Peer Review File

Article information: http://dx.doi.org/10.21037/atm-21-965

Reviewer A

OVERALL IMPRESSION: This Original Research article describes the methodology of a machine-learning classification of stone composition using single-energy CT in a single-center study. The purpose of this study has been extensively investigated in the literature, and the novelty of this work is marginal.

Responses to reviewer A:

Comment 1： Page 3, line 61. The authors describe their method as "novel, simple and non-invasive". However, there is extensive literature on classification of renal stones based on their composition, including using single or dual energy CT, as well as using AI-based methods, thus making the novelty of the proposed approach marginal. Additionally, the proposed method requires manual segmentation of the renal stone in each plane, which contradicts the claim of simplicity. On the contrary, the need for manual segmentation makes the technique subject to considerable inter- and intra-observer variability and inefficient for clinical adoption.

Reply 1: Thank you for your comments. This description may not be accurate, so we have revised it.

Changes in the text: We removed this sentence -- "which is a novel, simple and non-invasive method of stone analysis". (see Page 3, line 61)

Comments 2: Page 6, Lines 141-144. Sorting the stones by their main component ignores mixed stones. A Stone with 40% COM, 30% COD and 30% APA is not the same as a stone with 100% COM. The authors imply they classified them the same way. This greatly confounds the interpretation of the reported results.

Reply 2: Thank you for your comments. At present, there is no unified standard for classification of mixed stones.We should explain why we chose this classification method.

Reasons:

①There is no standardized system for classifying stones, and there is controversy about how best to classify mixed-component stones.

【】Thongprayoon C, Krambeck AE, Rule AD. Determining the true burden of kidney stone disease. Nat Rev Nephrol. 2020 Dec;16(12):736-746. doi: 10.1038/s41581-020-0320-7. Epub 2020 Aug 4. PMID: 32753740.

② Most reports on stone composition in the literature focus on the principal

component. It is a simple and useful approach from an epidemiological point of view. Mixed stones are more common clinically than pure stones, and the treatment of mixed stones is selected based on the major component.

【】Cloutier J, Villa L, Traxer O, Daudon M.Kidney stone analysis: "Give me your stone,I will tell you who you are!". World J Urol.2015;33:157-69.

【】Zhang GM, Sun H, Xue HD, et al. Prospective prediction of the major component of urinary stone composition with dual-source dual-energy CT in vivo. Clin Radiol 2016;71:1178-83.

③ The following paper classifies calculi in terms of main components:

【】Große Hokamp N, Lennartz S, Salem J, Pinto Dos Santos D, Heidenreich A, Maintz D,

Haneder S. Dose independent characterization of renal stones by means of dual energy computed tomography and machine learning: an ex-vivo study. Eur Radiol. 2020 Mar;30(3):1397-1404. doi: 10.1007/s00330-019-06455-7. Epub 2019 Nov 26. PMID: 31773296.

Changes in the text: In the method section, we added relevant literature to explain this classification method. (see Page 6, line 126-129)

Comments 3: Page 6, line 147. Please describe which metrics of stone morphology were recorded and how they were derived.

Reply 3: We used CT images to classify the morphology of stones, and we were mainly concerned with staghorn and non-staghorn stones.

Changes in the text: we have modified our text as advised (see Page 6, line 138-142)

Comments 4: Page 6, starting line 161. The use of Bf40 and Br40 reconstruction kernels on the CT data is a limitation of this work. These kernels are not adequate for quantitative tasks. They have edge enhancement features that alter the CT numbers are the edges of the stones. Depending on the size of the stones, this can substantially alter the measured stone attenuation from truth.

Reply 4: Thank you for your comments. The scan and reconstruction parameters described in the text are those routinely used in our hospital and may indeed have some shortcomings. We will add your opinion to the Limitations part**.**

Changes in the text: we have added this to the Limitations part (see Page 16, line 346-348)

Comments 5: Page 7, line 171-175: of the 6 steps described, only the latter can be considered machine-learning. This paragraph should be removed.

Reply 5: Ok, we agree to remove this paragraph.

Changes in the text: we have removed this paragraph. (see Page 8)

Comments 6: Page 7, lines 178-188. This is a significant limitation of the study. All stones were segmented by a single user, slice by slice. Therefore, the accuracy and robustness of the resulting model is intrinsically tied to the reliability and reproducibility of this manual segmentation.
Reply 6: The intraclass and interclass correlation coefficient (ICC) was used to evaluate the consistency of the extracted features, and features with ICC > 0.75 were considered to be consistent, reproducible and stable. In this step, a total of 8 features with ICC ≤ 0.75 were deleted, leaving 1210 features.
Changes in the text: This content has been described in the article (see Page 9, line 198-200) . At the same time, we added it to the Limitations part (see Page 16, line 350-353) .

Comments 7: Page 7, lines 194-195. This normalization task removes the information that is contained in the mean CT number of the stone, which is associated with stone type.
Reply 7: According to the AI standardization process, data standardization should be carried out before ROI feature extraction. As you say, it may remove some information that is contained in the mean CT number of the stone. Therefore, we will add this to the

Limitations part**.**
Changes in the text: we have added this to the Limitations part (see Page 16, line 348-350)

Comments 8: Page 11, line 283. Mixed stones are mentioned for the first time in the study population. It is unclear how mixed stones were defined, how they were characterized, and their impact on the study results. The authors need to demonstrate that accuracy of their methods in pure stones before they even try to address mixed stone. Characterizing a stone as COM when it included other important minerals is not a correct result.
Reply 8: Mixed calculi are more common in clinical practice. If we only study pure calculi, the sample size may be insufficient. As for the answer in Comments 2, there is no unified standard for classification of mixed stones. The following paper classifies mixed stones in terms of their main components:
【】Große Hokamp N, Lennartz S, Salem J, Pinto Dos Santos D, Heidenreich A, Maintz D, Haneder S. Dose independent characterization of renal stones by means of dual energy computed tomography and machine learning: an ex-vivo study. Eur

Radiol. 2020 Mar;30(3):1397-1404. doi: 10.1007/s00330-019-06455-7. Epub 2019 Nov 26. PMID: 31773296.
Changes in the text: In the method section, we added relevant literature to explain this classification method. (see Page 6, line 126-129)

Comments 9: Page 11, starting line 313. The comparisons to dual-energy CT are inadequate. Very few papers were referenced and the essentially 100% accuracy of differentiating UA stones from calcified stones is not mentioned. There are a great many in vivo studies and the dual-energy approach to stone composition is so accepted as accurate in vivo that it is already in mainstream clinical use. It is the reference standard that this technique needs to compare itself to and the authors do not do that adequately.
Reply 9: Thank you for your comments. We will add the literature on dual-energy CT in vivo trials and discuss it.
Changes in the text: We have modified our text as advised. (see Page 14, line 291-312)

Comments 10: Page 12, line 333. The manual segmentation and the lack of reproducibility test for this step in the model should be clearly listed as a limitation. All similar studies that are cited in the discussion involve a semi-automated segmentation approach.
Reply 10: We will add this to the Limitations part**.**
Changes in the text: we have added this to the Limitations part (see Page 16, line 350-353)

Comments 11: Page 12. There are a number of other limitations to the work that need to be described, some of which are noted above. The single most clinically relevant task for in vivo stone composition analysis is the identification of UA-based stones, which are amenable to medical treatment. The authors attempt to differentiate COM (n=373) vs. non-COM (n=170), however their non-COM populations contain two very different stones types: UA (n=114) and APA (n=56). UA and COM stones are extremely easy to separate in dual-energy CT. The important stones to separate are the hard to break COM and the easy to break APA. I think that the relatively good results are the result of having so manu UA stones in the non-COM population.
Reply 11: We have already mentioned this limitation in the article. Only the most common
COM stones were modeled in this study, other types of stones have not been analyzed due to the small sample size, and it is expected that more types of stones can be detected later after adding data from other centers.

Changes in the text: we have added this to the Limitations part (see Page 16, line 339-346)

Comments 12:REFERENCES

1.A number of additional in vivo dual-energy CT papers should be added showing the excellent accuracy (~100% in all but the largest patients) of separating UA from Ca-based stones.

2.Ferrero et al has reported on the radiomic properties of different stone types in work related to estimating stone fragility. This is directly related to the clinical task under study and should be included in the discussion.

Reply 12: Thank you for your comments. Relevant literature will be added as required.

Changes in the text: We have added relevant literature as required (see Page 14, line 299-312, 323-326)

Reviewer B

The authors describe an AI model to recognize CaOx Monohydrate stones on unenhanced CT images. This is an interesting topic and I do believe this technique will lead to improvements in daily practice. However, I do have two major concerns about this paper.

Comment 1： The majority of kidney stones have mixed compositions. The authors describe that they sorted the stones according to the main composition. What was the cut-off to decide if a stone was included in the CaOx Mono category or in the non-CaOx Mono category? This can be a major limitation and should be addressed in the discussion, because there is a big difference between a 100% CaOx Mono stone and a 50% CaOx Mono - 50% non-CaOx Mono stones. Ideally only pure stones would have been used (100% CaOx Mono versus 100% non-CaOx Mono). They do, however, mention mixed cases in the discussion, but not in relation to their AI model or which consequences this may have.

Reply 1: Thank you for your comments. At present, there is no unified standard for classification of mixed stones.We should explain why we chose this classification method.

Reasons:

①There is no standardized system for classifying stones, and there is controversy about how best to classify mixed-component stones.

【】Thongprayoon C, Krambeck AE, Rule AD. Determining the true burden of kidney stone disease. Nat Rev Nephrol. 2020 Dec;16(12):736-746. doi: 10.1038/s41581-020-0320-7. Epub 2020 Aug 4. PMID: 32753740.

② Most reports on stone composition in the literature focus on the principal component. It is a simple and useful approach from an epidemiological point of view. Mixed stones are more common clinically than pure stones, and the treatment of mixed stones is selected based on the major component.

【】Cloutier J, Villa L, Traxer O, Daudon M.Kidney stone analysis: "Give me your stone,I will tell you who you are!". World J Urol.2015;33:157-69.

【】Zhang GM, Sun H, Xue HD, et al. Prospective prediction of the major component of urinary stone composition with dual-source dual-energy CT in vivo. Clin Radiol 2016;71:1178-83.

③ The following paper classifies calculi in terms of main components:

【】Große Hokamp N, Lennartz S, Salem J, Pinto Dos Santos D, Heidenreich A, Maintz D,
Haneder S. Dose independent characterization of renal stones by means of dual energy computed tomography and machine learning: an ex-vivo study. Eur Radiol. 2020 Mar;30(3):1397-1404. doi: 10.1007/s00330-019-06455-7. Epub 2019 Nov 26. PMID: 31773296.

Changes in the text: In the method section, we added relevant literature to explain this classification method. (see Page 6, line 126-129)

Comment 2： Another big concern is how the authors corrected for bias in patients with multiple stones and possible different stone compositions in these multiple stones? As generally not all stones are analyzed and moreover only part of a stone that is send for analysis will be evaluated. The authors do not describe this possible bias or how they bypassed this problem.

Reply 2: Thank you for your comments. Since the most harmful stones are often treated first in clinical practice, for a particular patient with multiple stones, we defined all stones in the body as one sample, named after the main component of the surgically obtained specimen.

Changes in the text: In the method section, we explained this classification method. (see Page 6, line 129-132)

Comment 3： Abstract:

- In the sentence "Urolithiasis is a global disease with a high incidence and recurrence rate, and stone composition is closely related to the choice of treatment, surgery, and preventive measures." surgery is a choice of treatment and should be omitted.
- The authors describe that CaOx Mono stones are hard and difficult to pulverize. However, stones are not pulverized but dusted, popcorned or fragmented.
- The abbreviation AUC (area under the curve) is only described in the manuscript and not in the abstract.

Reply 2: Thank you for your comments. We will revise them as advised.
Changes in the text: we have modified our text as advised (see Page 2-3, line 36,38, 57).

Comment 4： Introduction:
- Omit "or kidney non-function" in the sentence "Severe cases may be combined with long-term, chronic obstructive hydronephrosis, eventually causing varying degrees of kidney function damage, or kidney non-function, and even endanger the lives of patients."
- according to the authors, what is the difference between choice of treatment and surgical method in the sentence "Calculi composition is closely related to the choice of treatment, surgical method and preventive measures."?
-"For larger stones that are difficult to crush, ...". Urolithiasis are not crushed. Please correct.
- "In terms of crystal properties, most of COM are brown, mulberry-shaped, hard and not easy to crush; COD is mostly white, with needle-like protrusions visible on the surface, brittle and easy to be broken." Please refer to the Daudan classification when describing morphological characteristics of urolithiasis, as CaOx Di stones are not predominantly white and CaOx Mono stones can have a white appearance. Furthermore, correct the term 'crush'.
- Please correct the term 'pulverized' in the sentence "COM is the most common stone and difficult to be pulverized in clinical practice, and it is of great clinical significance to clarify its chemical composition before surgery to facilitate the correct choice of treatment and surgery." Furthermore, once again I wonder how surgery and choice of treatment differ? Isn't surgery a form of treatment?
- Lithotripsy is a form of surgery. Please correct in the sentence "However, these methods can only be performed in vitro, and stone specimens must be obtained by lithotripsy, surgery, or after being passed out of the body on its own."

Reply 4: Thank you for your comments. We will revise them as advised.
Changes in the text: we have modified our text as advised (see Page 4-5, line 75-77, 82, 92-95, 95-97, 102).

Comment 5： Methods:
-Please clarify how a nephrostomy tube or ureteral drainage tube placement influences stone composition and thus why it should be an exclusion criteria.
-The authors included a total of 543 patients. Why 543? Was this number chosen at random or did the authors perform a POWER-analysis?
Reply 5: Thank you for your comments. Nephrostomy tube or ureteral drainage tube

placement and overlaping with the calculi (which may affect stone segmentation and feature extraction) is the exclusion criteria. 543 is the number of cases we finally collected according to the inclusion and exclusion criteria.
Changes in the text: we have modified our text as advised (see Page 6, line 134-136,143).

Comment 6： Results
-clinical information of included patients is redundant in my opinion, as it is not used for the AI model.
-24 cases of antler-shaped stones and 349 cases of non-antler-shaped stones. However in table 1 the authors describe staghorn and non-staghorn stones, which is the correct denomination. Please correct.
Reply 6: The clinical information can reflect the epidemiological characteristics of calculi in this region, and the corresponding analysis is also made in the discussion section (see Page 12, line 254-290).
Changes in the text: we have modified our text as advised (see Page 9, line 234).

Comment 7： Discussion:
-"Patients with urolithiasis have higher levels of COM in their urine and higher levels of COD in the urine of normal subjects,". I do not understand this sentence. I suppose the authors mean 'than in normal subjects'. Please correct or clarify.
- Use staghorn or antler-shaped consequently please: "It can be seen that pure COM rarely forms antler-shaped stones."
Reply 7: Thank you for your comments. We will revise them as advised.
Changes in the text: we have modified our text as advised (see Page 10, line 256-257, 277).