Abdominal Radiology (2021) 46:2097–2106
https://doi.org/10.1007/s00261-020-02865-0

KIDNEYS, URETERS, BLADDER, RETROPERITONEUM

Prediction of burden and management of renal calculi from whole kidney radiomics: a multicenter study

Fatemeh Homayounieh[1] · Ruhani Doda Khera[1] · Bernardo Canedo Bizzo[1] · Shadi Ebrahimian[1] · Andrew Primak[2] · Bernhard Schmidt[3] · Sanjay Saini[1] · Mannudeep K. Kalra[1]

Received: 9 September 2020 / Revised: 6 November 2020 / Accepted: 11 November 2020 / Published online: 26 November 2020

Abstract

Purpose To assess if autosegmentation-assisted radiomics can predict disease burden, hydronephrosis, and treatment strategies in patients with renal calculi.

Methods The local ethical committee-approved, retrospective study included 202 adult patients (mean age: 53 ± 17 years; male: 103; female: 99) who underwent clinically indicated, non-contrast abdomen-pelvis CT for suspected or known renal calculi. All CT examinations were reviewed to determine the presence ($n = 123$ patients) or absence ($n = 79$) of renal calculi. On CT images with renal calculi, each kidney stone was annotated and measured (maximum dimension, Hounsfield unit (HU), and combined and dominant stone volumes) using a HU threshold-based segmentation. We recorded the presence of hydronephrosis, number of renal calculi, and treatment strategies. Deidentified CT images were processed with the radiomics prototype (Radiomics, Frontier, Siemens Healthineers), which automatically segmented each kidney to obtain 1690 first-, shape-, and higher-order radiomics. Data were analyzed using multiple logistic regression analysis with areas under the curve (AUC) as output.

Results Among 202 patients, only 28 patients (18%) needed procedural treatment (lithotripsy or ureteroscopic stone extraction). Gray-level co-occurrence matrix (GLCM) and gray-level run length matrix (GLRLM) differentiated patients with and without procedural treatment (AUC 0.91, 95% CI 0.85–0.92). Higher-order radiomics (gray-level size zone matrix – GLSZM) differentiated kidneys with and without hydronephrosis (AUC: 0.99, $p < 0.001$) as well those with different stone volumes (AUC up to 0.89, 95% CI 0.89–0.92).

Conclusion Automated segmentation and radiomics of entire kidneys can assess hydronephrosis presence, stone burden, and treatment strategies for renal calculi with AUCs > 0.85.

Keywords Renal calculi · Radiomics · CT · Disease burden · Hydronephrosis

Abbreviations

CT	Computed tomography
IRB	Institutional ethical board

Sanjay Saini and Mannudeep K. Kalra have contributed equally to this work.

✉ Fatemeh Homayounieh
fhomayounieh@mgh.harvard.edu

Ruhani Doda Khera
rdodakhera@mgh.harvard.edu

Bernardo Canedo Bizzo
bbizzo@mgh.harvard.edu

Shadi Ebrahimian
sebrahimian@mgh.harvard.edu

Andrew Primak
andrew.primak@siemens-healthineers.com

Bernhard Schmidt
bernhard.schmidt@siemens-healthineers.com

Sanjay Saini
ssaini@mgh.harvard.edu

Mannudeep K. Kalra
mkalra@mgh.harvard.edu

[1] Department of Radiology, Massachusetts General Hospital, Harvard Medical School, 75 Blossom Court, Room 248, Boston, MA 02114, USA

[2] Siemens Medical Solutions USA Inc, Malvern, PA, USA

[3] Siemens Healthcare GmbH, Forchheim, Germany

ROC	Receiver operating characteristic
AUC	Area under the curve
DICOM	Digital imaging and communication in medicine
kV	Kilovolt
GLCM	Gray-level co-occurrence matrix
GLRLM	Gray-level run length matrix
GLSZM	Gray-level size zone matrix
NGTDM	Neighboring gray-tone difference matrix
GLDM	Gray-level dependence matrix
HU	Hounsfield unit
Idmn	Inverse difference moment normalized
Imc1	Informational measure of correlation 1
Imc2	Informational measure of correlation 2
Idn	Inverse difference normalized

Introduction

Prevalence statistics from the eastern and western hemispheres suggest increasing renal calculi burden over the last three decades [1–4]. In China, the prevalence rate of renal calculi rose from 5.95% in 1991–2000 to 10.63% in 2011–2016 [1]. The population-adjusted stone prevalence rate in the United States also increased to 10.3% in 2007–2010 [2]. England reported a 20% increase in kidney stones over 7 years between 2006–2007 and 2013–2014 [3]. The rising burden of kidney stone increased intervention treatments and the financial costs associated with 149%, 86%, and 26% increase in the rate of percutaneous lithotripsy, ureteroscopy, and shockwave lithotripsy [3]. Data suggest that kidney stone is a high prevalence disease with costs comparable to the combined cost of prostate and bladder cancer in the United Kingdom [5]. Analysis of the United States data found 2 million healthcare visits and 177,000 hospitalizations for kidney stones responsible for more than $2-billion burden in the year 2000 [6]. Furthermore, multivariable analyses suggest that obesity, diabetes, dietary habits, and lifestyle changes are strongly associated with kidney stone disease [6, 7].

CT is often considered the imaging modality of choice for patients with renal colic and calculi. CT assessment of presence, burden, and complications of renal calculi enables physicians to determine treatment strategies for a condition that can result in what many patients describe as the worst pain they have ever experienced. There is a growing consensus that abdomen-pelvis CT for primary evaluation of renal calculi should be performed with low radiation dose technique (< 3 mSv) [8–10].

Typical radiological interpretation of renal calculi involves qualitative description along with the measurement of their largest diameter [11]. Recent studies have reported on the role of stone-specific radiomics and artificial intelligence (AI)-based algorithms for detection, differentiation (from phleboliths and between different types of calculi), and burden of renal calculi [12–15]. To our best knowledge, evaluation of the stone disease burden and its effect on prediction presence of obstructive uropathy and treatment strategies has not been assessed. We hypothesized that automatically segmented whole kidney radiomics could assess the stone burden and predict hydronephrosis and treatment strategies from CT images. The purpose of our study was to determine if autosegmentation-assisted radiomics can predict disease burden, hydronephrosis, and treatment strategies in patients with renal calculi.

Methods

Approvals and Disclosures

Before project initiation, we obtained approval from the human research committee with a waiver of written informed consent due to the retrospective nature of the study. Per regulations, the study was Health Insurance Portability and Accountability Act (HIPAA) compliant. All CT examinations were deidentified and coded to protect patient data privacy. Two study coauthors are employees of Siemens Healthineers (AP, BS); they did not participate in subject recruitment, data collection, or data analyses parts of the study. A study coauthor (MKK) has received institutional research grants from Siemens Healthineers and personal remuneration from Globus Medical Imaging for unrelated work.

Patients and Stone Measurements

We reviewed our radiology archive and electronic medical record (EPIC, EPIC System Corporation) to identify 238 adult patients (> 18-years) who underwent clinically indicated, unenhanced abdomen-pelvis CT examinations for renal colic or calculi between May and June 2019. The exclusion criteria were the presence of ureteric and/or vesical calculi without calculus within their kidneys ($n = 29$ patients), large renal cysts (> 3 cm) ($n = 4$ patients), in situ pigtail or nephrostomy ($n = 2$ patients), and partial nephrectomy ($n = 1$ patients). The final sample size was 202 adult patients (mean age 53 ± 17 years; 103 males and 99 females) (Fig. 1). All patients were scanned in one of the three hospitals—Massachusetts General Hospital (Site A, $n = 132$ patients), Brigham and Women's Hospital (Site B, $n = 41$), or Newton-Wellesley Hospital (Site C, $n = 29$). Of these, 123 abdomen-pelvis CT demonstrated renal calculi, whereas the remaining 79 patients did not have renal calculi.

In addition to each patient's age and gender, we recorded medical and invasive therapies (such as laser lithotripsy,

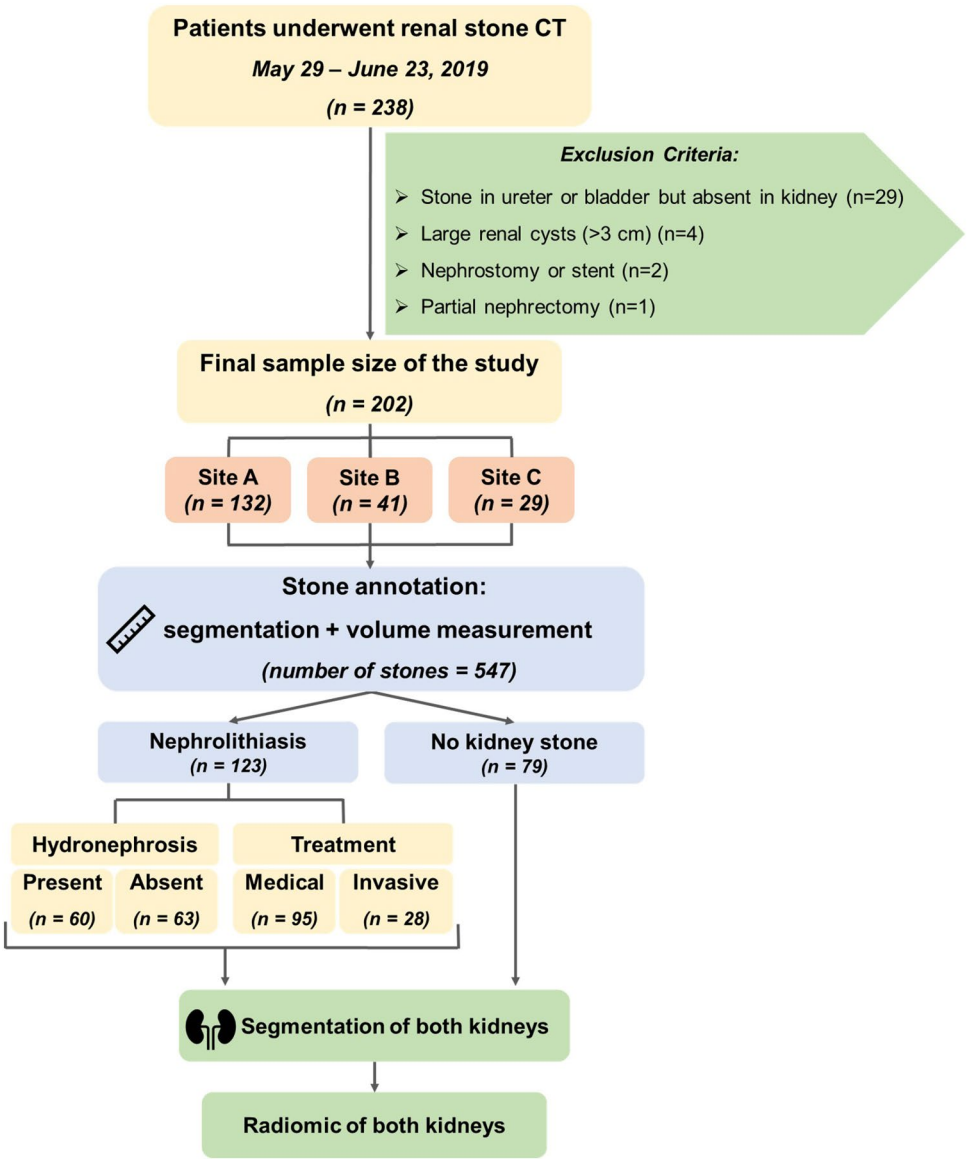

Fig. 1 Flow diagram summary of patient selection, exclusion criteria, and distribution of patients with renal calculi, hydronephrosis, and invasive procedural treatment

extracorporeal shockwave therapy, interventional, and surgical stone removal) within 10 months of abdomen-pelvis CT exam.

Renal stone protocol CT

All abdomen-pelvis CT examinations for evaluation of renal colic or calculi were performed on one of the three multidetector-row CT (MDCT) scanners: 96-detector-row, third-generation, dual-source MDCT (Siemens Force, Siemens Healthineers, Forchheim, Germany; $n = 106$ patients), 256-detector-row single-source MDCT (GE Revolution, GE Healthcare, Waukesha, Wis.; $n = 83$ patients), and 64-detector-row single-source MDCT (Philips IQon Spectral CT, Philips Healthcare, Eindhoven, The Netherlands; $n = 13$ patients). The following scan parameters were used on all CT scanners: 100–120 kV, automatic exposure control (Smart mA, GE; 3D Modulation, Philips; Care Dose 4D, Siemens), 0.5-s gantry rotation time, and 0.9–1:1 pitch. Reconstructed section thickness was 1–1.25 mm with a 0.8–1 mm overlap and soft tissue reconstruction algorithm using vendor-specific iterative reconstruction techniques (ASIRv, GE; iDose 4, Philips; ADMIRE, Siemens).

Qualitative evaluation

A senior radiologist (MKK, with more than 15-year experience) reviewed all abdomen-pelvis CT images for the presence of renal calculi and hydronephrosis on a PACS workstation (Visage 7.1 Enterprise Imaging Platform, Visage Imaging GmbH, Berlin, Germany). Upon detection, the radiologist clicked in the center of each renal calculus, which

triggered the platform to segment and annotate the calculus using a region growing feature based on the isolation of CT voxels with attenuation values (Hounsfield Unit, HU) similar to the center of the calculus. The radiologist reviewed each segmented contour and edited the segmentation margins if necessary (in < 20% of the cases). The radiologist also selected side of each calculus and the absence or presence of hydronephrosis. The platform separately estimated the volume of each calculus from the number and size of segmented voxels. The dominant calculus was defined as the calculus with the largest volume. Stone burden represented the volume of all calculi in each kidney.

Radiomics prototype

Another study coauthor (FH, a physician with 2-year postdoctoral research experience in image processing) separately processed all 202 abdomen-pelvis CT examinations from a standalone radiomics prototype (FRONTIER, Siemens Healthineers). The prototype automatically identifies and segments the entire kidney volume with a single click on the Kidney Isolation icon (Fig. 2). The segmented renal volume included the entire renal parenchyma and the intrarenal portion of the pelvicalyceal system. When present, all renal calculi were included within the segmentation contours. The user can modify the isolated contours if needed. We did not need to correct any segmented renal contour in our study. Upon confirmation of the contours, the prototype estimated 1690 first-, shape, and higher-order radiomics for each kidney. The estimated radiomics are described elsewhere (https://pyradiomics.readthedocs.io/en/latest/changes.html, accessed on February 22, 2020). These included first-order, shape, gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), neighboring gray-tone difference matrix (NGTDM), gray-level dependence matrix (GLDM), as well as squares, square roots, logarithm, exponentials, Log with five (0–4 mm) sigma level, and 3D wavelet transform (multidimensional decomposition of image—multidimensional signal processing) of radiomics other than the shape features.

Statistical analysis

All statistical analyses were performed with the statistical program (R Statistical Computing: https://www.R-project.org, R Foundation for Statistical Computing, Vienna, Austria, accessed on 3.18.2020) incorporated within the radiomics prototype used in our study. Data were analyzed on both per patient and per kidney levels to assess the ability of radiomics to predict stone burden, hydronephrosis, and treatment method using multiple logistic regression analyses. For the most suitable multiple logistic regression model, the prototype uses *t* test or analyses of variance (ANOVA) and corrects the *p* values for multiple testing with the Benjamini–Hochberg false discovery rate (FDR). On the set of statistically significant features with corrected $p < 0.05$, a

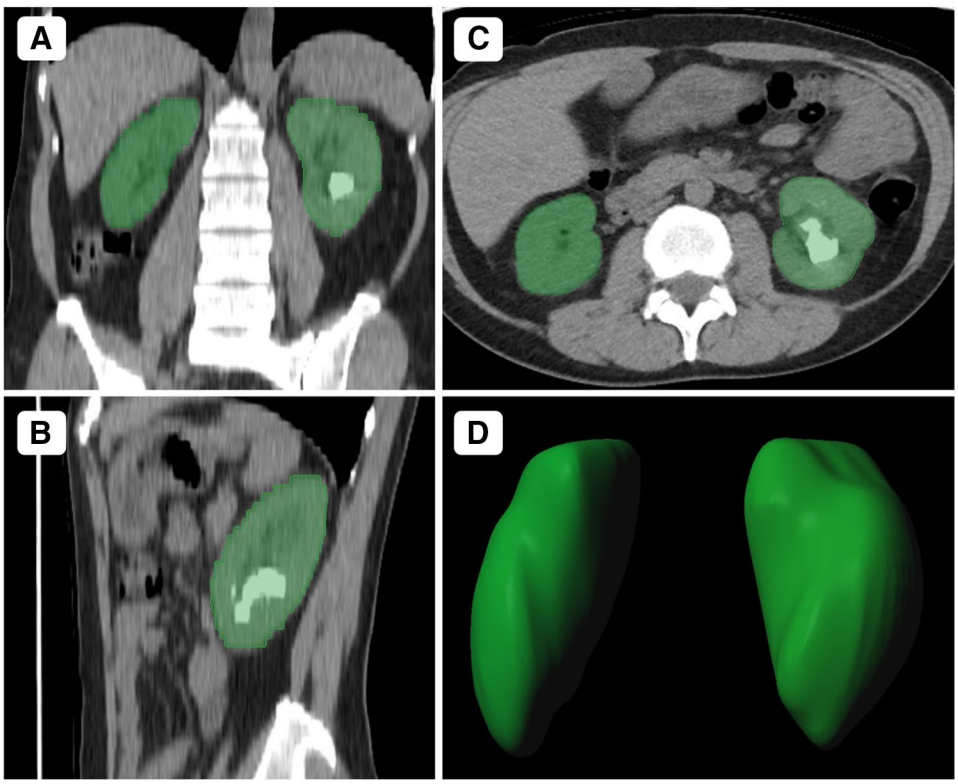

Fig. 2 Non-contrast abdomen-pelvis CT of a 50-year-old female with a staghorn left renal calculus. Multiplanar reformat (**a**–**c**) and volume rendered (**d**) images demonstrate accurate volumetric autosegmentation of both kidneys

minimum redundancy maximum relevance (mRMR) feature selection helps eliminate irrelevant and redundant features and limit selection to four features.

The receiver operating characteristics area under the curve (AUC) was set as the output parameter with both 95% confidence interval (CI) applied and *p* values.

Volumes of renal calculi were classified into four quartiles to assess if radiomics can differentiate between adjacent quartiles of stone burden using analysis of variance (ANOVA). The inter-CT vendor comparison was limited to two vendors (GE and Siemens) since very few patients were scanned on the third scanner (Philips). In addition to the independent analysis of the entire and the individual hospital site data, we applied a machine learning (ML)-based random forest classifier (with 10-fold cross validation using 100 trees and a split quality measure of Gini impurity for identifying most relevant radiomics) to assess predictive value of radiomics at Sites B and C when Site A data were used as a training dataset. When comparing performance across the three healthcare sites, we did not apply any specific selection method for selecting radiomic features. For multiple statistical testing, false discovery rate (FDR) Benjamini–Hochberg's correction was applied to the *p* values. Radiomics with the highest AUC and *p* values < 0.05 were deemed as the most important predictors.

Results

Detection of renal calculi

Out of 202 patients, 79 patients (39%) did not have any renal calculi. Among 123 patients (61%, 123/202) with a total of 547 renal calculi, 61 patients (50%, 61/123) had bilateral calculi while remaining patients (left: 28%, 35/123; right: 22%, 27/123) had unilateral calculi. Thus, renal calculi were present in 184/404 segmented kidneys.

Among the 202 patients from all three sites, radiomics (a combination of 3D log sigma of short-run low gray-level emphasis, exponential of run variance, run entropy (GLRLM features), and GLCM maximal correlation coefficient) could differentiate patients with and without renal calculi (AUC 0.84, 95% CI 0.78–0.89, $p < 0.003$). The AUC (AUC: 0.9, 95% CI 0.85–0.93, $p < 0.0001$) was higher for detecting renal calculi at the kidney level with a combined inclusion of GLCM inverse difference moment normalized, NGTDM exponential of coarseness, and GLRLM 3D log sigma of short-run low gray-level emphasis.

The ability of radiomics to differentiate kidneys with or without calculi did not change based on the CT vendor (GE CT: AUC 0.92, 95% CI 0.87–0.97; Siemens CT: AUC 0.93, 95% CI 0.9–0.95; $p > 0.9$) (Table 1). Likewise, radiomics did not change substantially across the three hospitals for the detection of renal calculi (AUC 0.88–0.93, $p > 0.8$) (Table 2). When site A was used as a training set, the ML-based assessment had AUCs of 0.85 (site B) and 0.91 (site C) for differentiating kidneys with and without calculi ($p > 0.8$).

Presence of hydronephrosis

Of the 184 kidneys with calculi, hydronephrosis was present in 61 kidneys (right kidney—30; left kidney—29; both kidneys—1). There was no significant difference ($p = 0.78$) between the volume of dominant renal calculi in patients with (104 ± 186 mm^3) and without (123 ± 512 mm^3) hydronephrosis. AUCs for the differentiation of patients with hydronephrosis with radiomics are summarized in Table 3. Among 404 kidneys with and without renal calculi, radiomics enabled detection of hydronephrosis with an AUC of 0.89 (95% CI 0.8–0.89, $p < 0.003$) with a combination of higher-order statistics (Table 3). For 184 kidneys with calculi, the AUC (AUC 0.85, 95% CI 0.77–0.87, $p < 0.006$) had a non-significant decrease for differentiating kidneys with and without hydronephrosis ($p = 0.6$). Figure 3 demonstrates most relevant features for differentiating patients with hydronephrosis.

Detection of hydronephrosis with radiomics did not change between scanners from different vendors (GE CT: AUC 0.91, 95% CI 0.79–0.92; Siemens CT: AUC 0.9,

Table 1 Performance of radiomics on scanners from two CT vendors

	GE CT scanners		Siemens CT scanners	
	AUC	95% CI	AUC	95% CI
Patient management (in all 202 patients)	0.94	0.87–0.97	0.93	0.79–0.95
Hydronephrosis (in all 202 patients)	0.91	0.79–0.92	0.9	0.81–0.92
Hydronephrosis (in 123 patients with calculi)	0.86	0.77–0.89	0.86	0.81–0.92
Renal calculi detection (in all 404 kidneys)	0.92	0.87–0.97	0.93	0.9–0.95
Median volume of calculi (in all 404 kidneys)	0.95	0.9–0.99	0.93	0.91–0.97

The table summarizes the best area under the curves (AUC) of radiomics for predicting invasive procedural management of renal calculi, presence of hydronephrosis, presence of renal calculi, and median volume of renal calculi (Key: *CI* confidence interval)

Table 2 Performance of radiomics at the three independent participating sites (Sites A, B, C) and with machine learning-based radiomics at two sites (ML-Site B and ML-Site C)

	AUC (95% CI)			Machine learning	
	Site A	Site B	Site C	ML-Site B	ML-Site C
Patient management (in all 202 patients)	0.93 (0.84–0.95)	0.99 (0.99–1.0)	0.96 (0.84–0.96)	0.87	0.8
Hydronephrosis (in all 202 patients)	0.88 (0.79–0.9)	0.9 (0.74–0.95)	0.996 (0.94–1.0)	0.72	0.84
Hydronephrosis (in 123 patients with calculi)	0.87 (0.78–0.89)	0.85 (0.75–0.98)	0.99 (0.95–1.0)	0.71	0.8
Renal calculi detection (in all 404 kidneys)	0.93 (0.87–0.95)	0.88 (0.87–0.93)	0.93 (0.89–0.98)	0.85	0.91
Median volume of calculi (in all 404 kidneys)	0.96 (0.86–0.98)	0.99 (0.98–0.99)	0.99 (0.92–1.0)	0.93	0.88

The table summarizes the best area under the curves (AUC) of radiomics for predicting invasive procedural management of renal calculi, presence of hydronephrosis, presence of renal calculi, and median volume of renal calculi (Key: *CI* confidence interval)

Table 3 Summary of the best radiomics for predicting invasive procedural management of renal calculi, presence of hydronephrosis, presence of renal calculi, and median volume of renal calculi

	Best features from multiple logistic regression	AUC	95% CI
Patient management (in all 202 patients)	Imc2 (GLCM) + Imc1 (GLCM) + Minimum (1st Order) + Short-run low gray-level emphasis (GLRLM)	0.91	0.85–0.92
Hydronephrosis (in all 202 patients)	Dependence non-uniformity (GLDM) + Small area emphasis (GLSZM) + Size zone non-uniformity (GLSZM)	0.89	0.8–0.89
Hydronephrosis (in 123 patients with calculi)	Dependence non-uniformity (GLDM) + Small area emphasis (GLSZM) + Dependence non-uniformity normalized (GLDM) + Size zone non-uniformity (GLSZM)	0.85	0.77–0.87
Renal calculi detection (in all 202 patients)	Short-run low gray-level emphasis (GLRLM) + Run variance (GLRLM) + Run entropy (GLRLM) + MCC (GLCM)	0.84	0.78–0.89
Renal calculi detection (in all 404 kidneys)	Idmn (GLCM) + Coarseness (NGTDM) + Short-run low gray-level emphasis (GLRLM)	0.9	0.85–0.93
Median volume of calculi (in all 404 kidneys)	Idn (GLCM) + Sum entropy (GLCM)	0.93	0.9–0.95

AUC area under the curve, *CI* confidence interval, *GLCM* gray-level co-occurrence matrix, *GLRLM* gray-level run length matrix, *GLSZM* gray-level size zone matrix, *GLDM* gray-level dependence matrix, *NGTDM* neighboring gray-tone difference matrix, *Idmn* inverse difference moment normalized, *Imc1* informational measure of correlation 1, *Imc2* informational measure of correlation 2, *Idn* inverse difference normalized

95% CI 0.81–0.92; $p > 0.9$) (Table 1). Likewise, there was no significant difference in performance of radiomics for detecting hydronephrosis in patients from different hospitals (Site A, AUC 0.88, 95% CI 0.79–0.9; Site B, AUC 0.9, 95% CI 0.74–0.95; Site C 0.996, 95% CI 0.94–1.0). With ML-based random forest classifier, the respective AUCs at site B and C were 0.72 and 0.84 for detection of hydronephrosis (Table 2).

Patient management

Of the 202 patients, 28 patients had invasive procedural treatment for renal calculi (25 patients with lithotripsy and three patients with flexible ureteroscopy). The remaining patients were managed with conservative medical treatment. Radiomics predicted patients managed with invasive procedural treatment therapy with the best differentiating features represented by a combination of the exponential of the first-order minimum, GLRLM 3D log sigma of short-run low gray-level emphasis, and GLCM inverse difference moment normalized 1 and 2 (AUC 0.91, 95% CI 0.85–0.92, $p < 0.02$). There was no difference in performance of radiomics in predicting invasive treatment in those scanned with different CT vendors (GE CT: AUC 0.94, 95% CI 0.87–0.97; Siemens CT: AUC 0.93, 95% CI 0.79–0.95 or at different hospitals (Site A, AUC 0.93, 95% CI 0.84–0.95; Site B, AUC 0.99, 95% CI 0.99–1.0; Site C 0.96, 95% CI 0.84–0.96) (Table 1). The respective AUCs for ML-based prediction of patient management were 0.87 and 0.80 at sites B and C.

Stone burden

The average volume of dominant renal calculi was 158 ± 468 mm^3 with a skewed preponderance of kidneys with larger calculi (147/184 kidneys (80%) with calculi volume > 158 mm^3 and 37/184 kidneys (20%) with calculi volume < 158 mm^3). The quartile volumes of renal calculi were < 12.7 mm^3 (1st quartile), 12.7–33.75 mm^3 (2nd quartile), 33.75–101.1 mm^3 (3rd quartile), and > 101.1 mm^3 (4th quartile). Radiomics differentiated between adjacent

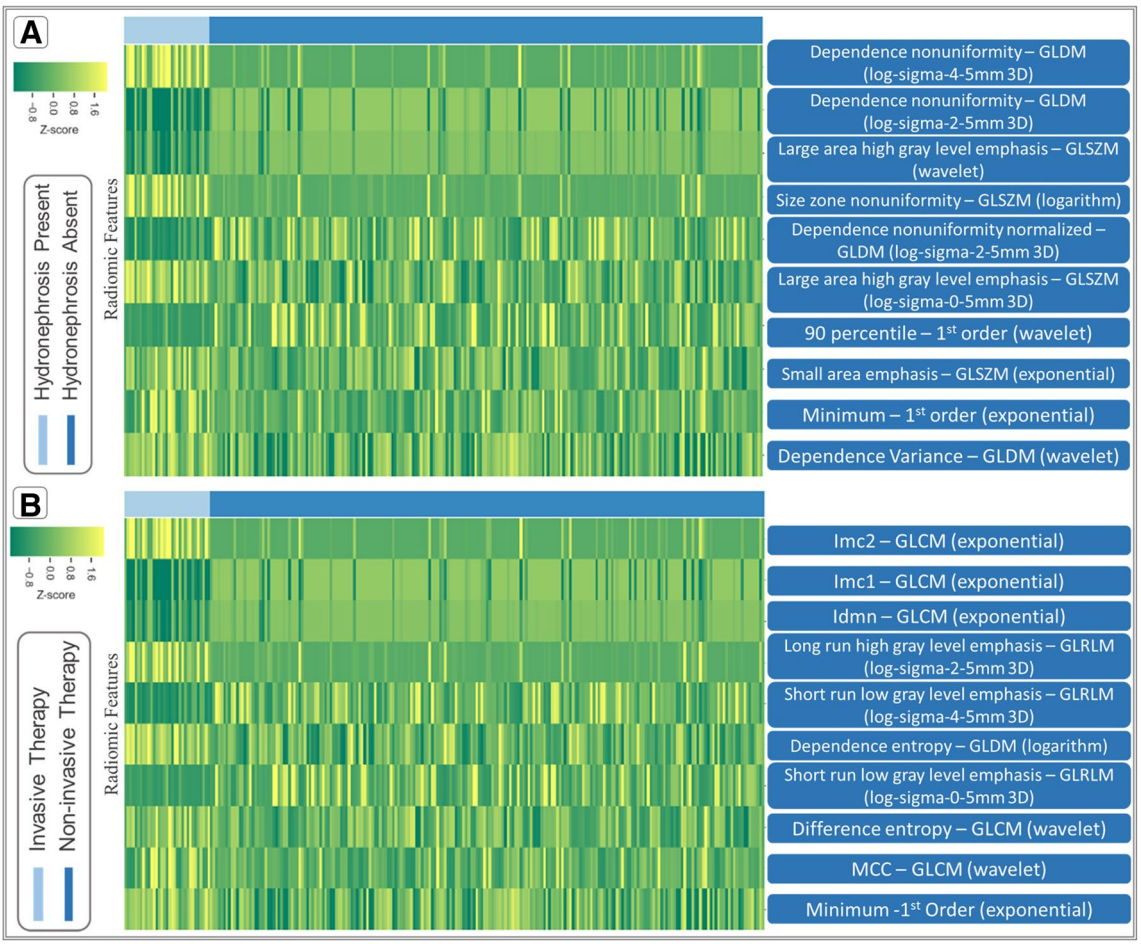

Fig. 3 Most relevant features for differentiating presence of hydronephrosis in those with renal calculi (**a**) and predicting the need of invasive procedural management (**b**)

quartiles of renal calculi volume (1st and 2nd quartiles: AUC 0.84, 95% CI 0.84–0.86; 2nd and 3rd quartiles: AUC 0.88, 95% CI 0.77–0.90; 3rd and 4th quartiles: AUC 0.89, 95% CI 0.89–0.92). The most significantly different radiomics between the adjacent quartiles estimated with ANOVA were first-order maximum, range, NGTDM, GLCM Idn, and GLCM Idmn (all features, $p < 0.0001$).

The renal calculi in patients with invasive treatment were significantly larger than those managed conservatively (641 ± 1090 vs. 53 ± 78 mm^3, $p < 0.0001$).

The performance of radiomics did not vary based on CT vendors (GE CT: AUC 0.95, 95% CI 0.9–0.99; Siemens CT: AUC 0.93, 95% CI 0.91–0.97) or at different hospitals (Site A, AUC 0.96, 95% CI 0.86–0.98; Site B, AUC 0.99, 95% CI 0.98–0.99; Site C 0.99, 95% CI 0.92–1) (Table 1). The application of ML-based radiomics for differentiating patients with median volume of renal calculi had AUCs of 0.93 (site B) and 0.88 (site C).

Discussion

The robust performance of radiomics for assessing renal calculi in our study (AUC 0.8–0.95) is comparable to prior publications on radiomics [12, 13]. De Perrot et al. used a machine learning-based radiomics model for differentiating renal calculi from phleboliths [12]. A senior radiologist identified and segmented the region of interest with either finding. Their model then estimated radiomics over the identified calculi or phlebolith and obtained an overall accuracy of 0.85 and AUC of 0.9. In another study, Cui et al. used a region-based image segmentation technique to extract regions of interest over renal calculi [13]. A radiologist verified these regions before the application of radiomics over the segmented calculi. The authors reported that radiomics differentiated infection stones from non-infection stones with an accuracy of 90.7% [13].

Instead of estimating radiomics with manual or region-based segmentation of renal calculi, we automated the isolation of kidneys and applied radiomics over the entire renal volume to predict hydronephrosis, stone burden, and invasive treatment of renal calculi. Cui et al. have reported on the use of CT texture analysis for predicting the ease of stone fragmentation with extracorporeal shockwave lithotripsy [14]. In their ex vivo study, mean CT attenuation number ($r^2 = 0.806$, $p = 0.028$) and entropy ($r^2 = 0.804$, $p = 0.04$) were strongly correlated with the number of shocks required for stone fragmentation [14]. Xun et al. reported that using machine learning-based radiomics limited to renal calculi achieved an AUC of 0.947 for predicting a stone-free rate of flexible ureteroscopy for stone removal in an internal validation set of 43 patients [15]. Mannil et al. have also reported an AUC of 0.81 with combination of three-dimensional radiomics and stone size for predicting success of shockwave lithotripsy in patients with renal calculi [16]. Prediction of calculi managed with lithotripsy or ureteroscopic stone removal with whole renal volume radiomics in our study was comparable to prior studies (AUC 0.91) [14, 15].

The automatic segmentation and inclusion of the entire renal volume enabled us to apply radiomics beyond renal calculi to the entire renal volume and obtain a reliable and generalizable prediction on stone burden and need for invasive treatment procedures. Other studies beyond kidneys have also reported on organ radiomics for assessing pancreatic ductal adenocarcinoma from entire pancreas [17], diffuse liver diseases from whole liver [18], COVID-19 pneumonia from whole lung [19, 20], and white matter hyperintensities from whole brain [21]. Due to the need for efficient image interpretation and the well-known complexity of radiomics (1690 features in our prototype), autosegmentation of entire organ-based regions of interest as well as estimation and analyses of radiomics is essential for bringing them into the clinical workflow. Our prototype achieved both these steps with minimal user input.

Although detecting clinically important renal calculi is usually not an interpretation dilemma, based on the number and shape of renal calculi, both manual and subjective assessment of their size or dimension can be time consuming and prone to intra- and inter-observer variations [22]. The assessed prototype addresses these variations and inefficiency in the evaluation of renal calculi. Although the volume of renal calculi was not related to the presence of hydronephrosis, whole kidney radiomics had a robust performance likely from the inclusion of radiomics from regions beyond the renal calculi. A high AUC (> 0.8) for radiomics in patients from all three participating sites and abdomen-pelvis CT performed with three different scanner types supports the generalizability of radiomics in assessing renal calculi. With a phantom study and extracted renal calculi, Moen et al. reported that radiomics can reliably predict stone fragility despite changes in acquisition (single energy or dual energy modes) or reconstruction parameters (section thickness and reconstruction techniques) [23]. Generalizability is critical as prior oncologic studies have voiced concerns that variations in scanners and acquisition parameters can limit application of radiomics [24–26].

With an automated segmentation of renal contours, extraction of radiomics, and embedded statistical analyses, radiomics prototypes like the one used in our study can be most useful in assessing burden of kidney stones in a more quantitative and reproducible manner. This can help determine the suitable treatment pathway for the patient as well as serve as a useful method of following patients for changes in stone burden over time.

Our study has limitations. First, there was an unequal distribution of patients across different participating sites and scanners due to differences in patient volumes at these sites. Second, the scan protocols and practices at the three sites are independent, but all sites are in the same metropolitan region, and therefore, likely share similar patient attributes. Third, we did not include ureteral and bladder calculi within this study since the applied prototype does not perform the segmentation of these structures. Fourth, we evaluated the radiomics of the entire kidneys rather than the calculi alone. Although evaluation of entire renal volume did limit our ability to extract specific radiomics and dimensions of renal calculi, this did not limit the ability to predict stone burden and outcomes in our patient group. Due to the lack of targeted radiomics limited to the calculi's volume rather than the entire renal volume, we cannot comment upon the relative performance of radiomics from the calculi versus the entire kidneys.

The fifth limitation of our study cohort pertains to the fact there was few patients ($n = 28$ patients) who had invasive treatment of renal calculi. Sixth, we used volumetric estimation of renal calculi to classify the stone burden instead of a two-dimensional measure of renal calculi, which are often used to determine the need and type of invasive treatment. However, prior studies have reported that volumetric estimation of lesions is a better and more reliable predictor of its size than unidimensional measures [22]. Seventh, we excluded patients with focal renal lesions (except hydronephrosis). Therefore, it is not possible to conclude on the performance of radiomics in the presence of other focal and diffuse renal lesions. It is likely but unproven from our study that prediction of hydronephrosis will be limited in the presence of large renal cysts, which are frequent in older patients and those on dialysis. Finally, none of the patients included in our study had abnormal or anomalous renal location, shape, or fusion. It is not unlikely that the performance of the autosegmentation feature varies with such anatomic variations.

In summary, the machine learning-based autosegmentation and radiomics can help detect and quantify renal calculi burden, assess the presence of hydronephrosis, and predict the need for invasive treatment options. The performance of radiomics was robust across all three participating sites and two CT vendors.

Funding Not applicable.

Data availability Upon our institution's policy, we cannot share patients' data.

Compliance with ethical standards

Conflict of interest Two study coauthors are employees of Siemens Healthineers (AP, BS), and they ensured veracity of technical description of the prototype; they did not participate in subject recruitment, data collection, or data analyses parts of the study. A study coauthor (MKK) has received institutional research grants from Siemens Healthineers and personal remuneration from Globus Medical Imaging for unrelated work.

Disclosure Fatemeh Homayounieh, Ruhani Doda Khera, Bernardo Canedo Bizzo, Shadi Ebrahimian, and Sanjay Saini: Nothing to disclose. Andrew Primak and Bernhard Schmidt: Employee of Siemens Healthineers. Mannudeep K. Kalra: Received institutional research grant from Siemens Healthineers and personal remuneration from Globus Medical Imaging for unrelated work.

Ethical approval IRB approved.

Consent to participate Waived by IRB.

Consent for publication Approved.

References

1. Wang W, Fan J, Huang G, et al. Prevalence of kidney stones in mainland China: A systematic review. Sci Rep. 2017;7:41630. Published 2017 Jan 31. https://doi.org/10.1038/srep41630
2. Scales CD Jr, Smith AC, Hanley JM, Saigal CS; Urologic Diseases in America Project. Prevalence of kidney stones in the United States. Eur Urol. 2012;62(1):160-165.
3. Rukin NJ, Siddiqui ZA, Chedgy ECP, Somani BK. Trends in Upper Tract Stone Disease in England: Evidence from the Hospital Episodes Statistics Database. Urol Int. 2017;98(4):391-396. doi:https://doi.org/10.1159/000449510
4. Kittanamongkolchai W, Vaughan LE, Enders FT, et al. The Changing Incidence and Presentation of Urinary Stones Over 3 Decades. Mayo Clin Proc. 2018;93(3):291-299. doi:https://doi.org/10.1016/j.mayocp.2017.11.018
5. Geraghty RM, Cook P, Walker V, Somani BK. (2020), Evaluation of the economic burden of kidney stone disease in the UK: a retrospective cohort study with a mean follow-up of 19 years. BJU Int, 125: 586-594. doi:https://doi.org/10.1111/bju.14991
6. Litwin MS, Saigal CS, Yano EM, et al. Urologic diseases in America Project: analytical methods and principal findings. J Urol. 2005;173(3):933-937.
7. Fwu CW, Eggers PW, Kimmel PL, et al. Emergency department visits, use of imaging, and drugs for urolithiasis have increased in the United States. Kidney Int. 2013;83:479-86.
8. Coursey CA, Casalino DD, Remer EM, et al. ACR Appropriateness Criteria® acute onset flank pain–suspicion of stone disease. Ultrasound Q. 2012;28(3):227-233. doi:https://doi.org/10.1097/RUQ.0b013e3182625974
9. Türk C, Petřík A, Sarica K, et al. EAU Guidelines on Diagnosis and Conservative Management of Urolithiasis. Eur Urol. 2016;69:468-74.
10. Gershan V, Homayounieh F, Singh R, et al. CT protocols and radiation doses for hematuria and urinary stones: Comparing practices in 20 countries. Eur J Radiol. 2020;126:108923. doi:https://doi.org/10.1016/j.ejrad.2020.108923
11. Miller NL, Lingeman JE. Management of kidney stones. BMJ. 2007;334(7591):468-472. doi:https://doi.org/10.1136/bmj.39113.480185.80
12. De Perrot T, Hofmeister J, Burgermeister S, et al. Differentiating kidney stones from phleboliths in unenhanced low-dose computed tomography using radiomics and machine learning. Eur Radiol. 2019;29(9):4776-4782.
13. Cui X, Che F, Wang N, et al., "Preoperative Prediction of Infection Stones Using Radiomics Features from Computed Tomography," in IEEE Access, vol. 7, pp. 122675-122683, 2019.
14. Cui HW, Devlies W, Ravenscroft S, et al. CT Texture Analysis of Ex Vivo Renal Stones Predicts Ease of Fragmentation with Shockwave Lithotripsy. J Endourol. 2017;31(7):694-700. doi:https://doi.org/10.1089/end.2017.0084
15. Yang X, Mingzhen C, Ping L, et al. A Novel Clinical-Radiomics Model Preoperatively Predicted the Stone-Free Rate of Flexible Ureteroscopy Strategy in Kidney Stone Patients (January 15, 2020). Available at SSRN: https://ssrn.com/abstract=3520022 or http://dx.doi.org/10.2139/ssrn.3520022
16. Mannil M, von Spiczak J, Hermanns T, Poyet C, Alkadhi H, Fankhauser CD. Three-Dimensional Texture Analysis with Machine Learning Provides Incremental Predictive Information for Successful Shock Wave Lithotripsy in Patients with Kidney Stones. J Urol. 2018;200(4):829-836. doi:https://doi.org/10.1016/j.juro.2018.04.059
17. Chu LC, Park S, Kawamoto S, et al. Utility of CT Radiomics Features in Differentiation of Pancreatic Ductal Adenocarcinoma From Normal Pancreatic Tissue. AJR Am J Roentgenol. 2019;213(2):349-357. doi:https://doi.org/10.2214/AJR.18.20901
18. Homayounieh F, Saini S, Mostafavi L, et al. Accuracy of radiomics for differentiating diffuse liver diseases on non-contrast CT [published online ahead of print, 2020 Jun 26]. Int J Comput Assist Radiol Surg. 2020;https://doi.org/10.1007/s11548-020-02212-0. https://doi.org/10.1007/s11548-020-02212-0
19. Fu L, Li Y, Cheng A, Pang P, Shu Z. A Novel Machine Learning-derived Radiomic Signature of the Whole Lung Differentiates Stable From Progressive COVID-19 Infection: A Retrospective Cohort Study [published online ahead of print, 2020 Jun 16]. J Thorac Imaging. 2020;https://doi.org/10.1097/rti.0000000000000544. https://doi.org/10.1097/rti.0000000000000544
20. Homayounieh F, Ebrahimian S, Babaei R, et al. CT Radiomics, Radiologists and Clinical Information in Predicting Outcome of Patients with COVID-19 Pneumonia. Radiol Cardiothorac Imaging. 2020;2(4):e200322. Published 2020 Jul 23. https://doi.org/10.1148/ryct.2020200322
21. Shu Z, Xu Y, Shao Y, Pang P, Gong X. Radiomics from magnetic resonance imaging may be used to predict the progression of white matter hyperintensities and identify associated risk factors. Eur Radiol. 2020;30(6):3046-3058. doi:https://doi.org/10.1007/s00330-020-06676-1
22. Selby MG, Vrtiska TJ, Krambeck AE, et al. Quantification of asymptomatic kidney stone burden by computed

tomography for predicting future symptomatic stone events. Urology. 2015;85(1):45-50. doi:https://doi.org/10.1016/j.urology.2014.08.031

23. Moen T, Ferrero A, McCollough C. Robustness of Textural Features to Predict Stone Fragility Across Computed Tomography Acquisition and Reconstruction Parameters. Acad Radiol. 2019;26(7):885-892. doi:https://doi.org/10.1016/j.acra.2018.08.010
24. Limkin EJ, Sun R, Dercle L, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Ann Oncol. 2017;28(6):1191-1206. doi:https://doi.org/10.1093/annonc/mdx034
25. Liu Z, Wang S, Dong D, et al. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. Theranostics. 2019;9(5):1303-1322. Published 2019 Feb 12. https://doi.org/10.7150/thno.30309
26. Homayounieh F, Singh R, Nitiwarangkul C, et al. Semiautomatic Segmentation and Radiomics for Dual-Energy CT: A Pilot Study to Differentiate Benign and Malignant Hepatic Lesions. AJR Am J Roentgenol. 2020;215(2):398-405. doi:https://doi.org/10.2214/AJR.19.22164