Towards an automated classification method for ureteroscopic in vivo kidney stone images using machine learning techniques

A. Martínez-Gonzalez [1] D.-H. Trinh [2] J. Hubert [3] V. Estrade [3] C. Daul [2] G. Ochoa-Ruiz [4]

Abstract

Urolithiasis is a common disease around the world and its incidence has been growing every year. The accuracy of the diagnosis is crucial for the prescription of an appropriate treatment that can eliminate the stones and diminish future relapses. This paper presents an effective supervised learning method to automate and improve the accuracy of the classification of kidney stones, as well as a dataset consisting of kidney stone images captured with ureteroscopes. In the proposed method, the image features that are visually exploited by urologists to distinguish diferent types of kidney stones are analyzed and encoded as vectors and fed to Random Forest classifier. The obtained classification results (89% accuracy) outperforms previous methods by more than 10% and shows that the implementation of automated image analysis techniques is feasible in the uretoscopic practice.

1. Introduction

Kidney stones that cannot drain naturally from the urinary tract are destroyed during an endoscopic intervention called ureterorenoscopy. Digital endoscopes allow for the insertion of a laser to fragment and remove the stones from the urinary tract (a laser lithotripsy technique called "dusting"). The biochemical composition of some of the collected fragments is systematically analyzed to understand the metabolic cause of the kidney stone formation. Moreover, morpho-constitutional analyses carried out under the microscope and infrared spectrometer can reveal the composition of pure or multi-layered stones and allow to establish treatments (e.g., diets, drug treatment) to reduce relapses in terms of stone formation (Daudon & Jungers, 2012).

[*]Equal contribution [1]Universidad Autónoma de Guadalajara, Mexico [2]Université de Lorraine and CNRS), Nancy, France [3]IADI-UL-INSERM (U1254), Nancy, France [4]Tecnológico de Monterrey,Mexico. Correspondence to: G. Ochoa-Ruiz <gilberto.ochoa@tec.mx>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 108, 2020.

Although this widespread technique is useful, it presents two drawbacks: (1) removing all stone fragments can be a tedious procedure lasting from thirty minutes to an hour, and (2) morpho-constitutional analytical results are usually available several weeks after the endoscopy, hindering the immediate treatment that is usually recommended depending on the kidney stone type (Cloutier et al., 2015).

Therefore, an intra-operative morphological and automated classification of kidney stones could alleviate these problems, since a description of the visual aspect (e.g., morphology, colors and textures) of their surface and cross-section is available for each class for the images acquired during in vivo ureteroscopic interventions (see the tables given in (Estrade et al., 2017) and Fig.1). However, this operator-dependent recognition requires a great deal of experience due to the high inter-class similarities and intra-class variations, and can only be achieved by few specialists, whereas the urologists deal with urolithiasis in a daily basis.

Thus, in this work we present an machine learning-based approach for classifying in vivo images obtained using ureteroscopes (URF-V and URF-V2 endoscopes from the Olympus). We have build a dataset consisting of 125 images of the most prevalent kidney stones, divided in three classses (i.e., Whewellite, Weddellite, and Uric Acid) which was validated and labeled by urologists with expertise in the domain. We carried a thorough feature analysis to determine the most discriminant information from the samples and performed an ablation study using various classifiers to validate the feature extraction. We then chose the best performing model (based on a Random Forest classifier) and compare it with other works in the state of the art.

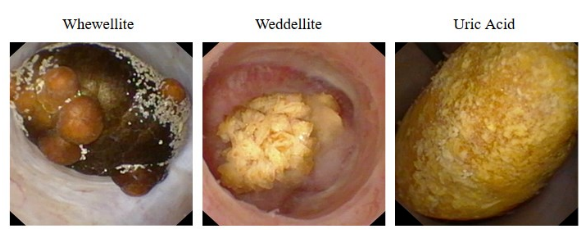

Figure 1. In vivo kidney stone images belonging to three different classes: Whewellite, Weddellite and Uric Acid.

2. Related Works and Motivation

Few automated methods for classifying kidney stones images have been published to date (Kazemi & Mirroshandel, 2017) and those using image analysis techniques are rather scant due to the lack of large image databases. For instance, the work of (Serrat et al., 2017) presented a method in which texture features (coded as Local Binary Patterns, LBP) and color features (RGB histograms) were extracted and injected into a Random Forest classifier. Although the average accuracy of this approach was relatively low (63%), the results showed that the texture and color information are discriminant enough to automate the classification (Pless et al., 2019). A continuation of the previous work improved the classification results by using deep learning techniques (Torrell Amado, 2018), as did the work presented in (Black et al., 2020). However, the two latter approaches yielded limited improvements due to the small size of the dataset.

Another major limitation of these previous works lies in the fact that the utilized images were neither acquired in vivo (in the patient) nor with an ureteroscope. Fragments of kidney stones were placed into a closed device allowing for strictly controlled illumination conditions. The stones were also placed in such a way that the surface and cross-section of the fragments were fully visible in the captured images.

Conversely, in clinical scenarios the imaging conditions are highly uncontrolled: the acquisition angles and distances are somewhat random since the ureteroscope position with respect to a fragment cannot be easily controlled; the images can be affected by motion blur and specular reflections; and the scene illumination depends notably on the ureteroscope position and the surrounding tissue. To the best of our knowledge, no other works have tackled the problem of analyzing and automatically classifying images acquired under the aforementioned conditions.

In this sense, our main contributions are threefold: 1. A new dataset of in vivo kidney stones images obtained using digital endoscopes. 2. A thorough analysis of the most important characteristics needed for classifying samples with high inter-class similarities and intra-class variations and 3. A method based on machine learning techniques to demonstrate the feasibility of an automated method to tackle this problem in a clinical setting.

3. Proposed Method and Results

It has been proven (Estrade et al., 2017; Serrat et al., 2017) that textures and colors are the most discriminant features to visually differentiate the kidney stone classes. Therefore, texture and color features were extracted using Scikit-Learn and gathered in descriptor vectors of 40 components each, where 10 components correspond to texture information and the other 30 to color information.

Table 1. Average classification accuracy of the tested model.

Sub-dataset	Classifier	Features	Accuracy
Section	RF	LBP + e(HSV)	89%
Surface	RF	LBP + e(HSV)	79%

The surface and cross-section images were used as two separate training and testing phases. Patches of 200×200 pixels were cropped from the original images to increase the number of images in the sub-datasets (781 in total). The patches only contain information of the 3 kidney stone classes (the background such as organs tissue is not visible in the cropped images) as it has been observed that both texture and color patterns are very similar for different regions of the kidney stone surfaces and cross-sections (Estrade et al., 2017) (Serrat et al., 2017).

We made use of a Random Forest (RF) classifier due to their accurate performance with small datasets ($<$1,500 samples) and good performance with stone-textured images, as shown in (Serrat et al., 2017). Other models were tested as well, but we decided to present only the results obtained using RF for conciseness, which are summarized in Table I. Scikit-Learn's Random and Grid Search were used to find the most efficient parameters. For the sub-dataset containing the surface images, a Random Forest with 160 Decision Trees (DTs) was built, while for the cross-section sub-datset, only 40 DTs were needed to obtain the most accurate results without overfitting. Due to the reduced size of the dataset, the K-fold cross validation strategy was used to train and test the classification models.

As it can be observed in Table I, we obtained very good results for both surface and section images; in contrast, (Serrat et al., 2017) obtained an average precision 63%, whereas (Torrell Amado, 2018) obtained an improvement of about 10% making use of a deep neural net.

4. Discussion and Future Work

In this paper we presented a method for classifying in vivo kidney stone images. Compared to previous works, our dataset is encompassed by more challenging images acquired via uretoscopes. Despite of issues like widely varying illumination conditions and blurring, we have demonstrated that an effective feature extraction and model selection can yield superior results. The images were obtained by urologists in several hospitals and the results presented here have been corroborated with a team of experts, who consider that the presented method has a strong clinical potential.

As future work, we plan to incorporate other types of renal calculi not acquired at the time of this study, and to explore the use of deep learning architecture making use of a few-shot learning approach.

References

Black, K. M., Law, H., Aldoukhi, A., Deng, J., and Ghani, K. R. Deep learning computer vision algorithm for detecting kidney stone composition. *BJU International*, 2020 (2020), 2020. doi: 10.1111/bju.15035.

Cloutier, J., Villa, L., Traxer, O., and Daudon, M. Kidney stone analysis: ”give me your stone, i will tell you who you are!”. *World Journal of Urology*, 33(2):157–169, 2015. ISSN 1433-8726.

Daudon, M. and Jungers, P. *Stone Composition and Morphology: A Window on Etiology*, pp. 113–140. Springer London, London, 2012. ISBN 978-1-4471-4387-1.

Estrade, V., Daudon, M., Traxer, O., and Méria, P. Pourquoi l'urologue doit savoir reconnaître un calcul et comment faire ? les bases de la reconnaissance endoscopique. *Progrès en Urologie - FMC*, 27(2):F26 – F35, 2017. ISSN 1761-676X.

Kazemi, Y. and Mirroshandel, S. A. A novel method for predicting kidney stone type using ensemble learning. *Artificial intelligence in medicine*, 84:117–126, 2017.

Pless, M. S., Williams, J. C., Andreassen, K. H., Jung, H. U., Osther, S. S., Christensen, D. R., and Osther, P. J. S. Endoscopic observations as a tool to define underlying pathology in kidney stone formers. *World Journal of Urology*, 37(10):2207–2215, 2019. ISSN 1433-8726.

Serrat, J., Lumbreras, F., Blanco, F., Valiente, M., and López-Mesas, M. mystone: A system for automatic kidney stone classification. *Expert Systems with Applications*, 89:41 – 51, 2017. ISSN 0957-4174.

Torrell Amado, A. Metric learning for kidney stone classification. *Universitat Autònoma de Barcelona. Escola d'Enginyeria*, 2018.