Modeling Semantics from Image Data: Opportunities from LIDC

Daniela S. Raicu[a], Ekarin Varutbangkul[a], Jacob D. Furst[a], Samuel G. Armato III[c]
[a]Intelligent Multimedia Processing Laboratory, School of Computer Science, Telecommunications, and Information Systems, DePaul University, Chicago, IL 60604, USA;
draicu@cs.depaul.edu, evarutbangkul@students.depaul.edu, jfurst@cs.depaul.edu
[c]Department of Radiology, The University of Chicago, Chicago, IL 60637, USA
s-armato@uchicago.edu

Abstract

While the advances in computed tomography (CT) technology allow better detection of pulmonary nodules by generating higher-resolution images, the new technology also generates many more individual transverse reconstructions. As a result, the efficiency and accuracy of the radiologists interpreting these images is reduced. Double reading by two trained human observers has been shown to improve the detection of lung cancer on chest radiographs by a 3% - 30% increase in sensitivity. Given the increased cost of interpretation of double reading and the variation among radiologists' interpretation, t*he objective of this paper is to develop computer-aided tools that could be used as "second readers"* when interpreting lung images by apriori rating the nodules based on automatically discovered image-semantics mappings. To attain the objective of this paper, we will test the working hypothesis that there is enough information in the low-level image features that can capture certain semantic meanings associated with the visual appearance of the nodules. The working hypothesis will be tested using data mining techniques. The rationale is that successful completion of the proposed research will reduce the semantic gap in the medical image indexing and retrieval community, in particular, for lung image interpretation. The acquisition of the mappings between the two types of features is also critical to the development of *visual ontology for lung interpretation* that can be used to automatically annotate new images (based on low-level image features) and provide *context-sensitive tools for pulmonary nodule retrieval.*

Keywords: Computed Tomography, lung nodules, low-level features, semantic gap, logistic regression, decision trees, support vector machine, visual ontology

1. Introduction

The explosion of the medical imaging technologies has generated mountains of data; depending on the size of the institution, a radiology department can perform between 100 and 5,000 examinations daily, generating a myriad of images, patient data, report text, findings, and recommendations [10]. New digital image management systems, Picture Archiving and Communication Systems (PACS), have been developed for image acquisition, storage, transmission, processing and display of images for their analysis and further diagnosis. Availability of digital data within the PACS raises a possibility of health care and research enhancements associated with manipulation, processing and handling of data by computers, that is a basis for *computer-assisted radiology* development.

In general, radiology data is well organized but poorly structured, and structuring this data prior to knowledge extraction is an essential first step in the successful mining of radiological data [10]. Compared to text, radiology images are enormous in size, highly variable over time, and quite difficult to mine. Therefore, image processing and data mining techniques are necessary for structuring, management, retrieval, and interpretation of image data.

In this paper, we present *a framework for modeling the associations between image content and radiologists' subjective assessments for lung nodule interpretation.* It is known that double reading by two trained human observers has been shown to improve the detection of lung cancer on chest radiographs by a 3% - 30% (mean 13%) increase in sensitivity [43]. Given the increased cost of interpretation of double reading and the variation among radiologists' interpretation, we expect that the proposed framework will be used as a "*second reader*" when interpreting lung images by *apriori* rating the nodules based on discovered associations.

The paper is organized as follows. We present a literature review relevant to our work in Section 2, the National Institutes of Health (NIH) Lung Image Database Consortium (LIDC) dataset used in this study in Section 3, the proposed framework (as outlined in Figure 1) in Section 4, and our preliminary models and visual ontologies for lung nodule interpretation in Section 5. We conclude the paper by summarizing our findings and presenting future avenues for modeling image semantics in the medical field.

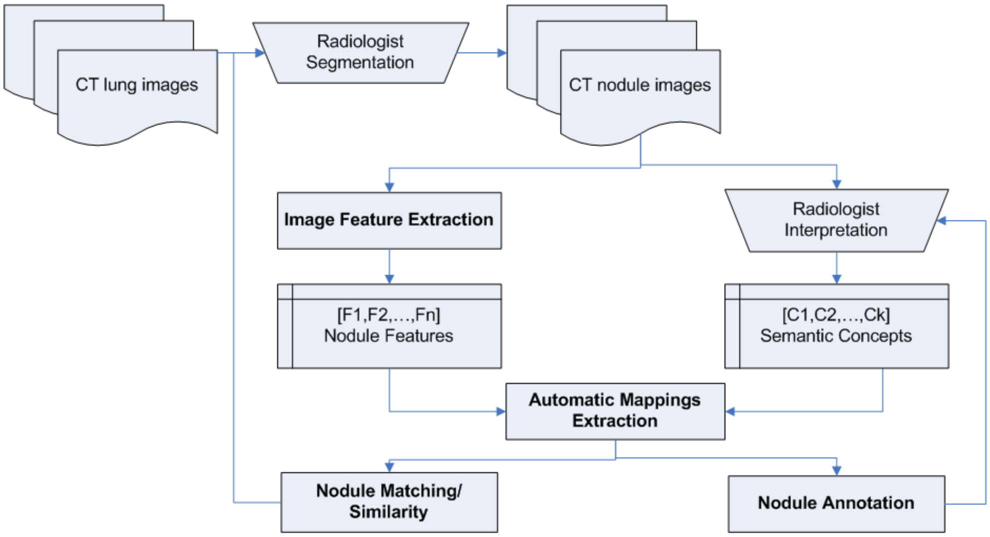

CT lung images
Radiologist Segmentation
CT nodule images
Image Feature Extraction
Radiologist Interpretation
[F1,F2,…,Fn] Nodule Features
[C1,C2,…,Ck] Semantic Concepts
Automatic Mappings Extraction
Nodule Matching/ Similarity
Nodule Annotation

Figure 1: Diagram of the proposed mapping framework

2. Related Work

Diagnostic decision-making in medical imaging by radiologists has been augmented by computer-aided diagnosis (CAD) systems. Typically, a CAD system marks the location of a suspicious region and then makes a diagnosis based on the low-level image features calculated for the region of interest. Increasing the sensitivity (the ratio between true positives and all positives) and reducing the number of false positives motivates current research ranging from image segmentation and feature extraction to data mining and knowledge discovery. While beneficial as a tireless and increasingly accurate screening tool, CAD systems rarely offer supporting guidance about their diagnostic decision rationale or, when they do, this guidance does not match the perceptual tasks used by the radiologist in forming their diagnosis [9]. Recently, several approaches have been proposed to study the relationships between the human interpretation of the regions of interest and the computer-calculated image features with the final goal of integrating these relationships in the CAD systems.

In the following subsections, we will provide a literature review for 1) existent lung CAD systems, 2) lung image datasets, 3) ontologies for medical interpretation, and 4) state-of-the-art research on modeling the relationships between image content and semantics.

2.1. Low-level Image Features and CAD for Lung

Several computer-aided diagnosis (CAD) systems have been developed to help estimate the probability of lung cancer based on nodule characteristics (such as nodule size, shape, texture and internal structure). McNitt-Gray et al. [33, 34] used nodule size, shape and co-occurrence texture features as nodule characteristics to design a linear discriminant analysis (LDA) classification system for malignant versus benign nodules. Lo et al. [28] used direction of vascularity, shape and internal structure to build an artificial neural network (ANN) classification system for prediction of the malignancy of the nodules. Armato et al [3] used nodule appearance and shape to build an LDA classification system to classify pulmonary nodules into malignant versus benign classes. Takashima et al. [53, 54] used shape information to characterize malignant versus benign lesions in the lung.

Beside these systems that are based on just nodule characteristics, there are several studies that also incorporated some clinical information (such as age, gender, history of smoking, and history of cancer). Gurney [13, 14] designed a Bayesian classification system based on clinical information in addition to radiological information. Matsuki et al. [31] also used clinical information in addition to sixteen features scored by radiologists to design an ANN for malignant versus benign classification. Aoyama et al. [2] used two clinical features in addition to forty-one image features to estimate the likelihood of malignancy for pulmonary nodules on low-dose CT images. Li et al. [24] used two clinical features (age and gender) in addition to fifty-six image features in their CAD system which assists radiologists in improving biopsy recommendations for small lung nodules. For more literature review on the analysis of computed tomography scans of the lung, we recommend the survey by Sluimer et al. [49].

2.2. Lung Image Benchmarks

In all these CAD studies, most authors created their own datasets with their own ground truth for evaluation. The use of different datasets makes the comparison of these CAD systems not feasible and therefore, there is an immediate need for reference datasets that can provide a common ground truth for the evaluation and validation of these systems.

The NIH Lung Image Database Consortium (LIDC) has created a dataset [4] to serve as an international research resource for development, training, and evaluation of CAD algorithms for detecting lung nodules on CT scans. Liu and Li [27] have already used the LIDC dataset to propose a new method for nodule detection based on gradient and intensity combined level set methods that generate stable and accurate segmentation results for complex organic structures like lung bronchia and nodules. While most nodule segmentation algorithms were designed for solid pulmonary nodule, Tachibana and Kido [52] developed an automated volumetric segmentation algorithm of small pulmonary nodules with ground-glass opacity and used lung images from LIDC to evaluate their algorithm. Opfer and Wiemker [41] also used lung images from LIDC to validate their general tumor segmentation approach which combines energy minimization methods with radial basis function surface modeling techniques.

In this paper, we propose to use the LIDC dataset for modeling the radiologists' nodule interpretations based on image content with the final goal of reducing the variability among radiologists and improving their interpretation efficiency. To our best knowledge, this is the first use of the LIDC dataset for the purpose of modeling lung nodule semantics.

2.3. Ontologies for Medical Image Interpretation

Besides creating medical image datasets to be used as benchmarks for the evaluation of the CAD systems, first steps have also been taken in creating frameworks for a common language when making diagnoses.

There are several ontologies available in the domain of biomedicine, which capture and represent the concepts and their relationships in that domain. The most notable, multi-purpose, and widely used is the National Library of Medicine's "Unified Medical Language System" [26], which integrates a large number of distinct source terminologies. The concepts are categorized in several semantic categories such as: organisms, anatomical structures, biological functions, chemicals, events, physical objects, etc. Rosse et al. [46] and Brinkley et al. [7] have argued that UMLS lacks the required granularity, adequate semantic types and relationships to comprehensively represent anatomical concepts, and hence they created the foundational model of anatomy as an enhancement to UMLS. Along the same lines, Leroy and Chen [23] developed a tool (Medical Concept Mapper) based on the UMLS and WordNet [36] that connects patient information to human-created ontologies. Hu et al. [19] built a breast imaging ontology based on the breast imaging reporting and data system (BI-RADS). Furthermore, Kahn et al. [20] constructed an ontology to represent radiological procedural knowledge for picture archiving and communication systems (PACS) integration. While all these ontologies are constructed based on textual concepts, there is some initial work that makes use of the image content as described in the following subsection.

2.4. Mappings between Image Features and Semantic Interpretations

Barb et al. [5] proposed a framework that uses semantic methods to describe visual abnormalities and exchange knowledge in the medical domain. Ebadollahi et al. [11] proposed a system to link the visual elements of the content of an echocardiogram (including the spatial-temporal structure) to external information such as text snippets extracted from diagnosis reports. Ogiela and Tadeusiewicz [40] has published some initial results to automatically generate linguistic descriptors of lesions in the coronary artery and in the urinary track, based on vessels' patters detected with image processing. More information about ontology-based object learning and recognition can be found in the work proposed by Maillot [30] and Mezaris et al. [35]. Our previous work [45, 56] can be considered one of the initial steps in the direction of mapping lung nodule image features to perceptual categories encoding the radiologists' knowledge for lung interpretation.

These studies provide strong support for the working hypothesis that *low-level image features can be automatically linked to the medical concepts used for diagnosis*. Our recent findings that we will present in Section 5 based on the proposed mapping framework presented in Section 4 also support the working hypothesis and increase the probability that it will prove to be valid.

3. Lung Image Database Consortium (LIDC) Dataset

The LIDC [4] has developed a data collection process to identify, in thoracic CT scans, lesions that are considered by radiologists to belong to one of three categories: 1) nodules greater than or equal to 3 mm in maximum diameter but less than 30 mm (regardless of presumed histology), 2) nodules less than 3 mm (but only if not clearly benign), and 3) non-nodules greater than or equal to 3 mm. The images and associated XML files that contain the radiologists' annotations have been made publicly available through a web-based archive. To date, this archive contains 85 CT scans with associated XML files. In the proposed research, the images that contain lesions marked as nodules > 3 mm by LIDC radiologists, along with the nodule outlines and nodule characteristics provided by the LIDC radiologists will be used. The nine nodule characteristics and their possible ratings are shown in Table 1; the table also reports our interpretation for these characteristics based on our CAD systems review for lung nodules.

Table 1: LIDC nodule characteristics with corresponding notes and possible ratings; a '.' represents a rating on the scale for which a definition label was not provided.

Characteristic	**Notes and References**	**Possible Ratings**
Calcification	Calcification appearance in the nodule - The smaller the nodule, the more likely it must contain calcium in order to be visualized [21]. Benignity is highly associated with central, non-central, laminated, and popcorn calcification [18, 58].	1. Popcorn; 2. Laminated 3. Solid 4. Non-central 5. Central 6. Absent
Internal structure	Expected internal composition of the nodule	1. Soft Tissue 2. Fluid 3. Fat 4. Air
Lobulation	Whether a lobular shape is apparent from the margin or not - lobulated margin is an indication of benignity [49].	1. Marked 2. . 3. . 4. . 5. None
Malignancy	Likelihood of malignancy of the nodule - Malignancy is associated with large nodule size while small nodules are more likely to be benign [18, 48]. Most malignant nodules are non-calcified [34] and have spiculated margins [32].	1. Highly Unlikely 2. Moderately Unlikely 3. Indeterminate 4. Moderately Suspicious 5. Highly Suspicious
Margin	How well defined the margins of the nodule are	1. Poorly Defined 2. . 3. . 4. . 5. Sharp
Sphericity	Dimensional shape of nodule in terms of its roundness	1. Linear 2. . 3. Ovoid 4. . 5. Round
Spiculation	Degree to which the nodule exhibits spicules, spike-like structures, along its border - Spiculated margin is an indication of malignancy [58, 59].	1. Marked 2. . 3. . 4. . 5. None
Subtlety	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
Texture	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 [48].	1. Non-Solid 2. . 3. Part Solid/(Mixed) 4. . 5. Solid

It is important to notice that the LIDC did not impose a forced consensus; rather, all of the lesions indicated by the radiologists were recorded and are available to users of the database. Accordingly, each lesion in the database considered to be a nodule > 3 mm could have been marked as such by only a single radiologist, by two radiologists, by three radiologists, or by all four LIDC radiologists. Therefore, there can be up to 4 different boundaries/images of a nodule marked by up to 4 radiologists on a slice as presented in Figure 2. If the nodule appears on X slices, there can be up to 4*X images for the nodule in the dataset.

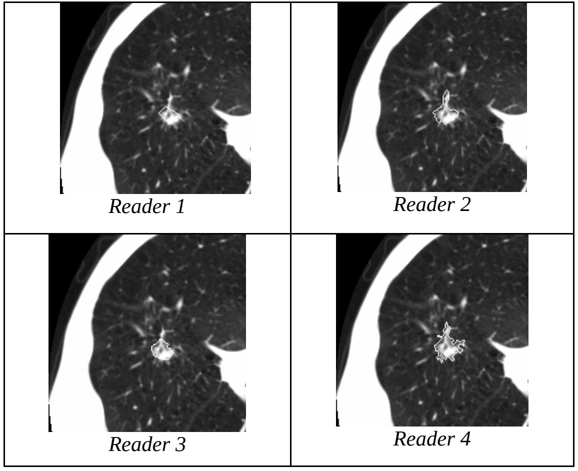


Figure 2: An example of four different boundaries of a nodule on a slice marked by 4 different radiologists

From the current 85 cases available, 60 cases had 149 nodules greater than or equal to 3 mm in maximum diameter which generated 1989 nodule images. From all nine semantic characteristics, we focused on the relationships between the image content and the radiologists' subjective assessments with respect to seven semantic concepts: *subtlety, lobulation, margin, sphericity, malignancy, texture, and spiculation.* Calcification and internal structure were not considered since most of the ratings for them were dominated by only one rating ('no calcification' appears in the nodule, and the internal composition of the nodule is 'soft tissue').

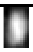
4. Proposed Data Mining Framework for Image Semantics Modeling

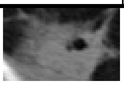
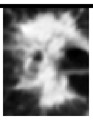
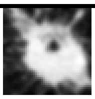
First, we quantify the lung nodule images using a set of low-level image features that satisfy the main requirements for feature extraction [29]: a) completeness/expressiveness (features should be a rich enough representation of the image contents to reproduce the essential information); b) compactness (the storage of the features should be compact to allow efficient access) and c) tractability (the distance between features should be efficient to compute). Second, we model the uncertainty in the radiologists' interpretations (characteristics/annotations) through probabilistic data mining approaches. In other words, instead of predicting just a single rating per characteristic (for example, rating '1' - 'extremely subtle' for "subtlety"), we propose to generate probabilities for all possible ratings (in this example, probabilities for all five ratings of the subtlety). Table 2 shows an example of probabilistic predictions obtained using decision trees. Finally, in the third stage, a visual-based ontology is constructed using the probabilistic mappings learned in the previous stage.

Table 2: Predicted ratings (first column), probabilistic rules (second column), and annotated nodules (third column)

Ratings	**Probabilistic Low-level image features-based rules**	Examples of nodules rated by the rule on the left

1
(Extremely Subtle)
	IF (MinorAxisLength <= 0.14683) AND (MaxIntensity <= 0.226443) THEN
	subtlety = 1 with *Pr (1) = 1.00*
		Nodule# 65

4
(Moderately Obvious)
	IF (MinorAxisLength <= 0.14683) AND (MaxIntensity > 0.226443) THEN
	subtlety = 4 with *Pr (4) = 0.94 & Pr (5) = 0.06*
		Nodule# 42
		 Nodule# 47
		 *Nodule# 105*

5 (Obvious)	IF (MinorAxisLength > 0.14683) THEN subtlety = 5 with *Pr (5) = 0.99 & Pr (4) = 0.01*	 *Nodule# 24*  *Nodule# 46*  *Nodule# 68*

4.1. Low-level Image Feature Extraction

We propose to extract four types of features (Table 3) that encode the a) shape, b) size, c) intensity and d) texture of the region of interest (nodule) while satisfying the main requirements for feature extraction mentioned above. 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 [12, 59].

Table 3: The entire set of image features extracted from each lung nodule's boundary.

Shape Features	Size Features	Intensity Features	Texture Features
Circularity Roughness Elongation Compactness Eccentricity Solidity Extent RadialDistanceSD	Area ConvexArea Perimeter ConvexPerimeter EquivDiameter MajorAxisLength MinorAxisLength	MinIntensity MaxIntensity MeanIntensity SDIntensity MinIntensityBG MaxIntensityBG MeanIntensityBG SDIntensityBG IntensityDifference	11 Haralick features calculated from co-occurrence matrices (Contrast, Correlation, Entropy, Energy, Homogeneity, 3rd Order Moment, Inverse variance, Sum Average, Variance, Cluster Tendency, Maximum Probability)
			24 Gabor features which are mean and standard deviation of 12 different Gabor images (orientation = 0°, 45°, 90°, 135° and frequency = 0.3, 0.4, 0.5)
			5 MRF features which are mean of 4 different response images (orientation = 0°, 45°, 90°, 135°), along with the variance response image

<u>*Shape Features*</u>

We use 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 circumference of the equivalent area circle by the actual perimeter of the nodule. *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.

<u>*Size Features*</u>

We use the following seven features to quantify the size of the nodules: 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.

<u>*Intensity Features*</u>

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.

<u>*Texture Features*</u>

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 hierarchal 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 work [22], in this research we focus on three well-known texture analysis techniques: *co-occurrence matrices (a statistical-based method), Gabor filters (a transform-based method), and Markov Random Fields (a model 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 (0º, 45º, 90º, and 135º) 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 descriptors [16] 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 filtering [1] 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 (0º, 45º, 90º, and 135º) 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 [1], 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.

Markov Random Fields (MRFs) is a model based method which captures the local contextual information of an image [6]. We use the algorithm devised by Cesmeli [8] to calculate five features correspond to 4 orientations (0°, 45°, 90°, 135°) along with the variance. We calculated feature vectors for each pixel by using 9×9 estimation window. The mean of 4 different response images and the variance response image are used as our 5 MRF features.

At the end of the image feature extraction process, each nodule image is encoded using a set of sixty-four image features $f_i, i = 1,\dots 64$ and nine radiologist annotations $c_j, j = 1,\dots 9$ (semantic concepts). Therefore, the nodule representation is given by the vector representation $\left[f_1, f_2,\dots f_{64}, c_1, c_2,\dots c_9\right]$; Figure 3 shows an example of feature values for a nodule representation.

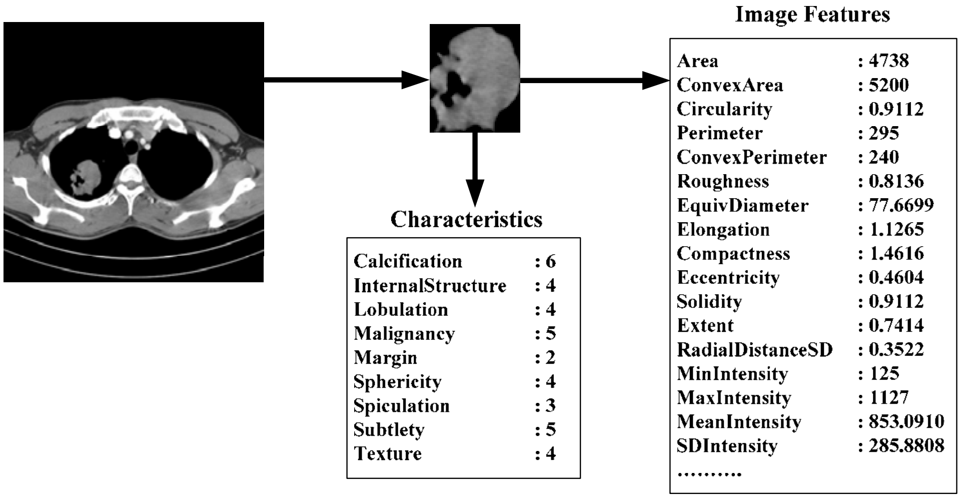


Image Features

Area	: 4738
ConvexArea	: 5200
Circularity	: 0.9112
Perimeter	: 295
ConvexPerimeter	: 240
Roughness	: 0.8136
EquivDiameter	: 77.6699
Elongation	: 1.1265
Compactness	: 1.4616
Eccentricity	: 0.4604
Solidity	: 0.9112
Extent	: 0.7414
RadialDistanceSD	: 0.3522
MinIntensity	: 125
MaxIntensity	: 1127
MeanIntensity	: 853.0910
SDIntensity	: 285.8808
……….	

Characteristics

Calcification	: 6
InternalStructure	: 4
Lobulation	: 4
Malignancy	: 5
Margin	: 2
Sphericity	: 4
Spiculation	: 3
Subtlety	: 5
Texture	: 4

Figure 3: An example of nodule characteristics assigned by a radiologist and low-level features

4.2. Mappings between Image Features and Semantic Interpretations

Three approaches whose output can be expressed in terms of probabilities are investigated with respect to their power to model the radiologists' subjective annotations through a set of objective image features: logistic regression, decision trees, and support vector machine (SVM). All three approaches are appropriate for ordinal response variable (as it is the case for the radiologists' annotations) and support several types of explanatory variables (binary, categorical, and continuous - low-level image features). Each one of the approaches has some advantages over the others. Logistic regression takes into account the order among the values for the categorical response variable, while decision trees and SVM separate well non-linear data. Feature selection is automatically performed in decision tree learning, while it is optional in logistic regression and SVM. Each one of the three approaches has the property of modeling probabilistically the ratings for each of the seven characteristics as we describe below.

Logistic Regression

Logistic regression [51] is a statistical data analysis technique that can be used to predict a categorical response variable based on a set of explanatory variables. The explanatory variables can be binary, categorical, continuous, or any combination of these types. The response variable can be an ordinal or nominal variable. When the response variable is nominal (such as calcification) the generalized logits model can be applied. For an ordinal response variable (such as malignancy, lobulation, spiculation, etc) the cumulative logits model will be more appropriate. Furthermore, when predicting the response variable, the probabilities of all possible values are included in the outcome of the logistic model. Given that all the semantic concepts we focus on in this paper are ordinal variables, we explain below the cumulative logits model.

In cumulative logits model, the cumulative probabilities $\theta_l = \sum_{j=1}^{l} \pi_j$, l = 1,…(r-1); and its logits are calculated, where r is the number of ratings (all characteristics except calcification and internal structure have r =5) and π_j denotes the probability of rating j.

$$logit(\theta_l) = \ln \left[\frac{\theta_l}{1-\theta_l} \right] = \alpha_l + \sum_i \beta_i f_i \qquad (1)$$

where regression coefficient β_i is the same for all logits, but the intercepts α_l for different logits are not the same. From each of $logit(\theta_l)$ we can compute each cumulative probability θ_l

$$\theta_l = \frac{exp(log\,it(\theta_l))}{1 + exp(log\,it(\theta_l))} \qquad (2)$$

and from all θ_l we can compute response probabilities for all categories, since $\sum_{j=1}^{r} \pi_j = 1$. Then the highest response probability can be used for deciding the predicted category. Furthermore, the goodness-of-the-fit of the regression model is measured through the coefficient of determination Nagelkerke's R^2 ; the higher the value of R^2 is, the better the regression model will fit the data.

Decision Trees

Decision tree learning [37] is a data mining technique that can be used to map the low-level representation of the data to the high-level representation of the data encoded through class or category labels. The low-level features are sorted based on some criterion that quantifies the discrimination power of the features with respect to the given classes. The tree will be formed by placing the features with the highest discrimination power at the top and the features with lower discrimination power towards the bottom of the tree. Each internal node in the tree is a test of an attribute and branches from the node correspond to the possible values of the attribute. Therefore, leaf nodes represent classifications and branches represent conjunctions of attributes that lead to those classifications. The leaf nodes can also produce probabilistic classifications by dividing the number of cases for a certain class under the leaf node by the total number of cases grouped under the corresponding node. The complexity of the tree measures the tradeoff between a small tree that generates a reasonable number of leaf nodes (or rules) and low classification errors obtained on the training set. We use the minimum description length (MDL) principle [15] to encode the complexity of the tree as follows:

$$Cost(tree, data) = Cost(tree) + Cost(data | tree) \qquad (3)$$

$$Cost(tree) = (i * log_2 m) + (j * log_2 k) \qquad (4)$$

$$Cost(data | tree) = (e * log_2 n) \qquad (5)$$

$Cost(tree)$ is the cost of encoding all the nodes in the tree and $Cost(data|tree)$ is encoded using the classification errors the tree commits on the training set. Each internal node of *i* internal nodes is encoded by the ID of the splitting attribute. The cost of encoding each attribute is $log_2\ m$ bits, where *m* is the number of attributes. Each leaf node of *j* leaf nodes is encoded using the ID of the class it is associated with. The cost of encoding each class is $log_2\ k$ bits, where *k* is the number of classes. Each error of e errors is encoded by $log_2\ n$ bits, where *n* is the number of instances in the training set.

The decision tree algorithm used in this study is C4.5 pruned tree (J48 in WEKA [57]) with the minimum objects per each leaf node being equal to 2 (best accuracies from all experiments with 2, 3, 4, and 5 objects per node) and the feature selection criterion for growing the tree being the information gain [37]:

$$IG(S,A) = Entropy(S) - \sum_{v \in A} \frac{|S_v|}{|S|} Entropy(S_v) \qquad (6)$$

where *v* is a value of attribute *A*, $|S_v|$ is the subset of instances of *S* where *A* takes the value *v*, and $|S|$ is the number of instances, and

$$Entropy(S) = \sum_{i=1}^{C} p_i\ log_2\ p_i \qquad (7)$$

where p_i is the proportion of instances in the dataset that has the target attribute as *i* from *C* categories. We did not perform any apriori discretization of our image features as a preprocessing step since the ranges that might work for predicting one of the radiologists' characteristics might not work in predicting others and C4.5 can handle continuous data as well as discrete data by creating a threshold and splitting data whose attribute value is above the threshold from data whose attribute values is less than or equal to the threshold [44].

Support Vector Machine

Support vector machine (SVM) [17] is a supervised learning technique that performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. When data are not linearly-separable, SVM uses a kernel function to transform the data from a highly-dimensional input space into a new feature space in which the data can be linearly separable. The kernels that are usually used to transform the original feature space are: polynomial function, radial basis function (RBF), or sigmoid function.

In the SVM approach, a predictor variable is called an *attribute*, and a transformed attribute that is used to define the hyperplane is called a *feature*. A set of features that describes one instance is called a *vector*. The vectors at the margin are called *support vectors*. SVM seeks for an optimal separating hyperplane which have the maximum margin. When the target variable has more than two classes, SVM can handle the problem by two approaches: 1) "one against all" where we classify one class from all other classes; and 2) "pair-wise" where n*(n-1)/2 models are constructed where n is the number of classes.

In this study we used the sequential minimal optimization (SMO) algorithm [47] implemented in WEKA [57] which handle multi-class problems by using pair-wise classification. We used polynomial kernel with the exponent equals to 2 or "quadratic kernel" which provided the best results compared to the results when the exponent was set to be 1 or 3 or the results when the RBF kernel was used. Logistic models were fitted to SMO's output to express it as posterior probabilities [50] for each characteristic.

Image- Semantics Mappings Evaluation

The performance metric for evaluating the results is accuracy (hit ratio). When evaluating the three approaches, the variability among radiologists with respect to 1) the detection of a lesion (present/absent at a certain location), 2) boundary delineation (same/multiple contours for the lesion), and 3) lesion multi-level interpretation (for instance, two radiologists identify the lung nodule's texture as solid and the other two as part-solid) are also analyzed to see how they influence the classification performance. While it is expected the classification performance to improve as we build classifiers on images on which more agreement exists, the generalizability power (high accuracy on testing data) of the classifiers will decrease given the smaller number of images available for training and testing as we move from at least two radiologists' agreement to at least three agree. Therefore, we will use the leave-one out evaluation technique [15] designed to produce reliable results on small data sets.

Visual-based Ontology

In addition to the prediction models, both logistic regression and decision trees produce rankings of the image features with respect to their classification power for each annotation. For the logistic regression, the model coefficients whose $p\_values \le \alpha$ ($\alpha = 0.5$) will generate the most important image features; for the decision tree approach, the features ranked with the highest information gain [15] will generate the most important features. Unlike logistic regression and decision trees, the SVM model is hard to understand given the non-linear kernel applied to classify the data.

After the most important features (f_1, …,f_k) are identified along with their appropriate values/ranges (Range$_1$,…Range$_k$) for each semantic characteristic (in particular each rating), the visual-based ontology is constructed as indicated in the diagram from Figure 4.

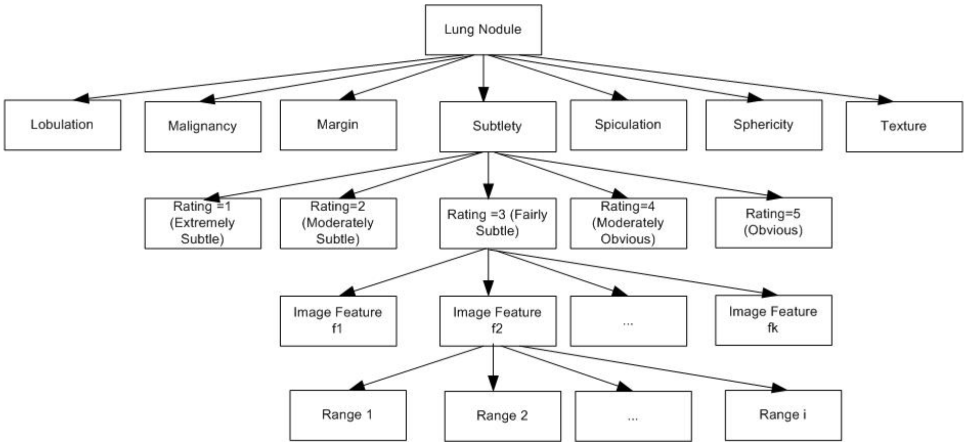


Figure 4: The diagram of the visual-based ontology for lung nodule interpretation; only the relationships for subtlety are shown, but all the other characteristics can be expanded similarly

The visual ontology can be used to 1) *show the computer interpretation of the corresponding radiologist rating in terms of a set of uniform and objective image features, 2) automatically annotate new images (based on low-level image features), and 3) provide context-sensitive tools for pulmonary nodule retrieval* (for instance, if a pulmonary nodule is rated as highly spiculated, then the system will use only the corresponding image features for that rating to retrieve the most similar nodules based on image content). When integrating the visual ontology in the radiology lexicon, RadLex, this can help increase the efficiency of individual readers (radiologists) and reduce the inter-reader variability in interpreting lung CT studies. While there are some research studies [38, 39, 25, 55, 60] that attempt to find the best combination between image features and computer-based similarity measures used to retrieve the most similar images with the final goal of resembling the human (radiologist) perception of image similarity, our proposed work attempts to find the direct mappings from image features to radiologists' annotations (spiculation, lobulation, texture, margin, etc) with the final goal of providing recommendations for automatic image annotation and markup.

5. Preliminary Results

The results presented in this section are obtained on the 149 unique nodules that generated 1989 nodule images. The data was exempt from human subjects' research regulation since the LIDC dataset was completely de-identified. When evaluating the results, besides considering the variation in the data with respect to radiologists' agreement, we also look at the results for the following two situations. First, the noise in the data was eliminated by removing the end caps (the first and the last slices of each nodule marked by each radiologist; in this study, we loose information for some nodules which have only one or two slices outlined by a radiologist, but we plan in a future study to remove only the end caps of nodules which have at least three slices outlined by a radiologist. Second, the bias in the data was eliminated by using only one slice (the one with the largest nodule area) as the representative image for each nodule marked by each radiologist. Given the large number of experiments performed to compare the three approaches for the seven characteristics, in Table 7 to Table 15 we also report the overall accuracy calculated as the average of the seven accuracies weighted by the number of nodule images for each characteristic. Furthermore, although we got best results when all 4 radiologists agreed, we do not consider these models for building the ontology since 1) the models are based on very small data sets (consisting of two or three nodules) and the agreement among radiologists was with respect to only one rating (lobulation = 1, margin = 5, subtlety = 5, and texture =5) or two ratings (malignancy = {3 or 5}, sphericity = {3 or 4}, and spiculation = {1 or 5}) per characteristic.

Logistic Regression Results

From Table 4 to Table 6, we found that the goodness-of-fit of the regression models (Nagelkerke's R^2) increase when there are more agreements among radiologists. For example, R^2 went up from 0.39 to 0.61 when we considered the nodules on which at least two radiologists agreed on malignancy instead of considering all nodules; furthermore, it went up to 0.990 when at least three radiologists agreed and 0.999 when all 4 radiologists agreed for the same feature (malignancy). Furthermore, Table 5 shows that the goodness-of-fit of the models increase after removing end caps and Table 6 shows even better fits for the data when only one single slice with the largest nodule area was selected.

Table 4: Nagelkerke's R^2 from **logistic regression**, multiple slices per nodule *before* removing end caps; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (1989 images, 149 nodules, 60 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	0.1504	0.3361 (943, 73, 42)	**0.8125** (331, 20, 19)	- (20, 2, 2)
Malignancy	0.3908	0.6182 (1090, 73, 42)	**0.9987** (295, 19, 15)	0.9999 (85, 3, 3)
Margin	0.3975	0.5172 (981, 77, 42)	**0.9602** (452, 17, 12)	- (128, 3, 3)
Sphericity	0.2577	0.5245 (1028, 77, 45)	**0.7929** (400, 27, 20)	0.9999 (28, 2, 2)
Spiculation	0.1384	0.3464 (1145, 77, 43)	**0.6829** (427, 27, 24)	0.9999 (40, 3, 3)
Subtlety	0.4338	0.5087 (1308, 70, 41)	**0.8724** (775, 25, 21)	- (452, 10, 9)
Texture	0.3528	0.5452 (1333, 79, 43)	**0.9592** (801, 34, 24)	- (482, 11, 11)

Table 5: Nagelkerke's R^2 from **logistic regression**, multiple slices per nodule *after* removing end caps; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (1259 images, 112 nodules, 52 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	0.2359	0.3742 (600, 55, 36)	**0.9659** (214, 14, 14)	- (4, 1, 1)
Malignancy	0.3947	0.6489 (723, 59, 37)	**0.9996** (175, 13, 11)	0.9999 (61, 3, 3)
Margin	0.5129	0.6815 (630, 54, 36)	**0.9987** (345, 16, 11)	- (104, 3, 3)
Sphericity	0.3405	0.6637 (645, 55, 38)	**0.9937** (237, 23, 19)	0.9999 (12, 2, 2)
Spiculation	0.2165	0.4624 (759, 59, 39)	**0.9436** (263, 21, 20)	0.9999 (16, 3, 3)
Subtlety	0.5016	0.5940 (932, 58, 38)	**0.9991** (600, 23, 20)	- (361, 10, 9)
Texture	0.4363	0.6627 (898, 61, 39)	**0.9958** (566, 31, 24)	- (377, 11, 11)

Table 6: Nagelkerke's R^2 from **logistic regression**, single slice per nodule; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (379 images, 149 nodules, 60 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	0.2707	0.6878 (190, 73, 42)	**0.9997** (63, 20, 19)	- (8, 2, 2)
Malignancy	0.5296	0.8196 (187, 73, 42)	**0.9998** (61, 19, 15)	0.9999 (12, 3, 3)
Margin	0.4981	0.8413 (186, 77, 42)	**1.0000** (56, 17, 12)	- (12, 3, 3)
Sphericity	0.3748	0.8092 (197, 77, 45)	**0.9998** (85, 27, 20)	0.9999 (8, 2, 2)
Spiculation	0.2911	0.6790 (200, 77, 43)	**0.9995** (87, 27, 24)	0.9999 (12, 3, 3)
Subtlety	0.5779	0.8018 (194, 70, 41)	**0.9999** (88, 25, 21)	- (40, 10, 9)
Texture	0.4932	0.8639 (222, 79, 43)	**0.9992** (120, 34, 24)	- (44, 11, 11)

We used further the logistic regression to predict the radiologist interpretations. In Table 7 to Table 9, the classification accuracy of the models from logistic regression using leave-one-out cross validation are presented.

Table 7: Classification accuracy (hit ratio) from **logistic regression**, multiple slices per nodule *before* removing end caps, leave-one-out cross validation; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (1989 images, 149 nodules, 60 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	35.34%	42.42% (943, 73, 42)	**62.54%** (331, 20, 19)	- (20, 2, 2)
Malignancy	50.73%	67.98% (1090, 73, 42)	**88.47%** (295, 19, 15)	100.00% (85, 3, 3)
Margin	43.44%	58.82% (981, 77, 42)	**86.50%** (452, 17, 12)	- (128, 3, 3)
Sphericity	40.77%	54.77% (1028, 77, 45)	**72.00%** (400, 27, 20)	64.29% (28, 2, 2)
Spiculation	42.79%	51.09% (1145, 77, 43)	**62.53%** (427, 27, 24)	95.00% (40, 3, 3)
Subtlety	55.81%	72.02% (1308, 70, 41)	**94.39%** (775, 25, 21)	- (452, 10, 9)
Texture	61.69%	77.34% (1333, 79, 43)	**97.13%** (801, 34, 24)	- (482, 11, 11)
Average	47.22% (1989)	61.82% (1118)	**80.51%** (497)	92.16% (51)

Table 8 shows that the classification accuracies of the models improve after removing end caps. However, in Table 9 the classification accuracies of the models for single slice per nodule per radiologist are not as good as we expected especially when at least 3 radiologists agreed on the same ratings. One possible reason is that the number of samples in the data sets is too small compared with the number of features; Peduzzi et al. [42] reported a simulation study on how the events per variable (EPV) ratio affects the variability of the coefficients in logistic regression and suggested a rule of thumb that logistic models should be used with a minimum of 10 EPV.

Table 8: Classification accuracy (hit ratio) from **logistic regression**, multiple slices per nodule *after* removing end caps, leave-one-out cross validation; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (1259 images, 112 nodules, 52 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	37.17%	46.17% (600, 55, 36)	**74.77%** (214, 14, 14)	- (4, 1, 1)
Malignancy	52.10%	71.65% (723, 59, 37)	**97.71%** (175, 13, 11)	98.36% (61, 3, 3)
Margin	47.58%	68.25% (630, 54, 36)	**92.46%** (345, 16, 11)	- (104, 3, 3)
Sphericity	44.16%	60.62% (645, 55, 38)	**75.11%** (237, 23, 19)	50.00% (12, 2, 2)
Spiculation	41.06%	50.99% (759, 59, 39)	**74.14%** (263, 21, 20)	87.50% (16, 3, 3)
Subtlety	67.59%	80.79% (932, 58, 38)	**98.67%** (600, 23, 20)	- (361, 10, 9)
Texture	66.00%	77.51% (898, 61, 39)	**98.76%** (566, 31, 24)	- (377, 11, 11)
Average	50.81% (1259)	66.55% (741)	**87.37%** (343)	89.89% (30)

Table 9: Classification accuracy (hit ratio) from **logistic regression**, single slice per nodule, leave-one-out cross validation; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (379 images, 149 nodules, 60 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	34.04%	**47.78%** (180, 73, 42)	41.27% (63, 20, 19)	- (8, 2, 2)
Malignancy	47.49%	**56.68%** (187, 73, 42)	31.15% (61, 19, 15)	100.00% (12, 3, 3)
Margin	40.11%	52.69% (186, 77, 42)	**57.14%** (56, 17, 12)	- (12, 3, 3)
Sphericity	40.11%	55.84% (197, 77, 45)	**58.82%** (85, 27, 20)	87.50% (8, 2, 2)
Spiculation	40.63%	54.50% (200, 77, 43)	**62.07%** (87, 27, 24)	83.33% (12, 3, 3)
Subtlety	44.85%	53.61% (194, 70, 41)	**93.18%** (88, 25, 21)	- (40, 10, 9)
Texture	53.56%	74.32% (222, 79, 43)	**94.17%** (120, 34, 24)	- (44, 11, 11)
Average	42.97% (379)	56.95% (195)	**62.54%** (80)	90.62% (11)

Decision Trees Results

A similar evaluation was performed for the decision tree approach. In Table 10 to Table 12, the classification accuracies of the models from decision trees using leave-one-out cross validation are presented.

Table 10: Classification accuracy (hit ratio) from **decision trees**, multiple slices per nodule *before* removing end caps, leave-one-out cross validation; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (1989 images, 149 nodules, 60 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	42.23%	65.75% (943, 73, 42)	**90.03%** (331, 20, 19)	- (20, 2, 2)
Malignancy	46.41%	74.22% (1090, 73, 42)	**89.49%** (295, 19, 15)	98.82% (85, 3, 3)
Margin	43.34%	72.27% (981, 77, 42)	**90.04%** (452, 17, 12)	- (128, 3, 3)
Sphericity	41.88%	63.72% (1028, 77, 45)	**88.75%** (400, 27, 20)	96.43% (28, 2, 2)
Spiculation	44.19%	65.94% (1145, 77, 43)	**75.18%** (427, 27, 24)	97.50% (40, 3, 3)
Subtlety	53.49%	77.60% (1308, 70, 41)	**93.55%** (775, 25, 21)	- (452, 10, 9)
Texture	59.78%	81.25% (1333, 79, 43)	**96.63%** (801, 34, 24)	- (482, 11, 11)
Average	47.33% (1989)	72.13% (1118)	**89.10%** (497)	98.04% (51)

Table 11: Classification accuracy (hit ratio) from **decision trees**, multiple slices per nodule *after* removing end caps, leave-one-out cross validation; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (1259 images, 112 nodules, 52 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	47.50%	73.83% (600, 55, 36)	**93.46%** (214, 14, 14)	- (4, 1, 1)
Malignancy	49.32%	76.35% (723, 59, 37)	**94.86%** (175, 13, 11)	98.36% (61, 3, 3)
Margin	47.82%	79.05% (630, 54, 36)	**93.91%** (345, 16, 11)	- (104, 3, 3)
Sphericity	43.69%	71.78% (645, 55, 38)	**85.23%** (237, 23, 19)	91.67% (12, 2, 2)
Spiculation	48.21%	67.46% (759, 59, 39)	**81.37%** (263, 21, 20)	100.00% (16, 3, 3)
Subtlety	64.89%	84.01% (932, 58, 38)	**97.50%** (600, 23, 20)	- (361, 10, 9)
Texture	65.45%	81.51% (898, 61, 39)	**98.41%** (566, 31, 24)	- (377, 11, 11)
Average	52.41% (1259)	76.79% (741)	**92.11%** (343)	97.75% (30)

Table 12: Classification accuracy (hit ratio) from **decision trees**, single slice per nodule, leave-one-out cross validation; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (379 images, 149 nodules, 60 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	27.44%	57.22% (180, 73, 42)	**68.25%** (63, 20, 19)	- (8, 2, 2)
Malignancy	42.22%	68.98% (187, 73, 42)	**90.16%** (61, 19, 15)	91.67% (12, 3, 3)
Margin	35.36%	61.83% (186, 77, 42)	**82.14%** (56, 17, 12)	- (12, 3, 3)
Sphericity	36.15%	63.45% (197, 77, 45)	**71.76%** (85, 27, 20)	87.50% (8, 2, 2)
Spiculation	36.15%	69.50% (200, 77, 43)	**78.16%** (87, 27, 24)	66.67% (12, 3, 3)
Subtlety	38.79%	65.46% (194, 70, 41)	**94.32%** (88, 25, 21)	- (40, 10, 9)
Texture	53.56%	81.08% (222, 79, 43)	**98.33%** (120, 34, 24)	- (44, 11, 11)
Average	38.52% (379)	67.20% (195)	**83.30%** (80)	81.25% (11)

Table 11 shows that the classification accuracies of the models improve after removing end caps. The classification accuracies of the models for single slice per nodule per radiologist (Table 12) also dropped but not as much as logistic regression. With at least 3 agreements, most characteristics were predicted with accuracy higher than 80% except for lobulation, sphericity, and spiculation. These three characteristics are related to margin and shape of the margin and we expect that the image features used in this study were not able to capture the shape and boundary properties of the nodules.

Comparing Table 9 and Table 12, we noticed that decision trees provide higher accuracy than logistic regression. One possible explanation is the large number of variables compared with the number of cases. We performed a step-wise feature selection for the logistic model, but the full-model still produced the best results for the logistic model. Another explanation is given by the non-linearity of the data (given the complex structure of the image features encoding the nodules' visual appearance) that is better handled by the decision tree approach versus the logistic regression approach.

SVM Results

A similar evaluation was performed for SVM. Tables 13 to 15 show the classification accuracies of the models from SVM using leave-one-out cross validation. For multiple slices per nodule as presented in Table 13 and removing end caps as presented in Table 14, SVM performs better than decision trees at 95%

significance level (except multiple slices per nodule when at least 2 radiologists agreed which SVM still performs better than decision trees at 90% significance level). For single slice per nodule (the bias in the data is eliminated), Table 15 shows that there are no significant differences between the SVM and decision tree classification results at a .95 confidence level. The highest accuracies are obtained on the nodules for which there is agreement among at least three radiologists.

Table 13: Classification accuracy (hit ratio) from **SVM**, multiple slices per nodule *before* removing end caps, leave-one-out cross validation; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (1989 images, 149 nodules, 60 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	46.76%	73.91% (943, 73, 42)	**96.37%** (331, 20, 19)	- (20, 2, 2)
Malignancy	55.51%	77.25% (1090, 73, 42)	**96.61%** (295, 19, 15)	100.00% (85, 3, 3)
Margin	47.71%	71.76% (981, 77, 42)	**96.02%** (452, 17, 12)	- (128, 3, 3)
Sphericity	48.97%	71.40% (1028, 77, 45)	**89.50%** (400, 27, 20)	100.00% (28, 2, 2)
Spiculation	50.73%	70.83% (1145, 77, 43)	**92.27%** (427, 27, 24)	100.00% (40, 3, 3)
Subtlety	59.02%	77.98% (1308, 70, 41)	**94.84%** (775, 25, 21)	- (452, 10, 9)
Texture	67.12%	84.02% (1333, 79, 43)	**98.63%** (801, 34, 24)	- (482, 11, 11)
Average	53.69%	75.73% (1118)	**94.89%** (497)	100.00% (51)

Table 14: Classification accuracy (hit ratio) from **SVM**, multiple slices per nodule *after* removing end caps, leave-one-out cross validation; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (1259 images, 112 nodules, 52 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	50.04%	80.67% (600, 55, 36)	**96.73%** (214, 14, 14)	- (4, 1, 1)
Malignancy	58.06%	88.77% (723, 59, 37)	**99.43%** (175, 13, 11)	100.00% (61, 3, 3)
Margin	52.74%	83.65% (630, 54, 36)	**96.81%** (345, 16, 11)	- (104, 3, 3)
Sphericity	51.79%	78.76% (645, 55, 38)	**93.25%** (237, 23, 19)	100.00% (12, 2, 2)
Spiculation	53.14%	71.54% (759, 59, 39)	**97.72%** (263, 21, 20)	100.00% (16, 3, 3)
Subtlety	69.18%	85.41% (932, 58, 38)	**99.00%** (600, 23, 20)	- (361, 10, 9)
Texture	71.09%	85.97% (898, 61, 39)	**98.59%** (566, 31, 24)	- (377, 11, 11)
Average	58.01%	82.36% (741)	**97.36%** (343)	100.00% (30)

Table 15: Classification accuracy (hit ratio) from **SVM**, single slice per nodule, leave-one-out cross validation; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Entire dataset** (379 images, 149 nodules, 60 cases)	**At least 2 radiologists agreed**	**At least 3 radiologists agreed**	**All 4 radiologists agreed**
Lobulation	32.98%	67.78% (180, 73, 42)	**84.11%** (63, 20, 19)	- (8, 2, 2)
Malignancy	46.97%	71.66% (187, 73, 42)	**98.36%** (61, 19, 15)	100.00% (12, 3, 3)
Margin	41.16%	81.57% (186, 77, 42)	**92.86%** (56, 17, 12)	- (12, 3, 3)
Sphericity	43.27%	67.51% (197, 77, 45)	**85.88%** (85, 27, 20)	87.50% (8, 2, 2)
Spiculation	43.27%	71.00% (200, 77, 43)	**86.21%** (87, 27, 24)	91.67% (12, 3, 3)
Subtlety	44.06%	70.62% (194, 70, 41)	**97.73%** (88, 25, 21)	- (40, 10, 9)
Texture	54.62%	76.58% (222, 79, 43)	**100.00%** (120, 34, 24)	- (44, 11, 11)
Average	43.76%	72.45% (195)	**92.16%** (80)	93.75% (11)

Visual Ontology Results

Since logistic regression did not provide high classification accuracies and SVM did not provide significantly better results than decision trees, we consider only the important image features for each characteristic based on the results of the decision trees which are easier to interpret and understand. Furthermore, we consider the mappings for single slice per nodule per radiologist with at least 3 agreements, since 1) there is no bias from the representation of many instances (slices) of each nodule marked by each radiologist, and 2) the models with at least 3 agreements provide us the highest classification accuracies.

From the decision rules for the seven radiologists' characteristics learned by decision trees, we constructed a visual ontology for lung nodule interpretation based on the most important low-level image features. Since the entire ontology is too large to be presented in one diagram, we present the diagram of each characteristic as one separate figure but they can be linked together at the end to form the entire visual ontology as follows: the "lung nodule" concept will be at the top of the ontology, followed by the seven semantic concepts on the second level and then, Figures 5 to 11 will generate the rest of the levels of the ontology diagram. Furthermore, each one of these figures is composed of all possible ratings for the appropriate characteristic along with rules that correspond to each rating. Given a nodule image, the ontology can provide probabilistic ratings for each characteristic.

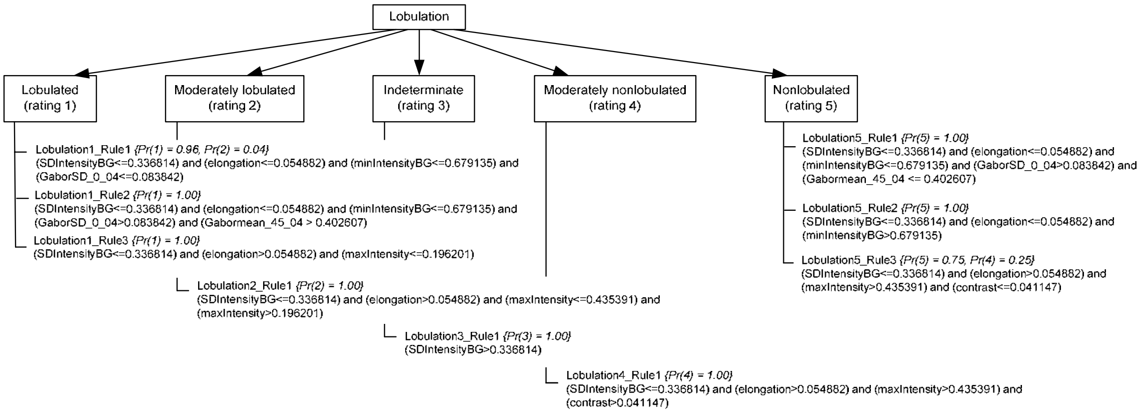

Lobulation
Lobulated (rating 1)
Moderately lobulated (rating 2)
Indeterminate (rating 3)
Moderately nonlobulated (rating 4)
Nonlobulated (rating 5)
Lobulation1_Rule1 {Pr(1) = 0.96, Pr(2) = 0.04}
(SDIntensityBG<=0.336814) and (elongation<=0.054882) and (minIntensityBG<=0.679135) and (GaborSD_0_04<=0.083842)
Lobulation1_Rule2 {Pr(1) = 1.00}
(SDIntensityBG<=0.336814) and (elongation<=0.054882) and (minIntensityBG<=0.679135) and (GaborSD_0_04>0.083842) and (Gabormean_45_04 > 0.402607)
Lobulation1_Rule3 {Pr(1) = 1.00}
(SDIntensityBG<=0.336814) and (elongation>0.054882) and (maxIntensity<=0.196201)
Lobulation2_Rule1 {Pr(2) = 1.00}
(SDIntensityBG<=0.336814) and (elongation>0.054882) and (maxIntensity<=0.435391) and (maxIntensity>0.196201)
Lobulation3_Rule1 {Pr(3) = 1.00}
(SDIntensityBG>0.336814)
Lobulation4_Rule1 {Pr(4) = 1.00}
(SDIntensityBG<=0.336814) and (elongation>0.054882) and (maxIntensity>0.435391) and (contrast>0.041147)
Lobulation5_Rule1 {Pr(5) = 1.00}
(SDIntensityBG<=0.336814) and (elongation<=0.054882) and (minIntensityBG<=0.679135) and (GaborSD_0_04>0.083842) and (Gabormean_45_04 <= 0.402607)
Lobulation5_Rule2 {Pr(5) = 1.00}
(SDIntensityBG<=0.336814) and (elongation<=0.054882) and (minIntensityBG>0.679135)
Lobulation5_Rule3 {Pr(5) = 0.75, Pr(4) = 0.25}
(SDIntensityBG<=0.336814) and (elongation>0.054882) and (maxIntensity>0.435391) and (contrast<=0.041147)

Figure 5: A part of a visual ontology for lung nodule interpretation with regard to lobulation

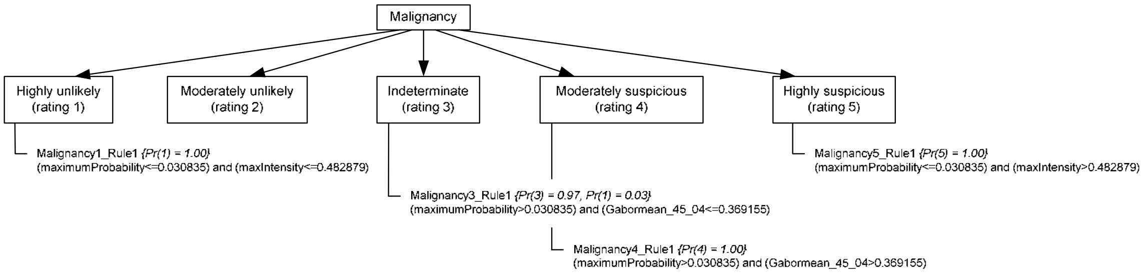

Malignancy
Highly unlikely (rating 1)
Moderately unlikely (rating 2)
Indeterminate (rating 3)
Moderately suspicious (rating 4)
Highly suspicious (rating 5)
Malignancy1_Rule1 {Pr(1) = 1.00}
(maximumProbability<=0.030835) and (maxIntensity<=0.482879)
Malignancy3_Rule1 {Pr(3) = 0.97, Pr(1) = 0.03}
(maximumProbability>0.030835) and (Gabormean_45_04<=0.369155)
Malignancy4_Rule1 {Pr(4) = 1.00}
(maximumProbability>0.030835) and (Gabormean_45_04>0.369155)
Malignancy5_Rule1 {Pr(5) = 1.00}
(maximumProbability<=0.030835) and (maxIntensity>0.482879)

Figure 6: A part of a visual ontology for lung nodule interpretation with regard to malignancy

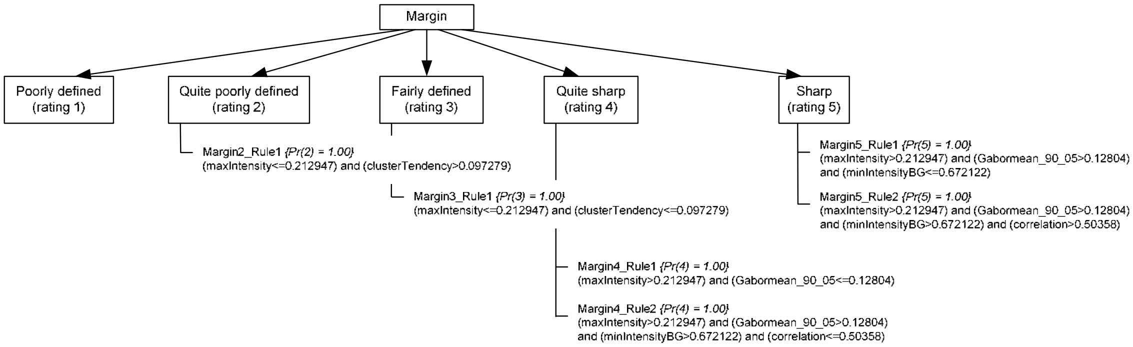

Margin
Poorly defined (rating 1)
Quite poorly defined (rating 2)
Fairly defined (rating 3)
Quite sharp (rating 4)
Sharp (rating 5)
Margin2_Rule1 {Pr(2) = 1.00}
(maxIntensity<=0.212947) and (clusterTendency>0.097279)
Margin3_Rule1 {Pr(3) = 1.00}
(maxIntensity<=0.212947) and (clusterTendency<=0.097279)
Margin4_Rule1 {Pr(4) = 1.00}
(maxIntensity>0.212947) and (Gabormean_90_05<=0.12804)
Margin4_Rule2 {Pr(4) = 1.00}
(maxIntensity>0.212947) and (Gabormean_90_05>0.12804) and (minIntensityBG>0.672122) and (correlation<=0.50358)
Margin5_Rule1 {Pr(5) = 1.00}
(maxIntensity>0.212947) and (Gabormean_90_05>0.12804) and (minIntensityBG<=0.672122)
Margin5_Rule2 {Pr(5) = 1.00}
(maxIntensity>0.212947) and (Gabormean_90_05>0.12804) and (minIntensityBG>0.672122) and (correlation>0.50358)

Figure 7: A part of a visual ontology for lung nodule interpretation with regard to margin

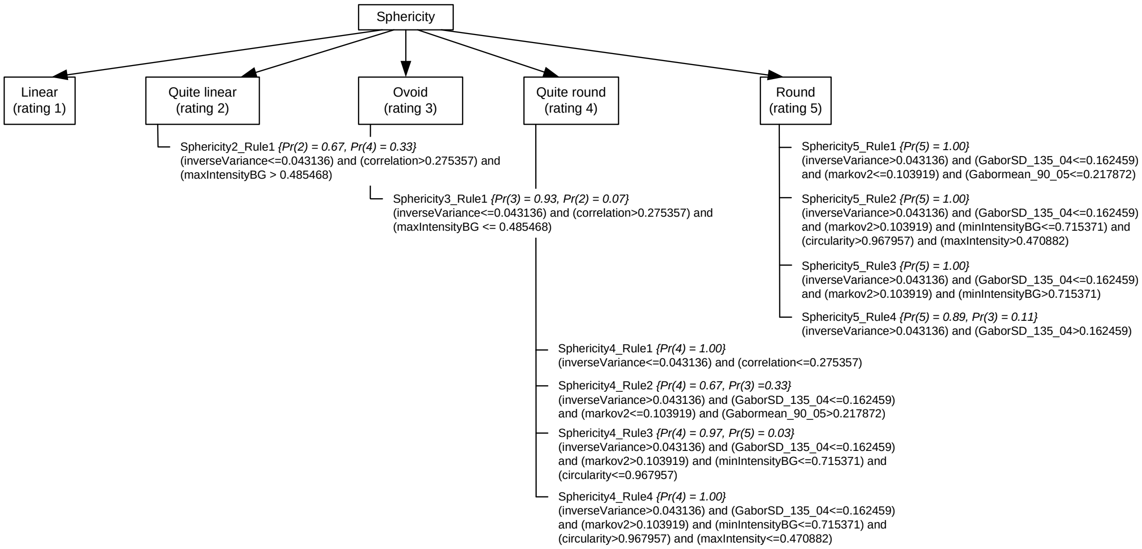

Sphericity
Linear
(rating 1)
Quite linear
(rating 2)
Ovoid
(rating 3)
Quite round
(rating 4)
Round
(rating 5)
Sphericity2_Rule1 *{Pr(2) = 0.67, Pr(4) = 0.33}*
(inverseVariance<=0.043136) and (correlation>0.275357) and
(maxIntensityBG > 0.485468)
Sphericity3_Rule1 *{Pr(3) = 0.93, Pr(2) = 0.07}*
(inverseVariance<=0.043136) and (correlation>0.275357) and
(maxIntensityBG <= 0.485468)
Sphericity5_Rule1 *{Pr(5) = 1.00}*
(inverseVariance>0.043136) and (GaborSD_135_04<=0.162459)
and (markov2<=0.103919) and (Gabormean_90_05<=0.217872)
Sphericity5_Rule2 *{Pr(5) = 1.00}*
(inverseVariance>0.043136) and (GaborSD_135_04<=0.162459)
and (markov2>0.103919) and (minIntensityBG<=0.715371) and
(circularity>0.967957) and (maxIntensity>0.470882)
Sphericity5_Rule3 *{Pr(5) = 1.00}*
(inverseVariance>0.043136) and (GaborSD_135_04<=0.162459)
and (markov2>0.103919) and (minIntensityBG>0.715371)
Sphericity5_Rule4 *{Pr(5) = 0.89, Pr(3) = 0.11}*
(inverseVariance>0.043136) and (GaborSD_135_04>0.162459)
Sphericity4_Rule1 *{Pr(4) = 1.00}*
(inverseVariance<=0.043136) and (correlation<=0.275357)
Sphericity4_Rule2 *{Pr(4) = 0.67, Pr(3) =0.33}*
(inverseVariance>0.043136) and (GaborSD_135_04<=0.162459)
and (markov2<=0.103919) and (Gabormean_90_05>0.217872)
Sphericity4_Rule3 *{Pr(4) = 0.97, Pr(5) = 0.03}*
(inverseVariance>0.043136) and (GaborSD_135_04<=0.162459)
and (markov2>0.103919) and (minIntensityBG<=0.715371) and
(circularity<=0.967957)
Sphericity4_Rule4 *{Pr(4) = 1.00}*
(inverseVariance>0.043136) and (GaborSD_135_04<=0.162459)
and (markov2>0.103919) and (minIntensityBG<=0.715371) and
(circularity>0.967957) and (maxIntensity<=0.470882)

Figure 8: A part of a visual ontology for lung nodule interpretation with regard to sphericity

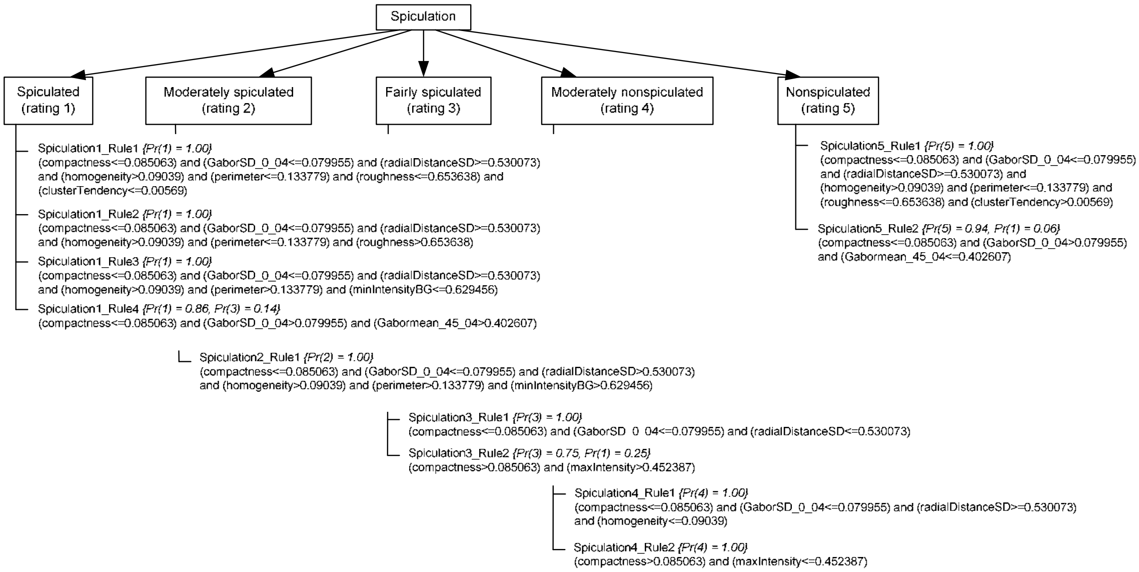

Spiculation
Spiculated
(rating 1)
Moderately spiculated
(rating 2)
Fairly spiculated
(rating 3)
Moderately nonspiculated
(rating 4)
Nonspiculated
(rating 5)
Spiculation1_Rule1 *{Pr(1) = 1.00}*
(compactness<=0.085063) and (GaborSD_0_04<=0.079955) and (radialDistanceSD>=0.530073)
and (homogeneity>0.09039) and (perimeter<=0.133779) and (roughness<=0.653638) and
(clusterTendency<=0.00569)
Spiculation1_Rule2 *{Pr(1) = 1.00}*
(compactness<=0.085063) and (GaborSD_0_04<=0.079955) and (radialDistanceSD>=0.530073)
and (homogeneity>0.09039) and (perimeter<=0.133779) and (roughness>0.653638)
Spiculation1_Rule3 *{Pr(1) = 1.00}*
(compactness<=0.085063) and (GaborSD_0_04<=0.079955) and (radialDistanceSD>=0.530073)
and (homogeneity>0.09039) and (perimeter>0.133779) and (minIntensityBG<=0.629456)
Spiculation1_Rule4 *{Pr(1) = 0.86, Pr(3) = 0.14}*
(compactness<=0.085063) and (GaborSD_0_04>0.079955) and (Gabormean_45_04>0.402607)
Spiculation5_Rule1 *{Pr(5) = 1.00}*
(compactness<=0.085063) and (GaborSD_0_04<=0.079955)
and (radialDistanceSD>=0.530073) and
(homogeneity>0.09039) and (perimeter<=0.133779) and
(roughness<=0.653638) and (clusterTendency>0.00569)
Spiculation5_Rule2 *{Pr(5) = 0.94, Pr(1) = 0.06}*
(compactness<=0.085063) and (GaborSD_0_04>0.079955)
and (Gabormean_45_04<=0.402607)
Spiculation2_Rule1 *{Pr(2) = 1.00}*
(compactness<=0.085063) and (GaborSD_0_04<=0.079955) and (radialDistanceSD>0.530073)
and (homogeneity>0.09039) and (perimeter>0.133779) and (minIntensityBG>0.629456)
Spiculation3_Rule1 *{Pr(3) = 1.00}*
(compactness<=0.085063) and (GaborSD 0 04<=0.079955) and (radialDistanceSD<=0.530073)
Spiculation3_Rule2 *{Pr(3) = 0.75, Pr(1) = 0.25}*
(compactness>0.085063) and (maxIntensity>0.452387)
Spiculation4_Rule1 *{Pr(4) = 1.00}*
(compactness<=0.085063) and (GaborSD_0_04<=0.079955) and (radialDistanceSD>=0.530073)
and (homogeneity<=0.09039)
Spiculation4_Rule2 *{Pr(4) = 1.00}*
(compactness>0.085063) and (maxIntensity<=0.452387)

Figure 9: A part of a visual ontology for lung nodule interpretation with regard to spiculation

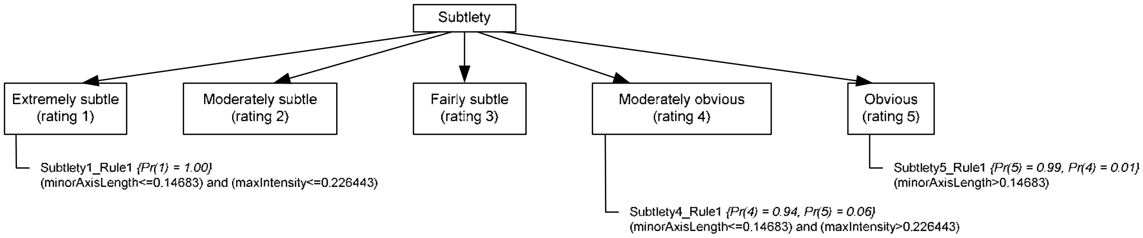

Subtlety
Extremely subtle
(rating 1)
Moderately subtle
(rating 2)
Fairly subtle
(rating 3)
Moderately obvious
(rating 4)
Obvious
(rating 5)
Subtlety1_Rule1 *{Pr(1) = 1.00}*
(minorAxisLength<=0.14683) and (maxIntensity<=0.226443)
Subtlety5_Rule1 *{Pr(5) = 0.99, Pr(4) = 0.01}*
(minorAxisLength>0.14683)
Subtlety4_Rule1 *{Pr(4) = 0.94, Pr(5) = 0.06}*
(minorAxisLength<=0.14683) and (maxIntensity>0.226443)

Figure 10: A part of a visual ontology for lung nodule interpretation with regard to subtlety

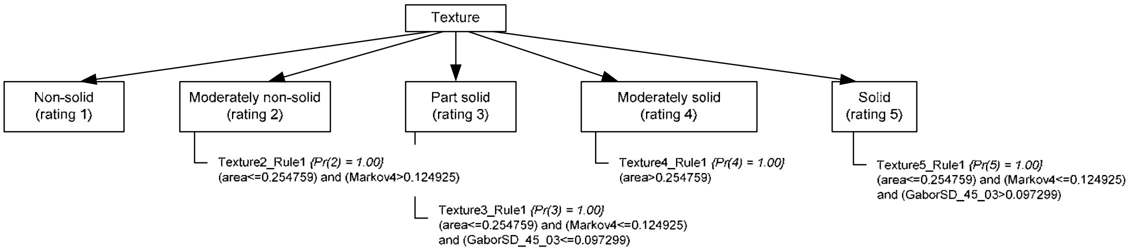

Texture
Non-solid (rating 1)
Moderately non-solid (rating 2)
Part solid (rating 3)
Moderately solid (rating 4)
Solid (rating 5)
Texture2_Rule1 {Pr(2) = 1.00}
(area<=0.254759) and (Markov4>0.124925)
Texture3_Rule1 {Pr(3) = 1.00}
(area<=0.254759) and (Markov4<=0.124925)
and (GaborSD_45_03<=0.097299)
Texture4_Rule1 {Pr(4) = 1.00}
(area>0.254759)
Texture5_Rule1 {Pr(5) = 1.00}
(area<=0.254759) and (Markov4<=0.124925)
and (GaborSD_45_03>0.097299)

Figure 11: A part of a visual ontology for lung nodule interpretation with regard to texture

Analyzing the ontology (in particular the rules under each one of the ratings), we noticed that not all the features are selected to encode the semantic characteristic. In Table 16 we list all the important features for each characteristic in descending order based on their information gain values. Table 16 also shows that all seven characteristics can be predicted by using at most 10 (less than 16%) from all 64 image features. Furthermore, most important features are found to be the ones generated by the gray-level intensity and texture features. However, there are some shape features that are important for some characteristics related to the shape of the nodule such as elongation for lobulation, circularity for sphericity, and compactness, RadialDistanceSD, and roughness for spiculation. There are only 2 important features for subtlety of the nodule, MinorAxisLength and MaxIntensity, which means that only size and gray-level intensity are enough to identify the subtlety of the nodule or that subtlety is independent of shape and texture of a nodule. The complexity of the ontology is related to the complexity of the decision tree. The optimal decision tree for the LIDC data is summarized in Table 17.

Table 16: Important features for each characteristic based on the information gain criterion

Characteristics	Important features	Total number of important features
Lobulation	SDIntensityBG, elongation, MinIntensityBG, MaxIntensity, GaborSD_0_04, contrast, Gabormean_45_04	7 (10.94%)
Malignancy	maximumProbability, MaxIntensity, Gabormean_45_04	3 (4.69%)
Margin	MaxIntensity, clusterTendency, Gabormean_90_05, MinIntensityBG, correlation	5 (7.81%)
Sphericity	inverseVariance, correlation, GaborSD_135_04, MaxIntensityBG, Markov2, Gabormean_90_05, MinIntensityBG, circularity, maxIntensity	9 (14.06%)
Spiculation	Compactness, GaborSD_0_04, maxIntensity, RadialDistanceSD, Gabormean_45_04, homogeneity, perimeter, roughness, MinIntensityBG, clusterTendency	10 (15.63%)
Subtlety	MinorAxisLength, MaxIntensity	2 (3.13%)
Texture	Area, Markov4, GaborSD_45_03	3 (4.69%)

Table 17: Number of nodes, number of rules, depth of the tree, and complexity of the tree for each characteristic

Characteristic	Number of nodes	Numbers of rules	Depth of the tree	Complexity of the tree
Lobulation	17	9	5	80.85
Malignancy	7	4	2	33.22
Margin	11	6	4	43.93
Sphericity	19	10	6	109.27
Spiculation	21	11	7	104.87
Subtlety	5	3	2	31.88
Texture	7	4	3	27.29

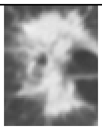
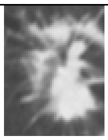
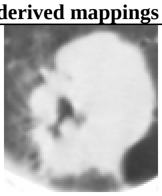
From Table 18, the classification accuracies from decision trees when using only important features for each characteristic are higher than or equal to the classification accuracies from decision trees when using all 64 image features. These results show that using only the important features for each characteristic from the visual ontology are sufficient to predict the rating for each characteristic.

Table 18: Classification accuracy (hit ratio) from **decision trees**, single slice per nodule when at least 3 radiologists agreed, leave-one-out cross validation when using all 64 image features and when using only important features for each characteristic from Table 16; the triplets represent the number of nodule images, number of nodules, and number of cases, respectively.

Characteristics	**Using all 64 image features**	**Using only important features**
Lobulation	68.25% (63, 20, 19)	80.95% (63, 20, 19)
Malignancy	90.16% (61, 19, 15)	91.80% (61, 19, 15)
Margin	82.14% (56, 17, 12)	85.71% (56, 17, 12)
Sphericity	71.76% (85, 27, 20)	72.94% (85, 27, 20)
Spiculation	78.16% (87, 27, 24)	78.16% (87, 27, 24)
Subtlety	94.32% (88, 25, 21)	96.59% (88, 25, 21)
Texture	98.33% (120, 34, 24)	98.33% (120, 34, 24)

As described in Section 4.2, the visual ontology can be used to improve the interpretation process by making objective recommendations based on the learned image-semantics mappings. In Table 19, we show an example where the ontology was used to recommend probabilistic ratings for each one of the seven semantic concepts characterizing the nodule of interest. If the recommended ratings do not correspond to the ones that the radiologists has in mind, then the radiologists can look at other nodule images rated similarly by the system and attempt to understand the computer interpretation/quantification of the nodule appearance. As a consequence, the radiologist could either recognize that his interpretation of a certain nodule appearance does not corresponds to the equivalent low-level image features used by the medical imaging community or, if he agrees on the recommended ratings, change his apriori ratings accordingly.

Table 19: An example of nodule image along with probabilistic ratings and recommended ratings for all characteristics and other nodule images that have the same ratings as the example image using the derived mappings

Nodule image of Interest			**Other nodule images with the same ratings using the derived mappings**		
 Nodule 1			 Nodule 2	 Nodule 3	 Nodule 4
Characteristics	Predicted Probabilistic Ratings for Nodule 1	Recommended Ratings for Nodule 1	Recommended Ratings for Nodule 2	Recommended Ratings for Nodule 3	Recommended Ratings for Nodule 4
Lobulation	Pr(4) = 1.00	4	**4**	3	**4**
Malignancy	Pr(1) = 1.00	1	**1**	**1**	**1**
Margin	Pr(4) = 1.00	4	**4**	**4**	5
Sphericity	Pr(4) = 0.97, Pr(5) = 0.03	4	**4**	**4**	**4**
Spiculation	Pr(3) = 0.75, Pr(1) = 0.25	3	**3**	**3**	**3**
Subtlety	Pr(5) = 0.99, Pr(4) = 0.01	5	**5**	**5**	**5**
Texture	Pr(4) = 1.00	4	**4**	**4**	5

6. Conclusions and Future Work

Based on our preliminary results, we learned that it is possible to probabilistically model lung nodule image semantics (lobulation, malignancy, margin, sphericity, spiculation, subtlety, and texture) using image content (shape, size, gray-level intensity, and texture). We also learned that certain classifiers perform better for certain characteristics than others; for example, SVM performed significantly better on lobulation and sphericity than decision trees and logistic regression. Therefore, as future work, we plan to investigate ensemble of classifiers to improve the accuracy of our probabilistic mappings. Furthermore, in addition to modeling the uncertainty in the radiologists' annotations, we plan to investigate various approaches on combining the radiologists' delineated boundaries and study the effect of these combinations on the image features and further, on the accuracy of the probabilistic mappings.

We showed that the mappings can be used to automatically build visual-ontologies for lung nodule interpretation. In the long term, the visual ontology can be integrated in the radiology lexicon, RadLex, to provide radiologists with both a set of uniform and objective image features and their relationship to nodules' semantics with the final goal of better interpretation and less variance across multiple readers.

References

1. Andrysiak T and Choras M, "Image retrieval based on hierarchical Gabor filters", International Journal Applied Computer Science, 15(4), 471-480, 2005
2. Aoyama M, Li Q, Katsuragawa S, Li F, Sone S, and Doi K, "Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose CT images", Medical Physics, 30(3): 387–394, 2003
3. 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
4. Armato SG, McLennan G, McNitt-Gray MF, Meyer CR, Yankelevitz D, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA, MacMahon H, Reeves AP, Croft BY, and Clarke LP, "Lung Image Database Consortium: Developing a resource for the medical imaging research community", Radiology, 232(3): 739–748, 2004
5. Barb AS, Shyu,C-R and Sethi Y P, "Knowledge Representation and Sharing Using Visual Semantic Modeling for Diagnostic Medical Image Databases, IEEE Transactions Information Technology in Biomedicine, 9(4), 2005
6. Bouman C, "Markov random fields and stochastic image models", in 1995 IEEE International Conference on Image Processing, 1995, Tutorial notes
7. Brinkley JF., Rosse C., "Imaging Informatics and the Human Brain Project: the Role of Structure, Review", Yearbook of Medical Informatics, 2002: pp.111-128, 2002
8. Cesmeli E and Wang D, "Texture segmentation using gaussian-markov random fields and neural oscillator networks", IEEE Transactions on Neural Networks, vol 12, 394-404, March 2001
9. Doi K, "Current status and future potential of computer-aided diagnosis in medical imaging," The British Journal of Radiology 2005
10. Dreyer KJ "The Alchemy of Data Mining", Imaging Economics, 2005
11. Ebadollahi S, Chang SF, Wu H, Takoma S, "Echocardiogram Video Summarization", SPIE Medical Imaging 2001, pp. 492-501, 2001
12. Goo JM, Lee JW, Lee HJ, Kim S, Kim JH, Im J, "Automated Lung Nodule Detection at Low-Dose CT: Preliminary Experience", Korean J Radiology, vol. 4, 211-216, 2003
13. Gurney J, "Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part I. Theory", Radiology, 186(2): 405–413, 1993
14. Gurney J, Lyddon D, and McKay J, "Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis. Part II. Application", Radiology, 186(2): 415–422, 1993
15. Han J, Kamber M, "Data Mining: Concepts and Techniques", Morgan Kaufman Publishers, Second Edition, 2006
16. Haralick RM, Shanmugam K, and Dinstein I, "Textural Features for Image Classification", IEEE Trans. On Systems, Man, and Cybernetics, 3(6): 610-621, 1973
17. Hearst MA, "Support vector machines", IEEE Intelligent Systems, 13(4): 18–28, 1998
18. Hollings N, Shaw P, "Diagnostic Imaging of Lung Cancer", European Respiratory Journal, vol. 19, 722-742, 2002
19. Hu B, Dasmahapatra S, Lewis P, Shadbolt N, "Ontology-based Medical Image Annotation with Description Logics", Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, 77-82, 2003
20. Kahn CE, Channin DS, Rubin DL, "An Ontology for PACS Integration", Journal of Digital Imaging, 19(4): 316-327, 2006
21. Ketai L, Malby M, Jordan K, Meholic A, Locken J, "Small Nodules Detected on Chest Radiography: Does Size Predict Calcification?", CHEST, vol. 118, 610-614, 2000
22. Lam M, Disney T, Pham M, Raicu D, Furst J, "Content-Based Image Retrieval for Pulmonary Computed Tomography Nodule Images", SPIE Medical Imaging Conference, San Diego, CA, February 2007
23. Leroy G and Chen H, "Meeting medical terminology needs-the ontology enhanced medical concept mapper", IEEE Transaction on Information Technology in Biomedicine,5(4), 2004

24. Li F, Li Q, Engelmann R, Aoyama M, Sone S, MacMahon H, Doi K, “Improving Radiologists’ Recommendations With Computer-Aided Diagnosis for Management of Small Nodules Detected by CT”, Academic Radiology, 13(8): 943-950, 2006
25. Li Q, Li F, Shiraishi J, Katsuragawa S, Sone S, Doi K, “Investigation of new psychophysical measures for evaluation of similar images on thoracic computed tomography for distinction between benign and malignant nodules”, Medical Physics, 30(10): 2584-93, 2003
26. Lindberg DAB and Humphreys BL, UMLS Knowledge Sources, 14thed. Bethesda, MD: Na. Library Med. 2003AB, 5(4): 261–270, 2003
27. Liu S, and Li J, “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
28. Lo S-C B, Li-Yueh Hsu MTF, and Lure Y-M F, and Zhao H, “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, 183–189, 2003
29. Lu G, "Design issues of multimedia information indexing and retrieval systems", Journal of Network and Computer Applications (Academic Press) , 22(3):175-198, July 1999
30. Maillot N, Ontology Based Object Learning and Recognition, PhD Thesis, 2005
31. Matsuki Y, Nakamura K, Watanabe H, Aoki T, Nakata H, Katsuragawa S, and Doi K, “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, 178(3): 657–663, 2002
32. Matsuoka S, Kurihara Y, Yagihashi K, Niimi H, Nakajima Y, “Peripheral Solitary Pulmonary Nodule: CT Findings in Patients with Pulmonary Emphysema”, Radiology, vol. 235, 266-273, 2005
33. McNitt-Gray MF, Hart EM, Wyckoff N, Sayre JW, Goldin JG, and Aberle DR, “A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results”, Medical Physics, 26(6): 880–888, 1999
34. McNitt-Gray MF, Wyckoff N, Sayre JW, Goldin JG, and Aberle DR, “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, 23(6): 339–348, 1999
35. Mezaris V, Kompatsiaris I, and Strintzis MG, "Region-based Image Retrieval using an Object Ontology and Relevance Feedback", EURASIP Journal on Applied Signal Processing, Special Issue on Object-Based and Semantic Image and Video Analysis, vol. 2004, no. 6, pp. 886-901, June 2004
36. Miller GA, “WordNet: A lexical database for English”, Communications ACM, 38(11): 39–41, 1995
37. Mitchell TM, Machine Learning, McGraw-Hill, 1997
38. Muramatsu C, Li Q, Schmidt R. Suzuki K, Shiraishi J, Newstead G, Doi K, “Experimental determination of subjective similarity for pairs of clustered microcalcifications on mammograms: observer study results”, Medical Physics, 33(9): 3460-8, 2006
39. Muramatsu C, Li Q, Suzuki K, Schmidt RA, Shiraishi J, Newstead GM, Doi K, “Investigation of psychophysical measure for evaluation of similar images for mammographic masses: preliminary results”, Medical Physics, 32(7): 2295-304, 2005
40. Ogiela M, Tadeusiewicz R. "Semantic-oriented syntactic algorithms for content recognition and understanding of images in medical databases", Proceedings of IEEE International Conference on Multimedia and Expo (ICME2001) 2001
41. Opfer R, Wiemker R, “A new general tumor segmentation framework based on radial basis function energy minimization with a validation study on LIDC lung nodules”, SPIE Medical Imaging Conference, San Diego, CA, February 2007
42. Peduzzi P, Concato J, Kemper E, Holford TR, and Feinstein A, “A simulation study of the number of events per variable in logistic regression analysis”, Journal of Clinical Epidemiology 49(12): 1373-1379, 1996
43. Queckel LG, Goei R, Kessels AG, van Engelshoven JM “Detection of lung cancer on the chest radiograph: impact on previous films, clinical information, double reading, and dual reading”, Journal of Clinical Epidemiology; 54: 1146-1150, 2001
44. Quinlan JR, “Improved Use of Continuous Attributes in C4.5”, Journal of Artificial Intelligence Research, 4: 77-90, 1996

45. Raicu DS, Varutbangkul E, Cisneros JG, Furst JD, Channin DS, Armato SG III, "Semantics and Image Content Integration for Pulmonary Nodule Interpretation in Thoracic Computed Tomography", SPIE Medical Imaging Conference, San Diego, CA, February 2007
46. Rosse C, Mejino JLV, Modayur BR, Jakobovits RM, Hinshaw KP, Brinkley JF, "Motivation and Organizational Principles for Anatomical Knowledge Representation: The Digital Anatomist Symbolic Knowledge Base", Journal of the American Medical Informatics Association, 5(1): p. 17-40, 1998
47. Schölkopf B, Burges C, and Smola A, eds., Advances in Kernel Methods - Support Vector Learning, MIT Press, pp. 185-208, 1999
48. Siegelman SS, Khouri NF, Leo FP, Fishman EK, Braverman RM, Zerhouni EA, "Solitary Pulmonary Nodules: CT Assessment", Radiology, vol. 160, 307-312, 1986
49. Sluimer I, Schilham A, Prokop M, and Ginneken B, "Computer Analysis of Computed Tomography Scans of the Lung: A Survey", IEEE Transactions on Medical Imaging, 25(4), 2006
50. Smola A, Bartlett P, Schölkopf B, Schuurmans D, eds., Advances in Large Margin Classifiers, MIT Press, pp. 61-74, 1999.
51. Stokes ME, Davis CS, Koch GG, Categorical Data Analysis Using the SAS System, 2nd Edition, SAS Publishing, NC, 2000
52. Tachibana R and Kido S, "Automatic segmentation of pulmonary nodules on CT images by use of NCI lung image database consortium", SPIE Medical Imaging Conference, San Diego, CA, February 2006
53. Takashima S, Sone S, Li F, Maruyama Y, Hasegawa M, and Kadoya M, "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, 180(5): 1255–1263, 2003
54. Takashima S, Sone S, Li F, Maruyama Y, Hasegawa M, Matsushita T, Takayama F, and Kadoya M, "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, 180(4): 955–964, 2003
55. Tourassi GD, Harrawood B, Singh S, Lo JY, Floyd CE, "Evaluation of information-theoretic similarity measures for content-based retrieval and detection of masses in mammograms", Medical Physics, 34(1): 140-50, 2007
56. Varutbangkul E, Raicu D and Furst J, "A Computer-Aided Diagnosis Framework for Pulmonary Nodule Interpretation in Thoracic Computed Tomography", DePaul CTI Research Symposium (CTIRS 2007), May 2007
57. Witten IH and Frank E, "Data Mining: Practical Machine Learning Tools and Techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005
58. Zerhouni EA, Stitik FP, Siegelman SS, Naidich DP, Sagel SS et al. "CT of the Pulmonary Nodule: A Cooperative Study", Radiology, vol. 160, 319-327, 1986
59. Zhao B, Gamsu G, Ginsberg MS, Jiang L, Schwartz LH, "Automatic Detection of Small Lung Nodules on CT Utilizing a Local Density Maximum Algorithm", Journal of Applied Clinical Medical Physics, 4(3): 248-260, 2003
60. Zheng B, Mello-Thomas C, Wang XH, Abrams GS, Sumkin JH, Chough DM, Ganott MA, Lu A, Gur D, "Interactive computer-aided diagnosis of breast masses: computerized selection of visually similar image sets from a reference library", Academic Radiology, 14(8): 917-27, 2007