European Journal of Radiology 74 (2010) 124–129

Contents lists available at ScienceDirect

European Journal of Radiology

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

Multilevel binomial logistic prediction model for malignant pulmonary nodules based on texture features of CT image☆

Huan Wang[a], Xiu-Hua Guo[a,*], Zhong-Wei Jia[a], Hong-Kai Li[b], Zhi-Gang Liang[c], Kun-Cheng Li[c], Qian He[d]

[a] *Department of Epidemiology and Health Statistics, School of Public Health and Family Medicine, Capital Medical University. Beijing 100069, China*
[b] *Department of Biochemistry and Molecular Biology, Basic Medical Sciences, Peking University Health Science Center, Beijing 100083, China*
[c] *Department of Radiology, Xuan Wu Hospital, Capital Medical University, Beijing 100050, China*
[d] *Department of Radiology, Friendship Hospital, Capital Medical University, Beijing 100053, China*

ARTICLE INFO

Article history:
Received 9 October 2008
Received in revised form 14 December 2008
Accepted 16 January 2009

Keywords:
Texture extraction
CT image
Small pulmonary nodules
Hierarchical data
Multilevel model

ABSTRACT

Purpose: To introduce multilevel binomial logistic prediction model-based computer-aided diagnostic (CAD) method of small solitary pulmonary nodules (SPNs) diagnosis by combining patient and image characteristics by textural features of CT image.
Materials and methods: Describe fourteen gray level co-occurrence matrix textural features obtained from 2171 benign and malignant small solitary pulmonary nodules, which belongs to 185 patients. Multilevel binomial logistic model is applied to gain these initial insights.
Results: Five texture features, including Inertia, Entropy, Correlation, Difference-mean, Sum-Entropy, and age of patients own aggregating character on patient-level, which are statistically different ($P < 0.05$) between benign and malignant small solitary pulmonary nodules.
Conclusion: Some gray level co-occurrence matrix textural features are efficiently descriptive features of CT image of small solitary pulmonary nodules, which can profit diagnosis of earlier period lung cancer if combined patient-level characteristics to some extent.

1. Introduction

Computer and medical image processing technology have been developing rapidly. Many new medical imaging processing methods have been applied to clinic with developing medicine. Image segmentation and texture feature extraction are important areas of research. To date, various types of research on medical images have been used in clinical auxiliary diagnosis [1–4].

Lung cancer is one of the most harmful forms of cancer [5], which is the leading cause of cancer death in many regions of the world [6]. The overall 5-year survival rate of lung cancer patients is only 14%, and remained at this level for the past two decades. However, when lung cancer is found at the early stage I or II, 5-year survival rates can be as high as 60–70% [7]. Early diagnosis of lung cancer was only 15% [8]. Early detection can be achieved in a population screening; the most common screenings for lung cancer make use of low-radiation dose computer tomography (CT) scans. Spiral CT is considered as an effective tool of screening and early diagnosis of lung cancer [9]. CT enables us to visualize lung anatomy in great detail and has been used to accurately diagnose lung diseases since the 1980s [10]. Aside from the primary research question whether lung cancer screening is effective or not, the optimal management of CT detected pulmonary nodules is also of major clinical relevance [11]. The high frequency of small pulmonary nodules incidentally detected on a spiral CT of the chest made for purposes other than lung cancer screening, raises the question of how clinicians and radiologists should deal with these nodules [12]. (Imaging, solitary pulmonary nodules (SPNs) are defined as being <30 mm in size, usually >10 mm, completely surrounded by pulmonary parenchyma [13].) Detecting and diagnosing SPNs, the most common manifestation of lung cancer, are critical since early identification of malignant nodules is crucial to the chance for successful treatment [14]. As a result, many computer-aided diagnosis (CAD) methods have been proposed for SPNs [15–17]. To improve the accuracy and efficiency of CT screening programs for the detection of early signs of lung cancer, a number of research projects are focusing on developing methods for computer-aided diagnosis to assist radiologists in diagnosing lung cancer, including works on image segmentation [18–20] and texture analysis [21–23] for efficient detection of lung cancer.

Multilevel modeling techniques are appropriate when there is correlation among clusters of subjects. It is the presence of within-

☆ Program of Funds: The program of Science and Technology Development of Beijing Municipal Commission of Education (Serial Number: KM200610025014); the program of Capital Medical University of Basic-Clinical Study (Serial Number: 2006JL57); the program of Natural Science Fund of Beijing (Serial Number: 7092010).
* Corresponding author. Tel.: +86 1083911508; fax: +86 1083911501.
E-mail address: guoxiuh@ccmu.edu.cn (X.-H. Guo).

0720-048X/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ejrad.2009.01.024

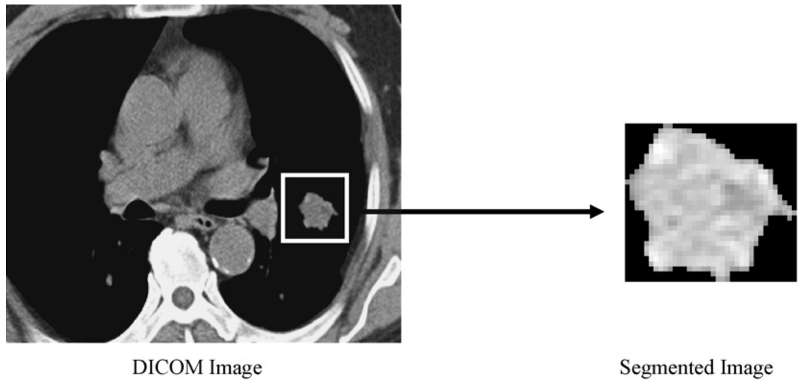

Fig. 1. Sample from CT scan and segmented image from the data set.

cluster correlation that justifies the use of a multilevel (hierarchical) model; without within-cluster correlation multilevel modeling does not provide benefit [24]. In the literature, multilevel models can be named multilevel linear models [25], mixed-effects models [26] or random-effects models [27], and so on. In this paper, the term “multilevel models” is selected.

Multilevel modeling techniques are commonly used in many fields, but it is applied to the area of CT images’ texture features analysis modeling firstly. One reason might be that the possible existence of multilevel structures in CT images data is commonly ignored. Alternatively, analysts may not be aware of the analysis technique or the motivation for adopting the technique. The authors identified there is correlation among CT images of one patient, so multilevel models were fitted to a two-level hierarchy and used to identify factors affecting texture features of benign and malignant CT images for individual casualties. By establishing a multilevel model of texture features of pulmonary nodules’ CT images can better describe the CT images of pulmonary nodules and profit early identification of small pulmonary nodules.

2. Materials and methods

2.1. Image database and texture features extraction

The digitized CT image set used in this study contains 2171 region of interests (ROIs) extracted from 185 patients with small solitary pulmonary nodules, with 61 related to benign nodule and 124 to malignant tumors. They are 107 men and 78 women (age range, 19–80 years; mean age, 58 years). The final diagnosis of 124 small peripheral lung cancers (diameter range, 1.0–3.0 cm; mean diameter, 2.0 cm) was determined by either an operation or biopsy. The diagnosis of benign nodule (diameter range, 1.2–3.0 cm; mean diameter, 2.1 cm) was confirmed either pathologically or after a 2-year follow-up. 124 small peripheral lung cancers shown in the images being proved by operation pathologically (including carcinoid, small-cell carcinoma, adenoma, sqamocellular carcinoma, etc.), and all the images being provided by the radiology department of Beijing Friendship Hospital as affiliated to Capital University of Medical Science.

CT scans of 149 nodules were obtained with an 64-slice helical CT scanner (GE/Light speed ultra System CT99, USA) using a tube voltage of 120 kV and current of 200 mA. Slice thickness and reconstruction intervals for routine scanning were 5 mm, and those for high-resolution CT were 1 mm. CT scans of the remaining 36 nodules were obtained with a single-slice helical CT scanner (Picker 2000, USA), with a slice thickness and reconstruction interval of 2 mm. CT images were displayed at fixed setting (lung window center, −600 Hounsfield units (HU); lung window width, 1600 HU; mediastinum window center, 40 HU; mediastinum window width, 400 HU). Data were reconstructed with a matrix of 512×512.

The structure of data from 185 patients is postulated as hierarchical data, which consists of two different levels: level-1 consisting of image-level characteristics and level-2 consisting of patient-level characteristics. Image-level characteristics contain detailed information associated with individual images such as Energy, Contrast, and Inverse Difference Moment (IDM); where as patient-level characteristics include sex and age.

ROIs (small pulmonary nodules) were segmented using gray level threshold algorithm [28]. Fig. 1 shows an example CT scan and a segmented slice of small pulmonary nodule. Using this segmentation algorithm, the 2171 small pulmonary nodules images were generated. Then, texture analysis (TA) was used to the processed images, which is a vital component of CAD because it is difficult to classify human tissues based on shape or gray-level information given in HU only [29]. Texture analysis methods can generally be grouped into structural or statistical methods, and co-occurrence is one category of statistical methods, which is a measure of the relative frequency or joint probability of two image properties occurring under predefined constraints, across the domain of an image. Gray-level co-occurrence matrix (GLCM) is the most widely used texture analysis method in biological imaging [30]. GLCM holds potential for analyzing segmented images of biogenic sedimentary structures because it can be used to analyze multi-scale differences in image texture [31]. Fourteen image-level and two patient-level variables are used as independent variables in the analysis, and the benign and malignant pulmonary nodules as the dependent variables, 1 if malignant, and 0 benign. Sex and age are patient-level variables, 1 if man, 0 woman and 1 if age >50.00, 0 otherwise. The descriptions of the fourteen image-level variables used in the study are provided in Table 1. Besides, Table 1 gives formulas of fourteen GLCM textural features in the study [32–34]. The Boundary Point of the fourteen image-level variables was got from analysis on the mean and standard deviation of texture features of benign and malignant pulmonary nodules. Image segmentation and texture feature analysis were analyzed in MATLAB (version 7.3.0), spending 10–13 and 10–15 s, respectively.

Energy is defined to measure the number of repeated pairs, which is expected to be high if the occurrence of repeated pixel pairs is high. Homogeneity measures the local homogeneity of a pixel pair. The homogeneity is expected to be large if the gray levels of each pixel are similar. IDM tells us about the smoothness of the image, like homogeneity. Inertia reflects the roughness of texture, which is expected to be low if the more elements are near to diagonal line of matrix when texture is rougher. Correlation is expected to measure the relevance on the gray of pixel. Contrast measures the local contrast of an image, which is expected to be low if the gray