Characterization of solid pulmonary nodules using three-dimensional features

Artit C. Jirapatnakul[a], Anthony P. Reeves[a], Tatiyana V. Apanasovich[b], Matthew D. Cham[c], David F. Yankelevitz[c], and Claudia I. Henschke[c]

[a]School of Electrical and Computer Engineering, Cornell University, Ithaca, NY;
[b]School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY;
[c]Department of Radiology, Weill Medical College of Cornell University, New York, NY

ABSTRACT

With the development of high-resolution, multirow-detector CT scanners, the prospects for diagnosing and treating lung cancer at an early stage are much improved. However, it is often difficult to determine whether a nodule, especially a small nodule, is malignant from a single CT scan. We developed a computer-aided diagnostic algorithm to distinguish benign from malignant solid nodules based on features that can be extracted from a single CT scan. Our method uses 3D geometric and densitometric moment analysis of a segmented nodule image and surface curvature from a polygonal surface model of the nodule. After excluding features directly related to size, we computed a total of 28 features. Prior to classification, the number of features was reduced through stepwise feature selection. The features are used by two classifiers, k-nearest-neighbors (k-NN) and logistic regression. We used 48 malignant nodules whose status was determined by biopsy or resection, and 55 benign nodules determined to be clinically stable through two years of no change or biopsy. The k-NN classifier achieved a sensitivity of 0.81 with a specificity of 0.76, while the logistic regression classifier achieved a sensitivity of 0.85 and a specificity of 0.80.

Keywords: CT, computer-aided diagnosis, pulmonary nodules, nodule characterization

1. INTRODUCTION

Lung cancer, the most common cause of cancer deaths today, typically presents as a pulmonary nodule in its earliest manifestation. The introduction of high-resolution, multi-row detector CT has allowed radiologists to detect more small nodules than previously possible with either chest radiographs or thick-slice CT. A majority of these small nodules are benign, but the status of these nodules is often difficult to ascertain, requiring follow up. This follow up typically consists of another CT scan at a later time to assess growth rate. A high growth rate is typically indicative of a malignant nodule; however, growth rate assessment requires a second CT scan which prolongs true diagnosis and exposes the patient to a second, possibly unnecessary dose of radiation. Ideally, there would be an accurate, minimally invasive method to diagnose a pulmonary nodule based on the low-dose CT scan used for screening.

There have been many proposed methods that rely on data from a single CT scan to classify a nodule as malignant or benign. Several studies have tested the classification performance of radiologists in conjunction with computer-aided diagnosis (CAD) algorithms. These studies report better performance when the radiologist is aided by a CAD scheme than the radiologist alone.[1,2] However, while these methods are helpful to the radiologist, they still require extensive radiologist interaction. Completely automated methods promise to alleviate much of the burden from the radiologist through the elimination of many benign nodules, allowing radiologists to concentrate their time on nodules that are more likely to be malignant. These automated methods, in general, train a model based on features extracted from a large, documented database of nodules. Individual methods differ in the features and classifier that are used. One method by Suzuki et al[3] utilized pixel values in a local window on a region of interest in a CT image in conjunction with a massively trained artificial neural network to distinguish between malignant and benign nodules. The researchers reported a sensitivity of 1.00 and a specificity of 0.48, with an area under the ROC curve of 0.88. A second study by Aoyama et al[4] also used neural networks, but utilized 41 features extracted from regions of interest containing a nodule. The effective diameter of the nodule was included among the features; the authors reported an area under the ROC curve of

Medical Imaging 2007: Computer-Aided Diagnosis, edited by Maryellen L. Giger, Nico Karssemeijer,
Proc. of SPIE Vol. 6514, 65143E, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.707814

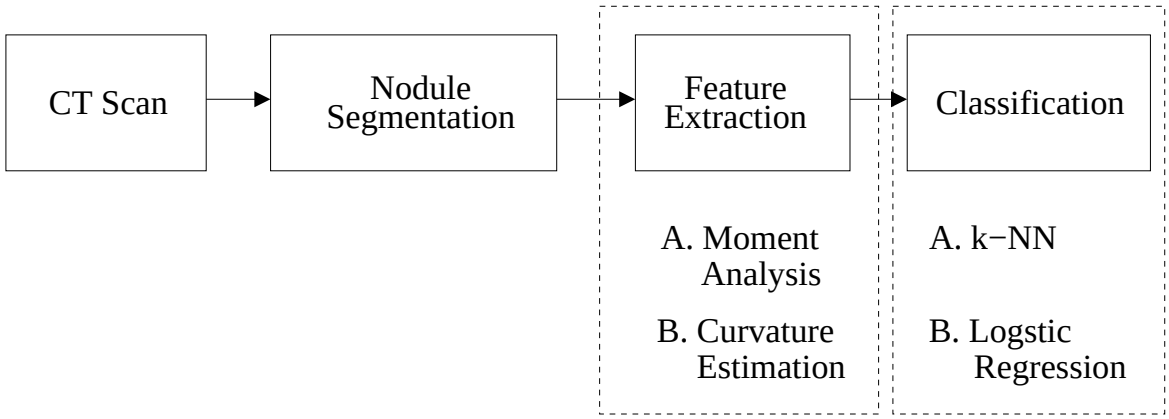


Figure 1. Schematic illustration of classification system.

0.85 using multiple slices. Aside from neural networks, other popular classifiers include logistic regression and linear discriminant analysis (LDA). Shah et al[5] extracted several two-dimensional features, including size-based features, and tested several classifiers, including a LDA classifier, a logistic regression classifier, a decision tree, and quadratic discriminant analysis. Using LDA, they achieved an area under the ROC curve of 0.92. The authors did another study[6] using a different dataset and a different set of features. Some of the features were based on the entire volumetric region of interest as opposed to a single slice, as in their previous paper. The authors again achieved an area under the ROC curve of 0.92 using a logistic regression classifier.

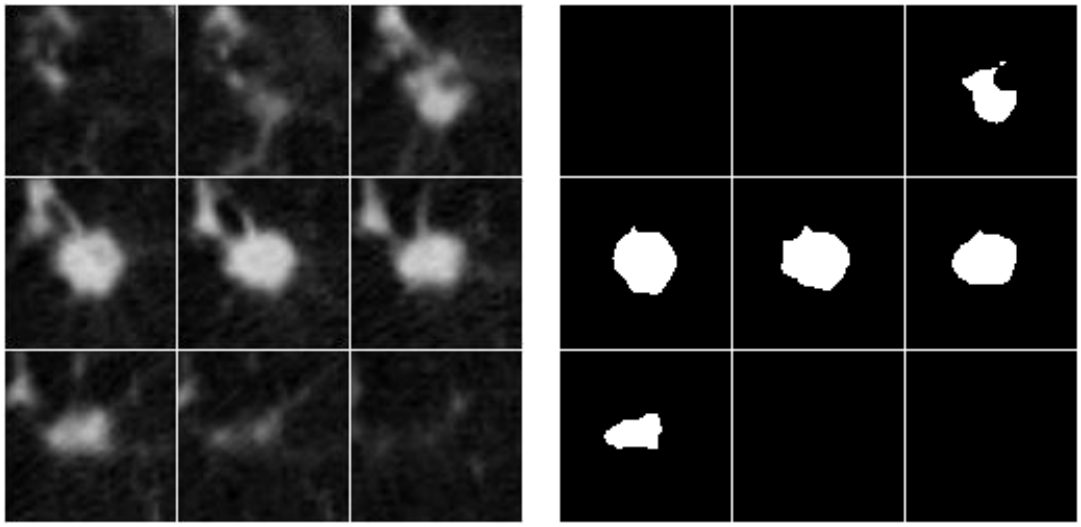
All of these studies, although they use different methods and different sets of features, all use size in some form as a feature in their system. As size is already known to be an important indicator of malignancy, we chose a novel set of features derived from 3D moment analysis and curvature estimation, taking care to exclude any features that are directly impacted by size, such as volume or mass. We compare the performance of two different classifiers, k-nearest neighbors (k-NN) and logistic regression, to determine if the classification task is better suited to a non-parametric or parametric algorithm.

2. METHODS

The classification system can be divided into three primary stages, as shown in Figure 1. The first stage consists of nodule segmentation to select the region from which to extract features. In the second stage, features are extracted from the region of interest and normalized if necessary. In our system, features are computed through the use of moment analysis and surface curvature estimation. The third and final stage classifies the nodule as malignant or benign; there are two separate classifiers tested in our system, k-NN and logistic regression. These stages are described in further detail in the sections below.

2.1. Nodule Segmentation

The first stage of the classification system is segmentation of the nodule. In this paper, segmentation is performed using an algorithm previously developed by Reeves et al.[7] An approximate size and location for the nodule in the CT image is computed based on an initial user-specified seed point using a template-matching method. From this size and location, a region of interest (ROI) is selected around the nodule, as shown in Figure 2a. The ROI is re-sampled into isotropic space by trilinear interpolation and a threshold is applied to the image to obtain a binary image. Morphological filtering using an algorithm by Kostis et al[8] is performed to remove any attached vessels, followed by juxtapleural detection and, if necessary, segmentation using an iterative algorithm that separates the nodule from the pleural surface using a clipping plane. The result of this algorithm is a binary segmented image of the nodule as shown in Figure 2b. A gray-scale image is obtained by using the binary image as a mask on the selected region of interest. These images are used in the feature extraction stage described in the following section.



(a) (b)

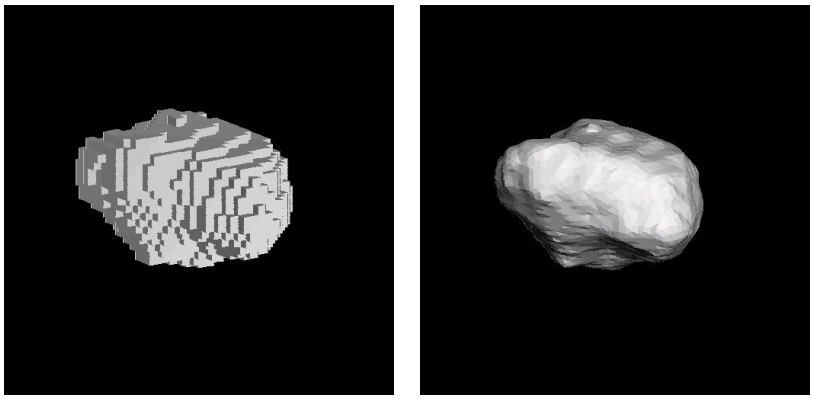
Figure 2. Segmentation of nodule, starting with a) region of interest and resulting in b) binary segmented image

2.2. Feature Extraction

There have been several studies regarding what features best differentiate malignant from benign nodules based on radiologists' observations. In an early paper by Siegelman et al,[9] the criteria for benign nodules included a representative CT number of at least 164 HU and smooth margins. Zwirewich et al[10] found that size (mean diameter), coarse spiculation, and lobulation were good indicators of malignancy. The authors also found that homogeneous attenuation within the nodule occurred with significant more frequency among benign lesions compared to malignant lesions. These two studies alone suggest that the maximum density, density distribution, size, and shape are all important features for nodule classification. Although the features noted here are not a comprehensive list of features studied in the literature, they serve to suggest some of the features that should be included in an automated nodule classification system.

We computed several 3D morphological, shape, and CT features from the segmented binary and grayscale images. These features are computed from moment analysis, curvature estimation, and analysis of CT gray-level data. In this paper, 3D geometric and densitometric moments were computed according to the algorithm described by Reeves et al.[11] Several descriptors of the general nodule shape can be easily derived from the moments, including compactness, sphericity, and aspect ratios.

The margin of a malignant nodule may contain irregularities such as spiculation or lobulation. Such irregularities tend to result in an uneven surface, whereas benign nodules typically have smooth surfaces. The irregularity of the surface can be described through analysis of surface curvature. In contrast to previous methods of surface curvature estimation that have used the values of the gray-level voxels directly,[12] we estimate curvature on a smoothed tessellated polygonal surface model of the nodule, as described by Kostis.[13] Estimating the curvature from the gray-level voxels introduces errors due to the fact that voxels are rectangular approximations of the nodule surface. A comparison of the nodule surface estimated using the voxels from the segmented image is shown in Figure 3a, with the same nodule surface estimated using a smoothed polygonal surface shown in Figure 3b. By using the smoothed polygonal surface model, we reduce the quantization error, ensuring a more accurate curvature estimate. The smooth polygonal surface model is obtained from the segmented binary image using the same method described in Reeves et al.[7] When computing curvature, we ignore regions of the nodule where there might be artifacts created by the removal of vascular structures and the pleural surface and normalize the curvature so that a sphere has a curvature of 1.



(a) (b)

Figure 3. Nodule surface represented using a) voxels and b) smoothed, tessellated polygonal surface. Curvature estimated from the tessellated polygonal surface should be closer to the actual curvature of the nodule surface than the voxel representation of the surface.

From these metrics, we obtained a total of 28 features. This set of features excluded features that were explicitly size-dependent, such as diameter or volume. To ensure that all the features were treated with equal weight, features were normalized to have a zero mean and standard deviation of 1.

2.3. Classification

Two different classifiers were tested in this study, k-nearest-neighbors (k-NN) and logistic regression. The k-NN classifier has several advantages, namely speed and an easy implementation of leave-one-out cross-validation, as compared to artificial neural networks. It is also a non-parametric classifier, so it does not make any assumptions about the underlying model of the data. The k-NN algorithm computes similarity between different nodules in the dataset using Euclidean distance in feature space. The best set of features was chosen using forward stepwise selection.

Logistic regression is a statistical modeling technique that is often used for medical applications. It offers the advantage of a robust statistical foundation as well as many tools to interpret the results. For logistic regression, the best set of features was selected via stepwise feature selection. An ROC curve was generated only for logistic regression, while sensitivity and specificity were computed for both classifiers. Leave-one-out cross validation was used to test both methods to make best use of the limited number of cases in our dataset.

2.4. Data

Cases were selected from a screening trial that had at least one solid nodule with known diagnosis on at least one thin-slice CT scan. Metastatic cancers and benign calcifications were excluded from the dataset. A total of 103 nodules (48 malignant and 55 benign) with CT scans of 1.0 mm, 1.25 mm, or 2.5 mm slice thickness were included in the dataset, with a size range of 1.9 mm to 29.9 mm, distributed as shown in Figure 4. Approximately 54% of the nodules were on 1.0 mm scans, 43% on 1.25mm scans, and 3% on 2.5mm scans. For those nodules on 2.5mm scans, the nodule appeared on at least three slices. Scans were obtained using either GE Medical Systems HiSpeed CT/i, Genesis HiSpeed, LightSpeed QX/i, or LightSpeed Ultra CT scanners. Malignant nodules were confirmed by biopsy or histopathologic analysis of resected tissue; benign nodules were confirmed by at least 2 years of no significant clinical change or negative histopathology.

3. RESULTS

The automated 3D segmentations for each of the 103 nodules were verified by visual inspection. The best value of k for the k-NN classifier was 3; other values of k were tested but resulted in worse performance. Table 1

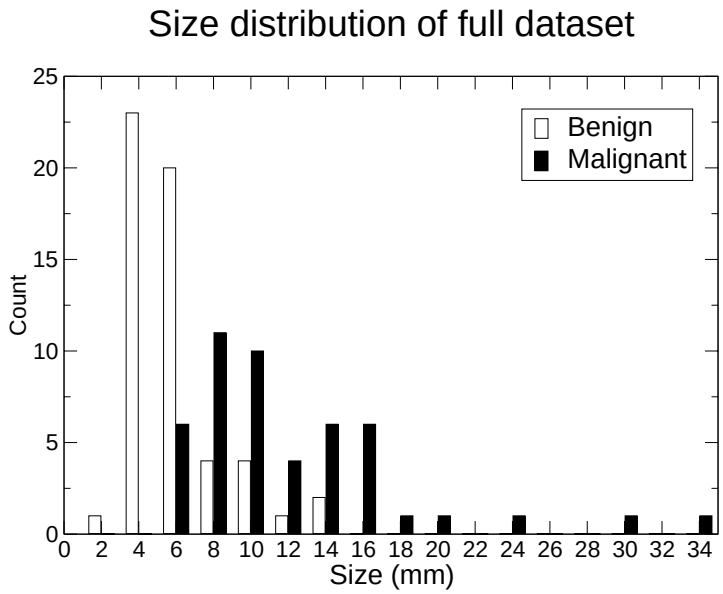


Figure 4. Size distribution of nodules in the dataset where size was determined through automated 3D segmentation. Labels on the size axis represent upper limit of nodule sizes included in the count.

Table 1. Summary of performance for both classifiers.

	Sensitivity	Specificity
Logistic regression	0.85	0.80
k-NN	0.81	0.76

summarizes the results of both classifiers determined through leave-one-out cross-validation, with a ROC curve for logistic regression shown in Figure 5. Although logistic regression achieved a slightly better performance than k-NN, given the size of the dataset, the results are essentially the same, as a 0.04 difference in sensitivity and specificity corresponds to only 2 nodules. From the process of feature selection, the best two features were mean curvature and the ratio between the x and y extent computed from the geometric moment analysis.

4. DISCUSSION

In this study, we developed a method for determining whether a solid nodule is malignant or benign on low-dose high-resolution CT images. Although several previous studies have been done on this topic, none have used the same set of features or utilized the same dataset. Our dataset has a large number of small nodules; approximately 78% of the nodules in our dataset are less than 10 mm. We also took care to eliminate size-dependent features, such as volume or diameter, from consideration. Despite this, the performance of our system was in a similar range as other studies; at a similar operating point of 0.80 specificity, Suzuki et al achieved approximately 0.80 sensitivity, as estimated from their ROC curve, while Aoyama et al achieved approximately 0.75 sensitivity, as estimated from their ROC curve. However, the results are not directly comparable due to differences in the distribution, size, and resolution of the nodules in the datasets.

The best features selected through stepwise feature selection were mean curvature and ratio between the x and y extent derived from geometric moment analysis. In the feature selection method used for logistic regression, feature sets with a large number of features were penalized, leading to only a few features being chosen. In forward stepwise feature selection for k-NN, the algorithm terminated after two features, indicating that adding additional features did not significantly improve performance. Although the number of features is low compared to the feature sets used in other studies, using a large number of features increases the chance of over-fitting to the training data. Each case is plotted in feature space in Figure 6. The two features separate a majority of the cases in the dataset, though there is overlap between the benign and malignant nodules that represent nodules

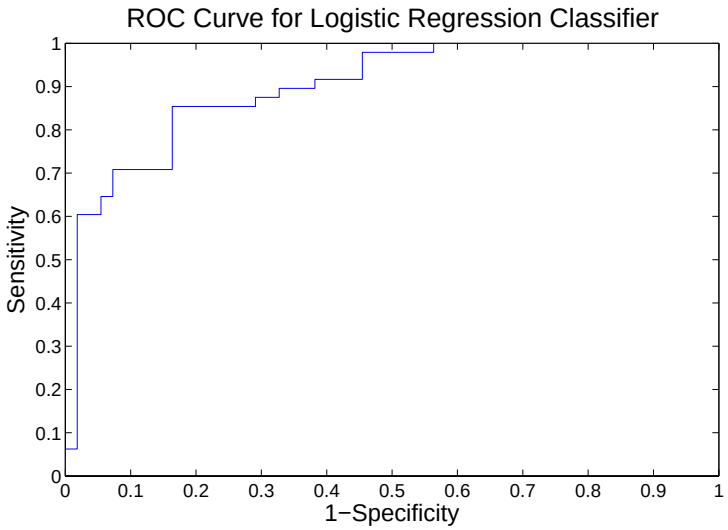


Figure 5. ROC curve for logistic regression using leave-one-out cross validation.

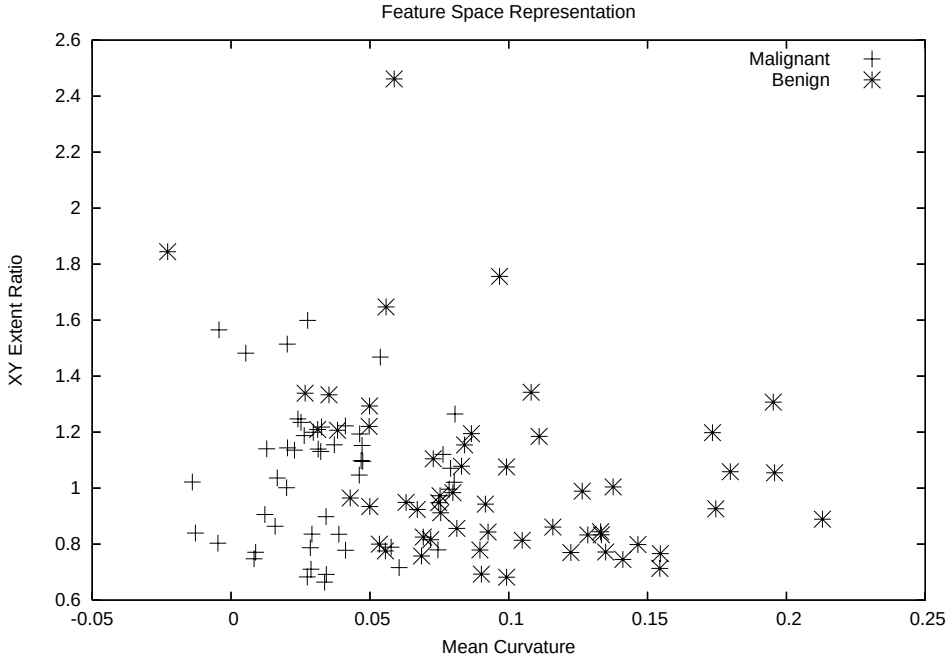
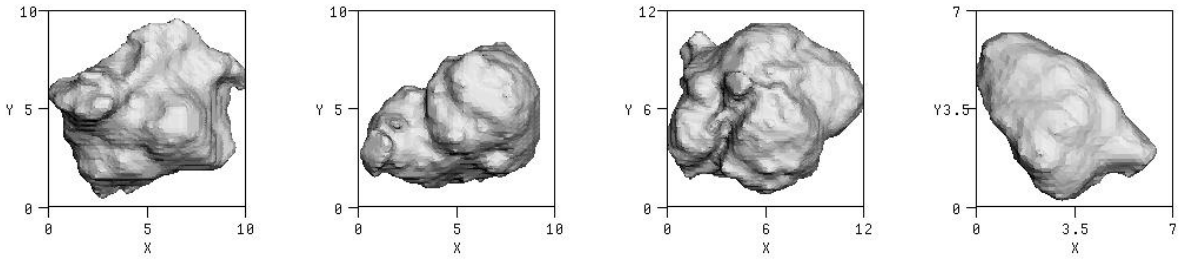


Figure 6. Plot of nodules in feature space. These two features separate most of the nodules, though there is some overlap between the two classes.

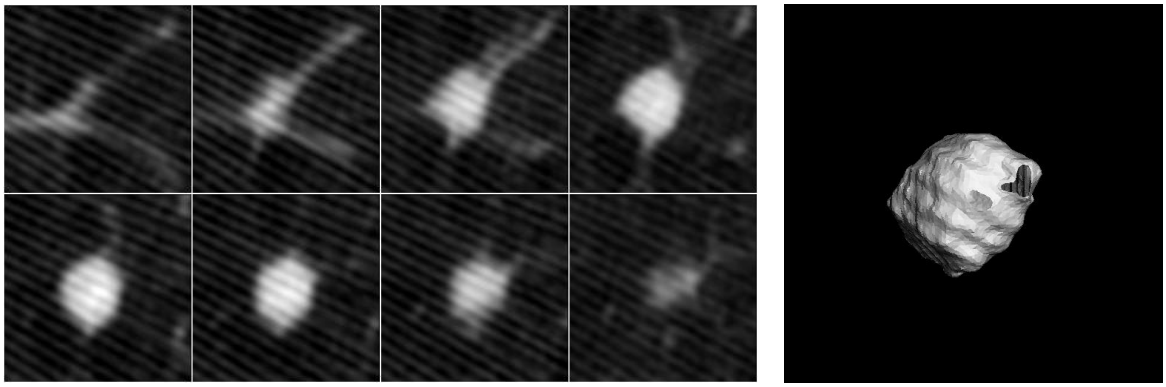
that will be misclassified. Mean curvature is an important feature, as the majority of the malignant nodules have a mean curvature of 0.075 or less. In our estimation of curvature, negative curvature indicates a concave surface, while positive curvature indicates a convex surface. A nodule with many surface irregularities tends to have a lower mean curvature due to the presence of concave features, while a nodule with a consistent surface would have a higher mean curvature.

Nodules that were correctly classified are shown in Figure 7a and b, with misclassified nodules shown in Figure 7c and d. These examples highlight the difficulty in nodule characterization based on only shape features. Although some nodules are misclassified because there are no discernible differences between the nodule and



(a) (b) (c) (d)

Figure 7. Light-shaded visualizations, axial view, of a) a malignant nodule correctly classified as malignant, b) a benign nodule correctly classified as benign, c) a malignant nodule incorrectly classified as benign, and d) a benign nodule incorrectly classified as malignant. All the nodules have very similar appearances, highlighting the difficulty in nodule classification.



(a) (b)

Figure 8. A benign nodule misclassified as malignant with image artifacts, with a) a montage of several slices of the region of interest and b) light-shaded visualization of the nodule. Note that the holes in the 3D visualization are portions of the surface removed in the process of vessel elimination.

nodules of the opposite class, other nodules are misclassified due to artifacts in the CT image. A montage of the region of interest for a nodule with streaking artifacts is shown in Figure 8. The streaking artifacts alter the margins of the nodule, causing ridge-like protrusions from the nodule surface. Since curvature is used as a feature in our classification system, the streaking artifacts significantly change the values of the curvature features for the nodule.

Two different classifiers were tested in our system, logistic regression and k-NN. Both classifiers exhibited similar sensitivity and specificity, with logistic regression performing slightly better than k-NN. On the plot of the nodules in feature space in Figure 6, there is good separation between the malignant and benign nodules based on the two features we selected. In this case, logistic regression will perform very well, as our results have shown. The k-NN classifier also does well, but since it is a majority voting scheme, it is more likely to incorrectly classify examples near class boundaries. Additional refinements to the k-NN classifier may be necessary for improved performance, such as a different similarity metric.

5. CONCLUSION

Accurate characterization of pulmonary nodules is important for the correct diagnosis and treatment of patients. We have developed a classification system that uses features derived from three-dimensional moment analysis and curvature estimation to classify nodules as malignant or benign. Two different classifiers were tested, logistic regression and k-NN, with similar performance. Our best system using logistic regression achieved a sensitivity of 0.85 with a specificity of 0.80. Our results suggest that 3D moment analysis and curvature analysis provide useful features for the characterization of nodules. Although the performance is promising, additional testing needs to be performed on a larger dataset.

REFERENCES

1. F. Li, M. Aoyama, J. Shiraishi, Q. Li, K. Suzuki, R. Engelmann, S. Sone, H. MacMahon, and K. Doi, "Radiologists' performance for differentiating benign from malignant lung nodules on high-resolution CT using computer-estimated likelihood of malignancy," *American Journal of Roentgenology* **183**, pp. 1209–1215, November 2004.
2. S. K. Shah, M. F. McNitt-Gray, K. R. D. Zoysa, P. Batra, A. Behrashi, K. Brown, L. E. Greaser, J. M. Park, D. K. Roback, C. Wu, E. Zaragoza, J. G. Goldin, R. D. Suh, M. S. Brown, and D. R. Aberle, "Solitary pulmonary nodule diagnosis on CT: Results of an observer study," *Academic Radiology* **12**, pp. 496–501, April 2005.
3. K. Suzuki, F. Li, S. Sone, and K. Doi, "Computer-aided diagnostic scheme for disctinction between and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," *IEEE Trans. on Medicial Imaging* **24**, pp. 1138–1150, Sept. 2005.
4. M. Aoyama, Q. Li, S. Katsuragawa, F. Li, S. Sone, and K. Doi, "Computerized scheme for determination of the likelihood measure of malignancy for pulmonary nodules on low-dose ct images," *Medical Physics* **30**, pp. 387–394, March 2003.
5. S. K. Shah, M. F. McNitt-Gray, S. R. Rogers, J. G. Goldin, R. D. Suh, J. W. Sayre, I. Petkovska, H. J. Kim, and D. R. Aberle, "Computer-aided diagnosis of the solitary pulmonary nodule," *Academic Radiology* **12**, pp. 570–575, May 2005.
6. S. K. Shah, M. F. McNitt-Gray, S. R. Rogers, J. G. Goldin, R. D. Suh, J. W. Sayre, I. Petkovska, H. J. Kim, and D. R. Aberle, "Computer aided characterization of the solitary pulmonary nodule using volumetric and contrast enhancment features," *Academic Radiology* **12**, pp. 1310–1319, October 2005.
7. A. Reeves, A. Chan, D. Yankelevitz, C. Henschke, B. Kressler, and W. Kostis, "On measuring the change in size of pulmonary nodules," *IEEE Transactions on Medical Imaging* **25**, pp. 435–450, April 2006.
8. W. J. Kostis, A. P. Reeves, D. F. Yankelevitz, and C. I. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images," *IEEE Transactions on Medical Imaging* **22**, pp. 1259–1274, October 2003.
9. S. Siegelman, N. Khouri, F. Leo, E. Fishman, R. Braverman, and E. Zerhouni, "Solitary pulmonary nodules: CT assessment," *Radiology* **160**(2), pp. 307–312, 1986.
10. C. Zwirewich, S. Vedal, R. Miller, and N. Muller, "Solitary pulmonary nodule: high-resolution CT and radiologic-pathologic correlation," *Radiology* **179**(2), pp. 469–476, 1991.
11. A. P. Reeves, R. J. Prokop, S. E. Andrews, and F. P. Kuhl, "Three-dimensional shape analysis using moments and fourier descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**, pp. 937–943, November 1988.
12. Y. Kawata, N. Niki, H. Ohmatsu, M. Kusumoto, R. Kakinuma, K. Mori, H. Nishiyama, K. Eguchi, M. Kaneko, and N. Moriyama, "Curvature based characterization of shape and internal intensity structure for classification of pulmonary nodules using thin-section CT images," in *Medical Imaging 1999: Image Processing*, K. Hanson, ed., **3661**, pp. 541–552, May 1999.
13. W. J. Kostis, *Three-dimensional computed tomographic image analysis for early cancer diagnosis in small pulmonary nodules.* PhD thesis, Cornell University, January 2001.