

A Model for the Relationship between Semantic and Content Based Similarity using LIDC

Grace Dasovich^a, Robert Kim^b, Dr. Daniela S. Raicu^c, and Dr. Jacob D. Furst^c

^aNorthwestern University, Evanston, IL 60201

^bJohns Hopkins University, Baltimore, MD 21218

^cSchool of Computing, CDM, DePaul University, Chicago, IL 60604

ABSTRACT

There is considerable research in the field of content-based image retrieval (CBIR); however, few of the current systems incorporate radiologists' visual impression of image similarity. Our objective is to bridge the semantic gap between radiologists' ratings and image features. We have been developing a conceptual-based similarity model derived from content-based similarity to improve CBIR. Previous work in our lab reduced the Lung Image Database Consortium (LIDC) data set into a selection of 149 images of unique nodules, each containing nine semantic ratings by four radiologists and 64 computed image features. After evaluating the similarity measures for both content-based and semantic-based features, we selected 116 nodule pairs with a high correlation between both similarities. These pairs were used to generate a linear regression model that predicts semantic similarity with content similarity input with an R^2 value of 0.871. The characteristics and features of nodules that were used for the model were also investigated.

Keywords: Content-based image retrieval, CT scans, lung nodules, semantic-based image retrieval

1. INTRODUCTION

Lung cancer has one of the shortest survival rates after diagnosis of all cancers.¹ Studies show that the presence of similar images assists radiologists in correctly diagnosing lung nodules as benign or malignant.² Content-based image retrieval (CBIR) systems, where image features of a query image are compared to a database of images, are employed to obtain similar images. One such system is BRISC, implemented in previous work by Lam et al.³ Lam computed 40 image features using three texture models: Haralick co-occurrence matrices, Gabor filters, and Markov random fields to construct a CBIR system for pulmonary nodules. Furthermore, Lam found that image features generated from Gabor texture model produced the best retrieval results.

However, content-based similarity obtained from computer algorithms does not necessarily correspond to a human perception of image similarity. In order to bridge this gap, Jabon et al.⁴ expanded the BRISC project by taking into account both content and semantic based features. Using the National Institutes of Health (NIH) Lung Image Database Consortium (LIDC), Jabon compared the content-based retrieval using 64 image features and the semantic-based retrieval using four radiologists' ratings on seven nodule characteristics. Jabon discovered that a substantial number of nodules recognized as similar semantically were also similar based on image features.

In this paper, we extend the work by Jabon by constructing a linear model between the two types of retrieval systems, content-based and semantic-based, for a certain number of LIDC lung nodule pairs. By establishing the linear relationship, we aim to predict the semantic similarity given only content-based features.

Further author information: (Send correspondence to Grace Dasovich)

Grace Dasovich.: E-mail: gracedasovich2010@u.northwestern.edu, Telephone: 651 338 4981

Robert Kim: E-mail: robert.kim@jhu.edu, Telephone: 805 220 8183

2. RELATED WORKS

Numerous Computer-Aided Diagnosis (CADx) systems based on low-level image features and lung nodule characteristics (such as nodule size, shape, calcification, and internal structure) have been developed to characterize lung nodules as benign or malignant. One such system was built by Armato et al.⁵ who set up an automated classification based on a linear discriminant analysis (LDA) to differentiate malignant and benign lung nodules in low-dose computed tomography (CT). In their study, the features shape characteristics of lung nodules were merged through a linear discriminant classifier to classify the nodules. Although these studies illustrate that low-level image features can be used to distinguish between malignant and benign nodules, it is important to incorporate radiologists' knowledge into the process and to understand the relationship between the image features and radiologists' annotations. Such understanding can not only improve diagnosis of malignant lung nodules, but also simplify and accelerate the radiology interpretation process as suggested by Kahn et al.⁶

In the medical imaging area, efforts to find the relationship between image features and subjective or semantic ratings were spearheaded by Barb et al.⁷ and Raicu et al.⁸ Barb developed a framework that manages visual content of lung pathologies. The framework named Evolutionary System for Semantic Exchange of Information in Collaborative Environments (ESSENCE) uses semantic methods to describe visual abnormalities and exchange knowledge in the medical domain. The framework largely consists of a semantic domain, a feature domain, and a preference domain. The semantic domain contains semantic ratings assigned by users (physicians), a feature domain stores image features extracted by feature extraction algorithms, and the preference domain contains user preferences. Fuzzy logic techniques were employed to map from low-level image features to high-level semantic terms and to retrieve images using both computed image features and physician-defined semantics. More recently, Raicu et al. developed two semi-supervised methods to predict semantic ratings for lung nodules from the Lung Image Database Consortium (LIDC) given low-level image features. Using an ensemble of classifiers from DECORATE⁹ and decision trees, they were able to improve the accuracy prediction of semantic ratings by 50 % on average despite the variability in the radiologists' interpretation.

While the above work focuses on the relationship between the image features and the semantic characteristics in terms of their prediction power (for example how well image features can be used to predict the spiculation perception of a lung nodule in Raicu et al.), there is some work that focuses on finding the same relationships but with respect to how closely the image features capture the human perception of similarity. For mammographic masses, much work has been done to establish the relationship between image features and radiologists' similarity perception. Muramatsu et al.¹⁰ used an ANN to find the relationship between the image features and the radiologists' ratings on mammograms. The subjective similarity ratings for 300 pairs of images with clustered microcalcifications were obtained from ten radiologists and the average values of these ratings were used as teaching data for a three-layer feed-forward ANN with a backpropagation algorithm. Seven image features were used as inputs and the ANN was trained to predict the semantic similarity called a psychophysical similarity measure. The correlation coefficient between the radiologists' ratings and the psychophysical similarity measure was 0.71.

More recently, preliminary work by Muramatsu et al.¹¹ links image features to the Breast Imaging Reporting and Data System (BI-RADS). Using an artificial neural network (ANN), Muramatsu et al. determined similarity measures between subjective features (BI-RADS descriptors assigned by radiologists) and objective features (computed image features) for pairs of breast masses. The ANN was trained with average ratings by 10 breast radiologists as teaching data and the BI-RADS lesion descriptors or image features as input data. Several feature combinations were tested, and the leave-one-out method was used to test the ANN. Muramatsu et al. found that when the BI-RADS descriptors were used as input data for the ANN, the correlation coefficient was decent. However, when the combinations of image features and the BI-RADS descriptors were used as input, the correlation coefficients were relatively high.

For pulmonary nodules, there is not much work done to investigate the correlation between the computer similarity results and the radiologists' similarity perception. Two notable works in this area are presented by Samala et al.¹² and Li et al.²

Samala used nonparametric correlation coefficients, multiple regression analysis, principal-component analysis, and artificial neural network analysis to investigate the optimum selection of image features. They used 42

cases (28 lung nodules and 14 non-nodules) from LIDC for the feature characterization, and a total of 11 features were computed. Correlation analysis and multiple regression analysis were used to find the relationship between radiologists’ ratings and the computed features, and a three-layer feed-forward neural network was used to classify the abnormal and normal lung nodules. The correlation coefficients between the radiologists’ annotations on 9 characteristics and 11 computed image features range from 0.2693 to 0.5178. Using the multiple regression analysis, Samala et al. discovered that the higher number of features, the higher squared multiple correlation coefficients (R^2).

Li computed similarity measures using four different techniques: feature-based, pixel-value-difference-based, cross-correlation-based, and ANN-based techniques. They discovered that the Artificial Neural Network (ANN) technique gave the highest correlation, 0.72. The input layer takes seven content-based image features, while the semantic similarity ratings by ten radiologists were used as a teaching signal. 240 pairs of lung nodules were used for the ANN, and the leave-one-out method was applied to verify the effectiveness of the ANN. The proposed approach is most similar to Li’s work, however, while Li investigates the absolute similarity for nodule images, we propose to investigate a relative similarity with respect to semantic concepts that are used by LIDC radiologists to interpret lung nodules in the process of diagnosis.

3. MATERIALS AND METHODS

3.1 Semantic Ratings and Low-Level Image Features

The data in this research comes from the Lung Image Database Consortium (LIDC). The original database contains 85 CT scans and 1989 images (each nodule having multiple slices) where up to four radiologists delineated the boundary of lung nodules with sizes between 3 mm and 3 cm and rated nine characteristics for each identified nodule (calcification, internal structure, lobulation, malignancy, margin, sphericity, spiculation, subtlety, and texture).¹³ See Appendix A for a description of each characteristic. Jabon⁴ reduced the data set containing 1989 images to 149 images by selecting only one instance per nodule with the greatest nodule area. Hence, each of the 149 images represents a distinct lung nodule.

Only seven of the characteristics were used because two of them, calcification and internal structure, had very little variation. The radiologists’ ratings were each confined into one value by taking the mode. When a unique mode did not exist, the median was found and rounded down to the nearest whole number.⁴ Previous work extracted 64 features based on four categories: texture (Gabor, Markov Random Fields, and Haralick Co-Occurrence), size, shape, and intensity³ shown in Table 1. Each feature for the 149 nodules was normalized using the Min-Max method (values range from 0 to 1) and the Z-Score method (the mean and the standard deviation of values are 0 and 1) separately to evaluate the effectiveness of each.

Table 1. All 64 computed image features

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	
RadialDistanceSD		SDIntensityBG	
		IntensityDifference	
			5 Markov features

3.2 Similarity Measures

For the multivariate linear regression model, two types of similarity measures were employed: content-based and semantic-based. To determine the content-based similarity value for a pair of nodules, three similarity measures

for the content-based features were used: Canberra,¹⁴ Manhattan, and Euclidean; the Euclidean distance gave the best correlation, hence this measure was used.

$$S_E(m, n) = \sqrt{\sum_{i=1}^{64} (f_i^m - f_i^n)^2} \quad (1)$$

$S_E(m, n)$ is the Euclidean distance between two nodules m and n using 64 content features f_i ($i = 1, 2, \dots, 64$). To determine the semantic-based similarity, a slightly modified cosine similarity was used. The semantic values exhibit a ceiling effect with some distance measures. The cosine similarity subtracted from 1 was used to minimize this effect on the ratings. The semantic similarity value of 0 for a pair of nodules represents an absolute similarity, while 1 indicates no similarity between the two nodules. Multiple radiologists' ratings for a nodule were collapsed to one rating as described in the previous section.

$$S_C(m, n) = 1 - \frac{\sum_{i=1}^7 c_i^m \cdot c_i^n}{\sqrt{\sum_{i=1}^7 (c_i^m)^2} \cdot \sqrt{\sum_{i=1}^7 (c_i^n)^2}} \quad (2)$$

$S_C(m, n)$ refers to the cosine difference between two nodules m and n using seven semantic features c_i ($i = 1, 2, \dots, 7$).

3.3 Pair Selection

Comparing the cosine and Euclidean values for all combinations of the 149 images (11,026 pairs total) gives a correlation coefficient of 0.11. It is hard to construct a regression model with this poor of a correlation, so a method of pair selection to obtain pairs that had a good correlation and were similar to one another was necessary. First, the k most similar images for each image using both semantic and content based image retrieval were retrieved. Then only the images with greater than or equal to m matches between both semantic and content similarities were found, and the correlation coefficient of these pairs was computed. The k was varied from 20 to 60, while m was changed from 1 to the greatest number of matches. k and m values where the correlation was the highest were selected.

3.4 Multivariate Regression Model

On the selected pairs (n) we performed a multivariate regression:

$$\begin{pmatrix} S_C(1, 2) \\ \vdots \\ S_C(n-1, n) \end{pmatrix} = \begin{pmatrix} d_1(1, 2) & d_2(1, 2) & \cdots & d_p(1, 2) \\ \vdots & \vdots & \ddots & \vdots \\ d_1(n-1, n) & d_2(n-1, n) & \cdots & d_p(n-1, n) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (3)$$

$S_C(i, j)$ is the cosine similarity for semantic characteristics for i and j nodules. $d_n(i, j)$ refers to the difference in the n^{th} feature between i and j nodules. While the regression model can be implemented using each one of the features per nodule as an input variable for the regression model, we propose to work with the differences such that the number of features is not artificially doubled when predicting the nodule pairs' similarity. Three different measures to encode the differences in feature values, $d_n(i, j)$ for $n = 1, 2, \dots, 63, 64$, were applied separately and compared:

$$d_n(i, j) = |f_n^i - f_n^j| \quad (4)$$

$$d_n^2(i, j) = (f_n^i - f_n^j)^2 \quad (5)$$

$$e^{d_n(i, j)} = e^{|f_n^i - f_n^j|} \quad (6)$$

$|f_n^i - f_n^j|$ refers to the distance between i and j nodules for n^{th} feature calculated as the absolute difference. β and ε are the correlation coefficients and the residuals, respectively. The coefficient of determination, R^2 , measures how well the regression model fits the data. This was implemented using MATLAB's `regress` function.

4. RESULTS

4.1 Pair Selection

The general trend for pair selection was an increase in correlation with increasing m until a certain drop off point, as demonstrated in Figure 1. The highest correlating group for each method of normalization was used. Using the Min-Max technique, the highest correlating pairs were obtained with $k = 30$ and $m > 15$, with a correlation coefficient of 0.401 and 104 pairs. Using the Z-Score normalization methods we obtained the highest correlation coefficient ($r = 0.699$) with $k = 60$ and $m > 38$, with 116 pairs consisting of 79 distinct nodules. The results are shown in Table 2. The Z-Score normalized pairs were chosen for further examination, since those pairs have higher correlation coefficient than the pairs obtained from the Min-Max method. Figure 2 shows examples of both semantic and content based image retrieval using an LIDC DICOM Analyzer created by Rick Brock.

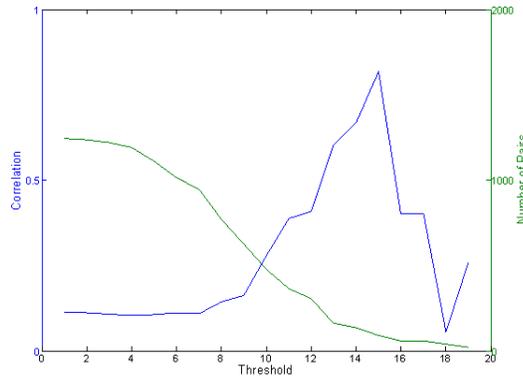


Figure 1. Correlation between image pairs for $k = 35$ while varying m and the number of pairs. The trend follows that correlation increases with increasing m until a point when too many pairs are lost.

Table 2. Values of k , m , correlation coefficient, and number of pairs obtained using the image features unnormalized, Min-Max normalized, and Z-Score normalized

	Unnormalized	Min-Max	Z-Score
k	50	30	60
m	27	15	38
Correlation	0.7870	0.4010	0.6990
Number of Pairs	112	104	116

4.2 Multivariate Linear Regression Analysis

The regression analysis returned correlation coefficients (β) for all 64 feature differences and the coefficient of determination, R^2 . Table 3 shows R^2 values for three types of feature differences, (5)-(7). Since the values for R^2 are very similar there is no significant difference in the way we choose to encode the differences in the individual features. We chose to implement our regression model using the absolute difference, $d(i, j)$. Figure 3 shows an example of retrieval using the predicted similarity with our model and the corresponding cosine values.

Table 3. The R^2 values obtained from the multivariate linear regression using the three difference measures

	$d(i, j)$	$d^2(i, j)$	$\exp(d(i, j))$
R^2	0.871	0.849	0.866

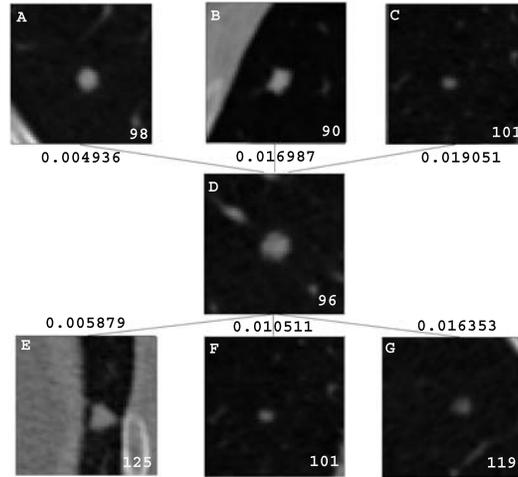


Figure 2. Nodule numbers are shown in the lower right corner of the images. Images D and K are query images, A, B, C, H, I, and J are each images three most similar images retrieved from SBIR (based on the cosine values), while E, F, G, L, M, and N are three most similar images retrieved from CBIR (based on the Euclidean distance values). Note that nodules 46 and 68 are retrieved by both CBIR and SBIR.

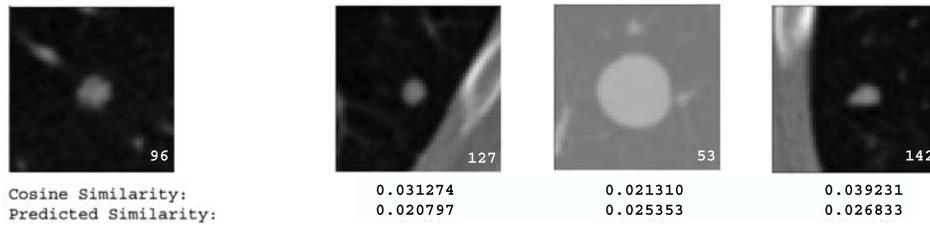


Figure 3. Example of cosine and predicted similarity for linear regression using nodules within the 116 selected pairs

The selected pairs ($n = 116$) had an R^2 of 0.871. We performed a leave-one-out validation. In this validation, one pair from the 116 pairs was left out and the multivariate linear regression was performed for the remaining 115 pairs, and this process was repeated for the rest of the pairs one by one. The range and the average of the R^2 values were (0.864, 0.886) and 0.872, respectively. The mapping of content similarity versus semantic similarity for these 116 pairs is shown in Figure 4 along with all 11,026 pairs. The coefficients from the regression determine which features are most important in differentiating the nodules we selected. Table 4 shows the most important ten features: two out of seven size features (29 %), six out of 40 texture features (15 %), and two out of nine intensity features (22 %) appear to be the most important. The coefficients for these ten features were significantly higher than the rest; the tenth coefficient was larger than the eleventh coefficient by 0.002. Thus the ten features were investigated. Nodule size (equivalent diameter) has the highest coefficient, and is therefore the most important feature for the determination of similarity measure. This is also supported by Li's work² which reports that effective diameter has the highest correlation coefficients among the features they used.

Performing the regression using these ten features gives an R^2 of only 0.665. Eleven of the least correlating features can be removed from the regression before the R^2 drops to 0.864; less than eleven removed maintains an R^2 close to 0.87. These features include seven texture features (one Markov, four Gabor, and two Haralick) three shape features (Solidity, Eccentricity, and Radial Distance STD) and one intensity feature (Intensity Difference).

Furthermore, we analyzed the two groups of nodules (79 nodules corresponding to the selected pairs and 70 nodules corresponding to the unselected pairs) based on the top ten correlating features and the seven semantic characteristics to understand the differences in these two groups. Figure 5 shows histograms for the two image features and Figure 6 shows the two characteristics that showed the most significant difference between the two groups of nodules.

Table 4. Top Ten Most Important Features for Z-score normalization

Correlation Coefficients	Features
0.1175	Equivalent Diameter
0.1085	Energy (Haralick)
0.0823	Gabor Mean 135° 0.5 Hz
0.0647	Convex Area
0.0467	Gabor Standard Dev. 135° 0.4 Hz
0.0322	Min Intensity Background
0.0295	Markov 4
0.0280	Variance (Haralick)
0.0265	Gabor Standard Dev. 45° 0.5 Hz
0.0238	Standard Dev. Intensity

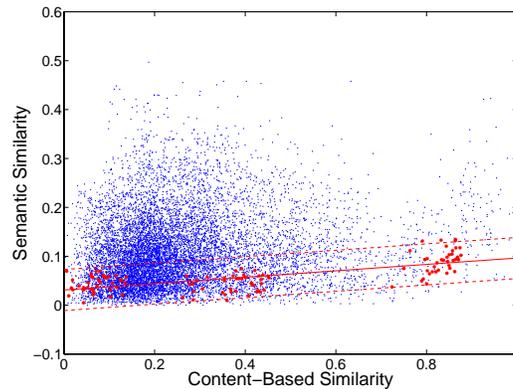


Figure 4. Mapping of content-based similarity and semantic-based similarity. Selected pairs are shown in red with all 11,026 pairs in blue. The 95% confidence intervals for the regression line are shown. The R^2 value is 0.871.

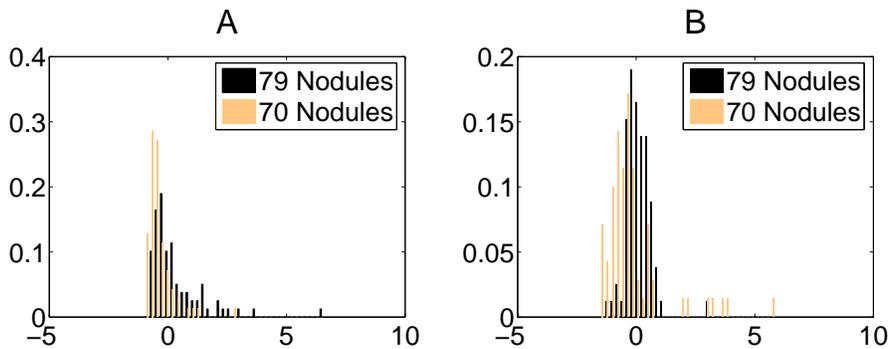


Figure 5. The two most important content-based features (z-score normalized) are **A** Equivalent Diameter and **B** Standard Deviation of Intensity. The frequencies were normalized as percentages.

Note that semantic ratings are discrete values while feature data is not. The selected group of nodules show a significantly larger diameter. The chosen nodules have diameters between 4.5 mm and 45.3 mm, with a mean of 10.4 mm and standard deviation of 6.5 mm. The nodules not chosen were between 3.0 mm and 24.3 mm, with a mean of 6.5 mm and standard deviation of 3.3 mm. The selected nodules also showed greater standard deviation of intensity. They also have more suspicions of malignancy, are more often rated as obvious (in the category of subtlety), more marked lobulation, and more marked spiculation than the unselected nodules.

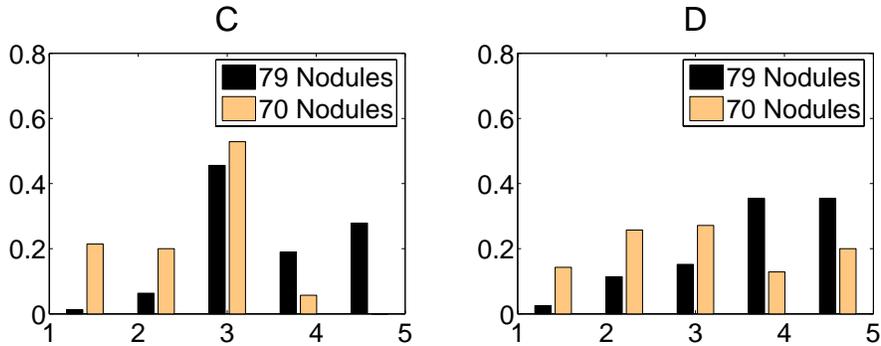


Figure 6. The two most significant semantic characteristics are **C** Malignancy and **D** Subtlety. The frequencies were normalized as percentages.

5. CONCLUSION

We were successful in intelligently selecting well-correlated images, implementing a linear regression model, and defining characteristics and features of nodules that will work with our model. This model was able to successfully interpret semantic similarity between two images based on the similarity of content features, with an R^2 of 0.871 for a specific set of images.

The challenge now will be to improve the accuracy of the model while extending it to a broader category of images. Our work thus far has been under the assumption that the relationship between content similarity and semantic similarity is a linear association, which has worked for a subset of data. To be practically implemented in medicine such a model must work for all images. Now we plan to investigate machine learning approaches, specifically an artificial neural network, to incorporate more images while maintaining a good accuracy for our prediction model. Also, we are planning to implement our system on the new LIDC data containing 929 distinct nodules.

A challenge for the proposed approach is the variability among radiologists. Since ratings from multiple radiologists were aggregated into one value, the lack of consensus among radiologists is hidden and therefore not taken into account when predicting the conceptual-based similarities. After creating a suitable model with our current data, we expect to take into account all four radiologists' ratings by using a probabilistic model, such as Jeffrey divergence and Kullback-Leibler divergence, which theoretically would increase the accuracy.

APPENDIX A. ALL 9 SEMANTIC CHARACTERISTICS AND POSSIBLE RATINGS

Characteristic	Description	Possible Ratings
Calcification	Calcification appearance in the nodule	<ol style="list-style-type: none"> 1. Popcorn 2. Laminated 3. Solid 4. Non-central 5. Central 6. Absent
Internal Structure	Expected internal composition of the nodule	<ol style="list-style-type: none"> 1. Soft Tissue 2. Fluid 3. Fat 4. Air
Lobulation	Whether lobular shape is apparent from margin or not	<ol style="list-style-type: none"> 1. Marked 2. . 3. . 4. . 5. None
Malignancy	Likelihood of malignancy	<ol style="list-style-type: none"> 1. Highly Unlikely 2. Moderately Unlikely 3. Indeterminate 4. Moderately Suspicious 5. Highly Suspicious
Margin	How well defined the margins are	<ol style="list-style-type: none"> 1. Poorly Defined 2. . 3. . 4. . 5. Sharp
Sphericity	Dimensional shape in terms of roundness	<ol style="list-style-type: none"> 1. Linear 2. . 3. Ovoid 4. . 5. Round
Spiculation	Degree of exhibition of spicules	<ol style="list-style-type: none"> 1. Marked 2. . 3. . 4. . 5. None
Subtlety	Contrast between nodule and surroundings	<ol style="list-style-type: none"> 1. Extremely Subtle 2. Moderately Subtle 3. Fairly Subtle 4. Moderately Obvious 5. Obvious
Texture	Internal density of nodule	<ol style="list-style-type: none"> 1. Non-Solid 2. . 3. Part Solid (Mixed) 4. . 5. Solid

REFERENCES

- [1] Landis, S. H., Murray, T., Bolden, S., and Wingo, P. A., "Cancer statistics 2000," *Ca-Cancer J. Clin.* **50**, 7–33 (2000).
- [2] Li, Q., Li, F., Shiraishi, J., Katsuragawa, S., Sone, S., and 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**, 2584–2593 (2003).
- [3] Lam, M., Disney, T., Pham, M., Raicu, D., and Furst, J., "Content-based image retrieval for pulmonary computed tomography nodule images," *SPIE Medical Imaging Conference* (February 2007).
- [4] Jabon, S. A., Raicu, D. S., and Furst, J. D., "Content-based versus semantic-based similarity retrieval: a LIDC case study," *SPIE Medical Imaging Conference* (February 2009).
- [5] Armato, S. G. I., Altman, M. B., Wilkie, J., Sone, S., Li, F., Doi, K., and Roy, A. S., "Automated lung nodule classification following automated nodule detection on ct: A serial approach.," *Medical Physics* **30**, 1188–1197 (2003).
- [6] Kahn, C., Channin, D., and Rubin, D., "An ontology for PACS integration," *Journal of Digital Imaging* **19**(4), 316–327 (2006).
- [7] Barb, A. S., Shyu, C., and Sethi, Y. P., "Knowledge representation and sharing using visual semantic modeling for diagnostic medical image databases," *IEEE Transaction Information Technology in Biomedicine* **9** (2005).
- [8] Raicu, D., Zinovev, D., Furst, J. D., and Varutbangkul, E., "Semi-supervised learning approaches for predicting lung nodules semantic characteristics," *Intelligent Decision Technologies Journal* **3**(2) (2009).
- [9] Melville, P. and Mooney, R., "Constructing diverse classifier ensembles using artificial training examples," *Proc of 18th Intl. Joint Conf. on Artificial Intelligence* , 505–510 (2003).
- [10] Muramatsu, C., Li, Q., Schmidt, R., Shiraishi, J., and Doi, K., "Investigation of psychophysical similarity measures for selection of similar images in the diagnosis of clustered microcalcifications on mammograms," *Med. Phys.* **35**, 5695–5702 (2008).
- [11] Muramatsu, C., Li, Q., Schmidt, R., Shiraishi, J., and Doi, K., "Determination of similarity measures for pairs of mass lesions on mammograms by use of bi-rads lesion descriptors and image features," *Acad. Radiol.* **16**, 443–449 (2009).
- [12] Samala, R., Moreno, W., You, Y., and Qian, W., "A novel approach to nodule feature optimization on thin section thoracic ct," *Academic Radiology* **16**, 418–427 (2009).
- [13] Armato, S. G., McLennan, G., McNitt-Gray, M. F., Meyer, C. R., Yankelevitz, D., and et al., "Lung image database consortium: Developing a resource for the medical imaging research community," *Radiology* **232**, 739–748 (2004).
- [14] Bugatti, P. H., Traina, A., and Traina, C., "Assessing the best integration between distance-function and image-feature to answer similarity queries," *ACM Symposium on Applied Computing* (March 2008).