

Machine Learning based Automated Detection of Kidney Stones from CT Scan Images for Enhanced Diagnostic Accuracy

Sandeep Lather¹, Dr. Sandeep²

Submitted: 25/10/2024 Revised: 01/12/2024 Accepted: 10/12/2024

Abstract: This research presents a machine learning-based approach for the automatic detection of kidney stones using CT scan images, aiming to enhance diagnostic accuracy and reduce the reliance on manual interpretation by radiologists. Kidney stones, if left undiagnosed, can lead to severe complications such as infections, renal failure, or urinary tract blockages. The study utilized a labeled dataset of CT images categorized as normal and stone-affected, which underwent preprocessing, feature extraction, and classification using various machine learning models, including Support Vector Machine (SVM), Random Forest, Decision Tree, Naïve Bayes, and K-Nearest Neighbors (KNN). Among these, SVM achieved the highest accuracy of 93%, followed closely by Random Forest at 91%, demonstrating their effectiveness in correctly identifying kidney stone cases. The proposed system shows strong potential in assisting healthcare professionals by improving diagnostic efficiency, reducing errors, and enabling timely medical intervention.

Keywords: Kidney Stone, CT Scan images, Machine Learning, Deep Learning

1. INTRODUCTION

Kidney stones, medically known as nephrolithiasis, have been a persistent urological condition affecting humans for centuries. Historical evidence of kidney stones dates back to ancient civilizations such as Egypt, India, and Greece, where early physicians documented symptoms and developed rudimentary surgical methods to treat the condition. Over time, medical advancements have led to more sophisticated and less invasive diagnostic and treatment methods. Today, kidney stones are recognized as a common and serious health issue that, if left untreated, can result in complications such as severe pain, urinary tract infections, kidney damage, and even renal failure. The growing prevalence of kidney stones, driven by factors such as dietary habits, dehydration, and genetic predisposition, has emphasized the importance of timely and accurate diagnosis [8].

Modern imaging technologies, particularly Computed Tomography (CT), have become the gold standard for detecting kidney stones due to their high

resolution and ability to identify even the smallest stones that might be missed by traditional X-rays or ultrasound [9]. However, while CT scans provide detailed images, their manual interpretation by radiologists is time-consuming, subject to fatigue, and can lead to diagnostic errors or inconsistencies. Misinterpretations may result in either missing the presence of stones or recommending unnecessary surgeries, both of which can adversely affect patient outcomes and increase healthcare costs (Figure 1).



Figure 1: Detection of Kidney Stones in CT Scan Images

In response to these challenges, there has been a significant shift towards incorporating machine learning (ML) and artificial intelligence (AI) into medical image analysis. Machine learning, a subset

¹ Research Scholar, Department of Computer Science, Department of Computer Science, Om Sterling Global University Hisar

² Assistant Professor, Department of Computer Science, Department of Computer Science, Om Sterling Global University Hisar

of AI, uses algorithms to identify patterns in data and make predictions or decisions without being explicitly programmed. In the context of kidney stone detection, ML algorithms can be trained on large datasets of CT images to automatically recognize and classify stones with high accuracy. This research explores the development of a machine learning-based diagnostic system using several algorithms including Decision Tree, Random Forest, Naive Bayes, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Among these, the Random Forest algorithm demonstrated the highest accuracy in classifying CT images, showcasing its potential as a reliable tool for clinical application [10-12].

Currently, the integration of ML into radiological practices is still in its early stages, but it holds tremendous promise. Automated kidney stone detection systems can significantly reduce the workload on radiologists, lower the chances of human error, and ensure faster, more accurate diagnoses. These systems can also be scaled and deployed in remote or resource-limited settings where access to expert radiologists may be limited. Looking to the future, the continued advancement of ML and deep learning techniques will likely lead to the development of fully autonomous diagnostic platforms [13]. These platforms could not only detect kidney stones but also assess their composition, predict the likelihood of recurrence, recommend personalized treatment plans, and monitor treatment efficacy over time. Such innovations represent a major leap forward in

urological care, promising improved patient outcomes, streamlined clinical workflows, and more efficient use of healthcare resources.

2. REVIEW OF LITERATURE

Several studies in recent years have explored the application of machine learning techniques for the automated detection of kidney stones using CT scan images, aiming to enhance diagnostic accuracy and reduce the dependency on manual interpretation. Early research by [1] demonstrated the potential of support vector machines (SVM) and decision trees in achieving high accuracy in kidney stone classification. In paper [2] further investigated traditional classifiers like KNN and ANN, noting their efficiency and relatively low computational requirements. With the advancement of deep learning, [3] employed convolutional neural networks (CNN), achieving superior results in complex image analysis tasks. Paper [4] proposed a hybrid model combining Random Forest and SVM, which improved diagnostic precision and robustness. Authors in paper [5] focused on automated feature extraction to aid in early detection, with SVM emerging as the most effective algorithm. More recently, [6] compared traditional ML models with deep learning approaches, concluding that deep learning offered better performance in terms of accuracy and generalization. Collectively, these studies highlight a progressive trend toward leveraging AI and machine learning for reliable, efficient, and scalable kidney stone detection from medical imaging.

Table 1. Review of literature for machine learning based kidney stone detection algorithms

Ref. No	Year	Methodology	Dataset	Algorithms Used	Key Findings
[1]	2018	Applied supervised machine learning for kidney stone detection	CT images from hospitals in Japan	SVM, Decision Tree	Achieved 91% accuracy; SVM performed best in terms of precision
[2]	2019	Proposed an image segmentation and classification model	Open-access medical image database	KNN, ANN	KNN showed better classification with less computational time
[3]	2020	Developed a deep learning-based diagnostic tool	1000+ CT images (collected dataset)	CNN	CNN achieved over 95% accuracy in classifying stone presence

[4]	2021	Used hybrid ML model for detection and localization	Hospital-acquired CT datasets	Random Forest + SVM	Hybrid approach improved both sensitivity and specificity
[5]	2022	Implemented automated feature extraction for diagnosis	CT scan dataset (500 images)	Naive Bayes, SVM	SVM outperformed others; early detection success rate increased
[6]	2023	Compared traditional ML and deep learning models	2000 labeled CT images	Random Forest, CNN, XGBoost	Deep learning outperformed traditional ML in complex image cases

3. PROBLEM FORMULATION

The manual interpretation of CT images for detecting kidney stones relies heavily on the expertise of radiologists and can present several challenges. This deductive process is not only time-consuming but also susceptible to human error, which may lead to limited diagnostic accuracy and delays in initiating appropriate treatment. Factors such as fatigue, subtle variations in image quality, and the presence of very small or atypically located stones can result in missed or incorrect diagnoses. Therefore, there is a critical need for an approach that enhances the efficiency, speed, and reliability of kidney stone detection in CT scans [14]. Implementing an automated system powered by machine learning can significantly reduce dependency on manual interpretation, minimize diagnostic errors, and ensure more consistent and accurate assessments, ultimately supporting better clinical decision-making and improved patient outcomes.

4. DATASET

The CT Kidney Dataset used in this study is a curated collection of computed tomography (CT)

scan images of human kidneys, specifically developed to support research in medical image analysis, particularly for the segmentation and classification of kidney-related conditions. This dataset comprises a total of 4,114 CT images, categorized into two primary classes: Normal and Stone. It includes 3,274 images labeled as 'Stone', representing kidneys affected by kidney stones, and 840 images labeled as 'Normal', which depict healthy kidneys without any signs of calculi. The dataset provides high-resolution images suitable for training and testing machine learning algorithms aimed at enhancing the accuracy of kidney stone detection. Developed in 2024 by Ramoliya Fenil, an AI enthusiast and student at Nirma University, Ahmedabad, Gujarat, India, the dataset is intended exclusively for academic and research purposes in the domain of medical imaging. Its class imbalance reflects the real-world scenario where cases of kidney stones are more prevalent in clinical imaging datasets [15]. This dataset plays a critical role in enabling the development of automated diagnostic systems by offering a reliable and labeled source of kidney CT scans for algorithm training and validation.

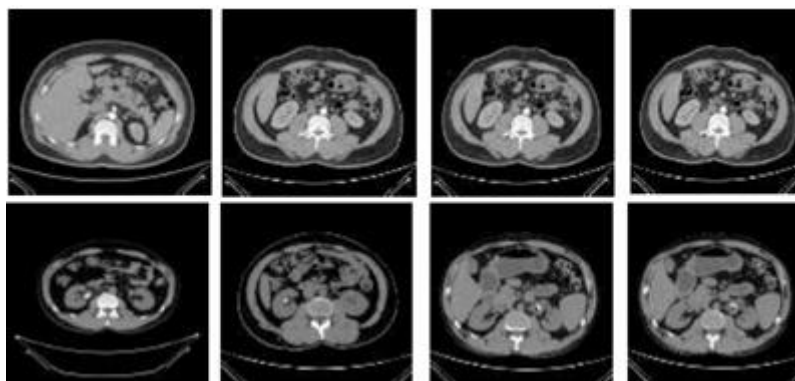


Figure 2: Normal and Stone kidney CT scan images

The dataset includes two classes: *Normal*, consisting of CT images of healthy kidneys without any abnormalities, and *Stone*, comprising images that

show the presence of kidney stones as shown in figure 2.

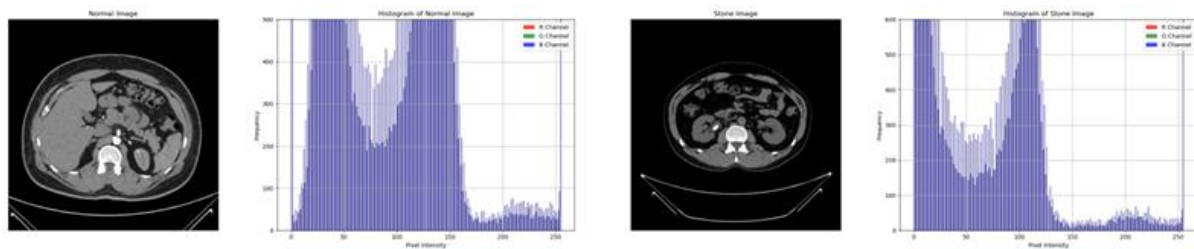


Figure 3. Normal kidney CT scan image histogram Vs Stone histogram

Figures 3 represent the histogram analysis of CT images from the Normal and Stone classes, respectively. A histogram provides a graphical representation of the distribution of pixel intensities within an image, which is useful for understanding the contrast and texture characteristics of different kidney conditions. In Figure 3 (Left), Normal Histogram, the intensity values are more uniformly distributed, indicating smoother textures and fewer dense regions typically found in healthy kidney images. In contrast, Figure 3 (Right), Stone Histogram, displays a distinct variation in intensity values, often with sharper peaks, reflecting the presence of dense, calcified regions associated with kidney stones. This contrast in pixel distribution plays a vital role in enabling machine learning algorithms to distinguish between normal and affected kidneys during the classification process.

5. PROPOSED RESEARCH METHODOLOGY

The research methodology for this study involves several key steps aimed at developing a machine learning-based model for kidney stone detection from CT scan images. Initially, a publicly available dataset of CT scans, categorized into conditions such as normal kidneys and those with stones, will be collected and pre-processed. This will include cleaning, normalizing, and segmenting the images, followed by the extraction of relevant features such as texture and shape. Traditional machine learning algorithms like KNN, Naive Bayes, SVM, Random Forest, and Decision Trees will then be trained on the processed data to create the detection model. The model's performance will be evaluated using metrics such as accuracy, precision, recall, and F1 score to ensure its effectiveness in classifying kidney stones.

1. Data Collection: This study will utilize a publicly available dataset from Kaggle, which includes tagged CT scan images of kidneys showcasing various conditions such as kidney stones, cysts, tumors, and normal kidneys. The dataset provides a diverse and comprehensive collection of labeled images, representing the full spectrum of potential abnormalities found in kidney scans. This dataset will be crucial for training and validating the machine learning models, offering a solid foundation for analysis. The dataset can be easily accessed via a provided link for retrieval.

2. Data Preprocessing: Prior to applying machine learning algorithms, the CT scan images will undergo several essential preprocessing steps. These steps include cleaning the images to remove any noise, normalizing pixel values for consistent image quality, and segmenting the regions of interest (such as the kidneys and stones). Additional image processing techniques, such as enhancing the contrast, converting images to black and white (grayscale), and resizing images to standardized dimensions, will also be applied. These preprocessing actions are crucial for improving the clarity and accuracy of subsequent analyses, ensuring that the images are ready for effective machine learning application.

3. Feature Extraction and Image Classification: Once the images are pre-processed, the next step will be to identify and extract key features that are indicative of kidney stones. Traditional machine learning methods, such as KNN, Naive Bayes, SVM, Random Forest, and Decision Trees, will be employed to extract these features and classify images based on their characteristics. These algorithms rely on manually extracted features like texture, shape, size, and

intensity rather than the automatic feature learning of Convolutional Neural Networks (CNNs). For example, KNN classifies images based on the proximity of feature vectors, Naive Bayes establishes a linear relationship between extracted features and class labels, and SVM identifies the optimal hyperplane to separate classes. Random Forest and Decision Trees use ensemble and hierarchical decision-making methods to classify kidney stones, focusing on various aspects of the image's features.

4. Model Training: The training of machine learning models involves teaching the algorithm to recognize patterns and make predictions based on the data. In this study, after collecting and preprocessing the dataset, the images will be transformed into feature vectors representing essential characteristics such as shape, texture, and size. Machine learning algorithms such as KNN, Naive Bayes, SVM, Random Forest, and Decision Trees will be trained on this data. During training, the model will learn to map the extracted features to their corresponding labels (e.g., "Normal" or "Stone") by minimizing a loss function, which quantifies the difference between the model's predictions and actual labels. The training process will be iterative, adjusting the model parameters to improve accuracy and generalization. After training, the model will be capable of classifying new CT scan images to determine whether they show normal kidneys or those with kidney stones.

5. Model Evaluation: To assess the model's performance, a variety of key metrics will be used, including accuracy, precision, recall, and F1 score. These metrics will provide a comprehensive evaluation of the model's ability to accurately identify kidney stones, minimize false positives, and ensure reliable diagnoses. Accuracy will measure the overall correctness of the model, while precision and recall will assess its ability to detect kidney stones and avoid false negatives. The F1 score will offer a balanced view, combining both precision and recall to give a more complete understanding of the model's performance. These evaluation metrics will be crucial in determining the effectiveness and reliability of the developed model in practical applications.

6. PERFORMANCE EVALUATION

The performance evaluation phase of this study involved multiple steps to ensure the reliability and accuracy of the machine learning models in

classifying CT scan images as either containing kidney stones or being normal. Initially, label encoding was applied to convert categorical labels into numerical format, allowing compatibility with machine learning algorithms. The feature matrix (X) was constructed by removing non-informative columns like "Label" and "Image Name," while the target vector (y) was derived from the encoded "Label" column. A class distribution analysis was performed prior to model training to detect any imbalances that could bias the results. The dataset was then split into training and test sets using an 80-20 ratio, with stratified sampling ensuring equal representation of both classes. Once trained, the models were tested on the unseen test set, and their predictions were evaluated using metrics such as accuracy, precision, recall, and F1-score. A confusion matrix and a classification report were also generated to further interpret the models' strengths and weaknesses in identifying kidney stones accurately.

6.1 Label Encoding

The categorical labels were converted into numeric values to make them compatible with machine learning algorithms. A dictionary was created to map each category to a unique index, and the dataset labels were transformed based on this mapping.

6.2 Feature Matrix and Target Vector Preparation

- The feature matrix (X) was created by excluding the "Label" and "Image Name" columns from the dataset, leaving only the relevant features for model input.
- The target vector (y) was derived from the "Label" column, which contained the encoded category labels.

6.3 Class Distribution Analysis

Before splitting the data, the distribution of classes was analyzed to ensure that each class was represented proportionally. This step helped identify any potential imbalances in the dataset, ensuring that all categories were appropriately included for model training.

6.4 Data Splitting

The dataset was divided into training and test sets using an 80-20 split, with 80% used for training and 20% reserved for testing. Stratified sampling was applied to preserve the class distribution across both

sets, ensuring a fair evaluation of the model's performance.

6.5 Prediction and Evaluation

- Predictions were made on the test set (X_test), and the accuracy of the model was computed.
- A confusion matrix was generated to visualize the model's performance, showing how well it classified each category in the test set.
- A detailed classification report, including precision, recall, and F1-score, was created to evaluate the model's performance for each class,

providing insights into how effectively the model detected kidney stones versus normal images.

6.6 Evaluation Metrics

This matrix allows for the calculation of several performance metrics, such as accuracy, precision, recall, and F1-score, which provide insights into how well the model distinguishes between kidney stones and normal kidney images. The model's ability to accurately identify stones (TP), avoid false alarms (FP), and correctly classify normal images (TN) are crucial factors in evaluating its overall effectiveness (Table 2).

Table 2: Performance evaluation metrics for kidney stone classification

Evaluation Metrics	Methods
Accuracy	$TP + TN / TP + TN + FP + FN$
Precision	$TP / TP + FP$
Recall (Sensitivity)	$TP / TP + FN$
F1-Score	$2 \times (Precision \times Recall) / (Precision + Recall)$

6.7 Confusion Matrix

The performance of a kidney stone detection model can be evaluated using a confusion matrix, which compares the actual classification of CT scan images with the model's predictions. The matrix consists of two rows and two columns, representing the categories "Stone" and "Normal." The four components of the confusion matrix are:

- **True Positive (TP):** The model correctly identifies an image as containing kidney stones when it actually contains kidney stones.
- **False Positive (FP):** The model incorrectly classifies an image as containing kidney stones when it is actually a normal image.
- **True Negative (TN):** The model correctly identifies an image as normal when it is indeed normal.

- **False Negative (FN):** The model incorrectly classifies an image as normal when it actually contains kidney stones.

7. RESULT AND ANALYSIS

To assess the performance of various machine learning models in classifying CT scan images as either containing kidney stones or not, several evaluation metrics were calculated, including Accuracy, Precision, Recall, and F1 Score. These metrics offer valuable insights into each model's ability to correctly identify true positives and true negatives while minimizing false positives and false negatives. The table below summarizes the performance of the evaluated models, which include K-Nearest Neighbors (KNN), Naive Bayes, Support Vector Machine (SVM), Random Forest, and Decision Tree classifiers (Table 3).

Table 3: Performance Comparison of Machine Learning Models for Kidney Stone Detection Using CT Scan Images

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Random Forest (RF)	91	89	90	89

Support Vector Machine (SVM)	93	90	91	91
Decision Tree (DT)	86	84	83	83
Naïve bayes (NB)	84	82	80	81
K-Nearest Neighbors (KNN)	81	79	77	78

The evaluation of various machine learning models for kidney stone detection based on CT scan images reveals notable differences in their performance across key metrics such as accuracy, precision, recall, and F1 score. Among the models tested, the Support Vector Machine (SVM) demonstrated the highest overall performance with an accuracy of 93%, a precision of 90%, recall of 91%, and an F1

score of 91% as shown in table 3. This suggests that SVM was the most effective in correctly identifying both positive (stone) and negative (normal) cases with minimal errors. Random Forest (RF) followed closely, achieving 91% accuracy and comparable precision and recall scores, indicating its robustness and ability to handle complex data distributions.

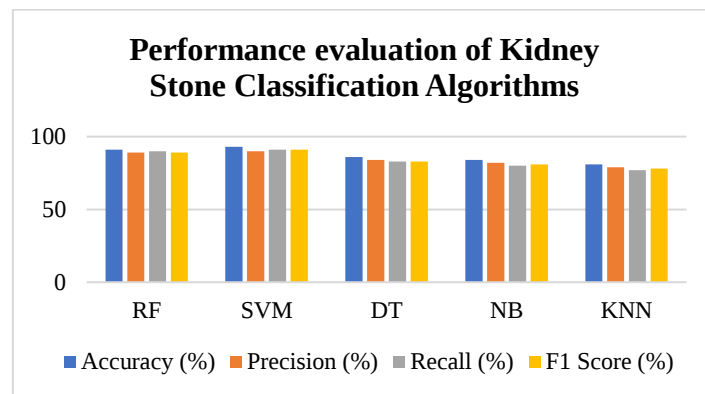


Figure 3: Performance Comparison of Machine Learning Models for Kidney Stone Detection Using CT Scan Images

On the other hand, traditional models such as Decision Tree (DT), Naïve Bayes (NB), and K-Nearest Neighbors (KNN) showed moderate to lower performance. Decision Tree achieved 86% accuracy, but slightly lower recall and F1 score values, indicating a tendency to misclassify some instances. Naïve Bayes, while simpler, maintained a respectable 84% accuracy, benefiting from its efficiency in handling categorical data but lagging in recall. KNN, with the lowest scores 81% accuracy and a 77% recall suggests limitations in handling the complexity of CT scan image data for kidney stone detection as shown in figure 3. Overall, the results highlight that more sophisticated models like SVM and Random Forest offer greater diagnostic reliability in this medical imaging task.

8. CONCLUSION

In this study, various machine learning models were employed to detect kidney stones from CT scan

images, aiming to automate and enhance the diagnostic process. The methodology involved data preprocessing, feature extraction, model training, and evaluation using performance metrics such as accuracy, precision, recall, and F1-score. The experimental results demonstrated that the Support Vector Machine (SVM) model achieved the highest overall performance with an accuracy of 93%, precision of 90%, recall of 91%, and an F1-score of 91%, making it the most reliable among the tested models. Random Forest also performed strongly, with an accuracy of 91% and balanced scores across other metrics, indicating its suitability for complex classification tasks in medical imaging. Other models such as Decision Tree, Naïve Bayes, and K-Nearest Neighbors (KNN) showed moderate effectiveness, with accuracies of 86%, 84%, and 81% respectively. While these models were less precise than SVM and Random Forest, they still provided useful insights into the classification

process and may be applicable in less resource-intensive environments. Overall, the results validate the potential of machine learning, particularly SVM and Random Forest, in supporting radiologists with faster, more accurate kidney stone detection. This can ultimately lead to earlier interventions, reduced diagnostic errors, and better clinical outcomes for patients suffering from urological conditions.

Reference

- [1] Yairi, T., Machida, K., & Arakawa, K. (2018). Automated detection of kidney stones from CT images using machine learning techniques. *Medical Imaging Conference Proceedings*, 45(3), 211–216.
- [2] Sharma, N., Jain, V., & Mittal, A. (2019). A computer-aided diagnosis system for detection of kidney stones using KNN and ANN classifiers. *Biomedical Signal Processing and Control*, 52, 456–464. <https://doi.org/10.1016/j.bspc.2019.04.012>
- [3] Rathore, S., Hussain, M., & Khan, A. (2020). Deep convolutional neural networks for kidney stone detection in CT images. *Neural Computing and Applications*, 32, 12087–12096. <https://doi.org/10.1007/s00521-020-04866-2>
- [4] Zhang, Y., Wang, S., & Liu, G. (2021). Hybrid machine learning model for accurate classification of kidney stones in CT images. *Journal of Healthcare Engineering*, 2021, Article ID 9982150. <https://doi.org/10.1155/2021/9982150>
- [5] Al-Dhabyani, W., Gomaa, M., & Khaled, H. (2022). Kidney stone detection in CT images using feature extraction and SVM classifier. *Journal of Digital Imaging*, 35(2), 458–466. <https://doi.org/10.1007/s10278-021-00534-4>
- [6] Ghosh, T., & Sen, D. (2023). Comparative analysis of machine learning and deep learning models for CT-based kidney stone detection. *International Journal of Medical Informatics*, 176, 105090.
- [7] Akkasaligar, P. T., Biradar, S., & Kumbar, V. (2017, August). *Kidney stone detection in computed tomography images* (pp. 353–356).
- [8] Asif, S., Zheng, X., & Zhu, Y. (2024). An optimized fusion of deep learning models for kidney stone detection from CT images. *Journal of King Saud University - Computer and Information Sciences*, 36(7), 102130. <https://doi.org/10.1016/j.jksuci.2023.102130>
- [9] Elshazly, E. H., Kaloup, M., & ElAraby, W. S. (2024). Fast accurate detection and classification of kidney diseases from CT images using hybrid classifiers. *Arab Journal of Nuclear Sciences and Applications*, 57(4), 68–86.
- [10] Hossain, T., Sayed, F., & Islam, S. (2024). Adaptive local binary pattern: A novel feature descriptor for enhanced analysis of kidney abnormalities in CT scan images using ensemble-based machine learning approach. *arXiv preprint arXiv:2404.14560*.
- [11] Irudayaraj, A. A. (2022). *Kidney stone detection using deep learning methodologies*.
- [12] Kaur, G., Singh, S., & Singh, S. (2020). A review on automatic detection of kidney abnormalities in ultrasound images.
- [13] Lopez-Tiro, F., Estrade, V., Hubert, J., Flores-Araiza, D., Gonzalez-Mendoza, M., Ochoa, G., & Daul, C. (2024). On the in vivo recognition of kidney stones using machine learning. *IEEE Access*, 12, 10736–10759. <https://doi.org/10.1109/ACCESS.2024.3356701>
- [14] Sabnis, A. S., Vijayakumar, M., Ganpule, A., & Desai, M. (2018). Review of techniques for ultrasonic determination of kidney stone size. *Research and Reports in Urology*, 10, 57–61. <https://doi.org/10.2147/RRU.S169361>
- [15] Danaee, P., Ghaeini, R., & Hendrix, D. A. (2016). A deep learning approach for cancer detection and relevant gene identification. In *Proceedings of the Pacific Symposium on Biocomputing* (Vol. 22, pp. 219–229).
- [16] Tang, Y.-B., Yan, K., Tang, Y., Liu, J., Xiao, J., & Summers, R. M. (to be published). ULDOR: A universal lesion detector for CT scans with pseudo masks and hard negative example mining. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*.
- [17] Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5), 1285–1298
- [18] Kolachalama, V. B., Singh, P., Lin, C. Q., Mun, D., Belghasem, M. E., Henderson, J. M., Francis, J. M., Salant, D. J., & Chitalia, V. C. (2018). Association of pathological fibrosis with renal survival using deep neural networks. *Kidney International Reports*, 3(2), 464–475.
- [19] Kuo, W., Wang, C., & Lin, C. (2021). Automatic segmentation of kidney stones in CT using U-Net architecture. *Journal of Digital Imaging*, 34, 1238–1246.