Asian Journal of Urology 10 (2023) 258–274

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ajur

Review

Transforming urinary stone disease management by artificial intelligence-based methods: A comprehensive review

Anastasios Anastasiadis [a], Antonios Koudonas [a], Georgios Langas [a], Stavros Tsiakaras [a,*], Dimitrios Memmos [a], Ioannis Mykoniatis [a], Evangelos N. Symeonidis [a], Dimitrios Tsiptsios [b], Eliophotos Savvides [c], Ioannis Vakalopoulos [a], Georgios Dimitriadis [a], Jean de la Rosette [d]

[a] *1st Department of Urology, Aristotle University of Thessaloniki, School of Medicine, "G.Gennimatas" General Hospital, Thessaloniki, Greece*
[b] *Neurology Department, Democritus University of Thrace, Alexandroupolis, Greece*
[c] *Department of Urology, Main Kinzig Kliniken, Gelnhausen, Germany*
[d] *Department of Urology, Istanbul Medipol Mega University Hospital, Istanbul, Turkey*

Received 12 September 2022; received in revised form 23 October 2022; accepted 10 February 2023
Available online 2 May 2023

KEYWORDS
Artificial intelligence;
Urolithiasis;
Endourology;
Machine learning;
Stone disease

Abstract *Objective:* To provide a comprehensive review on the existing research and evidence regarding artificial intelligence (AI) applications in the assessment and management of urinary stone disease.
Methods: A comprehensive literature review was performed using PubMed, Scopus, and Google Scholar databases to identify publications about innovative concepts or supporting applications of AI in the improvement of every medical procedure relating to stone disease. The terms ''endourology'', ''artificial intelligence'', ''machine learning'', and ''urolithiasis'' were used for searching eligible reports, while review articles, articles referring to automated procedures without AI application, and editorial comments were excluded from the final set of publications. The search was conducted from January 2000 to September 2023 and included manuscripts in the English language.
Results: A total of 69 studies were identified. The main subjects were related to the detection of urinary stones, the prediction of the outcome of conservative or operative management, the

* Corresponding author.
E-mail address: drstavros90@gmail.com (S. Tsiakaras).
Peer review under responsibility of Tongji University.

https://doi.org/10.1016/j.ajur.2023.02.002

optimization of operative procedures, and the elucidation of the relation of urinary stone chemistry with various factors.
Conclusion: AI represents a useful tool that provides urologists with numerous amenities, which explains the fact that it has gained ground in the pursuit of stone disease management perfection. The effectiveness of diagnosis and therapy can be increased by using it as an alternative or adjunct to the already existing data. However, little is known concerning the potential of this vast field. Electronic patient records, containing big data, offer AI the opportunity to develop and analyze more precise and efficient diagnostic and treatment algorithms. Nevertheless, the existing applications are not generalizable in real-life practice, and high-quality studies are needed to establish the integration of AI in the management of urinary stone disease.

1. Introduction

Artificial intelligence (AI) represents a scientific field where technology simulates human intelligent behavior and way of thinking, in order to process complex medical data and provide useful information to the physician [1]. Back in 1950, Alan Turing questioned whether computers could think like human beings. After the completion of the so-called "Turing test", Turing is considered the father of AI [2]. Although the idea of AI was conceived in the 21st century, it still addresses the interest of various scientific fields, including medicine, to date.

AI represents the process of optimizing the reward function of an algorithm [3]. Similar to the chess analogy process, where a large set of data from chess games is gathered and analyzed, AI makes large dataset analysis feasible and effective. Computer systems "learn" to recognize patterns through a variety of subdomains, specifically designed to meet the need of identifying and processing certain patterns. These patterns are typically represented by the so-called artificial neural networks (ANNs). A typical ANN comprises a number of neurons (the processors), which are connected in a logical manner. Input neurons are engaged by environmental sensors, while additional neurons are activated by weighted connections with previously active neurons. This process eventually configures the behavior of the respective ANN. The procedure of "learning" aspires to constitute an ANN able to carry out the foreseen behavior, such as remotely controlling a device [4].

Machine learning (ML) is a subtype of AI (Fig. 1) representing unsupervised learning based on the conclusions drawn by the statistical analysis of large and complex neural networks (NNs) [5]. ML opts to provide clinicians with possible rather than "correct" answers, as it describes the likelihood of a correlation between two or more variables. The emergence of algorithms capable of adjusting to advancing technologies and conditions, high-performance computing, access to personal scientific findings through cloud drives, as well as the utilization of open-access editing software are the four key points that ensure the proper and flawless operation of ML as an invaluable tool to its user [6].

Deep learning (DL), a subfield of ML, is considered a "self-taught" concept, based on the fundamentals of algorithm layering in order to build an ANN able to learn and decide for itself (Fig. 1) [7]. On the contrary, ML typically requires human intervention as part of its maintenance and proper function, whereas DL works as a human brain, independently and without external human intervention [8].

Convolutional neural network (CNN), a more complex type of ANN, is developed to incorporate spatial motives and features of an image by using fully connected, convolution, and pooling layers. It reduces the complexity of the model, requires less memory and computational energy, and provides the user with data previously inaccessible [9]. For example, image distortion in computed tomography (CT) examinations may impede the safe evaluation of the findings. To surpass such inconvenience, CNN focuses on certain regions of an image, isolates them from the neighboring structures, filters out undesired noise, and provides image information previously inaccessible [10].

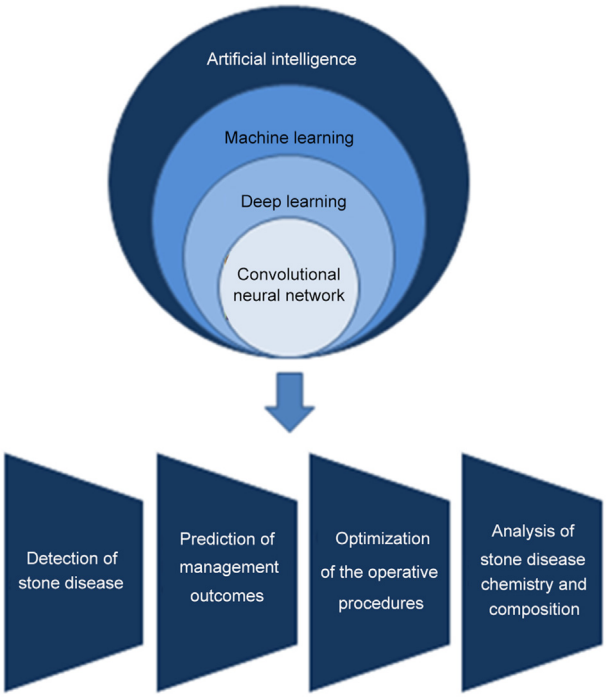

Figure 1 Subsets of artificial intelligence with emergent role in stone disease management.

In medicine, AI is further divided into two subcategories: virtual and physical [11]. The virtual part refers to the information and system-based learning incorporated to facilitate the optimal treatment. Electronic health records and treatment algorithms are created with the implication of AI and its subtypes. The physical part represents the technological advances, such as robotics and nanotechnology, developed to pursue maximum intraoperative efficiency and safety [12]. The development of such distinct AI types enables the use of a fully functioning artificial electronic brain, programmed to evaluate and analyze all tasks, regardless of their complexity.

This integrative review aimed to provide an overview of the contribution of AI to the diagnosis, evaluation, and treatment of stone disease and highlight its effect on the field of endourology. Electronic base searching (PubMed, Google Scholar, and Scopus) with proper terms resulted to a number of reports, which after screening from the authors were reduced to the 69 included studies of the current review (Fig. 2).

2. AI in the detection of stone disease

2.1. Optimization of stone disease detection by CT

During the last decades, the use of abdomen and pelvis CT to identify and diagnose urinary tract stones has increased. It is the most sensitive technique for the diagnosis of urolithiasis; thus, it has emerged as the examination of choice. This increasing tendency to unenhanced CT leads to the development of computer-assisted technologies to aid therapists in this time-consuming process (Table 1).

The AI-assisted workflow may aid in patient triage and streamlining as the usage of imaging keeps growing. Li et al. [13] used five state-of-art learning algorithms (three-dimensional [3D] U-Net, Res-U-Net, SegNet, DeepLabV3+, and UNETR) in order to address the automatic segmentation of kidneys and kidney stones in unenhanced CT images. They trained the segmentation networks independently (one-step direct segmentation) and dependently (two-step coarse-to-fine segmentation). The comparison of the above networks showed that the Res-U-Net network outperformed the other networks. The two-step segmentation showed better results than the one-step, although it has some limitations like the heavy computing requirements and the dependence on the precision of the first-step segmentation. Additionally, the segmentation of larger kidney stones had better results.

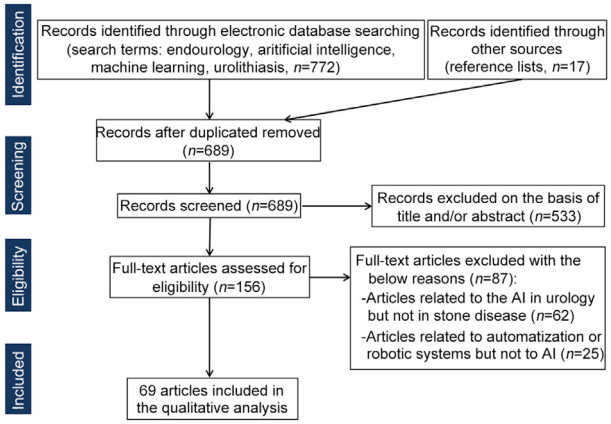

Figure 2 Flowchart of the literature selection process for articles.

Parakh et al. [14] focused on examining the diagnostic performance of pre-trained models improved with labelled CT images, across various scanners, for the identification of urinary stones using a cascading CNN on non-enhanced CT images. They developed two CNNs: CNN1 for identifying the anatomy of the urinary tract and CNN2 for urinary stone detection. They used the Inception-v3 CNN architecture which was pre-trained with ImageNet. The ImageNet pre-trained model was then fine-tuned in GrayNet, which contains a human anatomy image dataset (GrayNet pre-trained model). The CNN models for identifying the urinary tract and detecting stones were then weight initialized using the GrayNet pre-trained model. A high accuracy rate in detecting stones in the urinary tract was achieved (area under the curve [AUC] 0.954). In a study by Längkvist et al. [15], a DL CNN was used to distinguish ureteric stones from phleboliths based on thin slice CT images from a 465-patient database, a rather challenging task even for experienced radiologists. On a test set of 88 scans, the CNN model had a sensitivity of 100% and an average of 3.69 false positives per patient without using segmentation and anatomical information. Adding a probability distribution map of the stone's placements produced an average of 2.68 false positive tests per patient and 100% sensitivity [15]. Caglayan et al. [16] examined the effectiveness of a DL model in identifying kidney stones on unenhanced CT images in various planes based on stone size. In their retrospective study, 465 patients, who underwent CT scanning for kidney stones, formed three groups depending on the size of the kidney stones (Group 1 contained patients with renal stone sizes of 0—1 cm; Group 2 had sizes of 1 cm—2 cm; and Group 3 had sizes greater than 2 cm). The highest accuracy rates were in the sagittal plane images of 85%, 89%, and 93% for the three groups, respectively [16]. Jendeberg et al. [17] used a 2.5-dimensional CNN of 384 pelvic calcifications, on an unseen test set. They did a comparison between the CNN method, the assessments of seven radiologists, and a semi-quantitative method. The CNN performed better in comparison to the radiologists. It achieved an accuracy of 92% while the mean radiologist accuracy was 86%. In terms of distinguishing distal ureteric stones from phleboliths, the sensitivity, specificity, and AUC were 94%, 90%, and 0.95, respectively [17]. De Perrot et al. [18] investigated the performance of a ML model trained with radiomics, extracted from a cohort of 369 patients who underwent low-dose CT for acute flank pain. The ML model was then used to identify and distinguish phleboliths from ureteral stones in an independent testing cohort of 43 patients. The ML model achieved an AUC of 0.902, an accuracy of 85.1%, and a positive predictive value (PPV) and negative predictive value of 81.5% and 90.0%, respectively. They mentioned that one of the advantages of using radiomics and ML is its quantitative nature and reproducibility. Chak et al. [19] used an ANN with support vector machine (SVM) in order to detect kidney stones in pre-processed CT images. They trained the ANN system

Table 1 Summary of studies regarding AI in the detection of stone disease.

Study	Objective	Study design	AI-based outcome	Comparator arm outcome
Li et al. [13]	Detection of urinary stones by CT	Cross-sectional	Detection accuracy of 99.95%	Other algorithms with lower performance
Parakh et al. [14]	Detection of urinary stones by CT	Cross-sectional	High accuracy in stone detection (AUC of 0.954)	Other algorithms with lower performance
Längkvist et al. [15]	Detection of urinary stones by CT	Cross-sectional	Optimized accuracy with an AUC of 0.9971	No comparator
Caglayan et al. [16]	Detection of urinary stones by CT	Cross-sectional	Accuracy of 63%–93%, depending on imaging plane and stone size class	No comparator
Jendeberg et al. [17]	Differentiation of ureteral stones and pelvic phleboliths by CT	Cross-sectional	Accuracy of 92%	Mean radiologist accuracy: 86%; majority vote accuracy: 93%
De Perrot et al. [18]	Differentiation of ureteral stones and pelvic phleboliths by CT	Cross-sectional	Overall accuracy of 85.1% (AUC of 0.902)	Other algorithms with lower performance
Chak et al. [19]	Detection of urinary stones by CT	Cross-sectional	Accuracy of 95%–99%, depending on the number of features used by the algorithm	No comparator
G P et al. [20]	Detection of urinary stones by CT	Cross-sectional	Accuracy of 96.82%	No comparator
Elton et al. [21]	Detection of urinary stones by CT	Cross-sectional	High accuracy in stone detection (AUC of 0.95)	No comparator
Krishna et al. [22]	Differentiation of renal stones and renal cysts by US	Cross-sectional	Accuracy of 98.1%	No comparator
Balamurugan and Arumugam [23]	Differentiation of renal stones among other abnormalities in US	Cross-sectional	Accuracy of 95.83%	Other algorithms with lower performance
Selvarani and Rajendran [24]	Detection of renal stones by US	Cross-sectional	Accuracy of 98.8%	Other algorithms with lower performance
Viswanath et al. [25]	Detection of renal stones by US	Cross-sectional	Accuracy of 98.9%	Other algorithms with lower performance
Akkasaligar and Biradar [26]	Detection of renal stones by US	Cross-sectional	Accuracy of 96.8%	No comparator
Verma et al. [27]	Detection of renal stones by US	Cross-sectional	Accuracy of 89%	Other algorithms with lower performance
Kobayashi et al. [28]	Detection of radio-opaque urinary stones in KUB X-ray images	Cross-sectional	Sensitivity and PPV of 89.6% and 56.9% for the kidney, 92.5% and 87.6% for the proximal ureter, 59.1% and 50% for the mid-ureter, 80% and 55.8% for the distal ureter	No comparator
Aksakalli et al. [29]	Detection of radio-opaque urinary stones in KUB X-ray images	Cross-sectional	Precision of 78.4%	Other algorithms with lower performance

AI, artificial intelligence; AUC, area under the curve; CT, computed tomography; KUB, kidney-ureter-bladder; PPV, positive predictive value; US, ultrasound.

with gray level co-occurrence matrix (GLCM) features from the segmented images. These CT images were before processed and visual enhanced, while with the use of K-means segmentation method, the region of interest was determined. They claimed accuracy of 95% in the detection of renal stones in CT images of 25 patients. G P et al. [20] used a CNN (particularly, the Xception model), which was trained on the ImageNet database. The CT dataset consisted of 1453 CT kidney images, and they claimed accuracy of 96.82% with this DL model combination. In another study by Elton et al. [21], a DL-based system (a 13-layer CNN classifier) achieved a sensitivity of 86% at 0.5 false positives per scan in kidney stone identification in CT scan images. The system classified CT images from an external validation

set (CT scans from 6185 patients) with an AUC of 0.95, while sensitivity and specificity were 88% and 91%, respectively.

2.2. Optimization of stone disease detection by ultrasound (US)

In order to detect irregularity in the kidney using US images, Krishna et al. [22] suggested a field programmable gate array-based computer-aided detection algorithm. They first pre-processed (denoising) the US images and manually extracted the region of interest. From the segmented kidney region, Haralick (GLCM) and intensity histogram features were extracted. The algorithm was implemented in differentiating normal from abnormal kidney images first with the look-up table approach. A trained SVM with multi-layer perception (MLP) classifier was then used to distinguish the abnormal images of renal cysts and stones. The suggested algorithm identified successfully the precise abnormality shown on the renal US images with an accuracy of 98.1%, sensitivity of 100%, and specificity of 96.8%.

Balamurugan and Arumugam [23] proposed a novel US kidney disease prediction with the use of an ANN. They used 750 kidney US images, 80% of which were used for the training process and 20% for testing. The images were utilized in four types, *i.e.*, normal, cyst, stones image, and tumor image. For implementing the proposed technique, they have used the MATrix LABoratory platform (MathWorks, Natick, MA, USA). They pre-processed the US images, extracted the GLCM features, selected the important features with the use of the oppositional grasshopper optimization algorithm. The images were then classified as normal or abnormal with the ANN. They claimed a maximum accuracy of 95.83% and a maximum specificity of 97.22%.

For recognizing renal stones on US images, Selvarani and Rajendran [24] utilized the meta-heuristic SVM. They used an adaptive mean median filter (MathWorks, Natick, MA, USA) to eliminate speckle noises to the greatest extent documented in the literature. Utilizing standard K-means for segmentation, GLCM features were retrieved for classification and used by a meta-heuristic SVM classifier. The system showed a 98.8% accuracy after being trained on 250 US pictures (150 with stones and 100 without stones).

In another study by Viswanath et al. [25], a DL reaction-diffusion level set segmentation approach was designed to detect kidney stones in US images. They used a plain intensity filter for image processing and quality enhancement which showed better results than other pre-processing methods. Advanced wavelet sub-bands filters were used for feature extracting from the processed images based on the energy level. The proposed backpropagation, multi-layer perceptron, and SVM ANN (MLP-backpropagation ANN) proved to have an accuracy of 93.2%, with an average of 7.08 s for the image segmentation.

Akkasaligar and Biradar [26] used wavelet decomposition in 32 US images of kidneys for stone identification. The proposed method with the use of an ANN claimed to have an accuracy of 96.8%. In a study from Verma et al. [27] for stone detection in kidney US images, *k*th nearest neighbor (KNN) and SVM classification were used. The US images were at first processed and improved with Gaussian filter and unsharp masking. The analysis of kidney stone pictures then employed morphological procedures including erosion and dilation, followed by entropy-based segmentation to identify the region of interest and KNN and SVM classification approaches. The KNN accuracy was found to be 89% and that of SVM was 84%.

2.3. Optimization of stone disease detection by X-ray

Kidney-ureter-bladder X-ray imaging is not the method of choice for the detection of urolithiasis. The prerequisite of the stone radio-opacity and the subsequent low sensitivity represent the main disadvantages of the above modality. Conversely, X-ray is a low-cost, low-dose with high availability examination; thus it remains a diagnostic alternative for urinary tract stones. AI can improve the diagnostic accuracy of kidney-ureter-bladder X-ray, especially for radio-opaque stones (Table 1).

Kobayashi et al. [28] created a computer-aided detection system employing DL, to identify radio-opaque urinary tract stones on a plain X-ray. They collected 1017 plain X-rays of radio-opaque upper urinary tract stones, 827 of which were used for training purposes and 190 for testing. To account for the various picture sizes, all X-ray images were adjusted to 1328×1328 pixels, while for contrast improvement, histogram equalization was employed. The CNN architecture, which was used in the study, was a 17-layer Residual Network. As a consequence of the computer's prediction, the possibility that each pixel contained a stone was determined and shown in each input image as a heat map, with light red denoting a high chance of stone presence (100%) and dark green denoting zero probability (0%). For the performance evaluation, they evaluated the F score that demonstrates the balance of the accuracy and is the harmonic mean of the sensitivity and PPV. The 17-layer Residual Network model provided an answer in 110 ms for each X-ray image. The highest F score was 0.752, and the greatest sensitivity and PPV values were 87.2% and 66.2%, respectively. The sensitivity and PPV were at their lowest at mid-ureter when the study was constrained to proximal ureter stones, at 92.5% and 87.6%, respectively.

Aksakalli et al. [29] used different ML and DL methods to identify kidney stones in X-ray images. They evaluated various ML methods such as decision tree (DT), random forest, SVM, MLP, KNN, Naive Bayes (BernoulliNB), and deep NN (DNN) using CNN. They prepared the dataset of 221 X-ray images, 80% of which were used for training and 20% for testing purposes. Images were first scaled to predetermined sizes and then transformed to the grey level values. Then, data collection was produced by extracting the pictures' grey-level numerical values. There were several oversampling and under sampling techniques that were applied since this data set has unbalanced classes. The methods' performance metrics noticeably improve in this approach. The F1 score (harmonic average of the precision and recall values) was the most essential evaluation in their study. The DT approach seemed to outperform the other methods, showing the highest F1 score with a success rate of 85.3% at the synthetic minority over-sampling technique sampling method.

3. AI in the prediction of management outcomes

3.1. Prediction of outcome of conservative management

Symptomatic ureteral stones are one of the main reasons for attending the urology department. In this context, the clinical question of spontaneous stone passage (SSP) can be challenging and affects the patient's quality of life. The possibility of SSP varies depending mostly on stone size. In this scope, the need for faster and more accurately predicting algorithms emerges, in order to improve patient's quality of life with more efficient management and early intervention when needed (Table 2).

In 2000, Cummings et al. [30], addressing the SSP clinical question, used a commercially available ANN. They evaluated the data of 180 patients, 125 of which were used for training the NN and 55 for testing it. Feed forward-back propagation was used for error adjustment. In feature importance analysis, the duration of the symptoms had the highest weight. The accuracy rates of predicting the SSP were 100% (25/25) and 76% (42/55) for those who needed intervention.

Dal Moro et al. [31] used a prototype logistic regression (LR)-ANN-SVM prediction model for SSP for ureteral stones. They claimed that the addition of SVM improved the model performance in predicting the need for additional intervention. The SVM-based model provided a sensitivity of 84.5% and a specificity of 86.9%. Additionally, in this study, from the feature importance analysis, stone size, duration of symptoms, and stone position outclassed the other features in predicting SSP.

In another study, a prototype ANN was used for the SSP estimation and to assess the efficacy of the predictive factors [32]. A dataset of 192 patients was used and randomly divided into three groups (training, validation, and test). The ANN's performance for the estimation of SSP in the three groups was 99.2%, 85.5%, and 88.7%, respectively. The predictive factors that were more significantly correlated with the SSP were stone size, body weight, pain score, erythrocyte sedimentation rate, and C-reactive protein level.

Park et al. [33] used a ML and LR model for SSP estimation. They retrospectively reviewed medical data from 833 patients who attended the emergency department with unilateral ureteral stones. They evaluated the accuracy of the standard statistical approach (LR) with the DL method (MLP based on the Keras framework), in predicting SSP. MLP outperformed LR in SSP for 5–10 mm stones with a specificity of 100%. For predicting SSP, for ureteral stones of $<$5 mm, AUCs for MLP and LR were 0.859 and 0.847, while for the 5 mm–10 mm stones, AUCs were 0.881 and 0.817, respectively. The highest sensitivity observed was 90.5% for the LR model for the 5–10 mm stones. They claimed no significantly important difference between the two models and mentioned the need for image analysis for the improvement of the prediction.

3.2. Prediction of extracorporeal shockwave lithotripsy (ESWL) outcome

ESWL has gained widespread acceptance as a practical, non-invasive management method for urinary stones preferably smaller than 2 cm. The success and stone-free rates (SFRs) of the procedure are dependent on various factors such as initial stone size, location, number of stones, stone composition, and stone density. In the scope of predicting the outcomes and reducing the procedures' side effects, due to this variability, many mathematical and computational methods are being studied (Table 2).

In 2003, Poulakis et al. [34] reviewed the records of 680 patients (701 renal units) who underwent primary shock wave lithotripsy for lower pole renal calculi. In their univariate analysis, they assessed the important variables for lower pole calculi clearance after ESWL. They used an ANN to predict the SFR in the test group. They claimed an accuracy of 92% and an AUC of 0.936. As mentioned by the authors, urinary transport proved as the most important variable for lower pole SFR prediction by the ANN. In addition, infundibuloureteropelvic angle 2, caliceal pelvic height, and body mass index were also reported as significant factors for the ANN's performance.

Gomha et al. [35] compared an ANN model to a LR model. The two models were compared in their performance in predicting the stone-free status (SFS) at 3 months post ESWL, using 10 inputs for ANN and covariates for LR. Both models performed adequately, with ANN outperforming LR in predicting those who fail to respond to ESWL. Moorthy and Krishnan [36] used an ANN for the prediction of stone fragmentation, depending on different features extracted from non-enhanced CT images. By the method of first order, statistical features like mean, variance, skewness, and kurtosis were calculated and evaluated by ANN. The findings showed that, when the mean was taken into account as the feature of interest, the model prediction demonstrated a sensitivity of 80.7%. Moreover, 90% of the true positive and true negative instances were accurately identified. A specificity of 98.4% allowed for the identification of true negative patients.

In a study by Choo et al. [37], a ML model was proposed for detecting the SFS after a single ESWL session for ureteral stones. They defined success of the single ESWL session as the absence of fragments bigger than 2 mm on CT or plain X-ray images, 2 weeks after the procedure. The model showed improvement in accuracy performance when more factors were added. All potential combinations of factors were included in the construction of the decision models. With the use of DT analysis, the 15-factor model appeared to have an accuracy of 92.29% and an average receiver operating characteristic AUC of 0.951. In this study, the stone volume seemed to be the most relevant factor for the prediction of SFS after a single session of ESWL.

Seckiner et al. [38] aiming to predict the SFS and to aid in procedure planning, used a prototype ANN in a sample of 203 patients who had undergone ESWL. Both regression analysis and ANN were applied for the SFS determination.

Table 2 Summary of studies regarding AI in the prediction of management outcomes.

Study	Objective	Study design	AI-based outcome	Comparator arm outcome
Cummings et al. [30]	Prediction of SSP	Case-control	Accuracy of 76%	No comparator
Dal Moro et al. [31]	Prediction of SSP	Case-control	84.5% sensitivity and 86.9% specificity	Other algorithms with lower performance
Solakhan et al. [32]	Prediction of SSP	Case-control	Accuracy of 92.8%	Other algorithms with lower performance
Park et al. [33]	Prediction of SSP	Case-control	AUCs of 0.859 (stones of <5 mm) and 0.881 (stones of 5–10 mm)	AUC of 0.847 (stones of <5 mm) and 0.817 (stones of 5 mm–10 mm)
Poulakis et al. [34]	Prediction of lower pole clearance after ESWL	Case-control	Accuracy of 92%	No comparator
Gomha et al. [35]	Prediction of clearance after ESWL for ureteral stones	Case-control	Accuracy of 77.7%	Accuracy of 93.2%
Moorthy and Krishnan [36]	Prediction of renal stone fragmentation after ESWL	Case-control	Accuracy of 90%	No comparator
Choo et al. [37]	Prediction of clearance after ESWL for ureteral stones	Case-control	Accuracy of 92.29%	No comparator
Seckiner et al. [38]	Prediction of clearance after ESWL for renal stones	Case-control	Accuracy of 88.70%	No comparator
Mannil et al. [39]	Prediction of renal stones fragmentation after ESWL	Case-control	AUC of 0.85	Other algorithms with lower performance
Yang et al. [40]	Prediction of clearance after ESWL for renal or upper ureter stones	Case-control	AUC of 0.85 for stone-free status in an interval of 4 weeks; AUC of 0.78 for stone-free status after single session ESWL	Other algorithms with similar performance
Tsitsiflis et al. [41]	Prediction of complications after ESWL for renal or ureteral stones	Case-control	Accuracy of 81.43%	No comparator
Handa et al. [42]	Quantification of ESWL-induced renal injury by MRI	Experimental	Strong correlation between model prediction and morphology ($r=0.9691$)	No comparator
Aminsharifi et al. [43]	Prediction of multiple outcomes after PCNL	Case-control	Accuracy of 91.8%, 83% regarding stone clearance and need for blood transfusion; AUC of 0.915 for stone clearance	AUCs of 0.615 and 0.621 for stone clearance according to GSS and CROES nomograms
Shabaniyan et al. [44]	Prediction of multiple outcomes after PCNL	Case-control	Accuracy of 94.8% in prediction of the procedures' outcome, 85.2% accuracy in predicting the need for stent placement and 95% in predicting blood transfusion	Multiple decision support systems achieving higher performances in different parameters
Aminsharifi et al. [45]	Prediction of multiple outcomes after PCNL	Case-control	Accuracy of 82.8%, 92.5%–98.2%, 81.1%, and 85.8% for stone clearance, need for a second procedure, stent insertion by urine extravasation, and blood transfusion	No comparator
Geraghty et al. [46]	Prediction of multiple outcomes after PCNL	Case-control	Multiple classification models tested, highest accuracy of 99% and AUCs of 0.99–1.00 achieved for need for transfusion and infectious complications	No comparator

(*continued on next page*)

Table 2 (*continued*)

Study	Objective	Study design	AI-based outcome	Comparator arm outcome
Zhao et al. [47]	Prediction of stone clearance after PCNL	Case-control	AUC of 0.879	AUC of 0.800 for GSS; AUC of 0.844 for S.T.O.N.E. score
Chen et al. [48]	Prediction of sepsis after fURS or PCNL for proximal ureteral stones	Case-control	AUC of 0.874 for DNN model	AUC of 0.783 for LASSO model

AI, artificial intelligence; AUC, area under the curve; CROES, Clinical Research Office of the Endourological Society; DNN, deep neural network; ESWL, extracorporeal shockwave lithotripsy; fURS, flexible uretero-renoscopy; GSS, Guy's stone score; LASSO, least absolute shrinkage and selection operator; PCNL, percutaneous nephrolithotomy; SSP, spontaneous stone passage; S.T.O.N.E., stone size (S), tract length (T), obstruction (O), number of involved calices (N), and essence or stone density (E); MRI, magnetic resonance imaging.

They used 16 features for input values in the NN and the same parameters for the regression analysis. In this study, the features that proved to be significant for the stone-free outcome were stone size, number of ESWL sessions, stone location, infundibulopelvic angle, and skin-to-stone distance (SSD). The prediction accuracy rates for the NN for the three samples of the training group, validation group, and test group were 99.25%, 85.48%, and 88.70%, respectively.

In the same scope of predicting the successful SFS after ESWL, 3D texture features were tested using five ML algorithms [39]. Texture analysis (TA) identifies unique, measurable variations in stone features that a solely visual investigation could miss. Mannil et al. [39] in a preliminary study provided two crucial conclusions. First, 3D-TA offers additional reliable data on the efficacy of ESWL, in addition to previously known clinical factors, such as body mass index (BMI), SSD, and stone size. Second, mean CT attenuation values for kidney stones did not predict SWL success. The discriminatory accuracy increased further when 3D-TA characteristics were combined with clinical factors, with an AUC of 0.85 for 3D-TA features and SSD, 0.80 for 3D-TA features and BMI, and 0.81 for 3D-TA and stone size.

Yang et al. [40] trained a ML model trained with DT algorithms in a total of 358 non-contrast CT scans, from patients who underwent ESWL for renal and upper ureter stones. They used three DT-based algorithms, the random forest, the extreme gradient boosting trees (XGBoost), and the light gradient boosting method (LightGBM) for the prediction of SFS and one-session success. In this study, the DT-based algorithms were chosen due to their better performance in prediction in small sample training datasets. The authors claimed that median stone density and stone volume were the most important variables for the procedure outcome. The proposed models predicted the SFR with a maximum accuracy of 87.9% (LightGBM) and the one-session success with an accuracy of more than 77%.

Tsitsiflis et al. [41] created an ANN for the prediction and the procedure outcome when several ESWL features are evaluated. They collected medical data from 716 patients, of which 549 were used for training purposes and 167 for testing purpose. In the parameter importance analysis, diabetes and hydronephrosis had a significant predictive value for complication occurrence, while for the procedure outcome, the stone location was of the highest importance. The ANN model achieved a performance rate of 98.72% at the end of the training set, with a PPV of 83.82% and an accuracy of 81.43% in complication prediction.

In the scope of predicting the hemorrhagic injury after ESWL, Handa et al. [42] evaluated the injury lesion volumes in *ex vivo* kidneys by using a magnetic resonance imaging and multi-spectral neural network model. Quantifying renal damage lesion volumes as a result of ESWL was claimed to be quick and accurate with the use of magnetic resonance imaging and the multi-spectral neural network classifier.

3.3. Prediction of outcome of endourological procedures

Prediction of the SFR and other postoperative parameters of patients with kidney stones, planned for percutaneous nephrolithotomy (PCNL), is still under debate. Throughout the years, several scores have been developed in order to categorize renal calculi and estimate SFR. On the other hand, AI and its subtypes seem to present with comparable, or even superior results when it comes to postoperative outcome prediction (Table 2).

Aminsharifi et al. [43] developed an ANN model as a predictor of four outcomes after PCNL: SFR, need for repeat lithotripsy, need for ureteral stent insertion due to urine leakage, and need for blood transfusion. Sensitivity rates vary from 81% to 92%, with stone burden and morphometry being the most significant preoperative characteristics that affect all four outcomes. Therefore, the ANN model serves as a reliable tool for predicting certain postoperative features. Shabaniyan et al. [44] investigated several methods, of which the SVM is a frequently implemented ML model for regression and classification along with the sequential forward selection, which adds features to a list of empty candidates until a maximum number of features that do not reduce the criterion, and Fischer discriminant analysis is the optimal combination to improve system accuracy in predicting the four aforementioned outcomes. Aminsharifi et al. [45] also developed an SVM model and compared its results with respective ones from the Guy's stone score (GSS) and Clinical Research Office of the Endourological Society models. The AUC of the ML software (0.915) was significantly larger than the AUC of both GSS (0.615) and Clinical Research Office of the Endourological Society (0.621) nomograms. Furthermore, the ML model provided a sensitivity

and an accuracy rate of more than 82% and 80%, respectively, in successfully predicting all postoperative examined outcomes. Geraghty et al. [46] developed five ML models and DL models to assess the correlation between 43 patient preoperative characteristics and 11 outcomes. Of those outcomes, post-PCNL infection and the need for transfusion were successfully predicted in most of the patients. Each of the ML and DL implemented models achieved AUCs for these two outcomes larger than 0.90 and 0.77, respectively. Nonetheless, the authors underlined the inability of these models to safely process incomplete datasets and very rare outcomes, leading to poor predictions. They concluded that imaging findings may provide AI algorithms with data invaluable for the prediction of postoperative outcomes.

Zhao et al. [47] compared four ML models, including SVM, with the GSS and stone size (S), tract length (T), obstruction (O), number of involved calices (N), and essence or stone density (E) (S.T.O.N.E.) nephrolithometry score system. Although all four models successfully predicted SFRs after PCNL, they provided non-inferior results compared to GSS and S.T.O.N.E. nephrolithometry score AUCs for the ML models ranged from 0.803 to 0.879, while the AUCs of the GSS and S.T.O.N.E. nephrolithometry score were 0.800 and 0.844, respectively. Sensitivity values were also comparable. ML models were only superior to the S.T.O.N.E. nephrolithometry score system in terms of accuracy (minimum ML accuracy: 80.3%; S.T.O.N.E. accuracy: 78.8%). These findings come in line with previous author statements which summarized that there is a need for a nationwide collaboration among physicians, statisticians, and computer experts in order to optimize the use of AI in favor of its individual, especially the patients.

Chen et al. [48] used a DNN to estimate and predict the risk factors for sepsis after flexible ureteroscopic lithotripsy or PCNL. Medical data from 847 patients who met the inclusion criteria were reviewed. Sepsis incidence was estimated at 5.9% after flexible ureteroscopic lithotripsy or PCNL. Preoperative CT scans were performed on all patients. The least absolute shrinkage and selection operator model and DNN model were compared based on preoperative sepsis prediction in ureteral calculi patients. Although the least absolute shrinkage and selection operator model successfully predicted 26 variables, the DNN model outperformed it, with AUCs of 0.920 and 0.874 accounting for the internal and external validation, respectively.

4. AI for the optimization of the operative procedure

Results of the operative management of stone disease depend on a numerous set of factors, which frequently interact with each other, so that defining the optimal constellation of these factors remains a difficult task to perform. Especially the optimal setting of intraoperative factors comprises an essential step for the uncomplicated and successful completion of the respective procedures. AI has shown promising results in studying the complicated intraoperative environment to optimize and guide surgeons' preferences and strategy (Table 3).

4.1. Optimization of ESWL procedure

In 2003, Hamid et al. [49] collected the data from 82 patients with renal stones treated by ESWL. These data included various parameters, such as factors of urine chemistry, radiological features of stones, and ESWL settings. An ANN algorithm was constructed and trained through the introduction of the above data to optimize the prediction of the ESWL shockwave number needed for optimum fragmentation. The ANN algorithm identified all patients, who needed a number of ESWL shockwaves beyond standard protocol. Similarly, Goyal et al. [50] trained and tested an ANN algorithm through the data of 196 patients (training set) and 80 patients (validation set) with kidney stones treated by ESWL. The data included parameters, such as stone size and burden, number of ESWL sessions, and factors of urine chemistry. Shockwave power predicted by ANN algorithm and number for optimum stone fragmentation were very strongly correlated with the observed shockwave power and number needed during the ESWL procedures ($r^2=0.8343$, $r^2=0.9329$, respectively).

Mannil et al. [51] examined by CT a set of 34 urinary stones under *in vitro* settings in order to associate the stone features with the number of shockwaves needed for successful fragmentation. ML-based model managed to discriminate with a sensitivity of 94% and a specificity of 59% the subset of stones, which were fragmented with less than 72 shockwaves.

In a recent report, a DL model was constructed and trained to produce personalized ESWL protocols, which included the ESWL settings (power lever, shockwave rate, and total number) for each of the steps of every ESWL session [52]. The data used for the model were extracted from the best practices of ESWL treatments recorded in the International Stone Registry. The comparison of the DL model with standard methods and models for defining ESWL settings for each ESWL step showed the superiority of the DL model.

Muller et al. [53] trained a CNN algorithm to recognize a kidney stone in US images taken during the ESWL procedures of 11 patients. Subsequently, they investigated the effect of directing ESWL shockwaves through optical recognition of stone by the above algorithm. The study concluded that an AI-controlled ESWL procedure would demonstrate an increased stone hit rate of 75.3% and a mishit reduction by 67.1% compared to the operator-directed ESWL procedure. An increase in patient number for training the above algorithm would further improve the accuracy in directing ESWL shockwaves.

4.2. Optimization of endourological procedures

Regarding the PCNL procedure, Taguchi et al. [54] evaluated the feasibility of robot-assisted fluoroscopy guide puncture and compared the above method with the operator-dependent US-guided puncture. The system used for robot-assisted puncture was based on computer vision software and AI to analyze fluoroscopic images and direct needle puncture accordingly. The success rate for robotic-assisted puncture was 100%, which was superior to the

Table 3 Summary of studies regarding the contribution of AI in the optimization of the operative procedure.

Study	Objective	Study design	AI-based outcome	Comparator arm outcome
Hamid et al. [49]	Optimization of ESWL protocol	Cross-sectional	Accuracy of 75% for predicting shockwave number and 100% for predicting patients needing shockwave number beyond protocol	No comparator
Goyal et al. [50]	Optimization of ESWL protocol	Cross-sectional	Correlation coefficients for power level and shockwave number of 0.8343 and 0.9329, respectively	Correlation coefficients for power level and shockwave number 0.0195 and 0.5726, respectively
Mannil et al. [51]	Optimization of ESWL protocol	Experimental	AUC of 0.838 in the prediction of fragmentation with less than 72 shockwaves	Other multivariate models with lower performance
Chen et al. [52]	Optimization of ESWL protocol	Cross-sectional	Prediction accuracy values for power level, shockwave rate of 98.8%, 98.1%, respectively	Other multivariate models with lower performance
Muller et al. [53]	Optimization of ESWL protocol	Cross-sectional	Shockwave hit rate of 75.3%	Shockwave hit rate of 55.2%
Taguchi et al. [54]	Optimization of PCNL puncture	Experimental	Puncture success rate of 100%; puncture time of 35 s	Puncture success rate of 70.6%; puncture time of 46 s
Wang et al. [55]	Optimization of PCNL puncture	Experimental	Average recognition precision of 79% (SE: 4%) for cortex, 85% (SE: 6%) for medulla, and 91% (SE: 5%) for calyx	No comparator
Li et al. [56]	Optimization of PCNL puncture	Cross-sectional	ANN model achieved a better localization and puncture method selection compared to the MVRA model and the surgeon's experience	No comparator
Jeong et al. [57]	Optimization of RIRS safety profile	Experimental	Recognition of tissue exposure to laser energy with accuracy of 95% and latency time of 0.5 s	No comparator

AI, artificial intelligence; ANN, artificial neural network; ESWL, extracorporeal shockwave lithotripsy; MVRA, multiple variable regression analysis; PCNL, percutaneous nephrolithotomy; RIRS, retrograde intrarenal surgery; SE, standard error.

respective rate of 70.6% for operator-dependent puncture. Additional advantages of the method were the shortening of needle puncture time and total procedure duration.

Another report proposed an alternative method for guiding PCNL puncture based on AI and optical coherence tomography (OCT), a novel imaging modality, which can depict subsurface tissue at depth of several millimeters [55]. OCT can be applied through mini-probes, which can be adapted to the puncture needle of PCNL. In this study, a DL algorithm was constructed and trained to recognize the unique OCT patterns of renal cortex, medulla, and calyx with high accuracy (precision of 79%±4%, 85%±6%, 91%±5% for mean±standard error for cortex, medulla, and calyx, respectively). The above method proved to be effective in recognizing the type of tissue ahead of the PCNL needle and in guiding the puncture accordingly.

In 2016, Li et al. [56] performed a discrimination analysis to define the ideal imaging modality for performing PCNL puncture. An ANN algorithm was constructed and trained with the prospectively collected data from patients who had undergone PCNL surgery. The trained algorithm showed that the combination of X-ray with US guidance was beneficial in the case of complex or small stones, while the combination was not significantly better compared to one of the above methods for treating large or simple stones. Recently, Jeong et al. [57] developed a monitoring system for reducing tissue exposure to laser energy during retrograde intrarenal surgery, based on measuring the specific shockwave form produced during the interaction of laser with a soft (tissue) or hard (stone) material and AI. Shockwave measurement was performed through an accelerometer adapted to the ureteroscope and the data were produced in a simulated surgical environment. The above data were introduced in a ML model in order to enhance its predicting ability for tissue exposure and subsequent damage by laser energy [57]. The trained model was capable to warn surgeon of damage-inducing tissue exposure to laser energy with a latency time of 0.5 s and an accuracy of 95%. Further optimization of the model would shorten the latency time and increase its accuracy. Moreover, the same methodology seems applicable to additional operative procedures of endourology.

5. AI for the elucidation of stone disease chemistry and composition

Prediction of primary stone disease or recurrent development of urolithiasis is considered of utmost importance in the attempt to reduce its incidence and the subsequent consequences on renal function. AI has been proposed as one of the more effective methods to elucidate the complicated interaction between various clinical and biochemical factors, whose aberration is associated with predisposition for developing stone disease (Table 4).

5.1. Association of stone disease risk with blood, urine chemistry, and other clinical factors

According to a study comparing the data of 119 male patients with 96 controls by using ANN, calcium oxalate (CaOx) supersaturation and urea concentration in urine were the strongest predictors of developing calcium stones [58]. Another report on the predisposition for calcium stones compared the clinical and biochemical data of 119 male and 59 female patients with respective healthy controls by using ANN and concluded that the most important predisposing factors were CaOx supersaturation and 24 h-excreted urea [59]. Interestingly, the study emphasized that these predisposing factors were common for male patients regardless of family history and for female patients with a family history of stone disease [59].

Ensemble learning was applied in a recent study to extract the relevant features across a data set with 42 clinical and biochemical features and to construct a predicting model for developing stone disease with an accuracy of 97.1% [60]. Especially for kidney stones larger than 20 mm, another model based on ML algorithms defined hypertension, older age, decreased CaOx supersaturation, and a higher percentage of protein in stone composition as the strongest predictors for developing kidney stones of the above size category [61]. Importantly, in this report, AI-based prediction was adequate but so accurate as the result of LR. Aberrations of 24 h-excreted urine analytes are recognized as a strong predictor of stone disease, but the underlying clinical factors are yet mostly unknown.

An ensemble model, based on data from electronic health records, achieved to predict these aberrations with high accuracy [62]. The clinical factors, which were more strongly associated with these aberrations were BMI, age, and gender. Regarding the risk of recurrent stone disease, an ANN approach was applied on the clinical and biochemical data of 80 patients. Among the available parameters, natrium, kalium in serum and natrium, phosphorus, and CaOx in urine were included in a model, which predicted approximately 90% of the recurrence events [63].

Genetic factors comprise another recognized predisposing factor for stone disease. According to a study comparing genetic and environmental data of a patient group with respective controls, the application of ANN showed the highest discriminating ability among the comparing groups, since it classified successfully 89% of the participants [64].

Using microscopy images and an identification procedure based on a convolution neural network, another study reported an improved rate (74%) of recognition of CaOx crystals in urine sediment, which are considered as a predictive factor for developing CaOx stones, compared to the standard identification procedure [65].

Recently, another report, based on electron microscopy images and ML-based analysis, evaluated crystallization dynamics of CaOx under the effect of candidate crystallization inhibitors and revealed the effectiveness of myoinositol hexakisphosphate analogues, which can be applied for the prevention of CaOx nephropathies or stones [66].

5.2. Recognition of stone composition by CT

Besides the increased stone detection rate, radiographic features in CT examination represent a valuable source for the extrapolation of information relating to stone composition. This inherent advantage of CT examination can be further optimized through the application of AI.

According to a study, 32 *ex vivo* kidney stones with known composition were examined through a CT protocol, which produced 52 variables for each stone. After training five multiparametric algorithms, including ANN and SVM algorithms with data of these variables, distinguishing uric acid (UA) from non-UA stones achieved a maximum accuracy of 100%, while distinguishing between the non-UA subtypes was 75% accurate [67].

The same researchers applied another protocol of rapid voltage switching, single source, dual-energy CT on 38 *ex vivo* renal stones, which produced data of 17 variables for each stone. By using the same multiparametric algorithms, distinguishing between UA and non-UA stones was 100% accurate, while the best algorithm for distinguishing among non-UA subtypes was 88% accurate [68].

A third study on distinguishing between UA and non-UA stones was based on applying CT TA in order to train SVM classifiers and construct a respective model, which achieved an AUC of 0.965 (sensitivity of 94.4%, specificity of 93.7%) in the correct characterization of UA stones [69].

Große Hokamp et al. [70] examined 200 kidney stones of known composition with a spectral detector CT scanner, applying a number of different protocols. The produced data were used to train and validate a NN algorithm to distinguish among different stone compositions. The final model achieved accuracy rates of 91.1% on a per-voxel basis, and 87.1%—90.4% on independently tested acquisitions. Interestingly, the model was accurate in predicting the main component even in compound stones [70].

A recent publication by Tang et al. [71] reported the results on distinguishing between calcium oxalate monohydrate (COM) and non-COM stones by radiographic features of unenhanced and analysis through an AI model. The initial set of 1218 radiographic features was diminished to eight significant features, and the final model based on these features achieved an accuracy of 88.3% (AUC: 0.933; sensitivity: 90.5%; specificity: 84.3%). The above results were based on the *in vivo* radiographical characterization of stones, which makes the results applicable for the preoperative evaluation of patients [71].

Table 4 Summary of studies on the contribution of AI in the elucidation of stone disease chemistry and composition.

Study	Objective	Study design	AI-based outcome	Comparator arm outcome
Dussol et al. [58]	Risk factors for calcium stones	Case-control	Classification accuracy between stone formers and controls: 74.4%	75.8%
Dussol et al. [59]	Risk factors for calcium stones	Case-control	CaOx supersaturation and 24 h-urea for all men and women with a family history	No comparator
Kazemi and Mirroshandel [60]	Risk of nephrolithiasis	Cohort	Accuracy of 97.1%	Other classifiers with lower accuracy
Chen et al. [61]	Risk of forming renal stones of >2 cm	Cohort	AUC of 0.69	AUC of 0.74
Kavoussi et al. [62]	Prediction of 24 h urine abnormalities relevant for stone disease	Cohort	Higher accuracy in prediction of urine volume, uric acid, and natrium abnormalities	Higher accuracy in prediction of pH and citrate abnormalities
Caudarella et al. [63]	Risk of stone disease recurrence	Case-control	Accuracy of 88.8%	Accuracy of 67.5%
Chiang et al. [64]	Risk for stone disease	Case-control	Accuracy of 89%	Accuracy of 74%
Xiang et al. [65]	Identification of CaOx crystallization in urine sediment	Cross-sectional	Accuracy of 74%	Accuracy of 74%
Kletzmayr et al. [66]	Recognition of crystallization inhibition	Experimental	IP6 analogues inhibit effectively CaOx crystallization	No comparator
Kriegshauser et al. [67]	Stone composition by CT	Cross-sectional	Accuracy of 97% (UA instead of non-UA stones) and 72% among non-UA stones	Other multivariate models with lower performance
Kriegshauser et al. [68]	Stone composition by CT	Cross-sectional	Accuracy of 100% (UA instead of non-UA stones) and 88% among non-UA stones	Other multivariate models with lower performance
Zhang et al. [69]	Stone composition by CT	Cross-sectional	AUC of 0.965 (SD: 0.029) for UA instead of non-UA stones	Sensitivity of 94.4% and specificity of 93.7% for model using CT TA
Große Hokamp et al. [70]	Stone composition by CT	Cross-sectional	Accuracy of 91.1% on a per-voxel basis; accuracy of 87.1%–90.4% on independently tested acquisitions	No comparator
Tang et al. [71]	Stone composition by CT	Cross-sectional	Accuracy of 88.3% for COM instead of non-COM stones (AUC = 0.933)	No comparator
Black et al. [72]	Stone composition by visual image	Cross-sectional	Prediction precision for each stone composition from 71.43% (struvite) to 95% (COM stones)	No comparator
Lopez et al. [73]	Stone composition by visual image	Cross-sectional	Precision of 93%–98%, depending on stone type	Other multivariate models with lower performance
El Beze et al. [74]	Stone composition by visual image	Cross-sectional	PPV of 96%–99%, depending on stone type	PPV of 88%–99%, depending on stone type
Ochoa-Ruiz et al. [75]	Stone composition by visual image	Cross-sectional	Overall precision of 97%	Overall precision of 96%
Mendez-Ruiz et al. [76]	Stone composition by visual image	Cross-sectional	Overall accuracy of 74.38% and 88.52%, depending on the image capturing method	Overall accuracy of 45%
Kim et al. [77]	Stone composition by visual image	Cross-sectional	AUC of 0.98–1.00, depending on stone type	Other multivariate models with lower performance

(*continued on next page*)

Table 4 (*continued*)

Study	Objective	Study design	AI-based outcome	Comparator arm outcome
Fitri et al. [78]	Stone composition by microtomography	Cross-sectional	Overall accuracy of 99.59%	No comparator
Saçlı et al. [79]	Stone composition by dielectric properties	Cross-sectional	Overall accuracy of 98.17%	No comparator
Cui et al. [80]	Stone composition by Raman spectroscopy	Cross-sectional	Overall accuracy of 96.3%	No comparator
Onal and Tekgul [81]	Stone composition by smartphone microscopy	Cross-sectional	Overall accuracy of 88%	No comparator

AI, artificial intelligence; AUC, area under the curve; CaOx, calcium oxalate; CT, computed tomography; TA, texture analysis; COM, calcium oxalate monohydrate; IP6, myoinositol hexakisphosphate; PPV, positive predictive value; UA, uric acid; SD, standard deviation.

5.3. Recognition of stone composition by endoscopic images

Intraoperative images of stones can provide important visual information to the experienced endourologist relating to stone composition, which in turn can affect various surgery parameters, *e.g.*, the settings of the energy generator used for the lithotripsy. This intraoperative stone characterization may become more precise by the application of visual processing through AI methods. A study on the composition prediction based on macroscopic stone appearance reported an overall prediction rate of 85% for every composition class by training a CNN algorithm on a total of 63 kidney stones. The image of each stone sample of the above set was captured with a digital camera and depicted the surface and the inner core of each stone [72].

Lopez et al. [73] compared several classification methods for intraoperative recognition of stone type, based on 90 fragment surface images, 87 fragment cross-section images, and respective algorithms. The best classification method was based on a deep CNN algorithm and reached a precision of 98% for the four stone classes included. Another study compared the classification performance of simple algorithms, based on texture and color criteria, with DL-based methods by using a large number of stone surface and stone section images, captured through ureteroscopy. The above images represented six chemical classes of stones, and DL-based methods demonstrated a superior predicting ability for the four of the above classes, namely whewellite, UA, struvite, and cysteine [74].

Ochoa-Ruiz et al. [75] published a report aiming to examine the feasibility of predicting stone composition by capturing respective stone images during ureteroscopic procedures. A set of 94 surface images and 87 section images of stones with known composition was used to compare the predicting performance of six shallow ML methods and three DL algorithms. DL algorithms achieved superior performance, reaching a precision rate of 97% and demonstrating a slight advantage over the methods based on color and texture features.

According to a recent study, classical ML failed to perform adequately when tested in a different dataset than the one used for constructing and training the respective model. To form an algorithm with generalized efficacy in the classification of stone composition based on stone images, the author proposed the use of meta-learning, a subfield of ML, which can contribute to a significantly improved discriminating performance compared with DL [76]. Kim et al. [77] used the largest set of stone images captured by a digital camera, depicting 1332 stones of 31 different chemical compositions. Among them, images of 965 stones, representing the four more frequent classes, were used to construct and train a number of models, based on CNNs. The best-performing model demonstrated high accuracy for all of the included chemical classes, with an AUC value of at least 0.97 [77].

5.4. Recognition of stone composition by other innovative methods

Several studies reported significant results on discriminating stone composition by introducing the data produced through various innovative methods in AI-based models. These innovative methods were focusing mostly on the physical and chemical properties of urinary stones.

Fitri et al. [78] produced a set of 2430 images through microtomography of 30 urinary stones from different patients. A CNN algorithm was constructed to classify the above stone set into three main compositions, namely calcium, UA, and mixture stones. The final model reached an accuracy of 98.52% in classifying into the above classes. Saçlı et al. [79] exploited the dielectric properties of 105 kidney stones of known composition to build a dataset with the respective parameters. The above dataset was used to construct and train a classification ML-based model, which reached an accuracy of 98.17%. According to the authors, the method combines the high accuracy of the reference methods in the analysis of urinary stones with the reduced cost and the rapid processing procedure. Raman spectroscopy represents an innovative method for determining

stone composition by examining the spectral characteristics of the different components. According to a study, the respective measurements of a set of 135 kidney stones by Raman spectroscopy were processed through ML techniques to construct a number of models with high accuracy (96.3%) in classifying the stones into the four main chemical categories of the study [80]. The above procedure was considered beneficial compared to existing methods in terms of simplicity and cost efficiency. Onal and Tekgul [81] reported an innovative approach for determining stone composition by combining microscopy images of stones with a CNN model. A set of 37 surgically extracted kidney stones was examined with a smartphone microscope and microscopic images were captured from six different locations of each stone. A total of 222 images were used to train the classification model, which achieved an accuracy of 88% in predicting the stone composition among the four main classes of the study, namely CaOx, UA, cysteine, and struvite.

6. Conclusion

The above analysis of the available reports demonstrates that the application of AI in stone disease comprises a rapidly evolving management option, which still runs in the test phase. As expected, the evolvement of this application follows the progress of AI development through hardware and software innovations. Indeed, the number of available reports published in the decade from 2000 to 2010 was only nine, while in the next decade this number increased to 35 publications. For the current decade, 25 reports were published already in 2021 and 2022, which represents an acceleration in the respective research area. Regarding the coverage of the stone disease subfields, the majority of the main issues in stone disease diagnosis and management have already been targeted for further optimization by the application of AI. In our opinion, the further elucidation of the dynamic interplay among environmental, metabolic, and genetic factors in defining the individual risk for stone disease represents the top priority. The complicated data constellation of the above factors could be deployed to develop AI algorithms for the recognition of high-risk stone formers and the planning of targeted pharmaceutical prevention of stone recurrence. Another research gap is the further optimization of lithotripsy procedures, by applying AI algorithms to adjust intraoperative factors, such as the parameters of the laser generators, or the equipment selection depending on patient anatomic characteristics or stone physical characteristics.

Regarding the limitations of the available literature on the application of AI in stone disease, most of the studies only reported the results of the AI algorithm in a diagnostic or therapeutic procedure without comparison with the current standards in the respective subfields. Moreover, the results of the studies were mostly validated in the same patient cohorts, which limits their generalizability. The evidence level of the included studies does not exceed level 3 and prospective validations, or at least external validations are needed to test the applicability of AI in the real-life management of the stone disease. From the total of the included studies in the current review, it is obvious that only the diagnostical applications, which mostly are related to radiology are close to implementation in urological practice, while in many reports, the authors claimed that the function of their proposed model could be further optimized after the introduction of more data. Lastly, the heterogeneity of study design and outcome reporting makes the results of the included studies unsuitable for quantitative analysis and data pooling.

AI provides a learning capacity to computing systems, which creates new perspectives not only in several scientific domains but also in everyday life. In medicine, AI is considered an important decision-supporting factor with high potential in specific areas of health care. These areas include the clinical field, which embraces the diagnostic, prognostic, and predictive role of AI, the pharmaceutical field, which relates to the targeted discovery of drugs and the in-silico clinical trials, and the public health, where AI is applied in epidemic outbreak prediction and the configuration of public health policy [82]. Moreover, AI is considered one of the pillars for the wide application of precision medicine, since the continuously growing biomedical data, processed by the constantly evolving AI systems, are expected to result in new disease taxonomies, based on predictors of various nature, such as genomic, environmental [83]. The above effect of AI configures the future of urology, a specialty that is technology-driven and at the forefront of innovation. Data related to various urologic conditions are produced at an exponentially increasing rate, forming an amount described by the term "Big Data", which represents the data volume that is not manageable by traditional computing methods but only by the AI methods [84]. The combination of "Big Data" and AI is expected to create a different frame of urological practice, where the computing systems will be used to inform, enhance, and complement the clinicians' decisions, while the urological training will confront new challenges since the urologist will have to be familiar with the application of AI in his practice [85].

Regarding the current state of AI introduction in medicine, a review identified 222 AI-based medical devices approved in the USA, while the respective number for Europe was 240 [86]. According to the authors of the above report, there is an enormous increase in the AI-based device availability in recent years, since the numbers of approved devices in the years 2015, 2017, and 2019 were 9, 32, 77 for the USA and 13, 26, 100 for Europe, respectively. Radiology comprises the leading medical specialty for AI-based devices ($n=129$), followed by cardiology ($n=40$), and neurology ($n=21$) [86]. No urologic device in the approved status was reported in this study.

The wide application of AI in urology assumes not only the further improvement of AI methods but also the familiarity of the clinicians with AI, which depends on the trustworthiness, explainability, usability, and transparency of AI methods [87]. Moreover, a regulatory frame for the application of AI is needed from the public organizations, who are responsible for the legal regulations regarding health care provision. In this direction, Food and Drug Administration and European Commission have already published the legal conditions and prerequisites to protect human fundamental rights and guide the use of medical AI accordingly [88,89]. From the view of the public, a study

found that the participants had equal trust in AI versus clinicians regarding the diagnostic procedure, while the 94% of them would pay for a review of an imaging diagnostic procedure by an AI method [90]. On the contrary, the majority of the participants would feel uncomfortable with the perspective of automated robotic surgery. For the application of AI methods through "Big Data", large global corporations invest enormous amounts of money, while, at the same time, public funding moves still at low levels [91]. Since patients are interested in AI methods, at least for diagnostic purposes, and have the greatest trust in health providers, it is the responsibility of the latest and of the public regulatory authority to offer trustworthy and meaningful medical AI applications.

Author contributions

Study concept and design: Anastasios Anastasiadis, Jean de la Rosette, Georgios Dimitriadis, Ioannis Vakalopoulos, Antonios Koudonas, Georgios Langas, Stavros Tsiakaras.
Data acquisition: Anastasios Anastasiadis, Jean de la Rosette, Georgios Dimitriadis, Antonios Koudonas, Georgios Langas, Stavros Tsiakaras, Dimitrios Memmos, Ioannis Mykoniatis, Evangelos N. Symeonidis, Dimitrios Tsiptsios, Eliophotos Savvides.
Data analysis: Anastasios Anastasiadis, Jean de la Rosette, Georgios Dimitriadis, Antonios Koudonas, Georgios Langas, Stavros Tsiakaras, Dimitrios Memmos, Ioannis Mykoniatis, Evangelos N. Symeonidis, Dimitrios Tsiptsios, Eliophotos Savvides.
Drafting of manuscript: Anastasios Anastasiadis, Jean de la Rosette, Georgios Dimitriadis, Ioannis Vakalopoulos, Antonios Koudonas, Georgios Langas, Stavros Tsiakaras, Ioannis Mykoniatis.
Critical revision of the manuscript: Anastasios Anastasiadis, Jean de la Rosette, Georgios Dimitriadis, Ioannis Vakalopoulos, Antonios Koudonas, Georgios Langas, Stavros Tsiakaras.

Conflicts of interest

The authors declare no conflict of interest.

References

[1] Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Fam Med Prim Care 2019;8:2328—31.
[2] Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol 2019;28:73—81.
[3] Kueper JK. Primer for artificial intelligence in primary care. Can Fam Physician 2021;67:889—93.
[4] Schmidhuber J. Deep learning in neural networks: an overview. Neural Network 2015;61:85—117.
[5] Frankish K, Ramsey WM. The Cambridge handbook of artificial intelligence. Cambridge: Cambridge University Press; 2014. p. 151—66.
[6] Rowe M. An introduction to machine learning for clinicians. Acad Med 2019;94:1433—6.
[7] Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol 2020;9:14. https://doi.org/10.1167/tvst.9.2.14.
[8] Rabhi S, Jakubowicz J, Metzger MH. Deep learning versus conventional machine learning for detection of healthcare-associated infections in French clinical narratives. Methods Inf Med 2019;58:31—41.
[9] Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging 2018;9:611—29.
[10] Jin KH, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. IEEE Trans Image Process 2017;26:4509—22.
[11] Huang JA, Hartanti IR, Colin MN, Pitaloka DA. Telemedicine and artificial intelligence to support self-isolation of COVID-19 patients: recent updates and challenges. Digital Health 2022;8. 20552076221100634. https://doi.org/10.1177/20552076221100634.
[12] Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017;69:S36—40. https://doi.org/10.1016/j.metabol.2017.01.011.
[13] Li D, Xiao C, Liu Y, Chen Z, Hassan H, Su L, et al. Deep Segmentation Networks for segmenting kidneys and detecting kidney stones in unenhanced abdominal CT images. Diagnostics 2022;12:1788. https://doi.org/10.3390/diagnostics12081788.
[14] Parakh A, Lee H, Lee JH, Eisner BH, Sahani DV, Do S. Urinary stone detection on CT images using deep convolutional neural networks: evaluation of model performance and generalization. Radiol Artif Intell 2019;1:e180066. https://doi.org/10.1148/ryai.2019180066.
[15] Längkvist M, Jendeberg J, Thunberg P, Loutfi A, Lidén M. Computer aided detection of ureteral stones in thin slice computed tomography volumes using Convolutional Neural Networks. Comput Biol Med 2018;97:153—60.
[16] Caglayan A, Horsanali MO, Kocadurdu K, Ismailoglu E, Guneyli S. Deep learning model-assisted detection of kidney stones on computed tomography. Int Braz J Urol 2022;48:830—9.
[17] Jendeberg J, Thunberg P, Lidén M. Differentiation of distal ureteral stones and pelvic phleboliths using a convolutional neural network. Urolithiasis 2021;49:41—9.
[18] De Perrot T, Hofmeister J, Burgermeister S, Martin SP, Feutry G, Klein J, et al. Differentiating kidney stones from phleboliths in unenhanced low-dose computed tomography using radiomics and machine learning. Eur Radiol 2019;29:4776—82.
[19] Chak P, Navadiya P, Parikh B, Pathak KC. Neural network and svm based kidney stone based medical image classification. In: International conference on computer vision and image processing. Singapore: Springer; 2019. p158—73.
[20] G P VP, Reddy KVS, Kiruthik AM, ArunNehru J. Prediction of kidney stones using machine learning. Int J Res Appl Sci Eng Technol 2022;10:1037—44.
[21] Elton DC, Turkbey EB, Pickhardt PJ, Summers RM. A deep learning system for automated kidney stone detection and volumetric segmentation on noncontrast CT scans. Med Phys 2022;49:2545—54.
[22] Krishna KD, Akkala V, Bharath R, Rajalakshmi P, Mohammed A, Merchant S, et al. Computer aided abnormality detection for kidney on FPGA based IoT enabled portable ultrasound imaging system. IRBM 2016;37:189—97.
[23] Balamurugan SP, Arumugam G. A novel method for predicting kidney diseases using optimal artificial neural network in ultrasound images. IJIE 2020;7:37—55.
[24] Selvarani S, Rajendran P. Detection of renal calculi in ultrasound image using meta-heuristic support vector machine. J Med Syst 2019;43:1—9.
[25] Viswanath K, Anilkumar B, Gunasundari R. Design of deep learning reaction—diffusion level set segmentation approach

for health care related to automatic kidney stone detection analysis. Multimed Tool Appl 2022:1–43.
[26] Akkasaligar PT, Biradar S. Diagnosis of renal calculus disease in medical ultrasound images. In: 2016 IEEE international conference on computational intelligence and computing research (ICCIC). IEEE; 2016. p. 1–5. https://doi.org/10.1109/ICCIC.2016.7919642.
[27] Verma J, Nath M, Tripathi P, Saini K. Analysis and identification of kidney stone using *k*th nearest neighbour (KNN) and support vector machine (SVM) classification techniques. Pattern Recogn Image Anal 2017;27:574–80.
[28] Kobayashi M, Ishioka J, Matsuoka Y, Fukuda Y, Kohno Y, Kawano K, et al. Computer-aided diagnosis with a convolutional neural network algorithm for automated detection of urinary tract stones on plain X-ray. BMC Urol 2021;21:1–10.
[29] Aksakalli I, Kaçdioğlu S, Hanay YS. Kidney X-ray images classification using machine learning and deep learning methods. Balkan Journal of Electrical and Computer Engineering 2021;9:144–51.
[30] Cummings JM, Boullier JA, Izenberg SD, Kitchens DM, Kothandapani RV. Prediction of spontaneous ureteral calculous passage by an artificial neural network. J Urol 2000;164:326–8.
[31] Dal Moro F, Abate A, Lanckriet G, Arandjelovic G, Gasparella P, Bassi P, et al. A novel approach for accurate prediction of spontaneous passage of ureteral stones: support vector machines. Kidney Int 2006;69:157–60.
[32] Solakhan M, Seckiner SU, Seckiner I. A neural network-based algorithm for predicting the spontaneous passage of ureteral stones. Urolithiasis 2020;48:527–32.
[33] Park JS, Kim DW, Lee D, Lee T, Koo KC, Han WK, et al. Development of prediction models of spontaneous ureteral stone passage through machine learning: comparison with conventional statistical analysis. PLoS One 2021;16:e0260517. https://doi.org/10.1371/JOURNAL.PONE.0260517.
[34] Poulakis V, Dahm P, Witzsch U, De Vries R, Remplik J, Becht E. Prediction of lower pole stone clearance after shock wave lithotripsy using an artificial neural network. J Urol 2003;169:1250–6.
[35] Gomha MA, Sheir KZ, Showky S, Abdel-Khalek M, Mokhtar AA, Madbouly K. Can we improve the prediction of stone-free status after extracorporeal shock wave lithotripsy for ureteral stones? A neural network or a statistical model? J Urol 2004;172:175–9.
[36] Moorthy K, Krishnan M. Prediction of fragmentation of kidney stones: a statistical approach from NCCT images. Can Urol Assoc J 2016;10:E237–40. https://doi.org/10.5489/cuaj.3674.
[37] Choo MS, Uhmn S, Kim JK, Han JH, Kim D-H, Kim J, et al. A prediction model using machine learning algorithm for assessing stone-free status after single session shock wave lithotripsy to treat ureteral stones. J Urol 2018;200:1371–7.
[38] Seckiner I, Seckiner S, Sen H, Bayrak O, Dogan K, Erturhan S. A neural network-based algorithm for predicting stone-free status after ESWL therapy. Int Braz J Urol 2017;43:1110–4.
[39] Mannil M, von Spiczak J, Hermanns T, Poyet C, Alkadhi H, Fankhauser CD. Three-dimensional texture analysis with machine learning provides incremental predictive information for successful shock wave lithotripsy in patients with kidney stones. J Urol 2018;200:829–36.
[40] Yang SW, Hyon YK, Na HS, Jin L, Lee JG, Park JM, et al. Machine learning prediction of stone-free success in patients with urinary stone after treatment of shock wave lithotripsy. BMC Urol 2020;20:1–8.
[41] Tsitsiflis A, Kiouvrekis Y, Chasiotis G, Perifanos G, Gravas S, Stefanidis I, et al. The use of an artificial neural network in the evaluation of the extracorporeal shockwave lithotripsy as a treatment of choice for urinary lithiasis. Asian J Urol 2022;9:132–8.
[42] Handa RK, Territo PR, Blomgren PM, Persohn SA, Lin C, Johnson CD, et al. Development of a novel magnetic resonance imaging acquisition and analysis workflow for the quantification of shock wave lithotripsy-induced renal hemorrhagic injury. Urolithiasis 2017;45:507–13.
[43] Aminsharifi A, Irani D, Pooyesh S, Parvin H, Dehghani S, Yousofi K, et al. Artificial neural network system to predict the postoperative outcome of percutaneous nephrolithotomy. J Endourol 2017;31:461–7.
[44] Shabaniyan T, Parsaei H, Aminsharifi A, Movahedi MM, Jahromi AT, Pouyesh S, et al. An artificial intelligence-based clinical decision support system for large kidney stone treatment. Australas Phys Eng Sci Med 2019;42:771–9.
[45] Aminsharifi A, Irani D, Tayebi S, Jafari Kafash T, Shabanian T, Parsaei H. Predicting the postoperative outcome of percutaneous nephrolithotomy with machine learning system: software validation and comparative analysis with Guy's Stone Score and the CROES Nomogram. J Endourol 2020;34:692–9.
[46] Geraghty R, Finch W, Fowler S, Sriprasad S, Smith D, Dickinson A, et al. Use of internally validated machine and deep learning models to predict outcomes of percutaneous nephrolithotomy using data from the BAUS PCNL audit. medRxiv 2022. https://doi.org/10.1101/2022.06.16.22276481.
[47] Zhao H, Li W, Li J, Li L, Wang H, Guo J. Predicting the stone-free status of percutaneous nephrolithotomy with the machine learning system: comparative analysis with Guy's stone score and the STONE score system. Front Mol Biosci 2022;9:880291. https://doi.org/10.3389/fmolb.2022.880291.
[48] Chen M, Yang J, Lu J, Zhou Z, Huang K, Zhang S, et al. Ureteral calculi lithotripsy for single ureteral calculi: can DNN-assisted model help preoperatively predict risk factors for sepsis? Eur Radiol 2022;32:8540–9.
[49] Hamid A, Dwivedi U, Singh T, Gopi Kishore M, Mahmood M, Singh H, et al. Artificial neural networks in predicting optimum renal stone fragmentation by extracorporeal shock wave lithotripsy: a preliminary study. BJU Int 2003;91:821–4.
[50] Goyal NK, Kumar A, Trivedi S, Dwivedi US, Singh T, Singh PB. A comparative study of artificial neural network and multivariate regression analysis to analyze optimum renal stone fragmentation by extracorporeal shock wave lithotripsy. Saudi J Kidney Dis Transpl 2010;21:1073–80.
[51] Mannil M, Von Spiczak J, Hermanns T, Alkadhi H, Fankhauser CD. Prediction of successful shock wave lithotripsy with CT: a phantom study using texture analysis. Abdom Radiol (NY) 2018;43:1432–8.
[52] Chen Z, Zeng DD, Seltzer RG, Hamilton BD. Automated generation of personalized shock wave lithotripsy protocols: treatment planning using deep learning. JMIR Med Inform 2021;9:e24721. https://doi.org/10.2196/24721.
[53] Muller S, Abildsnes H, Østvik A, Kragset O, Gangås I, Birke H, et al. Can a dinosaur think? Implementation of artificial intelligence in extracorporeal shock wave lithotripsy. Eur Urol Open Sci 2021;27:33–42.
[54] Taguchi K, Hamamoto S, Okada A, Tanaka Y, Sugino T, Unno R, et al. Robot-assisted fluoroscopy versus ultrasound-guided renal access for nephrolithotomy: a phantom model benchtop study. J Endourol 2019;33:987–94.
[55] Wang C, Calle P, Ton NBT, Zhang Z, Yan F, Donaldson AM, et al. Deep-learning-aided forward optical coherence tomography endoscope for percutaneous nephrostomy guidance. Biomed Opt Express 2021;12:2404–18.
[56] Li G, Liu Z, Zhang Y, Chen Z, Hu P, Chen Q, et al. Discrimination analysis of B-mode ultrasonography and X-ray on the percutaneous nephrolithotomy localization of urinary stones: a prospective, controlled study. Int J Clin Exp Med 2016;9:2261–8.

[57] Jeong J, Chang K, Lee J, Choi J. A warning system for urolithiasis via retrograde intrarenal surgery using machine learning: an experimental study. BMC Urol 2022;22:1—8.
[58] Dussol B, Verdier JM, Le Goff JM, Berthezene P, Berland Y. Artificial neural networks for assessing the risk of urinary calcium stone among men. Urol Res 2006;34:17—25.
[59] Dussol B, Verdier JM, Le Goff JM, Berthezene P, Berland Y. Artificial neural networks for assessing the risk factors for urinary calcium stones according to gender and family history of stone. Scand J Urol Nephrol 2007;41:414—8.
[60] Kazemi Y, Mirroshandel SA. A novel method for predicting kidney stone type using ensemble learning. Artif Intell Med 2018;84:117—26.
[61] Chen Z, Prosperi M, Bird VG, Bird VY. Analysis of factors associated with large kidney stones: stone composition, comorbid conditions, and 24-h urine parameters—a machine learning-aided approach. SN Compr Clin Med 2019;1:597—602.
[62] Kavoussi NL, Floyd C, Abraham A, Sui W, Bejan C, Capra JA, et al. Machine learning models to predict 24 hour urinary abnormalities for kidney stone disease. Urology 2022;169:52—7.
[63] Caudarella R, Tonello L, Rizzoli E, Vescini F. Predicting five-year recurrence rates of kidney stones: an artificial neural network model. Arch Ital Urol Androl 2011;83:14—9.
[64] Chiang D, Chiang HC, Chen WC, Tsai FJ. Prediction of stone disease by discriminant analysis and artificial neural networks in genetic polymorphisms: a new method. BJU Int 2003;91:661—6.
[65] Xiang H, Chen Q, Wu Y, Xu D, Qi S, Mei J, et al. Urine calcium oxalate crystallization recognition method based on deep learning. In: 2019 international conference on automation, computational and technology management, ICACTM 2019; 2019. p. 30—3. https://doi.org/10.1109/ICACTM.2019.8776769.
[66] Kletzmayr A, Mulay SR, Motrapu M, Luo Z, Anders HJ, Ivarsson ME, et al. Inhibitors of calcium oxalate crystallization for the treatment of oxalate nephropathies. Adv Sci 2020;7:1903337. https://doi.org/10.1002/advs.201903337.
[67] Kriegshauser JS, Silva AC, Paden RG, He M, Humphreys MR, Zell SI, et al. *Ex vivo* renal stone characterization with single-source dual-energy computed tomography: a multiparametric approach. Acad Radiol 2016;23:969—76.
[68] Kriegshauser JS, Paden RG, He M, Humphreys MR, Zell SI, Fu Y, et al. Rapid kV-switching single-source dual-energy CT *ex vivo* renal calculi characterization using a multiparametric approach: refining parameters on an expanded dataset. Abdom Radiol (NY) 2018;43:1439—45.
[69] Zhang GMY, 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—9.
[70] Große Hokamp N, Lennartz S, Salem J, Pinto dos Santos D, Heidenreich A, Maintz D, et al. Dose independent characterization of renal stones by means of dual energy computed tomography and machine learning: an *ex-vivo* study. Eur Radiol 2020;30:1397—404.
[71] Tang L, Li W, Zeng X, Wang R, Yang X, Luo G, 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. https://doi.org/10.21037/atm-21-965.
[72] 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—4.
[73] Lopez F, Varelo A, Hinojosa O, Mendez M, Trinh DH, ElBeze Y, et al. Assessing deep learning methods for the identification of kidney stones in endoscopic images. Annu Int Conf IEEE Eng Med Biol Soc 2021;2021:2778—81.
[74] El Beze J, Mazeaud C, Daul C, Ochoa-Ruiz G, Daudon M, Eschwège P, et al. Evaluation and understanding of automated urinary stone recognition methods. BJU Int 2022;130:786—98.
[75] Ochoa-Ruiz G, Estrade V, Lopez F, Flores-Araiza D, Beze JE, Trinh DH, et al. On the *in vivo* recognition of kidney stones using machine learning. 2022. arXiv preprint arXiv:2201.08865.
[76] Mendez-Ruiz M, Lopez-Tiro F, El-Beze J, Estrade V, Ochoa-Ruiz G, Hubert J, et al. On the generalization capabilities of FSL methods through domain adaptation: a case study in endoscopic kidney stone image classification. 2022. arXiv preprint arXiv:2205.00895.
[77] Kim US, Kwon HS, Yang W, Lee W, Choi C, Kim JK, et al. Prediction of the composition of urinary stones using deep learning. Investig Clin Urol 2022;63:441—7.
[78] Fitri LA, Haryanto F, Arimura H, YunHao C, Ninomiya K, Nakano R, et al. Automated classification of urinary stones based on microcomputed tomography images using convolutional neural network. Phys Med 2020;78:201—8.
[79] Saçlı B, Aydınalp C, Cansız G, Joof S, Yilmaz T, Çayören M, et al. Microwave dielectric property based classification of renal calculi: application of a kNN algorithm. Comput Biol Med 2019;112:103366. https://doi.org/10.1016/j.compbiomed.2019.103366.
[80] Cui X, Zhao Z, Zhang G, Chen S, Zhao Y, Lu J. Analysis and classification of kidney stones based on Raman spectroscopy. Biomed Opt Express 2018;9:4175—83.
[81] Onal EG, Tekgul H. Assessing kidney stone composition using smartphone microscopy and deep neural networks. BJUI Compass 2022;3:310—5.
[82] Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. Am J Med 2019;132:795—801.
[83] Denny JC, Collins FS. Precision medicine in 2030—seven ways to transform healthcare. Cell 2021;184:1415—9.
[84] Hameed BZ, Dhavileswarapu S, Naik N, Karimi H, Hegde P, Rai BP, et al. Big data analytics in urology: the story so far and the road ahead. Ther Adv Urol 2021;13:1756287221998134. https://doi.org/10.1177/1756287221998134.
[85] John-Charles R, Tsanas A, Singh S. Rise of the machines: will artificial intelligence replace the urologist?. https://www.urologynews.uk.com/features/features/post/rise-of-the-machines-will-artificial-intelligence-replace-the-urologist. [Accessed 29 March 2021].
[86] Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015—20): a comparative analysis. Lancet Digit Health 2021;3:e195—203. https://doi.org/10.1016/S2589-7500(20)30292-2.
[87] Cutillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD. Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digit Med 2020;3:1—5.
[88] Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. https://www.fda.gov/media/145022/download. [Accessed July 5 2021].
[89] Stöger K, Schneeberger D, Holzinger A. Medical artificial intelligence: the European legal perspective. Commun ACM 2021;64:34—6.
[90] Stai B, Heller N, McSweeney S, Rickman J, Blake P, Vasdev R, et al. Public perceptions of artificial intelligence and robotics in medicine. J Endourol 2020;34:1041—8.
[91] Schoenthaler M, Boeker M, Horki P. How to compete with Google and Co.: big data and artificial intelligence in stones. Curr Opin Urol 2019;29:135—42.