


# Towards automatic recognition of pure and mixed stones using intra-operative endoscopic digital images

Vincent Estrade<sup>1</sup> , Michel Daudon<sup>2</sup>, Emmanuel Richard<sup>3</sup>, Jean-Christophe Bernhard<sup>1</sup>, Franck Bladou<sup>1</sup>, Grégoire Robert<sup>1</sup> and Baudouin Denis de Senneville<sup>4</sup>

<sup>1</sup>Department of Urology, CHU Pellegrin, Bordeaux, <sup>2</sup>Department of Multidisciplinary Functional Explorations, AP-HP, Tenon Hospital, INSERM UMRS 1155, Sorbonne University, Paris, <sup>3</sup>INSERM, BMGIC, U1035, CHU Bordeaux, University of Bordeaux, Bordeaux, and <sup>4</sup>IMB, UMR CNRS 5251, University of Bordeaux, Talence Cedex, France

## Objective

To assess automatic computer-aided *in situ* recognition of the morphological features of pure and mixed urinary stones using intra-operative digital endoscopic images acquired in a clinical setting.

## Materials and Methods

In this single-centre study, a urologist with 20 years' experience intra-operatively and prospectively examined the surface and section of all kidney stones encountered. Calcium oxalate monohydrate (COM) or Ia, calcium oxalate dihydrate (COD) or IIb, and uric acid (UA) or IIIb morphological criteria were collected and classified to generate annotated datasets. A deep convolutional neural network (CNN) was trained to predict the composition of both pure and mixed stones. To explain the predictions of the deep neural network model, coarse localization heat-maps were plotted to pinpoint key areas identified by the network.

## Results

This study included 347 and 236 observations of stone surface and stone section, respectively; approximately 80% of all stones exhibited only one morphological type and approximately 20% displayed two. A highest sensitivity of 98% was obtained for the type 'pure IIIb/UA' using surface images. The most frequently encountered morphology was that of the type 'pure Ia/COM'; it was correctly predicted in 91% and 94% of cases using surface and section images, respectively. Of the mixed type 'Ia/COM + IIb/COD', Ia/COM was predicted in 84% of cases using surface images, IIb/COD in 70% of cases, and both in 65% of cases. With regard to mixed Ia/COM + IIIb/UA stones, Ia/COM was predicted in 91% of cases using section images, IIIb/UA in 69% of cases, and both in 74% of cases.

## Conclusions

This preliminary study demonstrates that deep CNNs are a promising method by which to identify kidney stone composition from endoscopic images acquired intra-operatively. Both pure and mixed stone composition could be discriminated. Collected in a clinical setting, surface and section images analysed by a deep CNN provide valuable information about stone morphology for computer-aided diagnosis.

## Keywords

morpho-constitutional analysis of urinary stones, endoscopic diagnosis, automatic recognition, deep learning, aetiological lithiasis, #Urology, #EndoUrology, #KidneyStones, #UroStone

## Introduction

Modern endoscopic treatment of urinary stones now relies on laser (holmium:YAG) fragmentation of stones, which can be performed using 'popcorn' [1], 'dusting' modes [2], or more recently by means of thulium fibre laser (TFL) [3,4]. Laser fragmentation may destroy the morphology of the targeted

stone [5], however, analysis of stone morphology is crucial for the aetiological diagnosis of stone disease [6–8] and for the development of novel immediate postoperative treatment strategies that will eliminate potential residual stone fragments with a lower probability of relapse [9]. For example, calcium oxalate monohydrate (COM [or Ia]), or calcium oxalate dihydrate (COD [or IIb]), criteria would

support the prescription of an immediate diet containing potassium citrate, as reported in Soygür et al. [10]. Recognition of uric acid (UA [or IIIb]) morphological criteria would steer the medical decision towards postoperative urinary alkalinization with potassium citrate or sodium bicarbonate to dissolve residual fragments [11].

The complete morphological analysis workflow may typically include the following two complementary steps:

1. An intra-operative step, which is conducted by a urologist and involves an endoscopy-based examination of the morphology of entire stones *in situ* before their destruction. This step is commonly referred to as endoscopic stone recognition (ESR) [12]. Endoscopic images can be conveniently obtained before (surface image) and after (section images) fragmentation, thus providing valuable morphological information. Estrade et al. [12] recently showed that ESR allowed the identification of the following morphologies: COM, also referred to as types Ia, Ib, Id or Ie (subscripts in the Latin alphabet differentiate morphological subtypes, each being associated with a specific aetiology), COD (or IIa/IIb), UA (or IIIa/IIIb), carapatite (or IVa1), carapatite and struvite (or IVb), brushite (or IVd), and cystine (or Va).
2. A postoperative step, which is performed by a biologist and consists of collecting morpho-constitutional stone information based on both microscopic morphological, i.e., binocular magnifying glass, and spectrophotometric infrared recognition (Fourier transform infrared spectroscopy [FTIR] analysis) [6–8].

The international morpho-constitutional classification of urinary stones includes seven groups (denoted by roman numerals 'I' to 'VII'), each being associated with a specific crystalline type (I = whewellite, II = weddellite, III = uric acid and urates, IV = calcium and non-calcium phosphates, and V = cystine. Groups VI and VII are devoted to other stones). Each group comprises several subgroups that differentiate morphologies and aetiologies for a given crystalline type. Furthermore, urinary stones have mixed morphologies, that is, they include at least two morphologies (almost half of all cases are concerned). The interested reader is referred to Corrales et al. [13] for additional information about the international morpho-constitutional classification of urinary stones.

Recently, an artificial intelligence (AI) algorithm applied to various types of microscopic images of stones *ex vivo* proved to be a promising asset for automatic ESR using both peri- and postoperative images. While Serrat et al. [14] fed texture and colour features of stones into a random forest classifier, Black et al. [15] obtained much improved scores using a deep convolutional neural network (CNN). However, both approaches used *ex vivo* stone fragments placed in a controlled environment. Images were not disturbed by

motion blur, specular reflections or scene illumination variations, as occurs in common practice during an intra-operative endoscopic imaging session. More recent works demonstrated the potential of automated ESR approaches using *in vivo* images acquired in clinical conditions with ureteroscopes on three types of pure stones – Ia/COM, IIb/COD, and IIIb/UA – from 125 kidney stone images [16,17].

Morphological examinations of the entire stone before its destruction provide the best diagnostic agreement [6–8,12,13]. Moreover, Corrales et al. [13] showed that almost half of all urinary stones are of mixed morphologies with two or even three different crystalline components. AI applications must therefore be improved to meet this challenge. The present study has three objectives. Firstly, we aimed to report the preliminary results of the automatic ESR of the morphological components of both pure and mixed urinary stones (these mixed stones being composed of two morphologies in the scope of this study), *in situ*, and using intra-operative endoscopic images acquired in a clinical setting; the images used in this project were captured in an uncontrolled environment by means of ureteroscopes. In addition, the overall performance of a deep neural network was assessed in this setting. Secondly, we aimed to analyse the diagnostic scores computed from images obtained before and after laser fragmentation. Thirdly, we attempted to understand decisions made by a deep neural network in this setting. Specifically, a common problem of deep CNNs is their inability to explicitly display what the model has learned, hence, they are often named 'black box' algorithms. Predictions computed from deep CNNs are in turn hard to explain. Currently there is a growing interest in the development of robust validation procedures to address this key issue. In the present study, we illustrate the usefulness of providing deep CNN-algorithm based 'attention' maps to understand where the algorithm was 'looking' in the endoscopic image when it took its decision.

## Materials and Methods

### Study Design

A urologist (V.E.; 20 years' experience) prospectively examined the intra-operative endoscopic digital images of stones acquired between January 2018 and November 2020 in a single centre using a flexible digital ureterorenoscope (Olympus URF-V CCD sensor). The endoscopic examination included a visual observation of the stone surface. Then, a laser-induced stone split in two parts was performed (laser [holmium:YAG] parameters: frequency = 5 Hz; energy = 1.2–1.4 J; power = 6–7 W; pulse length = short; fibre diameter = 230 or 270 µm). A second visual observation of the section was then performed. An additional fragmentation session was carried out when needed, thus allowing fragmentation of all types of pure and mixed stones. Subsequently, ESR was

confirmed by means of microscopic observations of laser-fragmented stones based on both morphological (binocular magnifying glass) and infrared (FTIR) analyses. The study adhered to all local regulations and data protection agency recommendations (National Commission on Data Privacy requirements). Patients were informed that their data would be used anonymously.

### Morphological Criteria

Morphological criteria were collected and classified according to recommendations outlined in Estrade *et al.* [12]. Stones composed of Ia/COM, IIB/COD and IIIb/UA morphologies were selected. Hence, five morphology classes were included in this study, with three pure stones (Ia/COM, IIB/COD and IIIb/UA) and two mixed stones divided into two morphologies (Ia/COM + IIB/COD and Ia/COM + IIIb/UA).

### Computer-Assisted Endoscopic Stone Recognition Analysis

#### Generation of Annotated Datasets

Two annotated datasets were generated: the first comprised surface images (referred to as the 'surface dataset' hereafter) and the second contained section images (referred to as the 'section dataset' hereafter). All images were automatically cropped and resampled to an equal size of 256 by 256 pixels, then served as input of the automatic ESR algorithm.

#### Automatic Endoscopic Stone Recognition Algorithm

A deep CNN was trained to predict the composition of both pure and mixed stones. Used as a multiclass classification model, the deep CNN was a ResNet-152-V2 [18]. The optimizer algorithm for training the deep-learning model was Adam ( $\text{learning\_rate} = 0.001$ ) [19]. The loss function was a categorical cross-entropy ( $\cdot$ ). The batch size was eight, and 100 epochs were performed. To improve the ability of the network to generalize, the training dataset was expanded through data augmentation. In our implementation, horizontal/vertical flips and affine transformations, including random combinations of scaling ( $\text{range} = 0.3$ ), rotation ( $\text{range} = 50^\circ$ ), and translation ( $\text{range} = 0.2$  of total width/height) were applied during training.

Two networks were built separately: one using the surface dataset and the other using the section dataset.

#### Activation Maps

To explain the predictions of the deep CNN, coarse localization heat-maps were plotted to pinpoint key areas identified by the network. To this end, activation maps using the gradient-

weighted class activation mapping (or Grad-CAM) method were displayed, as discussed by Selvaraju *et al.* [20].

### Implementation Details

Our implementation was performed using TensorFlow 1.4 and Keras 2.2.4. The Keras image preprocessing tools available at <https://keras.io/api/preprocessing/image/> were applied for data augmentation.

### Statistical Analysis

#### Quantitative Assessment of Automatic Endoscopic Stone Recognition

For both the surface dataset and the section dataset, stones were randomly divided into complementary training (70%) and testing (30%) subsets (stratified split/no redundancy). A cross-validation step was repeated 10 times with randomly shuffled combinations for training and testing. The full process was also repeated with different random initialization seeds for the deep CNN algorithm. Average test metrics were reported for each step: accuracy, area under the ROC curve (AUROC), specificity, sensitivity, positive predictive value, negative predictive value, false predictive rate and false-negative rate. For additional information about these test metrics, see Kohavi *et al.* [21] and Cantor *et al.* [22]. Concerning mixed stones, test metrics were evaluated when at least one of the pure morphologies was predicted (note that mixed stones were composed of two pure morphologies in the scope of this study) and when both morphologies were predicted.

#### Qualitative Assessment of Activation Maps

A qualitative (visual) observation of the activation maps was carried out for surface and section images individually. The amount of correctly classified and misclassified images was calculated when the hot spots in the activation maps were located: (i) in the stone, (ii) outside the stone and (iii) over the tip of the endoscope.

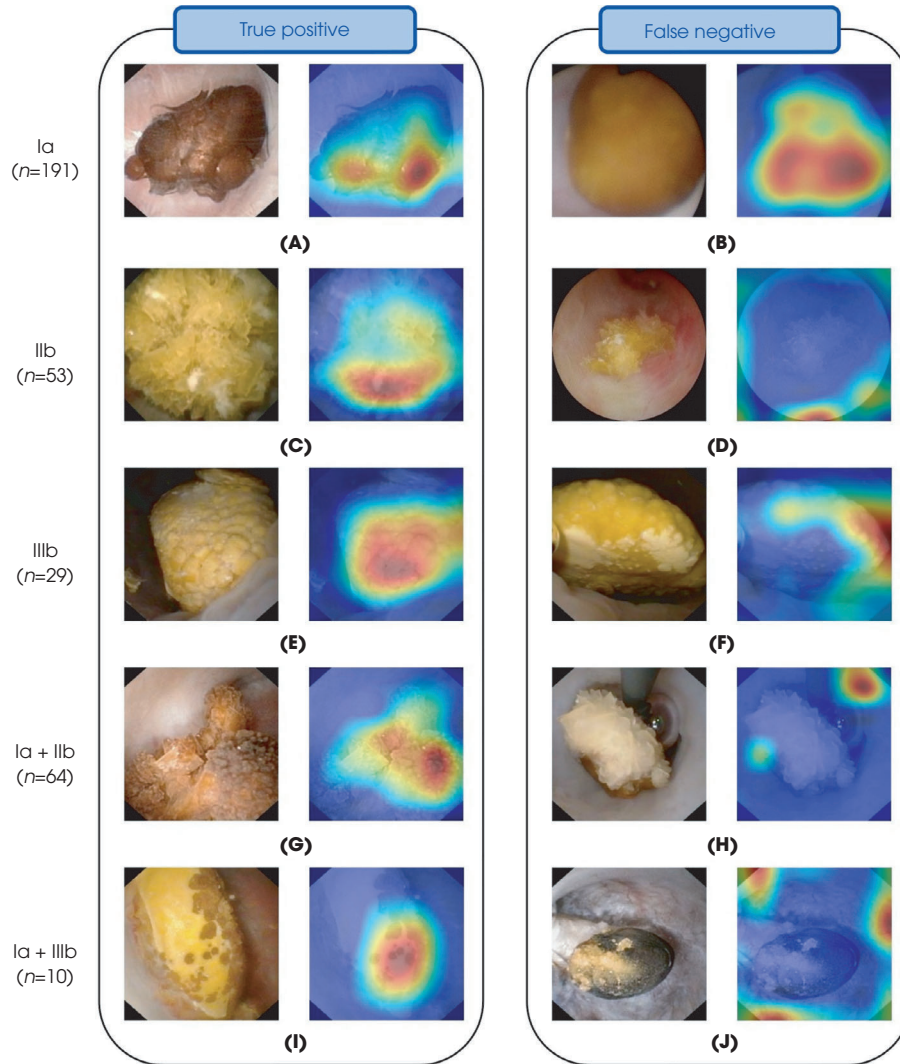
## Results

### Stone Characteristics

The study included 347 observations of stone surface (pure stones: Ia = 191/150 [number of images/number of unique stones], IIB = 53/48, IIIb = 29/23; mixed stones: Ia + IIB = 64/54, Ia + IIIb = 10/9) and 236 observations of stone section (pure stones: Ia = 127/96, IIB = 30/29, IIIb = 25/22; mixed stones: Ia + IIB = 31/26, Ia + IIIb = 23/15).

Figure 1 shows representative examples of *in situ* endoscopic images obtained for each pure stone morphology before laser

**Fig. 1** Representative automatic endoscopic stone recognition results obtained before laser fragmentation (surface image). Examples of both correctly (left panel) and misclassified images (right panel; type reported on far left is not recognised by network) are shown. In situ surface images (left image of each panel) are reported for each stone composition. Ia/calcium oxalate monohydrate, IIb/ calcium oxalate dihydrate and IIIb/uric acid pure morphologies are reported in first three rows. For each mixed stone (last two rows), a mixture of the corresponding pure morphologies is visible. Activation maps (right image of each panel) show areas where network concentrates attention.



fragmentation (surface image). Images acquired after laser fragmentation (section images) are shown in Fig. 2. The three pure stone morphologies (first three rows of Figs 1,2) had the following visual characteristics:

- *Ia/COM*: before laser fragmentation (1a): a smooth or mammillary, dark-brown surface; after laser fragmentation (2a): compact concentric layers with a radiating organization starting from a nucleus.
- *IIb/COD*: before laser fragmentation (1c): a yellowish or light-brown surface with smooth, long bi-pyramidal crystals (like small desert roses); after laser fragmentation (2c): a compact poorly organized pale brown-yellow crystalline section.
- *IIIb/UA*: before laser fragmentation (1e): a rough, porous surface with heterogeneous, beige to orange-red colour;

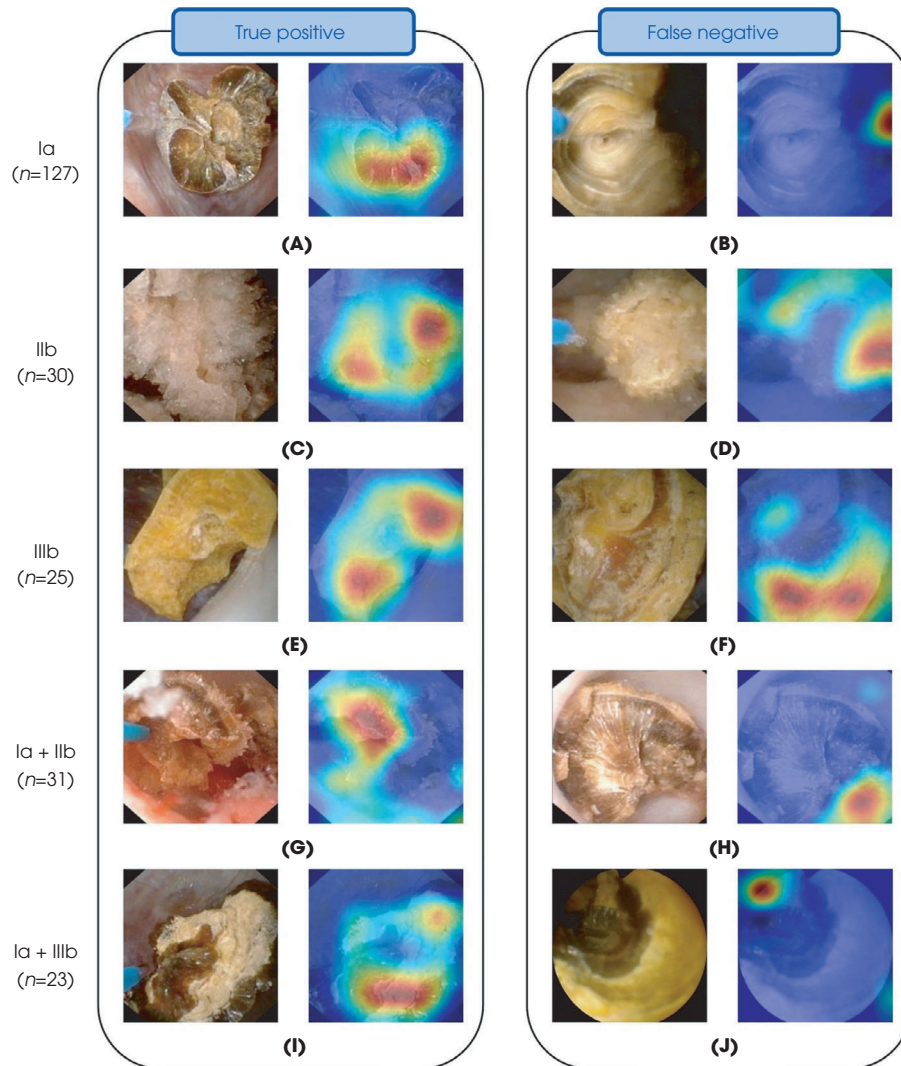
after laser fragmentation (2e): poorly organized, porous ochre to orange structure.

Representative examples of *in situ* endoscopic images of Ia + IIb and Ia + IIIb mixed stones are also shown (Figs 1,2, last two rows). For each, the combinations of the corresponding two of the three pure morphologies mentioned above are visible.

#### Diagnostic Performance of Automatic Endoscopic Stone Recognition

The testing subset of the surface dataset included 105 urinary stones (pure stones: Ia = 57, IIb = 16, IIIb = 9; mixed stones: Ia + IIb = 20, Ia + IIIb = 3). The testing subset of the section

**Fig. 2** Representative automatic endoscopic stone recognition results obtained after laser fragmentation (section images). Examples of both correctly (left panel) and misclassified images (right panel: type reported on far left is not recognised by network) are shown. In situ section images (left image of each panel) are reported for each stone composition. Ia/calcium oxalate monohydrate, IIb/calcium oxalate dihydrate and IIIb/uric acid pure morphologies are reported in first three rows. For each mixed stone (last two rows), a mixture of the corresponding pure morphologies is visible. Activation maps (right image of each panel) show areas where network concentrates attention.



dataset included 70 urinary stones (pure stones: Ia = 38, IIb = 9, IIIb = 7; mixed stones: Ia + IIb = 9, Ia + IIIb = 7).

Table 1 details the diagnostic performance of the deep CNN classifier for each tested pure type. The best sensitivity was obtained for the type IIIb using surface images (98% of IIIb stones correctly predicted). The most frequently encountered morphology was the type 'pure Ia'; it was correctly predicted in 91% and 94% of cases using surface and section images, respectively. On average, the accuracy was higher than 87% for both pure and mixed stones.

Table 2 details the diagnostic performance of the deep CNN classifier for each tested mixed type. Concerning Ia

+ IIb stones, Ia was predicted in 84% of cases using surface images, IIb in 70% of cases, and both types in 65% of cases. Concerning Ia + IIIb stones, Ia was predicted in 91% of cases using section images, IIIb in 69% of cases, and both types in 74% of cases. These findings are also displayed in the confusion matrices shown in Fig. 3: at least one of the two morphologies constituting mixed stones is preferably detected as a secondary choice, as indicated by off-diagonal values in Fig. 3. Overall, percentages of valid predictions using surface and section images were equal to 83% and 81%, respectively (see blue cells in confusion matrices in Fig. 3).

**Table 1** Diagnostic performance of implemented deep convolutional neural network classifier for pure stones (i.e., Ia/calcium oxalate monohydrate, IIb/calcium oxalate dihydrate and IIIb/uric acid morphologies).

Stone type	Accuracy, %	AUROC	Sensitivity, %	Specificity, %	PPV, %	NPV, %	FPR, %	FNR, %
<b>Surface</b>								
Ia	90 ± 3	0.90 ± 0.03	91 ± 5	90 ± 4	92 ± 3	90 ± 4	10 ± 4	9 ± 5
IIb	93 ± 2	0.86 ± 0.04	77 ± 7	95 ± 2	76 ± 9	96 ± 1	5 ± 2	23 ± 7
IIIb	99 ± 1	0.98 ± 0.02	98 ± 5	99 ± 1	90 ± 8	100 ± 0	1 ± 1	2 ± 5
<b>Section</b>								
Ia	94 ± 2	0.94 ± 0.02	94 ± 2	93 ± 5	94 ± 4	94 ± 3	7 ± 5	6 ± 2
IIb	94 ± 3	0.83 ± 0.09	69 ± 18	97 ± 2	77 ± 13	96 ± 3	3 ± 2	31 ± 18
IIIb	95 ± 2	0.78 ± 0.14	60 ± 30	97 ± 2	63 ± 27	97 ± 1	3 ± 2	40 ± 30

AUROC, area under the receiver operating characteristic curve; FNR, false-negative rate; FPR, false predictive rate; NPV, negative predictive value; PPV, positive predictive value. Results obtained using surface and section images are reported after cross-validation (averaged indicators shown with standard deviations). Accuracies, sensitivities, specificities, PPVs, and NPVs shown in percentages.

**Table 2** Diagnostic performance of implemented deep convolutional neural network classifier for mixed stones (i.e., Ia/calcium oxalate monohydrate (COM) + IIb/calcium oxalate dihydrate and Ia/COM + IIIb/uric acid morphologies).

Stone type	Predicted kidney type	Accuracy (%)	AUROC	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	FPR (%)	FNR (%)
<b>Surface</b>									
Ia + IIb	At least Ia	89 ± 2	0.88 ± 0.03	84 ± 4	91 ± 2	85 ± 4	91 ± 2	9 ± 2	16 ± 4
	At least IIb	90 ± 2	0.82 ± 0.03	70 ± 6	94 ± 2	70 ± 6	94 ± 1	6 ± 2	30 ± 6
	Both Ia and IIb	87 ± 3	0.78 ± 0.05	65 ± 10	92 ± 3	65 ± 8	92 ± 2	8 ± 3	35 ± 10
Ia + IIIb	At least Ia	94 ± 1	0.93 ± 0.02	89 ± 4	96 ± 1	91 ± 3	96 ± 2	4 ± 1	11 ± 4
	At least IIIb	98 ± 0	0.93 ± 0.04	86 ± 8	99 ± 0	87 ± 6	99 ± 0	1 ± 0	14 ± 8
	Both Ia and IIIb	98 ± 1	0.75 ± 0.14	50 ± 28	100 ± 1	71 ± 32	99 ± 1	0 ± 1	50 ± 28
<b>Section</b>									
Ia + IIb	At least Ia	91 ± 2	0.90 ± 0.02	86 ± 4	93 ± 2	86 ± 4	93 ± 2	7 ± 2	14 ± 4
	At least IIb	91 ± 2	0.78 ± 0.06	60 ± 13	95 ± 1	64 ± 8	94 ± 2	5 ± 1	40 ± 13
	Both Ia and IIb	88 ± 2	0.72 ± 0.08	51 ± 17	93 ± 2	51 ± 10	93 ± 2	7 ± 2	49 ± 17
Ia + IIIb	At least Ia	94 ± 2	0.93 ± 0.02	91 ± 3	95 ± 2	90 ± 4	96 ± 1	5 ± 2	9 ± 3
	At least IIIb	95 ± 2	0.83 ± 0.09	69 ± 18	97 ± 2	73 ± 13	97 ± 1	3 ± 2	31 ± 18
	Both Ia and IIIb	94 ± 2	0.85 ± 0.09	74 ± 18	97 ± 2	74 ± 16	97 ± 2	3 ± 2	26 ± 18

AUROC, area under the receiver operating characteristic curve; FNR, false-negative rate; FPR, false-predictive rate; NPV, negative predictive value; PPV, positive predictive value. Test metrics evaluated using surface and section images when at least one pure morphology is predicted (N.B. mixed stones were composed of two pure morphologies in this study) and when both morphologies were predicted.

## Qualitative Performance of Activation Maps

Figures 1 and 2 also show image areas where the classification network concentrated attention in the surface (Fig. 1) and section datasets (Fig. 2), respectively. Activation maps were overlaid on the digital endoscopic image in order to establish whether the classification model relied on relevant urological regions in the decision-making process. For example, a hot spot was usually observed Ia on a mammillary dark-brown area, which is a hallmark of Ia. Hot spots were found on characteristic stone features in 98% of the correctly classified images (using either surface or section images, see 'true-positive' columns in Figs 1,2). Hot spots outside the stone were found in 33% and 25% of misclassified surface and section images, respectively (Fig. 1d). Similarly, the tip of the endoscope was present in the image field of view in 5% and 2% of misclassified surface and section images, respectively (red arrow in Fig. 1h).

## Discussion

In the present study, we evaluated a deep-learning model to predict *in situ* the morphology of pure and mixed stones based on intra-operative endoscopic digital images acquired in a clinical setting. A learning curve is needed to acquire the ESR skills, which may limit its translation to practical use, especially when mixed stone morphologies are involved [12]. A computer-assisted approach can deliver reproducible results and minimizes operator dependency while assisting visual interpretation of stone morphologies.

As reported in the studies by Estrade et al. [812], ESR may be beneficial before fragmentation to preserve an aetiological approach in lithiasis. The motivation is twofold. Firstly, laser fragmentation (holmium:YAG and TFL), whether performed with 'popcorn' [1] or 'dusting' modes [2], irreversibly destroys the stone morphology. Postoperative FTIR examinations of the stone powder itself may not provide sufficient

**Fig. 3** Confusion matrices for implemented deep convolutional neural network classifier obtained using the surface **(A)** and section **(B)** datasets. Each column of the matrices represents an actual stone type, while each line represents a predicted type. Green diagonal cells show number (averaged by cross-validation) and percentage of correct predictions by trained network. Red off-diagonal cells correspond to wrongly predicted observations. Column on far right shows positive predictive value (green numbers) and false discovery rate (red numbers). Bottom row shows sensitivity (green numbers) and the false-negative rate (red numbers). Blue cell bottom right shows overall percentage of correct (green) and incorrect (red) predictions.

		Surface					
Predicted type	Ia	51.8 49.3%	3.5 3.3%	0.5 0.5%	0.8 0.8%	0 0.0%	92% 8%
	Ia+IIb	3.8 3.6%	13 12.4%	0.9 0.9%	2.2 2.1%	0.2 0.2%	65% 35%
	Ia+IIIb	0.3 0.3%	0.1 0.1%	1.5 1.4%	0.1 0.1%	0 0.0%	71% 29%
	IIb	1 1.0%	3 2.9%	0.1 0.1%	12.3 11.7%	0 0.0%	76% 24%
	IIIb	0.1 0.1%	0.4 0.4%	0 0.0%	0.6 0.6%	8.8 8.4%	90% 10%
		91% 9%	65% 35%	50% 50%	77% 23%	98% 2%	83% 17%
		Ia	Ia+IIb	Ia+IIIb	IIb	IIIb	
		Kidney type					

		Section					
Predicted type	Ia	37.8 53.3%	2.1 3.0%	0.5 0.7%	0.4 0.6%	0.1 0.1%	94% 6%
	Ia+IIb	1.4 2.0%	4.6 6.5%	0.5 0.7%	1.8 2.5%	0.7 1.0%	51% 49%
	Ia+IIIb	0.6 0.8%	0.8 1.1%	5.2 7.3%	0.1 0.1%	0.6 0.8%	74% 26%
	IIb	0 0.0%	1.2 1.7%	0.1 0.1%	6.2 8.7%	0.5 0.7%	77% 23%
	IIIb	0.3 0.4%	0.3 0.4%	0.7 1.0%	0.5 0.7%	3.9 5.5%	63% 37%
		94% 6%	51% 49%	74% 26%	69% 31%	60% 40%	81% 19%
		Ia	Ia+IIb	Ia+IIIb	IIb	IIIb	
		Kidney type					

information for the lithogenic stage [6–9]. Second, The infrared (IR) spectra can be modified when the stone fragmentation is achieved in dusting mode with high-frequency TFL [2–5]. This may in turn bias FTIR dust examinations: one can observe IR changes from COD towards COM, IR changes towards an amorphous phase in carapatite, IR changes towards a differing and amorphous crystalline phase in magnesium ammonium phosphate and IR changes from brushite towards carapatite [5].

We focused on pure and mixed stones involving Ia/COM, IIb/COD and IIIb/UA morphologies. This strategy was driven by the epidemiological distribution of the occurrence of urinary stones to obtain a sufficiently large population for an opposable statistical approach [23]. These morphological types cover almost 85% of the most common stones that urologists encounter in daily practice [6]. Our datasets of endoscopic images will be supplemented by more pure and mixed stone images in future studies (ongoing at our institution) in order to increase the number of morphologies to be predicted. In addition, the automatic ESR score will likely improve if the network is able to train on a larger set of data.

It should be noted that particles flying around in the saline may disturb the morphological stone examination. To properly recognize the colours and textures of the stone surface, it is necessary to wait a few tens of seconds until the saline solution has cleaned the urine in the kidney cavities. Then, once the stone is split into two parts, a few seconds' wait is again necessary before the saline solution has cleaned microparticles of the stone. In the present study, a ureteral

access sheath was used to improve saline flow. However, in practice, in the absence of a ureteral access sheath, only a few additional seconds are needed for complete cleaning using the saline serum. It should be emphasized that particles flying around in the saline were not present in the images used for training in the present study.

It should also be underlined that sufficient stability of the endoscopic video image is required for a short duration (5–10 s) to obtain good still frames. Any motion event is likely to hamper the image quality and, in turn, to bias the predictions of our trained network. In the present study, the trained urologist (V.E.) made several attempts to obtain sharp screenshot images (two attempts on average, maximum of four). Several strategies may be investigated in future works to improve the performance of the method when used on motion-corrupted endoscopic images. Enhanced high-quality images may be obtained from low-quality image series using dedicated motion-compensated super resolution techniques [24]. In addition, data augmentation techniques involving simulated blur and motion events may further improve the ability of the network to generalize for motion-corrupted endoscopic images [25].

Any unobserved events/image artifacts during the training step may, in turn, disturb the predictions of our trained network. In future studies, automated and reliable quality control on the input images must be developed in order to detect potential failure modes of the network. The urologist will then be advised to take laser-fragmented stones for a postoperative infrared – FTIR – examination in a dedicated laboratory.

Processing of both surface and section images provides valuable information about stone morphology for computer-aided diagnosis. In practice, recognition of the morphological hallmarks in surface images is easier than that in section images (or even at the nucleus), as shown in Estrade et al. [12] and Bergot et al. [26]. Consequently, our surface dataset was generally more densely populated than our section dataset (except for mixed Ia + IIb stones, for which section images better reveal the two morphological types). However, the diagnostic performance of the surface dataset was found to be comparable to that obtained in the section dataset (Fig. 3; blue cells). By contrast, the diagnostic performance of the section dataset was better than that obtained with the surface dataset for pure IIIb/UA (Table 1). Surface and section images may thus provide a source of cross-validation of the diagnosis according to both complementary and redundant information. We believe that paired surface and section images for each stone may be incorporated into the CNN in order to improve the accuracy of the predictions.

The annotated datasets must be accurate since any subjectivity in ESR or potential bias of the urologist may be transferred into the network model. A concordance study between endoscopic digital pictures and microscopy may provide a confirmed ESR image of stones corresponding to specific aetiologies or lithogenic mechanisms [12]. In addition to automatic ESR, activation maps could become an important tool to test whether, during the decision-making process, the classification model relied on relevant urological regions. Moreover, the present study indicated that a hot spot located outside the stone led to a misclassification.

Deep CNNs capable of processing a large number of specific images efficiently are paving the way for automatic ESR on videos, which would further improve the accuracy of classification scores. This will require the development of dedicated algorithms to remove 'on the fly' irrelevant areas of the image that are likely to bias the network, such as those around the endoscope tip and surrounding tissue, among others.

In conclusion, combined with endoscopic digital images confirmed according to the criteria published in Estrade et al. [12], AI is a good candidate for automatic ESR of the morphological features of pure and mixed urinary stones composed of two morphologies. This study is a preliminary step towards the automatic ESR of mixed stones of several morphologies. Activation maps may prove to be a great asset for urologists to understand intra-operatively the predictions made by the AI model. This is especially crucial in medical applications where model accuracy is paramount. Combined with didactic boards of confirmed endoscopic images, both computer-aided diagnosis and associated activation maps may be useful for urologists to recognize stones *in situ* using an endoscopic examination before destruction. The combination

of automatic intra-operative ESR and postoperative infrared (FTIR) examinations of laser-fragmented stones would improve the aetiological approach to lithiasis.

## Acknowledgements

Experiments presented in this paper were carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d'Aquitaine (see <https://www.plafrim.fr/>). The authors gratefully acknowledge the support of NVIDIA Corporation with their donation of a TITAN X GPU used in this research.

## Disclosures of Interest

None declared.

## References

- Emiliani E, Talso M, Cho S-Y et al. Optimal settings for the noncontact holmium: YAG stone fragmentation popcorn technique. *J Urol* 2017; 198: 702–6
- Doizi S, Keller EX, De Coninck V et al. Dusting technique for lithotripsy: what does it mean? *Nat Rev Urol* 2018; 15: 653–4
- Andreeva V, Vinarov A, Yaroslavsky I et al. Preclinical comparison of superpulse thulium fiber laser and a holmium: YAG laser for lithotripsy. *World J Urol* 2020; 38: 497–503
- Traxer O, Keller EX. Thulium fiber laser: the new player for kidney stone treatment? A comparison with Holmium:YAG laser. *World J Urol* 2020; 38: 1883–94
- Keller EX, De Coninck V, Doizi S et al. Thulium fiber laser: ready to dust all urinary stone composition types? *World J Urol* 2021; 39: 1693–8
- Daudon M, Dessombz A, Frochot V et al. Comprehensive morpho-constitutional analysis of urinary stone improves etiological diagnosis and therapeutic strategy of nephrolithiasis. *C R Chim* 2016; 19: 1470–91
- Cloutier J, Villa L, Traxer O et al. Kidney stone analysis: "give me your stone, I will tell you who you are!". *World J Urol* 2015; 33: 157–69
- Estrade V, Daudon M, Méria P et al. Why should urologists recognize urinary stones and how? The basis of endoscopic recognition. *Prog Urol – FMC* 2017; 27: F26–35
- Daudon M, Jungers P, Bazin D et al. Recurrence rates of urinary calculi according to stone composition and morphology. *Urolithiasis* 2018; 46: 459–70
- Soygür T, Akbay A, Küpeli S. Effect of potassium citrate therapy on stone recurrence and residual fragments after shockwave lithotripsy in lower caliceal calcium oxalate urolithiasis: a randomized controlled trial. *J Endourol* 2002; 16: 149–52
- Tsaturyan A, Bokova E, Bosshard P et al. Oral chemolysis is an effective, non-invasive therapy for urinary stones suspected of uric acid content. *Urolithiasis* 2020; 48: 501–7
- Estrade V, Denis de Senneville B, Meria P et al. Toward improved endoscopic examination of urinary stones: a concordance study between endoscopic digital pictures vs microscopy. *Br J Urol Int* 2020. <https://doi.org/10.1111/bju.15312>
- Corrales M, Doizi S, Barghouthy Y et al. Classification of stones according to Michel Daudon: a narrative review. *Eur Urol Focus* 2021; 7: 13–21
- Serrat J, Lumbreras F, Blanco F et al. MyStone: a system for automatic kidney stone classification. *Expert Syst Appl* 2017; 89: 45–51



- 15 Black KM, Law H, Aldoukhi A *et al.* Deep learning computer vision algorithm for detecting kidney stone composition. *Br J Urol Int* 2020; 125: 920–4
- 16 Martinez A, Trinh DH, El Beze J *et al.* Towards an automated classification method for ureteroscopic kidney stone images using ensemble learning. *Annu Int Conf IEEE Eng Med Biol Soc* 2020: 1936–9
- 17 Lopez F, Varela A, Hinojosa O *et al.* Assessing deep learning methods for the identification of kidney stones in endoscopic images. arXiv:2103.01146, 2021. Available at: <https://arxiv.org/abs/2103.01146>
- 18 He K, Zhang X, Ren S *et al.* Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, 2016: 770–8. <https://doi.org/10.1109/CVPR.2016.90>
- 19 Kingma DP, Adam BJ. A method for stochastic optimization. arXiv:1412.6980 [cs], 2017. Available at: <http://arxiv.org/abs/1412.6980>
- 20 Selvaraju RR, Cogswell M, Das A *et al.* Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *IEEE International Conference on Computer Vision (ICCV)*, 2017; 618–26. <https://doi.org/10.1109/ICCV.2017.74>
- 21 Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *IJCAI'95: Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Vol. 2, 1995: 1137–43
- 22 Cantor SB, Kattan MW. Determining the area under the ROC curve for a binary diagnostic test. *Med Decis Making* 2000; 20: 468–70
- 23 Daudon M, Traxer O, Lechevallier E *et al.* Epidemiology of urolithiasis. *Prog Urol* 2008; 18: 802–14
- 24 Almalioglu Y, Bengisu Ozyoruk K, Gokce A *et al.* EndoL2H: deep super-resolution for capsule endoscopy. *IEEE Trans Med Imaging* 2020; 39: 4297–309
- 25 Potmesil M, Chakravarty I. Modeling motion blur in computer-generated images. *SIGGRAPH Comput Graph* 1983; 17: 389–99
- 26 Bergot C, Robert G, Bernhard JC *et al.* The basis of endoscopic stones recognition, a prospective monocentric study. *Prog Urol* 2019; 29: 312–7

Correspondence: Vincent Estrade, Department of Urology, CHU Pellegrin, Place Amélie Raba Léon 33000, Bordeaux, France.

e-mail: [vincent.estrade@gmail.com](mailto:vincent.estrade@gmail.com)

Abbreviations: AI, artificial intelligence; AUROC, area-under-the-ROC curve; CCD, charge-coupled device; CNN, convolutional neural network; COD, calcium oxalate dihydrate; COM, calcium oxalate monohydrate; ESR, endoscopic stone recognition; FTIR, Fourier transform infrared spectroscopy; IR, infrared; LASER, light amplification by stimulated emission of radiation; TFL, thulium fibre laser; UA, uric acid.