Original Investigation

Computer-aided Diagnosis for Lung Cancer: Usefulness of Nodule Heterogeneity

Mizuho Nishio, MD, PhD, Chihiro Nagashima, MD

Rationale and Objectives: To develop a computer-aided diagnosis system to differentiate between malignant and benign nodules.

Materials and Methods: Seventy-three lung nodules revealed on 60 sets of computed tomography (CT) images were analyzed. Contrast-enhanced CT was performed in 46 CT examinations. The images were provided by the LUNGx Challenge, and the ground truth of the lung nodules was unavailable; a surrogate ground truth was, therefore, constructed by radiological evaluation. Our proposed method involved novel patch-based feature extraction using principal component analysis, image convolution, and pooling operations. This method was compared to three other systems for the extraction of nodule features: histogram of CT density, local binary pattern on three orthogonal planes, and three-dimensional random local binary pattern. The probabilistic outputs of the systems and surrogate ground truth were analyzed using receiver operating characteristic analysis and area under the curve. The LUNGx Challenge team also calculated the area under the curve of our proposed method based on the actual ground truth of their dataset.

Results: Based on the surrogate ground truth, the areas under the curve were as follows: histogram of CT density, 0.640; local binary pattern on three orthogonal planes, 0.688; three-dimensional random local binary pattern, 0.725; and the proposed method, 0.837. Based on the actual ground truth, the area under the curve of the proposed method was 0.81.

Conclusions: The proposed method could capture discriminative characteristics of lung nodules and was useful for the differentiation between malignant and benign nodules.

Key Words: CAD; Computer-aided diagnosis; CT; Lung cancer.

INTRODUCTION

In the United States, it was projected that 1,665,540 new cancer cases and 585,720 cancer deaths would occur in 2014, with 86,930 male and 72,330 female Americans dying from lung cancer (1). Lung cancer is the leading cause of cancer deaths. For most cancers, there have been notable improvements in survival rates over the past three decades, but lung and pancreatic cancers have shown the least improvement (1).

Computed tomography (CT) has high sensitivity in detecting lung nodules and improves the likelihood of detecting lung cancer at an early stage. However, it can be difficult for a radiologist or a pulmonologist to differentiate between malignant and benign lung nodules on CT. For example, in the National Lung Screening Trial, the rate of positive results was 24.2% with low-dose helical CT screening over all three rounds, and a total of 96.4% of the positive results in the low-dose CT group were false positives (2). False-positive findings can result in unnecessary follow-up CT scans, positron emission tomography scans, and invasive procedures such as bronchoscopy or surgical resection, raising concerns about the increased radiation or surgery risks for the patient (3).

Computer-aided diagnosis of lung nodules can enhance radiologists' workflow, assisting with the detection of lung nodules, differentiation between benign and malignant nodules, and monitoring of the treatment response. In terms of differentiating between benign and malignant nodules, computer-aided diagnosis for classification, often referred to as CADx, has been found to be useful for improving the accuracy of diagnoses by radiologists (4–14) (for an overview, see Refs. 4–6).

The purpose of this study was to develop and evaluate CADx classification of lung nodules on CT images. We focused on the classification of malignant and benign nodules using image features extracted from CT images. Recently, patch-based image processing has been shown to be useful in image denoising (15), organ segmentation (16), and CADx for Alzheimer disease (17). In this study, we proposed and evaluated a CADx method based on patch-based feature extraction without the need for segmenting lung nodules. Image convolution, pooling operations, and principal component analysis (PCA) were used to obtain discriminative feature of lung nodule using image patch. This method was easy to implement, and it did not require input of radiological or clinical findings.

Acad Radiol 2016; ■:■■–■■

From the Clinical PET Center, Institute of Biomedical Research and Innovation, 2-2 Minatojimaminamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.
Received May 3, 2016; revised October 14, 2016; accepted November 2, 2016.
Address correspondence to: M.N. e-mail: jurader@yahoo.co.jp

http://dx.doi.org/10.1016/j.acra.2016.11.007

MATERIALS AND METHODS

The institutional review board of our institution approved this study. Written informed consent was waived by the review board. Our CADx system was tested using chest CT images from the LUNGx Challenge, conducted as part of the 2015 SPIE Medical Imaging Conference. The LUNGx Challenge provided these images to participants as an opportunity for them to compare their CADx algorithms to those of others using the same datasets. Details of the LUNGx Challenge are available elsewhere (18).

CT Images

The LUNGx Challenge provided 60 test sets of chest CT images with 10 calibration sets. The organizers clarified that the lung nodules in the calibration sets were not necessarily representative of the difficulty level in the test sets. The test sets were used in the present study to train our CADx system and to evaluate its performance.

The following patient information and technical parameters of CT scan were obtained from the digital imaging and communication in medicine (DICOM) data of the test sets. In the test sets, the patients consisted of 23 men and 37 women; mean ± standard deviation age, 60.0 ± 13.7 years; age range, 18–84 years. The parameters of the CT scans were as follows: tube current, 240–500 mA; tube current-exposure time product, 200–325 mAs; tube potential, 120 or 140 kV; matrix size, 512 × 512; and slice thickness, 1 mm. Contrast-enhanced CT was performed for 46 of the 60 patients. The CT images included the whole lungs. In the LUNGx Challenge, 73 nodules revealed in the test sets were evaluated; these nodules were also assessed in the present study. The list of the evaluated nodules is available on the LUNGx Challenge website.

Evaluation by Radiologists and Construction of Surrogate Ground Truth

Because the ground truth for the lung nodules was not available from the LUNGx Challenge, a surrogate ground truth was constructed in the present study by radiological evaluation. Two board-certified radiologists (MN and CN), with 6 and 9 years of experience, respectively, independently evaluated the CT images without any clinical information. They evaluated the images radiologically and rated their suspicion of malignancy on a 10-point scale (1, benign nodule; 10, malignant nodule). The CT images were displayed with the lung window settings specified in the DICOM data, namely a window width of 2000 Hounsfield units (HU) and a window level of −450 HU. The radiologists were allowed to change the window setting and magnify or shrink the CT images during their evaluation of the lung nodules. If the mean score of the two radiologists was greater than 5.5, the lung nodule was assumed to be a malignant nodule; otherwise, it was considered to be a benign nodule.

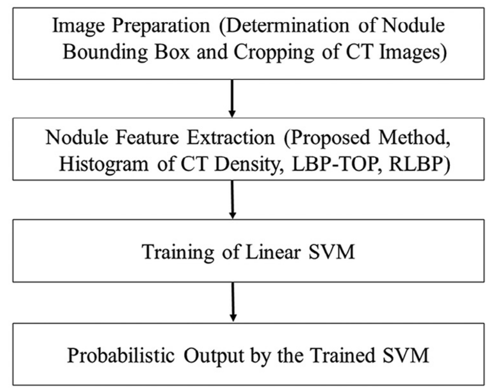

Figure 1. Schema of the computer-aided diagnosis for differentiation between benign and malignant nodules on CT images. LBP-TOP, local binary pattern on three orthogonal planes; RLBP, three-dimensional random local binary pattern; SVM, support vector machine.

Computer-aided Diagnosis

An overview of the CADx system is shown in Figure 1. After image preparation, the nodule features were extracted and the nodules were classified using machine learning algorithm. Because the lung nodule was not segmented in the present study, shape-related nodule features were not used in the CADx system.

The proposed CADx system being evaluated in this study used a novel patch-based feature extraction to capture the nodule characteristics on the CT images. Three other CADx systems were also evaluated for comparison: histogram of CT density, local binary pattern on three orthogonal planes (LBP-TOP) (19–21), and three-dimensional (3D) random local binary pattern (RLBP) (22).

Image Preparation

To prepare each set of CT images, a 3D cubic bounding box was set for each nodule such that the bounding box included the entire lung nodule and surrounding tissue. The location and size of the bounding box were visually validated by the radiologist with 9 years' experience (MN). The cropped CT images were analyzed as the input to the CADx system; areas of the images outside the bounding box were not assessed. Representative images of a lung nodule and bounding box are shown in Figure 2.

The Proposed CADx Method

To use CT density heterogeneity as the nodule feature, the proposed method used image convolution and pooling operations (23). In addition, PCA was used to calculate the kernels for image convolution. In the previous study (23), it was assumed that input to the system was two-dimensional (2D) images and that image sizes were identical. On the other hand, in the present study, we applied Labusch et al.'s method to