CUCKOO SEARCH: AN OPTIMIZED WAY FOR MAMMOGRAM FEATURE SELECTION

Deepika K. Nagthane[1], Dr. A.M.Rajurkar[2]
[1]Research Scholar at MGM's College of Engg.
[2]Prof. & Head in Dept. Comp. Sci. & Engg, Nanded, Maharashtra India

Abstract
One of the main reasons for increase in mortality rate of woman is breast cancer. Accurate early detection of breast cancer seems to be the only solution for the diagnosis. In the field of breast cancer research, many new computer aided diagnosis systems have been developed to reduce the diagnostic test false positives because of the subtle appearance of the breast cancer tissues. This paper investigates a new approach for the breast cancer classification system using mammogram images. The proposed method uses Cuckoo Search based feature selection algorithm to select optimal feature split points. Classification is based on decision tree classifier based on association rule agreements. The experimentation of proposed method on MIAS database resulted in classification of normal and abnormal cancerous mammogram The result proves the efficacy of the proposed method in classifying task.

Index Terms: Breast cancer, digital mammography, Cuckoo search, Decision tree classifier, Association rule

I. INTRODUCTION

In the modern hi-tech era, the innovations made in the field of medical diagnosis are astonishing. New technologies have been incorporated with medical tools for enhancing the diagnosis performance. Nonetheless, mortality rate of women because of the breast cancer is alarmingly increasing in the recent years [12-13].Breast cancer is considered to be the second leading cause for cancer death in women [3]. Accurate early detection of the breast tissue abnormalities is an effective way to reduce the high mortality rate in women. Since the early detection is concerned with the matter of life and death, numerous screening programmes have been initiated in the last few years [14]. Among them, mammography appears to be the apt screening practise. In mammography screening, X-ray images of the breast are taken and examined by the renowned cancer specialist doctors. The significant advantage of the mammography screening is detection of breast on asymptotic women [15]. And so, mammography screening is well-thought-out to be the best among the available and gold standard for breast cancer screening. In a survey, it have been found that nationwide mammogram screening programs reduced the breast cancer mortality of about 5% in developed countries [16]. Generally, masses and calcifications are the two abnormalities present in the mammogram images. Mammogram interpretation is a repetitive task which requires maximum attention for avoidance of misinterpretation [9]. The efficiency of the mammography screening varies because of different varying factors such as breast density, quality of mammogram, knowledge of radiologist etc. [17]. Double reading is an effective way to increase the cancer detection rate [17]. A valuable and lifesaving automated tool for breast cancer screening using mammogram image is obligatory. Hence in this paper an attempt is made to enhance the classification accuracy of mammogram.

This paper is organized as follows. Following the immense introduction about the breast cancer and diagnosis system, literature of the existing works related to the breast cancer diagnosis system using mammogram images is presented in section 2 highlighting the benefits and shortcomings. In section 3, detailed description about the proposed methodology is discussed. Section 4 discusses the experiments and results achieved. Finally, section 6 concludes the paper.

II. LITERATURE REVIEW

U. Raghavendra et al [1] have demonstrated an automated classification approach for the breast cancer diagnosis using the digitized mammogram images. Gabor wavelet transform method for the feature extraction and Locality Sensitive Discriminant Analysis method for feature data reductionis used in their approach. Subsequently, the reduced features were ranked based on their F-values and fed into the hybrid classifier for classification. The propose method had the ability to provide cross verification to manual diagnosis. But computational complexity of the classifier selection makes it unfeasible for real time applications. A diagnosis technique using association rule mining was presented by Aswini Kumar Mohanty et. al. [2] classification of mammogram images containing normal and cancerous breast tissues in three processing steps is done. In the first step, ROI was extracted, followed by feature extraction in the second step. In the third step, using the association rule mining, new rules were generated classifying the cancerous breast tissues. The efficiency of this method was found to be better. However, variation in the database affects its performance. A breast cancer classification method for diagnosis using data mining approach was proposed by Joana Diz et al. [3]. The mammogram features were extraced from the images and classified using data mining algorithm. The classified features were fed into hybridised classifier comprising different classifiers. Sami Dhabi et al. [6] have used curvelet transform for the feature extraction. After extraction, the best features for classification were selected based on the t-test ranking technique. Finally, K-nearest neighbour classifier was used classify and distinguish the normal and abnormal mammogram. The experimentation of their approach resulted in better accuracy but some information loss in curvelet based feature extraction degrades the performance. Chisako Muramatsu et al. [7] have proposed a new texture feature for the breast cancer diagnosis. The feature developed was ROI-based feature, namely, radial local ternary patterns (RLTP), which takes into account the direction of edge patterns with respect to the centre of masses for classification of ROIs for benign and malignant masses. Breast cacner diagnosis method using shape based features was demonstrated by Subrata Kar and D. Dutta Majumder et al. [8]. In this method, neuro fuzzy classifier was used to classify the breast tissues containing masses and microcalcification. The classification rate seems appealing, but the computational complexity seems knotty in this approach.

Kohei Nishimura et al. [9] investigated on a breast cancer diagnosis method by constructing Multidimensional scaling (MDS) similarity map. The similarity map was created by using similarity ratings. Using the similarity mapping, the discrimination of benign and malignant tumor tissues was performed in this method. One of the limitations in the similarity map was that, only a small number of cases in each pathologic type were used for the experiment. Mohamed Meselhy Eltoukhy et al. [10] have presented a method for breast cancer diagnosis in digital mammogram using SVM classifier. They have obtained the classification based on the following procedure: Initially, a matrix was constructed based on wavelet and curvelet energy of image, followed by feature extraction using statistical t test method. Then feature reduction was performed using a dynamic threshold value. Finally, SVM was used to classify the normal and abnormal breast cancer tissues and to distinguish between benign and malignant tumors. The experimentation of their method resulted in better classification accuracy.

III MAIN METHODOLOGY

The steps involved in the proposed methodology are as follows:

Step 1- Pre-process the mammogram.

Step2- Extract the features form mammogram.

Step3- Select relevant features form mammogram using cuckoo search algorithm.

Step4- Generate association rules form the extracted features.

Step5-Classifiy the mammogram as normal or cancerous.

Algorithm 1. System Workflow

As given in the algorithm above, mammogram images form MIAS [28] database are preprocessed by cropping and resizing. Pre-processing improves the accuracy and speed of the diagnosis process. In second step mammogram features are extracted for classification. Cuckoo search algorithm is used to select the relevant features from the selected features and association rules are generated which are further used for classification purpose.

A. Pre-processing

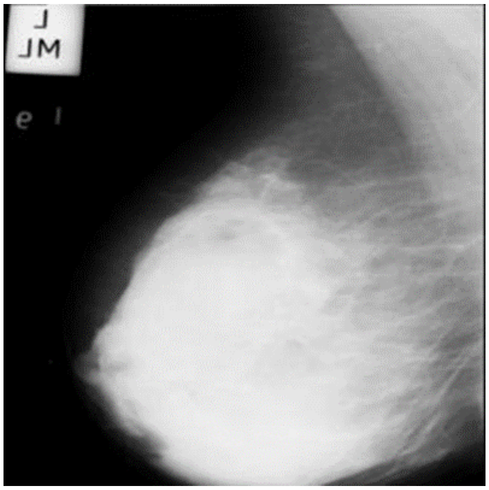


Figure 1. Mammogram before cropping and resizing

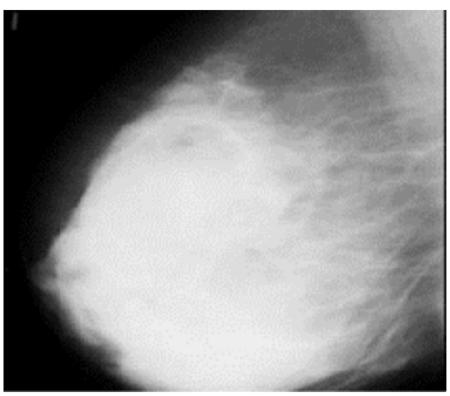


Figure 2. Mammogram after Pre-processing

A typical mammogram contains some labels and patient textual data which need to be removed for further processing.

B. Feature Extraction

Feature extraction is the crucial step in the proposed methodology. The nature of the different classes of the mammogram images is observed only based on the extracted features. Here, the texture features along with wavelet and PCA features of the images are extracted. The proper diagnosis procedure needs discriminative features which provide better classification. Moreover, the performance of the classifier depends on the choice of the feature extracted and the feature selected. Figure 3 depicts the feature extraction procedure in proposed diagnosis methodology.

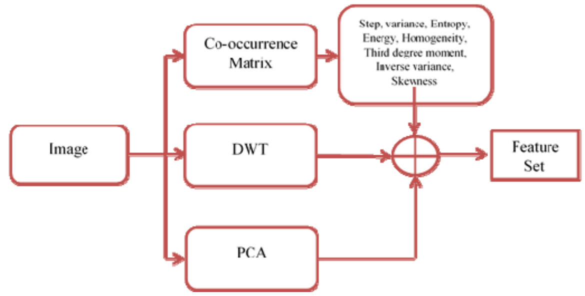


Figure 3. Feature Extraction

The textural features are considered significant because the difference in the density of the breast tissue is captured in the mammogram in the form of texture variations. Furthermore, there is no need of image segmentation for feature extraction. The texture features are extracted from the images from the co-occurrence matrixes [24]. The index of the co-occurrence matrices corresponds to the pixel value of the grey level image. After the co-occurrence matrix creation, the corresponding feature descriptors are calculated.

The wavelet transform is one which provides better representation of the image which makes it apt for analysis of the image. The wavelet feature has the advantage of keeping both the time and frequency information. The wavelets features are extracted from the image by the application of the discrete wavelet transform [22].

PCA feature is extracted from the image on application of PCA. The PCA as feature extraction provides uncorrelated feature of the image. In this approach, a single coefficient vector of size 1xM is computed using the principal components analysis for input image [23].

C. Feature Selection

The feature selection is the process of selecting the relevant feature from the continuous values of extracted features for the classification purpose. The feature selection reduces the vector size of the extracted feature increasing the speed of the process in the classification phase. In this approach, feature selection is performed using Cuckoo Search algorithm

Cuckoo search algorithm is a Meta heuristic optimization algorithm which is based on the brood parasitic behavior of the cuckoo bird and levy flight behavior of some birds and fruit flies. The cuckoo birds lay eggs in a randomly selected host bird's nest. The cuckoo birds remove the eggs of the host bird egg from the nest to increase the survival probability.

D. Association Rule Generation

An association rule is an expression, where P and Q are itemsets $P \Rightarrow Q$. The meaning of the rule is natural. For a transaction database D containing T transaction, the rule $P \Rightarrow Q$ expresses the fact that whenever a transaction T contain itemset P, then T also contains the Q (probably). In the proposed methodology, association rules are generated by submitting the transaction representation of each image to the apriori algorithm. Apriori is one of the association rule generation algorithm. The association rule generation using apriori algorithm is explained in [21]. In this work, the apriori algorithm scans the transaction database D and generates the rule. Furthermore, the rules are evacuated using the two key parameters lift and conviction ratio. They are formulated below;

Lift:

Lift of a rule of itemset P and Q is defined as;

$$\mathrm{Lift}\left(P \Rightarrow Q\right) = \frac{\mathrm{Support}(P \cup Q)}{\mathrm{Support}(P) \times \mathrm{Support}(Q)} \quad \text{----(1)}$$

Conviction Ratio:

The conviction ratio of a rule of itemset P and Q is defined as;

$$\mathrm{ConvictionRatio}(P \Rightarrow Q) = \frac{1 - \mathrm{Support}(Q)}{1 - \mathrm{Confidence}(P \Rightarrow Q)} \quad \text{----- (2)}$$

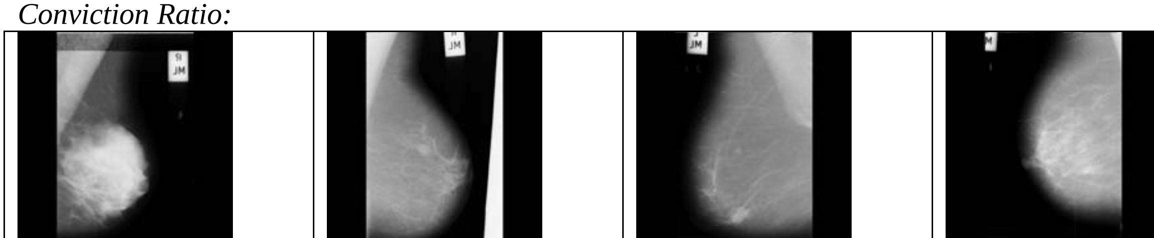
Here, $\mathrm{Support} = \frac{|P \cup Q|}{|R|}$ is defined as the support count of itemsets P and Q in transaction database to total number of transaction and $\mathrm{Confidence} = \frac{|P \cup Q|}{|P|}$ is defined as the support count of itemsets P and Q divided by the support count of itemset P in the transaction database. The association rules are generated limiting the conviction ratio and lift value to maximal values. Let R_k be the number of the association rule generated using the lift and conviction ratios.

E. Classification

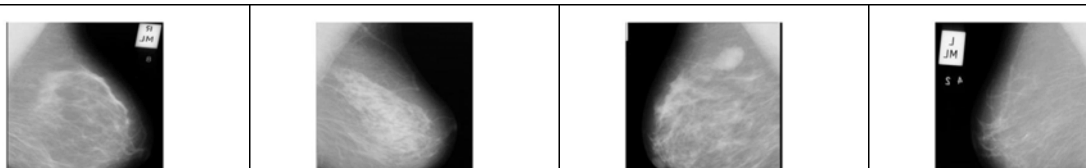
The classification is based on the decision tree classifier [21]. The test images are classified into normal and abnormal cancerous image based on the association rule generated from the features extracted from the training phase. Decision tree classifier is a useful tool for classification which classifies based on the rule generated from the input data. The significant advantage of decision tree classifier is ease of computation process.

IV EXPERIMENTAL RESULTS

Figure 4 depicts the experimental results of the proposed method.



(a)



(b)

{0-0} => {82978-101160} conviction: 0.73 Lift: 0.95 Sup: 0.76 {0-0} => {0-1} {4533-13227} conviction: 0.78 Lift: 0.94 Sup: 0.80 {82978-101160} => {0-0}{0-1} conviction: 0.83 Lift: 1.02 Sup: 0.64 {0-1} => {0-0}{8853441-9678100} conviction: 0.78 Lift: 0.97 Sup: 0.66 {0-0} => {10221349-11085993}{7054747-7151922} conviction: 0.53 Lift: 0.97 Sup: 0.55

(c)

Figure 4 Experimental Results

In figure 4, 4a represents the sample mammogram images of the breast considered for the experimentation. Figure 6b represents the pre-processed mammogram images. The pre-processing image alters from the original mammogram image in the main removing the unwanted portions of the image. Figure 6c represents the association rule generated from the feature vectors selected using cuckoo search algorithm. Finally with these rules mammograms are classified as normal or cancerous.

V CONCLUSION

In this paper a new breast cancer diagnosis methodology using Cuskoo search feature selection algorithm and Decision tree classifier is presented. The relevant features from the extracted features is selected using the cuckoo search algorithm which increases the speed of the whole diagnosis process by providing selected feature for the classification phase.

REFERENCES

[1] U. Raghavendra, U. Rajendra Acharya and Hamido Fujita, “Application of Gabor wavelet and Locality Sensitive Discriminant Analysis for automated identification of breast cancer using digitized mammogram images”, Applied Soft Computing, vol. 46, pp. 151-161, September 2016.

[2] Aswini Kumar Mohanty, Manas Ranjan Senapati and Saroj Kumar Lenka, “An improved data mining technique for classification and detection of breast cancer from mammograms”, Neural computing and applications, vol. 22, no. 1, pp. 303-310, May 2013.

[3] J. Dheeba, N. Albert Singh, and S. Tamil Selvi, “Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach”, Journal of Biomedical Informatics, vol. 49, pp. 45-52, June 2014.

[4] Marek R. Ogiela and Natalia Krzyworzeka, “Heuristic approach for computer-aided lesion detection in mammograms”, Soft computing, pp. 1-10, May 2016.

[5] Sami Dhahbi, Walid Barhoumi and Ezzeddine Zagrouba, “Breast cancer diagnosis in digitized mammograms using curvelet moments”, Computers in Biology and Medicine, vol. 64, pp. 79-90, September 2015.

[6] Chisako Muramatsu, Takeshi Hara, Tokiko Endo and Hiroshi Fujita, “Breast mass classification on mammograms using radial local ternary patterns”, Computers in Biology and Medicine, vol. 72, pp. 43-53, May 2016.

[7] Subrata Kar and D. Dutta Majumder, “An Investigative Study on Early Diagnosis of Breast Cancer Using a New Approach of Mathematical Shape Theory and Neuro-Fuzzy Classification System”, International Journal of Fuzzy Systems, vol. 18, no. 3, pp. 349-336, June 2016.

[8] Chisako Muramatsu, Kohei Nishimura and Tokiko Endo, “Representation of Lesion Similarity by Use of Multidimensional Scaling for Breast Masses on Mammograms”, Journal of Digital Imaging, vol. 26, no. 4, pp. 740-747, August 2013.

[9] Brahim Belhaouari Samir, Mohamed Meselhy Eltoukhy and Ibrahima Faye, “A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation”, Computerized Medical Imaging and Graphics, vol. 42, no. 1, pp. 123-128, January 2012

[10] Jun-Bao Li, Yun-Heng Wanga and Lin-Lin Tang, “Mammogram-based discriminant fusion

analysis for breast cancer diagnosis", Clinical Imaging, vol. 36, no. 6, pp. 710-716, December 2012.

[11] Mohamed Meselhy Eltoukhy, Ibrahima Faye and Brahim Belhaouari Samir, "Breast cancer diagnosis in digital mammogram using multiscale curvelet transform", Computerized Medical Imaging and Graphics, vol. 34, no. 4, pp. 269-276, June 2010.

[12] Jiri Grim, Petr Somol, Michal Haindl and Jan Dane, "Computer-Aided Evaluation of Screening Mammograms Based on Local Texture Models", IEEE Transactions on Image Processing, vol. 18, no. 4, pp. 765-773, April 2009.

[13] Karthikeyan Ganesan, U. Rajendra Acharya, Chua Kuang Chua, Lim Choo Min, "Computer-Aided Breast Cancer Detection Using Mammograms: A Review", In proceedings of IEEE world conference on Complex systems, pp. 626-631, November 2014.

[14] Fracheboud J, de Koning HJ et al., "Nationwide breast cancer screening programme fully implemented in The Netherlands", Journal Breast, vol. 10, no. 1, pp. 6-11, 2001.

[15] Andolina, V., & Lillé, S. (2011). Mammographic imaging: a practical guide. Lippincott Williams & Wilkins.

[16] Shradhananda Beura, Banshidhar Majhi and Ratnakar Dash, "Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer", Neurocomputing, vol. 154, pp. 1-14, April 2015.

[17] Anna N. Karahaliou, Ioannis S. Boniatis, Spyros G. Skiadopoulos, "Breast Cancer Diagnosis: Analyzing Texture of Tissue Surrounding Microcalcifications", IEEE Transactions on Information Technology in Biomedicine, vol. 12, no. 6, pp. 731-738, November 2008.

[18] Emina Alic kovic and Abdulhamit Subasi, "Breast cancer diagnosis using GA feature selection and Rotation Forest", Neural Computing and Applications, pp. 1-11, November 2015.

[19] Saba Bashir, Usman Qamar and Farhan Hassan Khan, "Heterogeneous classifiers fusion for dynamic breast cancer diagnosis using weighted vote based ensemble", Quality and Quantity, vol. 49, no. 5, pp. 2061-2076, August 2014.

[20] Xin-She Yang, Suash Deb, "Cuckoo Search via Levy Flights", World Congress on Nature & Biologically Inspired Computing, pp. 210 - 214, 2009.

[21] Philip H Swain, Hans Hauska, "The Decision Tree Classifier: Design and Potential", IEEE Transactions on Geoscience Electronics, Vol. 15, no. 3, pp. 142-147, 1977.

[22] A.N. Akansu, W.A. Serdijn, and I.W. Selesnick, "Wavelet Transforms in Signal Processing: A Review of Emerging Applications", *Physical Communication*, vol. 3, no. 1, pp. 1-18, 2010.

[23] A. A. Miranda, Y. A. Le Borgne, and G. Bontempi, "New Routes from Minimal Approximation Error to Principal Components", *Neural Processing Letters*, vol. 27, no. 3, pp. 197-207, 2008.

[24]J. C. Felipe, A. J. M. Traina, and C. Traina, "Retrieval by content ofmedical images using texture for tissue identification," in Proc. 16th IEEE Symp. Computer-Based Med. Systems, pp. 175–180, 2003.

[25] X. Wang, M. Smith, and R. Rangayyan, "Mammographic information analysis through association-rule mining," in Proc. IEEE CCGEI, pp. 1495–1498, 2004..

[26] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules", in Proc. Int. Conf. VLDB, pp. 487–499, 1994.

[27] Edward R. Omiecinski, "Alternative Interest Measures for Mining Associations in Databases", IEEE Transactions On Knowledge and Data Engineering, vol. 1, no. 5, pp. 57-69, 2003.

[28] MIAS database, http://www. mammoimage.org/databases/