

Semantics and image content integration for pulmonary nodule interpretation in thoracic computed tomography

Daniela S. Raicu^a, Ekarin Varutbangkul^a, Janie G. Cisneros^a, Jacob D. Furst^a,
David S. Channin^b, Samuel G. Armato III^c

^aIntelligent Multimedia Processing Laboratory, School of Computer Science, Telecommunications, and Information Systems, DePaul University, Chicago, IL 60604, USA;

^bDepartment of Radiology, Northwestern University Medical School, Chicago, IL 60611, USA;

^cDepartment of Radiology, The University of Chicago, Chicago, IL 60637, USA

ABSTRACT

Useful diagnosis of lung lesions in computed tomography (CT) depends on many factors including the ability of radiologists to detect and correctly interpret the lesions. Computer-aided Diagnosis (CAD) systems can be used to increase the accuracy of radiologists in this task. CAD systems are, however, trained against ground truth and the mechanisms employed by the CAD algorithms may be distinctly different from the visual perception and analysis tasks of the radiologist. In this paper, we present a framework for finding the mappings between human descriptions and characteristics and computed image features. The data in our study were generated from 29 thoracic CT scans collected by the Lung Image Database Consortium (LIDC). Every case was annotated by up to 4 radiologists by marking the contour of nodules and assigning nine semantic terms to each identified nodule; fifty-nine image features were extracted from each segmented nodule. Correlation analysis and stepwise multiple regression were applied to find correlations among semantic characteristics and image features and to generate prediction models for each characteristic based on image features. From our preliminary experimental results, we found high correlations between different semantic terms (margin, texture), and promising mappings from image features to certain semantic terms (texture, lobulation, spiculation, malignancy). While the framework is presented with respect to the interpretation of pulmonary nodules in CT images, it can be easily extended to find mappings for other modalities in other anatomical structures and for other image features.

Keywords: LIDC, shape, texture, stepwise multiple regression, correlation analysis

1. INTRODUCTION

Research studies have shown that double reading by two or more radiologists improves the detection of lung cancers, and the same have shown that interpretation performance varies greatly among radiologists.¹ Computer-aided diagnosis (CAD) systems can be used as a *second reader* to improve the overall accuracy of radiologists in this task. Since the CAD algorithms typically operate very differently from human perception it can be difficult at times to understand how certain findings were made.

The purpose of this research is to provide a quantitative approach for finding the relationships between image features and medical terms (semantic concepts) used for diagnosis and finding mappings from image features to semantic terms. Our approach can assist radiologists in interpreting lung nodules by providing either an initial or secondary estimate of the semantic values based on the calculation and subsequent analysis of image features. This can serve to increase the accuracy of a single reader and improve the consistency among multiple readers.

2. RELATED WORK

Several research studies have designed CAD systems that can help estimate the probability of cancer based on nodule characteristics (such as nodule size, shape, and internal structures) and clinical information (such as age, gender, history of smoking, and history of cancer).

McNitt-Gray et al.^{2, 3} used nodule size, shape and co-occurrence texture features as nodule characteristics to design a linear discriminant analysis classification system for malignant versus benign nodules. Lo et al.⁴ used direction of vascularity, shape and internal structure to build an artificial neural network classification system for prediction of the malignancy of the nodules. Armato and MacMahon⁵ used nodule appearance and shape to build a LDA classification system of pulmonary nodules in malignant versus benign classes. Takashima et al.^{6, 7} used shape information to characterize malignant versus benign lesions in the lung.

While these systems are based on just nodule characteristics, there are also studies that make use of clinical information. J. Gurney designed a Bayesian classification system^{8, 9} based on clinical information in addition to radiological information. Y. Matsuki et al.¹⁰ also used clinical information in addition to sixteen features scored by radiologists to design an ANN for malignancy versus benign classification. M. Aoyama et al.¹¹ used two clinical features in addition to forty-one image features to determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images. For a more detailed literature review on the analysis of computed tomography scans of the lung, we recommend the recent survey by I. Sluimer et al.¹²

All these studies are performed on different data sets and therefore, a comparison of the algorithms presented by these studies is not reliable. The necessity of creating benchmark data sets has been recognized and the Lung Image Database Consortium¹³ (LIDC) benchmark is a first step undertaken in this direction. The LIDC has been already used for the validation of two detection algorithms. Using the LIDC cases, Liu and Li¹⁴ also proposed a new method for nodule detection based on gradient and intensity combined level set methods that generated stable and accurate segmentation results for complex organic structures like lung bronchia and nodules.

All studies used a combination of features to characterize the size, shape and internal structure of the nodules. In that way, they indirectly encoded radiologists' knowledge about indicators of malignancy¹². However, the real question is if indeed the nodule characteristics as encoded by the CAD systems correspond to the same concepts as the ones used by the radiologists. In other words, it is important to have a correct mapping between the nodule image features and the semantic concept used by radiologist to annotate the nodule. For example, can we give a positive answer to the question: are the computerized texture features quantifying the texture as perceived by radiologists? In other words, is computed texture the same as the perceived texture? Furthermore, is the texture of a nodule perceived the same by two different radiologists? It was shown that when human ratings of nodule characteristics are used to train computer systems, such ratings are not always reliable and reproducible¹⁵⁻¹⁸. Moreover, in practice, physicians use several perceptual categories to make diagnoses. First steps have already been taken in creating frameworks for a common language when making diagnosis. Leroy and Chen¹⁹ developed a tool (Medical Concept Mapper) based on the Unified Medical Language System²⁰ (UMLS) and WordNet²¹ that connects patient information to human-created ontologies. Barb et al.²² proposed a framework that uses semantic methods to describe visual abnormalities and exchange knowledge in the medical domain. Our work can also be considered one of the initial steps in the direction of mapping image features to perceptual categories encoding the radiologists' knowledge for lung interpretation.

3. METHODOLOGY

In this section we present our proposed methodology for finding the mappings between the nodule image features and the physician annotations as summarized in Figure 1. In Section 3.1 we present the data set. In Section 3.2 we present the image processing algorithms used to perform automatic nodule feature extraction and in Section 3.3 we present the data analysis and machine learning algorithms used to map the image features to the physicians' annotations.

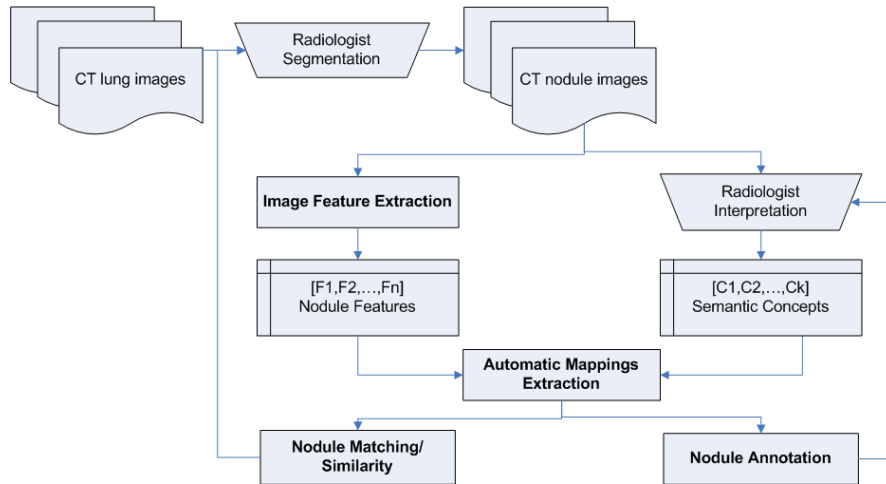


Figure 1: Diagram of the proposed mapping framework

3.1 Data Set

The data used in this study were generated from 29 cases of thoracic CT collected by the Lung Image Database Consortium (LIDC).¹³ In the LIDC's marking process, up to 4 radiologists marked the boundary of lung nodules with sizes between 3 mm and 3 cm for every slice on which the nodule appears and rated nine semantic characteristics for each identified nodule. The nine semantic characteristics chosen by the LIDC are shown in the Table 1, along with our notes and references on the meaning of these characteristics based on our literature review, and the range of possible scores from which the rating radiologists could choose. In the marking process, there is no forced agreement on the existence of a nodule, its location, its boundary, or its characteristics¹³.

Table 1: LIDC nodule characteristics with corresponding lung nodules

Characteristic	Notes and References	Possible Scores
<i>Calcification</i>	calcification appearance in the nodule - The smaller the nodule, the more likely it must contain calcium in order to be visualized. ²³ Benignity is highly associated with central, non-central, laminated, and popcorn calcification. ^{24,25}	<ol style="list-style-type: none"> 1. Popcorn 2. Laminated 3. Solid 4. Non-central 5. Central 6. Absent
<i>Internal structure</i>	expected internal composition of the nodule	<ol style="list-style-type: none"> 1. Soft Tissue 2. Fluid 3. Fat 4. Air
<i>Lobulation</i>	whether a lobular shape is apparent from the margin or not - lobulated margin is an indication of benignity. ¹²	<ol style="list-style-type: none"> 1. Marked 2. . 3. . 4. . 5. None
<i>Malignancy</i>	likelihood of malignancy of the nodule - Malignancy is associated with large nodule size while small nodules are more likely to be benign. ^{24,26} Most malignant nodules are non-calcified ³ and have spiculated margins. ²⁷	<ol style="list-style-type: none"> 1. Highly Unlikely 2. Moderately Unlikely 3. Indeterminate 4. Moderately Suspicious 5. Highly Suspicious
<i>Margin</i>	how well defined the margins of the nodule are	<ol style="list-style-type: none"> 1. Poorly Defined 2. . 3. . 4. . 5. Sharp
<i>Sphericity</i>	dimensional shape of nodule in terms of its roundness	<ol style="list-style-type: none"> 1. Linear 2. . 3. Ovoid 4. . 5. Round

<i>Spiculation</i>	degree to which the nodule exhibits spicules, spike-like structures, along its border - Spiculated margin is an indication of malignancy. ^{25,28}	1. Marked 2. . 3. . 4. . 5. None
<i>Subtlety</i>	difficulty in detection - Subtlety refers to the contrast between the lung nodule and its surroundings	1. Extremely Subtle 2. Moderately Subtle 3. Fairly Subtle 4. Moderately Obvious 5. Obvious
<i>Texture</i>	internal density of the nodule - Texture plays an important role when attempting to segment a nodule, since part-solid and non-solid texture can increase the difficulty of defining the nodule boundary. ²⁶	1. Non-Solid 2. . 3. Part Solid/(Mixed) 4. . 5. Solid

3.2 Feature extraction

In order to quantify the image content, we calculated four types of image features for each nodule: size, shape, intensity, and texture; this feature extraction stage generated 59 image features as presented in Table 2. The choice of these features was based on a literature review of the most common image features used for pulmonary nodule detection and diagnosis by existent CAD systems^{28, 29}.

Table 2: The entire set of image features extracted from each segmented lung nodule.

Shape Features	Size Features	Intensity Features	Texture Features
Circularity	Area	MinIntensity	11 Haralick features calculated from co-occurrence matrices (Contrast, Correlation, Entropy, Energy, Homogeneity, 3 rd Order Moment, Inverse Differential Moment, Variance, Sum Average, Cluster Tendency, Maximum Probability)
Roughness	ConvexArea	MaxIntensity	
Elongation	Perimeter	MeanIntensity	
Compactness	ConvexPerimeter	SDIntensity	
Eccentricity	EquivDiameter	MinIntensityBG	
Solidity	MajorAxisLength	MaxIntensityBG	
Extent	MinorAxisLength	MeanIntensityBG	24 Gabor features - mean and standard deviation of Gabor filters consistency of four orientations and three scales.
RadialDistanceSD		SDIntensityBG	
		IntensityDifference	

In order to quantify the shape of a nodule, we used eight common image shape features: circularity, roughness, elongation, compactness, eccentricity, solidity, extent, and the standard deviation of the radial distance. *Circularity* is measured by dividing the area of the region by the area of a circle with the same convex perimeter. *Roughness* can be measured by dividing the perimeter of the region by the convex perimeter. A smooth convex object, such as a perfect circle, will have a roughness of 1.0. The *eccentricity* is obtained using the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 (a perfect circle) and 1 (a line). *Solidity* is defined in terms of the convex hull corresponding to the region being the proportion of the pixels in the convex hull that are also in the region. *Extent* is the proportion of the pixels in the bounding box (the smallest rectangle containing the region) that are also in the region. Finally, the *RadialDistanceSD* is the standard deviation of the distances from every boundary pixel to the centroid of the region.

For the size of a nodule, the following seven features were found to be the most common ones: area, convexArea, perimeter, convexPerimeter, equivDiameter, majorAxisLength, minorAxisLength. The *area* and *perimeter* image features measure the actual number of pixels in the region and on the boundary, respectively. The *convexArea* and *convexPerimeter* measure the number of pixels in the convex hull and on the boundary of the convex hull corresponding to the nodule region. *EquivDiameter* is the diameter of a circle with the same area as the region. Lastly, the *majorAxisLength* and *minorAxisLength* give the length (in pixels) of the major and minor axes of the ellipse that has the same normalized second central moments as the region.

Gray-level intensity features used in this study are simply the *minimum*, *maximum*, *mean*, and *standard deviation* of the gray-level intensity of every pixel in each segmented nodule image and the same four values for every background pixel in the bounding box containing each segmented nodule image. Another feature, *intensityDifference*, is the absolute value

of the difference between the mean of the gray-level intensity of the segmented nodule image and the mean of the gray-level intensity of its background.

Normally texture analysis can be grouped into four categories: model-based, statistical-based, structural-based, and transform-based methods. Structural approaches seek to understand the hierarchical structure of the image, while statistical methods describe the image using pure numerical analysis of pixel intensity values. Transform approaches generally perform some kind of modification to the image, obtaining a new “response” image that is then analyzed as a representative proxy for the original image, and model-based methods are based on the concept of predicting pixel values based on a mathematical model. Based on our previous texture analysis work³¹, in this research we focus on two well-known texture analysis techniques: *co-occurrence matrices* (a statistical-based method), and *Gabor filters* (a transform-based method).

Co-occurrence matrices focus on the distributions and relationships of the gray-level intensity of pixels in the image. They are calculated along four directions (0°, 45°, 90°, and 135°) and five distances (1, 2, 3, 4 and 5 pixels) producing 20 co-occurrence matrices. Once the co-occurrence matrices are calculated, eleven Haralick texture descriptors³¹ are then calculated from each co-occurrence matrix. Although each Haralick texture descriptor is calculated from each co-occurrence matrix, we averaged the features by distance and then select the minimum value by direction resulting in 11 (instead of 11*4*5) Haralick features per image.

Gabor filtering³² is a transform based method which extracts texture information from an image in the form of a response image. A Gabor filter is a sinusoid function modulated by a Gaussian and discretized over orientation and frequency. We convolve the image with 12 Gabor filters: four orientations (0°, 45°, 90°, and 135°) and three frequencies (0.3, 0.4, and 0.5), where frequency is the inverse of wavelength. The size of each Gabor filter is set constant at 9×9. Our Gabor filter design is based on the work by Andrysiak and Choras,³² where Gabor filters were used to encode the image content for image retrieval. We then calculate means and standard deviations from the 12 response images resulting in 24 Gabor features per image.

At the end of the image feature extraction process, each nodule image is encoded using a set of fifty-nine image features $f_i, i = 1, \dots, 59$ and nine radiologist annotations $c_j, j = 1, \dots, 9$ (semantic concepts). Therefore, the nodule representation is given by the vector representation $[f_1, f_2, \dots, f_{59}, c_1, c_2, \dots, c_9]$; Figure 2 shows an example of feature values for a nodule representation. We should note here that each nodule might produce up to four images in the data set given that same nodule can be delineated differently by the four radiologists. Furthermore, since the same nodule can be rated differently by the four radiologists, for the same nodule image there might be up to four vector representations each representation encoding the rating of one of the radiologists.

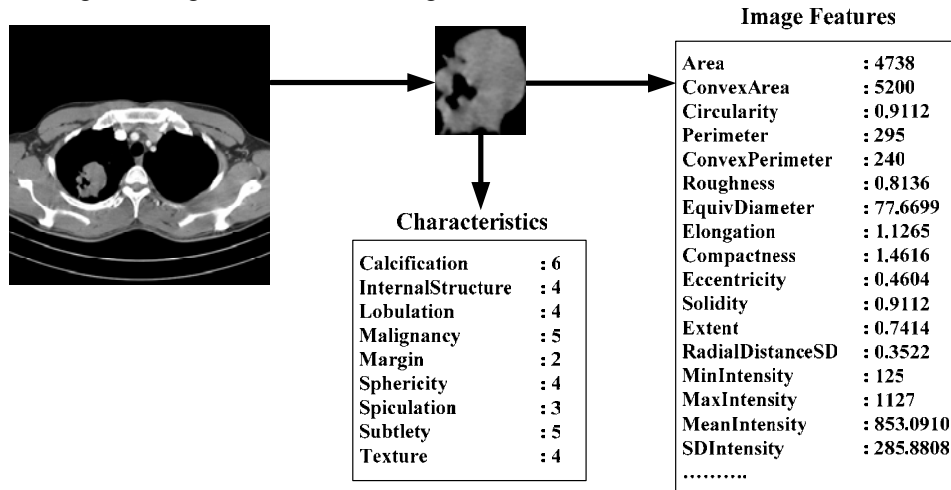


Figure 2: An example of nodule characteristics assigned by a radiologist and features extracted from the segmented nodule.

3.3 Automatic Mappings Extraction

In the image indexing and retrieval community, the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data has for a user in a given situation is known as *the semantic gap problem*. A look on the state-of-the-art in the image indexing and retrieval domain shows the importance and the complexity of the semantic problem. In this paper, we address this problem as encountered in the lung nodule image interpretation by proposing two basic statistical techniques to identify the relationships between the digital representation of a nodule image and its interpretation as perceived by up to four radiologists.

Because the radiologists were asked to provide quantified semantic descriptions, this data set lends itself well to basic statistical methods. First, we chose to concentrate on correlations among all pairs of semantic ratings as well as all (*semantic rating, image feature*) pairs. Using the nodule vector representation introduced in the previous section, the Spearman's rank correlation coefficient³⁴ among two semantic ratings c_i, c_j is defined as follows:

$$\rho(c_i, c_j) = 1 - \frac{6 \sum_{k=1 \dots n} (r_{c_i} - r_{c_j})^2}{n(n^2 - 1)}, \quad i, j = 1 \dots 9 \quad (1)$$

where n stands for the number of data points (nodule images), r_{c_i}, r_{c_j} are the ranks of the semantic ratings c_i, c_j , respectively. Similarly, the correlation among a semantic rating c_i and an image feature f_j is defined as follows:

$$\rho(c_i, f_j) = 1 - \frac{6 \sum_{k=1 \dots n} (r_{c_i} - r_{f_j})^2}{n(n^2 - 1)}, \quad i = 1 \dots 9, \quad j = 1 \dots 59 \quad (2)$$

where r_{f_j} is the rank of the image feature f_j . Values of the correlation coefficient ρ close to 1 will indicate a positive strong correlation between the considered pairs. Further, the pair-wise results for basic correlations can be used to compare our results to what has been reported in previous studies regarding the relationship between lung nodule semantic descriptions and physical descriptions of the nodule.

However, the correlation models can only account for pair-wise groupings of semantic ratings and thus provide limited predictive power. To investigate further the interaction of all images features and their combined relationship to the semantic ratings, we used a stepwise multiple regression.

Stepwise multiple regression analysis³³ was applied to generate prediction models M_i for each characteristic c_i based on all image features f_j :

$$M_i: \quad c_i = \beta_0 + \sum_{k=1, \dots, p} \beta_k f_k + \varepsilon_i \quad (3)$$

where p is the number of image features used in the regression model, β_i are the regression coefficients, and ε_i are the prediction errors per model. A model M_i was considered to be a good fit if the adj_R^2 (the square of the adjusted correlation coefficient) was greater than 0.8 which implies that the model captured more than 80% of the variance in the data:

$$adj_R^2 = 1 - (1 - R^2) \frac{(n-1)}{(n-p-1)} \quad (4)$$

The adjusted value of the correlation coefficient R^2 was used to take into account the multi-collinearity in the data (two or more image features are dependent). Unlike the R^2 , the adj_R^2 can decline if the addition of variable does not improve the regression model.

Besides modeling each semantic term or characteristic with respect to only the image features and thus including all the nodule images in the analysis, we also looked into the modeling of the relationships among the features and semantic terms when: 1) at least 2 radiologists and 2) at least 3 radiologists agreed on the same rating for the characteristic to be predicted. The adj_R^2 calculated for each characteristic was also used to measure the improvement in the predicted multiple regression models when there was better agreement among the radiologists. The stepwise feature selection incorporated into the regression model also provided a way to measure the contribution of each of the selected image features to predict the corresponding semantic characteristic. Furthermore, the linear combination of the most important features can be used to quantify the human perception of the corresponding visual characteristics as perceived by the radiologists and thus allows the integration of semantics and low-level image content.

4. RESULTS

From our experiments, we found many interesting correlations among nodule characteristics and image features, and statistically significant prediction models were obtained for several nodule characteristics.

4.1 Correlation analysis

For this work, we calculated 42 correlation coefficients for comparing semantic ratings. The data set for each rating was simply the set of scores that all radiologists provided for all nodules; the calcification and internal structure ratings were not included in the correlation analysis since they have categorical values. Since all ratings appear together for each nodule, the correspondence between features is trivial. We also calculated 413 correlation coefficients for each combination of semantic rating and image feature. While the number of coefficients was large, this kind of a basic statistical analysis gives us a very good picture of relationships between semantic concepts and between concepts and image features.

Given the variability in the radiologist ratings and the complexity of such an annotation process, we considered as meaningful associations even those associations that produced correlation values greater than 0.3. As shown in Figure 3, we found that:

1) Subtlety and malignancy (as assessed by the radiologists) are correlated with each other and both of them are correlated with many size features (Area, ConvexArea, Perimeter, ConvexPerimeter, EquivDiameter, MajorAxisLength, and MinorAxisLength). These findings were also supported by the literature review that malignancy is associated with the nodule size.^{25,27}

2) Sphericity is correlated with some shape features that are related to roundness of the region (Elongation, Eccentricity, Extent, and Circularity) as expected.

3) Margin and texture are correlated with each other and this correlation was in concordance with other research studies showing that texture plays an important role in nodule segmentation since ‘part-solid’ and ‘non-solid’ texture can increase the difficulty of defining the nodule boundary.³⁰ Furthermore, margin and texture are not directly correlated with any of the individual image features. Same pattern was noticed for lobulation and spiculation which were correlated with each other but not correlated with any of the individual image features.

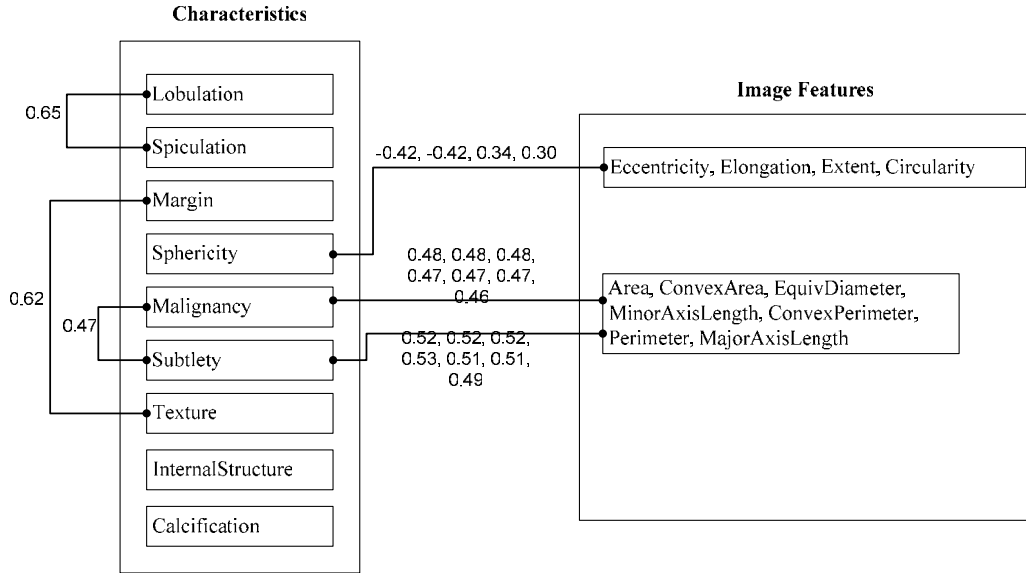


Figure 3: Correlations among nodule characteristics and image features

4.2 Stepwise multiple regression analysis

The adj_R^2 values of all regression models for all characteristics are presented in Table 3. The second column indicates the results on all nodules in the data set, while the third column shows results only for images on which at least two radiologists agreed on a rating for that nodule. The fourth column shows similar results for nodules on which at least three radiologists agreed on a rating. From the stepwise multiple regression analysis, we learned that the image features collected for this study can be used to produce good prediction models for *malignancy*, *lobulation*, *texture*, and *spiculation* (adj_R^2 values of 0.990, 0.877, 0.843, and 0.840 respectively); all the models were statistically significant ($\alpha = 0.05$) as well.

As we expected, the prediction models improved as the agreement among the radiologists improved; for example, the adj_R^2 went up from .310 to 0.641 when instead of considering all nodules we considered the ones on which at least two radiologists agreed on malignancy and further went up to 0.990 when at least three radiologists agreed for the same feature. The prediction models for the nodule data on which at least three radiologists agreed are presented in Figures 4, 5, 6, and 7. The regression coefficients and their corresponding p-values show the contribution and significance of each feature to the regression model; the large values for the F-tests show a strong support for the found linear regression models. The image features that show up in the models are the most important features and they are selected if the p-values for the tests showing their contributions to the model are less than 0.05.

Table 3: Adj_R^2 of the stepwise multiple regression models of each characteristic; in all cases, the number of images and the number of nodules included in the dataset are shown in parentheses. The largest adj_R^2 values; a cell for which there is a dash indicates that radiologist agreement occurred only for a single rating.

Characteristics	Entire dataset (1106 images, 73 nodules)	At least 2 radiologists agreed	At least 3 radiologists agreed
Calcification	0.397	0.578 (884, 41)	0.645 (644, 21)
Internal Structure	0.417	- (855, 40)	- (659, 22)
Lobulation	0.282	0.559 (448, 24)	0.877 (137, 6)
Malignancy	0.310	0.641 (489, 23)	0.990 (107, 5)
Margin	0.403	0.376 (519, 28)	- (245, 7)
Sphericity	0.239	0.481 (575, 27)	0.682 (207, 9)
Spiculation	0.320	0.563 (621, 29)	0.840 (228, 9)
Subtlety	0.301	0.282 (659, 25)	0.491 (360, 10)
Texture	0.181	0.473 (736, 33)	0.843 (437, 15)

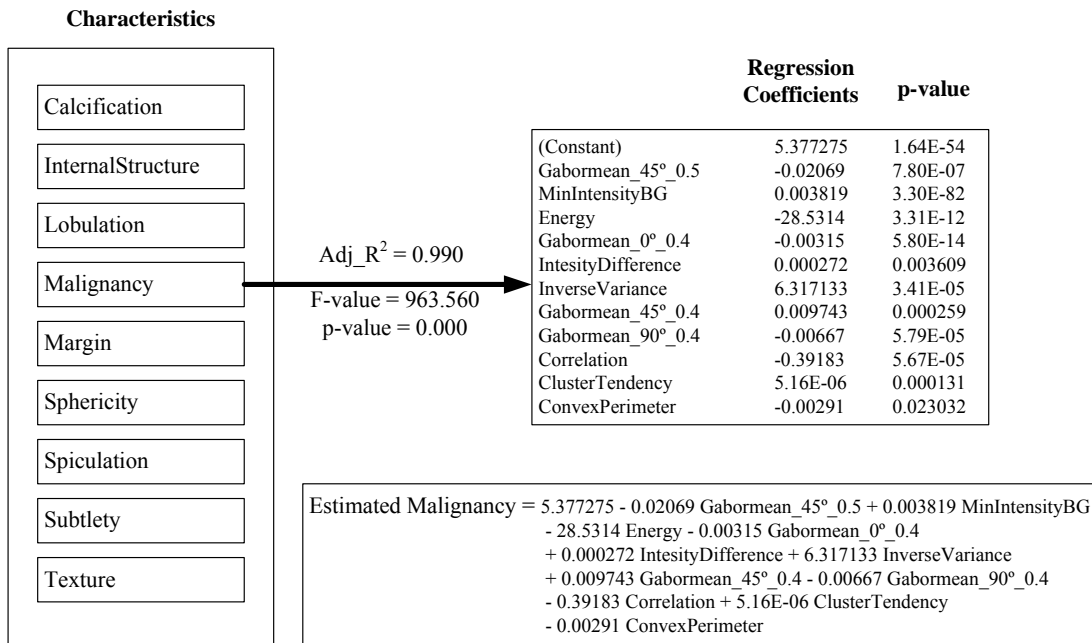


Figure 4: A prediction model for malignancy.

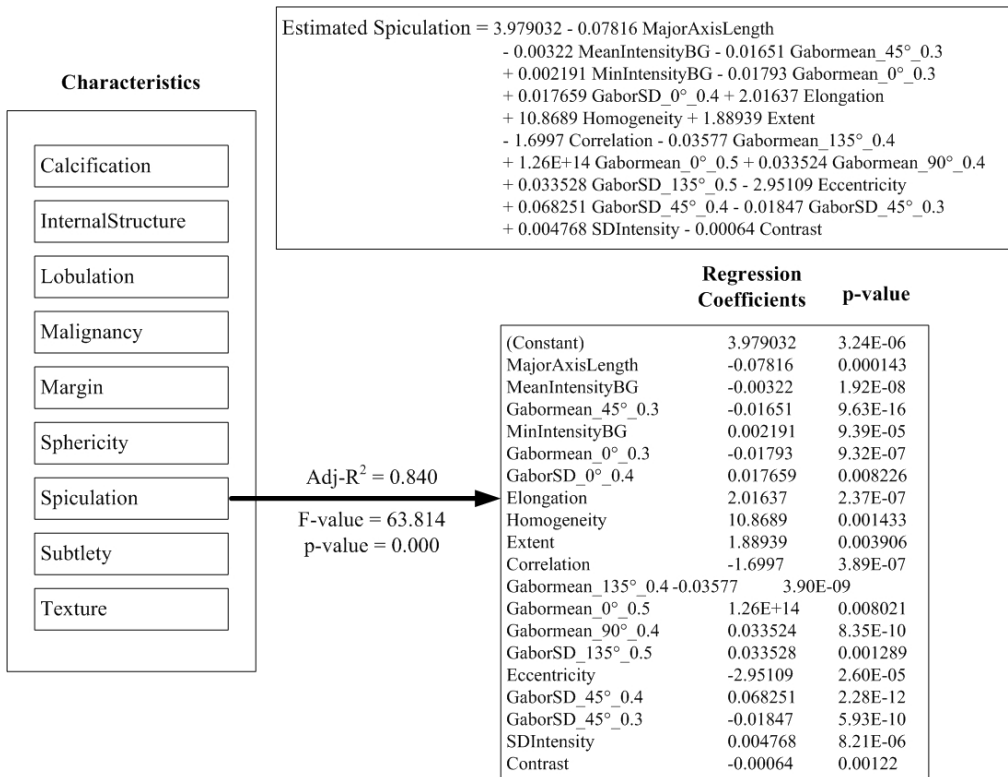


Figure 5: A prediction model for spiculation

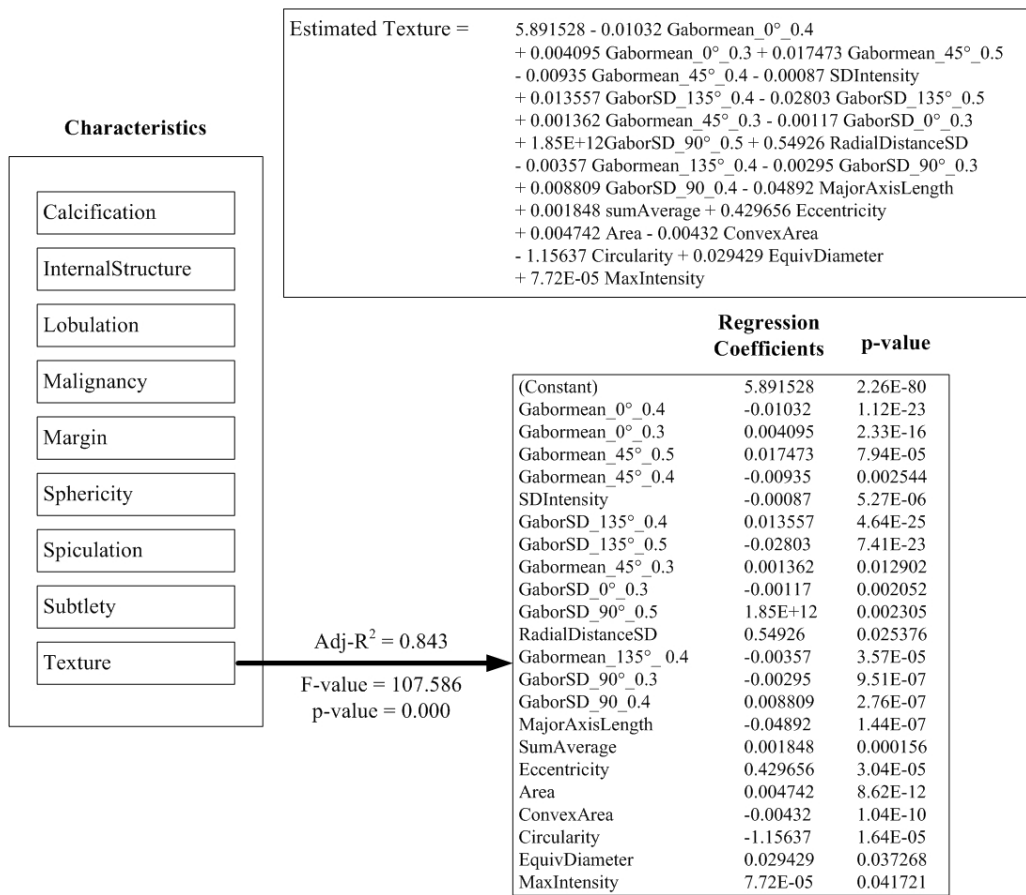


Figure 6: A prediction model for texture

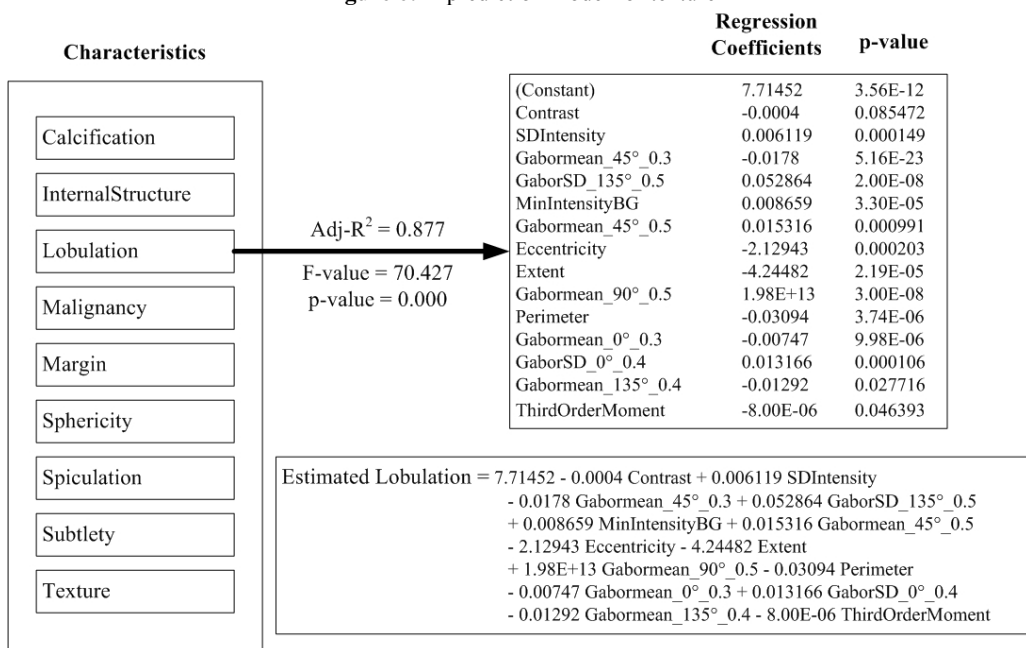


Figure 7: A prediction model for lobulation

5. CONCLUSIONS

In the past, researchers have developed several CAD systems for the detection and classification of pulmonary nodules. Most of these systems mimic domain knowledge in order to extract image content and they use a comparison with ground truth for diagnosis. They do so, however, in an algorithmic fashion that is only tenuously related to human perception and characterization of image features. Through the use of stepwise multiple regression, we proposed a quantitative model for finding the mappings between these two types of information. We found that the radiologists' perception with respect to lobulation, spiculation, texture and malignancy is captured with high accuracy based on the low-level image features used in this study. Our preliminary results are promising and can be considered the foundation of building computerized systems for detection, diagnosis, and medical image retrieval using radiologist-defined semantics. Although the approach is presented for the lung nodules, it can be easily extended to other modalities, anatomical structures and image features. Furthermore, we also provide an approach for finding the correlations among different semantic concepts used to describe the same visual pattern; in the long term, this approach can help create a domain-specific ontology of image feature descriptors.

REFERENCES

1. J. Burns, L. B. Haramati, K. Whitney, M. N. Zelefsky, "Consistency of Reporting Basic Characteristics of Lung Nodules and Masses on Computed Tomography," *Academic Radiology*, vol. 11, 23-237, 2004.
2. M. F. McNitt-Gray, E.M. Hart, N.Wyckoff, J.W. Sayre, J. G. Goldin, and D. R. Aberle, "A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results," *Medical Physics*, vol. 26, no. 6, pp. 880-888, 1999.
3. M. F. McNitt-Gray, N.Wyckoff, J.W. Sayre, J. G. Goldin, and D. R. Aberle, "The effects of co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography," *Computerized Medical Imaging and Graphics*, vol. 23, no. 6, pp. 339-348, 1999.
4. S.-C. B. Lo, M. T. F. Li-Yueh Hsu and, Y.-M. F. Lure, and H. Zhao, "Classification of lung nodules in diagnostic CT: An approach based on 3-D vascular features, nodule density distributions, and shape features," in *Proceedings of the SPIE*, vol. 5032, 2003, pp. 183-189.
5. Armato SG III, Altman MB, Wilkie J, Sone S, Li F, Doi K, Roy AS: Automated lung nodule classification following automated nodule detection on CT: A serial approach. *Medical Physics* 30: 1188-1197, 2003.
6. S. Takashima, S. Sone, F. Li, Y. Maruyama, M. Hasegawa, and M. Kadoya, "Indeterminate solitary pulmonary nodules revealed at population-based CT screening of the lung: using first follow-up diagnostic CT to differentiate benign and malignant lesions," *American Journal of Roentgenology*, vol. 180, no. 5, pp. 1255-1263, 2003.
7. S. Takashima, S. Sone, F. Li, Y. Maruyama, M. Hasegawa, T. Matsushita, F. Takayama, and M. Kadoya, "Small solitary pulmonary nodules (≤ 1 cm) detected at population-based CT screening for lung cancer: reliable high-resolution CT features of benign lesions," *American Journal of Roentgenology*, vol. 180, no. 4, pp. 955-964, 2003.
8. J. Gurney, "Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part I. Theory," *Radiology*, vol. 186, no. 2, pp. 405-413, 1993.
9. J. Gurney, D. Lyddon, and J. McKay, "Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. Application," *Radiology*, vol. 186, no. 2, pp. 415-422, 1993.
10. Y. Matsuki, K. Nakamura, H. Watanabe, T. Aoki, H. Nakata, S. Katsuragawa, and K. Doi, "Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution CT: Evaluation with receiver operating characteristic analysis," *American Journal of Roentgenology*, vol. 178, no. 3, pp. 657-663, 2002.
11. M. Aoyama, Q. Li, S. Katsuragawa, F. Li, S. Sone, and K. Doi, "Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images," *Medical Physics*, vol. 30, no. 3, pp. 387-394, 2003.
12. I. Sluimer, A. Schilham, M. Prokop, and B. Ginneken, "Computer Analysis of Computed Tomography Scans of the Lung: A Survey", *IEEE Transactions on Medical Imaging*, vol. 25, No. 4, 2006.
13. S. G. Armato, G. McLennan, M. F. McNitt-Gray, C. R. Meyer, D. Yankelevitz, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, A. P. Reeves, B. Y. Croft, and L. P. Clarke, "Lung Image Database Consortium: Developing a resource for the medical imaging research community," *Radiology*, vol. 232, no. 3, pp. 739-748, 2004.

14. S. Liu, and J. Li, "Automatic Medical Image Segmentation Using Gradient and Intensity Combined Level Set Method," The 28th IEEE EMBS Annual International Conference, 3118-3121, New York City, 2006.
15. G. D. Rubin, John K. Lyo, D.S. Paik, A. J. Sherbondy et al. "Pulmonary Nodules on Multi-Detector Row CT Scans: Performance Comparison of Radiologists and Computer-aided Detection", *Radiology* 2005;234:274-283
16. F. Li, Q. Li, R. Engelmann, M. Aoyama, S. Sone, H. MacMahon, K. Doi, "Improving Radiologists' Recommendations With Computer-Aided Diagnosis for Management of Small Nodules Detected by CT". *Academic Radiology*, Volume 13, Issue 8, Pages 943-950.
17. R. Wiemker, P Rogalla, T Blaffert, D Sifri, O Hay, et al. "Aspects of computer-aided detection (CAD) and volumetry of pulmonary nodules using multislice CT" *British Journal of Radiology* (2005) 78, S46-S56.
18. J. Burns, L. B. Haramati, K. Whitney, and M. N. Zelefsky, "Consistency of reporting basic characteristics of lung nodules and masses on computed tomography," *Academic Radiology*, vol. 11, pp. 233-237, 2004.
19. G. Leroy and H. Chen, "Meeting medical terminology needs-the ontology enhanced medical concept mapper," *IEEE Transaction on Information Technology in Biomedicine*", vol. 5, no 4, 2004.
20. D. A. B. Lindberg and B. L. Humphreys, *UMLS Knowledge Sources*, 14thed. Bethesda, MD: Na. Library Med., 2003. 2003AB. vol. 5, no. 4, pp. 261-270, Dec. 2001.
21. G. A. Miller, "WordNet: A lexical database for English," *Communications ACM*, vol. 38, no. 11, pp. 39-41, 1995.
22. A.S. Barb, C-R Shyu, and Y. P. Sethi, "Knowledge Representation and Sharing Using Visual Semantic Modeling for Diagnostic Medical Image Databases, *IEEE Transactions Information Technology in Biomedicine*, vol. 9, no.4, 2005
23. L. Ketai, M. Malby, K. Jordan, A. Meholic, J. Locken, "Small Nodules Detected on Chest Radiography: Does Size Predict Calcification?" *CHEST*, vol. 118, 610-614, 2000.
24. N. Hollings, P. Shaw, "Diagnostic Imaging of Lung Cancer," *European Respiratory Journal*, vol. 19, 722-742, 2002.
25. E. A. Zerhouni, F. P. Stitik, S. S. Siegelman, D. P. Naidich, S. S. Sagel et al. "CT of the Pulmonary Nodule: A Cooperative Study," *Radiology*, vol. 160, 319-327, 1986.
26. S. S. Siegelman, N. F. Khouri, F. P. Leo, E. K. Fishman, R. M. Braverman, E. A. Zerhouni, "Solitary Pulmonary Nodules: CT Assessment," *Radiology*, vol. 160, 307-312, 1986.
27. S. Matsuoka, Y. Kurihara, K. Yagihashi, H. Niimi, Y. Nakajima, "Peripheral Solitary Pulmonary Nodule: CT Findings in Patients with Pulmonary Emphysema," *Radiology*, vol. 235, 266-273, 2005.
28. B. Zhao, G. Gamsu, M. S. Ginsberg, L. Jiang, L. H. Schwartz, "Automatic Detection of Small Lung Nodules on CT Utilizing a Local Density Maximum Algorithm," *Journal of Applied Clinical Medical Physics*, vol. 4, Issue 3, 248-260, 2003.
29. J. M. Goo, J. W. Lee, H. J. Lee, S. Kim, J. H. Kim, J. Im, "Automated Lung Nodule Detection at Low-Dose CT: Preliminary Experience," *Korean J Radiology*, vol. 4, 211-216, 2003.
30. J.D. Furst, R. Susomboon, and D.S. Raicu, "Single Organ Segmentation Filters for Multiple Organ Segmentation", *IEEE 2006 International Conference of the Engineering in Medicine and Biology Society (EMBS'06)*, August 2006
31. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. On Systems, Man, and Cybernetics*, vol. 3, no. 6, 610-621, 1973.
32. T. Andrzykiak and M. Choras, "Image retrieval based on hierarchical Gabor filters," *International Journal Applied Computer Science*, vol. 15, no. 4, 471-480, 2005.
33. R. L. Ott and M.T. Longnecker, "An Introduction to Statistical Methods and Data Analysis", Duxbury Press (2000), 5th edition.
34. C. Spearman, "The proof and measurement of association between two things," *American Journal of Psychology*, vol. 15, 72-101, 1904.