

Predicting LIDC Diagnostic Characteristics by Combining Spatial and Diagnostic Opinions

William H. Horsthemke^{*}, Daniela S. Raicu, Jacob D. Furst
School of Computing, College of Computing and Digital Media,
DePaul University, Chicago, IL 60604, USA

ABSTRACT

Computer-aided diagnostic characterization (CADc) aims to support medical imaging decision making by objectively rating the radiologists' subjective, perceptual opinions of visual diagnostic characteristics of suspicious lesions. This research uses the publicly available Lung Image Database Consortium (LIDC) collection of radiologists' outlines of nodules and ratings of boundary and shape characteristics: spiculation, margin, lobulation, and sphericity. The approach attempts to reduce the observed disagreement between radiologists on the extent of nodules by combining their spatial opinion using probability maps to create regions of interest (ROIs). From these ROIs, image features are extracted and combined using machine learning models to predict a combined opinion, the median rating and a thresholded, binary version of their diagnostic characteristics. The results show slight to fair agreement—linear-weighted Kappa—between the CADc models and median radiologist opinion for the full scale five-level rating and fair to moderate agreement using a binary version of the median radiologist opinion.

Keywords: Characterization, Image Analysis, Feature Extraction, Region of Interest, Computer aided diagnosis

1. INTRODUCTION

Computer-aided decision support in medical imaging has focused primarily on the challenging problems of detecting and diagnosing suspicious lesions such as pulmonary nodules, which are often missed or misinterpreted by radiologists. Although detection (CADe) and diagnosis (CADx) offer valuable diagnostic information about the presence or absence of suspicious lesions and perhaps probabilities about the likelihood of malignancy, these CAD(x) systems rarely describe the lesion or offer additional information to support the radiologist in making their decision (Doi 2005). Towards describing medically meaningful diagnostic features to support the CADx decisions, this paper continues work on characterizing pulmonary nodules based upon the expert opinion of pulmonary radiologists to create an intermediate step between CADe and CADx for computer-aided diagnostic characterization (CADc).

This research examines several probabilistic regions of interest to determine which best represents the spatial location and extent of the nodule for extracting pixel-based features to predict a composite radiologist opinion on shape and boundary-based diagnostic characteristics: lobulation, sphericity, margin, and spiculation. The CADc ratings of these medically-meaningful image features can be used as evidence within the CAD(x) process or serve to annotate patient/nodule cases for computer- or semantic-based image retrieval (Wiemker, et al. 2008) (Horsthemke, Raicu and Furst 2009).

The seminal work on predicting radiologists' perception of diagnostic characteristics was performed by Nakamura where radiologists rated characteristics such as shape, margin irregularity, spiculation, lobulation, etc. on a scale of 1 to 5 (Nakamura, Yoshida and Engelmann 2000). Nakamura then extracted various statistical and geometric image features including Fourier and radial gradient indices and correlated these with the radiologists' ratings. They showed correlation between radial gradient indices with spiculation and other geometric features with shape, but concluded that there was poor predictive performance in predicting the radiologists' ratings due to the variability between radiologists.

^{*} Horsthemke@acm.org

Nakamura also described a two-stage CADx process where the first stage consists of radiologists rating nodule characteristics and the second stage applies those ratings towards predicting the radiologists' diagnoses. They reported poorer performance for predicting radiologists' own diagnosis from their ratings of diagnostic characteristics than from using raw image features to predict the radiologists' diagnoses and concluded that the variability of radiologists on diagnostic characteristics presents a larger challenge than the design and selection of image features for use in CADx.

Variability among radiologists on reporting diagnostic characteristics was studied by Burns et al. who demonstrated a lack of consistency in reporting pulmonary nodule characteristics and recommended standardizing the reporting criteria. Efforts to standardize reporting on pulmonary nodules are underway as part of the development of a general purpose radiology lexicon (Langlotz 2006).

The mammography community has standardized reporting based upon the BI-RADS assessment categories. Recent analysis indicates the effectiveness of this approach. Lazarus evaluated inter-observer variability in BI-RADS reporting and concluded that radiologists showed good agreement, the ratings had a high predictive value, and the results validated the use of the US BI-RADS lexicon (Lazarus, et al. 2006). These categories formed the target classification ground truth in a CADc study attempting to measure image features for classifying the BI-RADS categories of spiculation and circumscribed margin of breast masses (Sahiner, et al. 2008). In their successful study, Sahiner et al. demonstrated the promise of rating diagnostic characteristics when radiologists consistently rate diagnostic characteristics.

Initial work towards predicting radiologists' ratings on Lung Image Database Consortium (LIDC) characteristics by Raicu et al. applied an extended set of geometric features for roughness, eccentricity, solidity, extent, and radial standard deviation (Raicu, et al. 2007) (Varutbangkul, Raicu and Furst 2007). Their work demonstrated the challenge for predicting individual radiologists' ratings of LIDC diagnostic characteristics. Follow-up work extended the shape-based analysis of radiologist-drawn nodule outlines and applied Fourier-based shape descriptors and a variant of the radial gradient variant approach applied to outlines versus image gradients, the radial normal index, but presented no significant improvement in predicting individual radiologists' ratings of diagnostics characteristics due to the extent of their disagreement on ratings (Horsthemke, Raicu and Furst 2009).

Weimker et al. demonstrated the robustness of shape index features for measuring nodule spiculation (Wiemker, et al. 2008), then showed good correlation between shape index features and radiologists ratings spiculation using custom nodule segmentations of the Image Database Resource Initiative (IDRI), not yet publicly available (Wiemker, et al. 2009).

The shape of pulmonary nodules is used by radiologists in clinical practice and CADx to discriminate between benign and malignant diagnoses (Wormanns and Diederich 2004) (Armato III, et al. 2003). In selecting diagnostic characteristics for obtaining expert radiologists' ratings, the LIDC chose three (3) characteristics which capture boundary- and shape-based visual appearances: margin, spiculation, lobulation, and sphericity (Armato, et al. 2004). Early research on lung nodule detection and diagnosis selected several shape-based feature extraction methods. Giger et al. computed geometric features (effective diameter and degree of circularity) to detect suspicious nodules in chest x-rays (Giger, et al. 1990). Nakamura et al. selected low order features of the Fourier transformation of the nodule outline and applied the radial gradient index approach to measure the spiculation of the nodule (Huo, et al. 1995) (Nakamura, Yoshida and Engelmann 2000).

The prediction of the ratings of Lung Image Database Consortium (LIDC) diagnostic characteristics given by individual radiologists has shown limited success, mainly due to the disagreement among radiologists (Armato III, et al. 2009) (Raicu, et al. 2007) (Wiemker, et al. 2009). The research presented in this paper explores methods to reduce the variability between radiologists by combining both their ratings and outlines of the nodules and predicting a composite rating.

2. DATA

The Lung Image Database Consortium (LIDC) dataset serves as the source of images and radiologists' opinions (Armato, et al. 2004). The LIDC has developed a lung nodule collection and reporting protocol for four (4) radiologists to detect suspicious lesions between 3 and 30mm in diameter in thoracic CT scans. When radiologists identify a nodule, they draw an outline around the nodule and rate diagnostic characteristics on an ordinal scale of 1 – 5. The diagnostic

characteristics include texture, subtlety, spiculation, sphericity, margin, malignancy, lobulation, internal structure, and calcification. The LIDC ratings system offers labels for some ratings but not all, such as labeling the spiculation rating of 1 as "marked" and 5 as "none" while the 2-4 ratings are unlabelled and some characteristics have more labels. The LIDC protocol does not enforce consensus among the radiologists for detection, outlines, or ratings of nodules, thus each nodule may be marked by only one (1) or up to four (4) radiologists.

At the time of this study, the LIDC database contained 400 patient cases including 85 cases from an earlier release. This study considers only the most recently available 315 cases due to collection problems for the ratings of some characteristics in the earlier release. The database used in this study contained 832 nodules rated by at least one (1) radiologist.

3. METHODOLOGY

This research examines several probabilistic regions of interest to determine how well they represent the spatial location and extent of the nodule for extracting pixel-based features to predict a composite radiologist opinion on shape and boundary-based diagnostic characteristics: lobulation, sphericity, margin, and spiculation. The overall approach consists of three (3) major steps: 1) creation of regions of interest, 2) extraction of pixel-based image features, and 3) prediction of composite radiologist opinion on each diagnostic characteristic—using the median rating as well a binary version of the median.

3.1 Regions of Interest

The regions of interest are created from the largest representative slice of the nodule. The largest representative slice contains the largest number of agreed-upon pixels, the largest intersection of all radiologists' outlines—the slice with the largest TPM 100%. From the largest representative slice of the nodule, the radiologist-drawn outlines are combined using the probability map (p-map) approach described by Meyer et al. (Meyer, et al. 2006) then thresholded to create four probabilistic regions of interest, thresholded p-maps (TPMs): (25%--union, 50%, 75%, and 100%--intersection).

The p-map method considers each pixel within a radiologist's outline as a vote for including that pixel in the nodule, as illustrated in Figure 1. Included sets of pixels are accumulated and divided by the number of readers to create the nodule p-map representing the probability that any pixel is a member of the nodule. Using the nodule p-map, a set of ROIs are created using a threshold for membership of the pixel in the nodule as illustrated in Figure 1. For example, a 50% threshold p-map (TPM 50%) will include all pixels selected by at least 50% of the radiologists and contain pixels with p-map values of 50, 66, 75, and 100%. For this research, the TPM thresholds under study are 25%, 50%, 75%, and 100% where TPM 25% represents a logical OR, or UNION, of all pixels-where each pixel selected by at least one radiologist. The TPM 100% represents the logical AND, or INTERSECTION, of all pixels-where each pixel was selected by all radiologists. Example ROIs formed by thresholded p-maps are illustrated on an example nodule outlined by 4 radiologists in Figure 2, using the proportion notation where TPM 25% is $TPM_{0.25}$ and TPM 100 is $TPM_{1.0}$.

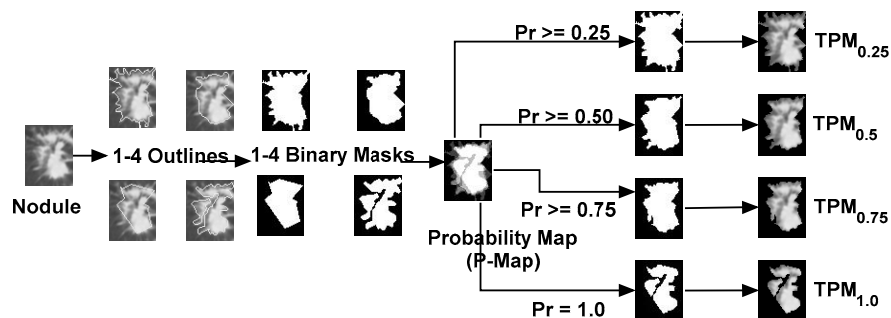


Figure 1. The Probability Map (p-map) is created by adding up all the pixel sets selected within the outlines of one or more radiologists, then dividing the accumulated count by the number of radiologists who rated the nodule, ranging from 1 to 4 in the LIDC. The illustrated p-map has 4 readers and pixel values of {0.0, 0.25, 0.50, 0.75, 1.0} representing the proportion of the four (4)

radiologists who selected that pixel. Thresholded p-maps are created by selecting pixels from the original DICOM image if at least a fixed proportion, a threshold, of radiologists included that pixel in their outlines.

In addition to the four (4) TPMs, a boundary region of interest is formed as the subtraction of the intersection from the union of all selected pixels (Horsthemke, Raicu and Furst 2009). Two other ROIs are derived from the union (TPM 25%) of selected pixels; one consists of a dilated version of the union and the other an enclosed circle around the union.

The boundary ROI is formed by removing the "interior"- intersection (TPM 100%)-of the nodule from the "exterior"—union (TPM 25%)—of the nodule, leaving only the boundary of the nodule. The boundary method uses morphological processing to extend the coverage (close intersection and dilate union) of the resulting boundary region and is illustrated in Figure 2.

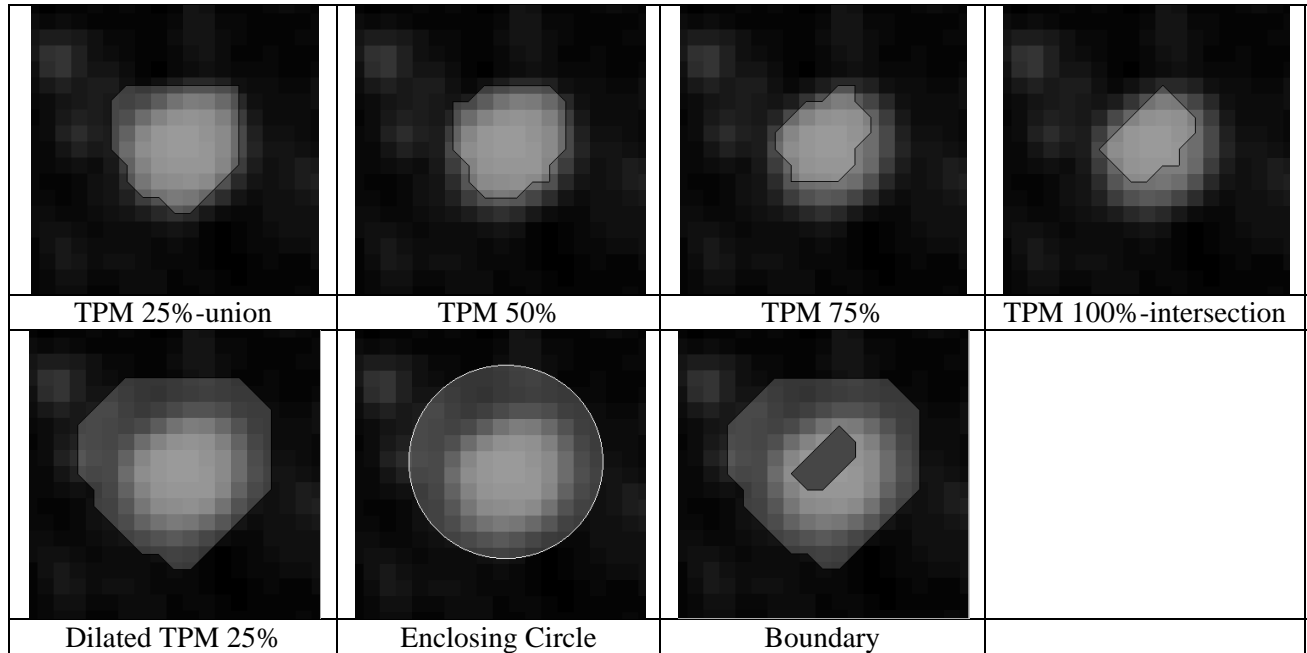


Figure 2. Probabilistic regions of interest formed from thresholded p-maps, including a dilated version of the union, an bounding circle, enclosing the union, and boundary method, formed from the removal of the intersection from the union of all pixels.

3.2 Image Features

Pixel-level image features are extracted from the set of ROIs, including radial gradient index (RGI) based upon first derivative of the image and second derivative features based upon the Hessian of the image: ShapeIndex and Curvedness (Koenderink and Van Doorn 1992). Other features include intensity and gradient entropy and Zernike moments(Chang 2005).

The radial gradient index was designed to measure spiculation along the borders of suspicious lesions in mammography by capturing the variability of the angles formed between a radial vector from the center of an object and the direction of the gradient at specific pixels (Huo, et al. 1995). The RGI algorithm adopted in this paper uses the gradient-magnitude normalized dot-product method which yields a single value for a set of pixels that ranges between -1 for a hole and +1 for a perfect circle (Kupinski, et al. 1995). Each pixel in the ROI is used for computing the RGI.

The ShapeIndex and Curvedness features were developed by Koenderink (J. Koenderink 1990) to classify the shape of objects. The ShapeIndex ranges from -1 for a cup shape, to 0 for a saddle, and to 1 for a cap. The Curvedness feature measures the magnitude of the curvedness at a point where zero (0) is flat. This methodology has been applied to reject false positives (Sahiner, et al. 2005) and has been correlated with the LIDC ratings for spiculation (Wiemker, et al. 2009).

The ShapeIndex and Curvedness features are derived from the Hessian matrix which represents the second-order partial derivative of the image. This paper uses the multi-scale version suggested by Frangi (Frangi, et al. 2001) to create the image Hessian matrix by convolving the image with Gaussian kernels of various scales. This research uses five scales with a sigma equal to 0.5, 1, 3, 5, and 7 mm. The features collected for building the prediction models include the mean, median, and standard deviation of both the ShapeIndex and Curvedness of the pixels contained in the ROI.

Entropy measures the uncertainty, disorder, or statistical randomness of an image based upon the probability density of the image and typically computed using the image histogram. Entropy is calculated as the negative sum of the product of the histogram counts multiplied by the logarithm of the histogram counts. The entropy is computed for both the intensity and the gradient of the pixels within the selected ROI.

Zernike moments offer a rotationally invariant method for capturing the shape of objects expressed as probability densities. To achieve rotational invariance, Zernike moments exploit the property that Zernike polynomials are orthogonal on the unit circle and thus invariant to rotation. To obtain scale and translation, invariant Zernike moments are normalized by dividing by the low-order regular moments. The current research uses the Zernike moment implementation by Chang (Chang 2005) which is based on the formulation by Teague (Teague 1980).

3.3 Ratings for Diagnostic Characteristics

This paper aims to predict a composite opinion from all radiologists' individual opinions by combining their ratings using techniques valid for the ordinal, non-interval, ratings system used in the LIDC. Two valid methods are median and majority (mode). The majority offers a useful interpretation as a voting method but presents challenges when the mode is undefined due to lack of agreement or multi-modal agreement. Given the problem with mode, the median of all radiologists' ratings for the nodule is used as the target prediction (category). To study the effect of labeling only two opposite ratings, a binary score for each diagnostic characteristic is created by threshold of the median rating (Petrick, et al. 2005) (Horsthemke, Raicu and Furst 2009). All four (4) possible binary thresholds for a 5-point rating scale are examined and the results of the best performing threshold are reported.

As illustrated in Figure 3, the overall methodology creates regions of interest containing pulmonary nodules, extracts image-based features, combines those features using machine learning, and predicts composite radiologists' ratings for diagnostic characteristics.



Figure 3. Illustration of overall approach to training computer-aided diagnostic characterization method.

3.4 Prediction Model

The image features are combined in a prediction/classification model using decision trees, a traditional machine learning approach, implemented as the J48 algorithm (Witten and Frank 2005). Decision trees are applicable to the ordinal (categorical) nature of the LIDC ratings collection methodology where the scores of 1-5 represent ordinal (Likert-style) ratings rather than interval or ratio data. After building the models, the evaluation uses 10-fold cross validation technique for performance analysis.

The prediction modeling adopts a classification approach using the decision tree methodology to predict both the full-scale median and binary thresholded version and reports classification performance using both accuracy and linear Kappa agreement. Accuracy is measured as the diagonal sum of the confusion/classification matrix divided by the

sum of the matrix. Linear Kappa represents agreement in excess of chance and measured using the method described by Kundel and Polansky (Kundel and Polansky 2003).

4. RESULTS

The predictive performance for full-scale ratings has accuracies between 40% and 80% (Figure 4) with Linear Kappa agreement between 0.1 and 0.3 (Figure 4), but with mixed results between the four diagnostic characteristics. The binary prediction performs better than the full-scale ratings with accuracies above 90% with Linear Kappa agreement between 0.1 and 0.4. Higher binary performance is expected, due to the lower number of prediction categories.

These results (Table 1) from the most recent 832 nodule, 325-patient LIDC dataset in 2009, show better predictive performance than previous work on either individual or combined opinions for the earlier LIDC dataset, the 147 nodule, 85-patient LIDC database (Raicu, et al. 2007) (Varutbangkul, Raicu and Furst 2007) (Horsthemke, Raicu and Furst 2009).

Table 1. The current approach has markedly better predictive performance—evaluated by accuracy and agreement—than prior work for predicting radiologists' opinions. The current research is more comparable to (Horsthemke, Raicu and Furst 2009) which also predicted combined radiologist opinion using the median, but that work considered only those nodules marked by at least 2 radiologists and reported un-weighted Kappa not linear-weighted Kappa. The work of (Varutbangkul, Raicu and Furst 2007) predicted individual radiologist opinion, thus multiple predictions per nodule, but reported only accuracy not Kappa agreement.

| | Current Results | | Horsthemke 2009 | | Varutbangkul 2007 | |
|-------------|-----------------|-------|-----------------|-------|-------------------|-------|
| | Accuracy | Kappa | Accuracy | Kappa | Accuracy | Kappa |
| Spiculation | 60% | 0.2 | 45% | 0.19 | 41% | N/A |
| Lobulation | 60% | 0.2 | 34% | 0.08 | 34% | N/A |
| Sphericity | 40% | 0.2 | 57% | 0.0 | 40% | N/A |
| Margin | 40% | 0.25 | N/A | N/A | 40% | N/A |

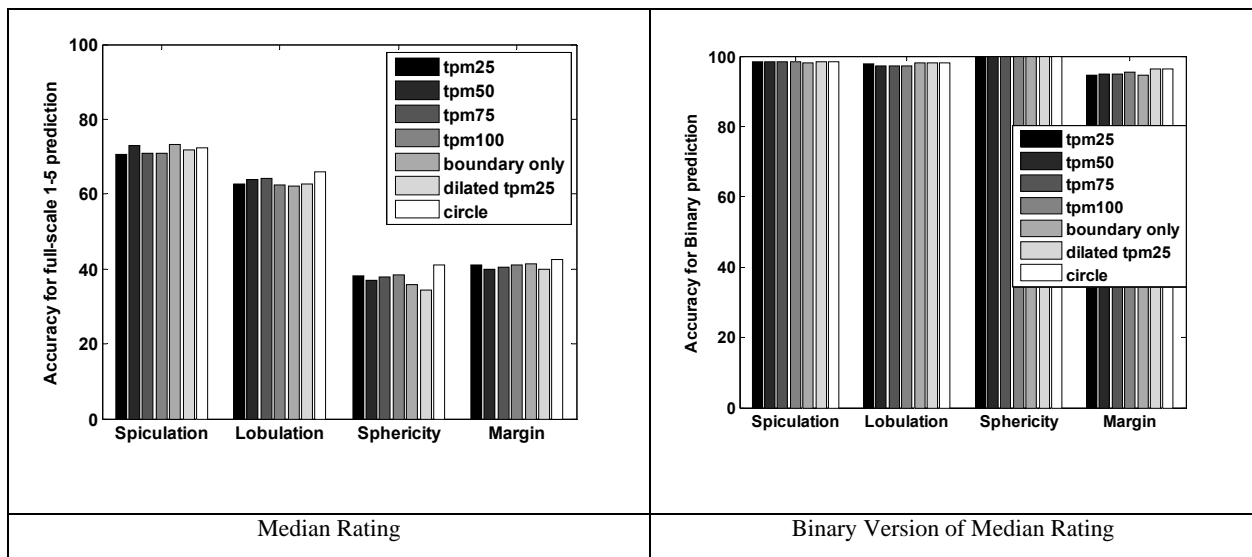


Figure 4 Accuracy for full-scale 1-5 prediction and Binary prediction across four characteristics (Spiculation, Lobulation, Sphericity, Margin) using seven different methods (tpm25, tpm50, tpm75, tpm100, boundary only, dilated tpm25, circle). The binary version consistently shows higher accuracy (above 90%) compared to the full-scale version (40-80%).

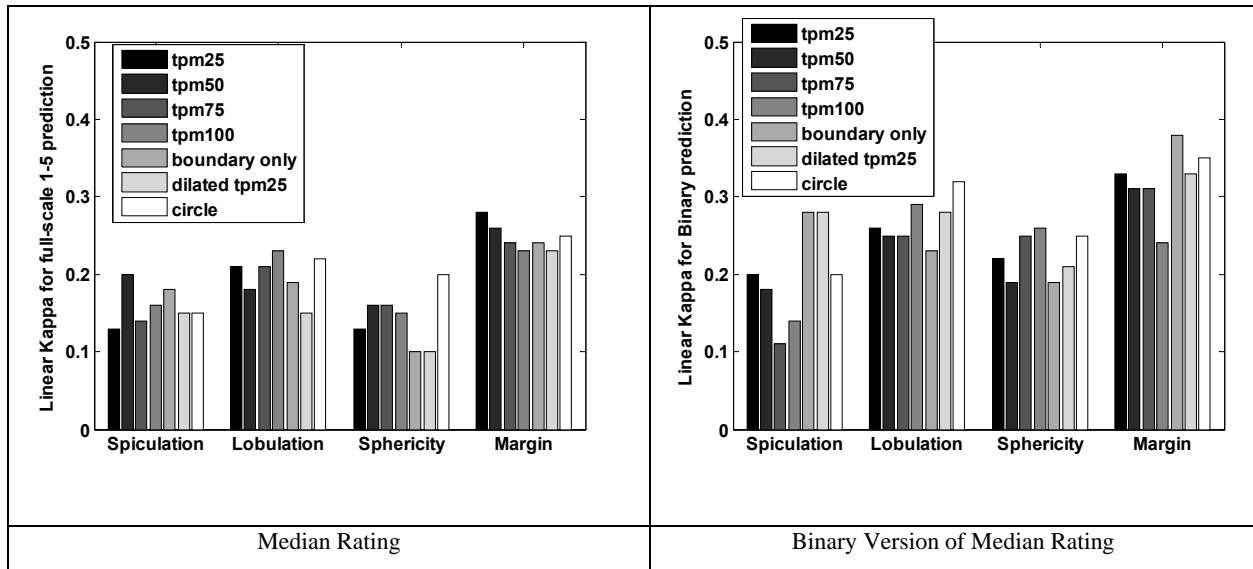


Figure 5. Agreement—measured with Linear Kappa—between CADc and radiologists for predicting full scale (1-5) median rating and binary version using different thresholded probability maps (TPMs) for creating regions of interest (ROI).

The predictive performance does not appear to depend upon the selection of the region of interest, either by varying the threshold for including pixels or selecting only the boundary pixels. The enclosing circle method for selecting a region of interest tends to outperform the other methods, although the performance difference is not substantially significant.

5. CONCLUSION

This paper considers which probabilistic, pixel-selection criteria best represents the LIDC pulmonary nodules and whether the extracted pixel-based image features sufficiently capture radiologists' perception of visual diagnostic characteristics of suspicious lesions.

From the measured image features extracted from the candidate regions of interest, the classification model predicts a combined radiologist opinion about medically meaningful diagnostic characteristic. This combined radiologist opinion is represented both in full scale using the median of the 5 ordinal categories or by a thresholded binary version representing a present or absent opinion about a feature, such as spiculated or not.

The predictive performance in terms of accuracy is good for both the 5-category median model and the thresholded binary version, but the observer agreement based upon Linear Kappa ranges from slight or poor to fair (Altman 1990).

Across diagnostic characteristics and ROIs, the predictive performance is mixed, but better than previous research on predicting both individual radiologists' ratings as well a composite approaches. The improved performance is expected in part due to the removal of patient cases which the LIDC observed as incorrectly recorded. The incorrect recording of these cases introduced inverted ratings for some diagnostic characteristics and presented additional challenges to modeling radiologist opinions on earlier research studies. The elimination of these known problem cases ensures that the datasets represent actual radiologist opinions and that the observed variability is due to opinion not data collection.

There is no consistent performance effect from varying the method for selecting the region of interest, although the enclosing circle ROI method tends to perform as well as and often better than the other methods. The enclosing circle method ROI contains the most pixels of any method since it encloses the union of all the pixels, the TPM 25%, in a

circle with an additional 1-3 mm expansion band, where the 1-3 mm band depends upon the maximum radius of the TPM 25%.

Although the performance of the enclosing circle method is not substantially significant in terms of improved predictive performance, the enclosing circle method uses less of the spatial information provided by radiologists. This method only requires a center location and nodule size to create a region of interest, suggesting that more detailed spatial information about the nodule may not be required for nodule characterization. However, the use of nodule outlines or 3D surface information is not used in the current approach but has been shown to provide useful information about the shape and margin characteristics (Wiemker, et al. 2009). To obtain the outline or 3D surface of the nodule will require segmenting the nodule and might not include or require any spatial information from a radiologist.

Future work looks toward adding more methods for feature extraction to increase the information extraction; expanding the region of interest to all slices (3D); and, when possible, applying the methods to future, expanded versions of the publicly available LIDC dataset.

REFERENCES

- [1] Altman, D. G. *Practical Statistics for Medical Research*. Chapman and Hall, 1990.
- [2] Armato III, S. G., et al. "Automated lung nodule classification following automated nodule detection on CT: A serial approach." *Medical Physics*, 2003: 1188-1197.
- [3] Armato III, S.G., et al. "Assessment of Radiologist Performance in the Detection of Lung Nodules: Dependence on the Definition of "Truth"." *Academic Radiology*, 2009: 28-38.
- [4] Armato, S. G., et al. "Lung Image Database Consortium: Developing a resource for the medical imaging research community." *Radiology*, 2004: 739-748.
- [5] Chang, K. Y. *IDEAL Pattern Recognition Toolbox*. 2005.
www.mathworks.com/matlabcentral/fileexchange/7972 .
- [6] Doi, K. "Current status and future potential of computer-aided diagnosis in medical imaging." *The British Journal of Radiology*, 2005.
- [7] Frangi, Alejandro F., Wiro J. Niessen, Paul J. Nederkoorn, Jeannette Bakker, Willem P.Th.M. Mali, and Max A. Viergever. "Quantitative analysis of vascular morphology from 3D MR angiograms: In vitro and in vivo results." *Magnetic Resonance in Medicine*, 2001: 311-322.
- [8] Giger, M. L., K. Doi, H. MacMahon, C. E. Metz, and F. F. Yin. "Pulmonary nodules: computer-aided detection in digital chest images." *RadioGraphics*, 1990: 861-865.
- [9] Horsthemke, William H., D. S. Raicu, and J. D. Furst. "Evaluation Challenges for Bridging Semantic Gap: Shape Disagreements in the LIDC." *International Journal of Healthcare Information Systems and Informatics*, 2009.
- [10] Horsthemke, William. H., D. S. Raicu, and J. D. Furst. "Characterizing pulmonary nodule shape using a boundary-region approach." *SPIE Medical Imaging*. Orlando: SPIE, 2009.
- [11] Huo, Z., et al. "Analysis of spiculation in the computerized classification of mammographic masses." *Medical Physics*, 1995.
- [12] Kawata, Y., et al. "Curvature-based characterization of shape and internal intensity structure for classification of pulmonary nodules using thin-section CT images." *SPIE Medical Imaging*. San Diego: SPIE, 1999.
- [13] Koenderink, J. J., and A. J. Van Doorn. "Surface shape and curvature scales." *Image and Vision Computing*, 1992: 557-565.
- [14] Koenderink, J. *Solid Shape*. Cambridge: The MIT Press, 1990.
- [15] Kundel, H., and M. Polansky. "Measurement of Observer Agreement." *Radiology*, 2003: 228-303.
- [16] Kupinski, M., M.L. Giger, P. Lu, and Z. Huo. "Computerized detection of mammographic lesions: Performance of artificial neural network with enhanced feature extraction." *SPIE Medical Imaging*. SPIE, 1995.
- [17] Langlotz, C. P. "RadLex: A New Method for Indexing Online Educational Materials." *RadioGraphics*, 2006: 1595-1597.
- [18] Lazarus, E., M. Mainiero, B. Schepps, S. Koelliker, and L. Livingston. "BI-RADS Lexicon for US and Mammography: Interobserver Variability and Positive Predictive Value." *Radiology*, 2006: 385-391.
- [19] Meyer, C. R., et al. "Evaluation of lung MDCT nodule annotation across radiologists and methods." *Radiology*, 2006: 385-391.

- [20] Nakamura, K., H. Yoshida, and R. Engelmann. "Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks." *Radiology*, 2000: 214:823–830.
- [21] Ochs, R., H. Kim, E. Angel, C. Panknin, M. McNitt Gray, and M. Brown. "Forming a reference standard from LIDC data: impact of reader agreement on reported CAD performance." *SPIE Medical Imaging*. San Diego, CA: SPIE, 2007.
- [22] Petrick, N., B. D. Gallas, F. W. Samuelson, R. F. Wagner, and K. J. Myers. "Influence of panel size and expert skill on truth panel performance when combining expert ratings." *SPIE Medical Imaging*. SPIE, 2005.
- [23] Raicu, D.S., E. Varutbangkul, J.G. Cisneros, J.D. Furst, D.S. Channin, and S.G. Armato III. "Semantics and Image Content Integration for Pulmonary Nodule Interpretation in Thoracic Computed Tomography." *SPIE Medical Imaging*. San Diego, CA: SPIE, 2007.
- [24] Sahiner, B., et al. "False-positive reduction using Hessian features in computer-aided detection of pulmonary nodules on thoracic CT images." *SPIE Medical Imaging*. San Diego: SPIE, 2005.
- [25] Sahiner, Berkman, et al. "Concordance of Computer-Extracted Image Features with BI-RADS Descriptors for Mammographic Mass Margin." *SPIE Medical Imaging*. San Diego: SPIE, 2008.
- [26] Teague, M. R. "Image analysis via the general theory of moments." *Journal of the Optical Society of America*, 1980.
- [27] van Ginnekin, Bram. "Supervised Probabilistic Segmentation of Pulmonary Nodules in CT Scans." *Medical Image Computing and Computer-Assisted Intervention*. Copenhagen: Springer, 2006. 912-919.
- [28] Varutbangkul, Ekarin, D. S. Raicu, and J. D. Furst. "A Computer-Aided Diagnosis Framework for Pulmonary Nodule Interpretation in Thoracic Computed Tomography." *DePaul CTI Research Symposium*. DePaul University, 2007.
- [29] Wiemker, R., R. Opfer, T. Bulow, S. Kabus, and E. Dharaiya. "Repeatability and Noise Robustness of Spicularity Features for Computer Aided Characterization of Pulmonary Nodules in CT." *SPIE Medical Imaging*. San Diego: SPIE, 2008.
- [30] Wiemker, R., M. Bergtholdt, E. Dharaiya, S. Kabus, and M. C. Lee. "Agreement of CAD features with expert observer ratings for characterization of pulmonary nodules in CT using the LIDC-IDRI database." *SPIE Medical Imaging*. Orlando: SPIE, 2009.
- [31] Witten, I. H., and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.
- [32] Wormanns, D., and S. Diederich. "Characterization of small pulmonary nodules by CT." *European Radiology*, 2004: 1380-1391.