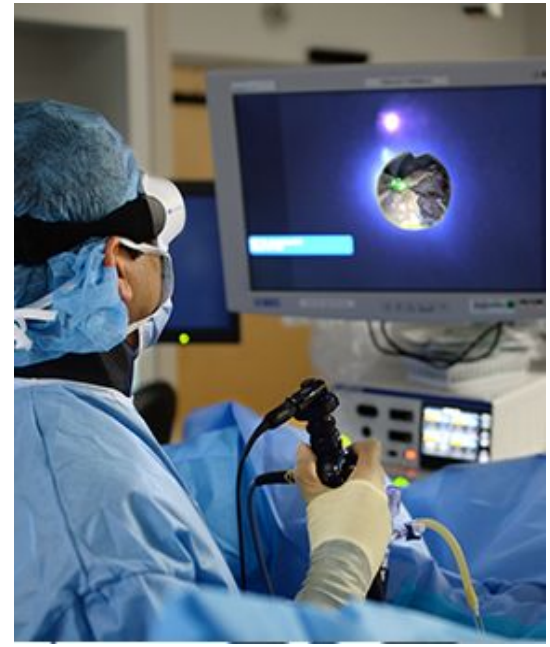
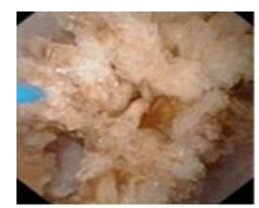
Assessing deep learning methods for the identification of kidney stones in endoscopic images





Francisco Lopez*, Andres Varela, Oscar Hinojosa, Mauricio Mendez, Dinh-Hoan Trinh, Jonathan ElBeze, Jacques Hubert, Vincent Estrade, Miguel Gonzalez, Gilberto Ochoa-Ruiz, Christian Daul

Tecnológico de Monterrey · CNRS · CRAN · CHRU Nancy · CHU Hôpitaux de Bordeaux · IADI · Université de Lorraine

Outline of the presentation

Introduction

→ Medical Context
→ Opportunities for AI systems

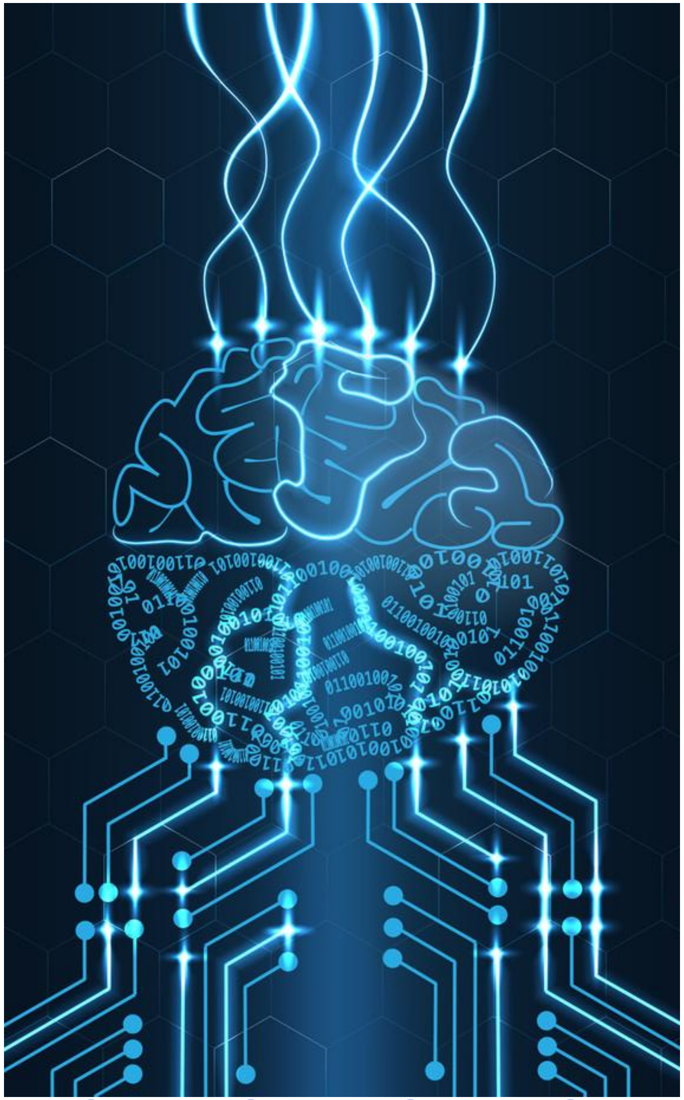
State of the Art

→ Existing Methods
→ Limitations

Proposed approach and contributions

→ Dataset of In Vivo Kidney Stones
→ Feature Extraction and Analysis
→ Classification Results

Future Work



Issues relating to kidney stones (i)

Urolithiasis refers to the formation of kidney stones (KS) in the urinary tract: (e.g., kidneys, ureters and bladder).

Common disease around the world and its incidence is growing every year.

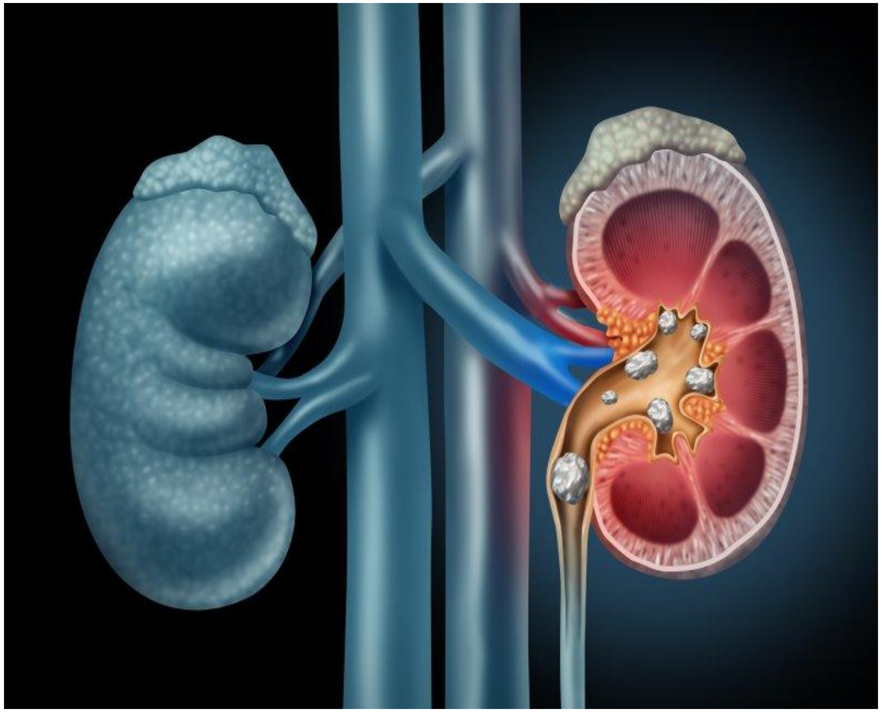
Common formation causes:

- Diet:
- Heredity
- Urinary tract infections
- Consumption of certain drugs

The accuracy of the diagnosis is crucial for the prescription of

* Appropriate treatment
* Diminish the risk of relapses.

Automatic endoscopic recognition of KS could improve this task!



Morpho-constitutional analysis (i)

INTERVIEW

Michel Daudon: "Urinary stones reflect the evolution of society"

Forty years of work in urolithiasis have given this French specialist international recognition

With the help of infrared and a 3D microscope, you can read all kinds of diseases on the stones

Michel Daudon, at the CRISTAL Laboratory of the Tenon Hospital in Paris. / CARME ESCALES

8 It is read in minutes

Un artículo de Carme escales

Paris
November 4, 2019
17:35

Since 1986, **Michel Daudon** directs the Center for Research and Scientific and Technical Information Applied to Lithiasis (CRISTAL Laboratory), a benchmark in Europe, located at the Tenon hospital in Paris. There this scientist, graduated in pharmacy and graduated in natural components of chemical structures, from his practice in analysis of calculations and the diploma that he created on this at the Sorbonne defends the need to analyze the stones made in the urinary tract to cure and prevent, with equal interest.

In 2019, ICONEKT encouraged medical community to use the Morpho-Constitutional Classification (MMC) approach proposed by **Michael Daudon**

Despite, growing evidence of

1. The robust diagnostic value of MMC,
2. The benefits in drawing up individualized treatment plans

This simple tool unfortunately still remains underexplored

James C. Williams "Trust my morphology", the key message from a kidney stone, Urolothiasis, 2021

https://www.elperiodico.com/es/sanidad/20191104/michel-daudon-los-calculos-urinarios-reflejan-la-evolucion-de-la-sociedad-7708327

Corrales, Mariela et al. 2021, Classification of Stones According to Michel Daudon: A Narrative Review, European Urology Focus

Morpho-constitutional analysis (ii)

There are various manuals and guidelines that help urologists to recognize urinary calculus, but the classification **accuracy remains very low!**

Microscopy — Surface | Section
Endoscopy — Surface | Section

Whewellite

Weddelitte

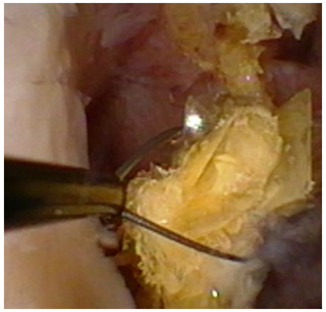
Acid Uric

V. Estrade et al, Pourquoi l'urologue doit savoir reconnaître un calcul et comment faire ? Les bases de la reconnaissance endoscopique, Progrès en Urologie - FMC, Volume 27, Issue 2, 2017,

How do experts identify KS composition? (i)

Morpho-constitutional analysis makes use of

- A microscope and an infrared spectrophotometer and/or
- A (visual) intraoperative morphological classification.



These methods are **time consuming, tedious and expensive**

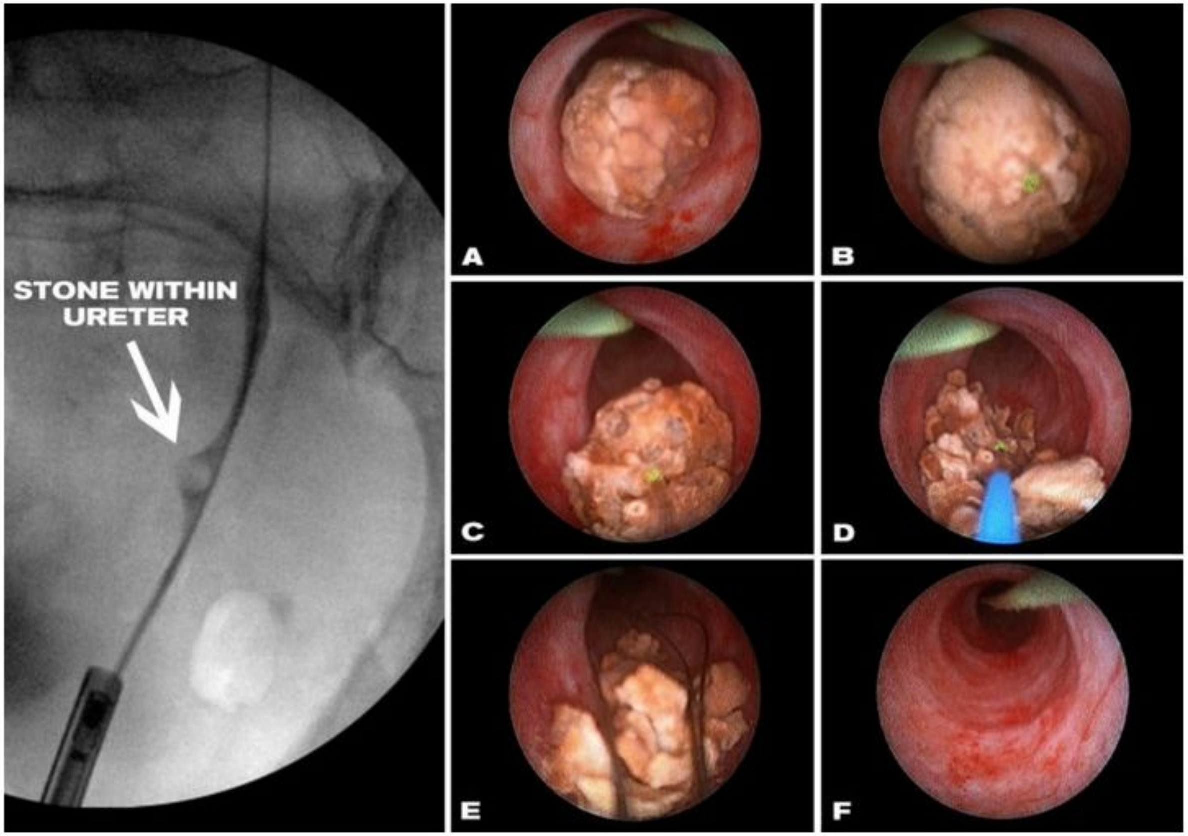
They require a **great deal of experience**.

Great opportunity for machine learning automatic identification!!!!

Jahrreiss V, Veser J, Seitz C, Özsoy M. Artificial intelligence: the future of urinary stone management? Curr Opin Urol. 2020 Mar;30(2):196-199. doi: 10.1097/MOU.0000000000000707. PMID: 31895075.

Ureteroscopy (i)

Kidney stone Removal Process



A: Kidney stone visualized with an endoscope. **B:** Preparing to laser stone. **C & D:** Laser pulverizing stone **E:** Removal of stone fragments with basket. **F:** Kidney stone-free ureter.

Weiss, B. Evaluation of dusting versus basketing — can new technologies improve stone-free rates?. *Nat Rev Urol* 13, 726–733 (2016)

Ureteroscopy (ii)

Advantages of dusting

- Dusting is a less invasive
- The interventions are less traumatic, easier and faster

Disadvantages of dusting

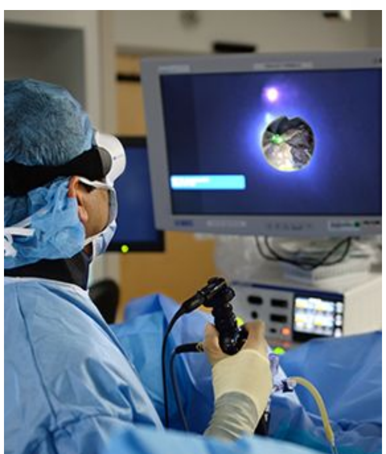
- Lower stone-free rates: some stones might remain
- Morpho-constitutional analysis is impossible

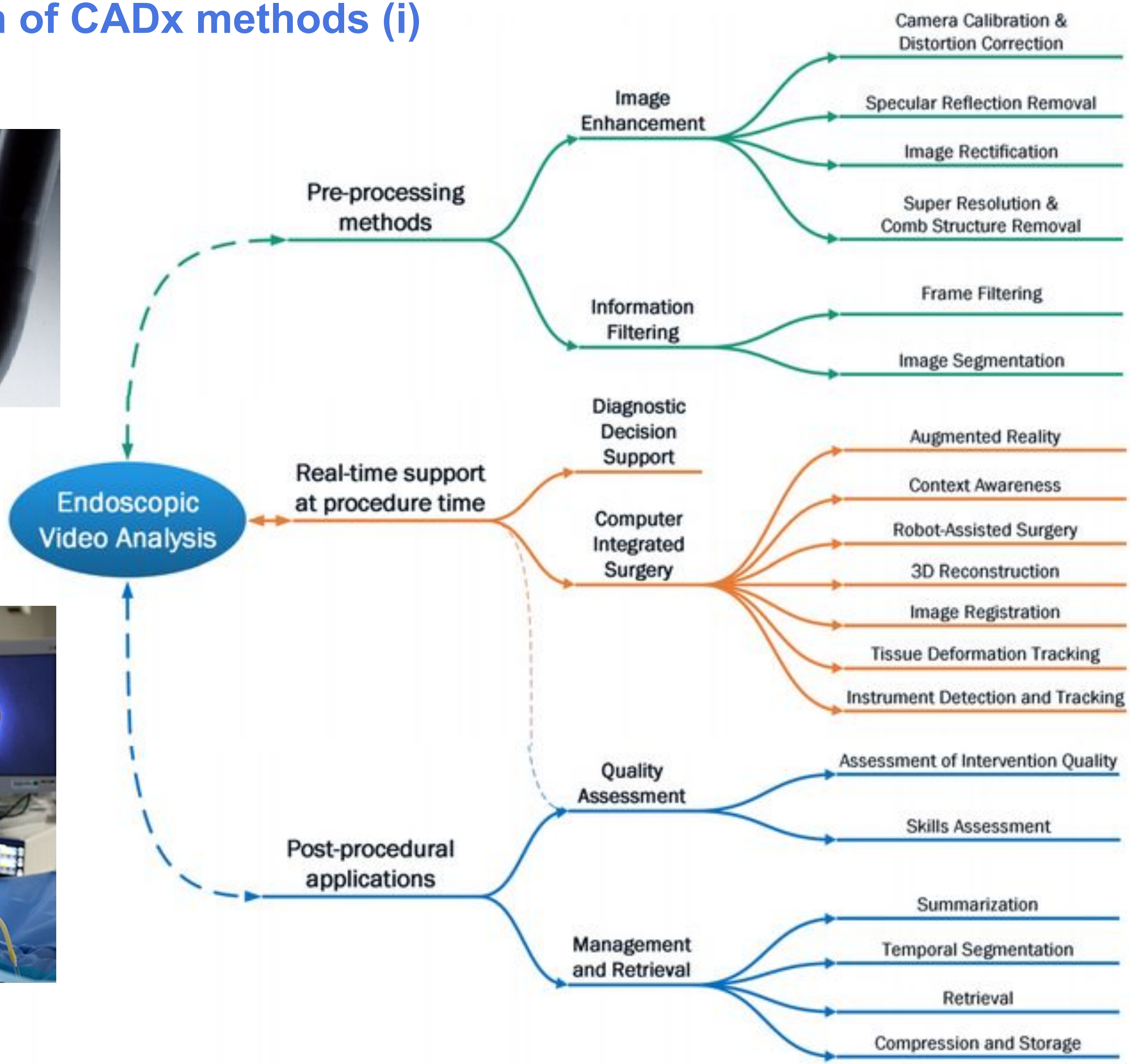
Interest of an image based classification

- **In vivo categorization of kidney stones**
- For diagnosis purposes, but also for adjusting laser settings

Keller EX, de Coninck V, Audouin M, et al. Fragments and dust after Holmium laser lithotripsy with or without "Moses technology": How are they different?. *J Biophotonics*. 2019;12(4):e201800227. doi:10.1002/jbio.201800227

Classification of CADx methods (i)





Motivation and contributions

1. Creation of a dataset of in vivo images
2. Optimized feature extraction methods
3. Machine Learning (shallow and deep learning based) methods for classification
4. Thorough comparison of shallow and deep machine learning methods

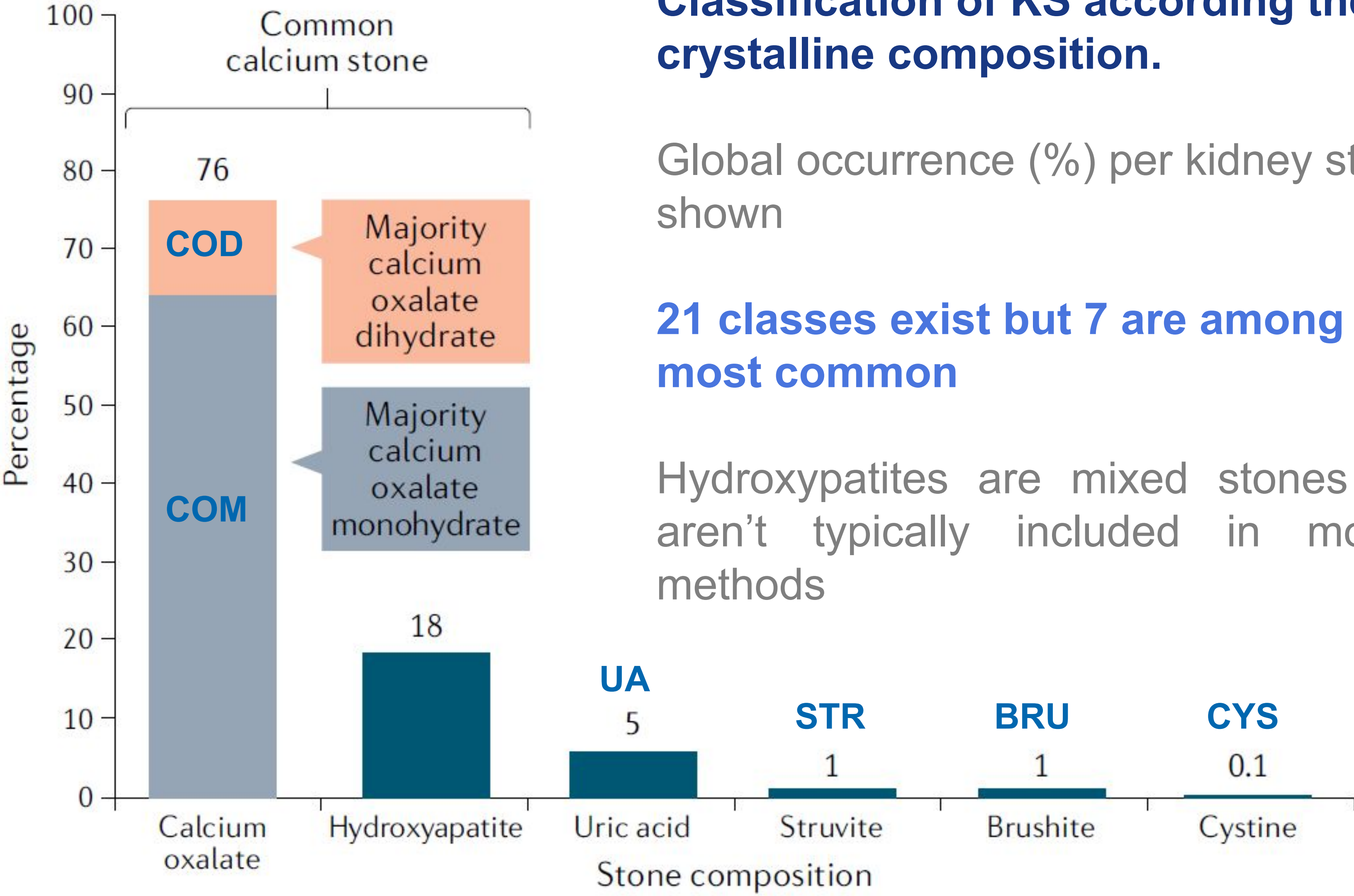
This work aims to build a software to aid urologists with the classification of kidney stones during ureteroscopies using standard hardware (i.e., ureteroscopes).

CHU Hôpitaux de Bordeaux

CHRU NANCY

Inserm

A. Martínez *et al.*, "Towards an automated classification method for ureteroscopic kidney stone images using ensemble learning," *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, QC, Canada, 2020



Classification of KS according their crystalline composition.

Global occurrence (%) per kidney stone is shown

21 classes exist but 7 are among the most common

Hydroxypatites are mixed stones which aren't typically included in most AI methods

Thongprayoon, C., Krambeck, A.E. & Rule, A.D. Determining the true burden of kidney stone disease. *Nat Rev Nephrol* (2020). https://doi.org/10.1038/s41581-020-0320-7

State of the Art

Deep learning–based image analysis for an automated kidney stone composition identification has attracted attention in recent years!

Reference/Feature	Kidney Stone Composition						Image Type		Acquisition
	UA	COM	COD	STR	CYS	BRU	Surface	Section	
Serrat et al 2017 [27]	✓	✓	✓	✓	✓		✓	✓	Ex vivo
Torrell et al 2018 [28]	✓	✓	✓	✓	✓	✓	✓		Ex vivo
Black et al 2020 [29]	✓	✓		✓	✓	✓	✓	✓	Ex vivo
Martinez et al 2020 [30]	✓	✓	✓				✓	✓	In vivo
This contribution	✓	✓	✓			✓	✓	✓	In vivo

Reference	Precision Per Class						Avg Precision	ML Method
	UA	COM	COD	STR	CYS	BRU		
Serrat et al 2017 [27]	65.0	55.0	69.0	50.0	N/A	N/A	63.0	Random Forest
Torrell et al 2018 [28]	76.0	66.0	80.0	61.0	N/A	N/A	74.0	Siamese CNN
Black et al 2020 [29]	94.1	95.0	N/A	71.4	75.0	75.0	85.0	CNN – ResNet101
Martinez et al 2020 [30]	88.0	84.0	90.0	N/A	N/A	90.0	91.0	Random Forest

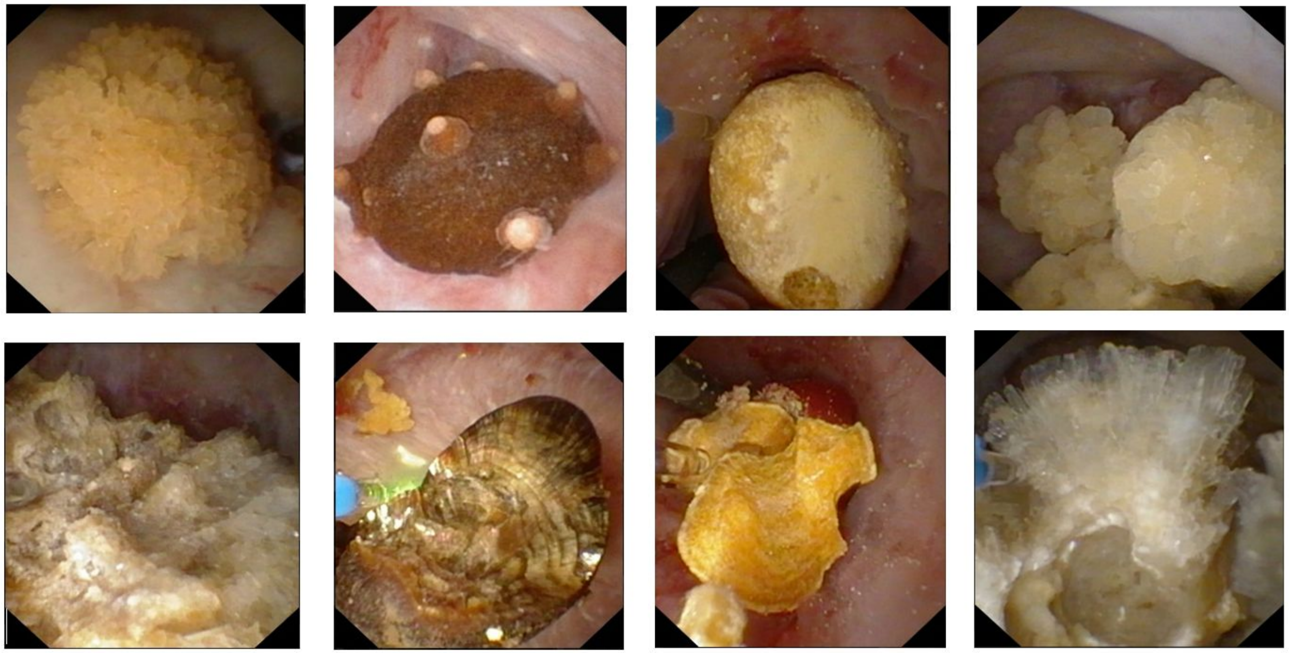
Previous contributions used images of extracted kidney stones captured in **highly controlled conditions.**

A. Torrell, Metric learning for kidney stone classification, BSc. thesis, Escola D'Enginyeria, Universitat Autonoma de Barcelona, 2018.

K. M. Black, H. Law, A. Aldoukhi, J. Deng and K. R. Ghani, Deep learning computer vision algorithm for detecting kidney stone composition, BJU International, 2020.

Dataset of In Vivo Kidney Stone Images

In vivo images collected by Dr. Vincent Estrade



Whewellite **Weddellite** **Uric Acid** **Brushite**

CHU Hôpitaux de Bordeaux

Acquisition Devices

URF-V/V2 Olympus
BOA Richard Wolf

Dataset Composition

94 surface images
87 cross-section images

Ethical Considerations

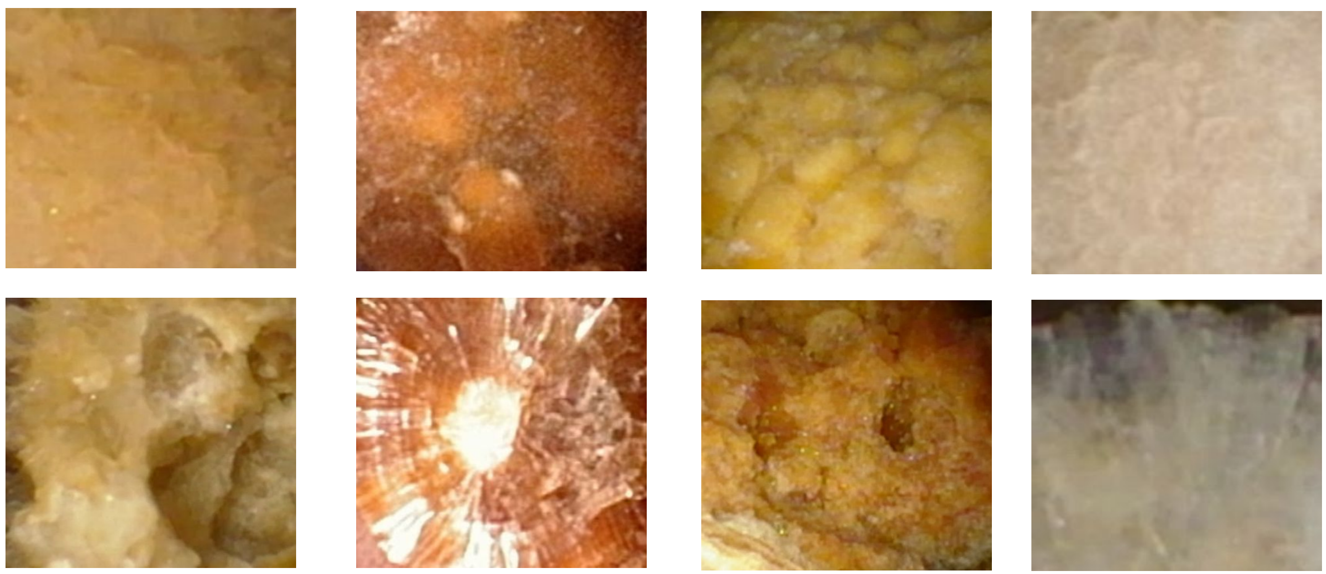
Patients agreed w/ their stones being used for this study!

Dataset processing (i)

Patches of 256x256 pixels were extracted to increase the number of samples in each class and **avoid overfitting of the models**

This size was chosen to obtain a **significant quantity of patches**, while keeping **relevant texture and color information** in each patch.

Very small overlap (< 20pixels) of patches of the same stone, no visible tissue/instruments



Surface

Section

Whewellite | Weddelitte | Uric Acid | Brushite

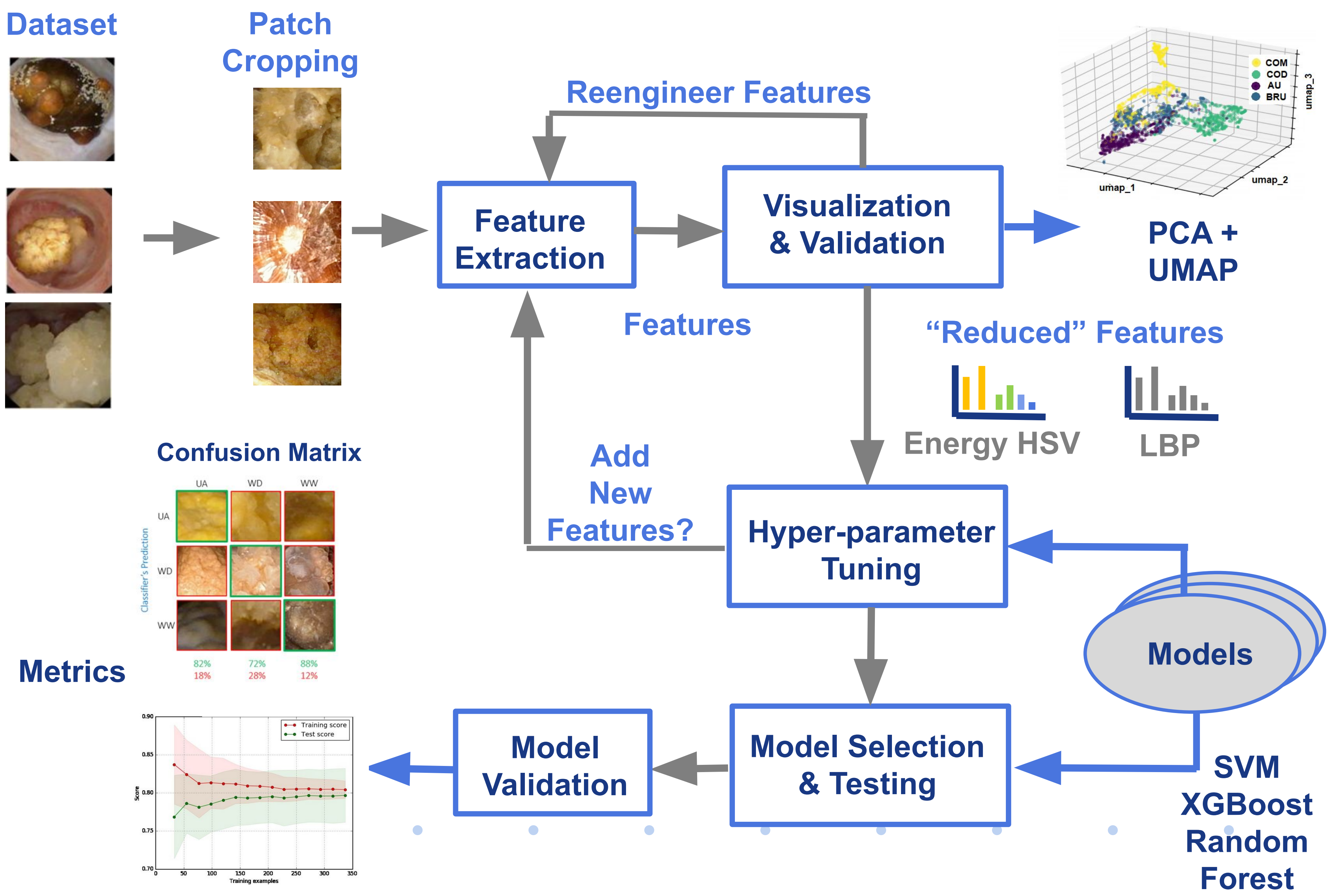
Barz, B.; Denzler, J. Do We Train on Test Data? Purging CIFAR of Near-Duplicates. *J. Imaging* 2020, *6*, 41.

Dataset processing (ii)

The proposed patch extraction procedure yielded a **new dataset**

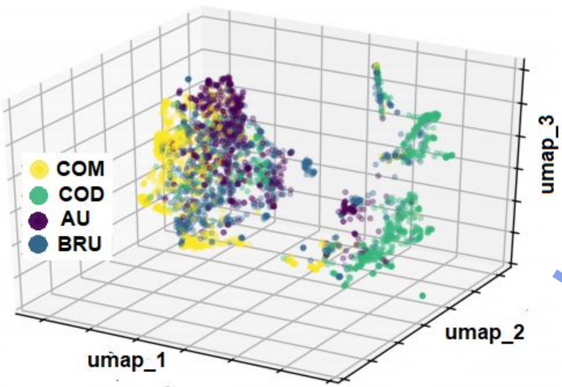
Stone Type		Acquired Images		Number of patches
View	Class	Number	Presence (%)	
Surface	Whewellite (COM)	30	31.9	870
	Weddelite (COD)	32	34.1	920
	Uric Acid	18	19.1	470
	Brushite	14	14.9	420
	Total	**94**	**100.0**	**2680**
Section	Whewellite (COM)	27	31.0	820
	Weddelite (COD)	28	32.2	780
	Uric Acid	18	20.7	460
	Brushite	14	16.1	410
	Total	**87**	**100.0**	**2470**

“Shallow” Feature Extraction and Analysis

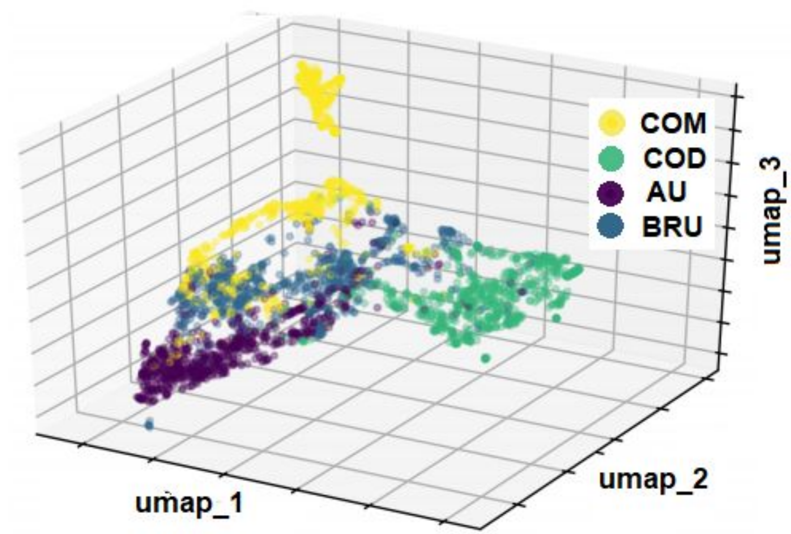


Visualization of embeddings using UMAP

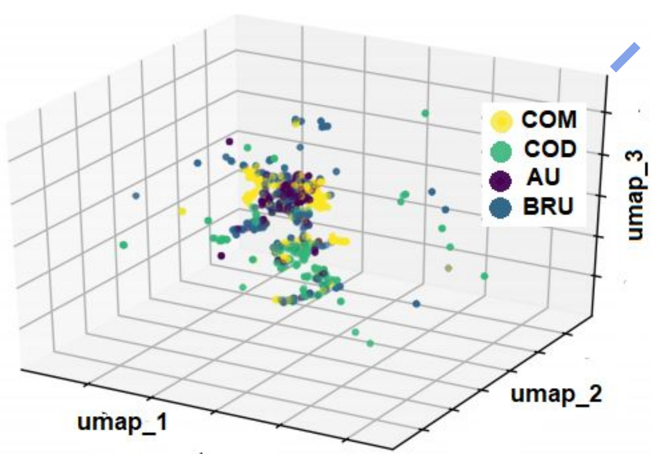
Handcrafted HSI + LBP feature vectors
for surface patches



Effect of mixing
Handcrafted HSI + LBP feature vectors
for surface and section patches



Handcrafted HSI + LBP feature vectors
for section patches



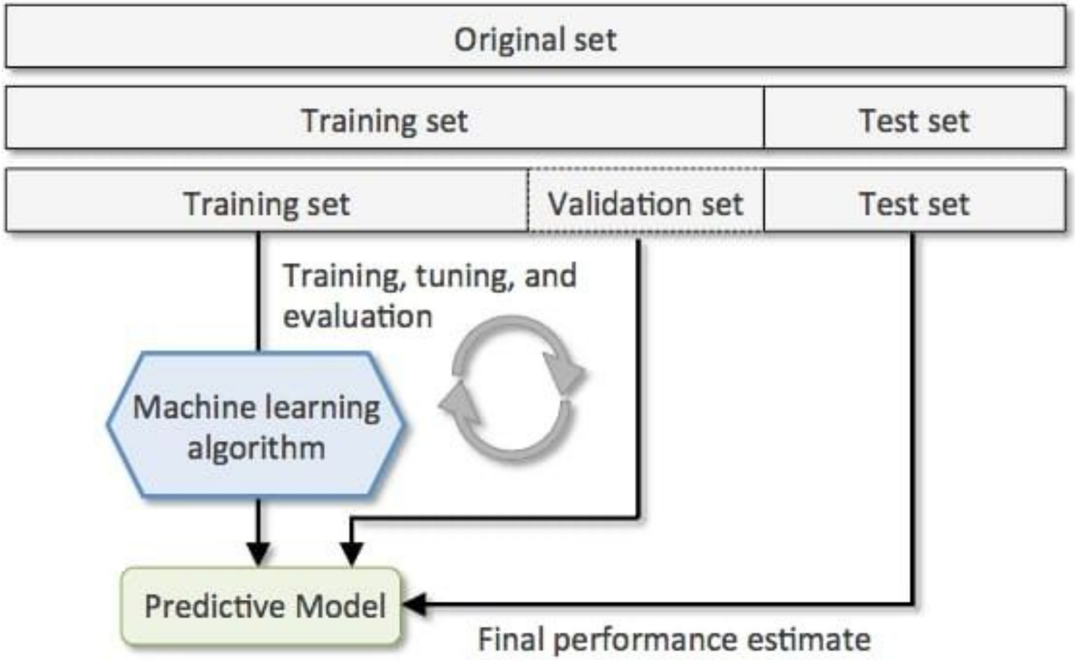
Can we improve the results
with a "better feature representations"?
Deep Learning…

“Shallow” Machine Learning Models Assessment

Tests with various shallow machine learning methods:

- **Classifiers: SVM, AdaBoost, Bagging, MLP, XGBoost, Random Forest**
- Use of INRIA’s library **SciKit Learn** for testing and visualization purposes

Hyper-parameter tuning using a combination of grid and random search based on an one-leave cross validation



[Scikit-learn: Machine Learning in Python](), Pedregosa *et al.* , JMLR 12, págs. 2825-2830, 2011.

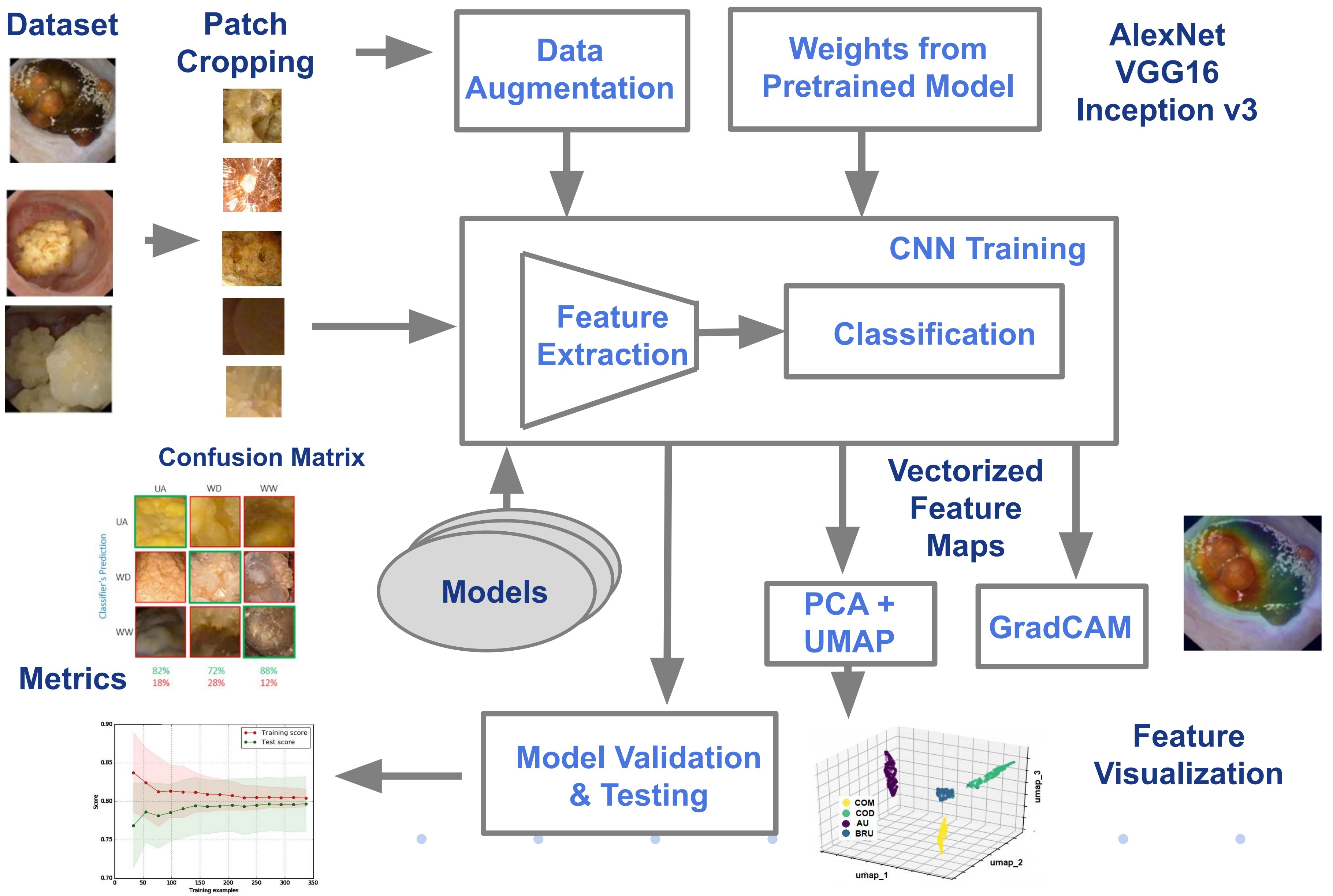
Classification Results – Shallow ML models

Method/ Metric	Surface			Section			Mixed		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
SVM	0.83	0.86	0.84	0.76	0.86	0.80	0.79	0.77	0.79
AdaBoost	0.83	0.86	0.84	0.81	0.85	0.85	0.81	0.81	0.81
Bagging	0.76	0.76	0.76	0.77	0.77	0.75	0.75	0.76	0.75
MLP	0.86	0.91	0.88	0.80	0.64	0.63	0.84	0.86	0.85
Random Forest	0.87	0.82	0.84	0.76	0.85	0.80	0.91	0.91	0.91
XGBoost	0.93	0.93	0.94	0.89	0.89	0.88	0.96	0.96	0.96

Combined Features – XGBoost – EMBC’20

Method/ Metric	Precision	Recall	F1
Surface	0.93	0.93	0.94
Section	0.89	0.89	0.88
Combined	**0.96**	**0.96**	**0.97**

"Deep" Feature Extraction and Analysis



Results w/ Deep Learning (Transfer Learning)

Results per class

Method/ Class	Whewellite (COM)		Weddelitte (COD)		Uric Acid (UA)		Brushite (BRU)	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Random Forest	0.84	0.86	0.90	0.95	0.88	0.67	0.90	0.92
XGBoost	0.92	0.96	0.91	0.91	0.97	0.96	0.96	0.94
AlexNet	0.93	0.98	0.95	0.85	0.88	0.92	0.93	0.92
VGG16	0.97	0.97	0.92	0.93	0.93	0.83	0.94	0.92
Inception v3	**0.98**	**0.97**	**0.93**	**0.96**	**0.95**	**0.90**	**0.96**	**0.95**

Results per patch type

Method/ Class	Surface			Section			Combined		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Random Forest	0.87	0.82	0.84	0.82	0.82	0.81	**0.91**	**0.91**	**0.91**
XGBoost	0.93	0.93	0.94	0.89	0.89	0.88	**0.96**	**0.96**	**0.96**
AlexNet	0.93	0.95	0.94	0.83	0.82	0.83	0.92	0.93	0.92
VGG16	0.95	0.94	0.95	0.91	0.92	0.92	0.95	0.96	0.95
Inception V3	**0.98**	**0.97**	**0.97**	**0.94**	**0.96**	**0.95**	**0.97**	**0.98**	**0.97**

Results w/ Deep (Transfer) Learning

XGBoost

Method	Precision	Recall	F1
Surface	0.93	0.93	0.94
Section	0.89	0.89	0.88
Combined	**0.96**	**0.96**	**0.97**

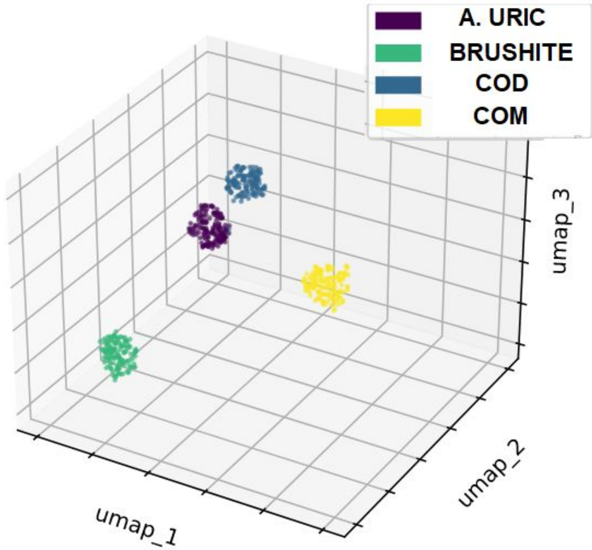
Inception v3

Method	Precision	Recall	F1
Surface	0.98	0.97	0.97
Section	0.94	0.96	0.95
Combined	**0.98**	**0.98**	**0.97**

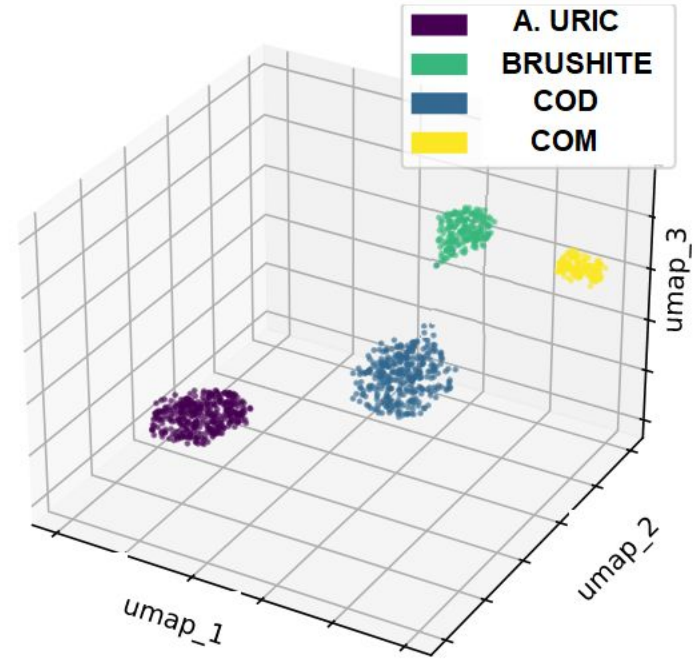
F. Lopez-Tiro et al 2021, *Assessing* deep learning methods for the identification of kidney stones in endoscopic images
IEEE Engineering in Medicine and Biology Conference (EMBC 2020)

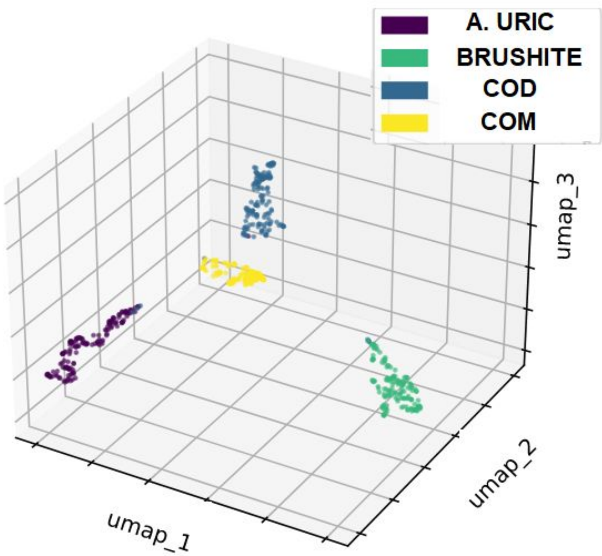
Results w/ Deep Learning – Results for Inception v3

Inceptionv3 Deep Features
for surface patches



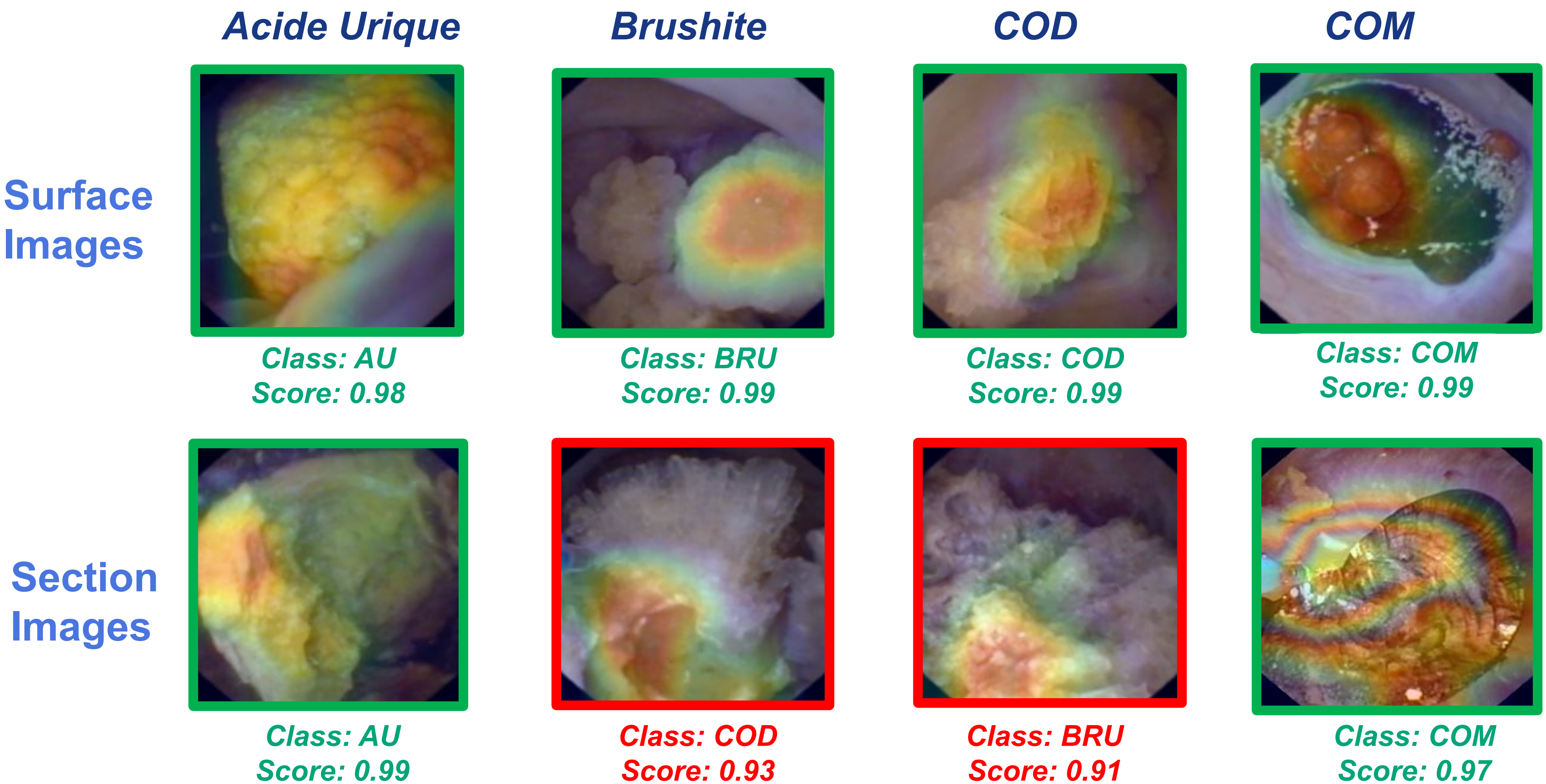
Inception v3 Deep Features
for mixed patches





Inception v3 Deep Features
for section patches

Understanding classification outcomes using GradCAM (using AlexNet)



Patch-wise these two samples are very similar, in terms of color and texture, leading to incorrect classifications

Ramprasaath R. Selvaraju et al 2020, Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization
International Journal of Computer Vision volume 128, pages336–359(2020)

Conclusion

One of the first attempts of kidney stone classification using in vivo images.

Although the images used for this work were captured in an uncontrolled environment, the classification accuracy was higher than for previous works dealing with ex-vivo images

However, the dataset need to be improved:

(1)Increasing the number of images
(2)Obtaining images from the rest of the existing classes
(3)Capturing images with diverse ureteroscopes

Once a bigger dataset is available, a more complex neural network could be implemented to potentially enhance the classification accuracy.

Problems such as image deblurring, few shot learning and real-time detection and/or instance segmentation will be explored

Thanks!!!

End


Francisco Lopez
Tecnológico de Monterrey
francisco.lopez@ieee.org
@friscolt