An Improved Decision Tree based Mammogram Image Classification of Breast Cancer for Decision Support Systems

N. Arivazhagan and V. Govindharajan

Abstract: Decision support system is used in medical field to enable the physician to make quick and efficient diagnosis of diseases using information and communication technology. Quick retrieval of image sample for comparison and classification is one of the important criteria for the design of such a system. Fast decision and accuracy of the classification is always trade off. The fast decision and retrieval usually suffer from a problem of less accuracy. Here in this paper a hybrid mechanism based image retrieval cum classification is proposed which will make quick decision with guaranteed level of accuracy. The proposed system is tested on mammogram medical images for finding early detection of cancer at various stages .The computational complexity analysis result on large image data base system proves the proposed scheme can be implementable in a typical practical decision support system.

Keywords: Mamogram Images, Decision support systems, Early cancer detection.

I. INTRODUCTION

The development towards medical image processing has achieved new heights. VOI (Volume of Interest) based techniques are the backbone technologies behind Multi dimensional image storage systems. A common approach for Image feature extraction is to segment the object into regions (objects) based on similar (RoI) region of interests. Guimond et al. [1] introduced the user-selected volume of interest (VOI) for the retrieval of pathological brain MRI images. The approaches, to VOI location was based on the transformation of the segmented images to the standard atlas and the research works, utilize the anatomical standardization procedure of the 3-D stereotactic surface projection transformation in the NEUROSTAT package[2-5]. There are ‘n’ number of research studies has been used to prove that decision support systems has been used to predict the level of malignance in cancer diseases. Konstantinos P. Exarchos et.al [6] states that there are three sources of data are employed, namely clinical, imaging, and genomic in pursue of the most prominent OSCC proliferating factors to identify oral cancer at early stages. Significant works has been done in the area of feature selection, the major frequent algorithms are correlation-based feature subset selection (CFS) [7] and the wrapper algorithm [8]. In 2018, an estimated 266,120 new cases of invasive breast cancer are expected to be diagnosed in women in the U.S., along with 63,960 new cases of non-invasive (in situ) breast cancer [9]. Most of the microwave breast cancer detection algorithms currently used in conjunction with actual hardware are burdened with significant computational times which currently limit their clinical utility. For example, it is not uncommon to wait tens of hours or even days for a single 3-D microwave tomographic image [10]. Image classification aims at assigning to each pixel, or to each region of interest (ROI) extracted from the image, a label associated with a class of object that can possibly be present in the analyzed scene [11]. Supervised learning techniques are used to analysis the breast images, Based on mammography images, the American College of Radiology has developed a method called BI-RADS (Breast Imaging-Reporting and Data System) for deciding whether biopsy of an identified suspicious lesion is indicated. Biopsy is the gold standard for breast cancer diagnosis; however it is an expensive, discomforting, and invasive procedure. Almost 80% of the biopsies carried out based on BI-RADS score turn out to be benign [12]. Hence, there exists a need for reducing the number of unnecessary breast biopsies and augment the current diagnosis methods. The machine learning framework are used to estimate the cancer likelihood (for SVM) and probability (in case of Random Forests). The estimated cancer likelihood is then used to generate a malignancy map.

Revised Manuscript Received on December 30, 2018.

N. Arivazhagan, Research Scholar, School of Computing, SRMIST, Chennai, India.

V. Govindharajan, Professor, School of Public Health Sciences, SRMIST, Chennai, India.

II. RELATED WORKS

The malignance level of cancer cells are monitored on a periodic basis. The ROI has to be measured in a regular clinical interval to prevent the early stage breast cancers, which is also used to cross measure the pharmological support and the growth of the cancer tissues. There are systems for fast data processing and new software paradigms yields a system able to image patients in several minutes[14]. Clinical decision support (CDS) tools help clinicians make detection and diagnostic decisions for complex diseases such as lung cancer [15], breast cancer [16], [17], and diabetes [18]. Linqi Song, has proposed CDS tool to aid physicians with making management decisions particularly in borderline cases. Radiological assessment of breast images are categorized using the Breast Imaging Report and Data System (BI-RADS) score. BI-RADS scores of 3 or 4 represent borderline cases associated with short interval followup or biopsy, respectively. Presently, many benign cases are being classified as BI-RADS 4A, which has raised the concern of overdiagnosis [19]. Content-based image retrieval (CBIR) has been a widely used image retrieval technique in various computer vision applications such as in medical field for getting the past patient details and the success of local binary pattern (LBP) in image processing applications led to lot of research in feature extraction based on spatial techniques [20].

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Clustering algorithms are one of the major contributions in finding solutions for health informatics [21].

III. SYSTEM APPROACH

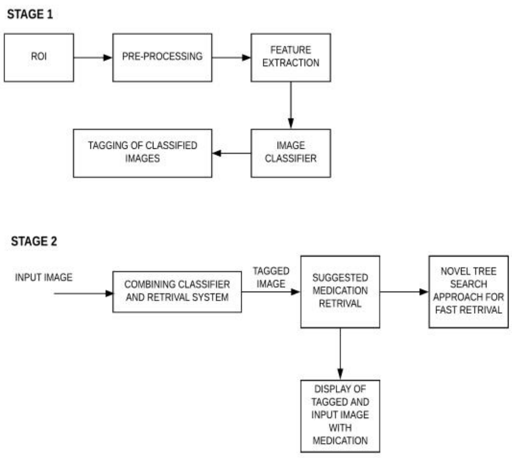

Figure 1: Stages of Decision Support system Classifier

The development of decision support system consists of two-stage process, as shown in Fig.1.

Stage 1 Process

In stage one ROI extraction, Preprocessing, Feature extraction and image classification has to be carried out with a tagging of classified images. This stage one process ensures for the given database of images, Tagged image-set for malignant and nonmalignant with two level of diseases will be available reliably on the testing face for medication recommendation.

Stage 2 Process

The performance of decision support system relays on the quick retrieval of tagged image samples, quick search algorithms, quick match processing. Based on this understanding the stage two processes is developed to realize quick retrieval and decision recommendation of the system. In order to reduce the matching and classification time without compromising much on accuracy. A decision tree based hierarchical classification is proposed. This approach will improve the accuracy by stage by stage, verifying the classification using different feature set at each stage.

The above approach combines this searching and matching process by tree mechanism.

IV. DECISION TREE APPROACH FOR HIERARCHICAL CLASSIFICATION

Decision tree mechanism is applied in order to support hierarchical classification. A regression tree is formed in order to classify the mammogram images as malignant and non malignant. At each level of tree a feature vector with a threshold value is used for make a branch. The regreesion tree follows a binary tree model at the leaf node of the tree that regression computation is carried out in order to predict the presence of cancer disease. Table 1. Presents the various feature applied on the regression tree with typical value and threshold value.

Table 1: Feature set and threshold value for regression tree.

Feature set	Typical or average value	Threshold value for non malignant
Kurtosis	10.23	10
Skewness	2.98	2
SD	25.2	24

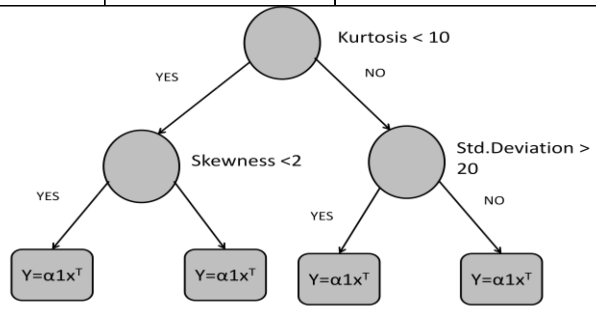

Figure 2: Regression Decision Tree for Malignant and Non-Malignant classification

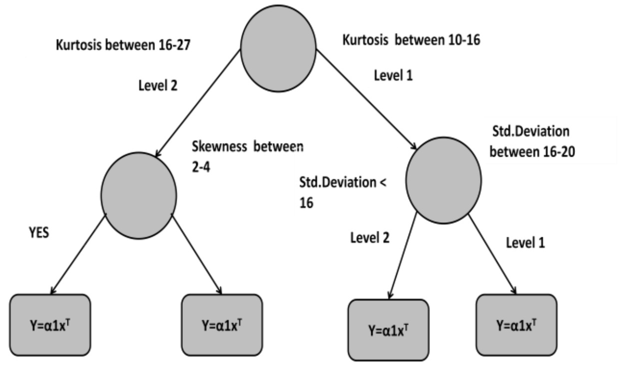

Figure 3: Regression Decision tree for two level breast cancers

Figure 2 shows the regression decision tree to classify the given image as malignant and non-malignant. The tree feature set called kurotias, skewness and standard deviation are used for traversal of the tree. The feature values are extracted for the given database and the average of the values as in the Table.1 is used to set the threshold value for decision making and traverse of the trees at each node.

Figure 3 shows the regression decision tree for malignant images to classify further in to two level of diseases. Based upon the computed reference values for malignant of two levels, this tree will make decision and traverse to finalize the level of malignant with more reliability.

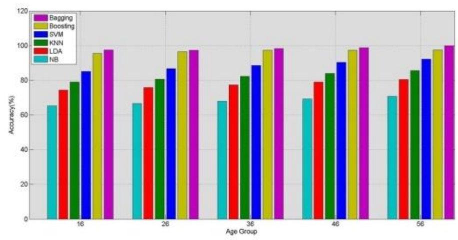

Figure 4: Accuracy results for Decision Support embeded Classification Algorithm

The observations on Fig.7 over Table.1 has been predicted, that the total observations are higher than the algorithms without NFe.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

On an average, after 'n' experimental iterations, the total accuracy has been improved from 95.8 % to 97.7% in proposed NFe when compared with ensemble algorithms. The test results have been assessed to find the accurate precision value through false positives and true positives. The total correctly predicted positive precision and recall achieves more than that of .80, which is given in Fig.5 and Fig.6.

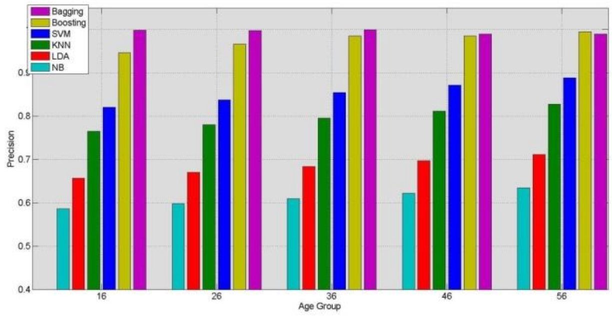

Figure 5: Precision Results using Decision Tree Algorithm

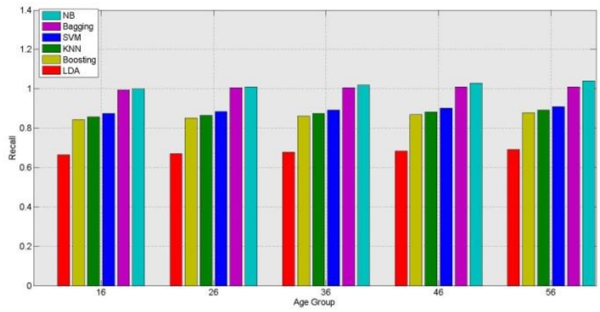

Figure 6: Recall Results using Decision Tree Algorithm

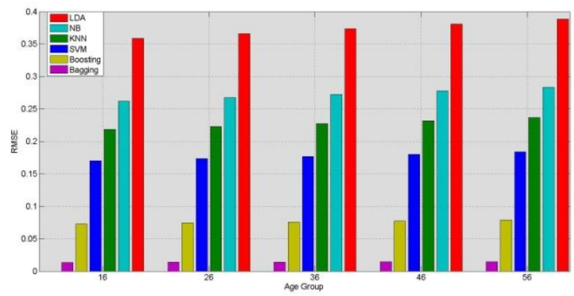

Figure 7: RMSE using Decision Tree Algorithm

After all operational functions, the RMSE (Root Mean Square Error) in Fig.7 has been reduced below the levels of existing algorithms given in Table.1. X axis is the age group for them the mammogram image samples are collected and Y axis is the efficiency parameter like accuracy of classification, Precision, recall and RMSE values.

REFERENCES

1. Guimond and G. Subsol, "Automatic MRI database exploration and applications," Int. J. Pattern Recognit. Artif. Intell., vol. 11, pp. 1345–1365, 1997.
2. Jinman Kim, Weidong Cai, Member, Dagan Feng, Hao Wu, "A New Way for Multidimensional Medical Data Management: Volume of Interest (VOI)-Based Retrieval of Medical Images With Visual and Functional Features", IEEE Transactions On Information Technology In Biomedicine, Vol. 10, No. 3, July pp. 598-608, 2006.
3. S. Minoshima, K. L. Berger, K. S. Lee, and M. A. Mintun, "An automated method for rotational correction and centering of three-dimensional functional brain images," J. Nucl. Med., vol. 33, pp. 1579–1585, 1992.
4. S. Minoshima, R. A. Koeppe, M. A. Mintun, K. L. Berger, S. F. Taylor, K. A. Frey, and D. E. Kuhl, "Automated detection of the intercommissural line for stereotactic localization of functional brain images," J. Nucl. Med., vol. 34, pp. 322–329, 1993.
5. S. Minoshima, R. A. Koeppe, K. A. Frey, and D. E. Kuhl, "Anatomic standardization: Linear scaling and nonlinear warping of functional brain images," J. Nucl. Med., vol. 35, pp. 1528–1537, 1994.
6. Konstantinos P. Exarchos, Yorgos Goletsis, Dimitrios I. Fotiadis, "Multiparametric Decision Support System for the Prediction of Oral Cancer Reoccurrence", IEEE Transactions on Information Technology In Biomedicine, Vol. 16, No. 6, pp.1127-1135, November 2012.
7. M. Hall, "Correlation-based feature selection for discrete and numeric class machine learning," in Proc. 17th Int. Conf. Mach. Learning, 2000, pp. 359–366.
8. R. Kohavi and G. John, "Wrappers for feature subset selection," Artif. Intell., vol. 97, pp. 273–324, 1997
9. https://www.breastcancer.org/symptoms/understand_bc/statistics
10. S. Y. Semenov, A. E. Bulyshev, A. Abubakar, V. G. Posukh, Y. E. Sizov, A. E. Souvorov, P. M. van den Berg, and T. C. Williams, "Microwave-tomographic imaging of the high dielectriccontrast objects using different image-reconstruction approaches," IEEE Trans. Microwave Theory Tech., vol. 53, no. 7, pp. 2284–2294, Jul. 2005.
11. Émilie Niaf, Rémi Flamary, Olivier Rouvière, Carole Lartizien, and Stéphane Canu, "Kernel-Based Learning from Both Qualitative and Quantitative Labels: Application to Prostate Cancer Diagnosis Based on Multiparametric MR Imaging", IEEE Transactions On Image Processing, vol. 23, no. 3, march 2014.
12. S. Poplack, P. Carney, and J. Weiss, "Screening mammography: Costs and use of screening-related Services1," Radiology, vol. 234, no. 1, pp. 79–85, 2005.
13. Nishant Uniyal, Hani Eskandari, Purang Abolmaesumi, Samira Sojoudi, Paula Gordon, Linda Warren, Robert N. Rohling, Septimiu E. Salcudean, and Mehdi Morad, "Ultrasound RF Time Series for Classification of Breast Lesions", IEEE Transactions On Medical Imaging, vol. 34, no. 2, pp. 652-662, 2015.
14. Tomasz M. Grzegorczyk*, Paul M. Meaney, Peter A. Kaufman, Roberta M. diFlorio-Alexander, Keith D. Paulsen, "Fast 3-D Tomographic Microwave Imaging for Breast Cancer Detection", IEEE Transactions on Medical Imaging, Vol. 31, No. 8, August 2012
15. M. N. Gurcan, B. Sahiner, N. Petrick, H. P. Chan, E. A. Kazerooni, P. N. Cascade, and L. Hadjiiski, "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system," Med. Phys., vol. 29, no. 11, pp. 2552–2558, 2002.
16. J. Tang, R. M. Rangayyan, J. Xu, I. E. Naqa, and Y. Yang, "Computeraided detection and diagnosis of breast cancer with mammography: Recent advances," IEEE Trans. Inf. Technol. Biomed., vol. 13, no. 2, pp. 236-251, Mar. 2009.
17. K. Ganesan, U. Acharya, C. K. Chua, L. C. Min, K. Abraham, and K. Ng, "Computer-aided breast cancer detection using mammograms: A review," IEEE Rev. Biomed. Eng., vol. 6, pp. 77–98, Mar. 2013.
18. M. F. Ganji and M. S. Abadeh, "A fuzzy classification system based on Ant Colony Optimization for diabetes disease diagnosis," Expert Syst. Appl., vol. 38, no. 12, pp. 14650–14659, 2011.
19. Linqi Song, William Hsu, Member, Jie Xu, Mihaela van der Schaar, "Using Contextual Learning to Improve Diagnostic Accuracy: Application in Breast Cancer Screening", IEEE Journal of Biomedical And Health Informatics, Vol. 20, No. 3, May 2016
20. G. Subash Kumar, V. Nagarajan, "Local curve pattern for content-based image retrieval", Pattern Analysis and Applications, Accepted: 9 July 2018.
21. RD Rajagopal, S Murugan, K Kottursamy, V Raju, 'Cluster based effective prediction approach for improving the curable rate of lymphatic filariasis affected patients', springer cluster computing, pp.1-9, 2018.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication