International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 05 | May-2018 www.irjet.net p-ISSN: 2395-0072

A Survey on Computer Aided Diagnosis System for Mammogram Classification

Sumit Kumar Sourav[1], Rajeev Ranjan[2], Manjunath M[3], Fazeel Ur Rehman Khan[4], Savitha G[5]

[1,2,3,4] Student BE final year, dept. of CSE BNMIT, Bangalore, India
[5]Professor, dept. of CSE BNMIT, Bangalore, India

Abstract: *Breast cancer is one of the crucially prevailing cancer among women. Early detection and diagnosis of breast cancer can be facilitating with mammography images since they are most cost effective and a good chance of recovery. Classification is an identification technique used to organize the data into categories. Classification algorithm identifies the severity of lymph's present in the breast. The entire study focuses on different classifier techniques which can be used after pre-processing and segmentation process to improve the accuracy result of the image and can be categorized as well. A survey on suitable techniques for mammogram images such as Curvelet Transform, K-nearest Neighbour, Deep learning method, Convolution Neural Network etc is done. For each classification, factors such as sensitivity, specificity and accuracy which are chosen according to their suitable scenarios.*

Keywords: Benign, Malignant, Micro calcification, Mammography, Neural Network.

1. INTRODUCTION

Breast cancer is one of the prominent reasons for deaths of women, and according to a 2016 estimate, 61,000 new cases of breast cancer are predicted. If breast cancer is detected earlier through mammographic screening, then the chances of survival are greater than 90%. To diagnosis breast cancer, the digital screening mammography is extensively utilized by radiologists as the most reliable and cost-effective method. For radiologists, the detection or interpretation of breast masses through digital mammography is a time-consuming task. Furthermore, the ability of mammography is limited in extremely dense breasts and detection accuracy is as low as 60%–70%. In this way, the computer-aided diagnosis (CAD) systems are advanced to support radiologists for the identification of benign and malignant masses. In the literature, there are many studies that suggested incorporating the CAD system into the diagnostic process of screening breast images. This step can increase the performance of routine diagnosis by decreasing inter-observer variation and providing the quantitative support for the clinical decisions. The current paper designates an effort to develop a CAD system for early detection of breast masses by using new features and machine learning techniques to assist radiologists.

The identification of benign and malignant masses in digital mammograms through a CAD system is one of the most important processes for radiologists due to the analysis of various breast tissue structures. An accurate classification allows medical experts to accurately differentiate between benign and malignant masses. Therefore, the primary aim of this paper is to develop an improved CAD system for effective diagnosis of breast masses. In the next section, background of this study are discussed in detail.

2. BACKGROUND STUDY

Several computer-aided diagnosis (CAD) systems were developed in the literature to provide a second opinion for the assistance of radiologists. The previous CAD systems were developed based on three main steps such as segmentation, feature extraction and classification. These steps are well-addressed in the past studies to find the discriminative features for classification of breast masses as benign or malignant. Unfortunately, these steps require a complicated task such as pre- and post-processing steps and depended on critical domain expert knowledge about image processing. Despite these methods, a few CAD systems have also been developed recently to overcome these problems.

3. MODULES IN MAKING CAD

a) Pre-processing of Mammograms

Mammograms show a projection of the breast that can be made from different angles. As described in paper [1][2] the two most common projections are medio-lateral oblique and craniocaudal. The advantage of the mediolateral oblique projection is that almost the whole breast is visible, often including lymph nodes. The main disadvantage is part of the pectoral muscle will be shown in upper part of the image, which is superimposed over a portion of the breast. The craniocaudal view is taken from above, resulting in an image that sometimes does not show the area close to the chest wall. In paper [1][2] they considered the earlier one for its advantage but pectoral muscle detection is one more difficult task in the breast segmentation process. Reason for detecting pectoral muscle is to remove. Suppression can help in some auto detect procedures such as finding bilateral asymmetry etc.

It is important to detect the pectoral muscle and defines the region of interest (ROI) for further analysis. This operation is important in medio-lateral oblique (MLO), where the pectoral muscle, slightly brighter compared to

the rest of the breast tissue, can appear in the mammogram. To detect the same, the mammograms are pre-processed so that the intensity should be proper to work on and to avoid distortion. Filters are used as primitive way to pre-process but now a days Curvelet Transform is very useful as it helps in avoiding data loss along the borders of breast unlike filters as they smooth's the mammograms causing data loss [3].

b) Clustering:

In paper [4] author suggested new method for breast image segmentation. In order to detect breast cancer based on micro calcification an adaptive k-means clustering algorithm is proposed. In this stage, the Region of Interest(ROI) is calculated, and the mammograms have a black region on the left and right end sides. So these regions are removed using threshold based morphology technique. Next the cropped image is clustered to find the ROI that has micro calcifications and tumors. k-means is one of the simplest learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters). The main idea is to define k centers, one for each cluster. These centers should be placed in a clever way because of different location causes different result. So, the better choice is to place them far from each other. The next step is to take each point belonging to a given data set and associate it to the nearest Centre. When no point is pending the first step is completed otherwise k new centroids are recalculated resulting from the previous step.

After these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop it may be noticed that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm aims at minimizing an objective function know as squared error function.

c) Segmentation:

Image segmentation is the fundamental approach of digital image processing. Among all the segmentation methods, Otsu method [5]

[6] is one of the most successful methods for image thresholding due to automatically compute the threshold. Thresholding helps to create binary images from grey-level image by converting all pixels value below a threshold to zero and the remaining pixels to one, Binary image help to remove all the background and leave only the breast prior to segment the tumor from the breast (It is necessary to work on tumor only to detect the benign or malignant breast). Tumor will be segmented by reconstructing the image from binary image.

d) Mammogram classification:

There are various algorithms for automated classification. A survey is made for several classification algorithms to compare their performance: Random forest (RF), The Naive Bayes (NB), C4.5, The multi-layer perceptron (MLP) and Decision Table (DT). Random Forest (RF) is an approach which has been proposed by Breiman for classification tasks. It mainly comes from the combination of tree-structured classifiers with the randomness and robustness provided by bagging and random feature selection [7]. The classification is performed by sending a sample down is each tree and assigning it the label of the terminal node it ends up in.

At the end the average vote of all trees is reported as the result of the classification. Random forest is very efficient with large datasets and high dimensional data [8].

The Naive Bayesian (NB) is based on the Bayesian theorem [9]. The Naïve Bayesian Classifier assumes that features are independent [10]. This method is important for several reasons. It is very easy to construct and does not need any complicated iterative parameter estimation schemes. This means it may be readily applied to huge data sets [11]. This classification technique analyses the relationship between each attribute and the class for each instance to derive a conditional probability for the relationships between the attribute values and the class [9].

C4.5 is an extension of ID3 algorithm that was designed by Quinlan to deal with issues that cannot be handled by the ID3 algorithm. These include avoidance of over fitting the data; reduced error pruning, rule post-pruning, handling continuous attributes and handling data with missing attribute values [12]. It attempts to build a decision tree with a measure of the information gain ratio of each feature and branching on the attribute which returns the maximum information gain ratio [13]. Pruning takes place in C4.5 by replacing the internal node with a leaf node thereby reducing the error rate [14].

The multi-Layer Perceptron (MLP) is a feed forward neural network consisting of an input layer of nodes, followed by two or more layers of perceptron, the last of which is the output layer. The layers between the input layer and output layer are referred to as hidden layers [14]. The major aim of MLP algorithms is to automatically learn and make intelligent decisions. It is known as feed forward because it does not contain any cycles and network output depends only on the current input instance.

It is known as feed forward because it does not contain any cycles and network output depends only on the current input instance. Learning take place by changing connection weights after each piece of data is processed, based on the amount of error in the target output as compared to the expected result. [15].

A Decision Table (DT) is the method used to build a complete set of test cases without using the internal structure of the program in question. In order to create test cases, we use a table to contain the input and output values of a program. Such a table is split up into four sections [16]. Two variants of decision table classifiers are available. The first classifier, called DTMaj (Decision Table Majority) returns the majority of the training set if the decision table cell matching the new instance is empty, that is, it does not contain any training instances. The second classifier, called

DTLoc (Decision Table Local), is a new variant that searches for a decision table entry with fewer matching attributes (larger cells) if the matching cell is empty. This variant therefore returns an answer from the native region [17].

4. DATA FLOW IN CAD:

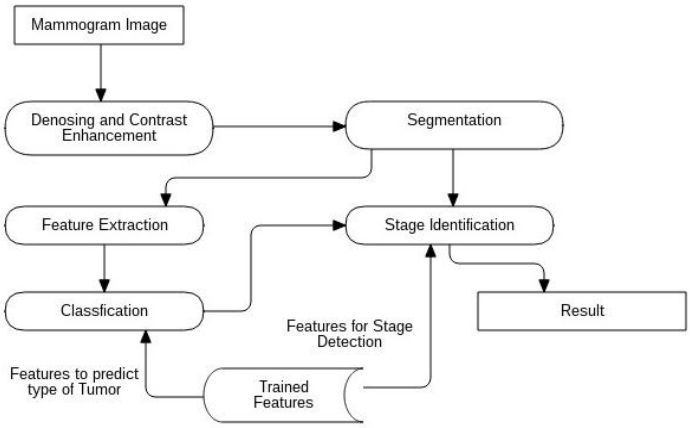

Fig 1: Modules and Interaction in CAD

5. STEPS IN MAKING CAD:

As describe in FIG 1 the general steps to be followed to construct CAD are as follows:

Step1: All mammogram images are preprocessed to extract region of interest. The region of interest is extracted from the abnormal images depending on the information contained.

Step2: To remove noise from mammogram images and improve the quality of the region of interest filters or Transform techniques are used.

Step3: The features are extracted from the normalized image regions using First-Order Statistics and Second-Order Statistics.

Step4: Finally, mammogram images is classified using number of classification algorithm such as (Random forest, The Naive Bayes (NB), C4.5, The multi-layer perceptron (MLP) and Decision Table (DT) classifier).

6. ADVANTAGE OF CAD:

The CAD systems can improve the diagnostic performance and, at the same time, reduce the radiologists' interpretation variability. The CAD system can play the role of a reference reader completely immune to human variability, i.e., it can reduce the variability which is dependent on the radiologists' interpretation that is subject to variations inherent to the human perception and to the decision making process. The rate of detection of architectural distortion by CAD is low, increasing for masses, and achieving almost 100% for micro calcifications, helping radiologists in reducing errors in the detection. of carcinomas, potentially avoiding unnecessary biopsies. The CAD system constitutes an active area of investigation and development in radio diagnosis. Its technological features and software versions have undergone swift changes.

7. CONCLUSION

Breast cancer is one of the most common cancers among women around the world. In this paper, a survey is made to describe the techniques used in making a CAD for mammogram classification. This paper suggested to use the Computer aided diagnosis systems to help the doctors in detection/diagnosis of abnormalities faster than the traditional procedures. The automated mammogram classification method suggested based on first-order statistics used for features extraction post of tumor segmented. Number of features selected to distinguish between benign and malignant breast cancer, such as mean, entropy, standard deviation, and variance.

8. REFERENCES

[1] Mustra M and Grgic M. “Robust automatic breast and pectoral muscle segmentation from scanned mammograms“, Signal Processing 2013; 93: 2817-2827.

[2] Dos Santos RLC, da Costa VMA and Schiabel H. “Mammography images restoration by quantum noise reduction and inverse MTF filtering” IEEE Brazilian SympComput Graphics Image Process (SIBGRAPI) 2009.

[3] Shanthi DR and M.L. Valarmathi; “Comparison of Wavelet, Contourlet and Curvelet Transform with Modified Particle swarm Optimization for Despeckling and Feature Enhancement of SAR Image” 978-1-4673-4862-1/13/$31.00 2013 IEEE

[4] Bhagwati Charan Patel and G. R. Sinha, “An Adaptive K-means Clustering Algorithm for Breast Image Segmentation”, International Journal of Computer Applications Volume 10, No 4. DOI: 10.5120/1467- 1982

[5] Liu Jian-zhuang and Li Wen-qing “The automatic thresholding of graylevel pictures via two-dimensional Otsu method” Automatica Sinica.1993,19(1),pp.101-105

[6] WU Cheng-Mao, TIAN Xiao-Ping and TAN Tie-Niu.Fast "Iterative Algorithm for Two-Dimensional Otsu thresholding Method" PR&AI.2008,21(6), pp.746~757

[7] Mariana R. Mendoza, Guilherme C. da Fonseca, Guilherme Loss-Morais, Ronnie Alves, Rogerio Margis and Ana L. C. Bazzan, "Predicting Human MicroRNA Target Genes with a Random Forest Classifier", plos , 2013.

[8] Blagojce Jankulovski, Ivan Kitanovski, Katarina Trojacanec, Ivica Dimitrovski and Suzana Loskovska, "Mammography Image Classification Using Texture Features ", The 9th Conference for Informatics and Information Technology , 2012.

[9] RK Kavitha1 and Dr. D Dorai Rangasamy, "Predicting Breast Cancer Survivability Using Naïve Baysein Classifier and C4.5 Algorithm ", Elysium journal Engineering Research and management, 2014

[10] Lavneet Singh and Girija Chetty, "A Comparative Study of MRI Data using Various Machine Learning and Pattern Recognition Algorithms to Detect Brain Abnormalities ", Proceedings of the Tenth Australasian Data Mining Conference 2012.

[11] Xindong Wu, Vipin Kumar, J. Ross Quinlan, and Joydeep Ghosh, "Top 10 algorithms in data mining", SpringerVerlag London Limited,2007.

[12] K.Rajesh and S. Anand," Analysis of SEER Dataset for Breast Cancer Diagnosis Using C4.5 Classification Algorithm", International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 2, 2012

[13] Lipo Wang and Xiuju Fu, "Data Mining with Computational Intelligence", Springer - Verlag Berlin Heidelberg, 2005.23

[14] G. Sujatha and Dr. K. Usha Rani, "Evaluation of Decision Tree Classifiers on Tumor Datasets", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 2013.

[15] Komal Sharma, AK winder Kaur and Shruti Gujral, "Brain Tumor Detection based on Machine Learning Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 103 – No.1, October 2014.

[16] Vijayarani M and Muthulakshmi, "Evaluating The Efficiency Of Rule Techniques For File Classification", International Journal of Research in Engineering and Technology, eISSN: 2319-1163 | pISSN: 2321- 7308, 2013.

[17] Vasileios A. Tatsis, "Evaluating data mining algorithms using molecular dynamics trajectories", Int. J. Data Mining and Bioinformatics, Vol. 8, No.2, 2013.

[18] Sampat M.P, Bovik A.C, Whitman G.J and Markey M.K. "A model-based framework for the detection of speculated masses of mammography" Med.Phys.2008, 35, 2110–2123.

[19] Bellotti R, De Carlo F, Tangaro S, Gargano G, Maggipinto G, Castellano M, Massafra R, Cascio D, Fauci F and Magro R, et al. "A completely automated CAD system for mass detection in a large mammographic database." Med.Phys.2006, 33, 3066–3075.

[20] Samulski M and Karssemeije N. "Optimizing case-based detection performance in Multiview CAD system for mammography. IEEE Trans. Med. Imaging 2011, 30, 1001–1009.

[21] Eltoukhy M.M, Faye I and Samir B.B. "A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation." Comput. Biol.Med.20124 2, 123–128.

[22] Wang H, Li J.B, Wu L and Gao H. "Mammography visual enhancement in CAD-based breast cancer diagnosis" Clinic Imaging 2013, 37, 273–282.

[23] Vadivel A and Surendiran B. "A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories" Comput. Biol. Med.2013,43, 259–267.

[24] Keleş A and Yavuz U. "Expert system based on neuro-fuzzy rules for diagnosis breast cancer" Expert Syst. Appl.2011, 38.

[25] Ramos R.P, do Nascimento M.Z and Pereira D.C. "Texture extraction: An evaluation of ridgelet wavelet and co-occurrence-based methods applied to mammograms." Expert Syst.Appl.2012, 39, 11036–11047.

[26] Jian W, Sun X and Luo S. "Computer-aided diagnosis of breast micro calcifications based on dual-tree complex wavelet transform" Biomed. Eng. Online2012,11, 1–12.

[27] Zhang X, Homma N, Goto S, Kawasumi Y, Ishibashi T, Abe M, Sugita N and Yoshizawa M. "A hybrid image filtering method for computer-aided detection of microcalcification clusters in mammograms" J. Med Eng 2013,1-8

[28] Huang Y.J, Chan D.Y, Cheng D.Y, Cheng D.C, Ho Y.J, Tsai P.P, Shen W.C and Chen R.F. "Automated feature set selection and its application to MCC identification in digital mammograms for breast cancer Detection" Sensors 2013, 13, 4855–4875