Journal of Biosciences and Medicines, 2026, 14(11), 462-469
https://www.scirp.org/journal/jbm
ISSN Online: 2327-509X
ISSN Print: 2327-5081

# Research Progress of Radiomics in the Diagnosis and Treatment of Urolithiasis

**Xiaoxi Yu, Youkong Li***

Department of Urology, Jingzhou Hospital Affiliated to Yangtze University, Jingzhou, China
Email: *310416986@qq.com

**How to cite this paper:** Yu, X.X. and Li, Y.K. (2026) Research Progress of Radiomics in the Diagnosis and Treatment of Urolithiasis. *Journal of Biosciences and Medicines*, **14**, 462-469.
https://doi.org/10.4236/jbm.2026.1411033

**Received:** October 17, 2026
**Accepted:** November 20, 2026
**Published:** November 23, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

## Abstract

Radiomics derives high-throughput quantitative descriptors from routine imaging and, in urolithiasis, offers information that complements—rather than simply replaces—stone size and density. This review examines four clinical applications: detection and segmentation, stone-composition prediction, treatment-response prediction, and recurrence prediction. The literature is unevenly distributed: composition prediction and stone-free rate (SFR) after extracorporeal shock wave lithotripsy (ESWL) have attracted the most attention, while other treatment scenarios remain supported mainly by exploratory models. Across studies, inconsistent segmentation and reporting practices continue to limit cross-study comparison.

## Keywords

Radiomics, Urolithiasis, Stone Composition, Extracorporeal Shock Wave Lithotripsy, Retrograde Intrarenal Surgery, Percutaneous Nephrolithotomy

## 1. Introduction

Urolithiasis is among the more common conditions encountered in urology, with an estimated global prevalence of approximately 10.85% [1]. Clinical decisions still rest largely on stone size, density, and other morphological descriptors, both for anticipating composition and for judging likely treatment response. Radiomics takes a different route: it extracts high-dimensional quantitative features from medical images, applies machine-learning methods for feature selection and model construction, and thereby supplies information that is not readily captured by conventional indicators [2]. In stones, such features can reflect gray-level uniformity, local density heterogeneity, and boundary sharpness, and may therefore carry indirect signals about microstructure and compositional difference [3]. The present

*Corresponding author.

review summarises how radiomics has been applied across clinical domains in urolithiasis and highlights the design choices that account for many of the discrepancies among published findings.

## 2. Research Progress in Detection Performance, Segmentation Methods, and Stone Burden Quantification in Urolithiasis

Detection and segmentation are best viewed as upstream steps in any radiomics pipeline rather than as evidence of radiomics performance per se. A systematic review of 32 studies reported that artificial intelligence (AI) models achieved sensitivity of 88% - 98%, specificity of 85% - 97%, and AUC of 0.85 - 0.96 for stone detection [4]. A separate meta-analysis of 45 studies, 10 of which focused on non-contrast computed tomography (NCCT), found a pooled sensitivity of 96% (95% CI 92% - 98%) and a pooled specificity of 98% (95% CI 97% - 99%); external validation, however, was available in only 4 studies (1056 images), where pooled sensitivity was 95% (95% CI 92% - 97%) and the lower 95% CI bound for specificity fell to 70% [5]. These figures should not be overread as radiomics-specific evidence: most detection models are broader AI or deep-learning systems, and their role here is to establish the feasibility of the prerequisite detection step. Convolutional neural network-radiomics hybrid frameworks have also been described with respect to segmentation accuracy and diagnostic speed [4].

Segmentation is not a neutral preprocessing step. The region of interest (ROI) may be confined to the stone core or may deliberately include the high-density halo artefact adjacent to the stone, and studies differ in how completely they report slice thickness and reconstruction algorithms. Automated segmentation performs reasonably well in overall terms—Dice scores of 87.56% for UNet++ [6] and 87.69% for ResU-Net [7]—but performance is less secure for small stones (<3 mm), with a Dice score of 73.7% reported for small stones in one methodological study [6]. Whether semi-automated and fully automated masks can be used interchangeably has received surprisingly little attention.

Mask variation propagates into feature values, although not uniformly. Texture descriptors such as skewness and kurtosis appear more sensitive to segmentation differences, whereas volume and HU are comparatively stable. One study reported inter-observer intraclass correlation coefficients (ICCs) of 0.751 - 0.998 for extracted features, which suggests that the sensitivity of individual features to segmentation variation is uneven [8]. The Image Biomarker Standardisation Initiative (IBSI) offers a framework for standardising feature definitions, but its uptake in urolithiasis radiomics remains limited. Such differences matter because they can affect the comparability of feature values across studies.

For stone burden quantification, automatically segmented volume, maximum diameter, and stone number have shown high agreement with manual measurements, with only small systematic bias on Bland-Altman analysis [6]. The authors rightly cautioned, however, that statistical agreement is not the same as clinical

interchangeability; the consequences of individual errors near clinical decision thresholds have not been evaluated.

## 3. Research Progress of Radiomics in Stone Composition Prediction in Urolithiasis

Composition matters clinically because management pathways diverge. Uric acid stones may be treated by oral chemolysis; cystine stones respond relatively poorly to ESWL and often require retrograde intrarenal surgery, percutaneous nephrolithotomy, or laser lithotripsy; infection stones demand complete clearance together with appropriate antimicrobial management; and calcium oxalate and other stones are managed according to size, location, and anatomy. Against this background, radiomics studies have concentrated on identifying uric acid stones and on discriminating cystine, calcium oxalate, and infection stones [3] [9].

For uric acid stone identification, a multicentre study developed a combined radiomics-clinical model and reported AUCs of 0.942, 0.944, and 0.957 in the training, internal test, and external validation sets, respectively (external validation 95% CI 0.905 - 1.000); Shapley additive explanations (SHAP) identified the radiomics prediction probability, urinary pH, serum uric acid, body mass index, and serum creatinine as the main contributors [9]. A single-centre NCCT study using a LightGBM model that combined radiomics with clinical features reported a test-set AUC of 0.805 (95% CI 0.666 - 0.944), with decision curve analysis showing net benefit across a range of threshold probabilities [8]. Two discrepancies are worth noting. First, external validation was a distinguishing feature of the multicentre study, whereas most single-centre studies did not attempt it. Second, the incremental value of combining radiomics with clinical variables was inconsistent: in one study the radiomics-only model reached an AUC of 0.634 and the combined model 0.805 in the test set, while elsewhere the gain from combination was more modest [8].

Dual-energy CT is attractive because it has a material-decomposition basis for separating uric acid from non-uric acid stones, yet it has a recognised blind spot for ammonium urate stones: 62% of ammonium urate stones have been reported to be misclassified as uric acid stones by dual-energy CT (DECT). Zheng *et al.* addressed this gap by constructing a radiomics model to distinguish ammonium urate from uric acid stones, with internal and external validation AUCs of 0.895 (95% CI 0.796 - 0.995) and 0.870 (95% CI 0.769 - 0.972), respectively [10]. Most composition studies still rely on single-energy CT, and studies designed specifically around DECT-ambiguous cases remain scarce in the retrieved literature. How radiomics should relate to DECT in this setting is therefore a question for further study rather than a settled issue.

## 4. Research Progress of Radiomics in the Treatment of Urolithiasis

### 4.1. Non-Invasive Treatment: ESWL and Conservative Management

Prediction of SFR after ESWL is the most mature treatment-response application

in urolithiasis radiomics, but the evidence base is not as robust as the AUC figures might suggest. A systematic review and meta-analysis of 5 studies reported a pooled AUC of 0.883 (95% CI 0.840 - 0.927) with low heterogeneity ($I^2$ = 24.3%); all included studies were nevertheless judged to have a high risk of bias, methodological compliance was moderate, and the stability of the pooled estimate may be affected by study design [11]. In a single-centre study of 329 patients with ureteral stones undergoing a single ESWL session, Yang *et al.* reported an internal test-set AUC of 0.888 (95% CI 0.822 - 0.949) without external validation; SHAP analysis identified the radiomics score and proximal ureteral diameter as major contributing variables [12]. That ranking indicates that both contributed to prediction, but it does not by itself establish redundant information between the radiomics score and traditional anatomical parameters. A more consequential source of heterogeneity lies in outcome definitions and treatment parameters: SFR has been defined variously as "no visible residual fragments" or "residual fragments ≤ 5 mm," assessment windows have ranged from postoperative day 1 to 3 months, and ESWL energy level, shock wave number, and machine type were either not incorporated or incompletely reported in most models. These differences can affect event counts and reported model performance, and existing pooled estimates cannot yet be translated directly into specific clinical thresholds.

Conservative management poses a different question: whether spontaneous stone passage (SSP) occurs during observation, a outcome whose ascertainment depends on follow-up imaging and timing. Study designs must therefore contend with competing events such as loss to follow-up and subsequent intervention. Panthier *et al.* studied 428 patients with acute ureteral colic and reported a validation-set AUC of 0.79 (95% CI 0.68 - 0.88) for SSP prediction using a random forest model; they also suggested that a hybrid model incorporating stone location might improve performance [13]. Maximum stone diameter and location remain the traditional predictors of SSP. Radiomics models have rarely been compared head-to-head with a "stone size + location" reference model or evaluated by net reclassification improvement, and the feasibility of automated analysis in this scenario deserves further observation.

### 4.2. Invasive Treatment: RIRS and PCNL

Invasive treatment is less well studied than ESWL. Radiomics prediction studies for RIRS are fewer than those for ESWL, and PCNL studies are fewer still. Both RIRS and PCNL assess residual stones at relatively early postoperative time points, and outcome definitions are shaped by residual-stone thresholds and imaging modality. RIRS outcomes are additionally influenced by laser equipment, lithotripsy strategy, and operator technique, whereas PCNL studies often use anatomical scoring systems such as the Guy's stone score and the S.T.O.N.E score as comparators [14].

RIRS evidence remains predominantly single-centre. Xun *et al.* constructed a clinical-radiomics nomogram for 264 patients with isolated kidney stones under-

going RIRS and reported a training-set AUC of 0.94 and an internal validation-set AUC of 0.947, without external validation [15]. That AUC is higher than in most ESWL studies, but the validation set consisted of internal samples, and reproducibility in external populations remains unknown. A systematic review of 5 retrospective cohorts found substantial differences among studies in populations, outcome definitions, and follow-up windows, with some pooled $I^2$ values of 94.6% - 96.9%, and concluded that the clinical interpretability of pooled estimates is limited [16]. The key unresolved issue is whether radiomics features retain predictive value after controlling for operational variables such as laser equipment, lithotripsy strategy, and operator technique; few studies have addressed this.

PCNL has only a handful of radiomics studies. Two 2024 studies reported a validation-set AUC of 0.944 for a clinical-radiomics nomogram and AUCs of 0.77 - 0.85 for machine-learning models, respectively [14] [17]; the nomogram study also reported that its AUC exceeded that of the Guy's stone score (0.786) and the S.T.O.N.E score (0.714) [14]. With so few studies, differences in scoring-system comparators and model type inevitably limit cross-study comparison.

## 5. Recurrence Prediction: Correspondence between Single Time-Point Imaging and Longitudinal Outcomes

Recurrence prediction is the least developed of the four domains. In one study with external validation, radiomics features extracted from stone ROIs were used to construct a nomogram incorporating stone history and lower calyx stone location; the radiomics model achieved an AUC of 0.760 (95% CI 0.636 - 0.864) in the external test set, the combined model 0.786 (95% CI 0.684 - 0.878), and time-dependent ROC curves showed discrimination at 3, 5, and 7 years [18]. A machine-learning study of 4246 patients used clinical data, demographic information, and CT findings as inputs, without explicit extraction of radiomic features, and its random forest model reached an AUC of 0.64 for recurrence prediction [19]; this therefore provides broader AI/clinical-ML evidence rather than radiomics-specific evidence. A further study evaluated changes in stone burden on follow-up CT and found that automated threshold volumetry had higher diagnostic accuracy than radiomics analysis [20]; this is a longitudinal stone-burden assessment task and should be distinguished from prediction of new recurrence. Taken together, these results suggest that radiomics performance in recurrence prediction may vary considerably with outcome definition, input variables, and follow-up imaging assessment.

The temporal logic of recurrence prediction is awkward. Except for [20], which assessed longitudinal stone-burden change, the prediction models in [18] [19] used single time-point CT before treatment, whereas recurrence unfolds during follow-up; the two are separated by a temporal gap. Recurrence risk is also associated with metabolic abnormalities, urinary stone-forming propensity, and other factors that cannot be directly obtained from texture features of a single CT [21]. Definitions of recurrence differ as well: the nomogram study defined recurrence as new

stones >2 mm on follow-up imaging, but postoperative stone-free status was primarily confirmed by ultrasound, and the authors noted that some "recurrences" may in fact represent residual fragments missed on initial postoperative imaging; interval-censored models were not used to address the temporal uncertainty introduced by follow-up intervals. Recurrence definitions, follow-up imaging modalities, and temporal modelling methods therefore remain unresolved across studies.

## 6. Conclusions

Taken together, the radiomics literature in urolithiasis is unevenly distributed. Composition prediction and SFR prediction after ESWL account for the largest share of studies, and both address decision scenarios that already have established clinical assessment methods—urinary pH, dual-energy CT, and stone size and density. By contrast, RIRS/PCNL selection, timing of discontinuation of conservative management, and recurrence risk stratification remain relatively understudied. Most studies report their own model AUCs, whereas reports of net reclassification improvement or incremental benefit on decision curves are fewer. A further cross-cutting limitation is calibration: most prediction studies summarised above reported discrimination (AUC) but did not report calibration plots, calibration slope/intercept, or calibration at clinically relevant thresholds; where calibration was not reported, predicted probabilities should not be assumed to support threshold-based decisions. Studies linking different domains are also relatively few, with composition prediction models and treatment-response models usually constructed separately; segmentation methods, external validation, outcome definitions, and reporting practices continue to differ across studies.

On this basis, radiomics is perhaps best regarded as one candidate component of a multi-parameter decision-making framework in urolithiasis management, rather than as a stand-alone replacement for existing assessment. Its incremental contribution requires further observation through prospective studies and decision-impact studies. Future work might usefully concentrate on standardising segmentation and reporting, designing integrated studies that link composition prediction with treatment-response prediction, and conducting validation studies with treatment selection changes or outcome improvement as endpoints.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

[1] Vera-Ponce, V.J., Sanchez-Tamay, N.M., Ballena-Caicedo, J., Zuzunaga-Montoya, F.E., De Carrillo, C.I.G. and Poemape Mestanza, R.L. (2025) Global Prevalence of Urolithiasis: A Meta-Analysis Accounting for Methodological Heterogeneity. *Frontiers in Urology*, **5**, Article ID: 1705953. https://doi.org/10.3389/fruro.2025.1705953

[2] Fan, B. and Liu, C. (2025) Research Progress of Radiomics in Urolithiasis. *Journal of*

*Clinical Urology*, **40**, 472-475.

[3] Lim, E.J., Castellani, D., So, W.Z., Fong, K.Y., Li, J.Q., Tiong, H.Y., *et al.* (2022) Radiomics in Urolithiasis: Systematic Review of Current Applications, Limitations, and Future Directions. *Journal of Clinical Medicine*, **11**, Article No. 5151. https://doi.org/10.3390/jcm11175151

[4] Alam, K., Alam, F. and Ullah, A. (2025) A Review on Kidney Stone Detection and Diagnosis Using CT Kub Imaging: Bridging Emerging Technologies and Clinical Practice. *BMC Journal of Medical Sciences*, **6**, 94-103. https://doi.org/10.70905/bmcj.06.02.0547

[5] Hajalamin, M., Msherghi, A., Alhammadi, I., Sultan, A., Alatefi, D., Wahhab, A., *et al.* (2026) The Diagnostic Performance of Machine Learning Based Detection of Urinary Tract Stones: A Systematic Review and Meta-Analysis. *European Journal of Radiology*, **200**, Article ID: 112836. https://doi.org/10.1016/j.ejrad.2026.112836

[6] Kim, Y.I., Song, S.H., Park, J., Youn, H.J., Kweon, J. and Park, H.K. (2023) Deep-learning Segmentation of Urinary Stones in Noncontrast Computed Tomography. *Journal of Endourology*, **37**, 595-606. https://doi.org/10.1089/end.2022.0722

[7] Chen, Z., Xiao, C., Liu, Y., Hassan, H., Li, D., Liu, J., *et al.* (2025) Comprehensive 3D Analysis of the Renal System and Stones: Segmenting and Registering Non-Contrast and Contrast Computed Tomography Images. *Information Systems Frontiers*, **27**, 97-111. https://doi.org/10.1007/s10796-024-10485-y

[8] Huang, Y., Li, N., Han, X., Xu, S., Zhang, G. and Zhu, X. (2025) Radiomics Application Using Non-Contrast Computed Tomography for Predicting Uric Acid Kidney Stones. *BMC Medical Imaging*, **25**, Article No. 433. https://doi.org/10.1186/s12880-025-01965-x

[9] Shi, Z., Xu, Z., Zhang, L., Zheng, W., Liu, X. and Guo, L. (2025) Noninvasive CT Radiomics-Clinical Model Accurately Classifies Anhydrous Uric Acid Stones: A Multicenter Study. *World Journal of Urology*, **44**, Article No. 37. https://doi.org/10.1007/s00345-025-06152-9

[10] Zheng, J., Zhang, J., Cai, J., Yao, Y., Lu, S., Wu, Z., *et al.* (2024) Development of a Radiomics Model to Discriminate Ammonium Urate Stones from Uric Acid Stones in Vivo: A Remedy for the Diagnostic Pitfall of Dual-Energy Computed Tomography. *Chinese Medical Journal*, **137**, 1095-1104. https://doi.org/10.1097/cm9.0000000000002866

[11] Talyshinskii, A., Ali, S., Gauhar, V., Güven, S., Somani, B.K., Sarica, K., *et al.* (2025) CT-Based Radiomics for Predicting Stone-Free Rate after ESWL: A Systematic Review and Meta-Analysis. *World Journal of Urology*, **43**, Article No. 566. https://doi.org/10.1007/s00345-025-05933-6

[12] Yang, R., Zhao, D., Ye, C., Hu, M., Qi, X. and Li, Z. (2025) Predicting ESWL Success for Ureteral Stones: A Radiomics-Based Machine Learning Approach. *BMC Medical Imaging*, **25**, Article No. 268. https://doi.org/10.1186/s12880-025-01817-8

[13] Panthier, F., Khene, Z., Velinova, D., Berthe, L., Crawford-Smith, H., Allen, S., *et al.* (2026) Spontaneous Stone Passage Prediction in Acute Ureteric Colic Using Radiomics and Machine Learning. *BJU International*, **138**, 841-851. https://doi.org/10.1111/bju.70408

[14] Zou, X.C., *et al.* (2024) Prediction of Preoperative the Stone-Free Rate in Percutaneous Nephrolithotomy Based on CT Clinical-Radiomics Nomogram: A Comparative Study with Guy's Stone Score and S.T.O.N.E Score. https://doi.org/10.21203/rs.3.rs-3767593/v1

[15] Xun, Y., Chen, M., Liang, P., Tripathi, P., Deng, H., Zhou, Z., *et al.* (2020) A Novel

Clinical-Radiomics Model Pre-Operatively Predicted the Stone-Free Rate of Flexible Ureteroscopy Strategy in Kidney Stone Patients. *Frontiers in Medicine*, **7**, Article ID: 576925. https://doi.org/10.3389/fmed.2020.576925

[16] Ghazwani, Y., Alghafees, M., Alshasha, M., Brayan, F., Alsayyari, A. and Alyami, A. (2026) Can AI and Predictive Models Accurately Predict Stone-Free Status? A Systematic Review and Meta-Analysis. *Canadian Journal of Urology*, **33**, 291-308. https://doi.org/10.32604/cju.2026.077411

[17] Zou, X.C., Luo, C.W., Yuan, R.M., Jin, M.N., Zeng, T. and Chao, H.C. (2024) Develop a Radiomics-Based Machine Learning Model to Predict the Stone-Free Rate Post-Percutaneous Nephrolithotomy. *Urolithiasis*, **52**, Article No. 64. https://doi.org/10.1007/s00240-024-01562-7

[18] Lei, Y., Zhong, J., Chai, C.A., Chen, T., Lin, Z., Chen, X., *et al.* (2025) Pre-Operatively Predicting Kidney Stone Recurrence: Integrating Radiomic Features and Clinical Variables Using Machine Learning. *BMC Medical Imaging*, **25**, Article No. 489. https://doi.org/10.1186/s12880-025-02043-y

[19] Homayoun, H., Mousavirad, S.J., Zareian Baghdadabad, L., *et al.* (2026) Machine Learning-Based Prediction of Urolithiasis Recurrence Using Patient's Clinical Data, Demography, and CT Findings. *Urology Journal*, **22**, 289-300.

[20] Kaviani, P., Froelich, M.F., Bizzo, B., Primak, A., Dasegowda, G., Garza-Frias, E., *et al.* (2025) Assessing Change in Stone Burden on Baseline and Follow-Up CT: Radiologist and Radiomics Evaluations. *Journal of Imaging*, **12**, Article No. 13. https://doi.org/10.3390/jimaging12010013

[21] Skolarikos, A., Somani, B., Neisius, A., Jung, H., Petřík, A., Tailly, T., *et al.* (2024) Metabolic Evaluation and Recurrence Prevention for Urinary Stone Patients: An EAU Guidelines Update. *European Urology*, **86**, 343-363. https://doi.org/10.1016/j.eururo.2024.05.029