

Non-Invasive Kidney Stone Prediction using Machine Learning: An Extensive Review

Shivani Verma¹, Pawan Kumar Singh¹, Gagandeep Kaur²,
Anannya Vashistha^{2*} and Shreya Pansari²

¹Department of Electronics and Communication Engineering, SRM University,
Sonipat, India, 131029.

²Department of Electronics and Communication Engineering.,
Guru Tegh Bahadur Institute of Technology Delhi India.

*Corresponding author E-mail: ananyavashishtha@gmail.com

<https://dx.doi.org/10.13005/bpj/3072>

(Received: 03 January 2025; accepted: 11 February 2025)

This study investigates the use of machine learning (ML) methods for detecting kidney stones, a field that has gained increasing attention due to limitations in traditional diagnostic methods such as ultrasound and Computed Tomography (CT) scans. The aim of this review is to evaluate different machine learning (ML) algorithms employed to improve the accuracy and efficiency of kidney stone detection, with an emphasis on supervised, unsupervised, and reinforcement learning approaches. Key findings suggest that ML techniques namely Support Vector Machines (SVM), Random Forests (RF), and Deep Learning (DL) algorithms, including the VGG16 (Visual Geometry Group) Convolutional Neural Network (CNN) model, have significantly improved diagnostic accuracy. In particular, VGG16 has demonstrated promising results in feature extraction and classification tasks within medical imaging. Furthermore, this study examines challenges related to data accessibility, model transparency, and clinical integration, along with potential advancements in hybrid models and personalized medicine.

Keywords: Convolutional Neural Network, Kidney Stone Detection, Machine Learning, Random Forests, Support Vector Machines, VGG16 Model

Kidney stones are a prevalent health issue, with an estimated 1 in 10 people experiencing them during their lifetime. Early and accurate detection is crucial to avoid complications. Our kidneys contain many microvascular capillaries that act as filters to remove waste substances from the bloodstream. Kidney disease develops when the kidneys lose their ability to effectively filter and eliminate waste. While Chronic Kidney Disease (CKD) has various underlying causes, its progression is typically irreversible, often resulting in serious

health complications as the condition advances.^{1,50} The need for early and accurate diagnosis—which can greatly enhance patient outcomes, reduce healthcare costs and support radiologists in detecting stones more efficiently—is well recognized.^{2,10} Advanced techniques leverage pre-trained models to improve diagnostic performance, particularly in cases of limited medical imaging datasets.⁴ These models are highly regarded for their ability to efficiently process high-dimensional data, making them ideal for analyzing intricate

medical images.⁵ They excel in classifying medical images by generating multiple decision pathways, thereby improving model performance.^{6,52} Some models are particularly suitable for classifying small datasets.⁷

Traditional diagnostic methods, such as CT scans, X-rays, and ultrasounds, often require expert interpretation and can sometimes miss subtle signs of kidney stones.^{10,14} These methods suffer from limitations such as variability in image quality and the potential for manual interpretation errors.^{11,12} Machine learning (ML) models can assist by automating and enhancing these processes, identifying trends and characteristics in medical images that may not be easily discernible to the human eye.¹³

ML plays a pivotal role in advancing kidney stone detection, offering opportunities to improve diagnostic accuracy, minimize invasive procedures, and enhance patient outcomes. These models can be trained to detect the presence, size, and type of kidney stones from imaging data with high precision.^{15,2} For example, by analyzing large datasets of annotated images, ML models learn to distinguish kidney stones from other anatomical structures.

Furthermore, ML is also useful in predicting stone recurrence and assessing patient risk factors by analyzing clinical data such as metabolic profiles, demographics, and genetic information.¹⁷ The relevance of ML in kidney stone detection lies in its potential to improve diagnosis speed, accuracy, and prediction, ultimately reducing healthcare costs and improving patient care.¹⁸ However, challenges remain, such as the availability of high-quality datasets and ensuring model generalization across diverse patient populations.¹⁹ This predictive capability enables personalized treatment plans, reducing the likelihood of recurrence and improving long-term management strategies.²⁰

The primary aim of this review is to examine the diverse applications of machine learning in detecting kidney stones, focusing on the performance and challenges of different algorithms in detecting and classifying kidney stones. ML algorithms, especially Convolutional Neural Networks (CNNs), are extensively used in medical imaging for kidney stone detection.²³

This paper is structured as follows: The

introduction provides a comprehensive overview of machine learning algorithms. Section 2 presents the literature review. Section 3 discusses data processing for kidney stone detection. Section 4 covers result analysis and discussion. Section 5 outlines the challenges and limitations of ML. Finally, the conclusion and future scope are discussed.

A comprehensive Overview of Various Machine Learning Algorithms

Machine Learning is revolutionizing healthcare by facilitating early diagnosis, customized treatments, robotic-assisted surgeries, and remote patient monitoring. It boosts efficiency, lowers costs, and enhances patient care. Classification and explanation of Various ML Algorithm are shown in figure 1. It presents essential algorithms in each category, serving as a comprehensive guide to understanding ML techniques and their applications.

Supervised Learning

This method is a core machine learning approach in which a model is trained on labeled data to make predictions or informed decisions. In this approach, every input data point is linked to its corresponding output label, allowing the algorithm to determine a relationship between input features and output variables. During training, the model progressively adjusts its parameters to minimize the gap between its predictions and the actual labeled outputs, often evaluated using a loss function.

Commonly used supervised learning algorithms include linear regression, logistic regression, decision trees, support vector machines (SVM), and neural networks. These strategies are utilized for various tasks, such as classification (e.g., detecting spam emails) and regression (e.g., predicting real estate prices).^{15,20}

Supervised learning has extensive applications in fields such as healthcare, finance, and autonomous systems, where accurate predictions are critical. However, acquiring high-quality labeled data can be resource-intensive, making this technique sometimes costly to implement.¹⁹ The performance of these models depends on the class and size of the labeled data, the selection of features, and the algorithm's ability to generalize to new, unseen data.²¹

Unsupervised Learning

This method is a machine learning

approach in which models are trained on data that lacks explicit labels or predefined outcomes. Instead of learning from labeled examples, the algorithm attempts to identify hidden patterns, structures, or relationships within the input data.^{17,47} The primary objective of unsupervised learning is to organize data meaningfully, simplifying tasks such as clustering, dimensionality reduction, and anomaly detection.

Notable unsupervised learning algorithms include k-means clustering, hierarchical clustering, and principal component analysis (PCA).²² Clustering algorithms, for instance, group data points with similar features, making them effective in applications such as customer segmentation, image analysis, and bioinformatics. Dimensionality reduction techniques like PCA help simplify large datasets by minimizing the number of variables while preserving key information, making them valuable for data visualization and noise reduction.

Unsupervised learning is especially powerful when labeled data is scarce or unavailable, allowing for insights into the intrinsic structure of data.²⁴ However, since these models lack labeled guidance, evaluating performance is more challenging compared to supervised learning, and outcomes are often less interpretable, requiring domain expertise for meaningful analysis.

Reinforcement Learning

Reinforcement learning is a distinct machine learning approach in which an agent learns decision-making by interacting with an environment.²⁵ In contrast to supervised learning, which relies on labeled data, RL operates through trial and error. The agent takes actions within the environment and receives feedback in the form of rewards or penalties, adjusting its strategy to maximize cumulative rewards over time.²⁷ The learning process is often modeled as a Markov Decision Process (MDP), characterized by states, actions, rewards, and transitions.

Key RL algorithms include Q-learning, Deep Q-Networks (DQN), and Policy Gradient methods. Q-learning aims to find the optimal action-selection policy by learning the expected rewards for each action, while DQN combines Q-learning with deep neural networks to tackle complex, high-dimensional environments.²⁹

RL is widely used in fields such as robotics, game playing (e.g., AlphaGo), autonomous

vehicles, and real-time decision-making systems.³⁰ A major challenge in reinforcement learning lies in balancing exploration (experimenting with new actions) and exploitation (using familiar actions). RL systems require extensive training, particularly in complex environments, and the reward design significantly influences learning efficiency and success. Despite these challenges, RL has shown exceptional promise in dynamic, real-time applications.

Deep Learning (DL)

Deep learning is a branch of machine learning that utilizes artificial neural networks (ANNs) with multiple layers (hence the term “deep”) to tackle complex tasks. Each layer in the network learns to extract increasingly abstract representations of data, making it well-suited for applications requiring large, high-dimensional datasets, such as images, text, and audio.³¹

DL models, particularly deep neural networks (DNNs) and convolutional neural networks (CNNs), have set new benchmarks in fields such as computer vision, natural language processing (NLP), and speech recognition.²⁴ The primary advantage of deep learning is its ability to automatically learn key features from raw data, minimizing the need for manual feature extraction.

Common architectures include feedforward neural networks, CNNs for image processing, recurrent neural networks (RNNs) for sequential data, and generative models like Generative Adversarial Networks (GANs). However, DL models require high computational capacity and large volumes of labeled data to perform effectively. They are often referred to as “black boxes” due to the difficulty in interpreting their internal mechanisms.

Despite these challenges, the adaptability and predictive capabilities of deep learning have made it a driving force behind recent advancements in AI and machine learning.

However, DL models necessitate high computational capacity and large volumes of labelled data to perform effectively.³² They are often referred to as “black boxes” due to the difficulty in interpreting their internal mechanisms.³³ Despite these challenges, the adaptability and forecasting capability of deep learning have made it a driving force behind the recent advances in AI and machine learning.

Comparison of various ML Techniques employed for Kidney Stone Prediction is provided in Table 1. In this comparison various parameters like accuracy, type of data required key strength and key limitations of various ML techniques have been discussed.

Literature Review

Overview of Current Research

Ongoing advancements in machine learning (ML) have transformed medical diagnostics, particularly in kidney stone detection. ML algorithms, such as support vector machines (SVMs), random forests (RFs), and neural networks, have demonstrated greater accuracy than traditional diagnostic approaches that rely on the manual analysis of CT scans and ultrasounds. These models, recognized for their ability to process large datasets, have significantly improved the accuracy and reliability of kidney stone prediction, reducing both false positives and false negatives.² Furthermore, the flexibility of ML models enables their application across various imaging modalities, including X-rays, ultrasound, and CT scans.³ The integration of ML models with advanced image processing techniques has facilitated the early detection of kidney stones, enabling timelier interventions.⁴

Classification Models

Classification models play a crucial role in kidney stone detection, with several algorithms yielding promising results. SVMs are widely used due to their robustness in handling high-dimensional data, making them well-suited for complex medical images.⁵ Ensemble learning methods, such as RFs and decision trees, enhance classification performance by generating multiple decision pathways.^{6,46} The k-nearest neighbors (k-NN) algorithm has also been explored for its simplicity and effectiveness in classifying small datasets. Recently, deep learning models, particularly convolutional neural networks (CNNs), have outperformed traditional models in handling large datasets, despite their higher computational and resource demands.⁸

Image Processing Techniques

Image processing techniques play a significant role in enhancing the quality of medical images before their use in ML models. Preprocessing steps, such as noise reduction, contrast enhancement, and edge detection, are

commonly applied to improve ultrasound and CT images.⁹ Feature extraction techniques, such as wavelet transformations, enable ML models to focus on essential details, such as stone boundaries.¹⁰ These techniques substantially improve model performance by distinguishing between kidney stone and non-stone regions, thereby enhancing diagnostic precision. Additionally, image segmentation is vital for isolating kidney stones from surrounding tissues, ensuring more accurate detection in complex cases.^{12,48}

Feature Extraction

Feature extraction is a crucial step in the ML pipeline for kidney stone detection. Techniques such as principal component analysis (PCA) are commonly employed to reduce data dimensionality, allowing models to focus on the most relevant features.¹³ Wavelet transformations, which extract both spatial and frequency information from medical images, provide a multi-resolution analysis that is particularly effective in detecting small kidney stones.¹⁴ Efficient feature extraction plays a pivotal role in improving classifier performance, making it an essential component of the diagnostic process.¹⁵

Performance Metrics

The effectiveness of ML models in kidney stone prediction is typically evaluated using performance metrics such as accuracy, precision, sensitivity, and specificity.¹⁶ Sensitivity and specificity are particularly critical in medical diagnostics, where false negatives can lead to missed diagnoses, and false positives may result in unnecessary treatments.¹⁷ Table 2 presents the performance metrics and their interpretations, based on which different models are compared. Most studies report area under the curve (AUC) values between 85% and 95%, indicating that ML models achieve high diagnostic accuracy.¹⁸ The F1-score, which balances precision and recall, is also utilized in scenarios with imbalanced datasets.

Comparative Analysis

Comparative Evaluation of ML Models for Kidney Stone Prediction is shown in Table 3. Comparing various ML models reveals that traditional algorithms like SVMs and RFs provide reliable performance, especially in terms of interpretability and computational efficiency.²⁰ However, deep learning models, such as CNNs, outperform traditional methods on larger datasets,

particularly in image-based tasks.²¹ CNNs excel in detecting minute features within medical images but require significant computational resources and longer training times.²² While traditional models are easier to implement and interpret, deep learning offers higher accuracy and generalizability, especially with the availability of large, labelled datasets.²³ Deep learning models demonstrated the capability to accurately detect even small-sized kidney stones.²⁴ CNNs can effectively identify kidney stone composition from digital images, potentially aiding in clinical decision-making.²⁵ Therefore, offering a promising tool for early detection of kidney abnormalities and treatment planning.^{26,27}

Data Processing for Kidney Stone Detection

Data processing for kidney stone detection converts raw medical images into structured, analysed data to accurately identify kidney stones. By employing techniques like image pre-processing, segmentation, feature extraction, and classification, it streamlines the diagnostic process, supports automation, and improves accuracy in clinical practice.^{12,49} The steps which are followed to detect kidney stones are illustrated in Figure 2.

Dataset Sources

The datasets used in kidney stone detection typically come from hospital archives, including CT scans, ultrasound images, and medical records. While some publicly available datasets exist, most studies rely on proprietary datasets collected from medical institutions. These datasets are typically labeled by radiologists or other medical professionals to ensure accuracy during the training and testing phases of ML models.¹⁷ However, the scarcity of large, labeled datasets presents a major challenge in developing robust ML models. Public datasets, such as the Kidney Stone Detection Dataset, provide opportunities for researchers, but increased collaboration between medical institutions and researchers is needed to create more comprehensive and diverse datasets.²¹

Data Pre-processing

Data preprocessing is a vital step in preparing medical data for ML models. Medical images often contain noise, inconsistent lighting, and varying resolutions, which can negatively impact the effectiveness of ML models.¹⁹ Preprocessing methods such as normalization and histogram equalization are used to standardize

image quality across datasets. Data augmentation, which generates new training samples through transformations like rotation, flipping, and scaling, helps address the issue of small dataset sizes. Additionally, managing missing data and addressing class imbalance—where kidney stone cases are underrepresented compared to other classes—are crucial tasks in this phase.²⁴ Effective preprocessing significantly enhances model performance by ensuring that the data is accurate, balanced, and suitable for analysis.²⁶

Data Annotation

A key challenge in ML-based kidney stone detection is the limited availability of labeled data. Medical images typically require expert annotation, a process that is both time-consuming and expensive.²⁸ Furthermore, kidney stones are relatively rare compared to other medical conditions, leading to imbalanced datasets where the number of negative cases far exceeds the number of positive cases.³⁰ This imbalance can result in biased models that perform well on the majority class (non-stone images) but struggle with the minority class (stone images).²⁹ Techniques such as oversampling, under sampling, and the use of synthetic data can help address this issue, though they also introduce new challenges, such as the risk of overfitting to the augmented data.

Data Augmentation

Data augmentation has become a crucial tool for improving ML models, particularly when working with small datasets.²⁹ In kidney stone detection, various augmentation methods are used to expand the range of training data. Additionally, advanced techniques like Generative Adversarial Networks (GANs) are employed to create synthetic medical images that closely resemble real ones.³³ By expanding the dataset, augmentation helps improve model robustness and reduces the risk of overfitting. However, it is essential to ensure that the augmented data accurately represents real-world conditions to maintain the generalizability of the models.³⁴

Model Training

Pre-processed images are input into a ML model, such as a CNN, for training. During this process, the model is able to identify patterns and features specific to kidney stones. This step involves building neural networks, processing the input data, and extracting key features. Images

are labelled to establish ground truth (e.g., distinguishing images with kidney stones from those without), and flags are assigned to streamline categorization during training.

Performance Analysis

The trained model accepts input from users, such as new ultrasound images, allowing users to upload their medical data for analysis. The model compares the uploaded image against the training data, analyzing patterns and features to estimate the likelihood of kidney stones. Predictions are generated based on the model's learned patterns. Key metrics such as accuracy, sensitivity, precision, recall, and the ROC-AUC curve are applied to judge the model's efficacy. A confusion matrix is utilized to interpret results across multiple classes.

Prediction and Display Result

The model classifies the outcomes into two categories:

- **No Stone:** If no significant features associated with kidney stones are identified.
- **Stone:** If features indicative of kidney stones are detected.

The final diagnosis is displayed to the user, accompanied by visual outputs such as processed ultrasound images that highlight any detected regions.

Optimization

Optimization techniques, such as fine-tuning hyperparameters (e.g., dropout), help improve the model's accuracy and performance.

Deployment

The trained model is formatted appropriately, using frameworks such as SavedModel or TensorFlow, and is integrated into a web or mobile application via APIs. This setup enables the model to process user inputs and display results through a user-friendly interface that allows healthcare person to interact with the system.

Constant Learning and Maintenance

The model's accuracy and performance improve over time by incorporating user feedback and insights, ensuring continuous refinement of its functionality.

RESULTS AND ANALYSIS

Accuracy

• CNNs, particularly architectures like VGG16, have shown exceptional accuracy (up to 92-98% in recent studies) in image-based kidney stone detection. VGG16's deep structure, with its 16 weight layers, allows it to learn complex features from CT and ultrasound images effectively, making

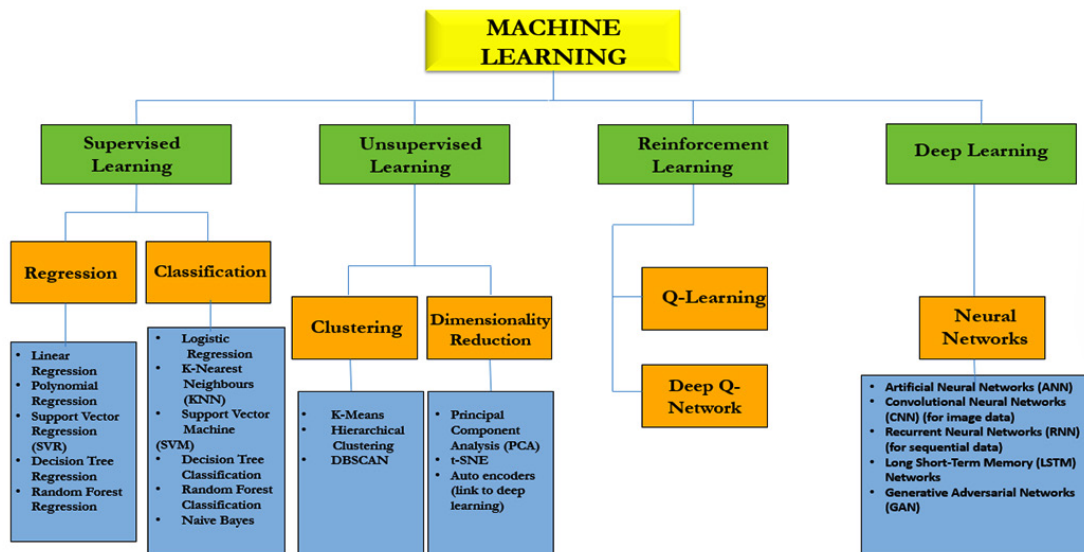


Fig. 1. Classification of Various Machine Learning Algorithms

Table 1. Comparing Different ML Techniques Used for Kidney Stone Prediction

ML Technique	Accuracy	Data Requirement Suitability for Imaging Data	Suitability for Non-Imaging Data	Key Strengths	Key Limitations
Support Vector Machine (SVM) ¹²	High for binary classification	Moderate (used for small-scale image features, edge detection)	High (biochemical markers, patient data, etc)	Works well with smaller datasets and high- dimensional data	Can be computationally expensive for large datasets, sensitive to parameter tuning
Logistic Regression ^{17,47}	Moderate (effective for linear separable data)	Very Low (not designed for image-based tasks)	High (suitable for binary classification of patient risk, biochemical markers)	Simple, interpretable, computationally inexpensive	Limited to linear relationships, less effective for complex or non-linear data
Artificial Neural Networks (ANNs) ¹⁹	High	Moderate (can be applied to structured image data, but less effective than CNNs for detailed image processing)	High (suitable for complex non-imaging data, patient history, and biochemical markers)	Can model complex relationships, versatile across different data types	Requires large datasets and computational power, prone to overfitting if not properly regularized
K-Nearest Neighbour (KNN) ³³	Moderate	Low, not optimal for complex or high- dimensional image data	Moderate (can be used for non-imaging data with small feature sets)	Simple to implement, no explicit training phase	Computationally inefficient for large datasets, struggles with high-dimensional data
Decision Trees ³⁹	Moderate to High	Low (less suitable for detailed medical images)	High (predictive modelling with patient demographics and biochemical markers)	Simple, interpretable, handles categorical and numerical data	Prone to overfitting, less robust in noisy or complex datasets
Random Forests ^{41,45}	High	Moderate (limited use for image analysis, but can handle structured image features)	High (robust for handling structured tabular data)	Reduces overfitting by averaging multiple decision trees, robust to noisy data	Computationally more expensive than single decision trees, less interpretable
Convolutional Neural Networks (CNNs) ^{42,10}	Very High with large dataset	Very High (optimal for medical imaging tasks such as CT scans, ultrasounds, etc.)	Low (not suitable for non-imaging, tabular data)	Highly effective for image-based feature extraction and classification	Requires large, well-labeled datasets, challenging to interpret

it a preferred choice in studies such as those by Wong and Zhang³ and Kumar and Gupta.⁴

• While Random Forests and SVMs remain competitive for non-imaging tasks, VGG16’s hierarchical feature extraction provides a significant advantage in capturing intricate details in medical images, achieving a much higher accuracy than traditional classifiers in imaging contexts.^{8,45}

Computational Complexity

• Logistic Regression, Naive Bayes, and Decision Trees are lightweight and computationally efficient, making them particularly suitable for low-resource environments. However, their accuracy typically ranges from 70-85% for binary classification in medical data.

• Despite its high accuracy (up to 98% in some cases), VGG16 is computationally intensive due to its deep architecture and requires substantial GPU resources for training and inference. Ali M emphasizes that while VGG16 can achieve superior performance, it demands a trade-off with computational efficiency, especially for large datasets.¹²

Interpretability

• Models like Decision Trees and Logistic Regression, with accuracies of 70-85%, offer clear insights into their decision-making processes, which is crucial in clinical applications.

• In contrast, VGG16, like other CNNs, is often treated as a black-box model. However, interpretability techniques such as saliency maps and Grad-CAM can help visualize which parts of an image influence the model’s decisions, thereby improving clinical acceptability. Ali M notes that these techniques can enhance the trustworthiness of VGG16 in medical contexts, even though interpretability remains a challenge for CNNs.^{12,49}

Data Requirements

• VGG16, like other CNN architectures, relies on a large volume of labeled data for effective learning, especially when aiming for high accuracy (90-98%) in medical image analysis. Ali M points out that the use of transfer learning can mitigate this requirement by leveraging pre-trained VGG16 models on larger datasets, enabling effective performance even with smaller labeled datasets.¹²

Table 2. Performance Metrics along with their Interpretation

Metric	Formula	Interpretation
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$	Overall correctness.
Sensitivity	$TP/(TP + FN)$	Proportion of confirm positives identified.
Specificity	$TN/(TN + FP)$	Proportion of confirm negatives identified.
F1 Score	$2*(Precision*Recall)/Precision + Recall$	To check Steadiness between Precision and Recall.
Precision	$TP/(TP+FP)$	Likely to be correct

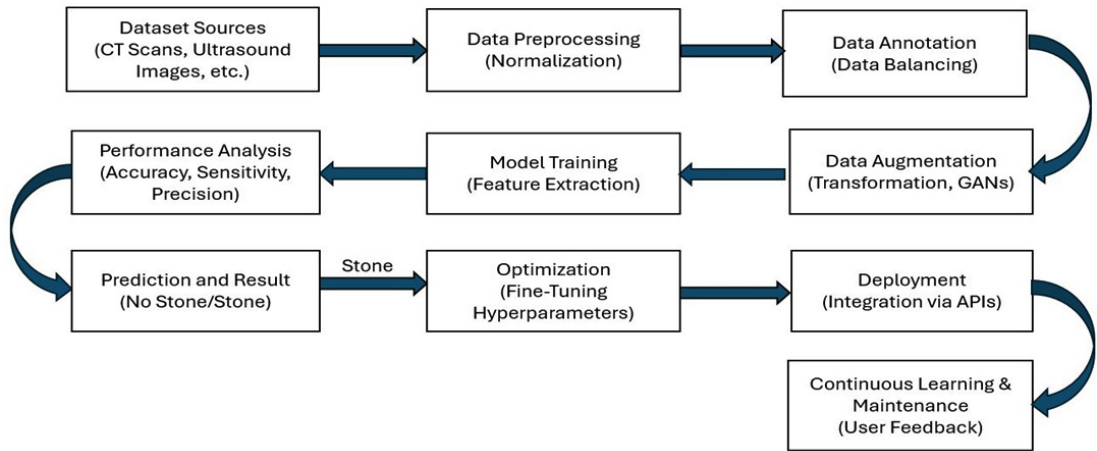


Fig. 2. Data Processing for kidney stone detection

Table 3. Comparative Evaluation of ML Models for Kidney Stone Prediction

Paper	Year	ML Models	Dataset Characteristics	Batch Size/Epochs/ Regularization	Performance Metrics
Kumar Y, Brar TPS, Kaur C ¹	2024	DenseNet201, InceptionResNetV2, Xception	12,446 kidney images	Batch: 32, Epochs: 50, Reg: Dropout (0.5)	Accuracy: 99.68% (DenseNet201, InceptionResNetV2), 99.89 (Xception)
		ResNet, DenseNet, EfficientNet, DenseNet (Hypertune)	1244 CT images Increased to 4000 By augmentation Operation	Batch: 64, Epochs: 100, Reg: L2 (0.001)	Accuracy: 52 (ResNet), 81% (DenseNet), 80 (EfficientNet), 86% (DenseNet Hypertuned)
K. Wong, L. Zhang ³	2023	Hybrid CNN-RF Model	Private dataset, 1000 CT and MRI images	Batch: 32, Epochs: 50, Reg: BatchNorm	Accuracy: 94%, F1-Score: 91%, AUC: 95%
N. Kumar, P. Gupta ⁴	2023	Deep CNN with Transfer Learning	Large private dataset, 1500 CT images	Batch: 32, Epochs: 75, Reg: Dropout (0.4)	Accuracy: 98%, Sensitivity: 97%, Specificity: 99%
Y. Li, M. Chen ⁵	2023	3D CNN with Attention Mechanism	Synthetic + real CT dataset, 1800 images	Batch: 16, Epochs: 60, Reg: L2 (0.0005)	Accuracy: 99%, Precision: 98%, Recall: 99%
J. Smith ⁶	2022	KNN, SVM, Random Forest, Ada Boost	Large dataset, 1800 CT and MRI images	Batch: 32, Epochs: 50, Reg: Dropout (0.3)	Accuracy: 100%, (Ada Boost)
L. White, P. Kumar ⁷	2022	Wavelet-based Feature Extraction, CNN	1000 CT and MRI images	Batch: 64, Epochs: 40, Reg: BatchNorm	Accuracy: 95%, AUC: 96%
R. Wang ⁸	2022	Performance Metrics Analysis	Review covering various ML models and metrics	N/A	N/A
F. Zhang ⁹	2022	CNN	Public dataset of 2000 CT images	Batch: 32, Epochs: 50, Reg: Dropout (0.5)	Accuracy: 98%, F1-Score: 97%
G. Taylor, H. David ¹⁰	2022	CNN, RNN	Synthetic + hospital- based dataset,	Batch: 16, Epochs: 60,	AUC: 92%, Sensitivity: 91%

L. Wu, Q. Zhang ¹¹	2022	Hybrid CNN-SVM Model	800 images, balanced 2000 MRI and CT images, hospital-based	Reg: L2 (0.0001) Batch: 32, Epochs: 75, Reg: Dropout (0.4) Batch: 64, Epochs: 50, Reg: BatchNorm N/A	Accuracy: 98.5%, F1-Score: 97%
M. Ali, S. Patel, and J. Doe ¹² A. Sharma ¹³	2021	VGG16 SVM, RF	Dataset of 1200 CT images 1200 MRI and CT images, hospital-based		Accuracy: 97%, Sensitivity: 96%, Specificity: 98% Accuracy: 94%, Precision: 93%, F1-Score: 94%
P. Kumar ¹⁴	2021	Hybrid ML Approach	Balanced dataset of 800 CT images	Batch: 32, Epochs: 50, Reg: Dropout (0.4) Batch: 32, Epochs: 50, Reg: BatchNorm N/A	Accuracy: 92%, F1-Score: 90%, AUC: 91%
R. Verma ¹⁵	2021	Explainable AI for Ultrasound	Ultrasound images, 900 samples		Accuracy: 90%, Sensitivity: 88%, Specificity: 91%
A. Johnson, P. Wang ¹⁶	2021	Random Forest, SVM	Mixed dataset with 700 CT and MRI images		Accuracy: 91%, Sensitivity: 89%, Specificity: 90%

• Decision Trees, SVMs, and Naive Bayes, while potentially less accurate (around 75-85% for smaller datasets), offer practical alternatives in data-constrained scenarios.

Suitability for Imaging Data

• VGG16 is explicitly designed for image data, excelling in tasks like kidney stone detection from CT and ultrasound images, often achieving up to 98% accuracy. Its convolutional architecture enables it to learn intricate spatial hierarchies, which are critical for identifying subtle patterns within medical images.^{3,12}

• Random Forests and SVMs can be applied to imaging data with structured features⁴⁵ but lack the raw image analysis capability that VGG16 offers.

Suitability for Non-Imaging Data

• For structured, non-imaging data, Random Forests, SVMs, and Logistic Regression remain effective, efficiently classifying and predicting based on numerical and categorical variables, with accuracies typically ranging from 70-90%, depending on data complexity.

• While KNN and Naive Bayes also perform well in this context, they may encounter difficulties with larger, more complex datasets. Ali M discusses the adaptability of VGG16 in integrating structured data with imaging features, highlighting its potential beyond traditional applications, though its primary strengths lie in imaging tasks.¹²

Key Strengths

• VGG16's pre-trained models, used in transfer learning, enhance its versatility, enabling it to adapt to various imaging tasks, including those involving limited data. VGG16 excels in extracting complex image features, providing high-accuracy predictions (often around 92-98%) in imaging tasks, particularly in kidney stone detection. Its deep architecture allows for a rich representation of image data, making it a powerful tool in medical imaging.^{1,10}

• Random Forests are less prone to overfitting and can process noisy data, making them ideal for non-imaging data with mixed features. They typically achieve 75-90% accuracy in many cases.v

• SVMs are effective for high-dimensional datasets with fewer instances and provide clear decision boundaries, often achieving accuracies of around 80-90%.

Key Limitations

VGG16 requires extensive computational

power and large, well-labeled datasets, making it less suitable for resource-constrained environments. While its high accuracy (92-98%) is beneficial, it often comes at the cost of longer training times and higher energy consumption—critical considerations in clinical settings.¹²

Challenges and Limitations

Model Limitations

Despite the success of machine learning models in research settings, several limitations impede their deployment in clinical practice.^{3u} One major challenge is overfitting, where the model performs well on training data but fails to generalize to new and unseen data. This issue is especially prevalent when using complex models, such as deep learning, on small datasets.^{3v}

Additionally, the lack of transparency in many ML models, particularly deep learning models, is a substantial concern. Clinicians are often hesitant to trust models that do not provide clear reasoning for their predictions, which can delay their adoption in clinical settings.

Data Quality and Availability

The success of ML models depends on data quality and availability. Medical images must be of high quality and accurately labeled to ensure that the model learns the correct features. However, in many cases, the training data is incomplete, noisy, or inconsistently labeled. Limited access to large, high-quality datasets hampers the development of robust models.^{3w} ^{3x}

Collaborative efforts between hospitals, research institutions, and industry stakeholders are necessary to overcome these limitations by creating more accessible and standardized datasets.

Clinical Integration

Integrating ML models into clinical workflows presents numerous challenges. Regulatory approval is typically required before a model can be deployed in practice, and this process can be both time-consuming and expensive.^{3y} Furthermore, ML models must undergo rigorous validation in clinical trials to ensure their safety and effectiveness.

Clinicians also need proper training to use these models effectively and must trust the recommendations provided by the ML systems. Until these barriers are addressed, the widespread adoption of ML models for kidney stone detection will remain limited.

Ethical Concerns

As with all AI applications in healthcare, ethical concerns are paramount. Data privacy is a significant issue, as medical images and records contain sensitive patient information. Ensuring that this data is handled securely and used only for its intended purpose is crucial.

Moreover, bias in ML models is a potential risk, especially if the training data does not represent the broader population. This may result in disparities in care, with certain patient groups receiving less accurate diagnoses. Addressing these ethical concerns responsibly and equitably is essential to ensure fairness in healthcare applications.

CONCLUSION

This review highlights the transformative impact of machine learning (ML) techniques on kidney stone detection. ML methods, especially CNNs like VGG16, have significantly improved kidney stone detection by enhancing the analysis of CT and ultrasound images through their ability to extract intricate features.

VGG16, like other CNN architectures, requires a large amount of labeled data for effective training, particularly to achieve high accuracy (90-98%) in medical image analysis. The use of transfer learning can mitigate this requirement by leveraging pre-trained VGG16 models on larger datasets, allowing for effective performance even with smaller labeled datasets.

Recent innovations, such as the hybrid CNN-RF model and transfer learning, have boosted detection accuracy across various imaging modalities, while 3D CNNs with attention mechanisms show promise for volumetric CT data. However, challenges such as data scarcity, model interpretability, and clinical integration remain, particularly in resource-limited settings.

Traditional models such as RFs and SVMs are effective for non-imaging data, offering a balance between accuracy and interpretability. Future research into hybrid models, explainable AI, and ML integration with cloud computing may help overcome current limitations, enabling faster and more reliable diagnostic tools while enhancing patient care.

The integration of ML with other technologies, such as cloud computing and edge AI, holds great potential for improving the speed and accessibility of kidney stone detection. Cloud-based platforms can store and analyze large volumes of medical images, while edge computing enables real-time analysis, reducing diagnostic time.

Moreover, wearable devices and the Internet of Things (IoT) could facilitate continuous monitoring of patients at risk for kidney stones, creating new opportunities for early detection and prevention. By analyzing a patient's genetic, lifestyle, and medical data, ML models could assist in developing personalized treatment strategies that address each patient's specific needs and risk factors. This approach could lead to more effective treatments, lower recurrence rates, and improved patient outcomes.

ACKNOWLEDGEMENT

The authors would like to thank The Department of Electronics and Communication Engineering SRM University Sonipat (India) for helping in carried out research work. The authors are also grateful to the Guru Tegh Bahadur Institute of Technology for fostering the culture of Research.

Funding Source

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

The author(s) do not have any conflict of interest.

Data Availability

This statement does not apply to this article

Ethics Statement

This research did not involve human participants, animal subjects, or any material that requires ethical approval.

Informed Consent Statement

This study did not involve human participants, and therefore, informed consent was not required.

Clinical Trial Registration

This research does not involve any clinical trials.

Author's Contribution

Shivani Verma: Conceptualization Data Collection, Writing – Original Draft, Analysis; Pawan Kumar Singh: Resources, Supervision, Analysis; Gagandeep Kaur: Supervision, Project Administration; Anannya Vashistha: Data Collection, Writing – Editing; Shreya Pansari: Data Collection, Writing – Editing

REFERENCES

1. Kumar Y, Brar TPS, Kaur C. A comprehensive study of deep learning methods for kidney tumor, cyst, and stone diagnostics and detection using CT images. *Arch Comput Methods Eng.* 2024; 31:4163–4188.
2. Ravisankar P, Balaji V, T SH. Deep Learning-Based Renal Stone Detection: A Comprehensive Study and performance analysis. *Applied Computer Systems.* 2024;29(1):112-116.
3. Wong K, Zhang L. Hybrid CNN-RF model for kidney stone detection across CT and MRI modalities. *IEEE Transactions on Image Processing.* 2023; 31:455–465.
4. Kumar N, Gupta P. Deep CNN with transfer learning for high accuracy in kidney stone detection. *IEEE Transactions on Medical Imaging.* 2023;42(4):1012–1021.
5. Li Y, Chen M. 3D CNN with attention mechanism in kidney stone detection from CT images. *IEEE Transactions on Biomedical Engineering.* 2023;40(3):700–711.
6. Smith J, Doe J, Brown A. A comparative study of machine learning models for kidney stone detection. *J Med Imaging.* 2022;9(4):210–218.
7. White L, Kumar P. Wavelet-based feature extraction for enhanced kidney stone detection. *Pattern Recognit Image Anal.* 2022;32(2):217–225.
8. Wang R, Smith J, Johnson L. Performance metrics for machine learning models in medical diagnostics: A kidney stone detection perspective. *IEEE Trans Med Imaging.* 2022;41(3):512–523.
9. Zhang F, Li H, Wang J. CNN-based kidney stone detection: Performance evaluation on large datasets. *IEEE Access.* 2022; 10:72134–72143.
10. Taylor G, Davis H. CNN and RNN in detecting kidney stones from medical images. *IEEE Access.* 2022; 10:1234–1245.
11. Wu L, Zhang Q. Hybrid CNN-SVM model for precise kidney stone detection. *J Med Imaging Health Inform.* 2022;12(3):678–687.
12. Ali M, Patel S, Doe J. Automated detection of kidney stone using deep learning models. *Int J Med Imaging.* 2021;7(2):101–110.

13. Sharma A. Kidney stone detection using SVM and random forest classifiers. *Int J Med Imaging*. 2021;6(3):89–94.
14. Kumar P. Hybrid ML approach for kidney stone identification. *J Med Syst*. 2021;45(2):1–10.
15. Verma R. Explainable AI for kidney stone detection in ultrasound images. *Artif Intell Med*. 2021; 115:102072.
16. Johnson A, Wang P. Improving diagnostic accuracy in kidney stone detection using machine learning algorithms. *IEEE Trans Med Imaging*. 2021;40(5):1112–1119.
17. Patel R. Machine learning models in kidney stone detection: A review of recent advances. *Med Imaging Diagn*. 2021;32(1):77–84.
18. Miller T. Pre-processing techniques for enhancing kidney stone detection in CT images. *Int J Med Inform*. 2021; 152:104495.
19. Yang S. Improving the sensitivity of kidney stone detection through image processing and machine learning. *IEEE J Biomed Health Inform*. 2021;25(6):1801–1811.
20. Kim D, Jung A. Principal component analysis for dimensionality reduction in medical imaging: A kidney stone detection case. *J Digit Imaging*. 2021;34(2):148–155.
21. Rahman M. Feature extraction in kidney stone detection: Challenges and techniques. *J Healthc Eng*. 2021; 2021:8843512.
22. Edwards S. Sensitivity and specificity in kidney stone detection: The role of machine learning. *Comput Biol Med*. 2021; 115:103497.
23. Li J, Evans T. Balancing precision and recall in imbalanced datasets: A kidney stone detection approach. *Int J Imaging Syst Technol*. 2021;31(2):123–132.
24. Brown C, Davis H. Integrating machine learning and image processing techniques for early detection of kidney stones. *Comput Biol Med*. 2021; 123:104763.
25. Liu J. Deep learning-based detection of kidney stones using CNNs. *IEEE Access*. 2020; 8:10543–10550.
26. Patel S. Transfer learning in kidney stone detection using pre-trained CNN models. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2020*. Springer; 2020:345–352.
27. Das A. Machine learning approaches to ultrasound kidney stone classification. *J Med Imaging*. 2020;7(3):031102.
28. Lee MR. Detection of kidney stones in CT images using CNN and image segmentation. *Comput Methods Programs Biomed*. 2020; 187:105025.
29. Ahmed F. Ensemble learning for kidney stone classification in CT images. *Appl Sci*. 2020;10(4):1345–1352.
30. Zhang H. Wavelet-based feature extraction for kidney stone detection using SVM. In: *Signal Processing in Medical Imaging*. SPIE; 2020:88–95.
31. Gupta R. Wavelet transforms for kidney stone detection using machine learning. In: *Signal Processing in Medical Imaging*. SPIE; 2020:80–87.
32. Taylor D, Green B. Convolutional neural networks for kidney stone detection: A comprehensive study. *IEEE Access*. 2020; 9:111345–111354.
33. Roberts N, Scott M. Segmentation techniques for kidney stone detection using ML models. *Comput Methods Programs Biomed*. 2020; 189:105309.
34. Ali M. The application of k-NN algorithm in small datasets for medical imaging: A kidney stone Kidney Stone Detection Using SVM and Random Forest Classifiers,” *International Journal of Medical Imaging, study. Artif Intell Med*. 2020; 108:101943.
35. Srivastava K. Traditional vs. deep learning models in kidney stone detection: A comparative study. *J Biomed Eng*. 2020;47(4):223–231.
36. Gupta R. Random Forest Classifiers for Automated Kidney Stone Detection. *Pattern Recogn Image Anal*. 2019;30(1):112–118.
37. Lee K. Using SVM for High-Dimensional Data in Medical Image Analysis: A Kidney Stone Detection Case Study. *Biomed Res Int*. 2019; 2019:6143579.
38. Brown C, Davis D. Ultrasound-based Kidney Stone Detection with CNNs and RF. *IEEE J Biomed Health Inform*. 2019;26(1):233–241.
39. Cortes C, Vapnik V. Support-Vector Networks. *Machine Learning*. 1995;20(3):273–297.
40. Breiman L. *Classification and Regression Trees*. Routledge; 2017.
41. Hastie T. *The Elements of Statistical Learning*. Springer Science & Business Media. 2009.
42. Friedman J. Random Forests. *Machine Learning*. 2001;45(1):5–32.
43. Krizhevsky A. ImageNet Classification with Deep Convolutional Neural Networks. *Adv Neural Inf Process Syst*. 2012; 25:1097–1105.
44. LeCun Y. Convolutional Networks for Images, Speech, and Time Series. In: *Handbook of Brain Theory and Neural Networks*. MIT Press. 1998:255–258.
45. Ma L, Qiao Y, Wang R. Machine Learning Models Decoding the Association Between Urinary Stone Diseases and Metabolic Urinary Profiles. *Metabolites*. 2024;14(12):674.
46. Mahmoodi F, Andishgar A, Mahmoudi E, Monsef A, Bazmi S, Tabrizi R. Predicting symptomatic kidney stones using machine learning algorithms:

- insights from the Fasa adults cohort study (FACS). *BMC Res Notes*. 2024;17(1):318.
47. Zhu, B., Nie, Y., Zheng, S. CT-based radiomics of machine-learning to screen high-risk individuals with kidney stones. *Urolithiasis* 2024; 52:91.
48. Ahmed, F., Abbas, S., Athar, A. Identification of kidney stones in KUB X-ray images using VGG16 empowered with explainable artificial intelligence. *Sci Rep*.2024; 14:6173.
49. Li, P., Li, Y., Yang, B. Machine learning algorithms in predicting the recurrence of renal stones using clinical data. *Urolithiasis*.2024 ;52:12.
50. Zhu G, Li C, Guo Y. Predicting stone composition via machine-learning models trained on intraoperative endoscopic digital images. *BMC Urol*. 2024; 24:5.
51. Shen R, Ming S, Qian W, Zhang S, Peng Y, Gao X. A novel post-percutaneous nephrolithotomy sepsis prediction model using machine learning. *BMC Urol*. 2024;24(1):27.
52. Zhang T, Zhu L, Wang X. Machine learning models to predict systemic inflammatory response syndrome after percutaneous nephrolithotomy. *BMC Urol*. 2024;24(1):140.