Deep Learning Computer Vision Algorithm for Detecting Kidney Stone Composition

Kristian M Black[1] MD, MSCR, Hei Law[2], Ali Aldouhki[1] MBBS, MS,
Jia Deng[2] PhD, Khurshid R Ghani[1], MBChB, MS, FRCS(Urol)

[1]Department of Urology, University of Michigan, Ann Arbor, MI, USA
[2]Department of Computer Science, Princeton University, Princeton, NJ, USA

Authors contact information:

[1]Medical Sciences Unit I, University of Michigan Medical School, 1301 Catherine St, Room 4432, Ann Arbor, MI 48109

[2]Department of Computer Science, Princeton University 35 Olden Street, Princeton, NJ 08540-5233

[3] Medical Sciences Unit I, University of Michigan Medical School, 1301 Catherine St, Room 4432, Ann Arbor, MI 48109

[4] Department of Urology, Room 114W, NCRC Building 16, University of Michigan, 2800 Plymouth Road, Ann Arbor, MI 48109

Corresponding author:

Kristian M Black, MD/MSCR
Email: krismbla@med.umich.edu
Telephone: +1 318-505-3830
University of Michigan, Medical Sciences Unit I, 1301 Catherine St, Room 4432, Ann Arbor, MI 48109

Keywords:

Ureteroscopy, Laser lithotripsy, Holmium, Computer Vision, Artificial Intelligence, Deep Learning

Abstract: 247; **Word count:** 1547

Disclosures: KRG is a consultant for Boston Scientific and Lumenis; KRG has a scientific investigator grant from Boston Scientific. No funding was received for this study.

DR. KRISTIAN BLACK (Orcid ID : 0000-0002-3936-3752)
DR. KHURSHID R GHANI (Orcid ID : 0000-0002-6089-0733)

Article type : Original Article

Article Category: Translational Science

Abstract

Objectives: To assess the recall of a deep learning (DL) method to automatically detect kidney stones composition from digital photographs of stones.

Materials and Methods: 63 human kidney stones of varied compositions were obtained from a stone laboratory including calcium oxalate monohydrate (COM), uric acid (UA), magnesium ammonium phosphate hexahydrate (MAPH/struvite), calcium hydrogen phosphate dihydrate (CHPD/brushite) and cystine stones. At least two images of the stones, both surface and inner core, were captured on a digital camera for all stones. A deep convolutional neural network (CNN), ResNet-101 [ResNet, Microsoft], was applied as a multi-class classification model, to each image. This model was assessed using leave-one-out cross validation with the primary outcome being network prediction recall.

Results: The composition prediction recall for each composition were as follows: UA 94% (n=17), COM 90% (n=21), MAPH/Struvite 86% (n=7), Cystine 75% (n=4), CHPD/Brushite 71% (n=14). The overall weighted recall of the convolutional neural network's composition analysis was 85% for the entire cohort. Specificity and precision for each stone type were as follows: UA [97.83, 94.12], COM [97.62, 95] struvite [91.84, 71.43] cystine [98.31, 75], and brushite [96.43, 75].

Conclusion: Deep convolutional networks can be used to identify kidney stone composition from digital photos with good recall. Future work is needed to see if deep learning can be used for detecting stone composition during digital endoscopy. This technology may enable integrated endoscopic and laser systems that automatically provide laser settings based on stone composition recognition with the goal to improve surgical efficiency.

Introduction

There is increasing interest on optimizing holmium laser settings and techniques like Dusting[1], as ureteroscopy (URS) and laser lithotripsy has become the predominant surgical treatment for urinary stones in North America[2]. Currently, stones are fragmented by selecting pulse energy and frequency to break stones into either fine powder (dusting) or medium-sized fragments for extraction. Laser energy needed to ablate stones varies with stone composition and size [3]. Right now, surgeons manually choose laser settings based on a visual recognition of the stone type and its durility. However, if settings could be automatically calculated based on recognition of stone composition, this could improve the efficiency of lithotripsy. Furthermore, because stone samples are often extracted with baskets for composition analysis to guide management, an endoscopic visualization system reliably determines stone composition would have benefits in reducing operative time and surgical costs.

Computer vision together with deep learning may offer a solution to these unmet needs. Current state-of-the-art approaches to the image classification task, a computer vision task where the computers categorize images, use deep neural networks to extract patterns from an image and make prediction based on the patterns, permitting automatic prediction of outcomes. To date, several studies have demonstrated the value of DL for recognizing pathologic features in diseases such as melanoma and diabetic retinopathy [4,5]. With its emergence as a powerful tool for image-based analysis, we studied the recall of using convolutional neural networks (CNNs) for detecting the composition of five main categories of human kidney stones. Establishing this framework during URS could lead to the automatic selection of laser lithotripsy settings based on real-time stone composition analysis.

Materials and Methods

Human kidney stones of varied compositions were obtained from a stone laboratory in 2018 (Louis C. Herring and Company, Florida) including calcium oxalate monohydrate (COM), uric acid (UA), magnesium ammonium phosphate hexahydrate (MAPH/struvite), calcium hydrogen phosphate dihydrate (CHPD/brushite) and cystine stones. All stones included in this study were preserved in dry conditions in glass vials. Mean stone size was 5.7 mm (±3.5; range 2-18 mm). Dry stones were placed on a green non-reflective background and pictures were taken with a DSLR camera fitted with macro lens (55mm). At least two images of the stones, both surface and inner core, were captured on a digital camera for all stones. Using Photoshop (Adobe, CA), the green non-reflective background was manually removed from each photo using the mask function and saved as JPG files. This was followed by randomly generated computer automated cross-sectional cropping (Figure 1). While a pre-trained segmentation model such as UNet is powerful, it is more suitable for tasks which require pixel-wise predictions. Since we are classifying the stone composition, we picked the models that performed well on image classification. We applied a deep CNN, ResNet-101 [ResNet, Microsoft], as a multi-class classification model, to classify each image crop[6]. The average of the classification scores of 30 random crops was used for final prediction. We sampled crops of different sizes from an image and we resized all of them to size 96 x 96 before we fed them into the deep CNN. We whitened each crop using the RGB mean and standard deviation. The deep CNN was trained with the resized and whitened crops to predict stone composition. All stone images used in the training set were different from those used in the testing set. Since we only had limited amount of data, we used a ResNet-101 that was pretrained on the ImageNet classification dataset, a large-scale image classification dataset, to avoid overfitting. We replaced the fully connected layers in ResNet-101 with a fully connected layer of 128 channels with Batch Normalization and ReLU followed by another fully connected layer of 128 channels, which are both randomly initialized, and a softmax layer for predicting the stone composition. Hence, we used the cross-entropy loss function. During training, we fixed the weights in the convolution layers and only updated the weights in the fully connected layers. We used stochastic gradient descent (SGD) with a learning rate of 0.001, a momentum of 0.9 and a weight decay of 0.0001 to optimize the loss function for 2000 iterations across all stone types. We then reduced the learning rate to 0.0001 for another 2000 iterations. Hyperparameters were not chosen based on cross validation results. We used a batch size of 128. To account for our small image dataset, instead of dividing the images into test and train set, we assessed recall of the network using leave-one-out cross validation method, where we used all stones except one as the training set and tested the network on the remaining one. This was repeated until all stones were tested producing recall averages for each stone type. Because sample size varied between stone compositions, an overall weighted average was also calculated.

Results

A total of 63 stones were used including 17 UA, 21 COM, 7 struvite, 4 cystine, and 14 brushite stones comprising a total of 127 images (figure 2). Stone recognition prediction recall (sensitivity) varied by composition. UA stones had the highest recall at 94% followed by COM stones with 90%. Struvite and cystine stones were classified with moderate recall, correctly identified 86% and 75%, respectively. Lower predictive recall was seen for brushite stones (71%). Overall weighted prediction recall was 85%. Specificity and precision for each stone type are as follows: UA [97.83, 94.12], COM [97.62, 95] struvite [91.84, 71.43] cystine [98.31, 75] brushite [96.43, 75] (table 1). ROC curve, precision-recall curve, and our confusion matrix can be found in figure 3. The training loss per iteration plot for one of the cross-validation experiments is provided as a supplementary figure.

In an attempt to understand the accuracy of stone composition from endoscopic video images, we analyzed still images of 3 COM stones and 1 UA stone taken during a flexible URS case, through our deep CNN. The preliminary results, demonstrating feasibility, were as follows: recall for COM=0.67, precision for COM=0.71; recall for UA=1.0, precision for UA= 0.5.

Discussion

In this pilot study, we have shown that it is possible to predict kidney stone composition from digital photos using computer-vision and DL. Commonly encountered stones such as UA and COM, had higher accuracies than stones such as brushite and cystine. These stones have distinct visual appearances and are often the easiest for humans to identify. The lower prediction scores for other stone compositions may be a reflection of the visual heterogeneity of these stones. Brushite specifically has been noted as a difficult stone composition to classify with computer vision methods due to its high level of intraclass variability[7]. To our knowledge, this is the first report of using CNNs to predict kidney stone composition though other methods such as Raman spectroscopy and autofluorescence have been studied [8,9].

Only one prior study assessed image-based methods to determine kidney stone composition [7]. Serrat et al computed hand-crafted features from each image (e.g. local binary pattern and color histogram) and applied a traditional machine learning approach (random forest) to classify the features. They found an overall composition prediction accuracy of 63%. In their model, pH was also included as an additional feature to improve stone composition prediction. Using CNNs, we were able to produce higher accuracies without incorporating any hand-crafted features. The main advantage of DL is that the CNN autonomously learns to extract features useful for classifying stone composition directly from the image, eliminating manual feature extraction and user bias.

Our study has several limitations. We used only pure stones and additional studies are needed to predict the composition of mixed stones. We studied only still images, whereas during URS, imaging is video-based and includes body wall movement with blood/debris. Future studies will include predicting stone composition based on images taken during URS. Data reported in this study may serve as a benchmark for future comparisons of kidney stone composition detection during URS.

Our work lays the foundation for video-based recognition. Using CNNs, this would be feasible, as previously shown for recognizing and tracking surgical instruments in robotic surgery videos [10]. Another area of study is to determine if computer-vision can accurately detect stone size. Size information can provide feedback on when the optimal fragment size has been achieved for extraction or dusting. Lastly, we only had a limited set of stone imaging data. We hypothesize the recall of stone recognition will improve if the DL network is able to train on a larger set of data.

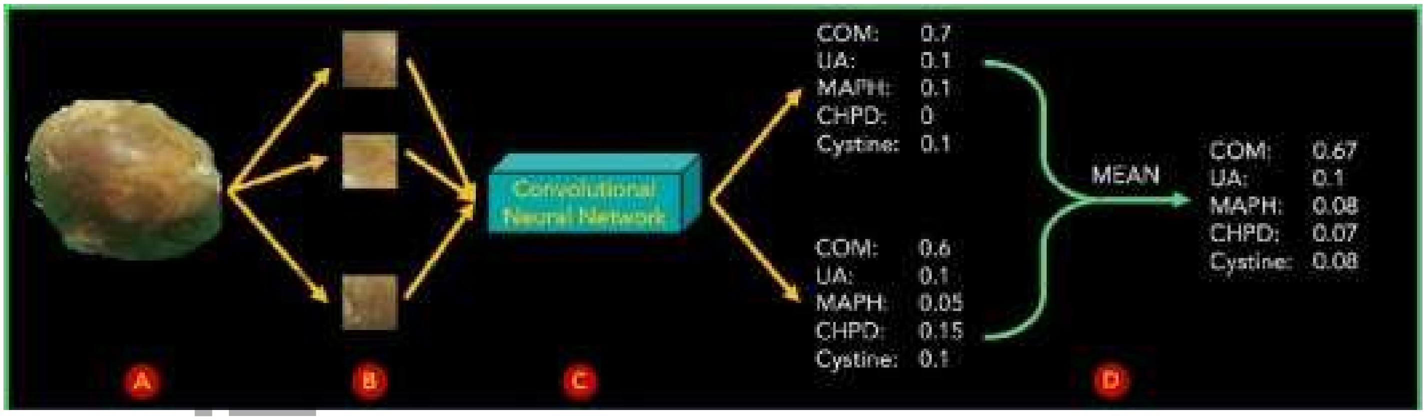
In conclusion, we have shown that a DL computer-vision algorithm can be used to detect the composition of commonly encountered kidney stones. In the future, digital endoscopic platforms that leverage AI and DL techniques could provide a cheaper and faster alternative to traditional stone analysis. Similar systems could be adapted to smartphones to allow office-based stone analysis. The ability to intraoperatively determine stone composition could result in the development of integrated endoscopic and laser systems that automatically provide laser settings based on computer-vision stone characterization with the goal to improve laser lithotripsy efficiency. One caveat that must be noted is the need for computer vision result verification and interpretation by a licensed clinician. While AI systems such as this have the ability to identify pathology, these technologies do not possess the capacity to consider clinical conditions that can impact the pathologic state and therefore should aid, not replace expert opinion.

REFERENCES

1. Dauw CA, Simeon L, Alruwaily AF, et al. Contemporary Practice Patterns of Flexible Ureteroscopy for Treating Renal Stones: Results of a Worldwide Survey. *J Endourol.* 2015;29(11):1221-1230.
2. Ordon M, Urbach D, Mamdani M, Saskin R, RJ DAH, Pace KT. The surgical management of kidney stone disease: a population based time series analysis. *J Urol.* 2014;192(5):1450-1456.
3. Vassar G.J. TJM, Glickman R.D. Holmium:YAG lithotripsy efficiency varies with energy density. *J Urol.* 1998;160(2):471-476.
4. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115-118.
5. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA.* 2016;316(22):2402-2410.
6. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27-30 June 2016, 2016.
7. Serrat J, Lumbreras F, Blanco F, Valiente M, López-Mesas M. myStone: A system for automatic kidney stone classification. *Expert Systems with Applications.* 2017;89:41-51.
8. Schutz J, Miernik A, Brandenburg A, Schlager D. Experimental Evaluation of Human Kidney Stone Spectra for Intraoperative Stone-Tissue-Instrument Analysis Using Autofluorescence. *J Urol.* 2019;201(1):182-187.
9. Miernik A, Eilers Y, Bolwien C, et al. Automated analysis of urinary stone composition using Raman spectroscopy: pilot study for the development of a compact portable system for immediate postoperative ex vivo application. *J Urol.* 2013;190(5):1895-1900.
10. Law H, Ghani K, Deng J. Surgeon Technical Skill Assessment using Computer Vision based Analysis. Proceedings of the 2nd Machine Learning for Healthcare Conference; 2017; Proceedings of Machine Learning Research.

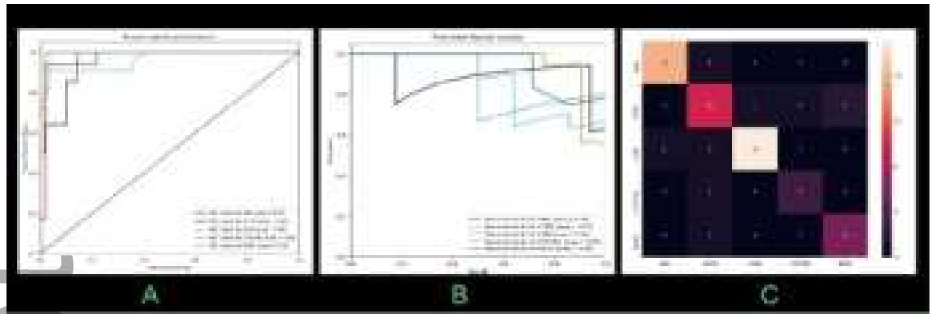
Table 1. Recognition performance measures by stone composition type for ResNet-101 CNN.

[redacted]			
[redacted]	Recall (Sensitivity)	Specificity	Precision (PPD)
[redacted]	94.12	97.83	94.12
[redacted]	90.48	97.62	95.00
[redacted]	71.42	91.84	71.43
[redacted]	75.00	98.31	75.00
[redacted]	85.71	96.43	75.00



bju_15035_f1.png

bju_15035_f2.png



bju_15035_f3.png