



OPEN

Identification of kidney stones in KUB X-ray images using VGG16 empowered with explainable artificial intelligence

Fahad Ahmed¹, Sagheer Abbas², Atifa Athar³, Tariq Shahzad⁴, Wasim Ahmad Khan¹, Meshal Alharbi⁵, Muhammad Adnan Khan^{6,7,8} & Arfan Ahmed⁹✉

A kidney stone is a solid formation that can lead to kidney failure, severe pain, and reduced quality of life from urinary system blockages. While medical experts can interpret kidney-ureter-bladder (KUB) X-ray images, specific images pose challenges for human detection, requiring significant analysis time. Consequently, developing a detection system becomes crucial for accurately classifying KUB X-ray images. This article applies a transfer learning (TL) model with a pre-trained VGG16 empowered with explainable artificial intelligence (XAI) to establish a system that takes KUB X-ray images and accurately categorizes them as kidney stones or normal cases. The findings demonstrate that the model achieves a testing accuracy of 97.41% in identifying kidney stones or normal KUB X-rays in the dataset used. VGG16 model delivers highly accurate predictions but lacks fairness and explainability in their decision-making process. This study incorporates the Layer-Wise Relevance Propagation (LRP) technique, an explainable artificial intelligence (XAI) technique, to enhance the transparency and effectiveness of the model to address this concern. The XAI technique, specifically LRP, increases the model's fairness and transparency, facilitating human comprehension of the predictions. Consequently, XAI can play an important role in assisting doctors with the accurate identification of kidney stones, thereby facilitating the execution of effective treatment strategies.

Keywords Artificial intelligence (AI), Machine learning (ML), Deep learning (DL), Convolutional neural network (CNN), Transfer learning (TL), VGG16, Kidney-ureter-bladder (KUB), Kidney stones, Explainable artificial intelligence (XAI), Layer-wise relevance propagation (LRP)

Urolithiasis, or kidney stones, is one of the most common urological conditions worldwide¹. Kidney stones are hard concretion or stone-like pieces that form in the kidneys due to dietary minerals in the urine². Symptoms, including flank pain, nausea, and vomiting, can indicate kidney stones³. While they can manifest in individuals of any gender, prevalence is higher in males, with approximately 7% of females and 13% of males experiencing them in their lifetime⁴. Factors such as dietary habits, sedentary lifestyle, diabetes mellitus, obesity, hypertension, and metabolic syndrome elevate the risk of stone formation⁵.

Medical professionals use imaging techniques to identify kidney stones removed by surgical intervention. After treatment, kidney stones may recur and develop into a chronic condition after treatment and kidney malfunctions can be life-threatening⁶. The ureter can become blocked depending on the size of the stone, causing significant pain, particularly in the lower back, although it can hurt the groin⁷. Older people are more likely to

¹School of Computer Science, National College of Business Administration and Economics, Lahore 54000, Pakistan. ²Department of Computer Sciences, Bahria University, Lahore Campus, Lahore 54000, Pakistan. ³Department of Computer Science, Comsats University Islamabad, Lahore Campus, Lahore 54000, Pakistan. ⁴Department of Computer Sciences, COMSATS University Islamabad, Sahiwal Campus, Sahiwal 57000, Pakistan. ⁵Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, 11942 Alkharj, Saudi Arabia. ⁶School of Computing, Skyline University College, University City Sharjah, 1797 Sharjah, UAE. ⁷Department of Software, Faculty of Artificial Intelligence and Software, Gachon University, Seongnam-si 13120, Republic of Korea. ⁸Riphah School of Computing and Innovation, Faculty of Computing, Riphah International University, Lahore Campus, Lahore 54000, Pakistan. ⁹AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar. ✉email: adnan@gachon.ac.kr; ara4013@qatar-med.cornell.edu

report atypical or no pain when passing a stone, making diagnosing kidney stone disease challenging in this demographic⁸. Different stages of disease evaluation employ various imaging techniques. The typical imaging techniques for examining kidney stones are sonography⁹, computed tomography (CT)¹⁰, and KUB X-ray imaging¹¹. Sonography, also known as ultrasonography or simply ultrasound, is a quick, safe, and easy procedure that can provide valuable evidence for a kidney stone diagnosis. Still, its sensitivity for detecting kidney stones is limited. CT can identify kidney stones and determine their number, location, and size; however, it involves exposure to ionizing radiation. KUB X-ray can also detect kidney stones and provide essential information regarding their classification, shape, number, position, and size. In this context, the most popular method is two plain KUB X-ray imaging, which is already available, less expensive, and exposes patients to less radiation than CT¹².

One of the most crucial stages in locating, measuring, and identifying the composition of kidney stones before and during treatment involves using KUB x-ray imaging, which is also employed to evaluate prognosis. Figure 1 displays samples of KUB X-ray images depicting kidney stones and normal images used in this article.

Nephrologists typically use KUB X-ray images to identify kidney stones. This information helps determine whether the individual is healthy or a patient requiring treatment. Directing an X-ray beam through the body obtains a KUB radiograph. The resulting image appears in shades of black and white, depending on the varying densities and X-ray absorption of different body parts. Muscles and fat appear grey due to medium densities, while bones appear white due to their high density. The low density of the air in the lungs makes them appear black on the radiograph.

Given the increasing prevalence and the complexity of diagnosing kidney stones, there is an urgent need for innovative diagnostic techniques. Traditional imaging techniques may produce low-quality images, making interpreting results easier. This limitation has led to the exploration of new methods, including using artificial intelligence (AI) to improve diagnostic accuracy. Integrating AI into novel diagnostic methodologies holds significant promise for refining diagnostic accuracy and facilitating therapeutic interventions¹³.

Machine learning (ML), a subfield of AI, is widely considered a powerful tool for enhancing disease prediction and diagnosis^{14,15}. Recently, there has been a substantial increase in the quantity and quality of research focused on utilizing ML for automatic disease identification. However, effective feature extraction methods are essential for improving ML models. The need for the manual formulation of complex hypotheses in traditional ML classifiers constitutes a disadvantage¹⁶. In contrast, deep neural networks (DNNs) can autonomously generate complex hypotheses, rendering them practical for learning nonlinear correlations¹⁶. This autonomous capability is part of why deep learning (DL), a subset of which includes DNNs, has historically diverged from traditional ML methods¹⁷. Due to their enhanced efficacy in processing large-scale data sets, their ability to extract hidden valuable knowledge from data, and to employ specific pre-trained networks, DL models are therefore frequently used in medical imaging systems.

DL can learn from and model vast amounts of data¹⁸. Due to their advanced information processing capabilities, DL models can effectively represent complex, high-dimensional datasets¹⁹. Deep models have been effectively applied in various applications, including lesion detection^{20,21}, classification^{22–25}, object tracking²⁶, image super-resolution reconstruction²⁷, image inpainting^{28–32}, and segmentation of medical images^{33,34}. Autoencoder (AE), Recurrent Neural Networks (RNN), Deep Belief Networks (DBN), Direct Deep Reinforcement Learning, Recursive Neural Networks, and Convolutional Neural Networks (CNN) are standard DL techniques³⁵. CNNs are frequently used in DL to automatically learn features, which are then used for classification and detection^{36,37}.

DL approaches find extensive use within the healthcare sector. However, these models, often called “black boxes,” challenge our understanding of the rationale behind their decisions or predictions, and the absence of interpretability can create issues. In this context, implementing explainable AI (XAI) techniques can enhance transparency and improve understanding of its decisions. This article introduces a model that suggests identifying kidney stones by applying transfer learning (TL) empowered by XAI.

The development of medical image processing methods has accelerated the introduction of smart prediction and diagnosis tools³⁸. AI can assist doctors in making better clinical decisions in specific functional areas of healthcare, such as radiography, or may even replace human judgment in certain circumstances³⁹. AI employs DL, ML, and other learning-based methods⁴⁰. Recent research has shown the utilization of DL techniques in specific applications, including an enhanced rime optimization-driven multi-threshold segmentation for COVID-19 X-ray images⁴¹, high-precision multiclass classification of lung diseases using customized MobileNetV2⁴², phase retrieval for X-ray differential phase contrast radiography with knowledge transfer learning⁴³, attention-based VGG-16 model for COVID-19 chest X-ray image classification⁴⁴, and pre-trained VGG-16 with CNN architecture for classifying X-ray images into normal or pneumonia categories⁴⁵. Researchers in the field of AI have created numerous ML and DL algorithms for detecting kidney stones over the past few decades.

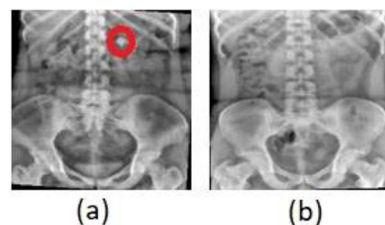


Figure 1. KUB x-ray sample images: (a) Kidney Stone; (b) Normal.

For the Computer-aided diagnosis (CAD) of kidney stones, Ishioka et al.⁴⁶ employed a CNN (ResNet) method utilizing over 1000 KUB x-ray images from three hospitals. The researchers used 190 as testing data and 827 as training data. The test dataset's precision, sensitivity, and F1 score were 0.49, 0.72, and 0.58, respectively. Chiang et al.⁴⁷ introduced an algorithm for detecting kidney stones using an artificial neural network (ANN) and discriminant analysis (DA) in conjunction with genetic polymorphisms and environmental factors such as milk consumption, water consumption, and outdoor activities. The research revealed that considering only genetic factors does not produce noticeable distinctions in the success of the models. However, considering the environmental and genetic factors, the ANN model outperforms the DA model with 89% accuracy.

Dussol et al.⁴⁸ implemented ANN models to examine 11 clinical and biochemical markers in 119 males with kidney stone formation and 96 males in the control group. Using linear discriminant analysis (LDA), they accurately identified 75.8% of the cases. Multivariate discriminant analysis (MVDA) accurately classified 74.4% of the patients.

In a parallel investigation, Cauderella et al.⁴⁹ implemented the ANN models in conjunction with traditional statistical methodologies to predict the recurrence of incidents within a five-year timeframe post-initial clinical diagnosis and metabolic assessment. They based their model on a dataset from 80 patients with kidney stone disease. Owing to its established reliability as a traditional statistical technique, logistic regression (LR) was selected as a comparison tool for ANN. The same training and testing sets as for ANN were used to create and test LR. The statistical software Statistical Package for the Social Sciences (SPSS) was used to develop LR. The ANN model demonstrated a predictive accuracy of 88.8%, significantly outperforming the LR model, which yielded an accuracy rate of 67.5%.

In a separate investigation conducted by Kumar and Abhishek⁵⁰, researchers made a comparative analysis to evaluate the diagnostic efficacy of three distinct neural network algorithms: Learning Vector Quantization (LVQ), Multilayer Perceptron (MLP), and Radial Basis Function (RBF). They compared the algorithms in terms of their level of accuracy, training dataset size, and the time required to construct a model. The MLP algorithm emerged as the most productive, with an accuracy of 92%, thereby establishing itself as an optimal tool for the early detection of kidney stones in patients and reducing the time required for diagnosis.

Ebrahimi and Mariano⁵¹ created a semi-automated program to enhance kidney stone detection in KUB computed tomography (KUB CT) analysis using image processing techniques and geometry principles. The program outlines and segments the kidney area, identifies kidney stones, and determines their size and position using pixel count metrics. An evaluation of the framework's performance on KUB CT scans from a cohort of 39 patients yielded a detection accuracy of 84.61%, indicating its potential to augment diagnostic precision in kidney stone identification. Kazemi and Mirroshandel⁵² proposed a novel method for predicting the chance of a kidney stone using ensemble learning. They sourced data from 936 patients diagnosed with nephrolithiasis at the Renal Center of the Razi Hospital in Rasht between 2012 and 2016. The ensemble-based model's final accuracy was 97.1%. Li and Elliot⁵³ conducted a study to assess the accuracy of natural language processing (NLP) in recognizing a group of patients (n = 1874) with positive CT KUB results for renal stones. The NLP achieved an accuracy rate of 85%.

De Perrot et al.⁵⁴ developed an ML algorithm that employs radiomics feature extraction from low-dose CT (LDCT) images to differentiate between kidney stones and phleboliths. This ML classification model, trained on radiomics characteristics, achieved an overall accuracy of 85.1% on the independent testing set. In another study, Kahani et al.⁵⁵ presented a classification technique for urinary stones utilizing KUB x-ray images. They employed the least absolute shrinkage and selection operator (LASSO) algorithm with ML classifiers. This methodology yielded a classification accuracy of 96% for kidney stones. Jungmann et al.⁵⁶ created an NLP technique trained on subjective assessment to automatically collect positive hit rates and clinical information to evaluate 1714 narrative LDCT reports. In 38% of occurrences, there was a minimum of one kidney stone, and in 45%, there was a minimum of one ureter stone.

Annameti Rohith et al.⁵⁷ developed a technique employing median and rank filters to increase the detection rate of identifying kidney stones in ultrasound images regarding accuracy and sensitivity. They evaluated the median and rank filters for their accuracies and sensitivities using a MATLAB simulation tool with a sample size 114 and a *p* value of 0.8. The median filter achieved an accuracy of 86.4%, the rank filter attained an accuracy of 82.2%, the median filter's sensitivity was 87.7%, and the rank filter's sensitivity was 82.5%. The median filter significantly outperformed the rank filter in both accuracy and sensitivity. Suresh and Abhishek⁵⁸ proposed image-processing techniques to detect kidney stones in KUB ultrasound images, including pre-processing, segmentation, and morphology. Their model achieved an accuracy of 92.57% in kidney stone detection.

To discriminate between distal ureteric calculi and phleboliths using the characteristics of non-contrast CT (NCCT) images, Jendenber et al.⁵⁹ trained and created a CNN model. They then compared their findings to the assessments of seven professional radiologists. The radiologists' accuracy was 86%, whereas the CNN model's was significantly higher at 92%. Cui et al.⁶⁰ proposed a DL and threshold-based model for detecting kidney stones. They performed experiments employing a small dataset of 625 CT images and achieved an accuracy of 90.30% and a sensitivity of 95.9%.

Yildirim et al.⁶¹ proposed a DL model for automated kidney detection utilizing 1799 coronal CT images. For kidney stone detection, they used XResNet-50. Using CT images to identify kidney stones, the designed automated model obtained a 96.82% identification rate. Tsitsiflis et al.⁶² constructed an ANN to evaluate extracorporeal shockwave lithotripsy (ESWL) parameters in patients with urinary lithiasis. Medical data from 716 patients were collected. 549 were used for training, 167 for testing, and 12 nodes were used as inputs for the ANN. The ANN achieved a testing accuracy of 81.43%.

Valencia et al.⁶³ introduced an image-processing methodology for detecting kidney stones in CT scans. The study comprised four steps: image preprocessing with a median filter, segmentation using the *k*-means clustering algorithm, kidney stone detection, and classification. The team gathered data from approximately 40 patients

diagnosed with kidney stone diseases, utilizing CT scans in a clinical setting. The novel approach in this study aimed to detect boundaries and segment areas and enhance kidney stone detection through pixel-level analysis. This methodology enables both the localization of kidney stones and the quantification of affected patients. The algorithm achieved an accuracy rate of 92.5%.

While existing literature has made valuable contributions to the field, some areas could benefit from further exploration (Table 1 outlines gaps identified in previous research). Given the identified research gaps, our proposed method aims to overcome these limitations and drive progress in kidney stone identification. The main motivations and innovations of our work are outlined below:

1. The studies encompassed in the review, ranging from references^{47–63}, have not incorporated data augmentation methodologies. Data augmentation methodologies improve model performance, reduce overfitting, and enhance the ability of the model to generalize to new, unseen data.
2. Previous literature^{47–63} may have yet to attain optimal accuracy in identifying and predicting kidney stones. Improved accuracy increases the chances of identifying kidney stones.
3. Current models could benefit from enhanced transparency and fairness to improve the interpretability of their predictions. A deeper understanding of the decision-making process and contributing factors is essential for achieving more transparent, fair, and effective diagnostic outcomes.

For this paper, the main contributions are as follows:

1. The proposed research introduces a novel deep TL model that autonomously extracts relevant features from KUB X-ray images. This model successfully identifies the presence of kidney stones in these images.
2. The proposed model uses various performance measures, including accuracy, misclassification rate, precision, sensitivity, specificity, false positive rate (FPR), false negative rate (FNR), and F1 Score. The evaluations show that the model performs reliably and commendably.
3. The study conducts a comparative analysis between the proposed model and existing methodologies documented in the literature^{47–63}. This evaluation reveals that the proposed model achieves higher accuracy than previous approaches, thus showcasing its superiority in kidney stone identification.
4. The research includes a technique called XAI, specifically layer-wise relevance propagation (LRP), to improve the transparency and fairness of the model's predictions. LRP helps clarify the reasoning behind the model's predictions, thereby promoting transparency and fairness in the kidney stone identification process.

The rest of the article is divided into the following sections: The proposed model's methodology is described in Section "Methodology". Simulation and results are presented in Section "Simulation and results". The conclusion is presented in Section "Conclusion". Limitations and future work are briefly discussed in Section "Limitations and future work".

Methodology

The proposed kidney stone identification model employs DL empowered with XAI (Fig. 2). The model consists of five layers and two phases: training and validation. During the training phase, Layer 1 is dedicated to acquiring raw kidney-ureter-bladder (KUB) x-ray images, categorized as either 'kidney stone' or 'normal.' These images are high-resolution JPEG files with dimensions exceeding 2000 × 2000 pixels. In Layer 2, raw data undergoes preprocessing per the requirements of the DL model. The images are resized to dimensions of 224 × 224 × 3 and converted into PNG format. In this context, '224 × 224' signifies length and width, and '3' denotes the number of channels. Following preprocessing, data is separated between training and testing, with 70% allocated for training and 30% for testing. The pre-trained VGG16 model is imported and customized to the DL model.

Layer 3 describes the predictions made by the DL model. While these predictions hold potential utility in decision-making, they do not offer insights into the model's reasoning, thus making it a 'black box.' To mitigate this, Layer 4 incorporates XAI into the model. This feature compares the DL model's predictions with the preprocessed data to furnish explanations. If the explanations are unfair, the model is retrained; otherwise, it is stored on the cloud.

Layer 5 represents the validation phase of the model, wherein the trained model is imported from the cloud to validate the pre-processed data acquired from various sources. The proposed model intelligently classifies the KUB x-ray images into two classes with explanations. Following the successful identification of kidney stones, the system saves the corresponding data.

Table 2 represents the pseudocode for the proposed kidney stone identification model.

KUB x-ray images dataset

KUB X-ray images were acquired from the Department of Urology and Kidney Transplantation at MAYO Hospital in Lahore, Pakistan. The dataset consists of 500 KUB X-ray images selected from patients who had undergone radiographic examinations for kidney stones between February 2021 and October 2022. The images were obtained through the anteroposterior (AP) view. Two radiology specialists examined the collected KUB X-ray images and determined the presence or absence of kidney stones. Of the 500 images, 250 were identified as exhibiting kidney stones, while the remaining 250 were not. Subsequently, the images were augmented into 14,265 KUB x-ray images. Within this augmented dataset, 8941 images displayed instances of kidney stones, while 5324 images represented normal cases. As mentioned in the methodology section, the dataset is divided

Author	Dataset	Dataset type	Technique	Outcomes	Limitations		
					Data augmentation	Improvable Accuracy	Use of XAI
Chiang et al. ⁴⁷	105 healthy controls and 151 patients with calcium oxalate stones	Handcrafted features	ANN and DA	ANN achieved an accuracy of 89%, while DA achieved 75%	No	Yes	No
Dussol et al. ⁴⁸	119 stone formers and 96 controls	Handcrafted features	ANN (LDA and MVDA)	ANN with LDA gives a high accuracy of 75.8% compared to ANN with MVDA (74.4%)	No	Yes	No
Cauderella et al. ⁴⁹	80 patient's data	Handcrafted features	ANN and LR	Comparing the performance of ANN with LR and then observing that ANN gives better accuracy of 88.8% compared to LR (67.5%)	No	Yes	No
Kumar and Abhishek ⁵⁰	Data from 1000 patients	Handcrafted features	LVQ, MLP, and RBF	The accuracy obtained by LVQ, MLP, and RBF is 84%, 92%, and 87%, respectively	No	Yes	No
Ebrahimi and Mariano ⁵¹	KUB CT scan slides from 39 patients	Image-based	Image processing techniques and geometry principles	Detect kidney stones with an accuracy of 84.61%	No	Yes	No
Kazemi and Mirroshandel ⁵²	Numeric characteristics from 936 patients	Handcrafted features	Ensemble learning model	Obtained an accuracy of 97.1%	No	Yes	No
Li and Elliot ⁵³	1874 CT KUB reports	Handcrafted features	NLP	An overall accuracy of 85% was attained by applying NLP to CT KUB reports	No	Yes	No
De Perrot et al. ⁵⁴	416 patient data	Handcrafted features	ML model	Using a machine learning model results in an overall accuracy of 85.1%	No	Yes	No
Kahani et al. ⁵⁵	KUB x-ray images	Image-based	LASSO with ML classifiers	Obtained an accuracy of 96%	No	Yes	No
Jungmann et al. ⁵⁶	1714 LDCT images	Image-based	NLP	Applying NLP to 1714 LCDT images achieves an overall accuracy of 72%	No	Yes	No
Annameti Rohith et al. ⁵⁷	114 ultrasound images	Image-based	Median and rank filters	When applied to 114 ultrasound images, the median filter gives an overall high accuracy of 86.4% compared to the rank filter (82.2%)	No	Yes	No
Suresh and Abhishek ⁵⁸	KUB ultrasound images	Image-based	Image processing techniques	The proposed model gives an accuracy of 92.57% for stone detection	No	Yes	No
Jendenber et al. ⁵⁹	NCCT images of 341 patients containing a distal ureteral stone, phlebolith, or both	Image-based	CNN	CNN differentiated stones and phlebolith with 92% accuracy	No	Yes	No
Cui et al. ⁶⁰	625 CT images	Image-based	DL and threshold-based model	Achieved an accuracy of 90.30%	No	Yes	No
Yildirim et al. ⁶¹	1799 coronal CT images	Image-based	XResNet-50	Using CT images, XResNet-50 demonstrated an accuracy of 96.82%	No	Yes	No
Tsitsiflis et al. ⁶²	Medical data of 716 patients	Feature-based data	ANN	Achieved a testing accuracy of 81.43%	No	Yes	No
Valencia et al. ⁶³	CT scans of around 40 patients	Image-based	Image processing	Achieved an accuracy of 92.5% for stone detection	No	Yes	No

Table 1. Limitations and outcomes of previous work.

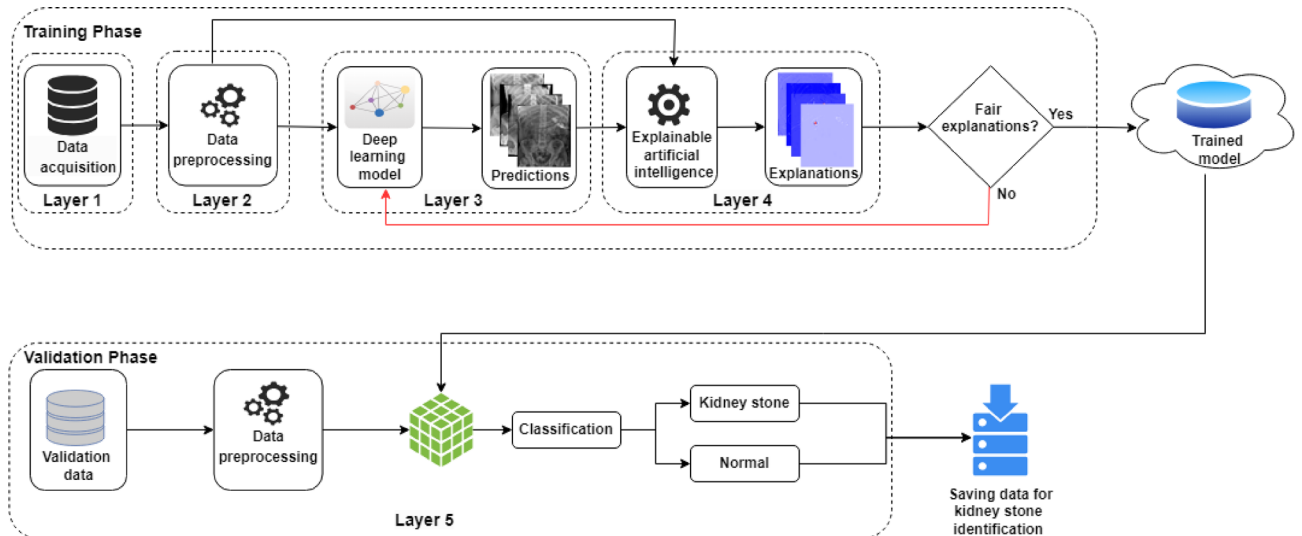


Figure 2. Proposed kidney stone identification model using DL empowered with XAI.

Pseudocode for proposed kidney stone identification model

Training Phase

function training_phase():

Layer 1: Acquiring Raw KUB X-ray Images

raw_images = acquire_raw_images(categories=['kidney stone', 'normal'])

Layer 2: Preprocessing

preprocessed_data = preprocess_data(raw_images, target_dimensions=(224, 224, 3), format='PNG')

Split data into training and testing sets

training_data, testing_data = split_data(preprocessed_data, split_ratio=0.7)

Layer 3: Import and customize the pre-trained VGG16 model

dl_model = import_and_customize_model('VGG16', num_classes=2)

Train the DL model

dl_model.train(training_data)

Layer 4: Explainable AI (XAI)

explanations = explain_model_predictions(dl_model, testing_data)

Check the fairness of explanations

if explanations_are_unfair(explanations):

Retrain the model

dl_model.retrain()

else:

Store the trained model on the cloud

store_model_on_cloud(dl_model)

Validation Phase

function validation_phase():

Layer 5: Import trained model from the cloud

dl_model = import_model_from_cloud()

Acquire preprocessed data from various sources

validation_data = acquire_preprocessed_data()

Classify KUB X-ray images with explanations

predictions_with_explanations = classify_with_explanations(dl_model, validation_data)

Save data after successful identification of kidney stones

save_identified_data(predictions_with_explanations)

Table 2. Pseudocode for proposed kidney stone identification model.

into 70:30. The number of training images is 9986 (kidney stone 6259, normal 3727), while the number of testing images is 4279 (kidney stone 2682, normal 1597).

TL

TL is a technique for applying a model's previously acquired knowledge to a new dataset⁶⁴. TL enables the utilization of highly competent, pre-trained networks rather than creating CNNs for each application. The core idea is that specific applications can be modeled by training a large model on a diverse and broad dataset. The initial layers will learn generic properties such as color, while later layers will serve particular applications. A pre-trained model, VGG16, is employed in this article to identify and predict kidney stones.

VGG16

In 2014, Simonyan and Zisserman introduced VGG16, a TL-based CNN model characterized by a sequential network structure⁶⁵. VGG16 is a deep CNN architecture with a total of 16 layers^{65,66}, which includes 13 convolutional layers and 3 fully connected dense layers (Fig. 3).

The original VGG16 model was initially trained to classify 1000 different object classes. However, the two classes of KUB x-ray images used in this study cannot be directly classified by the original VGG16 model. The current study introduces a model to classify KUB x-rays using a modified version of the VGG16 model (Fig. 4). This modified version of the VGG16 model enables the direct classification of the two KUB x-ray classes.

XAI

According to⁶⁷, explainability means the capacity to communicate how an AI decision has reached a broader range of end users in ways humans can comprehend. Many AI models, particularly those based on DL, have the potential to be challenging to understand. These models often involve millions of parameters and rely on complex patterns and correlations that are difficult to decipher. This complexity can raise concerns about bias, privacy, ethics, fairness, and transparency.

To address these concerns, XAI refers to the capability of AI systems to provide understandable and interpretable explanations for their decisions and actions, techniques that aim to enhance the comprehensibility and transparency of AI models. In this study, the LRP technique is used to determine which features of the DL model are responsible for specific predictions.

Layer-wise relevance propagation

For enhancing the explainability of networks utilizing the back-propagation algorithm, one of the principal algorithms employed is LRP⁶⁸. A backward propagation technique called LRP gives relevance scores to a model's

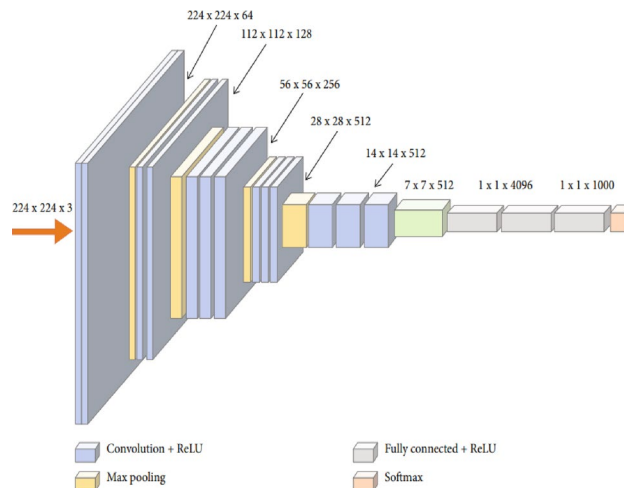


Figure 3. VGG16 original architecture⁶⁶.

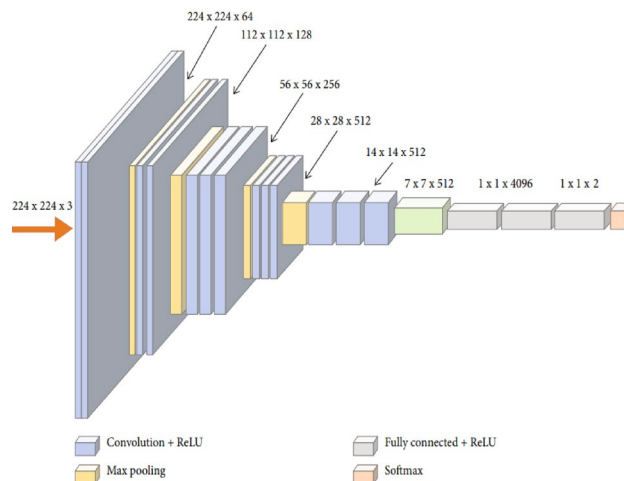


Figure 4. Modified version of the VGG16 model.

input features based on how much they contribute to the output. The most crucial neurons for the prediction are then identified through the model layers using the relevance scores. Additionally, LRP deals with the shortcomings of shattered gradients in gradient methods (Grad-CAM) and perturbation methods (occlusion maps)⁶⁹.

Simulation and results

We utilized Google Colab and PyTorch for simulation and obtaining results. Google Colab furnished the necessary computational resources, while PyTorch was an efficient framework for constructing and training DL models. Our performance assessment employed the metrics derived from Eqs. (1–8)^{70,71}, wherein Kp/Sp represents true positives, Km/Sm denotes true negatives, Ke/Se signifies false positives, and Kn/Sn indicates false negatives. The computed metrics encompassed accuracy, misclassification rate, precision, sensitivity, specificity, FPR, FNR, and F1 Score.

Accuracy Accuracy is the proportion of correctly classified instances out of the total predictions made by a model, often represented as a percentage.

$$\text{Accuracy} = \frac{(K_p/S_p) + (K_m/S_m)}{(K_p/S_p) + (K_m/S_m) + (K_e/S_e) + (K_n/S_n)} * 100 \quad (1)$$

Misclassification rate The misclassification rate is the proportion of incorrectly classified instances out of the total predictions, usually expressed as a percentage or a fraction.

$$\text{Misclassification rate} = \frac{(K_e/S_e) + (K_n/S_n)}{(K_p/S_p) + (K_m/S_m) + (K_e/S_e) + (K_n/S_n)} * 100 \quad (2)$$

Precision Precision measures the ratio of true positive predictions to the total positive predictions made by a model, emphasizing the accuracy of positive classifications.

$$\text{Precision} = \frac{(K_p/S_p)}{(K_p/S_p) + (K_e/S_e)} * 100 \quad (3)$$

Sensitivity Sensitivity calculates the proportion of true positive predictions relative to all actual positive instances, indicating a model's ability to identify positives correctly.

$$\text{Sensitivity} = \frac{(K_p/S_p)}{(K_p/S_p) + (K_n/S_n)} * 100 \quad (4)$$

Specificity Specificity quantifies the ratio of true negative predictions to all actual negative instances, measuring a model's capacity to identify negatives correctly.

$$\text{Specificity} = \frac{(K_m/S_m)}{(K_m/S_m) + (K_e/S_e)} * 100 \quad (5)$$

FPR FPR is the proportion of false positive predictions relative to all actual negative instances, demonstrating the model's tendency to misclassify negatives as positives.

$$\text{FPR} = \frac{(K_e/S_e)}{(K_e/S_e) + (K_m/S_m)} * 100 \quad (6)$$

FNR FNR calculates the ratio of false negative predictions to all actual positive instances, illustrating the model's likelihood to misclassify positives as negatives.

$$\text{FNR} = \frac{(K_n/S_n)}{(K_n/S_n) + (K_p/S_p)} * 100 \quad (7)$$

F1 Score The F1 Score is the harmonic mean of precision and sensitivity, providing a single metric that balances both aspects of classification accuracy.

$$\text{F1 Score} = \frac{2 * (\text{Precision} * \text{Sensitivity})}{\text{Precision} + \text{Sensitivity}} \quad (8)$$

For the model's training hyperparameters, we maintained the mini-batch size at 32, determined the optimal training epoch to be 10, applied a learning rate of 0.00001 during network training, and utilized the Adam optimization algorithm for the training process (Table 3 outlines each hyperparameter, accompanied by an explanatory note).

Subsequently, we tested the modified VGG16 model to analyze a dataset comprising 4279 KUB X-rays, aiming to distinguish between X-rays featuring kidney stones and those categorized as normal (Fig. 5; Table 4). Regarding kidney stone X-rays from KUB, the model identified 2612 X-rays as kidney stones (true positives). While mistakenly labeling 70 X-rays as normal (false positives). For normal X-rays of KUB, the model correctly identified 1556 X-rays as normal (true negatives) and erroneously labeled 41 X-rays as kidney stones (false negatives).

Table 4 illustrates the statistical significance of each criterion for the modified version of the VGG16, including accuracy, misclassification rate, precision, sensitivity, specificity, FPR, FNR, and F1 Score.

Sr. no.	Parameters	Value	Explanatory note
1	Size of the Mini Batch	32	A mini-batch size of 32 processes 32 data samples together in each training iteration
2	No. of Epochs	10	It means training the modified VGG16 for 10 complete passes through the entire dataset
3	Learning Rate	0.00001	A learning rate 0.00001 signifies a small step size for updating model parameters during training
4	Optimization Algorithm	Adam	The optimization algorithm "Adam" combines adaptive learning rates and momentum to update model parameters during training efficiently

Table 3. Training hyperparameters.

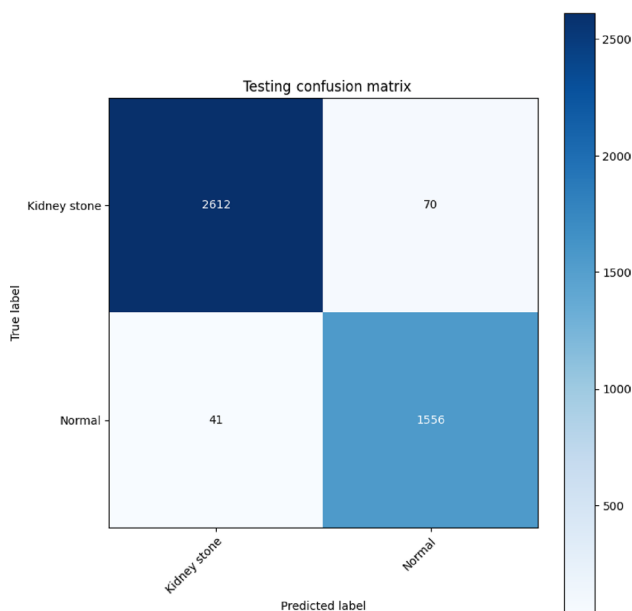


Figure 5. Testing confusion matrix for the modified VGG16 model.

Performance parameters	Testing
Accuracy	97.41%
Misclassification rate	2.59%
Precision	97.39%
Sensitivity	98.45%
Specificity	95.70%
FPR	4.30%
FNR	1.55%
F1 Score	0.98

Table 4. Statistical significance of each criterion for the modified VGG16 model.

Employing the LRP technique on the modified VGG16 model allowed us to pinpoint the regions in the KUB X-ray image that significantly contribute to the model's prediction of kidney stone presence. Notably, highlighted areas in KUB X-rays indicate the presence of kidney stones, while normal X-rays exhibit clarity and lack visible indications (Fig. 6).

Numerous ways have been utilized to detect kidney stones; nevertheless, TL is a revolutionary method for identifying the presence of kidney stones. Table 5 compares the proposed model's performance to previously reported state-of-the-art literature. The proposed model integrates modified VGG16 architecture with the XAI technique, significantly advancing kidney stone identification. This model distinguishes itself through exceptional performance, achieving a remarkable testing accuracy of 97.41% and an impressively low misclassification rate of 2.59%. Utilizing the XAI technique enhances the model's transparency and interpretability, addressing critical concerns related to the opacity of DL models. Additionally, the model benefits from a substantial dataset of 14,265 KUB x-ray images, enabling it to capture intricate patterns effectively.

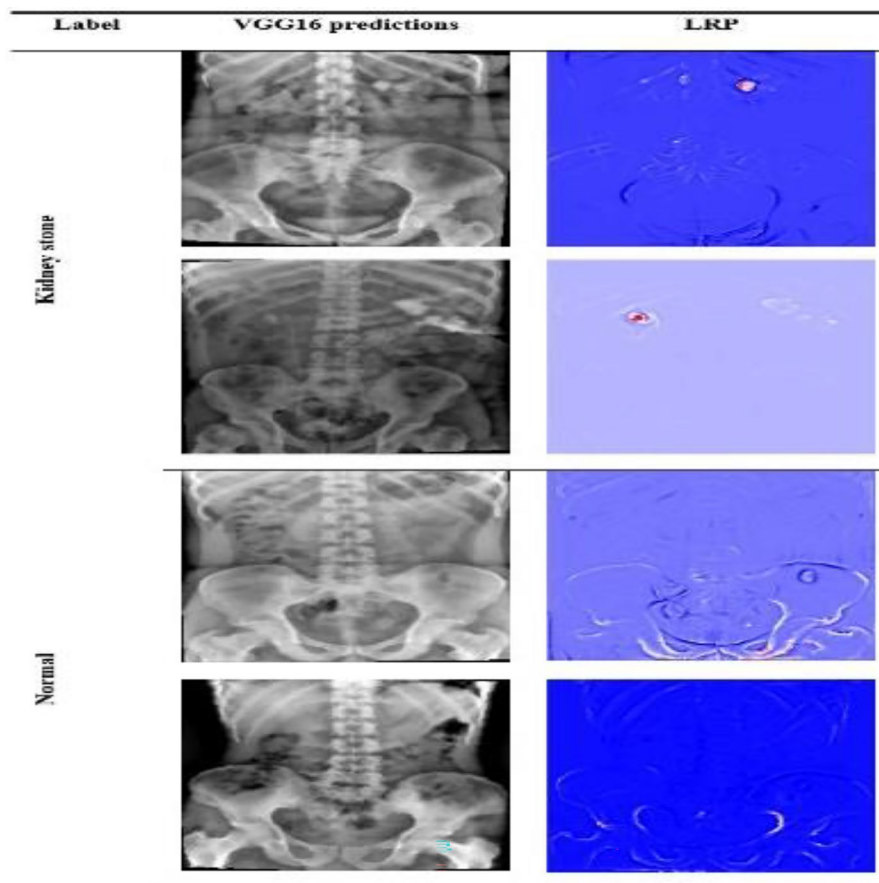


Figure 6. Explanations based on LRP for the modified VGG16 predictions.

Conclusion

Kidney stone formation can lead to a significant obstruction in renal function, consequently affecting human health and survival. As a result, the prompt identification and prediction of kidney stones assume critical importance. Recent technological advancements have enabled the broad integration of ML and DL methodologies into diagnosing kidney stones. In this study, we introduced and used a modified VGG16 model to identify kidney stones in KUB x-ray images. The results of our experiments show that the modified VGG16 model has an accuracy of 97.41% in identifying kidney stones within KUB x-ray images.

DL models like VGG16 can be perceived as “black boxes” because they lack transparency or prediction fairness. In addressing this issue, the study employs the XAI technique LRP to elucidate the model’s predictions, enhancing users’ comprehension of the rationale behind the decision-making process. This approach provides a transparent and effective solution for arriving at definitive diagnostic conclusions, reducing the time needed for diagnosis and enhancing diagnostic accuracy.

Limitations and future work

One of the critical limitations of our research is the availability of high-quality and diverse medical image data of KUB X-rays of kidney stones. The quality and diversity of the dataset are crucial in identifying kidney stones. In the future, overcoming this limitation will require continued efforts to collect, curate, and make a broader range of medical image data more readily available to improve model performance.

Even using XAI techniques such as LRP, the model’s interpretation may still be inconspicuous or might not give meaningful insight into the model’s decision-making. In the future, further research in advanced XAI techniques and methodologies will have the potential to visually enhance the transparency, fairness, and interpretability of the model’s predictions, allowing users to understand better and trust the model.

The development of AI-based medical diagnosis enables personalized and science-based approaches to medical care. However, ethical considerations must be carefully weighed; strategies must be developed to mitigate patient privacy data security and algorithmic bias and to minimize unintended consequences of AI-based medical diagnosis. Blockchain technology can address patient privacy and data security in future work by providing decentralized storage and secure access controls for patient data. With blockchain, the training of the AI models is transparent and auditable, improving algorithmic bias and enabling accountability, which is a cornerstone of trusted AI-based medical diagnosis.

Author	Year	Dataset type	Technique	Use of XAI	Accuracy	Misclassification rate
Chiang et al. ⁴⁷	2003	105 healthy controls and 151 patients with calcium oxalate stones	ANN and DA	No	ANN: 89% and DA: 75%	ANN: 11% and DA: 25%
Dussol et al. ⁴⁸	2006	119 stone formers and 96 controls	ANN (LDA and MVDA)	No	LDA: 75.8% and MVDA: 74.4%	LDA: 24.2% and MVDA: 25.6%
Cauderella et al. ⁴⁹	2011	80 patient's data	ANN and LR	No	ANN: 88.8% and LR: 67.5%	ANN: 11.2% and LR: 32.5%
Kumar and Abhishek ⁵⁰	2012	Data from 1000 patients	LVQ, MLP, and RBF	No	LVQ: 84%, MLP: 92%, and RBF: 87%	LVQ: 16%, MLP: 8% and RBF: 13%
Ebrahimi and Mariano ⁵¹	2015	KUB CT scan slides from 39 patients	Image processing techniques and geometry principles	No	84.61%	15.39%
Kazemi and Mirroshandel ⁵²	2018	Numeric characteristics from 936 patients	Ensemble learning model	No	97.1%	2.9%
Li and Elliot ⁵³	2019	1874 CT KUB reports	NLP	No	85%	15%
De Perrot et al. ⁵⁴	2019	416 patient data	ML model	No	85.1%	14.9%
Kahani et al. ⁵⁵	2020	KUB x-ray images	LASSO with ML classifiers	No	96%	4%
Jungmann et al. ⁵⁶	2020	1714 LDCT images	NLP	No	72%	28%
Annameti Rohith et al. ⁵⁷	2021	114 ultrasound images	Median and rank filters	No	Median filter: 86.4%, and rank filter: 82.2%	Median filter: 13.6%, and rank filter: 17.8%
Suresh and Abhishek ⁵⁸	2021	KUB ultrasound images	Image processing techniques	No	92.57%	7.43%
Jendenber et al. ⁵⁹	2021	NCCT images of 341 patients containing a distal ureteral stone, phlebolith, or both	CNN	No	92%	8%
Cui et al. ⁶⁰	2021	625 CT images	DL and threshold-based model	No	90.30%	9.70%
Yildirim et al. ⁶¹	2021	1799 coronal CT images	XResNet-50	No	96.82%	3.18%
Tsitsiflis et al. ⁶²	2022	Medical data of 716 patients	ANN	No	81.43%	18.57%
Valencia et al. ⁶³	2022	CT scans of around 40 patients	Image processing	No	92.5%	7.5%
Proposed model	2024	14,265 KUB x-ray images	VGG16 empowered with LRP	Yes	97.41%	2.59%

Table 5. Comparison of the proposed model with state-of-the-art literature.

The current study focused on developing and evaluating the proposed model. In the future, the proposed model's computational complexity and resource requirements will be analyzed to determine its size.

Data availability

The dataset & Simulation files used during the current study are available from the corresponding author upon reasonable request.

Code availability

Current code version: V1.0. Permanent link to code/repository used for this code version: https://github.com/deepfindr/xai-series/blob/master/05_lrp.py. Software code languages, tools and services used: Python 3 (Used in Google Colab). Compilation requirements, operating environments, and dependencies: CUDA GPU, Microsoft Windows. Support email for questions: ara4013@qatar-med.cornell.edu , adnan@gachon.ac.kr.

Received: 19 October 2023; Accepted: 6 March 2024

Published online: 14 March 2024

References

- Lang, J. *et al.* Global trends in incidence and burden of urolithiasis from 1990 to 2019: An analysis of global burden of disease study data. *Eur. Urol. Open Sci.* **35**, 37–46. <https://doi.org/10.1016/j.euros.2021.10.008> (2022).
- Vineela, T., Akhila, R. V. G. L., Anusha, T., Nandini, Y. & Bindu, S. Kidney stone analysis using digital image processing. *Int. J. Res. Eng. Sci. Manag.* **3**(3), 275–278 (2020).
- Alelign, T. & Petros, B. Kidney stone disease: An update on current concepts. *Adv. Urol.* **2018**, 1–12 (2018).
- Solie, I. & Situm, M. Kidney stones: Is there a way to see them better? In *2022 7th International Conference on Smart and Sustainable Technologies (SpliTech), Split/Bol, Croatia* 9–11 (2022).
- Caglayan, A., Horsanali, M. O., Kocadurdu, K., Ismailoglu, E. & Guneyli, S. Deep learning model-assisted detection of kidney stones on computed tomography. *Int. Braz. J. Urol.* **48**(5), 830–839. <https://doi.org/10.1590/S1677-5538.IBJU.2022.0132> (2022).
- Vinoth, R. & Bommannaraja, K. FPGA design of efficient kidney image classification using algebraic histogram feature model and sparse deep neural network (SDNN) techniques. In *2017 Conference on Emerging Devices and Smart Systems (ICEDSS), Mallasamudram, India* 1–6 (2017). <https://doi.org/10.1109/ICEDSS.2017.8073687>.
- Kidney Stones: Symptoms, Diagnosis & Treatment - Urology Care Foundation. Accessed Sep 19, 2023. [Online]. <https://www.urologyhealth.org/urology-a-z/k/kidney-stones>
- Krambeck, A. E. *et al.* Effect of age on the clinical presentation of incident symptomatic urolithiasis in the general population. *J. Urol.* **189**(1), 158–164. <https://doi.org/10.1016/j.juro.2012.09.023> (2013).

9. Ulsan, S., Koc, Z. & Tokmak, N. Accuracy of sonography for detection renal stone: Comparison with CT. *J. Clin. Ultrasound* **35**(5), 256–261. <https://doi.org/10.1002/jcu> (2007).
10. Odenrick, A., Kartalis, N., Voulgarakis, N., Morsbach, F. & Loizou, L. The role of contrast-enhanced computed tomography to detect renal stones. *Abdom. Radiol.* **44**(2), 652–660. <https://doi.org/10.1007/s00261-018-1778-7> (2019).
11. Sandhu, C., Anson, K. M. & Patel, U. Urinary tract stones—Part I: Role of radiological imaging in diagnosis and treatment planning. *Clin. Radiol.* **58**(6), 415–421. [https://doi.org/10.1016/S0009-9260\(03\)00103-X](https://doi.org/10.1016/S0009-9260(03)00103-X) (2003).
12. Thomson, J. M. Z., Glocer, J., Abbott, C., Maling, T. M. J. & Mark, S. Computed tomography versus intravenous urography in diagnosis of acute flank pain from urolithiasis: A randomized study comparing imaging costs and radiation dose. *Diagnostic Radiol.* **45**(3), 291–297 (2001).
13. Kim, H. S., Kim, E. J. & Kim, J. Y. Emerging trends in artificial intelligence-based urological imaging technologies and practical applications. *Int. Neurorol. J.* **27**(Suppl 2), S74–S81. <https://doi.org/10.5213/inj.2346286.143> (2023).
14. Kaur, M., Gianey, H. K., Singh, D., Sabharwal, M. & Science, C. Multi-objective differential evolution based random forest for e-health applications. *Mod. Phys. Lett. B* **33**(5), 1–13. <https://doi.org/10.1142/S0217984919500222> (2019).
15. Rahmani, A. M. *et al.* Machine learning (ML) in medicine: Review, applications, and challenges. *Mathematics* **9**(22), 1–52. <https://doi.org/10.3390/math9222970> (2021).
16. Dong, S., Wang, P. & Abbas, K. A survey on deep learning and its applications. *Comput. Sci. Rev.* **40**, 1–22. <https://doi.org/10.1016/j.cosrev.2021.100379> (2021).
17. Abdullah, A. A., Hassan, M. M. & Mustafa, Y. T. A review on bayesian deep learning in healthcare: Applications and challenges. *IEEE Access* **10**, 36538–36562. <https://doi.org/10.1109/ACCESS.2022.3163384> (2022).
18. Rao, T. V. N., Gaddam, A., Kurni, M. & Saritha, K. Reliance on artificial intelligence, machine learning and deep learning in the era of industry 4.0. *Smart Healthc. Syst. Des.* <https://doi.org/10.1002/9781119792253.ch12> (2022).
19. Alaskar, H. *et al.* Deep learning approaches for automatic localization in medical images. *Comput. Intell. Neurosci.* **2022**, 1–17. <https://doi.org/10.1155/2022/6347307> (2022).
20. Yan, K., Wang, X., Lu, L. & Summers, R. M. DeepLesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J. Med. Imaging* **5**(03), 1–11. <https://doi.org/10.1117/1.jmi.5.3.036501> (2018).
21. Kijowski, R., Liu, F., Caliva, F. & Pedoia, V. Deep learning for lesion detection, progression, and prediction of musculoskeletal disease. *J. Magn. Reson. Imaging* **52**(6), 1607–1619. <https://doi.org/10.1002/jmri.27001> (2020).
22. Talo, M., Baloglu, U. B., Yildirim, Ö. & Rajendra Acharya, U. Application of deep transfer learning for automated brain abnormality classification using MR images. *Cogn. Syst. Res.* **54**, 176–188. <https://doi.org/10.1016/j.cogsys.2018.12.007> (2019).
23. Ozturk, T. *et al.* Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput. Biol. Med.* **121**, 1–11. <https://doi.org/10.1016/j.compbiomed.2020.103792> (2020).
24. Kott, O. *et al.* Development of a deep learning algorithm for the histopathologic diagnosis and gleason grading of prostate cancer biopsies: A pilot study. *Eur. Urol. Focus* **7**(2), 347–351. <https://doi.org/10.1016/j.euf.2019.11.003> (2021).
25. Shkolyar, E. *et al.* Augmented bladder tumor detection using deep learning. *Eur. Urol.* **76**(6), 714–718. <https://doi.org/10.1016/j.eururo.2019.08.032> (2019).
26. Zhang, J., He, Y., Chen, W., Kuang, L.-D. & Zheng, B. CorrFormer: Context-aware tracking with cross-correlation and transformer. *Comput. Electr. Eng.* **114**, 109075. <https://doi.org/10.1016/j.compeleceng.2024.109075> (2024).
27. Chen, Y., Xia, R., Yang, K. & Zou, K. MICU: Image super-resolution via multi-level information compensation and U-net. *Expert Syst. Appl.* **245**, 123111. <https://doi.org/10.1016/j.eswa.2023.123111> (2024).
28. Chen, Y., Xia, R., Yang, K. & Zou, K. GCAM: lightweight image inpainting via group convolution and attention mechanism. *Int. J. Mach. Learn. Cybern.* <https://doi.org/10.1007/s13042-023-01999-z> (2023).
29. Chen, Y., Xia, R., Yang, K. & Zou, K. DGCA: High resolution image inpainting via DR-GAN and contextual attention. *Multimed. Tools Appl.* **82**(30), 47751–47771. <https://doi.org/10.1007/s11042-023-15313-0> (2023).
30. Chen, Y., Xia, R., Yang, K. & Zou, K. DARGs: Image inpainting algorithm via deep attention residuals group and semantics. *J. King Saud Univ. Comput. Inf. Sci.* **35**(6), 101567. <https://doi.org/10.1016/j.jksuci.2023.101567> (2023).
31. Chen, Y., Xia, R., Yang, K. & Zou, K. MFMAM: Image inpainting via multi-scale feature module with attention module. *Comput. Vis. Image Underst.* **238**(January 2024), 103883. <https://doi.org/10.1016/j.cviu.2023.103883> (2024).
32. Chen, Y., Xia, R., Yang, K. & Zou, K. DNNAM: Image inpainting algorithm via deep neural networks and attention mechanism. *Appl. Soft Comput.* **154**, 111392. <https://doi.org/10.1016/j.asoc.2024.111392> (2024).
33. Hesamian, M. H., Jia, W., He, X. & Kennedy, P. Deep learning techniques for medical image segmentation: Achievements and challenges. *J. Digit. Imaging* **32**(4), 582–596. <https://doi.org/10.1007/s10278-019-00227-x> (2019).
34. Roth, H. R. *et al.* Deep learning and its application to medical image segmentation. *Med. IMAGING Technol.* **36**(2), 63–71. <https://doi.org/10.1149/mit.36.63> (2018).
35. Shinde, P. P. & Shah, D. S. A review of machine learning and deep learning applications. In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India* 1–6 (IEEE, 2018).
36. Zhang, J., Lv, Y., Tao, J., Huang, F. & Zhang, J. A robust real-time anchor-free traffic sign detector with one-level feature. *IEEE Trans. Emerg. Top. Comput. Intell.* <https://doi.org/10.1109/TETCI.2024.3349464> (2024).
37. Islam, U., Al-atawi, A. A. & Alwageed, H. S. Detection of renal cell hydronephrosis in ultrasound kidney images: A study on the efficacy of deep convolutional neural networks. *PeerJ Comput. Sci.* **10**, 1–28. <https://doi.org/10.7717/peerj-cs.1797> (2024).
38. Kaur, M. & Singh, D. Fusion of medical images using deep belief networks. *Clust. Comput.* **23**(2), 1439–1453. <https://doi.org/10.1007/s10586-019-02999-x> (2020).
39. Jiang, F. *et al.* Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc. Neurol.* **2**(4), 230–243. <https://doi.org/10.1136/svn-2017-000101> (2017).
40. Abualigah, L. *et al.* Wind, solar, and photovoltaic renewable energy systems with and without energy storage optimization: A survey of advanced machine learning and deep learning techniques. *Energies* **15**(2), 1–26. <https://doi.org/10.3390/en15020578> (2022).
41. Li, Y. *et al.* CDRIME-MTIS: An enhanced rime optimization-driven multi-threshold segmentation for COVID-19 X-ray images. *Comput. Biol. Med.* **169**(February 2024), 107838. <https://doi.org/10.1016/j.compbiomed.2023.107838> (2024).
42. Shamrat, F. J. M. *et al.* High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. *Comput. Biol. Med.* **155**(January), 106646. <https://doi.org/10.1016/j.compbiomed.2023.106646> (2023).
43. Tao, S. *et al.* Phase retrieval for X-ray differential phase contrast radiography with knowledge transfer learning from virtual differential absorption model. *Comput. Biol. Med.* **168**(January 2024), 107711. <https://doi.org/10.1016/j.compbiomed.2023.107711> (2024).
44. Sitaula, C. & Hossain, M. B. Attention-based VGG-16 model for COVID-19 chest X-ray image classification. *Appl. Intell.* **51**, 2850–2863. <https://doi.org/10.1007/s10489-020-02055-x> (2021).
45. Naveen, P. & Diwan, B. Pre-trained VGG-16 with CNN Architecture to classify X-Rays images into Normal or Pneumonia. In *2021 International Conference on Emerging Smart Computing and Informatics (ESCI), IEEE, Pune, India* 102–105 (2021). <https://doi.org/10.1109/ESCI50559.2021.9396997>.
46. Ishioka, J. *et al.* Computer-aided diagnosis with a convolutional neural network algorithm for automated detection of urinary tract stones using kub. *J. Urol.* **201**(4), e845 (2019).
47. Chiang, D., Chiang, H. C., Chen, W. C. & Tsai, F. J. Prediction of stone disease by discriminant analysis and artificial neural networks in genetic polymorphisms: A new method. *BJU Int.* **91**(7), 661–666. <https://doi.org/10.1046/j.1464-410X.2003.03067.x> (2003).

48. Dussol, B., Verdier, J. M., Le Goff, J. M., Berthezene, P. & Berland, Y. Artificial neural networks for assessing the risk of urinary calcium stone among men. *Urol. Res.* **34**(1), 17–25. <https://doi.org/10.1007/s00240-005-0006-4> (2006).
49. 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.* **83**(1), 14–19 (2011).
50. Kumar, K. & Abhishek, A. Artificial neural networks for diagnosis of kidney stones disease. *Int. J. Inf. Technol. Comput. Sci.* **4**(7), 20–25. <https://doi.org/10.5815/ijitcs.2012.07.03> (2012).
51. Ebrahimi, S. & Mariano, V. Y. Image Quality Improvement in Kidney Stone Detection on Computed Tomography Images. *J. Image Graph.* **3**(1), 40–46. <https://doi.org/10.18178/joig.3.1.40-46> (2015).
52. Kazemi, Y. & Mirroshandel, S. A. A novel method for predicting kidney stone type using ensemble learning. *Artif. Intell. Med.* **84**, 117–126. <https://doi.org/10.1016/j.artmed.2017.12.001> (2018).
53. Li, A. Y. & Elliot, N. Natural language processing to identify ureteric stones in radiology reports. *J. Med. Imaging Radiat. Oncol.* **63**(3), 307–310. <https://doi.org/10.1111/1754-9485.12861> (2019).
54. De Perrot, T. *et al.* Differentiating kidney stones from phleboliths in unenhanced low-dose computed tomography using radiomics and machine learning. *Eur. Radiol.* **29**, 4776–4782 (2019).
55. Kahani, M., Hariri Tabrizi, S., Kamali-Asl, A. & Hashemi, S. A novel approach to classify urinary stones using dual-energy kidney, ureter and bladder (DEKUB) X-ray imaging. *Appl. Radiat. Isot.* **164**(May 2019), 109267. <https://doi.org/10.1016/j.apradiso.2020.109267> (2020).
56. Jungmann, F. *et al.* Towards data-driven medical imaging using natural language processing in patients with suspected urolithiasis. *Int. J. Med. Inform.* **137**, 1–21. <https://doi.org/10.1016/j.ijmedinf.2020.104106> (2020).
57. Rohith, A. Detection of kidney stones in ultrasound images using median filter compared with rank filter. *Rev. Gestão Inovação e Tecnol.* **11**(4), 1096–1111. <https://doi.org/10.47059/revistageintec.v11i4.2171> (2021).
58. Suresh, M. B. & Abhishek, M. R. Kidney stone detection using digital image processing techniques. In *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021, Coimbatore, India* 556–561 (IEEE, 2021). <https://doi.org/10.1109/ICIRCA51532.2021.9544610>.
59. Jendeberg, J., Thunberg, P. & Lidén, M. Differentiation of distal ureteral stones and pelvic phleboliths using a convolutional neural network. *Urolithiasis* **49**(1), 41–49. <https://doi.org/10.1007/s00240-020-01180-z> (2021).
60. Cui, Y. *et al.* Automatic detection and scoring of kidney stones on noncontrast CT images using S.T.O.N.E. nephrolithometry: Combined deep learning and thresholding methods. *Mol. Imaging Biol.* **23**(3), 436–445. <https://doi.org/10.1007/s11307-020-01554-0> (2021).
61. Yildirim, K. *et al.* Deep learning model for automated kidney stone detection using coronal CT images. *Comput. Biol. Med.* **135**, 1–7. <https://doi.org/10.1016/j.compbiomed.2021.104569> (2021).
62. Tsitsiflis, A. *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.* **9**(2), 132–138. <https://doi.org/10.1016/j.ajur.2021.09.005> (2022).
63. Valencia, F. A. C., Muñoz, J. J. A. M. & Montaña, F. M. J. Stone detection in kidney with image processing technique: CT images. *J. Posit. Sch. Psychol.* **6**(6), 7643–7653 (2022).
64. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C. & Liu, C. A survey on deep transfer learning. In *27th International Conference on Artificial Neural Networks, Rhodes, Greece* 270–279 (Springer, 2018). <https://doi.org/10.1007/978-3-030-01424-7>.
65. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In *Published as a Conference Paper at ICLR* 1–14 (2015).
66. Althubiti, S. A., Alenezi, F., Shitharth, S., Sangeetha, K. & Reddy, C. V. S. Circuit manufacturing defect detection using VGG16 convolutional neural networks. *Wirel. Commun. Mob. Comput.* **2022**, 1–10. <https://doi.org/10.1155/2022/1070405> (2022).
67. Doshi-Velez, F. & Kim, B. Towards a rigorous science of interpretable machine learning. arXiv Prepr., 1–13 (2017).
68. Bach, S. *et al.* On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE* **10**(7), 1–46. <https://doi.org/10.1371/journal.pone.0130140> (2015).
69. Huang, X., Jamonnak, S., Zhao, Y., Wu, T. H. & Xu, W. A visual designer of layer-wise relevance propagation models. *Eurographics Conf. Vis.* **40**(3), 227–238 (2021).
70. Seliya, N., Khoshgoftaar, T. M. & Van Hulse, J. A study on the relationships of classifier performance metrics. In *21st IEEE International Conference on Tools with Artificial Intelligence, Newark, NJ, USA* 59–66 (2009). <https://doi.org/10.1109/ICTAI.2009.25>.
71. Ahmed, F., Khan, W. A., Iqbal, M., Riad, A., Abazeed, A., Alrababah, H. & Khan, M. F. Rock-paper-scissors image classification using transfer learning. In *2023 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates* 1–6 (IEEE, 2023). <https://doi.org/10.1109/ICBATS57792.2023.10111433>.

Author contributions

F.A., A.A., and T.S. have collected data from different resources. S.A., F.A., and M.A.K. performed formal analysis and Simulation, S.A., W.A.K., M.A., and T.S. contributed to writing—original draft preparation, A.A., M.A.K., A.A., and W.A.K.; writing—review and editing, S.A., M.A.K., and A.A.; performed supervision, A.A., M.A., and F.A.; drafted pictures and tables, A.A., F.A., and S.A.; performed revisions and improve the quality of the draft. All authors have read and agreed to the published version of the manuscript.

Funding

This research work is supported by Qatar National Library.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.A.K. or A.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024