

이학석사 학위논문

**Computer-aided diagnosis for  
Pure Nodular Ground-Glass Opacity:  
Classification of Benign and Malignant  
Lesions using a Support Vector Machine**

순수 결절형 간유리 음영의 컴퓨터 보조진단:  
SVM을 이용한 양성 및 악성병변의 분류

2010년 7월

서울대학교 대학원

방사선응용생명과학전공

손 우 람

순수 결절형 간유리 음영의  
컴퓨터 보조진단: SVM을 이용한  
양성 및 악성병변의 분류

지도 교수 김 종 효

이 논문을 이학 석사 학위논문으로 제출함  
2010년 4월

서울대학교 대학원  
방사선응용생명과학전공  
손 우 램

손우람의 이학석사 학위논문을 인준함  
2010년 4월

위 원 장 \_\_\_\_\_ (인)

부위원장 \_\_\_\_\_ (인)

위 원 \_\_\_\_\_ (인)

## **Abstract**

# **Computer-aided diagnosis for Pure Nodular Ground-Glass Opacity: Classification of Benign and Malignant Lesions using a Support Vector Machine**

Wooram Son

College of Medicine, Radiation Applied Life Science

The Graduate School

Seoul National University

Lung cancer is one of the most prevalent diseases in the world. The widely use of computed tomography in clinical practice have increased the frequency of findings of subtle nodules or Ground-Glass Opacity (GGO). GGO may be observed in malignancies such as bronchioloalveolar carcinoma and adenocarcinoma, as well as in their putative precursors, such as atypical adenomatous hyperplasia and also may be seen in the presence of benign condition. According to the several studies of pathologically proven cases, Pure Nodular Ground-Glass Opacities (PNGGOs) has a considerable proportion of benign diseases. Hence, accurate differentiation of benign or malignant PNGGO is important in lung cancer diagnosis in order to reduce the unnecessary surgeries. In this study, we propose a CAD system to classify benign or malignant disease of PNGGOs from MDCT images. Our system

consists of following three steps: automated segmentation for PNGGO is performed using a modeling of ROI histogram with a Gaussian mixture. After that, Statistical features of segmented PNGGOs regions are extracted. Finally, Support vector machine as a classifier is applied with radial basis function kernel.

Experiment was performed employing 34 CT volume images with 36 nodules and 0.91 (leave-one-out validation) areas under the curve were achieved. Exhaustive search was performed to find the best combination of features which used as input variables for SVM. Of the statistical features, entropy, mean, absolute deviation, skewness, kurtosis, percentile at 25th, percentile at 50th, percentile at 75th and inter-quartile range were used as input variables for exhaustive search. Entropy, kurtosis, percentile at 50th were proven to be best combination with SVM classifier.

The result of our CAD system shows potential possibility of to classify between benign and malignant PNGGOs

**Keywords: Pure Nodular Ground-Glass Opacity (PNGGO), Support Vector Machine (SVM), Computer-aided Diagnosis (CAD).**

**Student ID: 2008-23287**

# Contents

<b>Abstranct</b> .....	<b>i</b>
<b>Contents</b> .....	<b>ii</b>
<b>List of tables</b> .....	<b>iv</b>
<b>List of figures</b> .....	<b>v</b>
<b>1. Introduction</b> .....	<b>1</b>
<b>2. Materials and Methods</b> .....	<b>3</b>
2.1. Patient Selection .....	3
2.2. CT examination.....	4
2.3. Computerized segmentation of PNGGOs.....	4
2.3.1. Pre-Processing.....	5
2.3.2. Histogram-modeling.....	6
2.3.3. Region-growing segmentation .....	8
2.3.4. Vessel-like structure removal .....	8
2.3.5. Evaluation of the proposed segmentation method.....	9
2.4. Computerized classification of PNGGOs .....	10
2.4.1. Statistical analysis .....	10
2.4.2. Classification of Benign and Malignant Lesion using a SVM .....	12
<b>3. Results</b> .....	<b>13</b>
3.1. Performance of segmentation method .....	13
3.2. Feature evaluation and selection.....	15
3.3. Classification performance of SVMs.....	17
<b>4. Discussion</b> .....	<b>19</b>
<b>5. Conclusion</b> .....	<b>21</b>
<b>6. Reference</b> .....	<b>22</b>
<b>Abstract (in Korean)</b> .....	<b>24</b>

## List of Tables

<b>Table 1. Quantity and Pathological diagnosis of PNGGO Lesions .....</b>	<b>3</b>
<b>Table 2. Segmentation performance of the proposed method. The mean and standard deviation of the three indices between the automatic and manual segmentation were described. ....</b>	<b>13</b>
<b>Table 3. The performance indices for the example 4 GGO nodules shown in figure 9. ....</b>	<b>15</b>
<b>Table 4. Features of benign and malignant PNGGOs. ....</b>	<b>16</b>
<b>Table 5. Result of leave-one-out trials – Diagnostic performance of SVMs in PNGGOs. ....</b>	<b>18</b>

# List of Figures

<b>Figure 1. An example of pure nodular ground-glass opacities in high-resolution CT. Left image shows benign lesion, Right image shows malignant lesion .....</b>	<b>2</b>
<b>Figure 2. Example of segmentation for PNGGO: (a) Original image, (b) Segmented lesion using the region-growing technique and histogram modeling, and (c) Final segmentation result after removal of vessel-like structures using a 3D morphological opening operation.....</b>	<b>5</b>
<b>Figure 3. Flowchart of proposed ground-glass opacity segmentation method.....</b>	<b>5</b>
<b>Figure 4. The CT number histogram of PNGGOs. a) benign lesion, b) malignant lesion .....</b>	<b>7</b>
<b>Figure 5. a) histogram of an ROI including PNGGO, b) estimated probability density functions of lung parenchyma and PNGGO. ....</b>	<b>8</b>
<b>Figure 6. 2-D scatter plot depicting distributions and relationships of all two combinations of three features.....</b>	<b>12</b>
<b>Figure 7. A screenshot of the developed PNGGO segmentation tool.....</b>	<b>13</b>
<b>Figure 8. An example of PNGGO segmentation. The upper images are the largest slice CT image of the PNGGO. The segmentation results are in the lower images.....</b>	<b>14</b>
<b>Figure 9. Automatic segmentation results are in the upper images. The lower images are comparing with manual segmentation performed by radiologist.....</b>	<b>14</b>
<b>Figure 10. ROC curves of the uses of three individual features and combination of all three features for PNGGO differentiation. ....</b>	<b>17</b>

# 1. Introduction

Recently, the pure nodular ground-glass opacity (PNGGO) which had not been detectable on plain chest radiography are commonly found with widespread use of computed tomography (CT) scans for early detection of lung cancer in clinical practice. (1) PNGGO is defined as nodule showing hazy increased attenuation within the lung that does not obscure the bronchial and vascular margins on CT. PNGGO, when persistently (either no change or an increase in diameter for  $\geq 3$  month) present on serial thin-section computed tomography (CT) scans, suggest the possibility of early lung cancer such as bronchioloalveolar cell carcinoma (BAC) or pulmonary adenocarcinoma, or a putative precancerous lesion such as atypical adenomatous hyperplasia (AAH). (2-4)

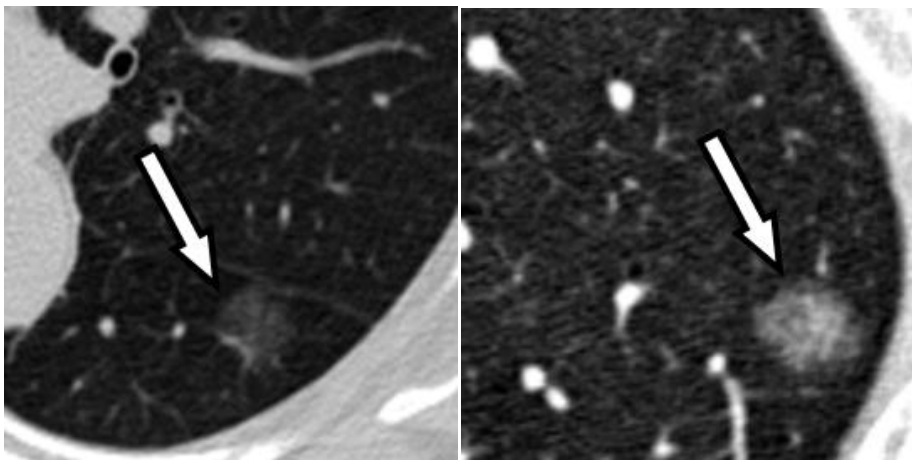
Previous studies of pathologically proven cases have shown part-solid nodules (PSN) have been reported to have a much higher malignancy rate than solid or PNGGO nodules, with the malignancy rate ranging from 62.5-89.6%. (2) It is shown that, PNGGOs has considerable proportion of benign diseases. Therefore accurate differentiation between malignant and benign lesions is important in order to reduce the number of unnecessary surgeries.

According to the result of early study from a report which was published regarding the predictive CT findings of malignancy in ground-glass nodules on thin-section chest CT, radiologists' performance is  $0.621 \pm 0.052$  (mean areas under the curve (AZ), without predictive information) and  $0.766 \pm 0.055$ ,  $P < 0.05$  (AZ, with predictive information) for PNGGOs. (3) Even though a few reports were published regarding the differentiation between malignant and benign persistent PNGGOs, (5-7) these studies were not able to provide practical differential points between them. For these reasons, recently, computerized quantification methods have tried in an attempt to differentiate



PNGGOs. (8-11) Nomori have used the CT number analysis to classify the malignant of GGOs, (8) but this study was constructed from both mixed-type and pure-type GGOs not only for PNGGOs. Reeves have tried the automated volumetric segmentation method to differentiate PNGGOs, (11) but this study did not provide classification scheme for malignant PNGGOs and did not allow the evaluation of differentiation between malignant and benign PNGGOs.

The purpose of this study was to evaluate the CT number features of PNGGOs and to develop the CAD system for differentiating malignant PNGGOs from benign PNGGOs.



**Figure 1. An example of pure nodular ground-glass opacities in high-resolution CT. Left image shows benign lesion, Right image shows malignant lesion**

## 2. Materials and Methods

### 2.1. Patient Selection

Approval for this study was obtained from our institutional review board. The requirement for informed consent was waived.

From April 2003 to March 2008, one chest radiologist (C.M.P) with 8 years of experience reading chest CTs selected patients who had PNGGOs on thin-section chest CT scans for lung cancer screening. Of these selected cases, we included cases that met the following criteria: 1) pathologically confirmed lesions, 2) PNGGOs smaller than 3 cm in size, 3) PNGGOs determined to be persistent on follow-up CT performed at least one month after the initial CT. Based on these criteria, 36 PNGGOs in 35 individuals were included in this study. There were xx men and xx women (age range, 30-71 years; mean age, 55.5 +- 11.2 years). The histological results were 8 adenocarcinomas (22 %), 10 BACs (28 %), 11 AAHs (31 %), and 7 FIFs (19%).

**Table 1. Quantity and Pathological diagnosis of PNGGO Lesions**

Malignant/Benign	Pathological Diagnosis	Quantity
Benign	atypical adenomatous hyper-plasia (AAH)	11
Benign	focal interstitial fibrosis (FIF)	7
Malignant	bronchioloalveolar carcinoma (BAC)	10
Malignant	adenocarcinoma (AD)	8

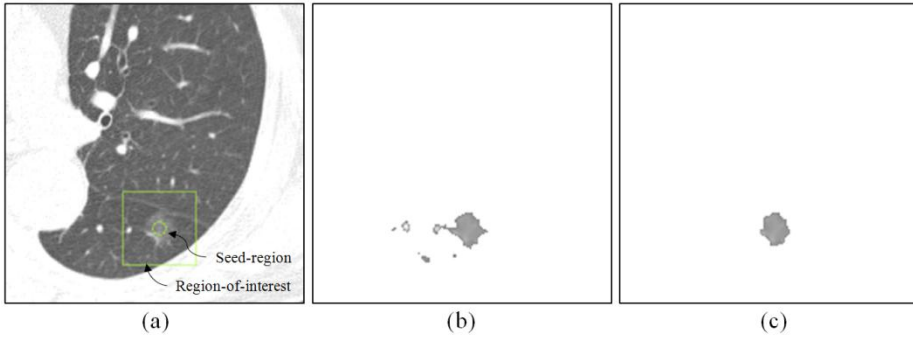
## **2.2. CT examination**

CT imaging was performed using one of four CT systems (LightSpeed Ultra, GE Medical Systems, Milwaukee, Wis.; Sensation 16, Siemens Medical Systems, Erlangen, Germany; Brilliance 64 and MX8000, Philips Medical Systems, Best, The Netherlands). CT parameters for thin-section helical CT were as follows: detector collimation, 1.0–2.5 mm; beam pitch, 0.75–1.75; reconstruction thickness, 1.0–2.5 mm; reconstruction interval, 1.0–2.5 mm; rotation time, 0.5–1.0 s; tube voltage, 120 kVp; and tube current, 150–200 mA. A reconstruction kernel and a high frequency algorithm were used. The mean interval between CT imaging and pathological confirmation was 20 days (range, 1–76 days).

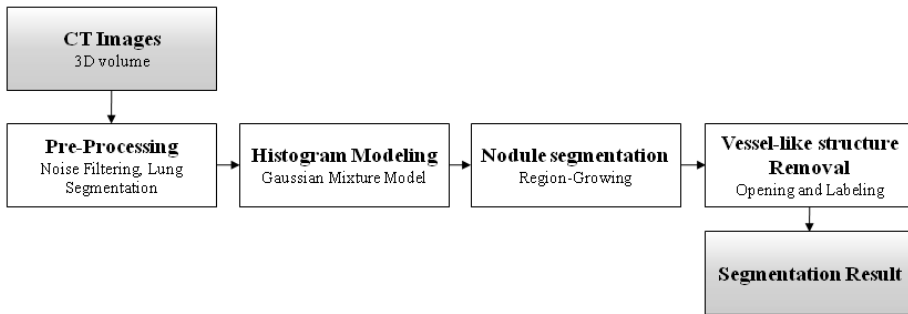
## **2.3. Computerized segmentation of PNGGOs**

Segmentation of PNGGOs was conducted using computer-aided diagnosis (CAD) software developed in-house. (10) Our proposed segmentation method consists of four steps, as shown in Figure 3, including pre-processing, histogram modeling, nodule segmentation and vessel-like structure removal.

Initially, one radiologist (C.M.P) specified the region of interest (ROI) around the seed area which included the target nodules. The threshold CT number between PNGGO and surrounding normal lung tissue was automatically determined by analyzing the histogram of the ROI. The Gaussian mixture model was used to estimate the distribution of PNGGOs in the histogram. Then, the CAD software automatically identified the PNGGO in the volume image from the surrounding normal lung tissue using the region-growing technique with the threshold CT number.



**Figure 2. Example of segmentation for PNGGO: (a) Original image, (b) Segmented lesion using the region-growing technique and histogram modeling, and (c) Final segmentation result after removal of vessel-like structures using a 3D morphological opening operation**



**Figure 3. Flowchart of proposed ground-glass opacity segmentation method.**

### 2.3.1. Pre-Processing

The objective of the pre-processing step is to remove quantum noises within the ROI and to remove soft tissues outside the lung which is necessary for the ROIs drawn near the lung boundary (for juxtapleural nodule segmentation). A median filter was used for noise filtering, and simple thresholding (above -204 HU) was used for removal of soft tissues.

### 2.3.2. Histogram-modeling

Because of hazy attenuation pattern in ground-glass opacity and subtle differences of CT number between normal tissues, determining an appropriate threshold value for segmentation of GGO nodules is particularly challenging. We propose a histogram modeling approach for automated selection of the threshold value between ground-glass opacity and surrounding normal lung tissue using a Gaussian mixture model. A clustering algorithm estimates the probability density functions of the ground-glass opacity and normal tissue iteratively. The optimal threshold CT number was then automatically determined by likelihood estimation.

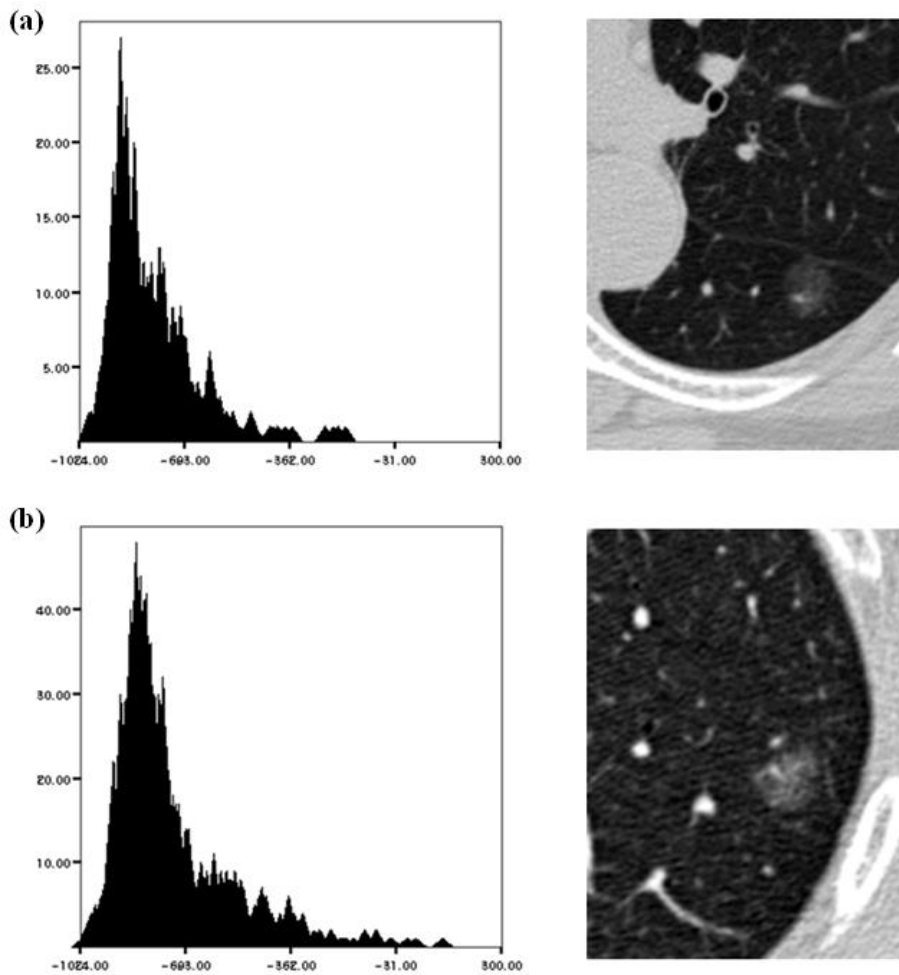
$$p(x_n | \omega_j, \theta) = p(x_n | \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x_n - \mu_j)^2}{2\sigma_j^2}\right) \quad (1)$$

$$p(x_n | \theta) = \sum_{j=1}^M p(x_n | \mu_j, \sigma_j^2) \alpha_j \quad (2)$$

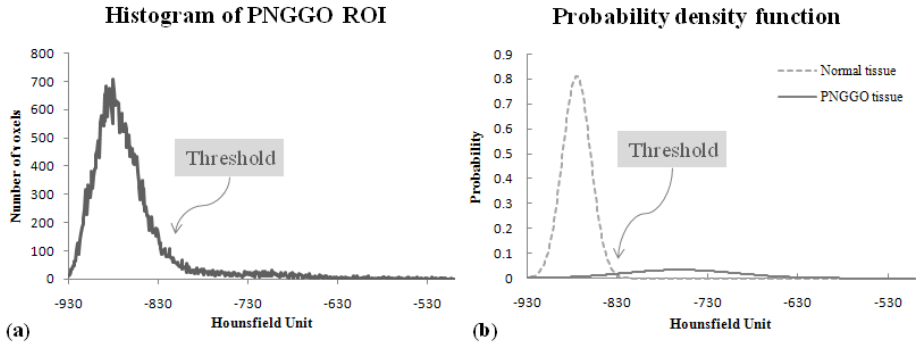
$$\frac{\partial}{\partial \mu_j} [\cdot] = 0, \quad \mu_j = \frac{\sum_{n=1}^N P(\omega_j | x_n, \theta) x_n}{\sum_{n=1}^N P(\omega_j | x_n, \theta)} \quad (3)$$

$$\frac{\partial}{\partial \sigma_j} [\cdot] = 0, \quad \hat{\sigma}_j^2 = \frac{1}{d} \frac{\sum_{n=1}^N P(\omega_j | x_n) \|x_n - \hat{\mu}_j\|^2}{\sum_{n=1}^N P(\omega_j | x_n)} \quad (4)$$

$$\frac{\partial}{\partial \alpha_j} [\cdot] = 0, \quad \hat{\alpha}_j = \hat{P}(\omega_j) = \frac{1}{N} \sum_{n=1}^N P(\omega_j | x_n) \quad (5)$$



**Figure 4. The CT number histogram of PNGGOs. a) benign lesion, b) malignant lesion**



**Figure 5. a) histogram of an ROI including PNGGO, b) estimated probability density functions of lung parenchyma and PNGGO.**

### 2.3.3. Region-growing segmentation

After histogram modeling, a region-growing technique is applied within the user specified ROI. The center of ROI was used as a seed point, and the threshold value was used to limit growing of the regions. Thus, the connected voxels having intensity values higher than the threshold value were captured as a candidate GGO nodule in this stage.

### 2.3.4. Vessel-like structure removal

Finally, morphological opening and labeling operations were applied to remove the unnecessary linear structures attached to GGO nodules which often results from region-growing segmentation step. The opening operation removes the attaching linear structures, and the following labeling technique removes small islands. Figure 2 (c) illustrates this procedure.

### 2.3.5. Evaluation of the proposed segmentation method

For quantitative evaluation of proposed segmentation method, we compared the segmentation results of the automated method with the segmentation results manually edited by a radiologist. For all 26 PNGGOs, a representative slice was selected that showed largest section nodule, and the segmentation result was given as a semi-transparent color overlay on top of CT image. Then, one radiologist (C.M.P) manually edited the segmentation result on a pixel-by-pixel basis. The added and removed pixels were painted in different colors (red for added, blue for removed).

After manual editing, segmentation accuracy of the proposed method was evaluated using the edited segmentation as a gold standard. The Hausdorff distance, Tanimoto coefficient and Dice similarity coefficient were used for this evaluation. (12, 13) The Hausdorff distance (HD) is a shape comparison metric based on binary image. The low Hausdorff Distance indicates high accuracy. The high Tanimoto coefficient and Dice similarity coefficient indicate high accuracy.

$$TanimotoCoefficient = \frac{N(A \cap B)}{N(A \cup B)} \quad (6)$$

$$DiceSimilarityCoefficient = \frac{2N(A \cap B)}{N(A) + N(B)} \quad (7)$$

$$HausdorffDistance = \max_{\sigma \in A} \left[ \min_{\sigma \in B} [d(a, b)] \right] \quad (8)$$



## 2.4. Computerized classification of PNGGOs

### 2.4.1. Statistical analysis

Statistical features for the segmented nodules were calculated using in-house CAD software. From the histogram of the segmented nodules, The CAD software quantified the following variables: (1) energy, (2) entropy, (3) mean CT number, (4) variance, (5) absolute deviation, (6) auto-correlation, (7) skewness, (8) kurtosis, (9) histogram at 25th percentile, (10) at 50th percentile, (11) at 75th percentile, (12) inter-quartile range, and (13) volume. The histogram percentile represents a CT number of the specific percentage of the pixels counted from the pixel with the lowest CT number.

Statistical differences between benign and malignant PNGGOs were analyzed using the independent sample t-test for individuals' statistical features. Features with a P-value less than 0.10 through univariate analysis were used as the independent variables for exhaustive search for optimal feature selection. Of the statistical features, entropy, mean, absolute deviation, skewness, kurtosis, 25th percentile, 50th percentile, 75th percentile and inter-quartile range were used as input variables for the exhaustive search method using a Support Vector Machine (SVM). Statistical analyses were performed using SPSS version 13.0 software for Windows (SPSS Inc., Chicago, Ill) and MedCalc 10.3 (MedCalc Software, Mariakerke, Belgium). A P-value less than 0.05 was considered to indicate statistical significance.

$$\text{Energy} = \sum_{i=0}^{L-1} [P(b_i)]^2 \quad (9)$$

$$\text{Entropy} = - \sum_{i=0}^{L-1} P(b_i) \cdot \log_2(b_i) \quad (10)$$

where  $L$  is amplitude range,  $b_i$  is intensity and  $P(b_i)$  is probability of  $b_i$ .

$$\text{Mean} = \frac{1}{N} \sum x_i \quad (11)$$

$$\text{Variance} = \frac{1}{(N-1)} \sum (x_i - \hat{\mu})^2 \quad (12)$$

$$\text{Absolute Deviation} = \frac{1}{N} \sum |x_i - \hat{\mu}| \quad (13)$$

$$\text{Auto Correlation} = \frac{\sum_{i=1}^n (x_i - \hat{\mu})(x_{i-1} - \hat{\mu})}{\sum_{i=1}^n (x_i - \hat{\mu})(x_i - \hat{\mu})} \quad (14)$$

$$\text{Skewness} = \frac{1}{N} \sum \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)^3 \quad (15)$$

$$\text{Kurtosis} = \left( \frac{1}{N} \sum \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)^4 \right) - 3 \quad (16)$$

where  $x_i$  are the elements of the dataset data.

$$\text{Histogram at } N^{\text{th}} \text{ percentile} = (1 - \delta)x_i + \delta x_{i+1} \quad (17)$$

$$\text{Inter Quartile Range} = \text{Percentile } 75^{\text{th}} - \text{Percentile } 25^{\text{th}} \quad (18)$$

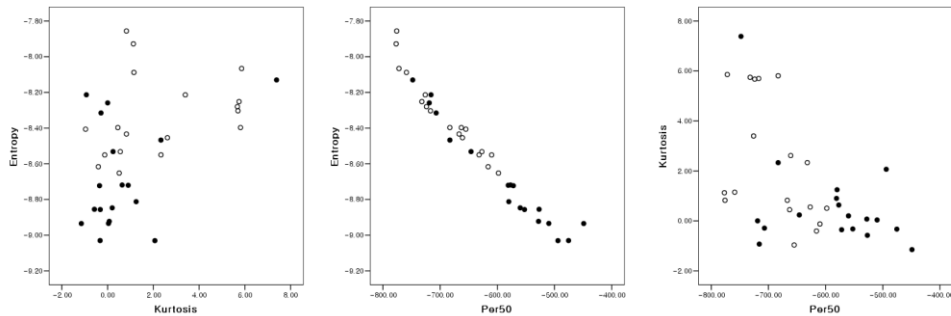
where  $i$  is  $\text{floor}((n-1)f)$  and  $\delta$  is  $(n-1)f - i$ .

$$\text{Volume} = (X\_SPACING \times Y\_SPACING \times \text{THICKNESS}) \times \text{CNT} \quad (19)$$

where  $X\_SPACING$  is  $x$ -axis pixel spacing,  $Y\_SPACING$  is  $y$ -axis pixel spacing,  $\text{THICKNESS}$  is slice thickness and  $\text{CNT}$  is the number of segmented voxels.

## 2.4.2. Classification of Benign and Malignant Lesion using a SVM

SVMs are a machine learning method for creating a classification function from a set of labeled training data. SVMs operate by locating a hyper-plane that attempts to split the training data into two categories. The hyper-plane is selected such that its distance to the nearest training data on either side of the surface is maximized. Classification of lesions can be performed as a two-class problem where the two classes are malignant and benign. The SVM-based classification used in our study was implemented using the libsvm library. Leave-one-out validation method and Receiver operating characteristic (ROC) analysis was also performed in order to evaluate the performance of the SVM classification model in discriminating between benign and malignant lesions.



**Figure 6. 2-D scatter plot depicting distributions and relationships of all two combinations of three features.**

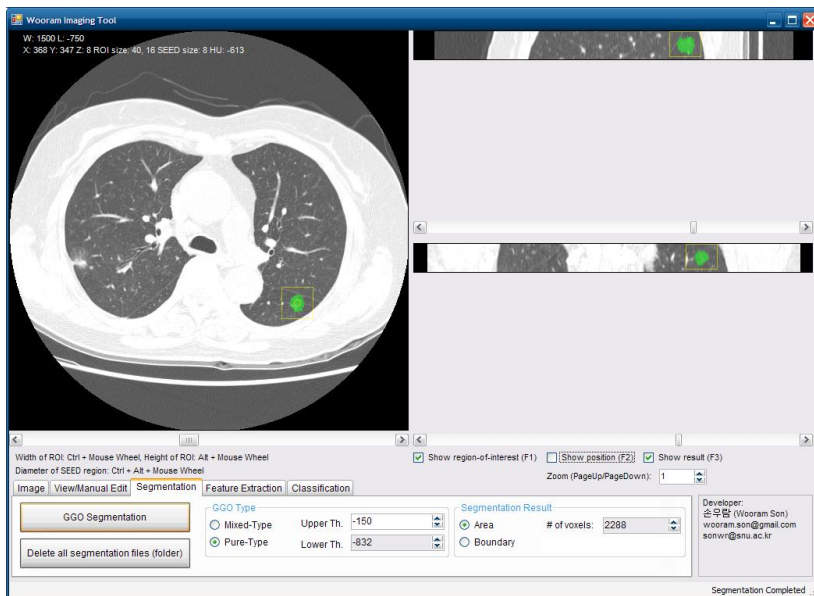
### 3. Results

#### 3.1. Performance of segmentation method

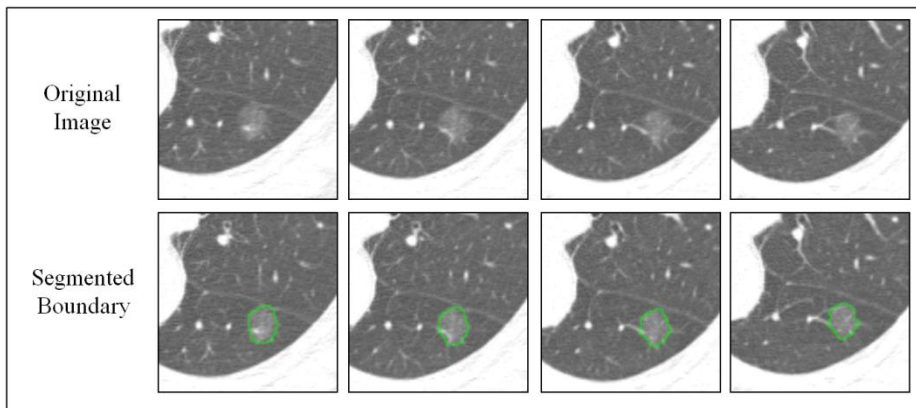
The results of segmentation accuracy using the Hausdorff distance, Tanimoto coefficient and Dice similarity coefficient are shown in Table 2.

**Table 2. Segmentation performance of the proposed method. The mean and standard deviation of the three indices between the automatic and manual segmentation were described.**

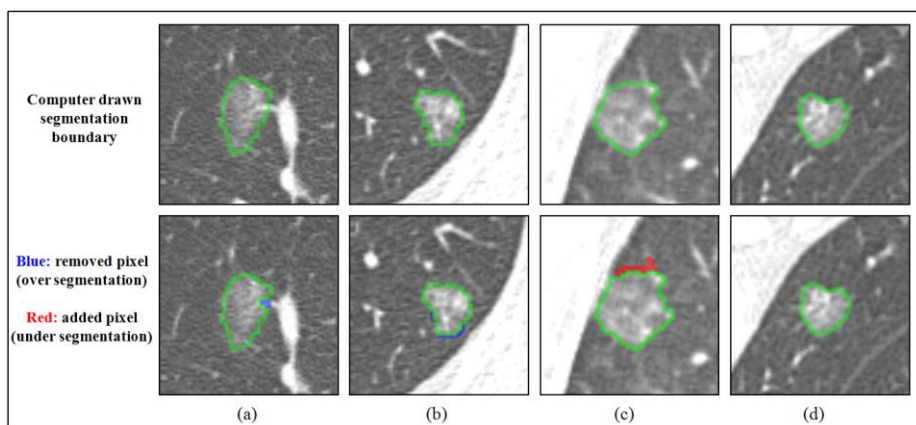
Performance Index	Mean (Standard deviation)
Hausdorff Distance	0.000105 (0.000150)
Tanimoto Coefficient	0.914693 (0.064465)
Dice Similarity Coefficient	0.954269 (0.036365)



**Figure 7. A screenshot of the developed PNGGO segmentation tool.**



**Figure 8. An example of PNGGO segmentation. The upper images are the largest slice CT image of the PNGGO. The segmentation results are in the lower images.**



**Figure 9. Automatic segmentation results are in the upper images. The lower images are comparing with manual segmentation performed by radiologist.**

**Table 3. The performance indices for the example 4 GGO nodules shown in figure 9.**

Performance Index	Figure 9a	Figure 9b	Figure 9c	Figure 9d
Hausdorff Distance	0.000091	0.000070	0.000105	0.000
Tanimoto Coefficient	0.974874	0.942177	0.927959	1.000
Dice Similarity Coefficient	0.987277	0.970228	0.962633	1.000

### 3.2. Feature evaluation and selection

Table 4 summarizes the predictability of all features in differentiating between benign and malignant PNGGOs along with the significant level and area under the ROC curve.

Entropy, mean CT number, 25th, 50th and 75th percentile of the histogram and inter-quartile range were found to have significantly larger feature values for malignant PNGGOs than benign PNGGOs. Skewness and kurtosis of the histogram showed significantly higher values for benign PNGGOs than malignant PNGGOs.

Characteristics with a P-value less than 0.10 through univariate analysis were used as the independent variables for exhaustive feature selection. Of these features able to differentiate between benign and malignant PNGGOs, entropy, kurtosis and 50th percentile of the histogram were proven through exhaustive search to be the best combination of input variables for the SVM classifier.

**Table 4. Features of benign and malignant PNGGOs.**

Characteristics	Benign PNGGO (n=18) <sup>†</sup>	Malignant PNGGO (n=18) <sup>†</sup>	AUC	P-value <sup>‡</sup>
Energy	0.0044 ±0.004	0.0029 ±0.0013	.713	0.1465
Entropy	-8.3319 ±0.2303	-8.6834 ±0.2914	.815	0.0003
Mean	-670.5682 ±61.1724	-575.7147 ±87.0633	.812	0.0006
Variance	21649.2575 ±11809.3177	27887.408 ±15089.1758	.608	0.1762
Abs Dev	107.8646 ±33.5533	128.8102 ±40.2401	.651	0.099
Autocorrelation	0.985 ±0.0205	0.9887 ±0.0208	.623	0.5951
Skewness	1.0798 ±0.6991	0.6068 ±0.5304	.691	0.0286
Kurtosis	2.2819 ±2.446	0.6206 ±1.927	.731	0.0301
Percentile at 25 <sup>th</sup>	-769.4028 ±32.35	-698.5139 ±64.3332	.781	0.0003
Percentile at 50 <sup>th</sup>	-688.6111 ±60.5884	-590.2778 ±91.5083	.816	0.0006
Percentile at 75 <sup>th</sup>	-598.5972 ±89.9644	-473.3194 ±118.6607	.809	0.0011
Inter-quartile Range	170.8056 ±64.0531	225.1944 ±81.094	.707	0.0322
Volume	1457.355 ±2051.0781	911.767 ±1124.6849	.503	0.3294

Characteristics with P-value less than 0.10 are marked with gray-shaded row

† Data are mean ± standard deviation.

‡ Independent sample T-test

### 3.3. Classification performance of SVMs

ROC analysis was conducted to evaluate performance of the SVM classifier in discriminating benign PNGGOs from malignant PNGGOs using entropy, kurtosis and histogram percentile at 50th features. When leave-one-out used as validation method, the area under the ROC curve (Az) was 0.914 (95% CI: 0.771 to 0.980). When 9-fold cross-validation used as validation method, the area under the ROC curve (Az) was 0.910 (95% CI: 0.802 to 1.000). When 4-fold cross-validation used as validation method, the area under the ROC curve (Az) was 0.870 (95% CI: 0.736 to 1.000). When 2-fold cross-validation used as validation method, the area under the ROC curve (Az) was 0.806 (95% CI: 0.657 to 0.954).

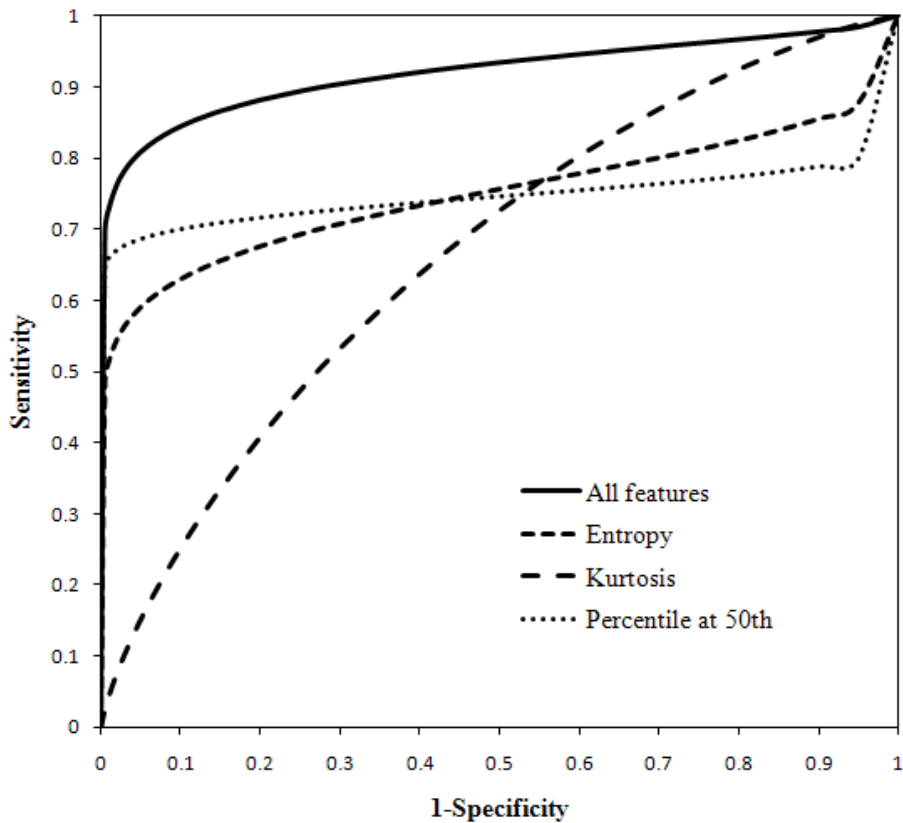


Figure 10. ROC curves of the uses of three individual features and combination of all three features for PNGGO differentiation.



**Table 5. Result of leave-one-out trials – Diagnostic performance of SVMs in PNGGOs.**

Features	Trials	Az	Standard Error	Significance level	95% CI
Entropy, Kurtosis, Percentile at 50 <sup>th</sup>	Leave-one-out	0.914	0.0505	0.0001	0.771 to 0.980
	9-fold cross-validation	0.910	0.055	0.0001	0.802 to 1.000
	4-fold cross-validation	0.870	0.068	0.0001	0.736 to 1.000
Entropy	Leave-one-out	0.759	0.089	0.008	0.584 to 0.934
	9-fold cross-validation	0.756	0.093	0.009	0.574 to 0.939
	4-fold cross-validation	0.759	0.088	0.008	0.586 to 0.932
Kurtosis	Leave-one-out	0.701	0.091	0.040	0.522 to 0.879
	9-fold cross-validation	0.710	0.090	0.031	0.534 to 0.886
	4-fold cross-validation	0.639	0.095	0.155	0.454 to 0.824
Percentile at 50 <sup>th</sup>	Leave-one-out	0.725	0.099	0.021	0.531 to 0.920
	9-fold cross-validation	0.721	0.101	0.024	0.524 to 0.918
	4-fold cross-validation	0.721	0.099	0.024	0.527 to 0.914

## 4. Discussion

We developed the CAD system for PNGGOs and evaluated the SVM classifier for differential diagnosis of PNGGOs using CT number features. The performance of our proposed differentiation scheme for PNGGOs was 0.91 Az (Sensitivity: 0.833, Specificity: 0.889) using ROC analysis. We have been performed study previously regarding the detection of PNGGOs and the predictive CT findings of malignancy in GGOs on thin-section chest CT previously. (3, 14) This study was our first research regarding the differentiation of benign and malignant PNGGOs using a CAD system. Our results show that our proposed classification scheme has promising potential for reliable differentiation of PNGGOs.

Nomori study which regarding the differential diagnosis of GGOs using CT number analysis by three-dimensional computerized quantification conclude that histogram percentile at 75th (0.906 Az for differentiation between AAH and BAC) with a cutoff value of -584 Hounsfield units (HU) was optimal. (8) However, in our experimental, histogram percentile at 75th shows poor performance (0.809 Az) than Nomori's result. It is presumed that because of the present study did not involve mixed-type GGOs. Mixed-type GGOs have a much higher malignancy rate than other types GGO nodules. Part-solid component of mixed-type GGOs have shown higher intensity values than non-solid regions of pure-type GGOs. Therefore, malignant mixed-type GGOs have shown higher histogram percentile at 75th than malignant PNGGOs.

SVMs have been shown to perform well as a classification model for medical imaging, (15, 16) and have also been shown to perform well when dealing with relatively small training sets. (16) This was particularly appealing given the inherent difficulty in acquiring large amounts of screening

data devoted exclusively to training. According to the Nomori's report, histogram analysis of CT number was useful for differentiation between benign and malignant GGOs. (8) Therefore, we used histogram-based features and evaluate these features. To find the best combination of features as an input variable of SVMs, we used univariate analysis and exhaustive feature selection method. For the evaluation of classification model, we used leave-one-out and cross-validation method.

Our study has several limitations. First is a small study population. Therefore, we need to further study for large study population. Second is inter-, intra-observer variation of semi-automated segmentation method. We measure the difference of classification performance between two segmentation results by two observers (C.M.P, W.R.S). W.R.S's result shows one more true-negative and one less false-positive than C.M.P's result. There are no significant different between two differentiation performances from ROC analysis.

## 5. Conclusion

In this paper, we studied the use of histogram-based features and SVMs as a classifier for differentiation between benign and malignant lesions from chest CT. We evaluate the performance of individual features and combination of features as an input variables of SVMs. Results from this study suggest that SVMs and histogram-based features as an input variables of the SVMs are a useful for CAD of PNGGOs. Future work will focus on the radiologist performance improvement with our proposed computer-aided diagnosis system.

## 6. Reference

1. Henschke C, Yankelevitz D, Mirtcheva R, McGuinness G, McCauley D, Miettinen O. CT screening for lung cancer frequency and significance of part-solid and nonsolid nodules. *Am Roentgen Ray Soc*, 2002; p. 1053-7.
2. Lee S, Park C, Goo J, et al. Transient Part-Solid Nodules Detected at Screening Thin-Section CT for Lung Cancer: Comparison with Persistent Part-Solid Nodules. *Radiology*. 2010.
3. Lee H, Goo J, Lee C, et al. Predictive CT findings of malignancy in ground-glass nodules on thin-section chest CT: the effects on radiologist performance. *European radiology*. 2008.
4. Patsios D, Roberts H, Paul N, et al. Pictorial review of the many faces of bronchioloalveolar cell carcinoma. *British Journal of Radiology*. 2007;80(960):1015.
5. Kim H, Shim Y, Lee K, Han J, Yi C, Kim Y. Persistent Pulmonary Nodular Ground-Glass Opacity at Thin-Section CT: Histopathologic Comparisons. *Radiology*. 2007;245(1):267.
6. Ohtsuka T, Watanabe K, Kaji M, Naruke T, Suemasu K. A clinicopathological study of resected pulmonary nodules with focal pure ground-glass opacity. *European Journal of Cardio-Thoracic Surgery*. 2006;30(1):160-3.
7. Nakajima R, Yokose T, Kakinuma R, Nagai K, Nishiwaki Y, Ochiai A. Localized Pure Ground-Glass Opacity on High-Resolution CT: Histologic Characteristics. *Journal of Computer Assisted Tomography*. 2002;26(3):323.
8. Ikeda K, Awai K, Mori T, Kawanaka K, Yamashita Y, Nomori H. Differential Diagnosis of Ground-Glass Opacity Nodules: CT Number Analysis by Three-Dimensional Computerized Quantification. *Chest*.

- 2007;132(3):984.
9. Son W, Park S, Park C, Kim J. Classification of Benign/Malignant PNGGOs using K-means algorithm in MDCT Images: A Preliminary Study. *IEICE Tech Rep.* 2009;108(385):257-60.
  10. Kim WSSJPCMPJMGJH. Automated volumetric segmentation method for computerized-diagnosis of pure nodular ground-glass opacity in high-resolution CT. *SPIE Medical Imaging. San Diego, California, USA2010*; p. 76241P.
  11. Browder W, Reeves A, Apananosovich T, Cham M, Yankelevitz D, Henschke C. Automated volumetric segmentation method for growth consistency of nonsolid pulmonary nodules in high-resolution CT. *SPIE Medical Imaging. San Diego, CA, USA: 2007, 2007*; p. 65140Y.
  12. Theodoridis S, Koutroumbas K. *Pattern Recognition.* 1999. Academic Press.
  13. Huttenlocher D, Klanderman G, Rucklidge W. Comparing images using the Hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence.* 1993;15(9):850-63.
  14. Kim K, Goo J, Kim J, et al. Computer-aided Diagnosis of Localized Ground-Glass Opacity in the Lung at CT: Initial Experience 1. *Radiology.* 2005;237(2):657-61.
  15. Chang R, Wu W, Moon W, Chou Y, Chen D. Support vector machines for diagnosis of breast tumors on US images. *Academic radiology.* 2003;10(2):189-97.
  16. El-Naqa I, Yang Y, Wernick M, Galatsanos N, Nishikawa R. A support vector machine approach for detection of microcalcifications. *IEEE Transactions on Medical Imaging.* 2002;21(12):1552-63.

## 초 록

폐암은 전세계적으로 유병율이 높은 질환으로써, 2008년 통계청의 자료에 의하면 암 사망원인 1위이다. 최근 조기 폐암 발견을 위한 CT 검사가 임상 현장에서 광범위하게 사용됨으로써, 과거 일반 X-선 흉부 영상에서는 보이지 않았던 결절형 간유리 음영이 흔히 발견되고 있다. 결절형 간유리 음영은 CT 영상에서 혈관을 가리지 않고 나타나는 간유리 형태의 음영영역으로 추적검사에서 없어지거나, 작아지지 않는 경우에는 세기관지 폐포성 선암이나, 폐선암의 가능성이 매우 높다.

추적검사서 없어지거나, 작아지지 않는 결절형 간유리 음영 중에 병변내부에 고형성 부분이 보이지 않는, 순수 결절형 간유리 음영은 국소 간질성 섬유증이나 비정형 선종성 증식증 등의 양성 병변의 가능성이 높아서 이들을 모두 절제할 경우 환자에게 불필요한 절제수술을 시행할 가능성이 높다. 하지만, 지속성 순수 결절형 간유리 음영 병변들 중에서 18% ~ 48%의 병변이 조기 폐암이란 점에서, 결코 간과할 수 없는 병변이다. 그러므로, 순수 결절형 간유리 음영으로 보이는 병변을, 악성 병변과 양성 병변으로 구분해 내는 일이 매우 중요하며, 이를 통해, 양성 병변의 불필요한 수술을 막을 수 있으며, 악성 병변의 경우, 조기 진단 및 치료를 통해, 환자의 삶의 질과 생명을 연장할 수 있다.

본 연구의 목적은 순수 결절형 간유리 음영의 양성과 악성 여부를 판별할 수 있는 컴퓨터보조진단 시스템의 개발이다. 우리가 제안하는 시스템은 간유리 음영 결절의 자동분할, 통계기반의 특징벡터추출 그리고 Support Vector Machine (SVM) 을 이용한

분류모델로 구성되어있다. 34 CT 볼륨 이미지에서 얻은 36개의 순수 결절형 간유리 음영에 대한 실험 결과 leave-one-out 검증기법과 ROC분석을 통하여 0.91 areas under the curve (AUC) 수치를 얻었다. SVM 분류모델의 입력으로써, 최적의 성능을 보여주는 특징벡터의 조합은 Entropy, Kurtosis, Histogram at 75<sup>th</sup> percentile로 나타났다. 이와 같은 실험결과는 순수 결절형 간유리 음영의 정확한 악성도 진단에 있어서 우리가 제안한 컴퓨터보조진단 시스템의 가능성을 보여주었고 향후 방사선과 의사에 의한 진단에 보조시스템으로써 상보적 역할의 가능성을 보여주었다.

**주요어:** 순수 결절형 간유리 음영, 컴퓨터보조진단, Pure Nodular Ground-Glass Opacity (PNGGO), Support Vector Machine (SVM), Computer-aided Diagnosis (CAD).

**학 번:** 2008-23287