Bayesian Classifier for Predicting Malignant Renal Cysts on MDCT: Early Clinical Experience

Youngjoo Lee[1]
Namkug Kim[2]
Kyoung-Sik Cho[2]
Suk-Ho Kang[1]
Dae Yoon Kim[2]
Yoon Young Jung[2]
Jeong Kon Kim[2]

[1]Department of Industrial Engineering, Seoul National University, Seoul, Republic of Korea.

[2]Department of Radiology, Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, 388-1 Poongnap-dong, Songpa-gu, Seoul 138-736, Republic of Korea. Address correspondence to J. K. Kim (kim.jeongkon@gmail.com).

OBJECTIVE. The objective of our study was to evaluate the feasibility and usefulness of the Bayesian classifier for predicting malignant renal cysts on MDCT.

MATERIALS AND METHODS. Ninety-three complicated cysts with pathologic confirmation were enrolled. Patient age and sex and seven morphologic features of the cysts including the maximum diameter, wall features, wall thickness, septa features, measurable enhancement of the wall and septa, presence of calcification, and presence of an enhancing soft-tissue component were used to train the Bayesian classifier. Four radiologists independently reviewed the MDCT images, and the probability of malignancy in each cyst was rated by the radiologists and the Bayesian classifier. The diagnostic performances of the radiologists' visual decisions and the Bayesian classifier were then compared using receiver operating characteristic (ROC) curve analysis. The sensitivity and specificity were also compared between the visual decisions and the Bayesian classifier.

RESULTS. The area under the ROC curve for predicting malignant renal cysts by the Bayesian classifier was greater than the visual decisions of three readers (reader 1, $p = 0.02$; reader 2, $p < 0.01$; reader 4, $p = 0.02$) and was similar to the visual decision of one reader (reader 3, $p = 0.51$). The specificity for predicting malignant renal cysts was greater by the Bayesian classifier than by the visual decisions in readers 2 ($p = 0.04$) and 4 ($p = 0.02$) and was similar in readers 1 ($p = 0.68$) and 3 ($p = 1.00$). In terms of sensitivity, there was no significant difference between the Bayesian classifier and the visual decisions in all four readers ($p > 0.05$).

CONCLUSION. For predicting malignant renal cysts on MDCT, the Bayesian classifier is feasible and may improve diagnostic performance.

A renal cyst is a very common disease entity that is noted in approximately half of patients older than 50 years [1]. Although most renal cysts are simple cysts and thus require no treatment unless they are symptomatic, complicated renal cysts are a diagnostic challenge because some are malignant—that is, cystic renal cell carcinoma [2]. Because clinical examination is inaccurate for differentiating benign from malignant renal cysts, imaging diagnosis plays a key role in differentiating malignant from benign cysts, and CT has been most widely used for this purpose.

To provide guidelines for managing patients with renal cysts, Bosniak [3] proposed a classification system for predicting the probability of malignant renal cysts according to the CT findings. However, the reliability and validity of this system are still controversial [4–15]. Because this classification system is based on the subjective decisions of observers, it is potentially limited in terms of interobserver agreement, particularly in differentiating category 2 lesions from category 3 lesions [12]. In addition, previous studies have shown its potentially limited validity for identifying malignant renal cysts because a considerable number of pathologically proven malignant cysts were classified as category 2 in the CT interpretations [4–14].

Inaccurate reader prediction of malignant renal cysts may be caused principally by two reader defects: inconsistent final decisions regarding the same findings and misperception of the imaging findings. Incorrect perception can be overcome by a certain degree of reader experience, consensus reading, and the use of MDCT. However, ensuring that readers make reliable decisions does not seem easy to achieve because decision making is a more complex process than per-

ception. The unreliability of readers' decisions can potentially be improved using the constant decision making provided by artificial intelligence. From this viewpoint, the Bayesian classifier, an outperforming artificial intelligence for medical decision making, is anticipated to improve the diagnostic performance of MDCT for diagnosing malignant renal cysts. With this hypothesis, we developed a Bayesian classifier for predicting malignant renal cysts using MDCT findings. The purpose of this study was to evaluate the feasibility and usefulness of the Bayesian classifier for predicting malignant renal cysts on MDCT.

Materials and Methods
Patients

A computerized search of our institution's medical records for patients who had undergone both MDCT and surgical excision of renal cysts from January 1999 to December 2006 generated a list of 146 patients. Of those 146 patients, reader 1, a board-certificated radiologist with 12 years of clinical experience with kidney CT interpretation at the time of the study, selected 101 patients with complicated cysts. A complicated cyst was defined as a cyst with at least one focus of septa, a solid nodule, and any calcification or wall thickening on MDCT. Of the 101 patients, four patients were excluded because they had not undergone contrast-enhanced CT and four were excluded because they had numerous renal cysts and identification of the surgically removed cysts was not possible. Finally, 93 patients (57 men and 36 women; mean age ± SD, 50 ± 12 years) were enrolled in this study.

MDCT Examination

The CT images were obtained using 4- and 16-MDCT scanners. All MDCT examinations consisted of unenhanced scans, corticomedullary phase scans, and nephrographic phase scans. The slice thickness for unenhanced scanning was 5 mm on both 4- and 16-MDCT. The slice thickness for corticomedullary and nephrographic phase scanning was 2.5 and 5 mm on 4-MDCT and 3 and 3 mm on 16-MDCT, respectively. All patients received 500–900 mL of oral contrast material (2% barium sulfate suspension [E-Z-CAT, E-Z-EM]) 30 minutes before CT examination. IV contrast material (iopromide [Ultravist 300, Bayer HealthCare]; or iopamidol [Iopamiro 300, Bracco]) was administered into an antecubital vein using a power injector (Percupump II, E-Z-EM) at a dose of 2.0 mL/kg to a maximum dose of 160 mL at a rate of 2.5–3.0 mL/s. The scanning delay for the corticomedullary phase was determined using the automatic bolus-triggering technique, which started scanning when the CT attenuation of a region of interest in the aorta at the renal artery level reached 100 HU; the scanning delay for the late nephrographic–early excretory phase was 120–150 seconds.

Database for the Bayesian Classifier

The flow of our study design in terms of collecting data and establishing the Bayesian classifier is illustrated in Figure 1. To train the Bayesian classifier and establish internal strategies for differentiating malignant from benign cysts, we used seven MDCT and two clinical features. In terms of the seven MDCT features of the cysts, readers 1 and 2 reviewed the MDCT images in a consensus fashion. Reader 2 is a board-certified radiologist

with 25 years of experience with kidney CT interpretation. Both readers were unaware of the information regarding pathologic diagnosis; however, they knew which kidney had been surgically removed. The two readers recorded the seven MDCT features for each cyst according to the image criteria of the Bosniak classification [16] and recorded each patient's age and sex. The MDCT features included the following: maximum cyst diameter; feature of the cyst wall (smooth or irregular); wall thickness (thick or thin based on visual inspection according to an arbitrary criterion of 3 mm); septa feature (no septa, a few hairline-thin septa, multiple hairline-thin septa, thickened smooth septa, or thickened irregular septa); enhancement of wall and septa (absence or presence); calcification (absence, fine or short segment of slightly thickened calcification, or thick and nodular calcification); and enhancing soft-tissue component (absence or presence).

Bayesian Classifier

We developed in-house software to determine the probability of malignant renal cysts according to the Bayesian classifier. A detailed explanation of the Bayesian classifier is presented in Appendix 1. Based on the training set of cases that included the nine features and the final pathologic diagnosis (benign or malignant), the Bayesian classifier was implemented. Note that the Bayesian classifier requires an assumption regarding the distribution of feature values of a specific class, so we assumed that feature values of each class followed a gaussian distribution. A sequential forward selection method was conducted to eliminate irrelevant features. To assess the cross-validation when selecting features, a leave-one-out method was used,
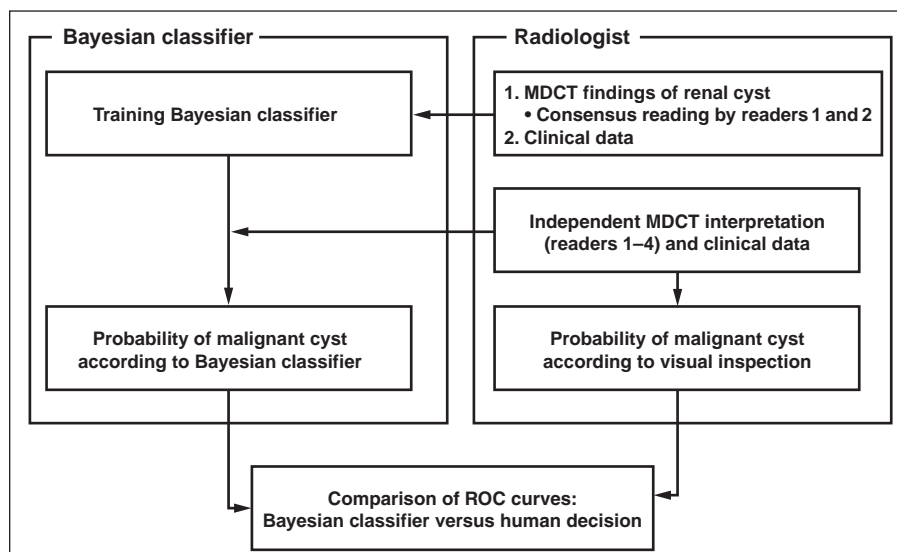

Fig. 1—Flow diagram shows study design. ROC = receiver operating characteristic.

which split the data into a training set ($n = N − 1$) and a test set ($n = 1$) that mutually exclusively and alternatively repeated the learning process until all data had been tested.

Predicting Malignant Renal Cysts

Seven months after the initial MDCT interpretations by readers 1 and 2, four radiologists independently reviewed the MDCT examinations of all study patients. The four readers included readers 1 and 2 who had participated in the initial MDCT interpretations, a board-certified radiologist with 6 years of experience with kidney CT interpretation (reader 3), and a resident radiologist with 3 years of experience with kidney CT interpretation (reader 4). These four readers recorded the seven MDCT features and patient age and sex. The seven MDCT features were the same as those used for the training set of the Bayesian classifier. The readers then rated the probability of malignancy in each cyst according to a 5-point scale: 1, definitely benign; 2, probably benign; 3, indeterminate; 4, probably malignant; and 5, definitely malignant.

The seven MDCT features and patient age and sex, all of which constituted the testing set of the Bayesian classifier, were transferred to the in-house software by a computer engineer. The software then determined the probability of a malignant cyst according to the Bayesian classifier. Finally, for each reader, there were two probability scores for a malignant renal cyst including a score rated by each radiologist's visual decision and that given by the Bayesian classifier.

Statistical Analysis

To compare the diagnostic accuracy of the radiologists' visual decisions with that of the Bayesian classifier for predicting malignant renal cysts, we used receiver operating characteristic (ROC) curve analysis. The area under the ROC curve (AUC) was compared both between each radiolo-

TABLE 1: Frequency of Malignant Renal Cysts According to the Malignancy Score Based on the Visual Decisions of Four Radiologists

Malignancy Score by Visual Decision[a]	No. of Pathologically Proven Malignant Cysts / No. of Cysts According to Malignancy Score Based on Visual Decision			
	Reader 1	Reader 2	Reader 3	Reader 4
1	0/4 (0)	0/8 (0)	1/9 (11)	0/5 (0)
2	6/28 (21)	4/18 (22)	2/14 (14)	1/14 (7)
3	3/13 (23)	3/10 (30)	7/17 (41)	5/19 (26)
4	20/24 (83)	18/25 (72)	22/32 (69)	32/41 (78)
5	23/24 (96)	27/32 (84)	20/21 (95)	14/14 (100)

Note—Data in parentheses are percentages of pathologically proven malignant cysts over cysts with a given malignancy score.
[a]The readers rated the probability of malignancy in each cyst according to a 5-point scale: 1, definitely benign; 2, probably benign; 3, indeterminate; 4, probably malignant; and 5, definitely malignant.

gist's visual decision and the Bayesian classifier and among readers in each method using the Z test. From this analysis, the optimal cutoff values that showed the greatest Youden's index [17] for differentiating benign from malignant cysts were extracted. Using those values, we calculated the sensitivity and specificity for differentiating benign from malignant renal cysts.

Results

Visual Decision

Of the 93 patients included in this study, 52 patients (38 men and 14 women; mean age, 47 years) had cystic renal carcinomas and 41 patients (19 men and 22 women; mean age, 53 years) had benign cysts on pathologic examination. The predictions of malignant renal cysts by each radiologist's visual decision and the corresponding pathology result are summarized in Table 1. The pathology reports indicated a malignant renal cyst in 16–25% of the cysts with a malignancy score of 1–3 on visual decision, whereas 78–90% of the cysts with a malignancy score of 4–5 were proven to be malignant cysts. The

AUCs of the visual decisions for predicting malignant renal cysts were 0.879 (95% CI, 0.795–0.938) by reader 1, 0.826 (95% CI, 0.733–0.897) by reader 2, 0.881 (95% CI, 0.797–0.939) by reader 3, and 0.835 (0.843–0.904) by reader 4 (Fig. 2). There was no statistically significant interobserver difference in the AUCs of visual inspection ($p > 0.05$).

From the ROC curves, a score of 4 was determined to be the optimal cutoff value for predicting malignant renal cysts in all readers; consequently, renal cysts with a score of 4 or 5 were considered to be malignant. The statistical parameters regarding the diagnostic accuracy according to this threshold are summarized in Table 2.

Bayesian Classifier

The AUC of the Bayesian classifier from the training set, which was based on the MDCT interpretations by readers 1 and 2, was 0.960 (95% CI, 0.908–0.986). Using the Bayesian classifier, the probability of a malignant renal cyst was calculated according to the MDCT interpretations of the four read-
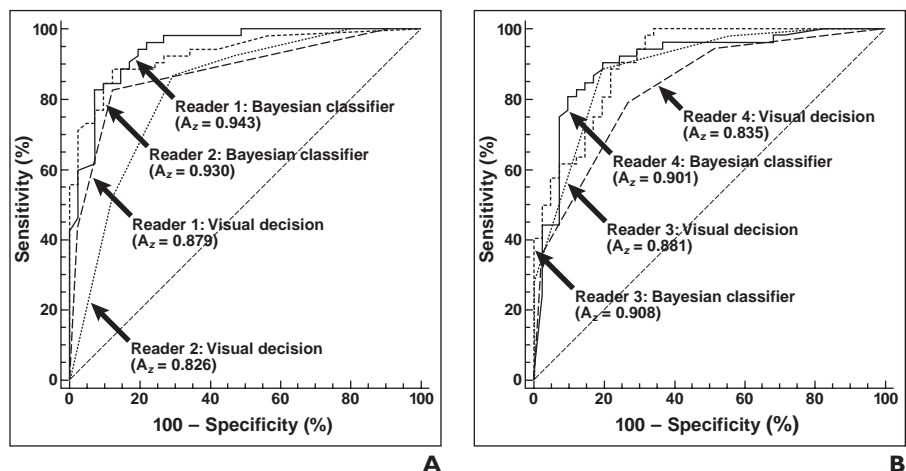
Fig. 2—Receiver operating characteristic (ROC) curves for readers.
A, Area under ROC curve (A_z) for predicting malignant renal cysts is significantly greater by Bayesian classifier (reader 1, 0.943; reader 2, 0.930) than by visual decisions of readers 1 and 2 (reader 1, 0.879; reader 2, 0.826).
B, A_z for predicting malignant renal cysts by Bayesian classifier (0.908) is similar to that by visual decision (0.881) of reader 3. A_z by Bayesian classifier (0.901) is significantly greater than by visual decision (0.835) of reader 4.

TABLE 2: Diagnostic Accuracy for Predicting Malignant Renal Cysts

Statistical Parameter	Reader 1		Reader 2		Reader 3		Reader 4	
	Visual Decision	Bayesian Classifier	Visual Decision	Bayesian Classifier	Visual Decision	Bayesian Classifier	Visual Decision	Bayesian Classifier
No. of true-positives	43	43	45	46	46	46	41	42
No. of true-negatives	36	38	29	36	33	32	30	37
No. of false-positives	5	3	12	5	8	9	11	4
No. of false-negatives	9	9	7	6	6	6	9	10
Sensitivity (%)	83	83	87	88	88	88	79	81
Specificity (%)	88	93	71	88	80	78	73	90

ers. The AUCs for predicting malignant renal cysts based on the Bayesian classifier were 0.943 (95% CI, 0.875–0.980) by reader 1, 0.930 (95% CI, 0.858–0.973) by reader 2, 0.908 (95% CI, 0.830–0.958) by reader 3, and 0.901 (0.822–0.954) by reader 4 (Fig. 2). The AUC of the Bayesian classifier was greater than that of the visual decisions by readers 1 ($p = 0.02$), 2 ($p < 0.01$), and 4 ($p = 0.02$) (Fig. 3). There was no statistically significant difference in the AUCs between visual inspection by reader 3 and the Bayesian classifier ($p = 0.51$). There was no interobserver difference in the AUCs of the Bayesian classifier ($p > 0.05$).

The optimal cutoff values of the Bayesian classifier for differentiating benign from malignant renal cysts were 0.721 for reader 1, 0.781 for reader 2, 0.723 for reader 3, and 0.723 for reader 4. The statistical parameters regarding the diagnostic accuracy according to these thresholds are summarized in Table 2. The specificity for predicting malignant renal cysts of the Bayesian classifier was greater than that of the visual decisions made by readers 2 ($p = 0.04$) and 4 ($p = 0.02$), whereas there was no difference between the Bayesian classifier and the visual decisions made by readers 1 ($p = 0.68$) and 3 ($p = 1.00$). For reader 2, eight benign renal cysts were correctly predicted by the Bayesian classifier but were incorrectly predicted by visual decision, whereas only one benign cyst was correctly predicted by visual decision but incorrectly predicted by the Bayesian classifier. For reader 3, seven benign cysts were correctly predicted by the Bayesian classifier but were incorrectly predicted by the visual decision, whereas no benign cyst was correctly predicted only by visual decision. The sensitivity for predicting malignant renal cysts was similar between the visual decisions and the Bayesian classifier in all readers ($p > 0.05$).

Discussion

Artificial intelligence, or "machine learning," is defined as follows [18]:

a field of science and engineering concerned with the computational understanding of what is commonly called intelligent behavior, and with the creation of artifacts that exhibit such behavior.

As the present practice of medicine encounters the challenge of acquiring, analyzing, and applying a large amount of information, machine learning has been introduced to help clinicians determine diagnoses, make therapeutic decisions, and predict outcomes [19]. Programs that enable computers to function in ways that make people seem intelligent are called artificial intelligence systems.

Among the various algorithms of machine learning, the Bayesian classifier has been shown to outperform most advanced and sophisticated algorithms for many medical as well as nonmedical diagnostic problems [20, 21]. Therefore, the Bayesian classifier has become a benchmark algorithm that is preferred to any other advanced method in many medical domains [20, 22].

In our study, the Bayesian classifier showed greater AUC than the visual deci-

sions by three readers and greater specificity than the visual decisions by two readers for predicting malignant renal cysts. None of the readers' visual decisions was superior to the Bayesian classifier in AUC, sensitivity, or specificity. Therefore, we suggest that the Bayesian classifier not only is feasible but also can improve the diagnostic accuracy for predicting malignant renal cysts.

Compared with the Bayesian classifier, the human decision-making process seems to be potentially limited because human decisions may be inconsistent when multiple parameters should be comprehensively evaluated to make a single decision [23]. Furthermore, decision making based on observer experience may be incomplete because humans cannot correctly remember the diagnoses of all previous cases. Last, various physical or emotional states can interfere with human decision making. From this viewpoint, decision making using the Bayesian classifier can overcome these human limitations and thereby provide better diagnostic performance.

An additional advantage of using the Bayesian classifier is that various context-sensitive classifiers can be produced by modifying the threshold value. The setting of a higher probability threshold improves the performance of the Bayesian classifier but leaves a larger amount of data unclassified. The posterior probability thresholding allows tuning of the sensitivity and specificity of the classifier according to the relative cost of false-positive and false-negative predictions. A probabilistic model is well suited for clinical practice because it allows a radiologist to make another decision with low confidence cases.

Despite the many advantages of the Bayesian classifier over the human decision-making process, our computerized evaluation system based on the Bayesian classifier has potential limitations. First, our computer-aided diagno-
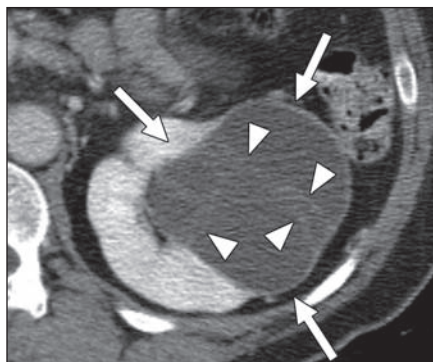

Fig. 3—MDCT of 49-year-old woman. Nephrographic phase MDCT image shows 7-cm cyst (*arrows*) with thickened wall and multiple thickened irregular septa in left kidney. According to visual decisions, probability score of malignancy was 4 by all readers. Malignant score by Bayesian classifier was 0.34 for reader 1 (benign), 0.66 for reader 2 (malignant), 0.80 for reader 3 (malignant), and 0.34 for reader 4 (benign). Pathology examination indicated benign renal cyst. Arrowheads = multiple thickened irregular septa.

sis system differs from those used in mammography and screening chest CT because our Bayesian classifier establishes a diagnosis based on findings obtained by subjective human observation. Therefore, our computerized evaluation has the pitfall that inaccurate perception of the MDCT findings may lead to an incorrect diagnosis. Second, in clinical situations, many findings cannot be classified simply as nominal or ordinal data and therefore cannot be adequately evaluated by the Bayesian classifier. For example, the septa score in our study cannot completely express the characteristics of septa, and some minute but important findings can be adequately reflected only by human interpretation. Last, when the database is not completely filled, application of the Bayesian classifier is complicated. Therefore, to achieve optimal decision making, the Bayesian classifier and the human process should be used simultaneously.

The decision by the Bayesian classifier is affected by the disease prevalence in a study population. For example, in a database in which the disease prevalence is very high, the output of the Bayesian classifier may shift toward a high positive rate and vice versa [24]. Therefore, the disease prevalence in the database should be made similar to the natural disease prevalence. In this context, our study may be limited because the proportion of malignant renal cysts in our study population was greater than the natural prevalence of malignant renal cysts because only patients who had undergone surgical excision of renal cysts were enrolled in our study.

There has been controversy with regard to the validity of the Bayesian classifier in a population in which the assumed distribution of class-specific features is different from the actual distribution of features. However, in a comprehensive article [25], investigators suggested that the Bayesian classifier can be valid even when the actual data do not follow the assumed distribution. Using this theoretic base, we conducted this study without considering the distribution of class-specific features.

There are limitations to this study. First, to select the study population, reader 1 chose complicated cysts from the primary candidates according to his MDCT observations. Although reader 1 is a board-certified radiologist, there is a risk that other readers might have disagreed with his primary decision regarding the determination of simple or complicated cysts in some patients.

Second, consensus observations by readers 1 and 2 were used for establishing the training data set of the Bayesian classifier. However, incorrect perception or misinterpretation of the MDCT findings is a potential risk and may have led to an inaccurate training set. To reduce this potential limitation, we used their consensus review rather than their individual reviews to generate a training set.

Third, readers 1 and 2 both participated in making the training set for the Bayesian classifier and in testing the Bayesian classifier. To minimize this potential limitation by eliminating the memory effect of both readers, the second review of the MDCT findings was performed 7 months after the initial MDCT interpretations. However, we admit that this time interval cannot change their inherent judgmental criteria.

Fourth, although this study included only patients who had undergone MDCT, examinations performed on both 4- and 16-MDCT units were included. To our knowledge, it is statistically impossible to compare the diagnostic accuracy between different populations. Two different types of MDCT scanners may produce different image quality and may need different scanning delays for the late nephrographic–early excretory phase, even though the difference should not have been substantial.

Last, unfortunately, the software for evaluating renal cysts according to the Bayesian classifier is not generally available and adaptable to any facility, although the concept and use of the Bayesian classifier are widely accepted. Therefore, clinical trials similar to ours are available only to institutions that can develop in-house software.

Although several studies have been performed to determine the optimal scanning delay for the nephrographic phase, determining the optimal scanning delay is still difficult because the onset of the nephrographic phase is highly dependent on the methods of contrast material administration and patient characteristics [26]. Israel and Bosniak [27] recommended a scanning delay of 90 seconds; however, other researchers have reported a mean scanning delay of 120–160 seconds for the nephrographic phase [28, 29]. In our study, the scanning delay for the nephrographic phase ranged from 120 to 150 seconds because the scanning delay was changed during the study period.

In summary, our study showed that the diagnostic accuracy of the Bayesian classifier was the same as or superior to the visual decisions by radiologists: The AUC and the specificity of the Bayesian classifier were greater than those of the visual decisions in some readers. Therefore, the Bayesian classifier is feasible and may improve the diagnostic performance in predicting malignant renal cysts on MDCT.

References

1. Kissane JM. The morphology of renal cystic disease. *Perspect Nephrol Hypertens* 1976; 4:31–63
2. Hartman DS. Cysts and cystic neoplasms. *Urol Radiol* 1990; 12:7–10
3. Bosniak MA. The current radiological approach to renal cysts. *Radiology* 1986; 158:1–10
4. Aronson S, Frazier HA, Baluch JD, Hartman DS, Christenson PJ. Cystic renal masses: usefulness of the Bosniak classification. *Urol Radiol* 1991; 13: 83–90
5. Bosniak MA. Cystic renal masses: a re-evaluation of the usefulness of the Bosniak classification system. (letter) *Acad Radiol* 1996; 3:981–984
6. Cloix P, Martin X, Pangaud C, et al. Surgical management of complex renal cysts: a series of 32 cases. *J Urol* 1996; 156:28–30
7. Curry NS, Cochran ST, Bissada NK. Cystic renal masses: accurate Bosniak classification requires adequate renal CT. *AJR* 2000; 175:339–342
8. Harisinghani MG, Maher MM, Gervais DA, et al. Incidence of malignancy in complex cystic renal masses (Bosniak category III): should imaging-guided biopsy precede surgery? *AJR* 2003; 180: 755–758
9. Koga S, Nishikido M, Inuzuka S, et al. An evaluation of Bosniak's radiological classification of cystic renal masses. *BJU Int* 2000; 86:607–609
10. Lang EK, Macchia RJ, Gayle B, et al. CT-guided biopsy of indeterminate renal cystic masses (Bosniak 3 and 2F): accuracy and impact on clinical management. *Eur Radiol* 2002; 12:2518–2524
11. Limb J, Santiago L, Kaswick J, Bellman GC. Laparoscopic evaluation of indeterminate renal cysts: long-term follow-up. *J Endourol* 2002; 16: 79–82
12. Siegel CL, McFarland EG, Brink JA, Fisher AJ, Humphrey P, Heiken JP. CT of cystic renal masses: analysis of diagnostic performance and interobserver variation. *AJR* 1997; 169:813–818
13. Spaliviero M, Herts BR, Magi-Galluzzi C, et al. Laparoscopic partial nephrectomy for cystic masses. *J Urol* 2005; 174:614–619
14. Wilson TE, Doelle EA, Cohan RH, Wojno K, Korobkin M. Cystic renal masses: a reevaluation of the usefulness of the Bosniak classification system. *Acad Radiol* 1996; 3:564–570

15. Brown WC, Amis ES, Kaplan SA, Blaivas JG, Axelrod SL. Renal cystic lesions: predictive value of preoperative computerized tomography. (abstr) *J Urol* 1989; 141:426A

16. Israel GM, Hindman N, Bosniak MA. Evaluation of cystic renal masses: comparison of CT and MR imaging by using the Bosniak classification system. *Radiology* 2004; 231:365–371

17. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. *Crit Care* 2004; 8:508–512

18. Shapiro S. Artificial intelligence. In: Shapiro S, ed. *Encyclopedia of artificial intelligence*. New York, NY: Wiley, 1992

19. Ramesh AN, Kambhampati C, Monson JR, Drew PJ. Artificial intelligence in medicine. *Ann R Coll Surg Engl* 2004; 86:334–338

20. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med* 2001; 23:89–109

21. Kononenko I, Bratko I, Kukar M. Application of machine learning to medical diagnosis. In: Michalski R, Bratko I, Kubat M, eds. *Machine learning, data mining and knowledge discovery: methods and applications*. New York, NY: Wiley, 1998

22. Spiegelhalter DJ, Dawid AP, Lauritzen SL, Cowell RG. Bayesian analysis in expert systems. *Statistical Sci* 1993; 8:277–283

23. Li M, Zhou ZH. Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. *IEEE Trans Systems Man Cybernetics A* 2007; 37:1088–1098

24. Jaynes ET. Prior probabilities. *IEEE Trans Systems Sci Cybernetics* 1968; 4:227–241

25. Schurmann J. *Pattern classification: a unified view of statistical and neural approaches.* New York, NY: Wiley 1996

26. Birnbaum BA, Jacobs JE, Langlotz CP, Ramchandani P. Assessment of a bolus-tracking technique in helical renal CT to optimize nephrographic phase imaging. *Radiology* 1999; 211:87–94

27. Israel GM, Bosniak MA. How I do it: evaluating renal masses. *Radiology* 2005; 236:441–450

28. Cohan RH, Sherman LS, Korobkin M, Bass JC, Francis IR. Renal masses: assessment of corticomedullary-phase and nephrographic-phase CT scans. *Radiology* 1995; 196:445–451

29. Birnbaum BA, Jacobs JE, Ramchandani P. Multiphasic renal CT: comparison of renal mass enhancement during the corticomedullary and nephrographic phases. *Radiology* 1996; 200:753–758

APPENDIX 1: The Bayesian Classifier

Assume data were labeled class C = benign, malignant and described in n-dimensional feature vector v. According to the Bayes theorem, the probability that a sample belongs to a specific class c is as follows:

$$P(C = c \mid v) = \frac{P(v \mid C = c)P(C = c)}{P(v)}$$

Bayes formula can be expressed informally in English by saying that

$$posterior = \frac{likelihood \times prior}{evidence}$$

Bayes formula shows that by observing the feature vector we can convert the prior probability $P(C = c)$ to a posteriori probability $P(C = c \mid v)$. We call $P(v \mid C = c)$ the likelihood of class c with respect to v, a term chosen to indicate that, other things being equal, the class c for which $P(v \mid C = c)$ is large is more likely to be the true class. Notice that it is the product of the likelihood and the prior probability that is most important in determining the posterior probability; the evidence factor, $P(v)$, can be viewed as merely a scale factor that guarantees that the posterior probabilities sum to 1, as all good probabilities must. If we have an observation v for which $P(C = \text{benign} \mid v)$ is greater than $P(C = \text{malignant} \mid v)$, we would decide that the true class of the observation is benign. Conversely, if $P(C = \text{malignant} \mid v)$ is greater than $P(C = \text{benign} \mid v)$, we would classify the observation as malignant.

If feature vectors in each class are normally distributed in n-dimensional space, the likelihood of class c can be written as

$$P(v \mid C = c) = \frac{1}{\sqrt{2\pi}^N |\Sigma_c|} \exp\left[-\frac{1}{2}(v - \mu_c)^t \Sigma_c^{-1}(v - \mu_c)\right]$$

where μ_c represents the mean of vectors in class c and Σ_c is the covariance matrix of vectors in class c. The prior probability of class c is simply the ratio of the number of data in class c to the number of whole data. The posterior probabilities of each class can be calculated by the likelihood product prior probability without the evidence because the evidence is the same in every class. Then sample data with feature vector v can be determined as a class that has the largest posterior probability.