I J C T A, 9(26) 2016, pp. 397-405

Early Detection of Cancer in MRI Lung Images using Classification and Regression Tree (CART) Method

P. Thamilselvan[1] and J.G.R. Sathiaseelan[2]

ABSTRACT

Image mining is one the leading research area in the field of computer science. In this process, lung cancer is one of the most deadly disease in human body. It is the second most dangerous disease in the world. This work is mainly planned to increase the classification accuracy rate of lung cancer tissue classification in magnetic resonance (MR) imaging and reduce the processing time using Classification and Regression Tree (CART) method. In recent years, the detection cancer in early stage is a challenging task in the field of medical. This early identification of lung tumor can develop the chance of survival among the people. In this paper, we improved Classification and Regression Tree to identify and classify MR images and this is also indented to reduce the processing time, higher classification rate and minimum error rate. This proposed work consists of two phases, such as feature extraction, classification. In the first phase, we have obtained the features of magnetic resonance images have been reduced using Principle Component Analysis (PCA) method. In the second phase, CART method has been implemented to classify subjects as benign or malignant magnetic resonance images.

Keywords: Image mining, CART, PCA, classification rate, MR images, feature extraction, detection of cancer.

1. INTRODUCTION

The early detection of cancer is very useful as a supporting decision tool to help physician in disease diagnosing in the field of medical [1]. The magnetic resonance images are mostly used in the medical field to diagnose the disease. Classification is one of the techniques used to classify the lung tissues in magnetic resonance images that is a usual problem for detecting abnormal structures in image mining. Classification methods have been applied in the analysis of functional structures involving clinical diagnosis, classification and segmentation. The classification process in machine learning and image mining is an eminent method for prediction and decision making. The various researcher addresses this kind of general problem like machine learning, artificial intelligence, decision making process, statistics and economics.

The different types of method have been used in the MR image recognition including decision tree [2], k nearest neighbor [3-5], linear discriminant analysis [6] and naive bayes [7]. Although much development is still difficult to achieve acceptable result in practical applications. Most of the researcher have shown that the different classifiers are achieves better classification result. The identification of lung tumor can be achieve in many ways, such as x-ray, magnetic resonance images and computed tomography. All these process consume a lot of resources in terms of money, time and accuracy. The researcher have proven that the non-invasiveness method of malignant cell analysis can assist in the successful diagnosis of lung cancer [8]. In this paper, we improved Computer Aided Diagnosis (CAD) system for identifying lung diseases and

[1] Research Scholar, Department of Computer Science, Bishop Heber College, Tiruchirappalli, TN, India. *E-mail: thamilselvan1987@gmail.com*

[2] Head, Department of Computer Science, Bishop Heber College, Tiruchirappalli, TN, India. *E-mail: jgrsathiaseelan@gmail.com*

improving the classification accuracy of CART method in magnetic resonance images. The computer aided diagnosis system can play important role in early detection of lung cancer. It helps as a valuable second opinion to clinicians observe patients during lung cancer detection [9]. A computer aided diagnosis system includes combination of artificial intelligent methods and image processing that can be used to identify abnormalities in medical images to improving the better performance in identification process. The CAD method could improve the efficiency of the diagnosis process by identifying lung cancer patients effectively and increase the diagnosis process. The process can be summarized as follows

1. Feature extraction using PCA method
2. Detection of malignant cell using classification and regression tree method;
3. Identify the accuracy of CART method to be used in the diagnosis process.

The objective of this work is improve the classification accuracy and reduce the processing time of proposed CART method. The rest of the work is structured as follows. Section 2 describes the background of the previous work in image classification analysis. Section 3 provides a details feature extraction using PCA method. Section 4 describes detection and classification of classification and regression tree method. Section 5 compares the proposed method about the different classification algorithms. Lastly, the conclusion and forthcoming works are discussed in section 6.

2. BACKGROUND WORK

Antonio *et al.* [10] present a method called support vector machine to developing automatic detection of lung nodules. The lung cancer nodules produce maximum death ratio in addition to one of the lowest survival ratio after disease diagnosis. The proposed support vector machine were achieves 97% of classification accuracy, 97% of specificity and 85% of sensitivity with false positive rate of 1.82.

Yang *et al.* [11] proposed a system based on the naive bayes classifier method for whole breast lesion detection. The detection rate of naive bayes method is achieves 92.1 %. This mechanism can improve the robustness and sensitivity of detection if the target lesion appears.

Jinsa *et al.* enhanced neural network method to classify lung cancer images for CT images. The whole lung is segmented from the parameters and the computed tomography images are measured from the segmented image. The neural network training function gives 93.3% classification accuracy [12].

The computerised lung nodule detection system can help identify the lung abnormalities in computed tomography images [13]. Lee *et al.* implemented random forest method based nodule classification by using clustering technique. This method shows more than 95% accuracy in their performance.

Magna *et al.* [14] considered the properties of a classifier based on group of artificial immune networks to identify mammography abnormalities for breast cancer detection. The proposed artificial immune networks shows that 90% accuracy in it classification performance.

3. EARLY DETECTION

Detection of early level lung disease is a challenging task in medical field. Automated disease diagnosis based on machine learning method it could be significant for early detection process of patients The image are collected from various hospitals in tiruchirappalli. The MRI images of more than 50 patients are collected considered between the ages of 18 to 65 years. The collected image is shown in figure 1.

The lung is detected from the MR images using classification and regression tree method. Cancer detection in preliminary level is directly linked with survival rate. Various biomarkers have been examined and classified

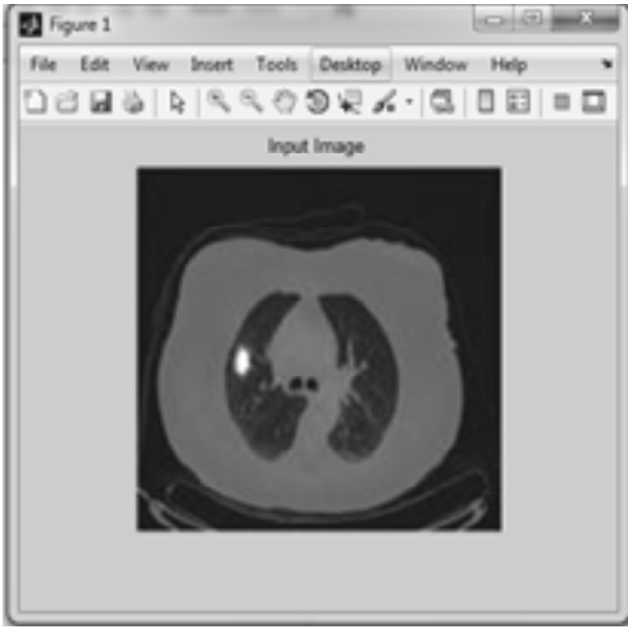

Figure 1: Collected input image

for monitoring alteration inside the cancerous tissues. The lung cancer may be seen in MR images, but disease is confirmed by operation which is usually performed by cytologist. It plays important role early detection of tumors. Feature extraction, cancer identification and classification are the main methods in this early detection process.

4. FEATURE EXTRACTION

Feature extraction methods evaluates the various medical images to extract the most important features of the image. In this work, Principal Component Analysis (PCA) has been used for extracting the features from the MRI lung images. The extracted features will be used in the diagnostic analysis for detecting the cancer cells from lung cancer images. The important problem in early diagnosis of lung cancer is associated with the skill of the CAD system to differentiate between benign and malignant cells. Accordingly, using this appropriate features we can eliminate or reduce number of misclassifications rates. In the background work, different method have been proposed depending on the implemented method. In our proposed work, we used the following features such as density, ration, border, curves and complexity. The extracted image by using principal component analysis is shown in figure 2.

A. Principal Component Analysis

The principal component analysis is well recognized tools for converting the input features into a new lower-dimension space. The Independent Component Analysis (ICA) and Principal Component Analysis

Figure 2: Various stages of feature extraction using PCA

(PCA) are two tools for converting the input features into lower dimensional feature space. In this work, PCA method has been implemented to extracting the image features. It is the most widely used technique to produce optimal resolution with low computational complexity. The main idea of this method is to minimize the dimensionality of the image and to improve the results more efficient and accurate classifier.

In overall, the exact feature extraction algorithm makes the classification process more efficient and effective. In this work, we investigate the effectiveness of Principal component analysis for feature reduction on the lung cancer detection and classification problem. The pseudo-code of PCA is given below:

```
Function [patterns, targets, UW, m, W] = PCAn (patterns, targets, dimension)
[r, c] = size (patterns);
If (r < dimension),
  Disp ('required dimension is larger than the data dimension.')
  Disp (['Will use dimension ' num2str(r)])
  Dimension = r;
End
m          = mean (patterns')';
S          = ((patterns - m*ones (1, c)) * (patterns - m*ones (1, c))');
[V, D]     = eig(S);
W          = V (:, r-dimension+1: r)';
U          = S*W'*inv (W*S*W');
 UW        = U*W;
Patterns   = W*patterns;
```

The size of the input matrix is reduced based on the dimension. The following steps are involved in extracting the PCA method of the input vector.

Step 1: Reshape the data points using Principal Component Analysis (PCA). If the required dimension is larger than the data dimension we can use dimension.

Step 2: Calculate the cov matrix and PCA matrixes.

Step 3: Calculate new patterns.

5. DETECTION AND CLASSIFICAITON

The classification and detection techniques plays a vital role in medical imaging, particularly in the detection and classification cancers [15]. This lung cancer classification is a serious task for a computer aided diagnosis system because it is the last step in a system to attain the good outcomes on the available features [16]. In this study we implemented Classification and Regression Tree (CART) algorithm for classification process and detecting cancerous cells.

A. Classification and Regression Tree

Classification and regression tree is a nonparametric arithmetic method and it produce multilevel structure of a tree. A common design of classification and regression tree output is presented in figure 3. The C&RT begin with one 'node', having the whole sample, called a root node. This method reviews all possible splits and selects the one from binary groups that is differ from other splitting variable. The root node then divided into two child node based on selected independent variable. Classification and regression tree only splits root node into two sub nodes.

The CART algorithm developed to provide information about the cancer tissues problem as benign tissues and malignant tissues in the input image dataset. It is important to provide additional information

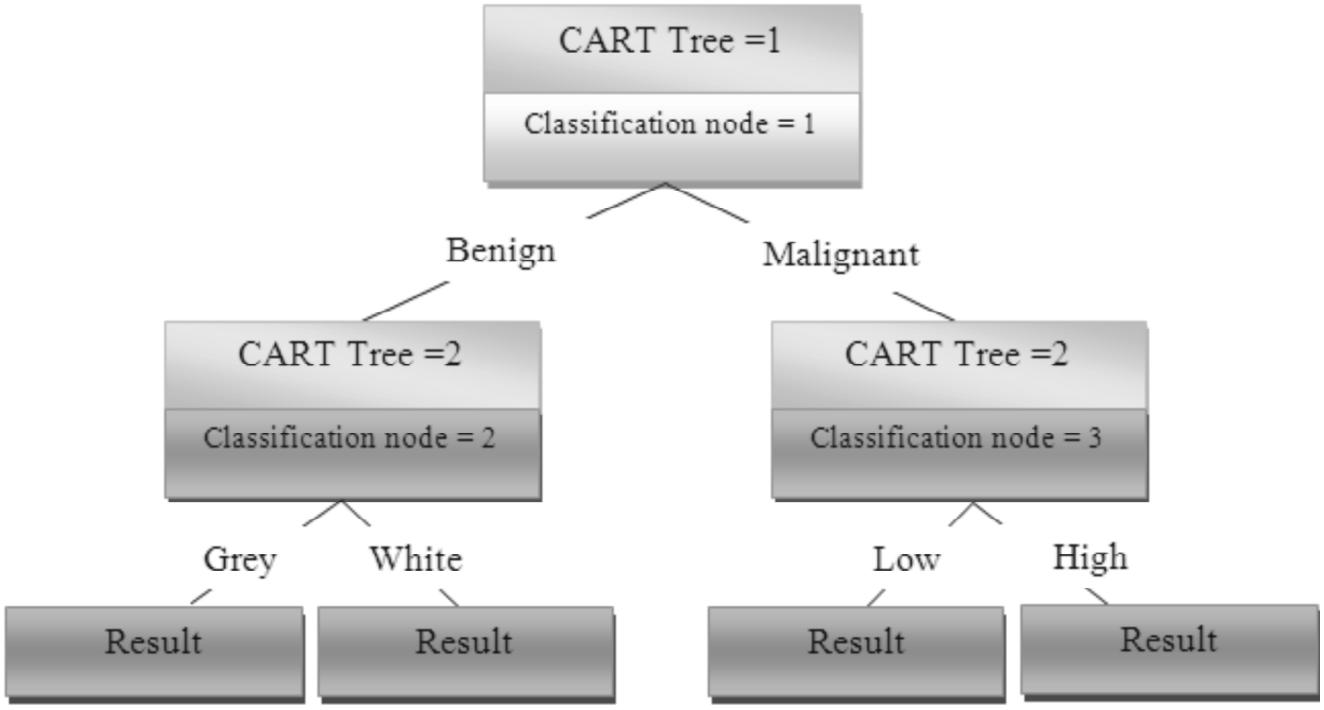

Figure 3: Block diagram of Classification and Regression Tree

about cancer tissues and achieve the high classification result. In order to find this cancer, the second node composed by second CART system. It is separated into two classification sub node: second classification node and third classification node. In order to identify information about level of cancer, the second classification node is implemented. The chance to discover in detail of the cancer can be identified in this classification node. In third classification node, benign (normal) tissue can be divided into benign grey and benign white.

Classification can be done by classification and regression tree (CART) by using following steps.

1. Input: features – train features

   Targets – train targets

   Parameters – misclassification or Gini (Variance)

   Region – decision region vector
2. Get the parameters
3. Create decision region
4. Preprocessing using PCA (Principal Component Analysis)
5. Build the tree recursively
6. Make the decision region according to the tree
7. Output: Decision Support

The main Pseudocode of Classification and Regression Tree is given below. To identify the performance of this proposed method, it is important to describe sensitivity, specificity and classification accuracy. The classification accuracy is calculated by total number of correctly classified samples divided by the total number of test samples.

```
Function delta = CART functions (split point, patterns, targets, dim, split type)
Uc    = unique (targets);
For i = 1: length (Uc),
in= find (targets == Uc(i));
Pr (i)= length (find (patterns (dim, in) >   split_point))/length (in);
Pl (i)= length (find (patterns (dim, in) <= split_point))/length (in);
end
switch split_type,
case 'Entropy'
Er              = sum (-Pr.*log (Pr+eps)/log (2));
El              = sum (-Pl.*log (Pl+eps)/log (2));
Case {'Variance', 'Gini'}
Er              = 1 - sum (Pr. ^2);
El              = 1 - sum (Pl. ^2);
case 'Missclassification'
Er              = 1 - max (Pr);
El              = 1 - max (Pl);
otherwise
error (‘possible splitting rules are: Entropy, Variance or Gini, Missclassification')
end
P               = length (find (patterns (dim, :) <= split_point)) / length (targets) ;
delta = -P*El - (1-P)*Er;
```

6. RESULT AND DISCUSSION

The experimental result used to validate the classification performance of proposed C&RT technique using magnetic resonance images. We accompanied a complete set of experiments to investigate the result of the detection classification process on the cancer tissues extraction. In the detection process, among more than 50 lung nodules were detected by MRI image respectively. The example of lung images are shown in Figure 1. In the figure, left lung nodule affected by the cancer. Example of preprocessing and feature extraction outputted by the principal component analysis method are shown in Figure 2. In this work, the detection of nodule for MR images was performed automatically by using MATLAB 10.0 software. The figure 4 shows detected cancer tissues in input image by using classification and regression tree.

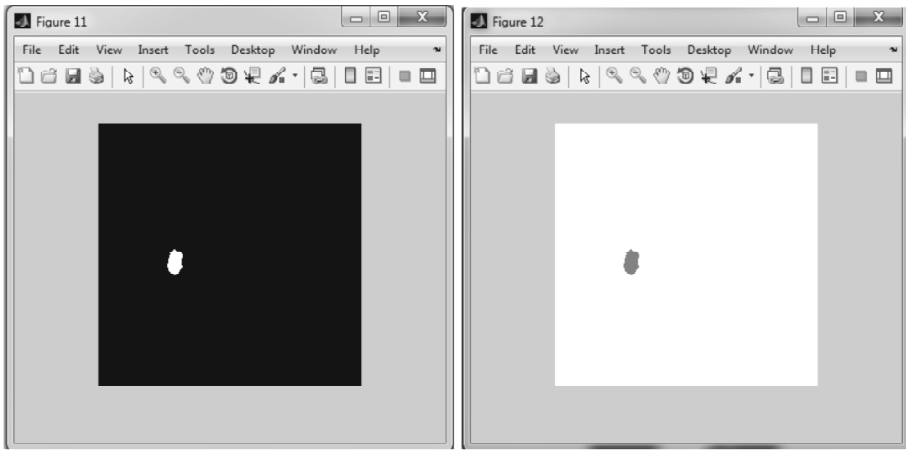

Figure 4: Detected cancerous tissues by using CART method

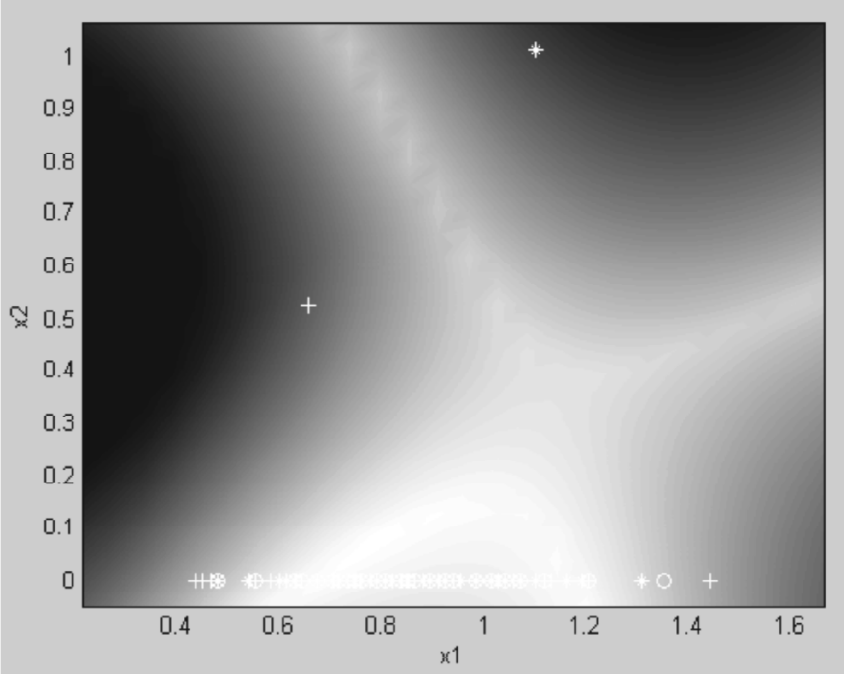

Figure 5: Classification result benign and malignant tissues

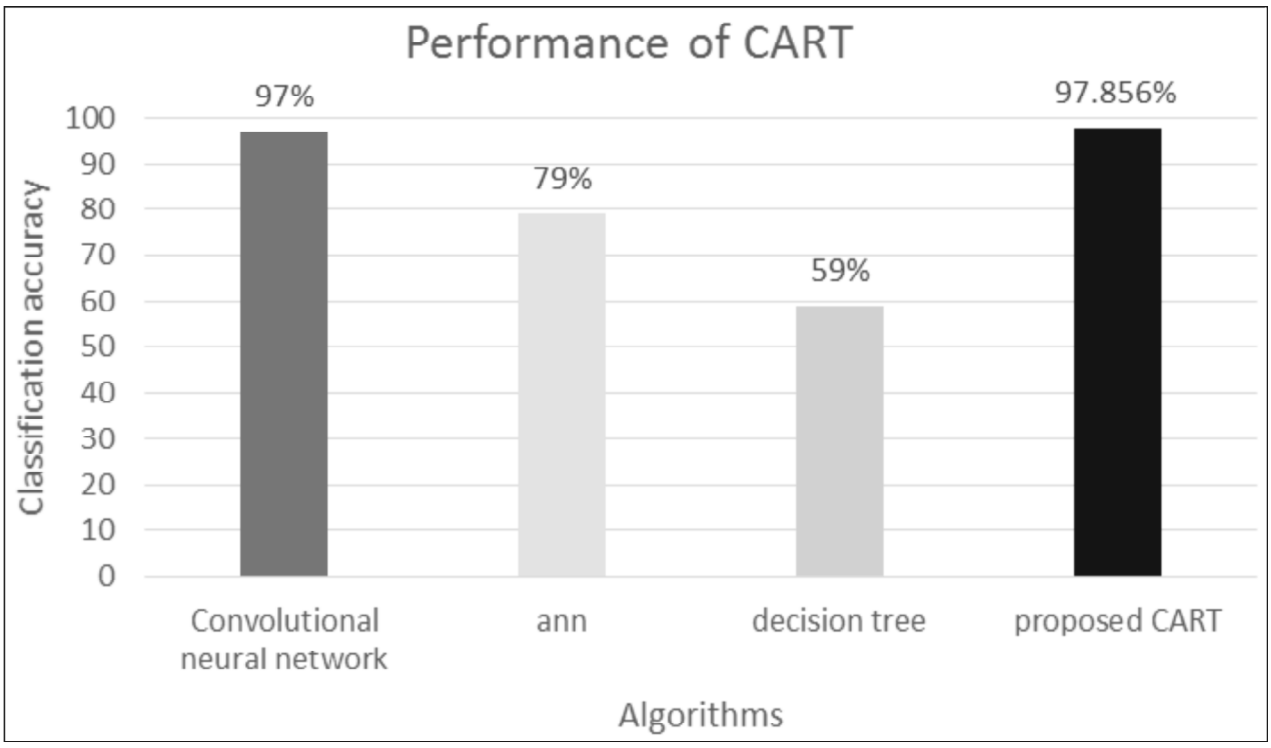

Figure 6: Performance analysis of proposed CART method.

The proposed method yields almost good detection result from the input image. The number of detected tissues in the opening identification was 35 from total 60 images and its classification accuracy is 97.856%, so good result was achieved.

The result of our proposed method for the detection of benign tissues and malignant tissues is shown in Figure 5. In the figure, + symbols indicates benign tissues and * symbols indicates malignant tissues in the input image.

We compared our proposed algorithm with different method to analysis performance of our method, it is shown in Figure 6. From this experimental result, our proposed Classification and Regression Tree method achieves better result when compare with CNN [17], artificial neural network [18] and decision tree [19].

CONCLUSIONS

In this research, the Classification and Regression Tree has been proposed to identify the cancer tissues and to improve the classification accuracy by using MRI images. From this research. The developed algorithm for early detection of lung cancer tissues shows better classification result when compare with other algorithms. The proposed CART algorithm yields 97.856% accuracy in their classification performance and it should thus be supportive for identifying early level lung cancer tissues detection. In future, this algorithm will be applied in large amount of data without accuracy loss and minimize the processing time.

REFERENCES

[1] Shah, A., Desai, SR., "Imaging in chest disease" *Medicine*, volume, **40(4)**, 177-185, 2012.

[2] Zrimec, T., Wong, J. S., "Improving computer aided disease detection using knowledge of disease appearance" *In Medinfo Proceedings of the 12th World Congress on Health Informatics; Building Sustainable Health Systems*, 1324-1324, 2007.

[3] Korfiatis, P. D, Anna N. Karahaliou, Alexandra D. Kazantzi, Cristina Kalogeropoulou, Lena I. Costaridou., "Texture Based Identification and Characterization of Interstitial Pneumonia Patterns in Lung Multidetector CT" *IEEE Trans. Inf. Technol. Biomed*, **14(3)**, 675-680, 2010.

[4] Nuzhnaya, T., Vasileios Megalooikonomou, Haibin Ling, Mark Kohn, Robert Steiner., "Classification of Texture Patterns in CT Lung Imaging" *SPIE Medical Imaging*, 796336, 2011.

[5] Sørensen L, Lo P, Ashraf H, Sporring J, Nielsen M, de Bruijne M., "Learning COPD Sensitive Filters in Pulmonary CT" *MICCAI*, 699-706, 2009.

[6] Way, T. W, Sahiner B, Chan HP, Hadjiiski L, Cascade PN, Chughtai A, Bogot N, Kazerooni E., "Computer-aided diagnosis of pulmonary nodules on CT scans: improvement of classification performance with nodule surface features" *Medical physics*, **36(7)**, 3086-3098, 2009.

[7] Song L, Xiabi Liu, Ling Ma, Chunwu Zhou, Xinming Zhao, Yanfeng Zhao., "Using HOG-LBP features and MMP learning to recognize imaging signs of lung lesions" *In Computer-Based Medical Systems (CBMS)*, 1-4, 2012.

[8] Fatma Taher, Naoufel Werghi, Hussain Al-Ahmad., "Computer Aided Diagnosis System for Early Lung Cancer Detection" *Algorithms*, **8**, 1088-1110, 2015.

[9] El-Baz, A, Beache, G.M, Gimel'farb, G, Suzuki, K, Okada, K, Elnakib, A, Soliman, A, Abdollahi, B., "Computer-Aided Diagnosis Systems for Lung Cancer: Challenges and Methodologies" *International Journal of Biomedical Imaging*, doi:10.1155/2013/942353, 2013.

[10] Antonio Oseas de Carvalho Filho, Wener Borges de Sampaio, Aristofanes Correa Silva, Anselmo Cardoso de Paiva, Rodolfo Acatauassu Nunes, Marcelo Gattass., "Automatic detection of solitary lung nodules using quality thresholdclustering, genetic algorithm and diversity index" *Artificial Intelligence in Medicine*, **60**, 165–177, 2014.

[11] Min-chun yan, chiun-sheng huang, jeon-hor chen, ruey-feng chang., "whole breast lesion detection using naive bayes classifier for portable ultrasound" *Ultrasound in Med.* and *Biol.*, **38(11)**, 1870–1880, 2012.

[12] Jinsa Kuruvilla, K. Gunavathi., "Lung cancer classification using neural networksfor CT images" *Computer methods and programs in biomedicine*, **113**, 202-209, 2014.

[13] S.L.A. Lee, A.Z. Kouzani, E.J. Hu., "Random forest based lung nodule classification aided by clustering" *Computerized medical imaging and graphics*, **34**, 535–542, 2010.

[14] Gabriele Magna, Paola Casti, Sowmya Velappa Jayaraman, Marcello Salmeri, Arianna Mencattini, Eugenio Martinelli, Corrado Di Natale., "Identification of mammography anomalies for breast cancer detection by an ensemble of classification models based on artificial immune system" *Knowledge-Based Systems*, **101**, 60–70, 2016.

[15] Ye X, Lin X, Dehmeshki J, Slabaugh G, Beddoe G., "Shape-Based Computer-Aided Detection of Lung Nodules in Thoracic CT Images" *IEEE Transaction Biomedical Engineering*, **56**, 1810–1820, 2009.

[16] Fatma Taher, Naoufel Werghi, Hussain Al-Ahmad., "Computer Aided Diagnosis System for Early Lung Cancer Detection" *Algorithms*, **8**, 1088-1110, 2015.

[17] Atsushi Teramoto, Hiroshi Fujita, Osamu Yamamuro, Tsuneo Tamaki., "Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique" *Medical Physics*, **43**, 2821-2827, 2016.

[18] Ling Ma, Xiabi Liu, Li Song, Chunwu Zhou, Xinming Zhao, Yanfeng Zhao., "A New Classifier Fusion Method based on Historical and On-line Classification Reliability for Recognizing Common CT Imaging Signs of Lung Diseases" *Computerized Medical Imaging and Graphics*, **14**, 1-20, 2014.

[19] Ming-Yih Lee, Chi-Shih Yang., "Entropy-based feature extraction and decision tree induction for breast cancer diagnosis with standardized thermograph images" *Computer methods and programs in biomedicine*, **100**, 269–282, 2010.