ARTICLE IN PRESS

Engineering Science and Technology, an International Journal xxx (2016) xxx–xxx

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: www.elsevier.com/locate/jestch

Full Length Article

# Automatic detection of solitary pulmonary nodules using swarm intelligence optimized neural networks on CT images

Ezhil E. Nithila *, S.S. Kumar

*Noorul Islam Centre for Higher Education, Kumaracoil 629180, India*

A R T I C L E I N F O

*Article history:*
Received 28 August 2016
Revised 8 December 2016
Accepted 8 December 2016
Available online xxxx

*Keywords:*
Back Propagation Neural Network
Computer-Aided Diagnosis
Particle swarm
Inertia weight

A B S T R A C T

Lung Cancer is one of the most dangerous diseases that cause a large number of deaths. Early detection and analysis will be the only remedy. Computer-Aided Diagnosis (CAD) plays a key role in the early detection and diagnosis of lung cancer. This paper develops a CAD system that focus on new heuristic search algorithm to optimize the Back Propagation Neural Network (BPNN) in characterizing nodule from non-nodules. The proposed CAD system consists of four main stages: (i) image acquisition (ii) lesion detection, (iii) texture feature extraction and (iv) tumor characterization using a classifier. The optimization mechanism employs Particle Swarm Optimization (PSO) with new inertia weight for NN in order to investigate the classification rate of these algorithms in reducing the problems of trapping in local minima and the slow convergence rate of current evolutionary learning algorithms. The experiments were conducted on CT images to classify into nodule and non-nodule from the tumor region of interest. The performance of the CAD system was evaluated for the texture characterized images taken from LIDC-IDRI and SPIE-AAPM databases. Due to improved inertia weight used in Particle Swarm (PS) the CAD achieves highest classification accuracy of 98% for solid nodules, 99.5% for part solid nodules and 97.2% for non solid nodules respectively. The experimental results suggest that the developed CAD system has great potential and promise in the automatic diagnosis of tumors of lung.



## 1. Introduction

Lung cancer is one of the most common internal malignancies worldwide [1]. The most critical aspect related to the occurrence of cancer is the use of tobacco. Persistent of cough, shortness in breathing, pain in the chest, and bronchitis were some of the symptoms exist when the cancer is in advanced stage. It was found that if the disease was diagnosed earlier, about one-third of cancers become preventable, another one-third become potentially curable. Effective implementation and knowledge from successful research can only pave the way to control the lung cancer [2].

In medical imaging, detection of lung lesion is one of the most challenging tasks [3]. Examination of lung tumor is performed with various imaging modalities such as Radiography, Computed Tomography (CT) and Positron Emission Tomography. Among them CT is most frequently used imaging modality due to low cost, quality and robustness. Due to non-pathological structures the radiologist finds difficult and time-consuming to distinguish some nodules in CT [4]. To overcome this, radiologists choose alternate choice to confirm their analysis. Computer- Aided Diagnosis (CAD) systems have been developed toward the intention in medical imaging modalities, to enhance the performance of the diagnosis and improve methodical decisions in clinical practice.

It also stresses that methods based on supervised classification generally outperform well compared with other standard methods [5]. According to researchers, learning a Neural Network (NN) is an imperative task to solve many classical problems. Learning is the process of modifying the weights in order to produce a network that performs some function.

Our study aims at examining the performance of supervised classifier for the detection of lesions even at small diameter. These classifiers are developed to improve the classification rate and reduce False Positives (FPs).

## 2. Related work

A number of literatures are available to state the problem which arises in automatic detection of pulmonary nodule. Some of the works are presented below:

The methodology developed by Jacobs et al. [6] aims to detect subsolid nodule automatically by means of 128 features based on intensity, shape, texture and context features. This methodology

* Corresponding author.
*E-mail address:* ezhilniche@gmail.com (E.E. Nithila).

Peer review under responsibility of Karabuk University.

http://dx.doi.org/10.1016/j.jestch.2016.12.006
2215-0986/

uses Gentle Boost (GB) classifier, Support vector machine (SVM) with radial basis as kernel function, linear discriminant classifier, k-nearest neighbor classifier and nearest mean classifier. The GB-10 presents better results with 80% sensitivity and 1 false positive/scan.

The work by Lee et al. [7] developed a two-step feature selection and Random Subspace Method (RSM) for the diagnosis of pulmonary nodules. To validate the performance a set of 216 features for 128 pulmonary nodules were used. The proposed methodology achieves a maximum $A_z$ of 8.89 for two-step feature selection and 8.66 for RSM.

The architecture of CAD system developed by Silva et al. [8] presents a lung nodule detection system by means of segmentation and classification with Lib SVM. The proposed methodology achieved 84.84% sensitivity, 96.15% specificity and 95.21% accuracy for 33 exams. Liu et al. [9] develop a new CAD system using proprietary databases. In their work, automatic segmentation of CT lung image was performed using thresholding and region growing algorithm. The lung nodule was detected using semi-supervised method called ADE-Co-Forest with a sensitivity of 85.6% and false positive of 13.4%. The proposed CAD shows better superiority when compared with SVM.

Choi et al. [10] developed a automatic CAD to improve the detection accuracy for solitary pulmonary nodules. Volumetric segmentation of lung was performed using optimal thresholding and neighborhood structures are eliminated using 3D-connected component analysis. For nodule detection a set of techniques such as multi-scale dot enhancement filtering and angular histograms are used to describe the detected features. Finally, a classification approach of SVM was used to obtain a sensitivity of 97.5% and 6.76% per scan of false positives.

To evaluate and compare the scheme of other published CAD schemes Silva et al. [11] developed quality threshold algorithm and region growing for image segmentation. The maximum classification sensitivity is 85.91%, specificity 97.7%, accuracy 97.55% with FP of 1.82 for SVM classifier. Tong et al. [12] has used a three step process to detect lung nodules. Firstly, an adaptive threshold algorithm was used to segment the lung region. Secondly, active contour model (ACM) was used to remove lung vessel and finally a Hessian matrix (selective shape filter) was used to detect the suspicious nodules. This method was able to produce an overall detection rate of 85%.

Marten et al. [13] evaluated and compared features such as nodule size, position, margin, matrix characteristics, vascular and pleural attachments with gold standard. Some author uses manually segmented lesion as gold standard and some others uses specialist references as gold standard. Azimifar et al. [14] used active contour modeling for image segmentation to produce a possible detection rate of 89%.

Hua et al. [15] proposes a deep learning approach in comparison with conventional CAD. The author finds some limitation in CAD. According to their view, the proposed learning approach has some intrinsic advantage in the case of performance evolution. Similarly in the place of NNs, Deep Belief Network (DBN) classifier was designed. Various techniques proposed in CAD were compared against the deep learning approach. The system shows an efficient output when compared with CAD techniques. The authors suggest that this method can be replaced in the place of CAD. Yet, the sensitivity and specificity of this method was 73.4% and 82.2% respectively. The proposed DBN was compared with Convolutional Neural Network (CNN), Scale Invariant Feature Transform (SIFT) and fractal method.

Firmino et al. [16] proposed a new CAD for the detection of pulmonary nodule. The images are taken from LIDC-IDRI databases and 420 cases are chosen. The nodule type such as solid, pleural, vascular and non solid are selected. The author uses watershed segmentation in order to detect the nodule from neighboring structures. Histogram of oriented Gradients (HOG) approach was used for extracting the features from the ROI. Various classifiers such as SVM and rule based method was used for reducing FPs. The system achieves an accuracy of 97%, 94.4% sensitivity with 7.04 FP/case.

Researchers interested in more specific topic about CT images can refer to publications by Irmark et al. for image enhancement algorithm [17], brain tumor [18], and Ertas et al. [19–21] for cortical bone simulation.

## 3. Materials and methods

The developed CAD consists of various stages of methodologies such as (1) CT lung image acquisition, (2) detection of lung lesions, (3) features extraction using texture analysis and (4) classification of nodule candidates using supervised algorithm. Fig. 1 displays the flowchart of proposed CAD.

### 3.1. CT lung acquisition

The Lung images are acquired from the Lung Imaging Database Consortium-Image Database Resource Initiative (LIDC-IDRI) and International Society for Optics and Photonics (SPIE) with the support of the American Association of Physicists in Medicine (AAPM) Lung CT challenge [22]. All the images are in DICOM format with the image size of $512 \times 512$ pixels. In LIDC-IDRI the thickness of slice varies from 1.25 to 2.5 mm range with the pixel size from 0.48 mm to 0.72 mm respectively [23]. Various parameters such as sphericity, lesion position, tumor types examined by 4 radiologists have been recorded [24]. 71 cases with 246 nodule images were annotated in this work. All the CT images in this datasets are scanned using GE medical systems, Philips, Siemens and Toshiba. Similarly in SPIE-AAPM about 35 cases with 28 nodule images are used. In total 548 slices with 274 each for nodule and non-nodule are used. The nodule sizes between 3 and 30 mm (11–113 pixels) are considered in this work. All images taken from the datasets are only based on solid nodule, part solid nodule and non solid nodule. Fig. 2 shows a sample of three input images. The first one was an image with solitary solid nodule, the second and third images are the examples for part solid nodule and non solid nodule respectively.

### 3.2. Detection of lung lesion

Detection of lung lesion alone of a CT image is a crucial step. The reason behind is that the intensity of bronchi and blood vessel around the tumor is similar. It is essential to segment the lung parenchyma from the surrounding structures such as heart, diaphragm, bones, skin, fat and tissue. From the segmented parenchyma, it becomes easier to segment the tumor candidates. The steps involved in detection of lesion as follows.

#### 3.2.1. Segmentation of lung parenchyma

Segmentation of Parenchyma from CT images only paves way to identify the tumor easier to diagnose lung disease. The main objective of this stage is to eliminate the mediastinum, thoracic wall that do not contribute in the reconstruction the parenchyma. The proposed work uses Contour Model (CM) for lung segmentation. Based on its operation, this model can be classified in two approaches. The edge based CM produces small variation in segmentation due to image gradients. If weak boundary exists, leakage problem arises too [25]. Presence of noise in the boundary cannot be tackled in edge-based approach. Region based CM does not produce boundary leakage problem and hinders the effect of problem