Asian Journal of Information Technology 15 (19): 3742-3747, 2016
ISSN: 1682-3915
© Medwell Journals, 2016

Investigation of Optimal Segmentation Algorithm for CT Lung Nodules Using CAD System

[1]K. Sakthivel, [1]S. Balu, [2]C. Nelson Kennedy Babu and [3]R. Balamurugan
[1]Department of CSE, K.S. Rangasamy College of Technology, Tamil Nadu, India
[2]Department of CSE, CMS Engineering College, Tamil Nadu, India
[3]Department of EEE, K.S. Rangasamy College of Technology, Tamil Nadu, India

Abstract: Computer Aided Diagnosis (CAD) has been playing a significant role in cancer detection for the past two decades. This research study mainly focuses on developing a CAD system for early detection of lung cancer with improved accuracy. The proposed system helps to reduce unnecessary biopsy and surgery. In this research, three methodologies namely Automatic Region Growing (ARG), Histogram based Earth-Mover's Distance (HEMD), Firefly Search Fuzzy C-Means (FSFC) algorithm have been developed for improving the accuracy of the computer aided diagnosis of lung cancer from CT images. The performance of three segmentation methodologies are evaluated and compared to prove that the proposed Firefly Search Fuzzy C-means methodology outperforms the existing algorithms.

Key words: Segmentation, computer aided diagnosis, region growing, earth movers distance, firefly search, fuzzy c-means

INTRODUCTION

The mortality rate due to lung cancer can be reduced if the cancer is diagnosed at an early stage. This has become possible with the tremendous growth of imaging techniques and information processing techniques which has led to the development of CAD systems to assist the radiologist in performing better diagnosis. A CAD system analyzes the data of the patient who is under investigation and suggests a diagnosis based on the analysis.

Lung cancer may be visualized as a mass or tumor on a chest computerized tomography of a patient. Tumors can be defined as benign or malignant. Benign tumors can be removed and does not extend to other parts of the body but malignant tumors are cancerous cells (Dwivedi *et al*., 2014). Recently detection of such lesions is the challenging tasks for the radiologist. Such lesions are not clearly detected because they may be masked by anatomical structure or poor quality images or poor decision criteria made by radiologist. The detection lung cancer at its early stages is very important inorder to survive patient's life because the symptoms occur in the advanced stages will make the low chances of survival.

The advancements in the medical field have led to introduce many imaging techniques for the diagnosis of lung cancer. Some imaging techniques are computed tomography, radiography, magnetic resonance imaging, and positron emission tomography. CT and radiography use ionization of radiation procedure whereas MRI and PET do not use the same (Hu *et al*., 2001). For the accurate diagnosis of the cancer, radiation procedures are best suitable. Radiologists are trained in such a way to perform radiation procedures. These procedures provide necessary information about the patient's symptoms and help the doctors to diagnosis them accurately (Zhang *et al*., 2015). The medical procedure will benefits greatly from any small risk of harm from the amount of radiation used.

Computer technology has been an aggressive impact on medical imaging. However, without computers modern radiological modalities like CT, Ultrasound, PET and MRI do not even exist. The change in the field of information technology along with medical imaging facilitates the upcoming decades are able to store large collection of images which become an important source of anatomical and research for the accurate diagnosis of cancer. Advancement of computer technology led to the development of such systems, namely CAD systems.

CAD is one of the major research areas in medical imaging and diagnostic radiology and these systems are developed as early as the 1980's. CAD systems are used by radiologists for detection and differential diagnosis of different types of abnormalities. The output of a CAD system can be used by radiologists along with laboratory

Corresponding Author: K. Sakthivel, Department of CSE, K.S. Rangasamy College of Technology, Tamil Nadu, India

results as a second opinion before making a diagnosis. CAD systems are developed based on an understanding of the radiological patterns which have to be observed in the image, for a particular disease.When a CAD system is developed, it should focus on the computer algorithms that should be developed with an eye on the type of medical image to be handled while simultaneously considering the radiological properties that the radiologist will look out for in a medical image.

It could be a great clinical importance if patient's chance of survival is increased by an effective CAD system (Parveen and Kavitha, 2012). The systems are developed to support the process of distinguishing malignant with benign lesions. The systems need to improve both the sensitivity and specificity in order to be successful for clinical use. CAD methods may use features extracted either by the computer or by the radiologist. For interpreting instantaneous data, radiologists employ many radiographic images and computer analysis is needed for determining the significant features.

In this research work the data is in the form of CT images of the lung and hence image analysis is done to perform diagnosis. The segmentation of such image is done by using the three algorithms developed in the CAD system and image processing is carried out to diagnose the abnormalities in the lungs.

Back ground : Image processing plays a vital role in diagnosis of diseases using CAD systems based on images. This research focuses on CAD systems for diagnosis of lung cancer. As with any system employing image analysis and image understanding, segmentation is one of the primary processes in a CAD system that works with images. Segmentation of lung cancer is a challenging task. The Region Of Interest (ROIs) can be extracted only if the lungs are correctly segmented (Sluimer *et al*., 2005). If the segmentation is incorrect, it may increase the false positives and false negatives thereby affecting the diagnostic accuracy of the system. Hence, it is important to use a robust and reliable technique for segmentation.

The use of Artificial Intelligence (AI) in diagnostic expert systems aims at making the systems to mimic the decision-making process of human experts. However, such systems can be used only to provide a "second opinion" to the physicians in improving diagnostic accuracy and cannot be considered as a replacement for a physician. AI techniques that are most frequently used in CAD systems in the literature are fuzzy logic, genetic algorithms, neural networks and ensemble approaches.

Image segmentation is more crucial part in obtaining information from the medical images. It is used for detecting the threatening diseases in medical field which lead to death. The lungs segmentation is a challenging problem due to the need of different scanners and scanning protocols for scanning the homogeneities in the lung region, pulmonary structures of similar densities such as arteries, veins, bronchi and bronchioles (Punithavathy *et al*., 2015). The segmentation of lung regions from CT images and chest radiographs by a particular technique can be addressed in terms of accuracy, processing time and automation level. The major classification of lung segmentation techniques are divided in to four categories: methods based on signal thresholding, deformable boundaries, shape models, edge based and region based.

To efficiently reduce the search space for lung nodules and to provide accurate segmentation some investigations are extended towards technical issues of the lung fields. The technical issues comprises of automation level of the technique, the sensitivity of the method to the scanning parameters, the competence of an algorithm to work with unusual image modalities, (e.g., CT, LDCT or CE-CT) and the capacity of the algorithm to give a proper lung segmentation in cases that are associated with the pathological lungs.

MATERIALS AND METHODS

In the present study, automatic classification of lung cancer nodules acquired by computed tomography is considered. The data is collected from the histology banks and samples are sorted across thousands of patients. These image data is pre-processed to extract the lung region alone. The ROI is determined by using segmentation algorithms and features are extracted for classifying the cancerous or non-cancerous regions.

The proposed computer aided diagnosis systems comprises of four phases for early detection of lung cancer and are as follows:

- Phase 1: Image denoising
- Phase 2: Segmentation of Lung Region and ROIs
- Phase 3: Feature Extraction from the Segmented Region
- Phase 4: Classification of malignant and benign nodules

Figure 1 shows the proposed system with all the phases of CAD system. In the initial phase of this CAD method, CT lung image is denoised using median filter. The first proposed methodology uses Automatic Region Growing algorithm for the segmentation of the lung. In the second methodology, histogram based Earth Mover's

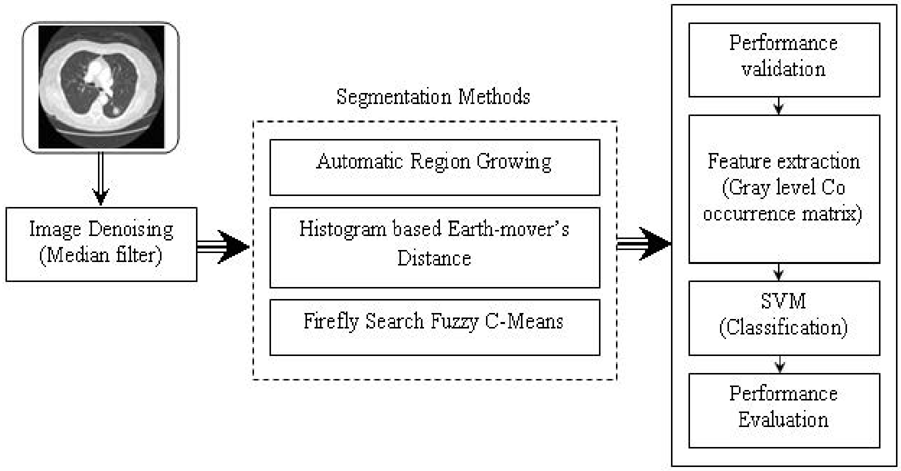

Fig. 1: Schematic flow of the proposed system

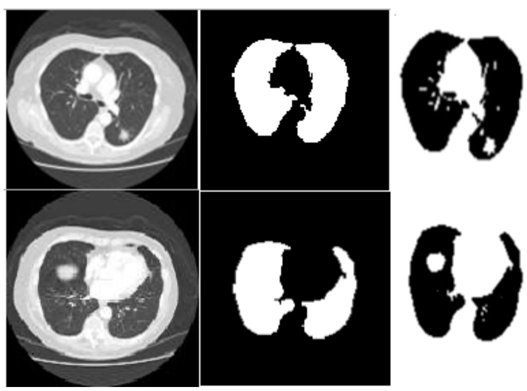

Fig. 2: a) Original image; b) Ground truth; c) Segmented lung CT images obtained using the ARG algorithm

distance algorithm is used for segmenting the lung nodules and morphological techniques are implemented to enhance the segmentation results. The third proposed methodology uses Firefly Search-Fuzzy C Means for segmentation.

Segmentation by ARG algorithm: Automatic region growing algorithm is proposed for segmenting lung CT images and it overcomes the drawback of conventional region growing techniques (Li *et al*., 2013a, b). The experiments were conducted on the proposed computer-aided diagnosis systems with the help of real time lung images. Nearly, 400 real time images of different categories of people were collected from well reputed hospital for the study. These images included both benign and malignant nodules. Malignant nodules are cancerous while benign nodules are not cancerous. Among these 400 nodules, 200 are malignant and 200 are benign. The size of the nodules ranges from 1.5-10 mm. In the first stage of this CAD system, image preprocessing technique is used to remove noise artifacts (Zhongming and Jun, 2015). The median filter is to remove salt and pepper noise which is a nonlinear digital filtering technique. Such noise reduction is a typical pre-processing step to improve the results of later processing.

Table 1: Segmentation Accuracy for CT lung images using ARG algorithm

Image	Solidity	Area	Perimeter	Total	Accuracy (%)
Image 1					
Segmented	0.8384	32364	1151.98	33516.8184	84.6499
Unsegmented	0.6820	37761	1832.94	39594.6220	
Image 2					
Segmented	0.8396	33364	1151.94	34513.7556	85.0141
Unsegmented	0.6860	38764	1832.98	40597.6660	
Average accuracy (%)					84.832

The next step is segmentation which finds all suspicious regions. The optimal thersolding converts the input image to bi-level image. The processing of image includes histogram analysis to find the threat point. The efficiency of the proposed segmentation algorithm is tested by comparing the results of the proposed algorithm with the significant properties like Solidity, Perimeter and Area. Figure 2 shows the CT images with ground truth using ARG algorithm and Table 1 gives the segmentation accuracy for CT images. The accuracy of the ARG algorithm is 84.83%.

Segmentation by HEMD algorithm: The Earth mover's distance method measures the dissimilarity between the two histograms. It can be applied to the many variable-size signatures which include histograms and also allows for partial matches in a natural way (Li *et al*., 2013b). If ground distance is metric and if the total weights of two signatures are equal, it follows a true

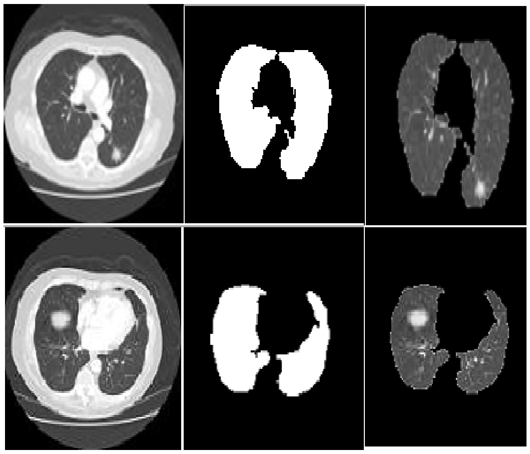

Fig. 3: a) Original image; b) Ground truth; c) Segmented lung CT images obtained using the HEMD algorithm

Table 2: Segmentation Accuracy for HEMD algorithm

Image	Solidity	Area	Perimeter	Total	Accuracy (%)
Image 1					
Segmented	0.7196	32564	1421.9230	33986.6426	90.6887
Unsegmented	0.7297	35763	1712.8054	37476.5351	
Image 2					
Segmented	0.8934	34674	1198.937	35873.8304	90.0988
Unsegmented	0.5230	37983	1832.540	39816.0630	
Average accuracy (%)					90.3937

metric. This allows endowing image spaces with a metric structure (Ling and Okada, 2007). The major components of this CAD system are Image Denoising by median filter, Segmentation using HEMD and Performance evaluation. The bright and dark regions are identified from the CT images using fuzzy logic technique. In this phase segmentation is done by using histogram based thresholding. The Earth Mover's distance measure is used for the segmentation. Morphological operations are used to extract the imperfections introduced during segmentation of the CT image.

The performance of this algorithm is evaluated and is given in Table 2. The segmentation accuracy obtained by using the HEMD algorithm for the two CT images shown in Fig. 3 is 90.39%

Segmentation by FSFC algorithm: Firefly Algorithm (FA) is a population based algorithm to find the global optima of objective functions based on swarm intelligence. Fuzzy clustering is the process of assigning the membership value and using them to assign elements to one or more clusters (Bezdek, 1981). The commonly used fuzzy method for clustering is Fuzzy C-Means (FCM) (Gong *et al.*, 2013). The aim of the objective function of FCM is to find the cluster centers and to produce class membership matrix which designates a membership to a data point, depending on the relativity of the data point to a particular class when compared to other classes. The performance of the FA in terms of obtaining near-optimal cluster centers values in the initialization phase in the FCM algorithm is investigated (Niazmardi *et al.*, 2013). A clustering approach based on Firefly Search with Fuzzy C-Means is proposed (Alomoush *et al.*, 2014). The proposed clustering method consists of two phases.

In order to determine the optimal cluster centers, firefly inspects the search space of the given dataset and then the values of the cluster centers will be obtained using the FA.

Starting the initialization of the Fuzzy C-Mean algorithm based on the evaluated results in the first phase and it overcome the drawbacks of Fuzzy C-Mean algorithm such as getting stuck in the local optimal and being susceptible to initialization sensitivity.

The performance of the FCSC algorithm is evaluated for the CT images shown in Fig. 4 and it segmentation accuracy is tabulated. Table 3 gives the segmentation accuracy for the ground truth images and it is 95.90%

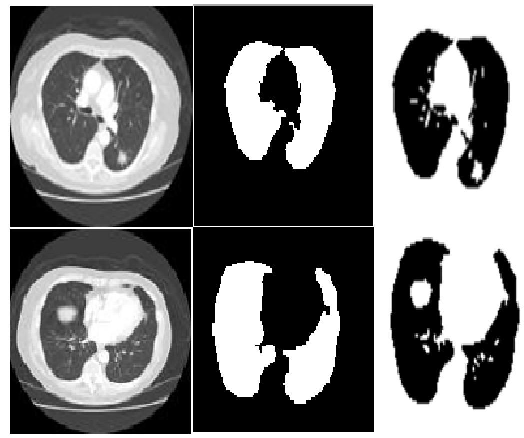

Fig. 4: a) Original image; b) Ground truth; c) Segmented lung Ground truth Images obtained after using the FCSC segmentation algorithm

Table 3: Segmentation Accuracy for proposed approach (FSFC)

Image	Solidity	Area	Perimeter	Total	Accuracy (%)
Image 1					
Segmented	0.8964	37788	1197.976	38986.8724	95.8315
Unsegmented	0.6640	38751	1830.980	40582.6440	
Image 2					
Segmented	0.9372	42573	1155.9340	43729.8712	95.9797
Unsegmented	0.6640	43732	1828.9084	45561.5724	
Average accuracy (%)					95.9056

RESULTS AND DISCUSSION

Figure 5 shows the comparison of segmentation accuracy for CT lung image 1 and 2. Figure 6 shows the optimal segmentation method compared with ARG, HEMD and FCSC algorithms. From the three proposed segmentation methodologies, FSFC methodology is proved to be the best with accuracy of 95.90%. Grey Level Co-occurrence Matrix (GLCM) is used for feature extraction and for classification, Support vector machine (SVM) is used. Support Vector Machine is utilized in this approach for automatic detection of cancer nodules. SVM is used mainly for its simplicity and accuracy. The SVM kernel function is Radial Basis Function (RBF).

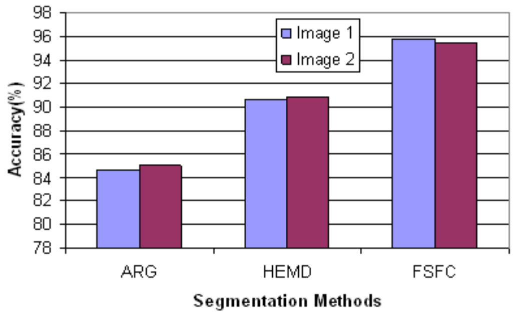

Fig. 5: Comparison of segmentation accuracy for the image 1 and 2

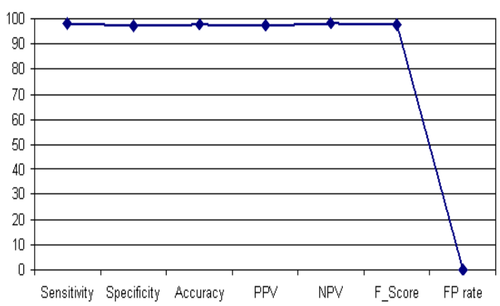

Fig. 6: Comparison of segmentation methods

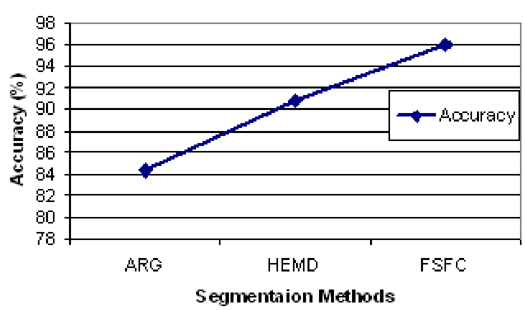

Fig. 7: Performance curve for FCSC CAD system

Table 4: Performance measures of proposed FCSC CAD system

CAD systems	Sensitivity (%)	Accuracy (%)	PPV (%)	NPV (%)	F-score (%)	FP rate (%)
FCSC CAD system	98	97	97.5	97.03	97.98	97.510.03

The classification accuracy of RBF kernel is high. The bias value and the error rate of RBF kernel are small when compared to other kernels. After the learning process is completed by providing several conditions, the proposed technique would be able to detect the presence of cancer in the lung region automatically. The performance of the proposed CAD system is evaluated using the following parameters based upon the confusion matrix:

- Accuracy
- Sensitivity
- Specificity
- Positive Predictive Value (PPV)
- Negative Predictive Value (NPV)
- F-score

Table 4 gives the performance measures parameters like Accuracy, Sensitivity, Specificity, PPV, NPV and F-Score in %, average false positive ratio obtained using RBF kernels of SVM for proposed automated CAD system and Fig. 7 shows the performance curve of the system. It is clear that when sensitivity and specificity increases, the presence of the malignant and benign nodules are classified correctly. The accuracy of the CAD system will also improve. If the sensitivity and specificity decreases, the accuracy of the CAD system will decrease which leads to misclassification of malignant and benign nodules.

CONCLUSION

In this study, three segmentation approaches viz, automatic region growing, histogram based earth-mover's distance, Firefly search fuzzy c-means algorithm have been presented for detecting the CT lung nodules. Initially the images are denoised and applied with the segmentation methods. The FSFC segmentation method is identified as the best when compared with the accuracy obtained during the segmentation of image by other methods. An automated CAD system is developed by using FCSC algorithm. The features are extracted from the resulted image of segmentation using grey correlation matrix and RBF based SVM kernel function is implemented. The performances are evaluated and the accuracy achieved is 97.5% with the average false positive ratio of 0.03%. From the observation, it is proved that the proposed CAD system performs better when compared with the other state-of-art CAD systems.

REFERENCES

Alomoush, W.K., S.N.H.S. Abdullah, S. Sahran and R.I. Hussain, 2014. Segmentation of MRI brain images using FCM improved by firefly algorithms. J. Applied Sci., 14: 66-71.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, Norwell, MA., USA., Pages: 256.

Dwivedi, S.A., R.P. Borse and A.M. Yametkar, 2014. Lung cancer detection and classification by using machine learning and multinomial Bayesian. IOSR. J. Elect. Comm. Eng., (IOSR.-JECE.), 9: 69-75.

Gong, M. Y. Liang, J. Shi and W. Ma and J. Ma, 2013. Fuzzy c-means clustering with local information and kernel metric for image segmentation. Image Proc. IEEE. Trans., 22: 573-584.

Hu, S., E.A. Hoffman and J.M. Reinhardt, 2001. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE. Trans. Med. Imaging, 20: 490-498.

Li, C., L. Yang, Z. Liu and K. Li, 2013a. A new automatic seeded region growing algorithm. Proceedings of the 2013 6th International Congress on Image and Signal Processing (CISP), December 16-18, 2013, Hangzhou, China, IEEE, ISBN: 978-1-4799-2763-0, pp: 543-549.

Li, P., Q. Wang and L. Zhang, 2013b. A novel earth mover's distance methodology for image matching with gaussian mixture models. Proceedings of the IEEE International Conference on Computer Vision, December 1, 2013, IEEE, Seoul, South Korea, pp: 1689-1696.

Ling, H. and K. Okada, 2007. An efficient earth mover's distance algorithm for robust histogram comparison. IEEE. Trans. Pattern Anal. Mach. Intell., 29: 840-853.

Niazmardi, S., S. Homayouni and A. Safari, 2013. An improved FCM algorithm based on the SVDD for unsupervised hyperspectral data classification. IEEE. J. Sel. Top. Appl. Earth Obs. Remote Sens., 6: 831-839.

Parveen, S.S. and C. Kavitha, 2012. A review on computer aided detection and diagnosis of lung cancer nodules. Int. J. Comput. Technol., 3: 393-400.

Punithavathy, K., M.M. Ramya and S. Poobal, 2015. Analysis of statistical texture features for automatic lung cancer detection in PET/CT images. Proceedings of the 2015 International Conference on Robotics Automation Control and Embedded Systems (RACE), February 18-20, 2015, IEEE, Chennai, India, pp: 1-5.

Sluimer, I., M. Prokop and B.V. Ginneken, 2005. Toward automated segmentation of the pathological lung in CT. IEEE. Trans. Med. Imaging, 24: 1025-1038.

Zhang, S., J. Huang, J. Zhang, Y. Liu and Y. Wang *et al*., 2015. Automatic immunoscoring of immunohistochemistry images of human lung cancer tissue samples: A screening tool. Proceedings of the 2015 Seventh International Conference on Measuring Technology and Mechatronics Automation, June 13-14, 2014, IEEE, Nanchang, China, ISBN: 978-1-4673-7142-1, pp: 217-220.

Zhongming, L. and W. Jun, 2015. The image segmentation algorithm of region growing and wavelet transform modulus maximum. Proceedings of the 2015 5th International Conference on Instrumentation and Measurement Computer Communication and Control (IMCCC), September 18-20, 2015, IEEE, Qinhuangdao, China, ISBN: 978-1-4673-7722-5, pp: 1171-1174.