Iranian Journal of Radiology
Volume 20 · Issue 2 · e134454

OPEN PEER REVIEW

A Study Toward Automatic Identification of Renal Stone Composition in Single-energy or Ultra-low-dose CT Scan Using Deep Neural Networks

Mahboobeh Sheikhi, Sedigheh Sina, Mehrnoosh Karimipourfard, Fereshteh Khodadadi Shoushtari

SUBMITTED	REVISED	ACCEPTED	DOI
28 Dec 2022	9 Apr 2023	11 Apr 2023	10.5812/iranjradiol-134454

About this report

This open peer review (OPR) report shows the public comments made on this article during peer review, by the reviewers, the associate editor and the editor-in-chief, for every revision round, together with the authors' replies. Comments appear only with the consent of the people who wrote them. Reviewers who chose to sign their report are named; the others are shown as "Referee". Comments are edited for language only, and confidential remarks to the editors are not included.

More about open peer review: brieflands.com/knowledgebase/opr

Contents

1 **Original submission** Adeleh Dadkhah, Referee 1, Associate Editor, Editor-in-Chief 2
2 **Revision 1** Adeleh Dadkhah, Referee 1, Associate Editor, Editor-in-Chief 4
Authors' reply to reviewers 5
Reviewer and editor comments 11
3 **Revision 2** Referee 1, Associate Editor, Editor-in-Chief 12
Authors' reply to reviewers 13
Reviewer and editor comments 14
4 **Revision 3** Editor-in-Chief 15
Authors' reply to reviewers 16
Reviewer and editor comments 21
5 **Revision 4** Editor-in-Chief 22
Authors' reply to reviewers 23
Reviewer and editor comments 24

1 Original submission

4 comments · 9 Jan 2023 – 1 Feb 2023

Adeleh Dadkhah Reviewer

MINOR REVISION 9 Jan 2023

Thank you to the authors for their efforts.
This article is about using artificial intelligence for the classification of kidney stone types using single-energy CT scans.
Although the subject is promising and the paper is well organized, there are some questions that the authors should clarify:

1. The manuscript must be revised grammatically. There are some typos that must be fixed in the revised version. For example:
"Dual-Energy Computed Tomography scans (DECT) is a non-invasive method…"
"Furthermore, using low-dose or ultra-low dose single energy CT scan is both more widely accessible and safer in terms of radiation exposure…"
etc.

2. Why is the Taguchi method used for optimization?
3. The equation mentioned on page 7 is attributed to [27]. But the reviewer could not find it in reference 27. It seems that [1- S\N…] at the beginning of the equation was added incorrectly. If 1 is the equation number, place it after the equation and in parentheses.
4. Page 10: Add numbers to equations.
5. As the reviewer understands it, the objective of this study is to determine renal stone types with the aid of ultra-low-dose CT scans and deep learning algorithms. If this is true, what is the role of Dual-Energy Computed Tomography scans in this study, given that the stone type is determined by the biochemical test as the gold standard method?

Referee 1 Reviewer

MINOR REVISION 14 Jan 2023

Thanks to the authors for considering IJR for their submission.
The results of this study show the efficacy of deep neural networks for the classification of kidney stone types using CT scans. Its structure is relatively appropriate for scientific writing, but it needs some revision as follows:
1. Abstract: It should be written according to the journal guidelines.
2. Abstract results: Descriptive findings about the size and types of stones should be added as the first sentence.
Full-Text
3. Introduction: It is very long and should be summarized.
4. The last section of the methods should contain the statistical analysis.
5. This section of the results describes methods and should be transferred to the statistical section of the methods: "This study was conducted using the MATLAB 2021a platforms' Deep Learning Toolbox on a system with the following setup: OS: Windows 10 (enterprise/home), CPU: Intel i7-10700 (@3.5 GHz), GPU: NVIDIA GeForce RTX 3080, 32GBs of RAM."
6. The indices in table 7 should be reported along with 95% confidence intervals. Moreover, precision should be changed to accuracy, and positive and negative predictive values should be added.
7. ROC curve analysis graphs should be added.

8. The published paper entitled "Dose independent characterization of renal stones by means of dual energy computed tomography and machine learning: an ex-vivo study" is a very good template for a more appropriate presentation of findings such as: 1) compositions and numbers of pure and compound stones, and 2) confusion matrices and receiver-operator-characteristic (ROC) for testing the network.

Associate Editor MAJOR REVISION 1 Feb 2023

The manuscript needs major revision based on the reviews and evaluation by the reviewers.

Editor-in-Chief MAJOR REVISION 1 Feb 2023

Dear authors,
Your article needs Major Revision.
Regards

2

Revision 1

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

STEP BY STEP response to reviewers:

We appreciate very much the comprehensive reviews provided by the referees.
The comments and recommendations have helped us to improve the quality of this article substantially.
We have provided point-by-point responses to each of the referee's concerns.

****

Reviewer 1:

****

Thanks for the authors' effort.
This article is about using artificial intelligence for the classification of kidney stone types by using Single energy CT scan.
Although the subject is promising and the paper is well organized; they are some questions that the authors should be clearer about them:

1. The manuscript must be revised grammatically. They are some typo that must be fixed in the revised version. For example:
"Dual-Energy Computed Tomography scans (DECT) is a non-invasive method…"
"Furthermore, using low-dose or ultra-low dose single energy CT scan is both more widely accessible and safer in terms of radiation exposure…"
etc.

ok, the typos are corrected.

2. Why Taguchi method is used for the optimization?

Thank you for your comment. Many studies have been conducted in the area of the Taguchi method, which focuses on training-procedure optimization. Taguchi optimized the number of hyperparameters simultaneously and the training procedure guide in an optimized manner. Method, provides a better trade-off between sufficient data samples and fewer experimental costs with an orthogonal table of limited factors and levels, and in situations where we have limited data, it can help achieve better network performance. The advantage of Taguchi's optimization technique is the ability of solving problems with a high degree of complexity using small number of experiments in the optimization process. Taguchi's method is easy to implement and converges to the desired goal quickly. To avoid sticking at a local optimum, improved Taguchi's method can be used to solve this problem, which implies increasing the initial level difference, using five level OA, and using random Taguchi's method. To address the aforementioned issue, we propose the Orthogonal Array Tuning Method (OATM) which can achieve a trade-off of the less tuning time and competitive performance.
There are many researches that use the Taguchi method in training procedure optimization which I addressed two of them below:

- Zhang, Xiang, et al. "Deep neural network hyperparameter optimization with orthogonal array tuning." International conference on neural information processing. Springer, Cham, 2019.
- Yu, Ling, et al. "Spectrum prediction based on Taguchi method in deep learning with long short-

term memory." IEEE Access 6 (2018): 45923-45933.
In this study, we want to focus on the group of hyperparameters optimization and also achieved the optimal arrangement.

3. The equation mentioned in page 7 is referred to [27]. But the reviewer could not reference it in 27. It seems that [1- S\N…] at the beginning of the equation is added wrongly. If the 1 is the equation numbers, mention it after the equation and between parentheses.

Ok, corrected

4. Page 10: Add numbers to equations.

Ok, corrected.

5. As the reviewer founded, the objective of this study is to determine renal stones type by the aid of ultra-low dose CT scan and deep learning algorithms. If this is true, what is the role of Dual-Energy Computed Tomography scan in this study while the stone type is determined by the biochemical test as gold standard method?

As stated in the title, and the objective part of the manuscript, the purpose of the manuscript is determining the stone type using ultra-low dose CT scans or single energy CT scans (the two single energy images from a DECT scan) by using deep network.

We have corrected the title, and the objective of the manuscript for more clarification of the subject, as follows:

The aim of this study is to specify the renal stone composition in Single-Energy or Ultra-low Dose CT scans by using Deep Neural Networks. In this study, we proposed a classification method based on a deep learning approach that emphasized on the use of CT images. Human kidney stones were placed on an anthropomorphic phantom, and several ultra-low, and DECT (a pair of SECTs) images were taken to collect input data. Human kidney stones were categorized based on biochemical reports as the gold standard. Deep neural networks were used for determination of the stone type by using the single energy CT or ultra-low CT.

Actually, both DECT, and lab results are the same, and they are used as the gold standard. We added this part to the manuscript:

“The results of the biochemistry analysis in the biochemical lab, and the DECT predictions were used as the gold standard in this study.”

****

Reviewer 2:

****

I read the manuscript, nice idea and a practical one but I think some improvement is needed,

First as I know the sample size, it should be larger than the present study for accuracy of the AI method.

We have used data augmentation to overcome the fact of short sample data. Data augmentation is a technique of artificially increasing the training set by creating modified copies of a dataset using existing data. It includes making minor changes to the dataset or using deep learning **to generate new data points.**

The following part was added to the manuscript.

The training and validation datasets were augmented (random reflection axis x,y ,random rotation 0°-20°,random rescaling 0.1-1) to improve the models' accuracy

A better title can help to be more interesting for the readers.

Ok, the title was changed.

Using very old references can be changed.

Ok, it was changed

Thanks for your attention, you are right, we changed old references, except one reference which is the highlight and prominent research in this issue, and we thought that is necessary to mentioned it. Based on your comments and suggestion, if it is necessary, we can change it.

****

Reviewer 3:

****

Thanks to the authors for considering IJR for their submission.

About this study it can be said that its results contain the efficacy of deep neural networks for the classification of kidney stone types by CT scan. According to scientific writing structure it is relatively appropriate; but it needs some revision as follows:

1. Abstract: It should be written according to journal guidelines

Ok, it was written according to journal guidelines

2. Abstract results: Descriptive finding about the size and types of stones should be added as first sentence.

Thanks for the good comment

This sentence was added:
53 uric acide, 55 Calcium- oxalate, and 38 Cystine stones with the diameters between 1 mm, and 3 mm was used for preparing the CT scans in this study.

Full-Text

3. Introduction: It is very long and should be summarized.

Ok, it was summarized.

4. Last section of method should contain statistical analysis.

Ok, we have added this part.

5. This section of results is method and should be transferred to the statistical section of methods: "This study was conducted using the MATLAB 2021a platforms' Deep Learning Toolbox on a system with the following setup: OS: Windows 10 (enterprise/home), CPU: Intel i7-10700 (@3.5 GHz), GPU: NVIDIA GeForce RTX 3080, 32GBs of RAM."

It was transformed.

6. Indices of table 7 should be reported along with %95 confidence intervals. Moreover, precision should be changed to the accuracy, and positive and negative predictive values should be added.

Ok, it was added.

7. Roc curve analysis graphs should be added.

OK, It was added.

8. Published paper entitled as "Dose independent characterization of renal stones by means of dual energy computed tomography and machine learning: an ex-vivo study" is very good template for more appropriate presentation findings like as: 1) compositions and number of pure and compound stones, and 2) confusion matrices and receiver-operator-characteristic (ROC) for testing of the network.

Ok, corrected.
All stones in this study were pure. Roc curve analysis graphs was drawn.

****

Reviewer 4:

****

This study reports the utilization of deep learning (DL) networks for the classification of renal stones in men and women. Such classification is of great importance to select suitable treatment/control/prevention methods.

The authors have then investigated and compared several pre-trained DL networks in terms of key evaluation metrics. The performance networks were further improved with a Taguchi optimization method. However, some points need to be clearly addressed by the authors. Upon correction, I would consider accepting the present contribution to the Iranian Journal of Radiology.

1- The role of ultra-low dose CT (120 kVp) in the study should be more elaborated. Could not the user determine the location of a stone only using dual-energy CT images?
Thanks for the good comment.
If we use the abdomen DECT for finding the position of the stone, the high dose is imposed to the patients. The role of ultra-low dose CT is determined in the manuscript as follows:
For renal stone type recognition using the DECT scanner, in the first step, a large field-of-view abdomen ultra-low CT scan is performed to determine the location of renal stone, and then dual-energy CT scans with small FOV are obtained from the kidney stone region. In this way, the DECT scan is not used for the entire abdomen, and does not impose high dose on the patients.

2- Fig. 7 & Table 7: From this figure and table, it came to the reviewer's attention that feeding a single energy CT scan (single channel) to the networks is adequate. While in Fig. 3 it seems that all three energy groups should be simultaneously fed (3-channel input) to the networks. Please comment.
You are right, Single channel is ok.
We have corrected Figure 3, each energy group was inserted separately. 80 or 120 or 135.

3- Among 80, 120, and 135 kVp, which energy is recommended from the perspective of radiation burden to the patient?
Based on the results shown in Table 7, the deep network has shown high accuracy in predicting the renal stone type, using the ultra-low CT images. Although the network showed the best performance for 80kVp, the ultra-low CT scan with the higher energy, 120 kVp, and lower tube current, 44 mA, is preferred from the view point of the dose delivered to the patient. Although the network showed the best performance for 80kVp, the ultra-low CT scan with the higher energy, 120 kVp, and lower tube current, 44 mA, is preferred from the view point of the dose delivered to the patient.
The dose reports of the scanner showed that the patient doses were increased in the dual imaging modality with respect to the specific and limited region of the patient's body compared to the patient's whole-body dose.

4- Please avoid short paragraphs.
Ok, Corrected.

5- English of the text mandates to be improved and reworded.
Ok, Corrected.

2 Revision 1

4 comments · 12 Feb 2023 – 27 Feb 2023

Adeleh Dadkhah Reviewer — ACCEPT — 12 Feb 2023

Thanks to the authors for their effort.

Referee 1 Reviewer — MINOR REVISION — 25 Feb 2023

Thanks to the authors for the revision, but some trivial changes should be made:
1. The conclusion should be based on the results in table 7. In my initial comments, I requested that positive and negative predictive values be added along with 95% confidence intervals (CIs), but these indices were not presented. Moreover, the one-way ANOVA p-value can be used for between-group comparisons of each measure. According to the CIs, only sensitivity should have a p-value less than 0.05, but there are no statistical differences in the other indices. Therefore, the conclusion should be based on the positive predictive values of the three methods.
2. This section of the results should be deleted: " In fact, the Receiver Operator Characteristic curve is a probability curve, and the area under the curve represents the measure, or degree of separability. The AUC of the ROC is usually used to determine, and compare the performance of the predictive models. The area under the curve is directly linked to the Mann-Whitney U-Statistic, which is proposed to be used as a non-parametric, robust alternative to Student's t-test. The AUC determines how much the model is capable of distinguishing between different classes. The higher the AUC of the ROC, the better the model is at classification."

Associate Editor — MINOR REVISION — 27 Feb 2023

The manuscript needs minor revision considering the reviews.

Editor-in-Chief — MINOR REVISION — 27 Feb 2023

Dear authors,
Your article needs Minor Revision.
Regards

3

Revision 2

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

Response to reviewer:

Thanks to the authors for edition; but, some trivial changes should be made:

Dear reviewer, thanks for your accurate review.

1. Conclusion should be based on results of table 7. In first time comments, i want to be added positive and negative predictive values along with % 95 confidence intervals(CIs); but these indices were not presented.

Thanks to the authors for edition; but, some trivial changes should be made:
1. Conclusion should be based on results of table 7. In first time comments, i want to be added positive and negative predictive values along with % 95 confidence intervals(CIs); but these indices were not presented.

OK, you are right. The positive, and negative values, also the accuracy values were added to table 7.

Moreover, one way ANOVA p-value can be used for between group comparisons of each measure. According to CIs, only sensitivity should have p-value less than 0.05; but other indices have not statistical differences. Therefore, conclusion should be based on positive predictive values of three methods.

OK, One-way ANOVA was used, and as you said, only sensitivity, and accuracy have significant differences. Therefore, we changed the discussion, and conclusion, and wrote that the three single energies could be used for stone type recognition by deep learning. The discussion, and conclusion was revised accordingly.

2. This section of results should be deleted: " In fact, the Receiver Operator Characteristic curve is a probability curve, and the area under the curve represents the measure, or degree of separability. The AUC of the ROC is usually used to determine, and compare the performance of the predictive models. The area under the curve is directly linked to the Mann-Whitney U-Statistic, which is proposed to be used as a non-parametric, robust alternative to Student's t-test. The AUC determines how much the model is capable of distinguishing between different classes. The higher the AUC of the ROC, the better the model is at classification."

Ok, this part was deleted from the result section.

3 Revision 2

3 comments · 4 Mar 2023 – 15 Mar 2023

Referee 1 Reviewer

ACCEPT 4 Mar 2023

Thanks to the authors for the appropriate revisions.
This manuscript is now ready for publication from a methodological point of view.

Associate Editor

ACCEPT 4 Mar 2023

Accept, considering the reviewers' review of the manuscript.

Editor-in-Chief

MINOR REVISION 15 Mar 2023

Dear authors,
Your article needs Minor Revision.
Regards

4

Revision 3

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

> Reviewers' Comments:

Reviewer 1:
Thanks to the authors for appropriate edition.
Currently, this manuscript is ready for publication according to methodology point of view.

Thanks a lot

Reviewer 2:
It's suggested that the authors add the naming of A and B for different parts of images of figure 2.

Ok, corrected

In fig 4, you didn't mention about image d) in the legend. pls, correct this.

Ok, corrected

It's suggested that the authors add the naming of A , B and C for different parts of images of figure 7.

Ok, corrected

> Associate Editor's Comments:

Associate Editor 1:
Accept considering the review of manuscript by reviewers.

Thanks

Associate Editor 2:
thank to authors for sending their work for IJR

Thanks for good comments and corrections.

134454
1-there are two “by” in title. It is better the authors rephrase it. For example, they can write “… using deep neural network”

Ok, corrected

2-it seems the authors should write in line 8 abstract “low accessibility” instead of “accessibility”

Ok, corrected

3-please add full term of abbreviations in first usage in abstract for example SNR.

Ok, corrected

4-in line 55, it is better to write “depends”

Ok, corrected

5-line 72, this part is unclear: “But although its benefits”

Ok, corrected

6-please add objective part to paper according to journal format.

Ok, corrected

7- please add full term of abbreviations in first usage in paper for example SECT, DLP, ….

Ok, corrected

8-the authors wrote: “calcium oxalate and calcium phosphate: 55 samples (CaC2O4·H2O -CaC2O4·2H2O)” while I think there is not chemical formula of calcium phosphate stone here.

Ok, corrected

9-in line 165, please correct table 3 to (Table 3) and in 167 (Table 4)

Ok, corrected

10-in line 168, please correct: “signal-to-noise ratio (SNR)” then in line 172, S/N ratio to SNR

Ok, corrected

11-what is Dicom line 183?

Ok, corrected

12-line 193, Bicubic

Ok, corrected

13-line 198, this part is unclear: “the type of the stones the CT scan software”

Ok, corrected

14-sensitivity is not probability of a positive test. It is probability of a positive result in a patient [or detection rate].

Ok, corrected

15-please correct in 225-6: “positive predictive value (PPV)” and (negative predictive value (NPV))

Ok, corrected

16-definition of PPV and MPV are unclear. [PPV is the rate of true positive among all positive results and NPV is the rate of true negative result among all negative results]

Ok, corrected

17-line 227, spec definition is wrong “probability of a negative test” please correct.

Ok, corrected

18- false positive rates (FPR or specificity) is wrong. ROC plots sensitivity versus 1-specificity.So this part should be written as: false positive rates (FPR or 1-specificity)

Ok, corrected

19-line 249, it is better as statistics.

Ok, corrected

20-line 250, please write ROC instead of Receiver Operator Characteristic

Ok, corrected

21-line 289, it should be 95% confidence interval

Ok, corrected

22-in line 313, is Cysteine correct? Cystine

Ok, corrected

23-please write in similar method for example Vgg and VGG, ResNet and Resnet…

Ok, corrected

24-please follow journal format in references, for example, journal names should be abbreviated.

Ok, corrected

25-please add full term of abbreviations in tables as footnote

Ok, corrected

26-as the journal format recommend or ask the authors to add maximum 6 tables, please integrate tables 1 and 2 together and represent as one table. [may be, the authors can integrate tables 3 and 4] or the authors can implement their own policy for this problem. Please note the same point for figures Ok, corrected

27-what do you mean by lower limit-upper limit in table 7? 95% confidence interval? If yes, please replace this term.

It was corrected. 95% CI.

28-i can not see the analysis and results based on the ANOVA?

Ok, added.

29-table 1 legend seems wrong: please consider this version:
Table 7: Diagnostic indices of each delivered energy level for detecting urinary stone compositions, with their 95% confidence intervals.

Ok, corrected.

30-please add explanation for figure 4 part d to its legend.

Ok, corrected

31- It's suggested that the authors add the naming of A and B for different parts of images of figure 2.

Ok, corrected

32- It's suggested that the authors add the naming of A , B and C for different parts of images of figure 7.

Ok, corrected

4 Revision 3

1 comment · 8 Apr 2023

Editor-in-Chief

MINOR REVISION 8 Apr 2023

Dear authors,
Your article needs Minor Revision.
Regards

5

Revision 4

AUTHORS' RESPONSE TO REVIEWERS

Authors' reply to reviewers

The next pages contain the authors' point-by-point reply to the comments of this revision round, as submitted with the revised manuscript. It is published as the authors wrote it.

> Reviewers' Comments:

> Associate Editor's Comments:

Associate Editor 1:
the authors have corrected many parts. some issues should be addressed:

1-Please change the sentence “The ROCs show the true positive rates (TPR or sensitivity) versus the false positive rates (1- specificity)” to “The ROCs plots sensitivity versus 1-specificity”

Ok, corrected

2-please correct definition of specificity as: “(refers to the probability of a negative test result among subjects without disease)”. [line 231]

Ok, corrected

3-please correct sdgm in table 1 to sgdm.

Ok, corrected

4-please add full term of abbreviations sgdm, adam and RMSPROP in table 2

Ok, corrected

5-please unify the writing style of sgdm, adam and RMSPROP in paper. [various forms in tables]

Ok, corrected

5 Revision 4

1 comment · 11 Apr 2023

Editor-in-Chief

ACCEPT 11 Apr 2023

Dear authors,
Your article is acceptable for publication in IJR.
Regards