Biomedical Signal Processing and Control 41 (2018) 140–151

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

An appraisal of nodules detection techniques for lung cancer in CT images

Muhammad Zia ur Rehman[a,b,*], Muzzamil Javaid[a], Syed Irtiza Ali Shah[a], Syed Omer Gilani[a], Mohsin Jamil[a], Shahid Ikramullah Butt[a]

[a] Department of Robotics and Artificial Intelligence, School of Mechanical and Manufacturing Engineering (SMME), National University of Science and Technology (NUST), H-12, NUST HQ, Islamabad, Pakistan
[b] Center for Sensory-Motor Interaction, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

ARTICLE INFO

Article history:
Received 19 April 2017
Received in revised form 2 October 2017
Accepted 26 November 2017

Keywords:
Medical image analysis
Lung cancer
Nodules detection
Chest CT images
Computer aided detection system

ABSTRACT

Lung cancer has a five-year survival rate of 17.7% which increases to 54.4% when it is diagnosed at early stages. Automated detection techniques have been developed to detect and diagnose nodules at early stages in computer tomography (CT) images. This paper presents a systematic analysis of the recent nodules detection techniques with the goal to summerize current trends and future challenges. The relevant papers are selected from IEEEXplore, science direct, PubMed, and web of science databases. Each paper is critically reviewed in order to summarize its methodology and results for further analysis. Our analyses reveal that several methods show potential progress in the field but still require an improvement to overcome many challenges like, high sensitivity with low false positive (FP) rate, detection of different nodules based on their size, shape, and positions, integration with electronic medical record (EMR) and picture archiving and communication system (PACS), and providing robust techniques that are successful across different databases. To overcome these challenges and developing a robust computer aided detection (CADe) system, it is believed that collaborative work is required among the developers, clinicians and other relating parties in order to understand particular issues and needs of a CADe system and develop automatic techniques to overcome these challenges with high processing speed, low cost of implementation and with software security assurance.

Contents

* Corresponding author at: Department of Robotics and Artificial Intelligence, School of Mechanical and Manufacturing Engineering (SMME), National University of Science and Technology (NUST), H-12, NUST HQ, Islamabad, Pakistan.
E-mail addresses: ziaurrehman@smme.edu.pk, zia@hst.aau.dk (M. Zia ur Rehman).

https://doi.org/10.1016/j.bspc.2017.11.017
1746-8094/

1. Introduction

Cancer is considered as a group of diseases which is initiated due to uncontrolled cell growth. It is one of the world's leading public health problem [1] with approximately 14 million new cases and 8.2 million deaths [2]. In US alone, the projected new cancer cases and deaths for year 2016 are 1,685,210 and 595,690 respectively [1]. Among other types, Lung cancer is a leading cause for cancer related deaths [3–5]. According to American Cancer Society, the projected new lung cancer cases and deaths for year 2016 are 224,390 and 158,080 respectively [1]. The survival rate is lower than many other types of cancer. According to American Lung association factsheet, five-year survival rate for lung cancer is 17.7%, However it increases to 54.4% when it is diagnosed at early stages and the cancer is localized, but only 15% of cases are diagnosed at early stages [6].

An early diagnosis of lung nodules can potentially improve the prognosis and hence can save many lives each year. The term nodules is used for the opacity with a diameter roughly between 3 and 30 mm [7] and it can be further categorized i.e. based on their location or position, they may be well circumscribed nodules having no attachment with neighboring structures, juxta-pleural and juxta-vascular nodules attached to lung parenchyma and vessels respectively [8]. They might be solid, sub-solid and non-solid [9]. Some of these nodules are shown in Fig. 4. Computer Tomography (CT) for early detection of these nodules is one of the most sensitive imaging modality [10]. Lung Cancer Screening Trials (LCST) screening results show that 5-year mortality for lung cancer reduce by about 20% in CT screening as compared with that of chest x-ray [11]. Other such studies conclude that nodules detection rate through CT is comparatively higher than analog radiography [12].

In clinical context, after performance of a patient's CT scan, radiologist have to interpret large number of images and based on nodule's morphology, his/her investigation may possibly result in no treatment or other clinical procedures [13]. This follow up of CT and clinical procedures directly effects the workload of a radiologist and along with other factors like fatigue, distraction and experience, it may result in error or misinterpretation of data [9]. Therefore, advanced image analysis systems are needed to help radiologists in interpreting diagnostic data, decision making and overcome human's memory limitation, fatigue and distraction. Such systems are generally called CAD systems and they have the potential to reduce workload and free up radiologists' time as they have to work with more complex tasks [14].

CADe (computer aided detection) systems are mainly focused to detect potential lesions in medical images while CADx (computer aided diagnosis) systems are used for characterization and further classification of lesions. Both systems are being active area of research [15], however this review focuses only on CADe systems which are generally aimed to assist radiologists in early detection of potential lesions with higher accuracy and less interpretation time [9]. In order to be considered worthful by most radiologists, they must meet some requirements including [9,13,16] higher sensitivity with a low false positive rate, high level of automation, computationally efficient, present low cost of implementation, detect different nodules based on their size, shape and location; and with software security assurance to avoid potential attacks.

One of the important aspect for comparison and generalization of developed algorithms are the use of databases for data collection. From the selected articles for this review, some researchers used the data from private databases or local hospitals while the rest have used data from some of publicly available databases like Lung Image Database Consortium (LIDC) [17,18], Early Lung Cancer Action Program [19], Database of Japanese Society of Radiological Technology (JSRT) [20], Automatic Nodule Detection 2009 (ANODE09) [21], Nederland-Leuvens Longkanker Screenings Onderzoek (NELSON) [22] and Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) [23,24]. These databases are being utilized for training and comparison of developed algorithms and their details are discussed in coming section.

Many researchers have previously reviewed the techniques utilized for the detection of potential nodules. Some of those papers include Lee et al. [8], Eadie et al. [14], Kenji Suzuki [15], El-baz et al. [25], Firmino et al. [9] and the recent Valente et al. [13]. Lee et al. [8] reviewed algorithms until 2009, Kenji Suzuki [15] up to 2010, El-baz [25] up to June 2012, Firmino [9] up to August 2013 and more recent Valente et al. [13] analyzed techniques up to December 2014. This paper is focused to select some of the good work and latest techniques that are incorporated into CAD system and hence it covers the articles from IEEEXplore, science direct, PubMed and web of science databases up to January 2016.

2. Design of CADe system

Several studies [26–29] showed that radiologists' performance in nodules detection improved with the use of CADe system. Therefore, it has been considered as a useful system for detection of potential nodules [15]. Analysis of the reviewed papers reveal that researchers utilized different structures in their nodule detection methods, however commonly used steps in CADe system are, data acquisition, preprocessing, segmentation of lungs, nodules detection and false positive reduction as shown in Fig. 1.

2.1. Data acquisition

Among other medical imaging modalities, CT is preferred for initial screening of lung nodules [8]. Data acquisition for CADe system refers to acquiring images of these modalities for further analysis of potential nodules. Ideally this step should be achieved through integration of CADe with a picture archiving and communication system (PACS) and electronic medical record (EMR) [13]. Integration of CADe with several PACS have many successful implementations [31] i.e. Bogoni et al. [26] showed that sensitivity of reader increases in such scenario with minimal impact on interpretation time. Along with having prior medical history of a patient, an EMR will further help radiologist in diagnosis.

As mentioned earlier in introduction section, collection of data is an important aspect for generalization and development of algorithms. Systems developed with private databases hampers the comparison among other such systems. Therefore, publicly available databases were created to allow researcher around the world for utilizing a large and diverse datasets in order to develop, train and evaluate their CADe algorithms [32]. Commonly utilized databases by researchers included in this work are JSRT, LIDC, ANODE09, ELCAP and LIDC-IDRI.

In 1998, JSRT in cooperation with JRS (Japanese Radiological Society) made the JSRT database which contains 247 digitized CXR with 154 nodules and 93 non-nodules that were reviewed by three thoracic radiologists [20]. The database also includes patient information along with each case i.e. patients gender, age, diagnosis, degree of subtlety and size of nodules along with its coordinates. The ELCAP database available since 2003, was made with collaboration between ELCAP and Vision and Image Analysis (VIA) research group and it contains 50 LDCT scans with the purpose to evaluate performance of different CADe systems [19,23]. The ANODE09 (Automatic Nodule Detection 2009) database [21], with initiative to evaluate automatic nodule detection algorithms, consist of 55 chest CT scans with 5 annotated and 50 non-annotated scans (for testing) that are provided by NELSON study. CT scans from NELSON study have been widely utilized in different projects associated with lung