Comparing fusion strategies for kidney stone composition identification

Elias Villalvazo-Avila[1], Francisco Lopez-Tiro[1], Daniel Flores-Araiza[1], Jonathan El-Beze[2], Jacques Hubert[2], Gilberto Ochoa-Ruiz[1], Christian Daul[3]

[1]Tecnologico de Monterrey, School of Engineering and Sciences, Mexico
[2]CHU Nancy, Service d'urologie de Brabois, Nancy, France
[3]Centre de Recherche en Automatique de Nancy, Université de Lorraine, France

Abstract

Identification of the type of kidney stones (i.e. morpho-constitutional analysis) is of paramount importance to prescribe an appropriate treatment and to prevent future relapses. However, this procedure is time-consuming, expensive, and requires a great deal of experience. Previous work has shown that the identification of kidney stones is a problem that can be solved with machine and deep learning strategies. However, most of these methods make use of single images (both spatial and temporal terms) and they do not follow the morphological approach to kidney stones identification: namely, they either take surface or section information separately and there is no clear approach for combining such information. Herein, we investigate means for producing a single description of kidney stone samples, in a manner that is more amenable or similar to the procedure the urologists and/or biologists follow when they perform this classification. We do so by exploring fusion methods based on multi-view learning, where we train a neural network for only surface images of the stone, and a second neural network is trained for section images of the stone. Several fusion techniques are tested to fully assess which combination outperforms the single-view models from the state of the art. Compared to the state of the art, the obtained results show an improvement of 10% in weighted precision and recall, and an increase in the robustness of classification of all stone types.

1. Introduction

Urolithiasis refers to the formation of kidney stones that cannot be expelled from the urinary tract. This is a medical condition that has been increasing over the last few years [17, 11]. Urolithiasis is caused by multiple factors, where diet is the most important, but also genetic inheritance, water intake, and a sedentary lifestyle could promote the formation of kidney stones [6]. Morpho-constitutional analysis (MCA) is the most important method for kidney stone identification. MCA is a combination of a visual examination under the microscope of the stone's texture, appearance, and color (surface and section views), and a biochemical analysis by Fourier Transform Infrared Spectroscopy (FTIR) [3]. If carried out properly, a timely treatment (diet adaptation, surgery) can be prescribed for each patient, reducing the risk of stone recurrence [3]. However, MCA has a major drawback: the results of this analysis are often available within a time frame of one or two months. For this reason, more urologists seek to visually identify the morphology of kidney stones only with the help of the image displayed on the screen [12] during the removal process (Endoscopic Stone recognition (ESR)). However, this visual analysis requires a great deal of experience due to the high similarities between stones that only a limited number of specialists can reliably identify [1].

Different Machine Learning (ML) approaches have been proposed [10, 5, 12, 14] for the classification of kidney stones, demonstrating that it is a problem that can be solved with traditional and deep learning techniques with very encouraging results. However, most of these models were trained on ex-vivo stones placed in controlled environments, whereas in reality, images may suffer from motion blur, reflections, illumination variations, as occurs in common practice during an endoscopic imaging session. Moreover, there is no ordered manner of mixing surface and section information for exploiting the visual information in a way that a specialist would do it. Besides, in most cases the amount of training data available is limited, thus these contributions use data augmentation techniques to increase the amount of input data, but some limitations have not been addressed. Nonetheless, these works [10, 5, 12] have demonstrated the potential of automatic ESR in an in-vivo dataset.

Multi-View (MV) classification is an area of ML that combines features from different sources or feature subsets, known as views, to identify objects with higher accuracy,

since diverse characteristics are extracted, synthesized and combined [8]. This variant of learning can improve the performance by optimizing multiple functions, one per view, and in that way, information can be obtained from different perspectives of the same data inputs. Moreover, MV can also be applied to Convolutional Neural Networks (CNNs) to boost the performance in situations where a single image does not yield sufficiently discriminative information for accurate classification by combining useful information from different views, so more comprehensive representations may be learned yielding a more effective classifier [13].

In this work, we leverage recent strides in deep learning that have sought to combine information from multiple views. Previous work [7, 15] demonstrates that MV learning can be applied in medical images, with promising results. We believe that such an approach can be beneficial/optimal for the real-time identification of kidney stones, by maximizing the amount of information that the model can use for classification. Through several experiments, we demonstrate that by combining the information of surface and section views in the same model, we can obtain a method that is more explainable and similar to what specialists do in clinical practice (i.e., MCA).

2. Materials and Methods

2.1. Kidney stone dataset

The ex-vivo dataset includes 305 kidney stone images acquired (two reusable digital flexible ureteroscopes from Karl Storz using video columns: Storz Image 1 Hub and Storz image1 S) and labeled manually by the urologist Jonathan El Beze[2]. To reproduce in-vivo conditions, the experimental setup used in this work consists of a small diameter tube where the inner walls were covered with a yellowish film to display the appearance of the urinary tract (for more details, see [4]). The ex-vivo dataset consists of three subsets: the first subset consist of 177 surface images, 128 section images for the second subset, and the third subset of 305 images (177 section + 128 surface) of the six kidney stone types with the highest incidence: Type Ia (Whewellite, WW), Type IIb (Weddellite, WD), Type IIIb (Acide Urique, AU), Type IVc (Struvite, STR), Type IVd (Brushite, BRU), and Type Va (Cystine, CYS). Images of this dataset are shown in Fig. 1.

Classification of kidney stones is not performed on whole images [14, 16, 2, 10]. Therefore, in this work as in previous works, patches of 256×256 pixels were cropped from the original images to increase the size of the training dataset (for more details, see [9]). However, the number of resultant patches for each class is imbalanced (due to the changing fragment sizes, image resolution, and the number of images in the original dataset). In order to balance the

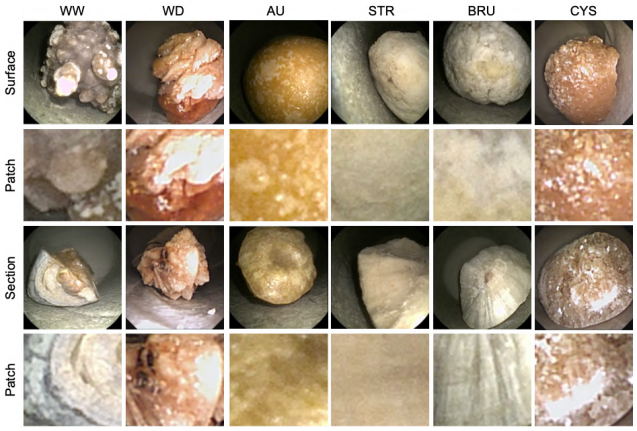


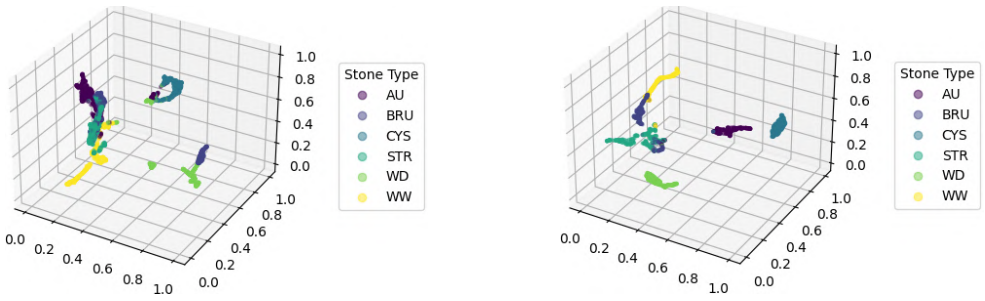
Figure 1: Examples of ex-vivo kidney stones images. From left to right: WW, WD, AU, STR, BRU, and CYS. The surface view is in the first row (top row), and their respective generated patches are in the second row. The section view is in the third row, and their respective generated patches are in the fourth row.

number of patches per class, a random sampling approach was used. This step yielded a total of 1000 patches per class (WW, WD, AU, STR, BRU, and CYS) and view (surface, section, and mixed). This new dataset was then split into 19200 images (80%) for training and validation, and 4800 images (20%) for the test. In order to limit the overfitting produced by the small size of the available training dataset, data augmentation was heavily performed. Additional patches were obtained by applying geometrical transformations (patch flipping, affine transformations, and perspective distortions). The number of patches increased from 19200 to 153600 using data augmentation (10% of the original patches were kept for test purposes). The patches were also "whitened" using the mean m_i and standard deviation σ_i of the color values I_i in each channel ($I_i^w = (I_i - m_i\sigma_i)$, with $i = R, G, B$).

2.2. Methods

2.2.1 Pre-training Stage

Previous approaches have used Deep Learning architectures such as AlexNet, or VGG16 for assessing the kidney stone classification task [9, 12]. For this contribution, the previously-mentioned architectures are used for the creation of the MV models. This network was trained on the entire training data, mixing both surface and section patches, and served as a baseline or comparison for the multi-view implementations introduced in this paper. Once trained, the feature extraction layers of this single-view network are frozen to ensure that each branch from the multi-view model extracts the same features and that any variation in the performance will rely on the unfrozen layers (fusion and fully-connected layers).



(a) SV-AlexNet (b) MV-AlexNet-max

Figure 2: Test results in (a) Single-View (SV) AlexNet architecture with no fusion, and (b) in Multi-View AlexNet architecture with a max-pool fusion layer.

2.2.2 Multi-View Model

The frozen layers (part A of the MV model) are duplicated. In this way, one copy will process only images of the surface of the stone, while the other copy will process images of the section view. These frozen layers are connected to a fusion layer, which will be responsible for mixing the information of the two views. In this work, the two late-fusion methods proposed in [8] are explored. Lastly, the output of the fusion layer is connected to part B of the multi-view model, which merely consists of the classifier. The proposed model is shown in Fig. 3. To make a direct comparison of the feasibility and performance of this architecture against previous works, we used the same hyper-parameters as [12]. Cross-entropy loss is used to compute the classification loss, and optimization is performed using Adam optimizer with a learning rate of $2e^{-4}$. Batch-size selected was 64 for both multi-view and single-view networks. The experiments and implementation of this network were performed using Pytorch v1.10.2.

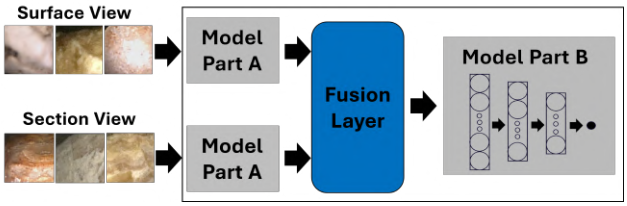


Figure 3: Proposed Multi-View model. Part A corresponds to feature extraction layers, Part B corresponds to classification layers. A fusion layer is added to combine information from different views.

3. Results and Discussion

Several experiments were performed to assess the ability of MV models to predict the kidney stone class, combining information from surface and section views, as done during an ESR procedure [5]. Precision (P) and Recall (R) metrics are determined for each class individually. The results reported for both fusion techniques in the multi-view models show that combining information from different classifiers yields significant results compared to the single-view classifiers. For instance, for a single-view AlexNet network, the results obtained were 0.84 and 0.83 for precision, and recall, respectively. In contrast, MV networks, independent of the fusion strategy, performed better compared to the other experiments. Table 1 shows the scores for all the models used for this work, and Fig. 2a, and Fig. 2b show how stone type clusters are distributed for both SV and MV networks. One disadvantage of using concatenation as fusion strategy is that the number of features of the first layers of the classifier increases considerably, limiting its implementation on systems with reduced memory.

Table 1: Weighted average metrics comparison for section, surface, and mixed patches. MV-AlexNet-max: Multi-View network with max-pool as fusion strategy. MV-VGG16-max: Multi-View network with max-pool as fusion strategy. MV-AlexNet-conc: Multi-View network with concatenation as fusion strategy. SV-AlexNet: Single-View AlexNet network. SV-VGG16: Single-View VGG16 network.

	Surface		Section		Mixed	
Classifier	**P**	**R**	**P**	**R**	**P**	**R**
MV-AlexNet-max	–	–	–	–	**0.95**	**0.94**
MV-VGG16-max	–	–	–	–	**0.94**	**0.94**
MV-AlexNet-conv	–	–	–	–	**0.94**	**0.93**
SV-AlexNet	0.77	0.71	0.88	0.87	0.84	0.83
SV-VGG16	0.79	0.70	0.89	0.89	0.83	0.81

4. Conclusion and future work

We showed that by mixing information from different views, it is possible to train more accurate models for predicting kidney stone composition from images obtained from ureteroscopy. Thus, AI technology can be included in the current stone removal workflow, speeding up preven-

tive diagnosis measures. However, we make use of a very-limited ex-vivo dataset in a simulated environment. We aim to solve this problem by applying metric learning in future work to tackle the amount of data that we require for training, as well as to increase inter-class separability.

Acknowledgments

The authors wish to thank the AI Hub and the CIIOT at ITESM for their support for carrying the experiments reported in this paper in their NVIDIA's DGX computer.

References

[1] C Bergot, G Robert, J-C Bernhard, J-M Ferrière, H Bensadoun, G Capon, and V Estrade. Base pédagogique de la reconnaissance endoscopique des calculs, étude prospective monocentrique. *Progrès en Urologie*, 29(6):312–317, 2019. 1

[2] Kristian M Black, Hei Law, Ali Aldoukhi, Jia Deng, and Khurshid R Ghani. Deep learning computer vision algorithm for detecting kidney stone composition. 2020. 2

[3] Michel Daudon, Paul Jungers, Dominique Bazin, and James C Williams. Recurrence rates of urinary calculi according to stone composition and morphology. *Urolithiasis*, 46(5):459–470, 2018. 1

[4] Jonathan El Beze, Charles Mazeaud, Christian Daul, Gilberto Ochoa-Ruiz, Michel Daudon, Pascal Eschwège, and Jacques Hubert. Evaluation and understanding of automated urolithiasis recognition methods. *BJU International*, 2022. 2

[5] Vincent Estrade, Michel Daudon, Emmanuel Richard, Jean-Christophe Bernhard, Franck Bladou, Gregoire Robert, and Baudouin Denis de Senneville. Towards automatic recognition of pure & mixed stones using intraoperative endoscopic digital images. *BJU International*, abs/2105.10686, 2021. 1, 3

[6] Justin I Friedlander, Jodi A Antonelli, and Margaret S Pearle. Diet: from food to stone. *World journal of urology*, 33(2):179–185, 2015. 1

[7] Krzysztof J. Geras, Stacey Wolfson, Yiqiu Shen, Nan Wu, S. Gene Kim, Eric Kim, Laura Heacock, Ujas Parikh, Linda Moy, and Kyunghyun Cho. High-resolution breast cancer screening with multi-view deep convolutional neural networks, 2018. 2

[8] D. Li and Y. Tian. Multi-view metric learning for multi-instance image classification, 2016. 2, 3

[9] Francisco Lopez, Andres Varelo, Oscar Hinojosa, Mauricio Mendez, Dinh-Hoan Trinh, Yonathan ElBeze, Jacques Hubert, Vincent Estrade, Miguel Gonzalez, Gilberto Ochoa, et al. Assessing deep learning methods for the identification of kidney stones in endoscopic images. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2778–2781. IEEE, 2021. 2

[10] Adriana Martínez, Dinh-Hoan Trinh, Jonathan El Beze, Jacques Hubert, Pascal Eschwege, Vincent Estrade, Lina Aguilar, Christian Daul, and Gilberto Ochoa. Towards an automated classification method for ureteroscopic kidney stone images using ensemble learning. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1936–1939. IEEE, 2020. 1, 2

[11] Anmar Nassir, Hesham Saada, Taghreed Alnajjar, Jomanah Nasser, Waed Jameel, Soha Elmorsy, and Hattan Badr. The impact of stone composition on renal function. *Urology Annals*, 10(2):215, 2018. 1

[12] Gilberto Ochoa-Ruiz, Vincent Estrade, Francisco Lopez, Daniel Flores-Araiza, Jonathan El Beze, Dinh-Hoan Trinh, Miguel Gonzalez-Mendoza, Pascal Eschwège, Jacques Hubert, and Christian Daul. On the in vivo recognition of kidney stones using machine learning. *arXiv preprint arXiv:2201.08865*, 2022. 1, 2, 3

[13] Marco Seeland and Patrick Mäder. Multi-view classification with convolutional neural networks. *PLOS ONE*, 16(1), 2021. 2

[14] Joan Serrat, Felipe Lumbreras, Francisco Blanco, Manuel Valiente, and Montserrat López-Mesas. mystone: A system for automatic kidney stone classification. *Expert Systems with Applications*, 89:41–51, 2017. 1, 2

[15] Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J. van Riel, Mathilde Marie Winkler Wille, Matiullah Naqibullah, Clara I. Sánchez, and Bram van Ginneken. Pulmonary nodule detection in ct images: False positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, 35(5):1160–1169, 2016. 2

[16] Alejandro Torrell Amado. Metric learning for kidney stone classification. 2018. 2

[17] Adie Viljoen, Rabia Chaudhry, and John Bycroft. Renal stones. *Annals of clinical biochemistry*, 56(1):15–27, 2019. 1