Author	Features Used	Features Elimination Technique	Classifiers
Arai	Wavelet, Variance, STD, CV, Mean, Centroid, Max, and 7 Hu	N/A	SVM
Khaoula	FIS and Zernike Moments	N/A	SVM
Pratiwi	GLCM (ASM, Correlation, Sum EnEopy, and Sum Variance)	T-test	Back-PNN and RBFNN
Setiawan	Laws’ texture, energy measures, and GLCM	T-test for GLCM only	ANN
Saad	GLCM, LAWS, Kurtosis, and Skewness	N/A	ANN and Adaptive boosting
Pavel	LBP	N/A	SvM
Yahia Osman	1st order statistics, histogram, and GLCM	T-Test	SVM and KNN
Author and date	**Accuracy**	**Sensitivity**	**Specificity**
Arai 2013	84.44%	90.00%	91.43%
Soulami 2017	86.36%	81.81%	90.9%
Pratiwi 2015	92.1%	97.22%	91.49%

Setiawan 2015	93.90%	91%	100%
Saad 2016	97.92%	64.33%	74.16%
Pavel 2016	84%	N/A	N/A
Yahia Osman 2020	94%	92%	100%
Our Study	95%	92.5%	100%

References

[1] Henry NL, Shah PD, Haider I, Freer PE, Jagsi R, Sabel MS. Chapter 88: Cancer of the Breast. In: Niederhuber JE, Armitage JO, Doroshow JH, Kastan MB, Tepper JE, eds. Abeloff's Clinical Oncology. 6th ed. Philadelphia, Pa: Elsevier; 2020.

[2] Jagsi R, King TA, Lehman C, Morrow M, Harris JR, Burstein HJ. Chapter 79: Malignant Tumors of the Breast. In: DeVita VT, Lawrence TS, Lawrence TS, Rosenberg SA, eds. DeVita, Hellman, and Rosenberg's Cancer: Principles and Practice of Oncology. 11th ed. Philadelphia, Pa: Lippincott Williams & Wilkins; 2019.

[3] National Cancer Institute. Physician Data Query (PDQ). Breast Cancer Treatment – Patient Version. 2021. Accessed at https://www.cancer.gov/types/breast/patient/breast-treatment-pdq on June 24, 2021.

[4] Patlak M, Nass SJ, Henderson IC, et al., "Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer: A Non-Technical Summary." Institute of Medicine (US) and National Research Council (US) Committee on the Early Detection of Breast Cancer; Washington (DC): National Academies Press (US); 2001

[5] Kohei Arai, Indra Nugraha Abdullah, Hiroshi Okumura, and Rie Kawakami, "Improvement of Automated Detection Method for Clustered Microcalcification Based on Wavelet Transformation and Support Vector Machine," International Journal of Advanced Research in Artificial Intelligence, vol. 2, no. 4, pp. 23-28, 2013.

[6] K. B. Soulami, M. N. Saidi and A. Tamtaoui, "A CAD system for the detection and classification of abnormalities in dense mammograms using electromagnetism-like optimization algorithm," 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Fez, Morocco, 2017, pp. 1-8, doi: 10.1109/ATSIP.2017.8075533.

[7] Mellisa Pratiwi, Alexander, Jeklin Harefa, Sakka Nanda, "Mammograms Classification Using Gray-level Co-occurrence Matrix and Radial Basis Function Neural Network", International Conference on Computer Science and Computational Intelligence (ICCSCI 2015), Procedia Computer Science 59 (2015), 83 – 91.

[8] Arden Sagiterry Setiawan, Elysia, Julian Wesley, Yudy Purnama, "Mammogram Classification using Law's Texture Energy Measure and Neural Networks," International Conference on Computer Science and Computational Intelligence (ICCSCI 2015), Procedia Computer Science 59 (2015) 92 – 97.

[9] A. K. Q. K. Ghada Saad, "ANN and Adaboost application for automatic detection of microcalcifications in breast cancer," The Egyptian Journal of Radiology and Nuclear Medicine, vol. 47, pp. 1803-1814, 2016.

[10] L. L. Pavel Kra'1, "LBP Features for Breast Cancer Detection," in International Conference on Image Processing, Arizona, 2016.

[11] Yahia Osman, Umar Alqasemi, "Breast Cancer Computer-Aided Detection System based on Simple Statistical Features and SVM Classification," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 1, Jan. 2020.

[12] http://www.eng.usf.edu/cvprg/mammography/database.html