Classification of Benign/Malignant PNGGOs using K-means algorithm in MDCT Images: A Preliminary Study

Wooram Son †, Sang Joon Park †, Chang Min Park ‡, Jin Mo Goo ‡, Jong Hyo Kim †

† Interdisciplinary Program in Radiation Applied Life Science,
Seoul National University of Medicine
101, Daehangno, Jongno-Gu, Seoul, 110-744 South Korea

‡ Department of Radiology, Seoul National University Hospital
101, Daehangno, Jongno-Gu, Seoul, 110-744 South Korea

E-mail: † {sonwr, lunao78, kimjhyo}@snu.ac.kr, ‡ cmpark@radiol.snu.ac.kr, jmgoo@plaza.snu.ac.kr

Abstract Lung cancer is one of the most prevalent diseases in the world. Recently, PNGGOs (Pure nodular ground-glass opacity) have been reported to increasing aspect for all CT-detected pulmonary nodules. Moreover, the malignancy rate of PNGGOs is a considerable proportion of benign diseases. In this study, we have developed a computerized classification scheme of PNGGOs malignancy. Segmentation of PNGGOs was performed semi-automatically. After that, the histogram based statistical features and region based features of benign and malignant GGO was extracted. Finally, K-means classifier was applied. Experiment was performed employing 12 CT image sets and 91.67% of accuracy was achieved.

Key words PNGGO, MDCT, Lung Cancer, Classification, K-means

1. Introduction

A PNGGO (pure nodular ground-glass opacity), also referred to as a ground-glass opacity nodule is defined as nodule showing hazy increased attenuation within the lung that does not obscure the bronchial and vascular margins on CT [1]. In the screening situations, PNGGOs have been reported to consist of 5.5-12% of all CT-detected pulmonary nodules [2]. For that reason, PNGGOs are one of the important features in lung cancer diagnosis of CAD (Computer-aided Diagnosis). Pulmonary PNGGOs are a kind of non-specific CT findings, however, when the lesions persist, these lesions have a possibility of malignancies including bronchioloalveolar carcinoma (BAC), and pulmonary adenocarcinoma (AD), or benign diseases including atypical adenomatous hyper-plasia (AAH) and focal interstitial fibrosis (FIF), but even most of these are benign. [2]

Some studies of pathologically proven cases have shown PNGGOs malignancy rates of 18% to 48% [3-6]; it is shown that, PNGGOs has considerable proportion of benign diseases. Therefore accurate estimation between malignant and benign lesions is important in order to reduce the number of unnecessary surgeries [2].

Therefore, the purpose of this study was to compare the statistical analysis and develop an automated scheme to classification of PNGGOs in the lung at CT following below step. Segmentation of PNGGOs is performed by threshold technique. After that, Euclidean distance between the histogram based statistical features of benign and malignant GGO is extracted. Finally, K-means classifier is applied.

2. Materials and Methods

2. 1. Patients

From July 2004 to July 2007, we selected patients with PNGGOs identified on thin-section CT. Inclusion criteria were PNGGOs no greater than 3cm on thin-section CT and pathologically confirmed cases. We identified 12 patients (four man, eight women; age range, 46-70 years; mean age, 58 years) having PNGGOs in the lung. Of these patients, 12 PNGGOs (malignant, 6 lesions; benign, 6 lesions) in 12 patients were confirmed pathologically (AAH, 5 patients; FIF, 1 patients; AD, 4 patients; BAC, 2 patient)

2. 2. PNGGOs Segmentation in ROI

The PNGGO regions were segmented by using a threshold and labeling technique [7]. The lower threshold values are automatically determined by analyzing the histogram of the marked ROI in CT images. After segmenting the GGO lesion, we manually correct the segmented region for removal vessels and produce the better segmentation result due to the large overlap of intensity values between PNGGO and surrounding vessels.

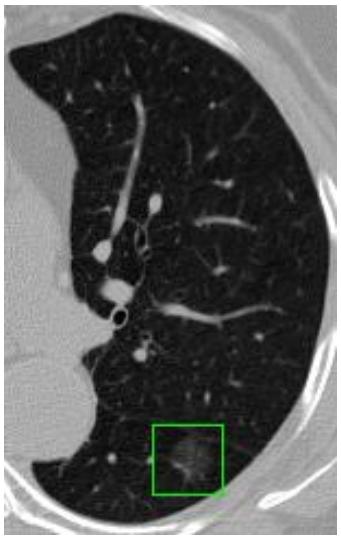
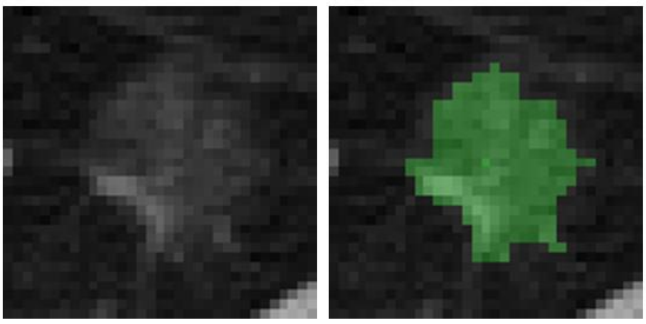


Figure 2. PNGGOs images with marked ROI.



(a) (b)

Figure 3. (a) PNGGO ROI, (b) Segmentation result.

2. 3. Statistical Feature Extraction

For each segmented regions of PNGGOs, we calculated eight statistical features as shown in Table 1.

Table 1. Statistical features and their abbreviations

Feature	Abbreviation
Mean	MEAN
Central Moment μ_2	CM 2
Central Moment μ_3	CM 3
Central Moment μ_4	CM 4
Energy	ENG
Entropy	ENT
Kurtosis	KURT
Skewness	SKEW

2. 4. K-means Clustering Algorithm

In this initial study, we used K-means clustering algorithms ([8], [9]) as the classification method. K-means minimizes the sum of the Euclidean distances from all feature vectors in a cluster to the cluster center.

The input value of K-means was statistical features of each segmented regions of PNGGOs with K=2. The K-means algorithm returns two clusters and two centers as hyper-spheres. Generally, malignant PNGGOs have higher CT numbers than benign PNGGOs [5]. Therefore, we expected 8 statistical features will be useful in classification into benign or malignant classes.

3. Results

A box ROI region (green rectangle) is manually marked then segmentation method is applied as described in Figure 4, 5.

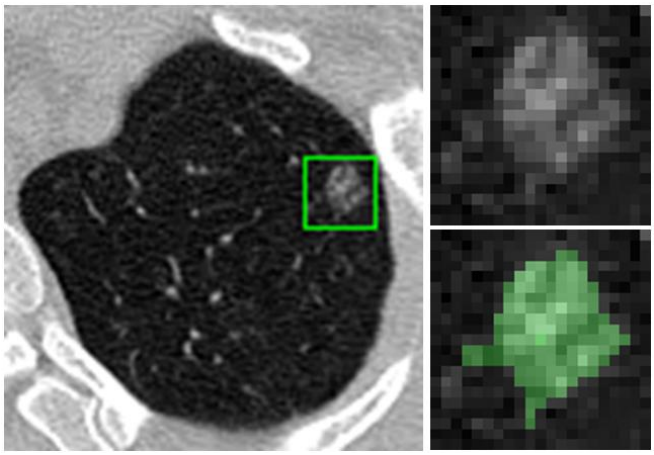


Figure 4. ROI region (left image), Cropped image (right-upper), Segmented image (right-lower)

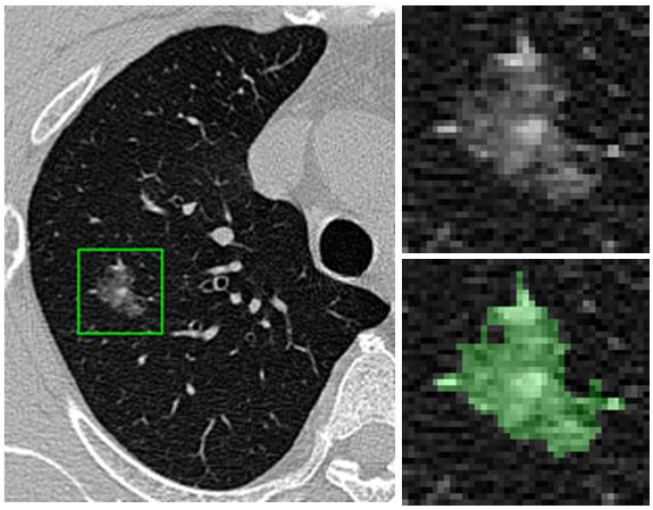


Figure 5. ROI region (left image), Cropped image (right-upper), Segmented image (right-lower)

We applied the PNGGO classification method to 12 CT images containing 12 nodules. The classification method detected the 6 of 6 benign nodules and the 5 of 6 malignant nodules, containing all of the benign nodules and one false negative malignant nodule as shown in Table 2. Therefore, the sensitivity was 83.3% and specificity was 100%, yielding 91.6% accuracy.

Table 2. Comparison of pathologic diagnosis and result of classification method

Case	Pathologic Diagnosis	Malignancy (=1)	Classifier result
1	AAH	0	0
2	AAH	0	0
3	AAH	0	0
4	FIF	1	1
5	AAH	0	0
6	AAH	0	0
7	FIF	0	0
8	AD	1	1
9	AD	1	1
10	BAC	1	0
11	BAC	1	1
12	AD	1	1

AAH: atypical adenomatous hyper-plasia

BAC: bronchioloalveolar carcinoma,

FIF: focal interstitial fibrosis

AD: adenocarcinoma

4. Conclusion

We applied the K-means clustering algorithm for automatic classification of malignant PNGGOs. Our automatic classification method consisted of two steps, PNGGOs segmentation and classification of malignant nodule. For PNGGOs segmentation, a box ROI was fist manually marked. Then, the marked regions were segmented by threshold technique. Finally, segmented regions were classified by K-means.

Our method applied to 12 CT images with 12 nodules, and 91.67% of accuracy was achieved. Our result showed that K-means classifier with statistical features can cluster the AAH, and classify the malignant or benign nodules.

5. Reference

[1] Austin, J., Muller, N., Friedman, P.e.a.: Glossary of terms for CT of the lung: recommendations of the nomenclature committee of the fleischner society. Radiology 200 (1996) 327–331

[2] C. Park, J. Goo, H. Lee, C. Lee, K. Kim, J.-G. Im. Persistent pure nodular ground-glass opacity in the lung: Comparison of malignancy and benignity. 2008 European Congress of Radiology.

[3] Nakajima, Ryu; Yokose, Tomoyuki; Kakinuma, Ryutaro; Nagai, Kanji; Nishiwaki, Yutaka; Ochiai, Atsushi. Localized Pure Ground-Glass Opacity on High-Resolution CT: Histologic Characteristics. Journal of Computer Assisted Tomography. 26(3):323-329, May/June 2002.

[4] Koei Ikeda, MD, PhD; Kazuo Awai, MD, PhD; Takeshi Mori, MD, PhD; Koichi Kawanaka, MD; Yasuyuki Yamashita, MD, PhD and Hiroaki Nomori, MD, PhD.: Differential Diagnosis of Ground-Glass Opacity Nodules. 2007 American College of Chest Physicians

[5] Nomori H, Ohtsuka T, Naruke T, Suemasu K. Differentiating between atypical adenomatous hyperplasia and broncholoalveolar carcinoma using the computed tomography number histogram. Ann Thorac Surg 2003; 76:867–871.

[6] Kim KG, Goo JM, Kim JH et al (2005) Computer-aided diagnosis of localized ground-glass opacity in the lung at CT: initial experience. Radiology 237:657–661

[7] Goo JM, Lee JW, Lee HJ, Kim S, Kim JH, Im JG. Automated lung nodule detection at low-dose CT: preliminary experience. Korean J Radiol 2003;4:211–216.

[8] Russ JC. The image processing handbook. 2nd ed. Boca Raton, Fla: CRC, 1995.

[9] J.B. MacQueen, Some methods for classification and analysis of multivariate observations. Proc. 5-th Berkeley Symposium on Mathematical Statistics and Probability, vol.1, pp.281–297, Berkeley, University of California Press, 1967.