



Stratification of cystic renal masses into benign and potentially malignant: applying machine learning to the bosniak classification

Nityanand Miskin¹ · Lei Qin² · Shanna A. Matalon¹ · Sree H. Tirumani³ · Francesco Alessandrino^{1,2} · Stuart G. Silverman¹ · Atul B. Shinagare^{1,2}

Received: 21 April 2020 / Revised: 14 June 2020 / Accepted: 23 June 2020 / Published online: 1 July 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Purpose To create a CT texture-based machine learning algorithm that distinguishes benign from potentially malignant cystic renal masses as defined by the Bosniak Classification version 2019.

Methods In this IRB-approved, HIPAA-compliant study, 4,454 adult patients underwent renal mass protocol CT or CT urography from January 2011 to June 2018. Of these, 257 cystic renal masses were included in the final study cohort. Each mass was independently classified using Bosniak version 2019 by three radiologists, resulting in 185 benign (Bosniak I or II) and 72 potentially malignant (Bosniak IIF, III or IV) masses. Six texture features: mean, standard deviation, mean of positive pixels, entropy, skewness, kurtosis were extracted using commercial software TexRAD (Feedback PLC, Cambridge, UK). Random forest (RF), logistic regression (LR), and support vector machine (SVM) machine learning algorithms were implemented to classify cystic renal masses into the two groups and tested with tenfold cross validations.

Results Higher mean, standard deviation, mean of positive pixels, entropy, skewness were statistically associated with the potentially malignant group ($P \leq 0.0015$ each). Sensitivity, specificity, positive predictive value, negative predictive value, and area under curve of RF model was 0.67, 0.91, 0.75, 0.88, 0.88; of LR model was 0.63, 0.93, 0.78, 0.86, 0.90, and of SVM model was 0.56, 0.91, 0.71, 0.84, 0.89, respectively.

Conclusion Three CT texture-based machine learning algorithms demonstrated high discriminatory capability in distinguishing benign from potentially malignant cystic renal masses as defined by the Bosniak Classification version 2019. If validated, CT texture-based machine learning algorithms may help reduce interreader variability when applying the Bosniak classification.

Keywords Cystic renal mass · Bosniak classification v2019 · Machine learning · Texture analysis

Introduction

Incidental cystic renal masses are common at CT, with an overall prevalence of 10–41% in adults [1–3] particularly those over the age of 50 [4]. While the vast majority of cystic renal masses are benign, some represent renal cell carcinoma (RCC). Although a common malignancy, (6th most common among men and 8th among women [5]) RCC is an uncommon cause of death, particularly when small [6–8]. Indeed, cystic RCC is a rare cause of mortality; the estimated 10-year risk of death from cystic RCC < 4 cm is ~0.2% [9]. In the pursuit of diagnosing cancer at an early, curable stage, imaging of indeterminate cystic masses that are highly likely benign often ensues. This leads to patient anxiety, and the unnecessary treatment of benign etiologies with subsequent procedural morbidity, loss of renal function

✉ Nityanand Miskin
nmiskin@partners.org

¹ Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA

² Department of Imaging, Dana-Farber Cancer Institute, Harvard Medical School, 450 Brookline Avenue, Boston, MA 02215, USA

³ Department of Radiology, University Hospital Cleveland Medical Center, 11100 Euclid Avenue, Cleveland, OH 44106, USA

and additional cost [10–14]. These data have prompted the need for increased specificity for the diagnosis of RCC [9, 15].

The Bosniak classification, widely used by radiologists and urologists [16], uses structural features to separate cystic renal masses into five classes. Bosniak I and II masses are reliably considered benign; Bosniak IIF, III, and IV masses are potentially malignant. Malignant entities, typically renal cell carcinoma, are found in approximately 10–20% of Bosniak IIF masses [17], 50% of Bosniak III masses [18] and 90% of Bosniak IV masses [18, 19].

The recently published update proposal, referred to as ‘Bosniak Classification version 2019, in part aims to improve specificity for the diagnosis of cystic RCC [15]. It also aims to address the often-cited limitation of interreader variability. Disagreements among readers have ranged from 6 to 75%, with the problem largely limited to Bosniak classes II, IIF and III [9, 12, 20–26]. An additional way to reduce interreader variability in cystic renal mass characterization might be to employ a machine learning algorithm, allowing greater objectivity in applying the Bosniak classification criteria. Texture analysis is a type of quantitative image processing in which the spatial interrelationships of pixel intensities are assessed [27]. Texture analysis and machine learning have been used to characterize and prognosticate solid renal masses including prediction of nuclear grade and histologic subtypes of RCC [28–32], and more recently to diagnose RCC among low attenuation renal masses [33]. Our purpose was to create a CT texture-based machine learning algorithm that distinguishes benign from potentially malignant cystic renal masses as defined by the Bosniak Classification version 2019.

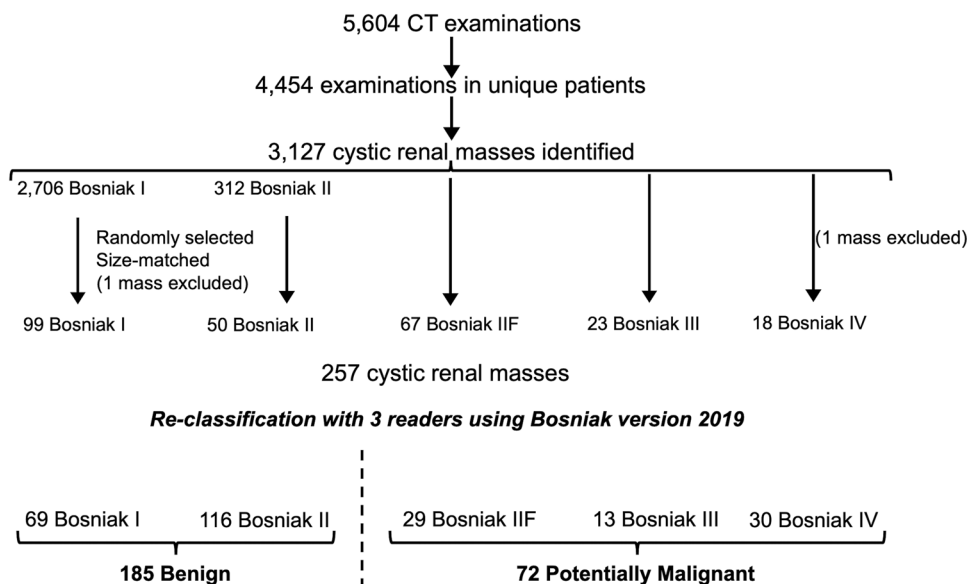
Materials and methods

Patients and setting

This was an Institutional Review Board-approved, Health Insurance Portability and Accountability Act-compliant, retrospective study, with informed consent waived. A search of our institution’s research database yielded 5604 CT examinations performed with renal mass or urography protocol between January 2011 and June 2018. All exams included 3 mm sections reconstructed with a 50% overlap before and 100-s (nephrographic phase) after IV administration 50–150 ml of iodinated contrast material (300–370 mg iodine/ml).

For patients with multiple exams within this time frame, the initial exam was selected; this yielded 4454 unique patients (Fig. 1). A single fourth year radiology resident (N.M.) reviewed the images and associated clinical radiology reports to select the largest mass with the highest Bosniak class from each kidney, using the original Bosniak classification. For example, if a patient had one Bosniak I mass and two Bosniak IIF masses in the right kidney and three Bosniak II masses in the left kidney, the larger of the two Bosniak IIF masses in the right kidney and the largest Bosniak II mass in the left kidney would be included in the study. Thus, a total 3127 cystic renal masses were selected from the 4454 patients, including 3018 Bosniak I and Bosniak II masses (benign group) and 109 Bosniak IIF, Bosniak III, and Bosniak IV masses (potentially malignant group). Mass size was determined by measuring the single largest axial diameter on nephrographic phase images. Size-matching was performed to prevent the predominance of sub-centimeter simple benign cysts (Bosniak I) in the

Fig. 1 Flowchart demonstrating how the study cohort of 257 cystic masses was derived and assigned into one of two groups, benign and potentially malignant. Initial assignment was based on clinical radiology reports and resident review using original Bosniak classification; final assignment was based on Bosniak Classification version 2019



benign group, given malignant cystic renal masses are usually greater than one centimeter in size.

Creation of study cohort

In order to create two groups with a comparable number of masses, a randomly selected sample of 100 size-matched Bosniak I and 50 size-matched Bosniak II masses was created in addition to the total 109 Bosniak IIF, Bosniak III, and Bosniak IV masses. Size-matching was performed for the benign group based on the proportion of masses within each of the following size ranges present in the potentially malignant group: < 1 cm, 1–2 cm, 2–3 cm, 3–4 cm, and > 4 cm. Therefore, 259 cystic renal masses (150 benign and 109 potential malignant) were included in the study. Two were excluded on subsequent review: one contained fat attenuation and thus represented an angiomyolipoma, another contained > 25% enhancing tissue, and therefore was considered a solid mass rather than a cystic mass as defined by the Bosniak Classification version 2019 [15]. Thus, the final patient cohort consisted of 257 cystic renal masses.

Cystic renal mass classification

Two fellowship-trained abdominal radiologists with 15 (S.H.T.), 13 (A.B.S.) years of radiology experience independently assigned a Bosniak Classification version 2019 class to each of 257 cystic renal masses [15]. For the 112 discrepant Bosniak class assignments between the two readers, a third fellowship-trained abdominal radiologist (S.A.M.) with six years of radiology experience independently assigned a Bosniak class. For the six masses with persistent discrepant assignments among the three readers, a fourth fellowship-trained abdominal radiologist (S.G.S.) with 33 years of radiology experience determined the Bosniak class by selecting one of the three assignments. The final study cohort consisted of 257 cystic renal masses, with 185 masses assigned as Bosniak Classification version 2019 I or II (benign group) and 72 assigned as Bosniak Classification version 2019 IIF, III, or IV) masses (potentially malignant group) (Fig. 1).

Texture analysis

A region-of-interest (ROI) that encompassed the entire mass on a single, 3.0 mm thick axial image from the nephrographic phase was created by the radiology resident. The single image selected was chosen to portray the feature associated with the highest Bosniak classification (e.g., enhancing septa, thick wall or nodule). Using a commercial software TexRAD (TexRAD, Feedback PLC, Cambridge, UK) six texture features: mean, standard deviation (SD), mean value of positive pixels (mpp), entropy, skewness, and kurtosis were extracted from the ROI. Wilcoxon signed-rank test was performed to determine the

association of each specific texture feature with benign versus potentially malignant group.

Machine learning algorithms

Three machine learning algorithms were selected, because they have been commonly used [34]: Support vector machine (SVM) with radial kernel, random forest (RF), and logistic regression (LR) were used to conduct supervised machine learning. Tenfold stratified cross validation method was used to train and estimate the machine learning algorithm performance. Because the size of the two groups was imbalanced with 185 benign and 72 potentially malignant masses, the data were partitioned randomly into tenfolds. In each fold, random sampling occurred within each group so as to ensure the proportion of benign to malignant cases found in the original distribution remained in each fold [35]. Ninefolds of data were used to build the machine learning algorithm and the remaining fold was used to test the performance of it. This process was repeated ten times with every fold being used as test data, and the results from the 10 test steps were aggregated and summarized. Prior to machine learning, feature reduction was implemented to remove highly correlated features. Pearson correlations between each pair of features were calculated and single features from pairs of features with Pearson correlations greater than 0.90 were removed. The remaining texture features were standardized to a mean of zero and a standard deviation of one prior to machine learning algorithm construction. For SVM and RF models, there are tuning parameters which control the model complexity. The best choice of these tuning parameters were selected by performing tenfold cross validation on the training data [36]. Delong's method was used to assess for significant differences in AUC values [37].

Receiver operating characteristics (ROC) curves from the aggregated tenfold cross validation were generated and the area under the curve (AUC) for each classifier was calculated. The optimal cutoff value was calculated based on Youden's index, where the cutoff value is the threshold that maximizes the distance to the identity line of the ROC curve, or equivalently, the value that maximizes the sum of sensitivity and specificity [38].

Statistical analysis was conducted using R version 3.3. The "Caret" package was used for machine learning algorithm creation. The "pROC" package was used for ROC analysis [39].

Results

Mass size

The Bosniak I and II masses (benign group) had an average size of 3.0 cm, with a standard deviation of 2.3 cm, with

Table 1 Texture features and association with benign versus potentially malignant in 257 cystic renal masses

Texture feature	Benign		Potentially malignant		P-value
	(Bosniak I, II)		(Bosniak IIF, III, IV)		
	Mean	SD	Mean	SD	
Mean	20.05	19.97	50.41	30.81	<0.0001
sd	19.2	7.31	33.94	28.9	<0.0001
Entropy	4.06	0.27	4.06	0.27	<0.0001
mpp	26.54	16.99	53.19	29.39	<0.0001
Skewness	0.44	0.81	0.76	1.23	0.244
Kurtosis	1.95	6	2.99	6.35	0.718

SD standard deviation, sd standard deviation texture feature, mpp mean value of positive pixels feature

a range of 0.7 cm to 10.9 cm. The Bosniak IIF, III and IV masses (potentially malignant group) had an average size of 3.4 cm, with a standard deviation of 2.2 cm, with a range of 0.9 cm to 11.7 cm. There was no significant difference in size between the two groups ($P=0.21$).

Texture feature associations

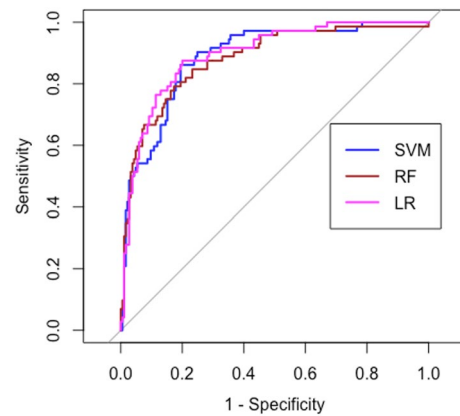
There was a significantly higher value for the texture features mean, sd, entropy, and mpp, among the Bosniak IIF, III, IV masses (potentially malignant group) (compared to the Bosniak I and II masses (benign group) ($P<0.0001$ for mean, sd, entropy, mpp)). The skewness and kurtosis texture features were not significantly different between the two groups ($P=0.244$, $P=0.718$, respectively) (Table 1). Since there was a strong positive correlation between mean and mpp ($r=0.99$), the feature mpp was removed from the group of texture features utilized in machine learning algorithm construction.

Machine learning algorithm performance

The performance of the three machine learning algorithms is displayed in Table 2 and Fig. 2. The RF, LR, and SVM machine learning algorithms demonstrated AUC of 0.88, 0.90, and 0.89, respectively, with mean and standard

Table 2 Performance of the three machine learning algorithms in differentiating benign from potentially malignant cystic renal masses ($n=257$)

Machine learning algorithm	Sensitivity	Specificity	PPV	NPV	AUC
Random forest	0.67	0.91	0.75	0.88	0.88
Logistic regression	0.63	0.93	0.78	0.86	0.9
Support vector machine	0.56	0.91	0.71	0.84	0.89

**Fig. 2** Performance of the three machine learning algorithms in lesion classification in 257 cystic renal masses. Receiver operator characteristic (ROC) curve of the three machine learning algorithms. SVM support vector machine, RF random forest, LR logistic regression

deviation of the individual folds for the RF, LR, and SVM algorithms as 0.89 ± 0.07 , 0.91 ± 0.06 , and 0.91 ± 0.06 , respectively. There was no significant difference among the three algorithms (RF vs. LR, $P=0.4611$; RF vs. SVM, $P=0.718$; LR vs. SVM, $P=0.572$).

The individual texture features alone, such as mean, SD, and entropy have high AUC performance in classification, but machine learning models slightly improved AUC further by combining all features into the models (Figs. 3, 4, 5). Performance of the LR model was significantly better than that of mean texture feature ($P=0.017$); however, there was

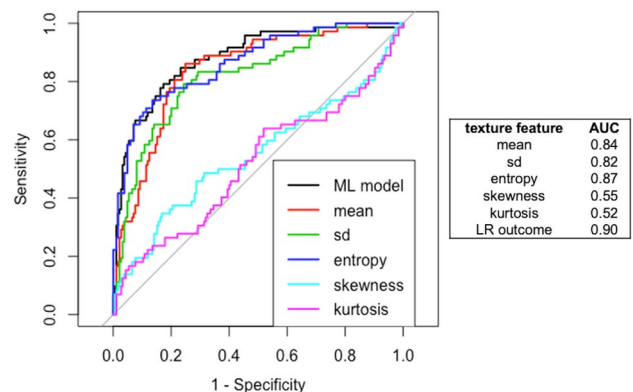
**Fig. 3** Performance of the individual texture features and logistic regression model in 257 cystic renal masses. Receiver operator characteristic (ROC) curve (left) demonstrating that the logistic regression (LR) machine learning algorithm improved the area under the curve (AUC) by combining all texture features. AUC performance for each feature is listed (right). Black is the color for the ROC curve for the LR machine learning algorithm, with the other colored ROC curves showing the performance of the individual texture features listed (right). For the other two machine learning algorithms, results were similar, and thus not shown. ML machine learning, sd standard deviation texture feature



Fig. 4 Axial CT images in nephrographic phase. **a** Low attenuation left interpolar renal mass containing thin calcification in its wall, characterized as a Bosniak II cyst. **b** Low attenuation left interpolar mass containing a short segment of thin calcification in the wall (arrow) and a single thin septation (arrowhead), characterized as a Bosniak II cyst. Both masses were correctly placed into the benign group by the logistic regression machine learning algorithm

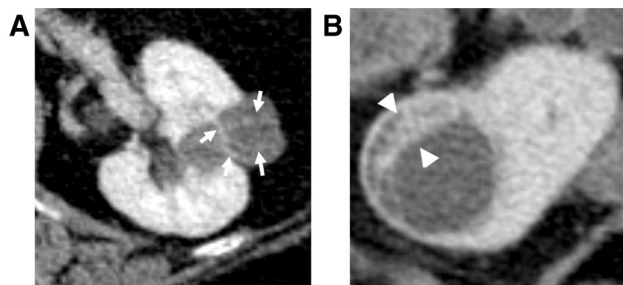


Fig. 5 Axial CT images in nephrographic phase **a** Low attenuation left interpolar mass, containing more than four thin septa (arrows), characterized as a Bosniak IIF cystic mass. **b** Low attenuation right lower pole renal mass containing an 8 mm (thick) septation (arrowheads), characterized as a Bosniak III cyst. Both masses were correctly placed into the potentially malignant group by the logistic regression machine learning algorithm

no significant difference between the LR model and entropy ($P=0.082$).

Discussion

Diagnosing renal cell carcinoma at a curable stage is an important goal, however, it is also important to reduce the overutilization of imaging, and the overdiagnosis and overtreatment of benign masses that is currently observed today [10–15]. Although the Bosniak classification is useful in distinguishing benign from malignant masses, there is marked interreader variability in the assessment of cystic renal masses, especially among Bosniak II, IIF and III masses [39]. Bosniak II masses are reliably considered benign and can be ignored, Bosniak IIF masses are often benign and generally followed, and Bosniak III masses historically have been surgically resected but are now being

increasingly followed [9, 40–42]. Interreader variability is a result of many factors. Included among them is how the imaging features are used to assign a particular Bosniak class. For example, whether septa are considered ‘thin’ (Bosniak II), ‘minimally thick’ (Bosniak IIF) or ‘thick’ (Bosniak III) depends on the definition of ‘thin’, ‘minimally thick’, and ‘thick’. Although Bosniak Classification version 2019 defines each (2 mm, 3 mm, and 4 mm, respectively), measurements can vary among readers [26]. Similarly, there may be interreader variability in the perceived number of septa. Although interreader variability may be lessened by these explicit definitions, it will likely persist to some degree. Texture analysis may address this problem, at least in part, by applying the same analysis to all masses; one potential source of interreader variability would be in how the ROI was placed. Nevertheless, we hypothesized that a combination of CT texture-based machine learning algorithms can be used to more objectively classify cystic renal masses into two groups, one with Bosniak I and II (which are reliably considered benign) and one with Bosniak IIF, III, and IV masses (which are potentially malignant) and possibly address the problem of interreader variability [43].

In this study, three CT texture-based machine learning algorithms demonstrated high discriminatory capability in distinguishing the group with Bosniak I and II masses from the group with Bosniak IIF, III, and IV masses. Our results demonstrated that there were significant differences in texture features mean, SD, entropy, mpp, and skewness between the two groups. Each of the RF, LR and SVM machine learning algorithms demonstrated high AUC (AUC 0.88, 0.90 and 0.89, respectively). The high performance of the three different algorithms using the six commonly used texture features suggests that their performance is robust and does not depend on statistical methods used.

The mean texture feature, which represents the average CT attenuation value of the pixels within a ROI [30], was one of the most predictive in distinguishing benign from potentially malignant cystic renal masses. This is explained in part by the fact that the Bosniak Classification is based on morphological features, such as the number of enhancing septa, the presence of enhancing thick walls or septa, and enhancing nodules, each of which increases CT attenuation. Higher mean texture values would be expected in masses with many enhancing septa, enhancing thick walls, and one or more enhancing nodules, all features of Bosniak IIF, III, and IV masses; attenuation of each of these features is higher than that of fluid.

First-order texture features were selected rather than second or higher order texture features because relative to second order features, they are easy to implement [44], and have been shown to demonstrate lower variability [45]. Since reproducibility is a known challenge with texture analysis [29], we sought to mitigate variability using first-order

features that were provided by commercially available software (TexRAD) rather than using home-grown software or second/higher order texture features.

There was also a significantly higher value for entropy among Bosniak IIF, III and IV masses. Entropy alone performed well in discriminating benign from potentially malignant masses, with AUC 0.87. Entropy represents the inherent irregularity in the gray level intensities of a mass [46]. Increased entropy, a measure of texture heterogeneity, would be expected in masses with many enhancing septa, enhancing thick walls, and one or more enhancing nodules. Since entropy performed well, in theory, it could be used alone to distinguish benign from potentially malignant cystic renal masses. However, we believe that other texture features add incremental value and help reduce reliance on a single texture feature. While none of the machine learning algorithms performed statistically significantly better than entropy alone, each algorithm was not computationally demanding and could be applied also. In particular, the LR algorithm trended towards better performance than entropy alone, and therefore may perform better in clinical practice.

There is little prior work demonstrating the utility of machine learning algorithms for characterizing cystic renal masses. Recently, Kim et al. [33] demonstrated the ability of a machine learning algorithm to diagnose RCC among low attenuation renal masses on non-contrast CT exams using a similar CT-based texture analysis, however, their algorithm did not address cystic renal masses detected at contrast-enhanced CT. Lee et al. [47] used a Bayesian classifier to predict malignancy among cystic renal masses. However, the Bosniak features used in their study were determined by radiologists' manual review of the images. Therefore, despite showing slightly increased specificity and similar sensitivity in predicting malignancy among cystic renal masses compared to individual radiologists, their methods were prone to interreader variability. Our machine learning algorithm was applied to the Bosniak classification also but because the texture analysis was performed directly on images and did not require radiological interpretation, our method was less affected by interreader variability. Finally, a common criticism of texture analysis and machine learning models is that sometimes these are difficult to understand and reproduce. Therefore, we included features derived from commercially available texture analysis software that uses only first-order statistical based texture parameters. The software and the texture features we used have been reported in the literature [29, 33, 48–51], and in our study demonstrated high discriminatory ability with all three tested algorithms.

We found that the algorithms demonstrated high specificity and relatively lower sensitivity. This would potentially impact clinical practice in the following way. The algorithms' high specificity means that radiologists may be more confident in recommending potentially malignant

masses be evaluated further. Relatively lower sensitivity means that some potentially malignant masses may be incorrectly classified as benign. However, given the current problem of overdiagnosis and overtreatment of cystic renal masses [9], the lower sensitivity may help reduce the unnecessary evaluation of masses. Overall, the algorithms could promote evaluating masses which are likely malignant, while ignoring masses which are likely benign.

There are several limitations to our study, including its single-center, retrospective design. The number of Bosniak I and II masses (185) and Bosniak IIF, III and IV masses (72) differed from the number of masses in each group determined by the radiology report review, 150 and 107, respectively. The masses were classified in the radiology reports by subspecialized attending radiologists using the original Bosniak classification [52, 53] (necessitated by the retrospective design of the study) and subsequently selected by a radiology resident. Each mass was then classified via a three-way attending radiologist consensus using Bosniak Classification version 2019. The resultant larger proportion of Bosniak I and II masses was a goal of the revised classification.

We obtained a 257 patient cohort and performed tenfold cross validations. We could not perform validation on an entirely separate set of masses due to a practical constraint; we had a relatively the small number of Bosniak IIF, III, and IV masses in our cohort. Therefore, we applied a tenfold cross validation which is an established method to validate the performance of a machine learning model in cohorts of limited size. We plan to test the performance of the algorithms on a separate, larger cohort of masses in the future.

Another limitation was that the texture analysis was based on a single axial CT image. Although the image was chosen to portray the feature associated with the highest Bosniak classification, a volumetric texture analysis would be more likely to capture all features pertinent to the Bosniak classification. However, drawing the ROIs around each image and the computations necessitated by such a machine learning algorithm would be time consuming, more computationally challenging, and thus not currently feasible for everyday clinical practice. We believe that the use of a single image that demonstrated the highest Bosniak class was a reasonable approach, and ultimately showed high discriminatory value in distinguishing benign cystic renal masses from potentially malignant ones. Future work could compare single image and volumetric analyses. A related limitation regarding texture analysis is that a single radiologist placed an ROI over the entire mass, and not a specific region that encompassed a feature described in the Bosniak classification (e.g., thick septa, enhancing nodule). We believe that using a standard ROI that encompassed the entire mass minimized the interreader variability that would result from having to select specific regions within each mass.

This study could represent the first of several steps in use of a CT texture-based model for cystic renal mass characterization. While our study used common texture features available in a commercial software to allow for higher reproducibility across different sites, future work will employ deep learning to assess the discriminatory potential of a multitude of higher order texture features. This CT texture-based technique could also be applied to pathological outcomes instead of Bosniak classification to determine if a lesion is benign or malignant.

In summary, a CT texture-based machine learning algorithm demonstrated high discriminatory capability in stratifying cystic renal masses as benign (Bosniak I, II) from potentially malignant (Bosniak IIF, III, IV), and if validated, may aid in reducing the interreader variability in characterizing cystic renal masses. Since nephrographic phase images, as opposed to non-contrast and excretory phase images, most closely resemble portal venous phase images, future studies could attempt to validate this algorithm on portal venous phase CT scans on which many renal masses are often initially detected.

References

1. Terada N, Arai Y, Kinukawa N, Yoshimura K, Terai A (2004) Risk factors for renal cysts. *BJU Int* 93:1300–1302. <https://doi.org/10.1111/j.1464-410X.2004.04844.x>
2. Carrim ZI, Murchison JT (2003) The prevalence of simple renal and hepatic cysts detected by spiral computed tomography. *Clin Radiol* 58:626–629
3. Suher M, Koc E, Bayrak G (2006) Simple renal cyst prevalence in internal medicine department and concomitant diseases. *Ren Fail* 28:149–152
4. Tada S, Yamagishi J, Kobayashi H, Hata Y, Kobari T (1983) The incidence of simple renal cyst by computed tomography. *Clin Radiol* 34:437–439
5. Siegel RL, Miller KD, Jemal A (2019) Cancer statistics, 2019. *CA Cancer J Clin* 69:7–34. <https://doi.org/10.3322/caac.21551>
6. Krajewski KM, Pedrosa I (2018) Imaging Advances in the Management of Kidney Cancer. *J Clin Oncol Off J Am Soc Clin Oncol* JCO2018791236. <https://doi.org/10.1200/JCO.2018.79.1236>
7. Davies L, Petitti DB, Woo M, Lin JS (2018) Defining, Estimating, and Communicating Overdiagnosis in Cancer Screening. *Ann Intern Med* 169:824. <https://doi.org/10.7326/L18-0517>
8. Esserman LJ, Thompson IM, Reid B (2013) Overdiagnosis and overtreatment in cancer: an opportunity for improvement. *JAMA* 310:797–798. <https://doi.org/10.1001/jama.2013.108415>
9. Schoots IG, Zaccari K, Hunink MG, Verhagen PCMS (2017) Bosniak Classification for Complex Renal Cysts Reevaluated: A Systematic Review. *J Urol* 198:12–21. <https://doi.org/10.1016/j.juro.2016.09.160>
10. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C (2004) Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* 351:1296–1305. <https://doi.org/10.1056/NEJMoa041031>
11. Sun M, Thuret R, Abdollah F, Lughezzani G, Schmitges J, Tian Z, Shariat SF, Montorsi F, Patard J-J, Perrotte P, Karakiewicz PI (2011) Age-adjusted incidence, mortality, and survival rates of stage-specific renal cell carcinoma in North America: a trend analysis. *Eur Urol* 59:135–141. <https://doi.org/10.1016/j.eururo.2010.10.029>
12. Sun M, Trinh Q-D, Bianchi M, Hansen J, Hanna N, Abdollah F, Shariat SF, Briganti A, Montorsi F, Perrotte P, Karakiewicz PI (2012) A non-cancer-related survival benefit is associated with partial nephrectomy. *Eur Urol* 61:725–731. <https://doi.org/10.1016/j.eururo.2011.11.047>
13. Tan H-J, Norton EC, Ye Z, Hafez KS, Gore JL, Miller DC (2012) Long-term survival following partial vs radical nephrectomy among older patients with early-stage kidney cancer. *JAMA* 307:1629–1635. <https://doi.org/10.1001/jama.2012.475>
14. Van Poppel H, Da Pozzo L, Albrecht W, Matveev V, Bono A, Borkowski A, Colombel M, Klotz L, Skinner E, Keane T, Marreud S, Collette S, Sylvester R (2011) A prospective, randomised EORTC intergroup phase 3 study comparing the oncologic outcome of elective nephron-sparing surgery and radical nephrectomy for low-stage renal cell carcinoma. *Eur Urol* 59:543–552. <https://doi.org/10.1016/j.eururo.2010.12.013>
15. Silverman SG, Pedrosa I, Ellis JH, Hindman NM, Schieda N, Smith AD, Remer EM, Shinagare AB, Curci NE, Raman SS, Wells SA, Kaffenberger SD, Wang ZJ, Chandarana H, Davenport MS (2019) Bosniak Classification of Cystic Renal Masses, Version 2019: An Update Proposal and Needs Assessment. *Radiology* 292:475–488. <https://doi.org/10.1148/radiol.2019182646>
16. Hu EM, Zhang A, Silverman SG, Pedrosa I, Wang ZJ, Smith AD, Chandarana H, Doshi A, Shinagare AB, Remer EM, Kaffenberger SD, Miller DC, Davenport MS (2018) Multi-institutional analysis of CT and MRI reports evaluating indeterminate renal masses: comparison to a national survey investigating desired report elements. *Abdom Radiol N Y* 43:3493–3502. <https://doi.org/10.1007/s00261-018-1609-x>
17. Hindman NM, Hecht EM, Bosniak MA (2014) Follow-up for Bosniak category 2F cystic renal lesions. *Radiology* 272:757–766. <https://doi.org/10.1148/radiol.14122908>
18. Hindman NM (2016) Cystic renal masses. *Abdom Radiol N Y* 41:1020–1034. <https://doi.org/10.1007/s00261-016-0761-4>
19. Silverman SG, Gan YU, Morteale KJ, Tuncali K, Cibas ES (2006) Renal masses in the adult patient: the role of percutaneous biopsy. *Radiology* 240:6–22. <https://doi.org/10.1148/radiol.2401050061>
20. Weibl P, Klatte T, Kollarik B, Waldert M, Schüller G, Geryk B, Remzi M (2011) Interpersonal variability and present diagnostic dilemmas in Bosniak classification system. *Scand J Urol Nephrol* 45:239–244. <https://doi.org/10.3109/00365599.2011.562233>
21. Siegel CL, McFarland EG, Brink JA, Fisher AJ, Humphrey P, Heiken JP (1997) CT of cystic renal masses: analysis of diagnostic performance and interobserver variation. *AJR Am J Roentgenol* 169:813–818. <https://doi.org/10.2214/ajr.169.3.9275902>
22. Siegel CL, Fisher AJ, Bennett HF (1999) Interobserver variability in determining enhancement of renal masses on helical CT. *AJR Am J Roentgenol* 172:1207–1212. <https://doi.org/10.2214/ajr.172.5.10227490>
23. Benjaminov O, Atri M, O'Malley M, Lobo K, Tomlinson G (2006) Enhancing component on CT to predict malignancy in cystic renal masses and interobserver agreement of different CT features. *AJR Am J Roentgenol* 186:665–672. <https://doi.org/10.2214/AJR.04.0372>
24. Kim DY, Kim JK, Min G-E, Ahn H-J (1987) Cho K-S (2010) Malignant renal cysts: diagnostic performance and strong predictors at MDCT. *Acta Radiol Stockh Swed* 51:590–598. <https://doi.org/10.3109/02841851003641826>
25. El-Mokadem I, Budak M, Pillai S, Lang S, Doull R, Goodman C, Nabi G (2014) Progression, interobserver agreement, and malignancy rate in complex renal cysts (\geq Bosniak category

- IIF). *Urol Oncol* 32:24.e21–27. <https://doi.org/10.1016/j.urolonc.2012.08.018>
26. Graumann O, Osther SS, Karstoft J, Hørlyck A (1987) Osther PJS (2015) Bosniak classification system: inter-observer and intra-observer agreement among experienced urologists. *Acta Radiol Stockh Swed* 56:374–383. <https://doi.org/10.1177/0284185114529562>
 27. Summers RM (2017) Texture analysis in radiology: Does the emperor have no clothes? *Abdom Radiol N Y* 42:342–345. <https://doi.org/10.1007/s00261-016-0950-1>
 28. Leng S, Takahashi N, Gomez Cardona D, Kitajima K, McCollough B, Li Z, Kawashima A, Leibovich BC, McCollough CH (2017) Subjective and objective heterogeneity scores for differentiating small renal masses using contrast-enhanced CT. *Abdom Radiol N Y* 42:1485–1492. <https://doi.org/10.1007/s00261-016-1014-2>
 29. Lubner MG, Stabo N, Abel EJ, Del Rio AM, Pickhardt PJ (2016) CT Textural Analysis of Large Primary Renal Cell Carcinomas: Pretreatment Tumor Heterogeneity Correlates With Histologic Findings and Clinical Outcomes. *AJR Am J Roentgenol* 207:96–105. <https://doi.org/10.2214/AJR.15.15451>
 30. Hodgdon T, McInnes MDF, Schieda N, Flood TA, Lamb L, Thornhill RE (2015) Can Quantitative CT Texture Analysis be Used to Differentiate Fat-poor Renal Angiomyolipoma from Renal Cell Carcinoma on Unenhanced CT Images? *Radiology* 276:787–796. <https://doi.org/10.1148/radiol.2015142215>
 31. Scrima AT, Lubner MG, Abel EJ, Havighurst TC, Shapiro DD, Huang W, Pickhardt PJ (2018) Texture analysis of small renal cell carcinomas at MDCT for predicting relevant histologic and protein biomarkers. *Abdom Radiol N Y*. <https://doi.org/10.1007/s00261-018-1649-2>
 32. Schieda N, Thornhill RE, Al-Subhi M, McInnes MDF, Shabana WM, van der Pol CB, Flood TA (2015) Diagnosis of Sarcomatoid Renal Cell Carcinoma With CT: Evaluation by Qualitative Imaging Features and Texture Analysis. *AJR Am J Roentgenol* 204:1013–1023. <https://doi.org/10.2214/AJR.14.13279>
 33. Kim NY, Lubner MG, Nystrom JT, Swietlik JF, Abel EJ, Havighurst TC, Silverman SG, McGahan JP, Pickhardt PJ (2019) Utility of CT Texture Analysis in Differentiating Low-Attenuation Renal Cell Carcinoma From Cysts: A Bi-Institutional Retrospective Study. *AJR Am J Roentgenol*. <https://doi.org/10.2214/AJR.19.21182>
 34. Erickson BJ, Korfiatis P, Akkus Z, Kline TL (2017) Machine Learning for Medical Imaging. *Radiogr Rev Publ Radiol Soc N Am Inc* 37:505–515. <https://doi.org/10.1148/rq.2017160130>
 35. He H, Ma Y (2013) Imbalanced learning: foundations, algorithms, and applications. John Wiley & Sons Inc, Hoboken, New Jersey
 36. James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning: with applications in R. Springer, New York
 37. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837–845
 38. Youden WJ (1950) Index for rating diagnostic tests. *Cancer* 3:32–35
 39. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. <https://doi.org/10.1186/1471-2105-12-77>
 40. Smith AD, Remer EM, Cox KL, Lieber ML, Allen BC, Shah SN, Herts BR (2012) Bosniak category IIF and III cystic renal lesions: outcomes and associations. *Radiology* 262:152–160. <https://doi.org/10.1148/radiol.11110888>
 41. Smith AD, Allen BC, Sanyal R, Carson JD, Zhang H, Williams JH, Collins C, Griswold M, Zhang X (2015) Outcomes and complications related to the management of Bosniak cystic renal lesions. *AJR Am J Roentgenol* 204:W550–556. <https://doi.org/10.2214/AJR.14.13149>
 42. Mousseian PN, Yamauchi FI, Mussi TC, Baroni RH (2017) Malignancy Rate, Histologic Grade, and Progression of Bosniak Category III and IV Complex Renal Cystic Lesions. *AJR Am J Roentgenol* 209:1285–1290. <https://doi.org/10.2214/AJR.17.18142>
 43. Han W, Qin L, Bay C, Chen X, Yu K-H, Miskin N, Li A, Xu X, Young G (2020) Deep Transfer Learning and Radiomics Feature Prediction of Survival of Patients with High-Grade Gliomas. *AJNR Am J Neuroradiol* 41:40–48. <https://doi.org/10.3174/ajnr.A6365>
 44. Varghese BA, Cen SY, Hwang DH, Duddalwar VA (2019) Texture Analysis of Imaging: What Radiologists Need to Know. *AJR Am J Roentgenol* 212:520–528. <https://doi.org/10.2214/AJR.18.20624>
 45. Foy JJ, Robinson KR, Li H, Giger ML, Al-Hallaq H, Armato SG (2018) Variation in algorithm implementation across radiomics software. *J Med Imaging Bellingham Wash* 5:044505. <https://doi.org/10.1117/1.JMI.5.4.044505>
 46. Parekh V, Jacobs MA (2016) Radiomics: a new application from established techniques. *Expert Rev Precis Med Drug Dev* 1:207–226. <https://doi.org/10.1080/23808993.2016.1164013>
 47. Lee Y, Kim N, Cho K-S, Kang S-H, Kim DY, Jung YY, Kim JK (2009) Bayesian classifier for predicting malignant renal cysts on MDCT: early clinical experience. *AJR Am J Roentgenol* 193:W106–111. <https://doi.org/10.2214/AJR.08.1858>
 48. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ (2017) CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. *Radiogr Rev Publ Radiol Soc N Am Inc* 37:1483–1503. <https://doi.org/10.1148/rq.2017170056>
 49. Zhang G-M-Y, Shi B, Xue H-D, Ganeshan B, Sun H, Jin Z-Y (2019) Can quantitative CT texture analysis be used to differentiate subtypes of renal cell carcinoma? *Clin Radiol* 74:287–294. <https://doi.org/10.1016/j.crad.2018.11.009>
 50. Bektas CT, Kocak B, Yardimci AH, Turkcanoglu MH, Yucetas U, Koca SB, Erdim C, Kilickesmez O (2019) Clear Cell Renal Cell Carcinoma: Machine Learning-Based Quantitative Computed Tomography Texture Analysis for Prediction of Fuhrman Nuclear Grade. *Eur Radiol* 29:1153–1163. <https://doi.org/10.1007/s00330-018-5698-2>
 51. Haider MA, Vosough A, Khalvati F, Kiss A, Ganeshan B, Bjarnason GA (2017) CT texture analysis: a potential tool for prediction of survival in patients with metastatic clear cell carcinoma treated with sunitinib. *Cancer Imaging Off Publ Int Cancer Imaging Soc* 17:4. <https://doi.org/10.1186/s40644-017-0106-8>
 52. Israel GM, Bosniak MA (2005) An update of the Bosniak renal cyst classification system. *Urology* 66:484–488. <https://doi.org/10.1016/j.urology.2005.04.003>
 53. Israel GM, Bosniak MA (2005) How I do it: evaluating renal masses. *Radiology* 236:441–450. <https://doi.org/10.1148/radiol.12362040218>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.