Method of detecting micro-calcification clusters using Contourlet transform and PCNN.

Kalyani S. Gaware

Kalyani Gaware, Computer Science and Engg/ D.I.E.M.S/ Dr.BAMU University, Aurangabad, India

ABSTRACT

Mammography analysis is an effective technology for early detection of breast cancer. Micro-calcification clusters (MCs) are a vital indicator of breast cancer, so detection of MCs plays an important role in computer aided detection (CAD) system, this paper proposes a new hybrid method to improve MCs detection rate in mammograms. Methods: The proposed method comprises three main steps: firstly, remove label and pectoral muscle adopting the largest connected region marking and region growing method, and enhance MCs using the combination of double top-hat transform and grayscale-adjustment function; secondly, remove noise and other interference information, and retain the significant information by modifying the contourlet coefficients using nonlinear function; thirdly, we use the non-linking simplified pulse-coupled neural network to detect MCs. Results: In our work, we choose 118 mammograms including 38 mammograms with micro-calcification clusters and 80 mammograms without micro-calcification to demonstrate our algorithm separately from two open and common database including the MIAS and JSMIT; and we achieve the higher specificity of 94.7%, sensitivity of 96.3%, AUC of 97.0%, accuracy of 95.8%, MCC of 90.4%, MCC-PS of 61.3% and CEI of 53.5%, these promising results clearly demonstrate that the proposed approach outperforms the current state-of-the-art algorithms. In addition, this method is verified on the 20 mammograms from the People's Hospital of Gansu Province, the detection results reveal that our method can accurately detect the calcifications in clinical application.

Keywords/ Index Term — Mammography,Micro-calcification clusters, Contourlet Transform, Simplified pulse-coupled neural network(SPCNN)

1. INTRODUCTION

Breast Cancer is found mostly in woman in world as per World Health Organisation (WHO). And it is most common type of cancer in woman and also using this cancer death happened in developed and under-developed countries [1]. As per report nearly 1.7 million new cases were analysis in 2012 year, and in that 12% are new cancer disease and 25% cancers are in woman with the 522,000 deaths cases [2]. Breast Cancer are categories i.e 80% cases was come from high income countries and below 40% was come from low income countries [3]. In order to improve the diagnosis and prognosis of breast cancer, early detection is becoming more and more important [2]. The ways of breast cancer detection and diagnosis can be concluded into breast self-examination (BSE), clinical breast exam (CBE), imaging or mammography and surgery. Among these methods, X-ray mammography as the most efficient and reliable early detection technique is widely used by radiologists; it can detect 85–90% of all breast cancers. Microcalcification clusters (MCs) are a major sign of breast cancer in mammography [3], the size, shape, texture and distribution of the micro-calcifications provide significant information for diagnosis, hence the accurate detection of MCs is a critical step in computer aided detection (CAD) system.

Although making research on MCs detection in CAD system has sustained for decades, the research of calcification detection still possess meaningful and challenging topic because of the inhomogeneous background and the high noise level in mammography. Various approaches have been suggested to detect MCs accurately. A variety of techniques have been used in different steps. For mammogram enhancement, variety attempts have been done, such as improved histogram equalization [4], image enhancement based on wavelet fusion [5], automated lesion intensity enhance [6], modified multi-fractal analysis [7], etc.; in the segmentation step, many techniques have been suggested, such as multi-stable cellular neural networks, geodesic active contours (GAC) technique associated with anisotropic texture filtering [8], case-adaptive decision rule method [9], new scale-specific blob detection technique [10], etc.; in the third step, select true MCs by extracting a group of features of micro-calcifications like moment-based geometrical features [11], wavelet feature and Gabor feature [12] and so on. These aforementioned techniques make great contributions, however because the MCs detection faces different difficulties, the hybrid detection algorithms combining different theories seems more popular. Yu and Huang [16] investigated the performance of MCs by adopting combined model-based and statistical textural features, 20 mammograms containing 25 areas of MCs from the MIAS database were used to test the performance, and a true positive rate of about 94% was achieved at the rate of 1.0 false positive per image, or the false positives per image could be reduced to 0.65 false positive per image at the rate of true positive about 90%. Malar et al. [17] exhibited the effectiveness of wavelet based tissue texture analysis for detecting MCs in mammograms using extreme learning machine (ELM), the sample image were collected from the MIAS database, and achieved relatively better classification accuracy (94%).

2. Related work

In machine learning, feature selection is the process of choosing a subset of relevant attributes from various candidate subsets, and it is a prerequisite for model building. Feature selection plays a vital role in creating an effective predictive model. There are several benefits to applying the feature selection methods :it (a) is effective and faster in training the machine learning algorithm, (b) reduces the complexity of a model and makes it easier to interpret, (c) improves the accuracy of a model if the right subset is chosen, and (d) reduces overfitting. Because the remay exista complex interrelation between the features, it is generally difficult to choose the best subset [25]. Different approaches have been proposed in the literature for breast cancer diagnosis [7, 17–20]. Usually, feature selection methods are classified into three general groups: filter, wrapper, and embedded methods [26]. The filter method primarily relies on general features, and it is generally used as a preprocessing step. The subset selection is independent of any specific learning approach. The wrapper approach uses machine learning techniques to choose the optimal subset of features. In other words, the selection of the best features is guided by the learning process, as shown in Figure 3. The forward feature selection, backward feature elimination, and recursive feature elimination are widely used as wrapper methods. Embedded methods combine the qualities of filter and wrapper methods. These are implemented by algorithms that have their own built-in feature selection methods. They perform variable selection as a par to the learning procedure and are usually specific to the given learning machines. The diagram on sequence of data is shown in Figure4. Wrapper methods were used to conduct the experiments in this study.

Recent years, there have a great deal of researches engaged in development of computerized methods for automatic detection of MCs, which potentially give assistance to radiologists in diagnosis of breast cancer. Although making research on MCs detection in CAD system has sustained for decades, the research of calcification detection still possess meaningful and challenging topic because of the inhomogeneous background and the high noise level in mammography. Various approaches have been suggested to detect MCs accurately. According to these studies, these methods can be roughly divided into classic methods and emerging methods. These classic methods can be decomposed into three steps; firstly, reduce the noise and enhance MCs; secondly, detect the MCs applying a specific segmentation technique; thirdly, select true MCs by diverse novel methods. A variety of techniques have been used in different steps. For mammogram enhancement, variety attempts have been done, such as improved histogram equalization [4], image enhancement based on wavelet fusion [5], automated lesion intensity enhance [6], modified multifractal analysis [7], etc.; in the segmentation step, many techniques have been suggested, such as multistable cellular neural networks, geodesic active contours (GAC) technique associated with anisotropic texture filtering [8], case-adaptive decision rule method [9], new scale-specific blob detection technique [10], etc.; in the third step, select true MCs by extracting a group of features of micro-calcifications like moment-based geometrical features [11], wavelet feature and Gabor feature [12] and so on. These aforementioned techniques make great contributions, however because the MCs detection faces different difficulties,the hybrid detection algorithms combining different theories seems more popular.

.

3. Proposed System

In this section we review the underwater image formation model and the classical dehazing method, respectively.

Pipeline is the process of tying together some ordered final modules into one to build an automated machine learning workflow. It provides high level abstraction of the machine learning process and significantly simplifies the complete workflow. Mostly, it is known as Extract, Transform, and Load (ETL) operations. Unfortunately, the performance of a machine learning algorithm is determined by number of hyperparameters, including the number of trees in a random forest, the depth, number of hidden layers in the neural network, learning rate, batch size, and degree of regularization. The purpose of the work is to optimize the list of data transformations and machine learning algorithms to accomplish the classification transformation. To determine the best combination of machine learning algorithm and data is difficult .As are sult of the growth of hyperparameter tuning, genetic programming (GP) [22] is proposed to optimize the data and the control parameters of the proposed model. The use of this a well-known evolutionary technique is necessary to find the best combination that leads to highest evaluation results. The GP generates randomly a fixed number of pipelines which constitute the members of the population. Each individual (pipeline) of the population was evaluated based on its fitness which is chosen in this work as the classification score. The implementation of pipelines is based to supervised models from scikit-learn library. The hyper parameters optimized in this work are the number of kennels function for all the classifiers except linear discriminant analysis.

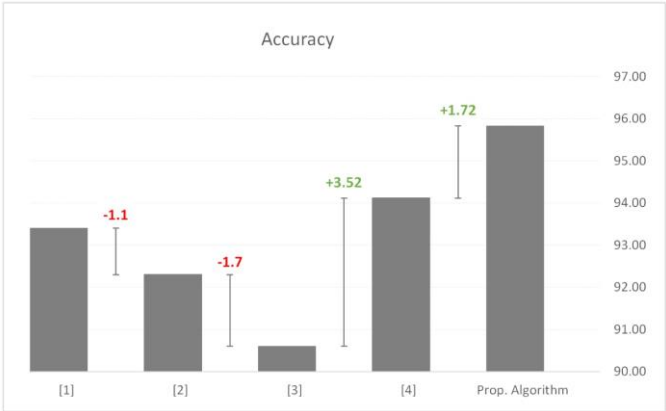

Fig 3.1 Accuracy of different algorithms.

As 1 algorithm is SVM (Support Vector Machine) where it shows -1.1 accuracy as compare to Proposed algorithm which is .Enhanced breast detection with feature selection and machine learning algorithm. As 2nd one is Receiver operating characteristic (ROC) curve which shows -1.7 accuracy as 3rd one is shear wave elastography (SWE) which shows +3.52 accuracy as 4th I compare which shows +1.72 accuracy. So we tried so many algorithms and then check so we got it our proposed algorithm has better accuracy as compare to other four algorithms accuracy.

The number of kernels function is chosen randomly.
In this work, many applied techniques were tested for the subsequent stages of processing and analysis of the breast cancer dataset

3.1.1 Stage 1: Preprocessing
As a part of this research, processing was performed on the raw breast cancer data to scale the features using the Standard Scaler module. Standardization of datasets is a common requirement for many machine learning estimators. It transforms the attributes to a standard Gaussian distributions based on (xi–mean(x))/ stdev(x) where stdev is the standard deviation. The Robust Scaler depends on the interquartile range to transform the features using (xi–Q1(x))/(Q3(x)–Q1(x)), where Q1,Q2 ,and Q3 represent quartiles. All the transformations used are included in scikit-learn machine learning library [27].

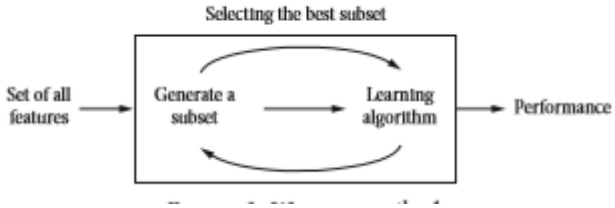

Fig 3.2 Wrapper methods

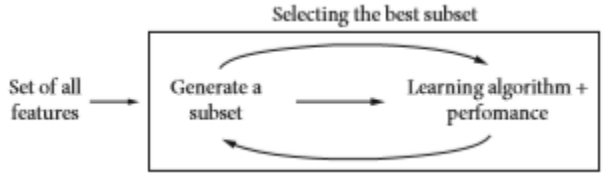

Fig 3.3 Embedded Methods

3.1.2 Stage 2: Features Selection
Usually, feature selection is applied as a preprocessing step before the actual learning. However, no algorithm can make good predictions without informative and discriminative features; therefore, to keep the most significant features and reduce the size of the dataset, we implemented PCA using randomized SVD [28]. The module used for feature selection was implemented in using the Python scikit-learn library. All selection strategies were based to many criteria to extract the best features. In our work, feature selection was based on the following modules: removing features with low variance, univariate feature selection, and recursive feature elimination.

3.1.3 Stage 3 : Machine Learning Algorithm
Usually, ensemble machine learning algorithms allow better predictive performance compared with a single model. This can be considered machine learning competition, where the winning solution was used as a model for breast cancer diagnosis. In this paper, the following heterogeneous ensembles machine learning algorithms were used to classify the given data set: support vector machine (SVM) [29], K-nearest neighbor (KNN) [30], decision tree (DT) [31], gradient boosting classifier (GB) [32], random forest (RF) [33], logistic regression (LR) [34], Ada Boost classifier (AB) [35], Gaussian Naive Bayes (GNB) [36], and linear discriminant analysis (LDA) [37].

3.1.4 Parameter Optimization
Genetic Programming (GP) is a type of evolutionary algorithm (EA) that generalizes the genetic algorithm. GP is a model for testing and selecting the best choice among a set of results. Based on biological evolution and its fundamental mechanism (mutation, crossover, and selection), GP generates a solution. The use of GP is the reason for its flexibility; it can model systems where the structure of the desired models and the key features are not known. In this paper, GP allowed the system to search for models from a range of possible model structures and optimizing the pipelines represented in tree structures for the classification problem. GP first generates a fixed number of pipelines based on the primitives described above, such as features selection decomposition. In other words, the sequence of operators evolves to produce machine learning pipelines that are evaluated to maximize the classification accuracy. Figure 1 depicts an example of a machine learning pipeline. After evaluation of the current pipelines machine learning, a new generation is created based on the highest previous pipelines. Each pipeline is considered an individual of GP. The GP is formed by the three main operators:

Mutation operator: changing hyper parameters or adding or removing a primitive preprocessing step such as Standard Scaler or the number of trees in a random forest.

Crossover operator: the crossover operator assumes that 5% of individuals will cross with each other using a 1-point crossover selected at random.

Selection operator: its main purpose is to select the top 20 individuals and make copies from them. To exchange information between the individuals of the population, the crossover or mutation operator can be applied. The subsequent stages of GP are given in Figure 3.2.

4. Conclusion

Original mammogram is preprocessed to remove label and pectoral muscle, then using the combination of double top-hat transforms and gray scale adjustment function to enhance micro-calcifications; subsequently, CT is adopted to removal some noises, backgrounds etc., and retain the significant information by nonlinear function; at last, we use the non-linking SPCNN to detect calcification clusters. In the first test, we proved that the three most popular evolutionary algorithms can achieve the same performance after effective configuration. The second experiment focused on the fact that combining features selection methods improves the accuracy performance. Finally, in the last experiment, we deduced how to automatically design the machine learning supervised classifier. Owing to the GP algorithm, we attempted to resolve the hyperparameter problem, which presents a challenge for machine learning algorithms. The proposed algorithm selected the appropriate algorithm from among the various configurations.

REFERENCES

[1] L. Zhengyou, G. Xiaoshan, A Segmentation Method for Mammogram X-ray Image Based on Image Enhancement with Wavelet Fusion, 2015.

[2] Z. Suhail, M. Sarwar, K. Murtaza, Automatic detection of abnormalities in mammograms, BMC Med. Imaging 15 (2015) 53.

[3] Z. Lifeng, C. Ying, Z. Fang, Z. Lu, Detection of clustered pleomorphic micro-calcifications in digital mammograms, in: Biomedical Engineering and Biotechnology (iCBEB), 2012 International Conference on, IEEE, 2012, pp. 768–771.

[4] A. J. Cruz and D. S. Wishart, “Applications of machine learning in cancer prediction and prognosis,” Cancer Informatics, vol. 2, pp. 59–77, 2006.

[5] J. G. Elmore, C. K. Wells, C. H. Lee, D. H. Howard, and A. R. Feinstein, "Variability in radiologists' interpretations of mammograms," New England Journal of Medicine, vol. 331, no. 22, pp. 1493–1499, 1994.

[6] E.Aliˇckovi´ c and A. Subasi, "Breast cancer diagnos is using GA feature selection and Rotation Forest, "Neural Computing and Applications", vol. 28, no. 4, pp. 753–763, 2017.

[7] Z. Yong, G. Dun-wei, and Z. Wan-qiu, "Feature selection of unreliable data using an improved multi-objective PSO algorithm," Neurocomputing, vol. 171, pp. 1281–1290, 2016.

[8] F. Zhang, B. Du, and L. Zhang, "Scene classification via a gradient boosting random convolutional network framework," IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 3, pp. 1793–1802, 2016.

[9] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees, Wadsworth, Belmont, CA, USA, 1984.

[10] K. P. Bennett, "Decision tree construction via linear programming," in Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society, pp. 97–101, Utica, IL, USA, 1992.

[11] Z.M. Hiraand D.F.Gillies, "Are view of feature selection and feature extraction methods applied on microarray data," Advances in Bioinformatics, vol. 2015, Article ID 198363, 13 pages, 2015.