ORIGINAL PAPER

Determination of urinary stone composition through computed tomography using artificial intelligence

Hector Gallegos[1,2*], Alejandro Rojas[2*], Enzo Castiglioni[3], Carolina Verdugo[3], Benjamin Barraza[3], Benjamin Vargas[3], Stefano Brusoni[1], Gastón Astroza[1]

[1]Department of Urology, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile
[2]Department of Urology, Hospital Sotero del Río, Santiago, Chile
[3]School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile
*Hector Gallegos and Alejandro Rojas have equal contribution.

Citation: Gallegos H, Rojas A, Castiglioni E, et al. Determination of urinary stone composition through computed tomography using artificial intelligence. Cent European J Urol. 2026; doi: 10.5173/ceju.2025.0186

Article history
Submitted: Aug. 16, 2025
Accepted: May 8, 2026
Published online: June 18, 2026

Corresponding author
Gastón M. Astroza
Department of Urology,
School of Medicine,
Pontificia Universidad
Católica de Chile,
Diagonal Paraguay 362
Santiago, Chile
gaeulufi@gmail.com

Introduction Urinary stone composition analysis is essential for guiding treatment and preventive strategies, but conventional methods such as Fourier-transform infrared spectroscopy (FTIR) are costly and limited in availability. Artificial intelligence (AI), particularly convolutional neural networks (CNNs), offers potential as a non-invasive alternative using computed tomography (CT) imaging. The aim of the study is to evaluate the performance of a CNN model in predicting urinary stone composition from non-contrast CT (NCCT) images.

Material and methods A retrospective analysis was conducted using an institutional database of patients with urinary stones surgically removed between 2020 and 2024. Stone composition was confirmed via FTIR. A total of 1,487 NCCT scan images were used for CNN development and internal validation, with 105 scan images reserved for external validation. Transfer learning with VGG16 architecture was applied, incorporating data augmentation and fine-tuning strategies. Performance metrics included accuracy, area under the ROC curve (AUC), and Cohen's κ coefficient.

Results The dataset comprised calcium oxalate monohydrate (74.9%), calcium oxalate dihydrate (17.6%), uric acid (2.7%), struvite (2.3%), and other compositions (2.5%). The CNN achieved an internal accuracy of 89.8% (AUC 0.92). In independent temporal validation the cohort accuracy was 85.4%, with an AUC of 0.81 and a Cohen's κ of 0.864. Misclassifications occurred mainly between calcium oxalate subtypes due to overlapping radiological features.

Conclusions CNN-based analysis of NCCT images provides accurate, non-invasive prediction of urinary stone composition, with performance comparable to conventional methods. This approach may improve preoperative planning and broaden diagnostic access, particularly in settings where FTIR is unavailable. Prospective multicenter studies are warranted to validate clinical implementation.

Key Words: urolithiasis ‹› artificial intelligence ‹› computed tomography

INTRODUCTION

Urinary stone disease is a common condition, with an estimated prevalence ranging from 5% to 14%, and a high recurrence rate [1]. Additionally, urolithiasis is associated with a high economic burden due to diagnostic and therapeutic costs, as well as indirect costs such as work absenteeism. Current clinical guidelines recommend the chemical analysis of urinary stones to assess the risk of recurrence, determine the likelihood of treatment success, and guide preventive strategies [2].

Calcium is present in around 75% of cases, being the most frequent component of urinary stones. About 60% are composed of calcium oxalate, while mixed compositions of calcium and hydroxyapatite account for 20%, uric acid for 10%, struvite for around 10%, brushite for 2%, and cystine for 1% [3]. One of the validated techniques for analyzing stone composition is infrared spectroscopy, which utilizes infrared light absorption bands generated by the energy released from directional atomic vibrations [4]. Although this technique demonstrates high sensitivity and specificity for determining urinary stone composition, its high cost and limited availability in most laboratories pose significant limitations [5].

On the other hand, artificial intelligence (AI) is rapidly evolving, and its implementation in medicine has led to significant improvements in clinical practice, particularly in urology and in the study of urinary stone disease [6]. AI is defined as the process of optimizing an algorithm's reward function. Computational systems "learn" to identify patterns through various subdomains specifically designed to address the need for recognizing and processing complex inputs. These patterns are typically represented by artificial neural networks (ANNs). A standard ANN consists of a series of interconnected neurons (processors), where input neurons are activated by environmental sensors, and subsequent neurons are activated through weighted connections from already activated neurons. This process ultimately determines the behavior of the neural network. The "learning" process aims to develop a neural network capable of performing a desired task, such as remotely controlling a device [7].

The convolutional neural network (CNN), an advanced variant of the artificial neural network, is specifically designed to recognize spatial patterns and features in images by implementing convolutional layers, pooling, and fully connected layers. This approach reduces model complexity, optimizes memory and computational power usage, and provides data that were previously inaccessible. For instance, image distortions in computed tomography (CT) scans can hinder accurate assessment. CNNs address this issue by focusing on specific regions of the image, isolating them from surrounding structures, removing unwanted noise, and extracting information that was not previously available [8].

The objective of our study is to determine the chemical composition of urinary stones by analyzing non-contrast computed tomography images using a trained convolutional neural network.

MATERIAL AND METHODS

Study design and dataset

This was a retrospective analysis of an institutional database of patients who underwent urinary stone analysis using FTIR between the years 2020 and 2024. Stone samples included both outpatient submissions and intraoperatively retrieved fragments obtained during endourological procedures. Patients with more than one surgical intervention were included provided that a new stone sample and a new imaging study were available for each procedure.

From this cohort, we included in the present study only those patients who had a NCCT performed at our institution within the time window ranging from 3 months before to 48 hours after the FTIR analysis.

For each patient, between four and five augmented images focusing exclusively on the stone were extracted. Image selection aimed to minimize surrounding tissue or anatomical structures to avoid potential noise or confusion during CNN training.

To prevent data leakage and ensure methodological rigor, dataset partitioning was performed at the patient (stone) level rather than at the image level. All images derived from the same patient and corresponding stone were strictly assigned to a single subset (training, internal validation, or external validation). The split was conducted prior to data augmentation to avoid artificially increasing similarity between subsets. Consequently, no augmented or original images from the same stone were shared across different data partitions. The external validation cohort consisted exclusively of patients who were not included in the training or internal validation sets, thereby ensuring full independence of the evaluation dataset.

An independent validation cohort was defined at the patient level and consisted of 105 NCCT images corresponding to 105 patients (each representing a unique stone event). These cases were not used during model training or internal cross-validation. The validation cohort was derived from the same institution and imaging infrastructure but from a non-overlapping subset of patients, ensuring independence at the patient level. All scans were acquired using standardized institutional NCCT protocols. Therefore, this evaluation represents an independent temporal validation rather than a multicenter external validation.

We developed a CNN model capable of determining the main chemical composition of urinary lithiasis, using the images of non-contrast computed tomography scans of the abdomen and pelvis. The dataset was divided into a training set (80%, n = 1190 images) and a test set (20%, n = 297 images). CT scans were initially retrieved from the institutional PACS system and manually exported as two-dimensional JPEG (JPG) images for model development.

The exported images corresponded to non-contrast abdominal CT slices centered on the stone. Because images were converted from DICOM to JPG format, original metadata such as pixel spacing and exact Hounsfield unit (HU) calibration were not preserved.

Prior to CNN training, images were resized to a standardized matrix dimension to ensure computational compatibility. Intensity values were normalized to a fixed range to facilitate model convergence. No additional windowing manipulations were applied beyond those present at the time of image export. Distribution of urinary stone types is shown in Figure 1.

Data preprocessing and augmentation

Data preprocessing and augmentation were performed using the ImageDataGenerator module from the Keras library. Applied transformations included shear range (0.2), zoom range (0.2), and random vertical/horizontal flip. Images were processed in batches of 3 (batch size = 3) during training and evaluation.

Convolutional neural networks model architecture

Transfer learning was employed utilizing the pre-trained VGG16 architecture initialized with ImageNet weights. Original, fully connected layers were excluded, and the convolutional layers of the pre-trained model were initially frozen to retain relevant general features [9].

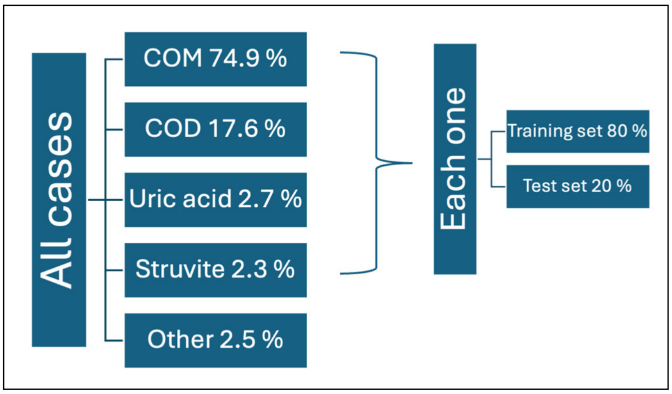

Figure 1. *Distribution of urinary stone types and corresponding data partitioning for model development. Each stone subtype – calcium oxalate monohydrate (COM, 74.9%), calcium oxalate dihydrate (COD, 17.6%), uric acid (2.7%), struvite (2.3%), and other compositions (2.5%) – was individually split into a training set (80%) and a test set (20%) for the development and internal validation of the convolutional neural network (CNN).*

Subsequently, a customized architecture was added, consisting of the following:

- One dense, fully connected layer with 128 neurons and Rectified Linear Unit (ReLU) activation.
- A dropout layer configured at 35% (0.35) to reduce overfitting.
- A final classification layer with Softmax activation for categorical discrimination between different stone composition groups.

A fine-tuning training phase was conducted by selectively unfreezing the last convolutional layers of the pre-trained model to optimize performance on the specific images used in this study.

Compilation and training

The model was compiled using the Adam optimizer with an initial learning rate of 0.0001, categorical cross-entropy loss function, and categorical accuracy as the evaluation metric during training and validation. Initially, training was conducted for 50 epochs, followed by implementation of Early Stopping, monitoring the "validation loss" with a patience of 5 epochs to prevent overfitting.

Internal validation strategy

Internal evaluation of the model was performed using a 10-fold cross-validation strategy, ensuring a robust and reliable assessment of its performance.

Software and computational environment

The entire development and training process was conducted using Python version 3.9 within the Spyder 6.0 environment and Anaconda Navigator software. TensorFlow 2.10 and Keras libraries were primarily utilized for model construction and training. The training was executed on a computer equipped with a 10-core GPU and 24 GB of RAM.

Class imbalance and training strategy

The distribution of stone composition in the dataset was markedly imbalanced, with calcium oxalate monohydrate accounting for approximately 75% of all cases. To preserve the natural epidemiological distribution of stone types, no explicit class-balancing techniques (e.g., class-weighted loss, oversampling, or undersampling) were applied; the model was trained on the original class frequencies. We acknowledge that this strategy favors ecological validity at the expense of potential bias toward the majority class.

RESULTS

A total of 2,516 patients had FTIR-confirmed stone composition. They were filtered by those who met the inclusion criterion: NCCT performed at our institution within a time window ranging from 3 months before to 48 hours after the FTIR analysis. A total of 336 patients met these criteria and were included (Table 1). From there, a total of 1,592 NCCT images scans with urinary stone disease were retrospectively reviewed. After preprocessing, data augmentation, and annotation, 1,487 scan images were allocated for training and internal validation of the CNN, and the remaining 105 scan images constituted the external validation set.

All cases were labeled according to the stone composition confirmed via Fourier transform infrared spectroscopy. The distribution of stone types in the dataset was as follows: calcium oxalate monohydrate (COM) 74.9%, calcium oxalate dihydrate (COD) 17.6%, uric acid 2.7%, struvite 2.3%, and other compositions 2.5%.

Following training, the model demonstrated robust performance metrics:

In the training set, the model achieved an accuracy of 89.8%, with an area under the receiver operating characteristic curve (AUC ROC) of 0.92.

In the validation set of 30 independent cases, the model maintained an accuracy of 85.4%, with a corresponding AUC ROC of 0.81 and a Cohen's κ coefficient of 0.864, indicating a strong agreement between model predictions and ground-truth labels.

Performance across stone types was consistent, with the highest precision observed for calcium oxalate monohydrate, the most prevalent and clinically relevant subtype. Notably, the model exhibited occasional misclassifications between COM and COD stones, likely due to their overlapping Hounsfield unit (HU) ranges and similar morphology on CT imaging.

Table 1. *Baseline demographic and anatomical characteristics*

Characteristic	n (%)
Mean age (years)	50
Sex	
Male	204 (60.7)
Female	132 (39.3)
Stone location	
Renal	120 (37.5)
Proximal ureter	65 (19.3)
Mid ureter	27 (8)
Distal ureter	124 (36.9)

DISCUSSION

Accurate determination of urinary stone composition is essential because it directly informs etiological evaluation, guides treatment decisions, and supports strategies to prevent recurrence. Current guidelines from the AUA and EAU recommend infrared spectroscopy or X-ray diffraction as the reference standards for stone analysis [2, 4]. However, these techniques are not universally available and may be costly, limiting their use in many clinical settings.

In recent years, artificial intelligence has rapidly expanded into clinical practice, particularly in urology. A recent systematic review analyzing 71 studies reported an overall accuracy of 88.2% for stone composition prediction, 96% for stone detection, and 87% for predicting spontaneous passage [6]. Similarly, Panthier et al. (2024) reported sensitivity ranging from 58.7% to 100%, specificity from 68.5% to 100%, and accuracy up to 99.95% for AI-based detection of urinary stones on CT imaging [10].

Within this context, our study evaluated the ability of a convolutional neural network to predict urinary stone composition using standard non-contrast CT images. Our model achieved high predictive performance, with strong agreement between predicted and true labels, as reflected by both accuracy and Cohen's κ. These results are consistent with prior studies in the field.

For example, Le et al. demonstrated that radiomics combined with machine learning could identify uric acid stones with up to 94% accuracy [11, 12], while Tang et al. reported an AUC of 0.933 for predicting calcium oxalate monohydrate stones using AI-based models [13]. Compared with these approaches, which rely on handcrafted radiomic features, our model uses a CNN applied directly to standard CT images. This may simplify implementation and improve scalability in routine clinical practice [14].

In addition, unlike studies based on intraoperative endoscopic images [15, 16], our approach enables preoperative prediction, which is more directly applicable to clinical decision-making and treatment planning.

Despite these encouraging results, several limitations should be acknowledged.

First, the dataset was markedly imbalanced, with calcium oxalate monohydrate accounting for approximately 75% of cases. While this reflects real-world epidemiology, it may bias the model toward the majority class and limit performance in less frequent stone types. Future studies should explore class-balancing strategies to improve model robustness across all compositions.

Second, although we performed an independent temporal validation, all data originated from a single institution using similar imaging protocols. Therefore, our findings require confirmation in true external multicenter datasets, including different scanners and acquisition settings, to ensure generalizability.

Third, the model was trained using manually selected two-dimensional CT slices rather than full three-dimensional volumes. Given that urinary stones are inherently heterogeneous 3D structures, this approach may not capture the full compositional complexity. In addition, manual slice selection introduces operator-dependent bias and limits reproducibility. Future work should focus on automated 3D segmentation and volumetric analysis.

Another important limitation relates to the reference standard. FTIR provides the overall composition of the entire stone, whereas the model was trained on selected CT slices representing only part of the stone. In mixed stones, this mismatch may introduce label noise and partially explain some misclassifications.

From a technical perspective, images were exported in JPEG format rather than being analyzed directly from DICOM files. As a result, important quantitative information such as Hounsfield units and pixel spacing was not preserved, and compression artifacts may have been introduced. This may affect reproducibility, and it highlights the need for future DICOM-based pipelines.

Additionally, we did not perform formal model interpretability analysis (e.g., Grad-CAM), which could help identify the image regions driving predictions. Incorporating explainability tools will be important to improve clinical trust and facilitate adoption.

Despite these limitations, our study demonstrates the feasibility of using deep learning to estimate urinary stone composition from routine CT imaging, achieving performance within the range reported in prior AI-based imaging studies [17].

From a clinical perspective, this approach may be particularly valuable in settings where FTIR is not readily available or is cost-prohibitive. In such contexts, AI-based tools could provide a low-cost, accessible alternative to support decision-making. Rather than replacing standard laboratory techniques, this technology should be viewed as a complementary tool that can enhance preoperative planning and expand access to compositional information.

Future research should focus on prospective, multicenter validation, inclusion of more diverse and mixed stone types, and integration of imaging with clinical and laboratory data into hybrid predictive models [18]. Ultimately, embedding these algorithms into radiology workflows or PACS systems could enable real-time, automated decision support in daily practice [19, 20].

CONCLUSIONS

The application of convolutional neural networks to non-contrast-enhanced computed tomography images demonstrates high diagnostic accuracy in predicting the chemical composition of urinary stones. The application of convolutional neural networks to non-contrast-enhanced computed tomography images demonstrates promising accuracy in estimating the chemical composition of urinary stones in a non-invasive manner. Rather than replacing established laboratory techniques such as FTIR or X-ray diffraction, this AI-based approach may serve as a complementary preoperative decision-support tool, particularly in settings where conventional compositional analysis is unavailable, delayed, or resource-constrained. Further prospective and multicenter validation is required before clinical implementation. By enabling early and personalized decision-making, it could improve surgical planning and facilitate preventive strategies tailored to stone type. Future efforts should aim to include a broader and more heterogeneous patient population, particularly underrepresented or mixed stone types, and to externally validate model performance in prospective, multicenter clinical trials. Additionally, integrating clinical and laboratory variables with imaging data may further enhance the predictive capability of these models. Ultimately, the integration of AI-based tools into radiological workstations or PACS could provide real-time decision support for urologists during diagnostic and therapeutic planning.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ETHICS APPROVAL STATEMENT

The ethical approval was not required.

References

1. Akram M, Jahrreiss V, Skolarikos A, et al. Urological Guidelines for Kidney Stones: Overview and Comprehensive Update. J Clin Med. 2024; 13: 1114.
2. Lemoine S, Dahan P, Haymann JP, Meria P, Almeras C; Lithiasis Committee of the French Association of Urology (CLAFU). 2022 Recommendations of the AFU Lithiasis Committee: Medical management – from diagnosis to treatment. Prog Urol. 2023; 33: 911-953.
3. Saigal CS, Joyce G, Timilsina AR; Urologic Diseases in America Project. Direct and indirect costs of nephrolithiasis in an employed population: opportunity for disease management? Kidney Int. 2005; 68: 1808-1814.
4. Türk C, Neisius A, Petrik A, Seitz C, Skolarikos A, Thomas K. EAU Guidelines on Urolithiasis. Eur Assoc Urol. 2020: 12-13.
5. Sánchez R, Navarro P, Troncoso MP, López C, Salvadó JA. Análisis de la composición química de litiasis urinaria de 649 pacientes mediante espectroscopía infrarroja, primera experiencia nacional [Chemical composition analysis of 649 urinary stones]. Rev Med Chil. 2021; 149: 1129-1133.
6. Altunhan A, Soyturk S, Guldibi F, et al. Artificial intelligence in urolithiasis: a systematic review of utilization and effectiveness. World J Urol. 2024; 42: 579.
7. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015; 61: 85-117.
8. Kyong Hwan Jin, McCann MT, Froustey E, Unser M. Deep Convolutional Neural Network for Inverse Problems in Imaging. IEEE Trans Image Process. 2017; 26: 4509-4522.
9. Müller L, Tibyampansha D, Mildenberger P, Panholzer T, Jungmann F, Halfmann MC. Convolutional neural network-based kidney volume estimation from low-dose unenhanced computed tomography scans. BMC Med Imaging. 2023; 23: 187.
10. Panthier F, Melchionna A, Crawford-Smith H, et al. Can Artificial Intelligence Accurately Detect Urinary Stones? A Systematic Review. J Endourol. 2024; 38: 725-740.
11. Le BD, Nguyen TA, Baek BH, Oh KJ, Park I. Accurate prediction of pure uric acid urinary stones in clinical context via a combination of radiomics and machine learning. World J Urol. 2024; 42: 150.
12. Black KM, Law H, Aldoukhi A, Deng J, Ghani KR. Deep learning computer vision algorithm for detecting kidney stone composition. BJU Int. 2020; 125: 920-924.
13. Tang L, Li W, Zeng X, et al. Value of artificial intelligence model based on unenhanced computed tomography of urinary tract for preoperative prediction of calcium oxalate monohydrate stones in vivo. Ann Transl Med. 2021; 9: 1129.
14. Zhang GM, Sun H, Shi B, Xu M, Xue HD, Jin ZY. Uric acid versus non-uric acid urinary stones: differentiation with single energy CT texture analysis. Clin Radiol. 2018; 73: 792-799.
15. Lim EJ, Castellani D, So WZ, et al. Radiomics in Urolithiasis: Systematic Review of Current Applications, Limitations, and Future Directions. J Clin Med. 2022; 11: 5151.
16. Zhu G, Li C, Guo Y, et al. Predicting stone composition via machine-learning models trained on intra-operative endoscopic digital images. BMC Urol. 2024; 24: 5.
17. McGrath TA, Frank RA, Schieda N, et al. Diagnostic accuracy of dual-energy computed tomography (DECT) to differentiate uric acid from non-uric acid calculi: systematic review and meta-analysis. Eur Radiol. 2020; 30: 2791-2801.
18. Wang Z, Yang G, Wang X, Cao Y, Jiao W, Niu H. A combined model based on CT radiomics and clinical variables to predict uric acid calculi which have a good accuracy. Urolithiasis. 2023; 51: 37.
19. Stone L. Assessing kidney stone composition using deep learning. Nat Rev Urol. 2020; 17: 192-193.
20. Popova E, Tkachev S, Shapoval A, et al. Kidney Stones as Minerals: How Methods from Geology Could Inform Urolithiasis Treatment. J Clin Med. 2025; 14: 997.