Automated Computer-Aided Lung Cancer Detection System

Ananya Bhattacharjee and Swanirbhar Majumder

Abstract Cancer is one of the most deadly diseases in the world. Cancer of the lung or lung cancer is one of the most commonly occurring variants of cancer. This is due to an increase in pollution, presence of harmful gases in the environment and work area, smoking etc. Lung cancer can be diagnosed in a better way only if pulmonary lung nodules are detected at an early stage of growth. Popularly used techniques used for assisting the computer-aided systems in various stages of image processing include image acquisition, pre-processing, candidate nodule detection, segmentation and post-processing. Several pre-processing techniques are employed like image enhancement, smoothing, edge detection and region of interest selection. The main goal of this paper is to establish such an efficient automated computer-aided detection system which is able to detect malicious nodules in the same manner as it is detected manually by a doctor. Soft computing will be employed for the decision-making process.

Keywords Lung cancer · Computer-aided detection · Nodule detection · LIDC-IDRI · CT images

1 Introduction

Lungs are the most important organs of our respiratory system. Lung cancer is the most dangerous and brutal disease. It is a disease whereof the many cells present in the lungs some abnormal cells rapidly multiply and grow into a tumour. Most of the lung cancer cases are diagnosed late because most of the people are not aware of its early symptoms. As they are camouflaged under common symptoms like

A. Bhattacharjee
Department of ECE, NERIST, Nirjuli, Auranachal Pradesh, India
e-mail: ananya0409@gmail.com

S. Majumder (✉)
Department of IT, Tripura University, Suryamaninagar, Tripura, India
e-mail: swanirbhar@ieee.org

Coughing/Shortness of breath, etc. Lung cancer is frequently diagnosed in advanced stage where probability of survival is too low. As a result, effective treatments are not possible on time. The major causes of lung cancer are related to the usage of tobacco and cigarette and among them, 80% of the total number of cases is due to tobacco usage. Cancer of the lung represents both the highest mortality rate as well as the smallest survival rates after diagnosis.

1.1 CAD System

CAD system helps doctors to correctly interpret medical images more accurately. This system is of two types: computer-aided detection (CADe) system and computer-aided diagnosis (CADX) system. The former is responsible for detecting lesions with the help of medical images whereas the later is responsible for measuring various lesion characteristics like determination of malignancy and stages of cancer. The main goals of CADe system are diagnosis accuracy, early detection of cancer and minimal time in evaluation by the radiologists.

Figure 1 represents various stages of Computer-aided diagnosis (CAD) system.

Pre-processing is the first stage of CAD system which consists of several methods like image enhancement, smoothening, edge detection and so on. The main purpose of nodule detection stage is to properly identify the presence of suspicious nodules in the analyzed images because patient's survival rate may be increased if these lung nodules are detected in an early stage but this is quite a tedious task. The segmentation substage is applied to separate the target region from other organs. Post-processing stage includes feature selection and classification.

1.2 Soft Computing

It is a branch of artificial computational intelligence which is able to analyze varied complex medical data by employing different kinds of optimization techniques in order to improve the diagnosis and detection of cancerous nodules. Its main methodologies are Fuzzy logic (FL), evolutionary algorithms such as Genetic algorithm (GA) and artificial neural networks (ANNs).

Fig. 1 CADe system block diagram
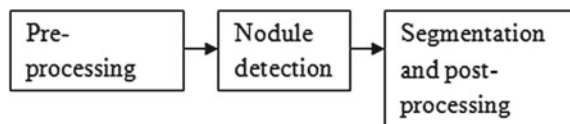
2 Review of Previous Work on Lung Nodules

Over more than a decade, various efforts have been done to develop an automated system which can detect suspicious lesions in thoracic CT and other types of imagery. In 1998, CT screening devices as well as a filtering technique known as "N-Quoit filter" are used [1]. In 2001, 2D as well as 3D feature analysis technique and linear discriminant-based classifier are used to differentiate actual nodules from false positive (FP) candidates [2, 3]. In 2002, directional gradient concentration features were developed to reduce FP generated [4]. In 2005, a surface normal overlap technique and lantern transform were developed in order to form a feature vector where rule-based classifier technique was processed to classify nodules and non-nodules [5]. In 2007, a dot enhancement filter was implemented for the selection of nodule candidate and a neural classifier was also used in order to reduce FP [6]. In 2009, temporal subtraction image was developed by implementing a technique based on artificial neural networks for detecting lung nodules [7]. In 2010, intensity thresholding as well as morphological processing were used for the detection of nodule candidates [8]. In 2011, Hopfield Neural Network (HNN) and Fuzzy C-Mean (FCM) clustering algorithm were used for segmentation in order to detect lung cancer. This will improve the survival rate of the patient. HNN provides better classification result than FCM [9]. In 2013, mean-shift methodology and geometric properties-based techniques like region of interest (ROI) were implemented [10]. In 2014, extraction and enhancement of pulmonary parenchyma were carried out and then nodule candidates are segmented. Micro-genetic algorithm was also employed in order to find out the best training model and SVM for final classification [11]. In 2015, much rarer larger nodules greater than 10 mm were mainly focused [12]. The main focus of this system was to include those large nodules which are attached to the pleural wall through the process of morphological processing. In 2016, the system presented cloud-based database was presented for detection of pulmonary nodules which is mainly characterized by 3D texture attributes [13]. Not only Structured Query Language (NoSQL) approach was developed which mainly consists of 838 nodules, 379 exams, 8237 images comprising of 4029 CT scans and 4208 manually segmented nodules. In 2017, a new segmentation algorithm named as PropSeg was proposed [14] which comprises of pre-processing, candidate detection by three levels of fuzzy C-means (FCM) clustering, segmentation and post-processing performed by morphological edge detection method. This system performed better than other techniques and so, it is suitable for lung-related disease detection. In [15], pattern recognition techniques were used. To classify malignant and benign nodules, phylogenetic diversity technique was used. To select the best model, genetic algorithm was used. It achieved accuracy of 95.52%, sensitivity of 93.1% and specificity of 92.26%. In [16], on the basis of detection accuracy, various lung cancer detection techniques were listed which were analyzed step by step and then overall limitations were pointed out.

Table 1 Database specifications

Sl. no.	Specifications	TAG description
1.	Manufacturer	GE medical systems
2.	Patient ID	LIDC-IDRI-00061
3.	Body part examined	Chest
4.	Scan options	Helical mode
5.	Bits allocated	16
6.	Bits stored	16
7.	Image type	Original/primary/Axial
8.	Manufacturer model name	Lightspeed 16
9.	Filter type	Body filter
10.	Modality	CT
11.	Slice thickness	2,500,000
12.	Exposure time	1160

3 Database Used

LIDC-IDRI is the database used which is an association of both the Lung Image Database Consortium (LIDC) and the Image Database Resource Initiative (IDRI). All the CT images are in the form of Digital Imaging and Communications in Medicine (DICOM). Each image is having a dimension of 512×512. It consists information about the nodule markings and size of the nodules is in between 3 and 30 mm. It contains 888 thoracic CT scans and section thickness is about lesser than or equal to 2.5 mm.
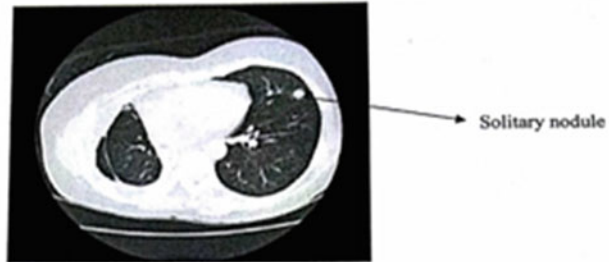
Its specifications are listed in Table 1.

4 Methodology

Working methodology is based on two approaches: manual detection and automated computer-aided detection. Manual detection is based on the marking of suspicious nodule which is certified by a doctor. This is shown in Fig. 2. The main aim of automated detection is to train CADe system in such a manner that malicious nodules detected by both doctor and automated system resemble each other.

Block diagram of the working procedure is shown in the above Fig. 3. After pre-processing, segmentation and feature extraction will be implemented. Segmentation will be implemented by applying multilevel thresholding and then several features like area, convex area, perimeter, solidity, etc., of the suspicious nodules will be extracted. Among various soft computing techniques, artificial neural

Fig. 2 Certification of suspicious nodules by a doctor

network (ANN) will be implemented for decision-making process in future [9]. ANN is useful for the development of algorithms for complex pattern recognition.

5   Result

Pre-processing is the initial stage of CAD system which includes image enhancement, smoothing, edge detection and Region of Interest (ROI) selection. Image enhancement is based on two approaches:

1. Contrast Enhancement with Morphological opening.
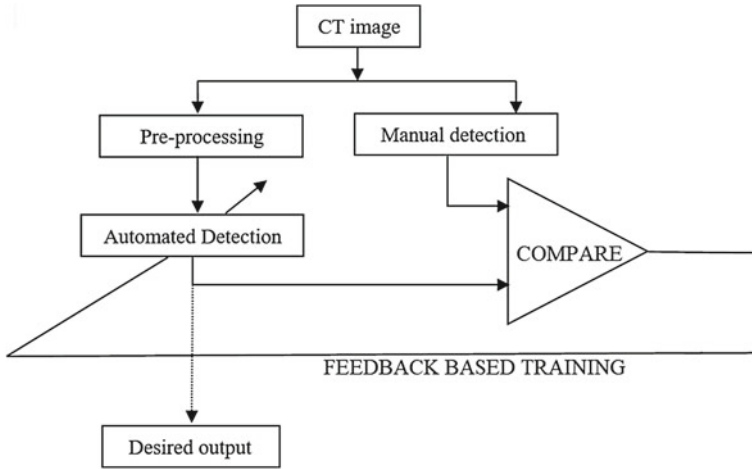2. Histogram equalized Enhanced image.

Fig. 3 Soft computing-based training of automated detection via manual detection data

The above two techniques are compared. Histogram equalization technique shows better enhancement result. Both of these techniques are clearly represented in Fig. 4a, b, respectively. Smoothing is obtained with the help of median filter. It is a non-linear filter that reduces noise in an image. Figure 4c represents smoothened image. Edge detection is performed on four operators: Sobel, Prewitt, Roberts and Canny. Among these, Canny edge detector provides the best edge details. Although Prewitt is simpler than Sobel, it produces noisier results which are not desirable. Roberts faces the problem of symmetry and detection of those edges which are multiples of 45°. Figure 5 represents all edge detection techniques.

ROI extraction algorithm is based on morphological filters which comprise operators of four types: erosion, dilation, opening and closing. All input CT image is fed into the morphological reconstruction block comprising of a marker and a mask as shown in Fig. 6a. This technique resulted in removal of those textures which are devoid of strong edges. So, these edges are extracted by applying canny edge detec-
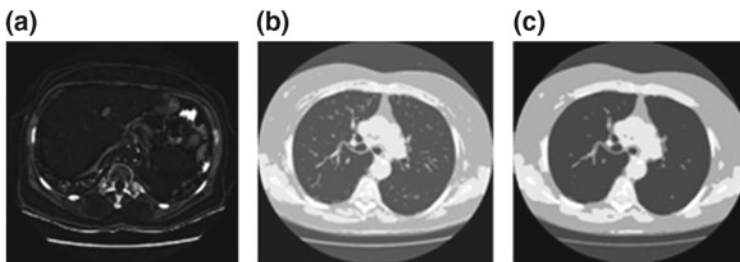

Fig. 4 Pre-processing stage **a** contrast-enhanced with morphological opening **b** histogram equalized image **c** median filter smoothened image
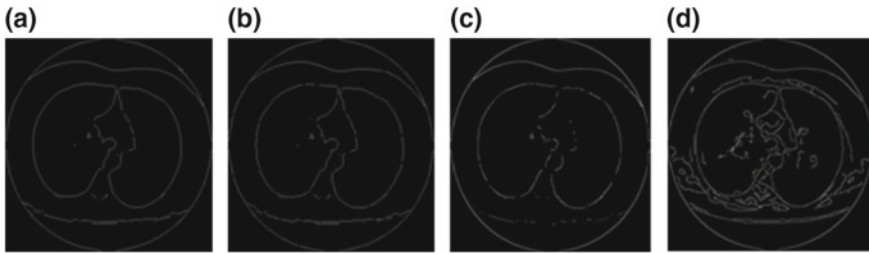
Fig. 5 Edge detection **a** Sobel **b** Prewitt **c** Roberts **d** Canny
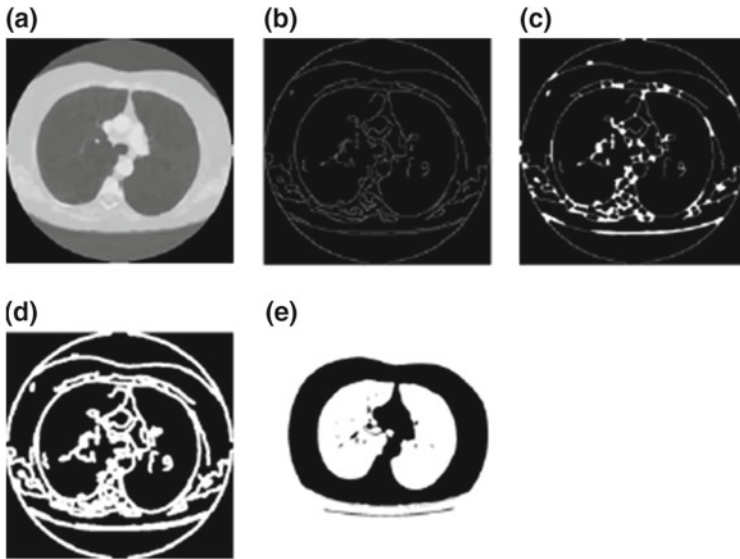

Fig. 6 ROI selection **a** reconstructed image **b** canny edge detected **c** morphologically closing operation on edge detected image **d** morphologically filtered image **e** ROI-extracted image obtained with threshold of 830 HU

tion. This is shown in Fig. 6b. However, some gaps in the edges are observed which is filled by morphological closing operation as shown in Fig. 6c Also, morphological filtering is applied to smooth the resultant image. This is shown in Fig. 6d. Finally, extraction of ROI is obtained by applying a threshold of 830 HU as shown in Fig. 6e.

6 Conclusion

Lung cancer presents the disease having one of the smallest survival rates after diagnosis. So, early diagnosis is much needed in this field to improve the condition of the patients suffering from this disease. Therefore, detection in initial stage has a

probability of being cured whereas at advanced stages it is life-threatening. The present work proposes a methodology for automatic detection of lung nodules using a combination of median filtering based smoothening, image enhancement, edge detection, histogram equalization in the pre-processing stage and then applying morphological operations to extract ROI. The authors are planning to perform segmentation and post-processing in the later stage with the help of artificial neural network.

References

1. Okumura T, Miwa T, Kako J, Yamamoto S, Matsumoto M, Tateno Y, Iinuma T, Matshmoto T (1998) Variable N-Quoit filter applied for automatic detection of lung cancer by X-ray CT. Comput Assist Radiol Surg (CAR98) 242–247
2. Armato III S, Giger M, Moran C, Blackburn J, Doi K, MacMahon H (1999) Computerized detection of pulmonary nodules on CT scans. RadioGraphics 19(5):1303–1311
3. Armato III S, Giger M, MacMahon H (2001) Automated detection of lung nodules in CT scans: preliminary results. Med Phys 28:1552–1561
4. Nappi J, Yoshida H (2002) Automated detection of polyps with CT colonography: evaluation of volumetric features for reduction of false-positive findings. Acad Radiol 386–397
5. Rubin G, Lyo J, Paik D, Sherbondy A, Chow L, Leung A, Mindelzun R, Schraedley-Desmond P, Zinck S, Naidich D (2005) Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. Radiology
6. Gori I, Fantacci M, Preite Martinez A, Retico A (2007) An automated system for lung nodule detection in low-dose computed tomography. In: Giger ML, Karssemeijer N (eds) Proceedings of the SPIE on medical imaging. computer-aided diagnosis, vol 6514, p 65143R
7. Miyake N, Kim H, Itai Y, Tan JK, Ishikawa S, Katsuragawa S (2009) Automatic detection of lung nodules in temporal subtraction image by use of shape and density features. In: Liao P, Bin-Yih, S (eds) Fourth international conference on innovative computing, information and control (ICICIC). Institute of Electrical and Electronics Engineers, pp 1288–1292
8. Messay T, Hardie RC, Rogers SK (2010) A new computationally efficient CAD system for pulmonary nodule detection in CT imaging. Med Image Anal 14:390–406
9. Taher F, Sammouda R (2011) Lung cancer detection by using artificial neural network and fuzzy clustering methods. In: IEEE GCC conference and exhibition (GCC), pp 295–298
10. Chama CK, Mukhopadhyay S, Biswas PK, Dhara AK, Madaiah MK, Khandelwal N (2013) Automated lung field segmentation in CT images using mean shift clustering and geometrical features
11. Antonio Oseas de Carvalho Filho, Wener Borges de sampaio, Aristofanes Correa Silva, Anselmo Cardoso de paiva, Rodolfo Acatavassu Nunes and Marcado galtass (2014) Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. Artif Intell Med 60:165–177
12. Setio AAA, Jacobs C, Gelderblom J (2015) Automatic detection of large pulmonary solid nodules in thoracic CT images. Med Phys 42(10)
13. Junior JRF, Oliveira MC, de Azevedo-Marques PM (2016) Cloud-based NoSQL open database of pulmonary nodules for computer-aided lung cancer diagnosis and reproducible research. J Digit Imaging. Springer
14. Abbas Q (2017) Segmentation of differential structures on computed tomography images for diagnosis lung-related diseases. Biomed Signal Process Control 33:325–334

15. Antonio Oseas de Carvalho Filho, Aristófanes Corrêa Silva, Anselmo Cardoso de Paiva, Rodolfo Acatauassú Nunes, Marcelo Gattass (2017) Computer-aided diagnosis of lung nodules in computed tomography by using phylogenetic diversity, genetic algorithm, and SVM. J Digit Imaging Soc Imaging Inf Med
16. Makaju S, Prasad PWS, Alsadoon A, Singh AK, Elchouemi A (2018) Lung cancer detection using CT scan images. In: 6th international conference on smart computing and communications, pp 107–114