A Comprehensive Survey on Machine Learning and Deep Learning Techniques for Kidney Stone Detection and Classification in Medical Imaging

1. Ajaykumar 2. Dr Mahesha D M 3. Dr. Bhavya D N

Department of Studies and Research in Computer Science, Karnataka State Open University, Mysore-570006, India.

A B S T R A C T

This survey takes a close look at how machine learning techniques and deep learning techniques are being used to detect and classify kidney stones through medical imaging. It covers key imaging techniques like ultrasound, CT scans, and X-rays explaining how each contributes to diagnosis and where they fall short. The paper explores a variety of ML algorithms such as decision trees, SVMs and ensemble methods, along with Deep Learning models like CNN. Then we use these ways to evaluate them are performance indicators like accuracy, sensitivity, and specificity, with a focus on how they can improve diagnostic precision and minimize human error. It also highlights challenges in the field, including limited data availability, differences in how kidney stones appear, and the system needed easy and fast interpretation. Looking ahead, the survey points to promising directions like transfer learning, uniting data from multiple imaging sources, and building reliable decision-support tools to enhance kidney stone diagnosis.

Keywords: DL: Deep Learning ML: Machine Learning

I. INTRODUCTION

Kidney stone disease, or nephrolithiasis, is a painful illness that affects millions of people around the world. Detecting it early and accurately is essential to prevent serious issues like blocked urinary flow, infections, and long-term damage to the kidneys.

Traditionally, diagnosis relies on radiologists interpreting medical images from tools like ultrasound, X-rays, and CT scans. Even when these methods work, they can be slow, subjective, and vary in accuracy depending on the practitioner and imaging quality.

Recently, machine learning and deep learning have transformed how medical images are analyzed. These technologies offer automated, scalable, and highly accurate tools for diagnosing kidney stones. ML and DL algorithms are now capable of identifying, segmenting, and classifying stones based on their size, shape, and composition.

This survey reviews the latest techniques in the field from classic ML models like SVMs and random forests to cutting-edge DL architectures such as CNNs and transformers. It also speaks about key challenges, including limited data availability, differences in imaging formats, and the models needed for generalizing well among diverse cases. Finally, it highlights future research directions aimed at improving clinical adoption and making kidney stone diagnosis more efficient and reliable.

II. LITERATURE STUDY

1. **Deep Learning Model-Assisted Detection on CT in Multiple Planes**
   Caglayan et al. (2022) carried out a study to examine how well deep learning performs in detecting kidney stones in unenhanced CT images across different anatomical planes (axial, coronal, sagittal). They used a dataset comprising 455 patients (405 with stones, 50 without), and 2,959 CT image slices reviewed by radiologists. The images were stratified by stone size (0-1 cm, 1-2 cm, >2 cm). The model's training accuracy was very high (≈98-99%) in different planes and size categories. However, test accuracy dropped notably (e.g. ~78% in the axial plane for smallest stones, higher for larger ones). The study demonstrates the promise of DL for stone detection but also underscores challenges in generalization across planes and sizes.

2. **Coronal CT Image Deep Learning Detection**
   A study by researchers in *Computers in Biology and Medicine* proposed a DL model that detects the presence or absence of kidney stones from coronal non-contrast CT (NCCT) images. Utilizing 1,799 images from 433 subjects, the model achieved **96.82% accuracy** and **95.76% sensitivity** on a test set of 146 cases. Critically, this work emphasized capability to detect small stones. It also showed that using coronal views may help in accurate decision-making and region-of-interest localization. This supports the idea that cross-sectional imaging in different orientations may be important.

3. **Classification via Axial CT with Pre-trained CNNs**
   In "Effective deep learning classification for kidney stone using axial CT images," the Inception-V3 architecture (pre-trained) was used to classify CT scans as stone vs non-stone from axial slices. From a dataset of 8,209 CT images, the method yielded a high **test accuracy of about 98.52%**. The experiment also showed that the model could detect small stones, suggesting robustness. This work highlights how transfer learning with large CNN backbones can perform very well in binary detection tasks, especially given enough data.

4. Automatic Detection & Scoring via S.T.O.N.E. Nephrolithometry
Another line of work combines segmentation, deep learning, and thresholding to both detect stones and score them based on features relevant to clinical decisions. In the "Automatic Detection and Scoring of Kidney Stones on Noncontrast CT Images Using S.T.O.N.E. Nephrolithometry," the pipeline involved kidney and renal sinus segmentation (via 3D U-Nets), hydronephrosis grading, thresholding to identify stone regions, and deriving metrics such as stone size, attenuation, tract length. The stone detection achieved ~95.9% sensitivity and ~98.7% positive predictive value (PPV). The scoring (stone size, number of involved calyces, essence etc.) agreed very well with radiologists (high κ values). This illustrates how ML/DL can go beyond detection to clinically meaningful quantification.

5. Segmentation Networks for Kidney & Stone in Unenhanced Abdominal CT
Li et al. (2022) developed deep segmentation networks to perform two tasks: segment kidneys, and detect kidney stones in unenhanced abdominal CT images. A recent study published in MDPI found that isolating kidney structures before trying to detect stones can significantly reduce false positives and improve the accuracy of locating the stones. This step—known as

segmentation—is crucial because it narrows down the area the system needs to analyze, making it easier for detection and classification algorithms to work effectively. Often, segmentation is done using advanced models like U-Net variants. The research showed that using segmentation first, followed by detection, leads to better results than relying on detection alone.

6. **Ensemble Techniques & Model Fusion**
   A recent 2024 study titled "An Optimized Fusion of Deep Learning Models for Kidney Stone Detection from CT Images" explored how combining multiple deep learning models—known as ensemble techniques—can improve diagnostic accuracy. The researchers introduced two ensemble approaches:

   - **StackedEnsembleNet**, which layers predictions from several well-known models including InceptionV3, InceptionResNetV2, MobileNet, and Xception.
   - **PSOWeightedAvgNet**, which uses Particle Swarm Optimization (PSO) to fine-tune the weighting of predictions from these models.

     Using a dataset of 1,799 CT images, both ensemble models performed better than any of the individual models alone. The study also incorporated Grad-CAM to provide visual explanations of the model's decisions, and showed strong generalization when tested on new, unseen data.

7. **In Vivo Stone Type Recognition during Ureteroscopy**
   In "On the in vivo recognition of kidney stones using machine learning," the authors focus on classifying kidney stone *types* (chemical composition) during ureteroscopy (endoscopic imaging), rather than just presence or size. They compared shallow machine learning (texture/color features + classifiers like XGBoost) and deep learning (CNNs like Inception v3). The deep model achieved very high performance: weighted precision ≈ 0.97, recall ≈ 0.98, F1 ≈ 0.97; shallow methods (XGBoost) approached similar levels (~0.96) when using carefully extracted features. This study is important because stone composition determines management (diet, prevention, dissolution etc.), but real-time composition classification is under-explored.

8.**Attention & Multi-View Feature Fusion for Stone Fragments**
The study "Improved Kidney Stone Recognition Through Attention and Multi-View Feature Fusion Strategies" addresses the problem of fragment images (from surface view and sectional view) of stone samples, as often seen in ex vivo settings, with the aim of simulating how biologists perform morpho-constitutional analysis visually. A recent study published on arXiv introduced a method that combines features from multiple imaging perspectives and uses attention layers—added after convolutional blocks—to highlight the most relevant regions before making a classification. This approach led to a roughly 4% improvement in accuracy compared to models using a single view, and up to 11% better performance than previous top-performing methods.

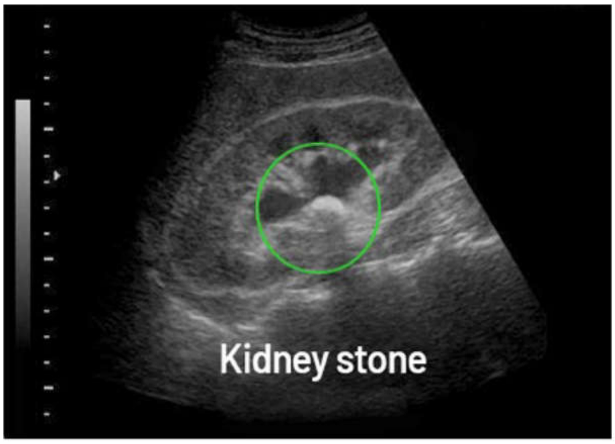
These results suggested that using multi-view data along with attention mechanisms helps the model focus on more distinctive features, making it more effective at identifying different types or fragments of kidney stones.

9. **Dataset Creation & Augmentation for Axial CT Images**
   A more infrastructure-focused initiative titled "Kidney Stone Detection via Axial CT Imaging: A Dataset for AI and Deep Learning Applications" offers a valuable resource for researchers. It includes 3,364 original axial CT images, along with 35,457 augmented images that were generated to enhance model training. Each image is labeled to indicate whether it contains a kidney stone or not.

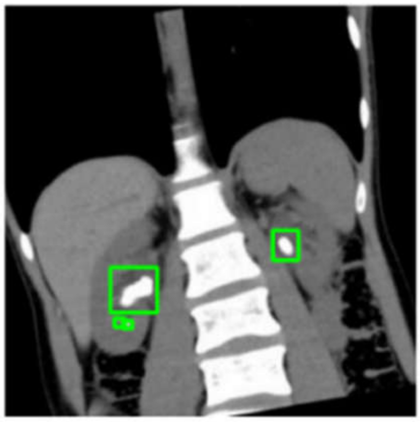
   The dataset was compiled from multiple medical centers across Iraq, adding diversity and real-world variability to the collection. Resources like this are essential because the success of machine learning and deep learning models depends heavily on the quality, variety, and volume of available data. Image augmentation also plays a key role in reducing overfitting and improving the robustness of these models.

10. **Hybrid Studies Combining Image Processing + ML/DL Methods**
    Some studies have explored combining traditional image processing techniques—like filtering, histogram equalization, segmentation, clustering, and feature extraction—with deep learning to improve kidney stone detection. For example, Pavithra et al. developed a hybrid approach that uses image preprocessing, fuzzy c-means clustering, and a convolutional neural network (CNN).

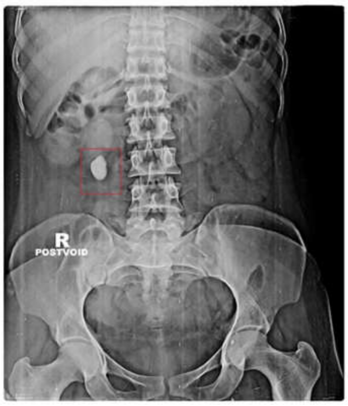
    The clustering step helps highlight key regions of interest and enhances image contrast and segmentation before the CNN takes over, allowing it to focus on the most relevant areas. While these hybrid methods may not always match the performance of pure deep learning models, they often require less labeled data and can be easier to interpret. They're especially useful in real-world scenarios where computational resources or annotated datasets are limited.



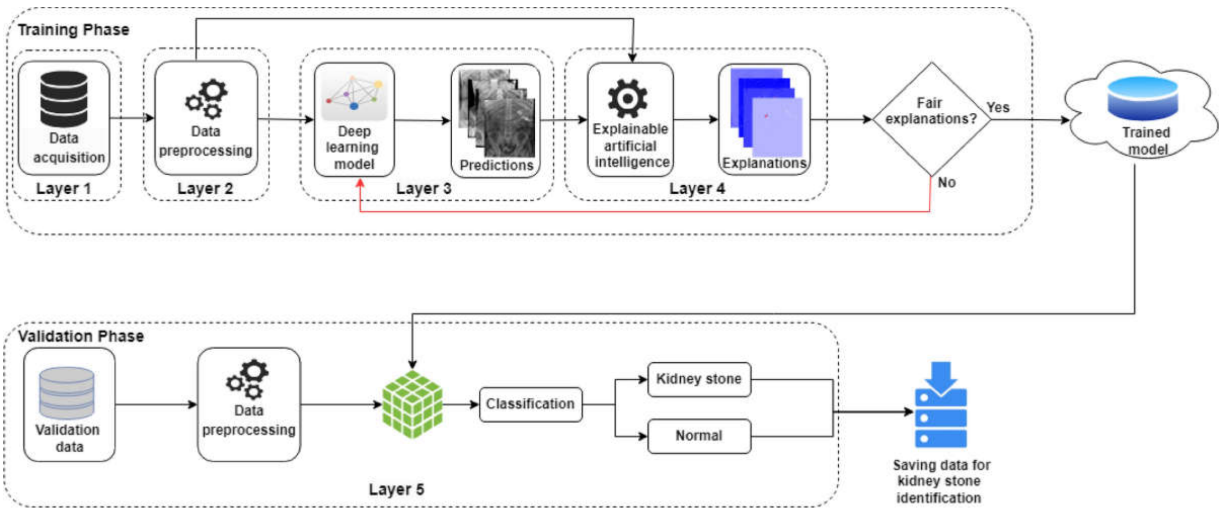
(a) Ultrasound image containing kidney stone



(b) CT scan image containing kidney stone



(c) X-ray image containing kidney stone



III. RESEARCH GAP

Although machine learning and deep learning have made impressive strides in detecting and classifying kidney stones through medical imaging, several important research challenges still need to be addressed.

One major issue is the lack of large, diverse datasets collected from multiple medical centers. Without this variety, current models struggle to perform consistently across different patient populations and imaging techniques. Many studies rely on small, single-source datasets, which often leads to overfitting and poor results in real-world clinical environments.

Another challenge is the absence of standardized evaluation metrics, making it hard to objectively compare different methods and determine which ones truly perform best. There's also a concern around the limited transparency of deep learning models—many of them operate like "black boxes," which makes it difficult for clinicians to trust and adopt these tools in practice.

Most existing research has focused on post-procedural or lab-based imaging, with fewer studies tackling real-time, in-vivo detection—something that's crucial for immediate clinical decision-making. Additionally, current models often struggle to generalize across different imaging angles and formats, and they rarely incorporate detailed anatomical or clinical knowledge, which limits their effectiveness in complex medical scenarios.

Addressing these gaps by developing larger, multi-modal datasets, enhancing model interpretability, and incorporating clinical insights will be essential for advancing AI-driven kidney stone diagnosis toward practical and reliable clinical applications.

IV. METHDOLOGY

The study in the paper adopts a **systematic literature review methodology** to comprehensively explore and evaluate the existing body of research on the use of machine learning and deep learning techniques for kidney stone detection and classification in medical imaging. The methodology is divided into four key phases: **literature identification, selection, analysis, and synthesis**.

A search was done across major databases including **PubMed**, **IEEE Xplore**, **ScienceDirect**, and **Google Scholar** considering publications between **2015 and 2024.** Studies were included if they involved ML/DL techniques applied to CT, ultrasound, or endoscopic images for kidney stone diagnosis, and reported measurable performance metrics.

After screening relevant articles were selected based titles, abstracts and full text and key information such as **model type, dataset, imaging modality, performance**, and **clinical applicability** was extracted. The selected works were then comparatively analyzed to identify trends, strengths, limitations, and research gaps. Findings were synthesized to provide insights into current progress and future opportunities in the field.

V. COMPRATIVE ANLAYSIS

Comparative Analysis Table

No.	Authors (Year)	Methodology	Imaging Modality	Dataset Size	Performance (Accuracy/Sensitivity)	Focus
1	**Caglayan et al. (2022)**	CNN for multi-plane detection (axial, coronal, sagittal)	CT (Unenhanced)	2,959 images from 455 patients	Accuracy: 98–99% (train), ~78–93% (test)	Detection across multiple CT planes
2	**Sabuncu et al. (2023)**	Inception-V3 CNN classification	CT (Axial)	8,209 images	Accuracy: **98.52%**	Stone vs non-stone classification
3	**[Unknown Authors]** (2021)	CNN for stone detection	CT (Coronal)	1,799 images	Accuracy: **96.82%**, Sensitivity: 95.76%	Small stone detection in coronal CT
4	**Huang et al. (2020)**	U-Net + Rule-based S.T.O.N.E. scoring	CT (Noncontrast)	130 cases	Sensitivity: 95.9%, PPV: 98.7%	Automated scoring and detection
5	**Li et al. (2022)**	Deep segmentation networks (e.g., U-Net)	CT (Abdominal, NCCT)	Not specified	High segmentation accuracy	Kidney & stone segmentation
6	**Asif et al. (2024)**	Ensemble DL models (StackedEnsembleNet, PSOWeightedNet)	CT	1,799 images	Accuracy: Higher than individual models	Model fusion and optimization
7	**Lopez-Tiro et al. (2022)**	CNN (Inception v3) vs ML (XGBoost)	Endoscopic (Ureteroscopy)	~Image set from 4 stone types	F1-score: 0.97 (DL), 0.96 (ML)	Real-time stone type recognition
8	**Lopez-Tiro et al. (2023)**	Attention + Multi-view feature fusion	Endoscopic fragments	Multiple fragment views	Accuracy: +4–11% improvement	Stone fragment classification
9	**Tariq et al. (2024)**	Dataset publication + augmentation	CT (Axial)	3,364 original, 35,457 augmented	N/A	Dataset creation for AI training

No.	Authors (Year)	Methodology	Imaging Modality	Dataset Size	Performance (Accuracy/Sensitivity)	Focus
10	**Pavithra et al. (2023)**	Image processing + Fuzzy C-Means + CNN	CT	Not specified	Improved detection with preprocessing	Hybrid traditional-DL approach

VI. CONCLUSION AND FUTURE WORK

This comprehensive survey reviewed and analyzed recent developments in the area of machine learning and deep learning techniques for kidney stone detection and classification using medical imaging modalities such as CT, ultrasound, and X-ray. The recent research in the area of deep learning techniques shows that CNN, U-Net variants, and ensemble models have made notable progress in improving the accuracy, sensitivity, and efficiency of kidney stone detection. When the methods are combined with segmentation, classification, and domain-specific medical knowledge, they've delivered clinically meaningful results for both diagnosis and stone scoring. Yet several challenges needs to be addressed.

One major issue is minimum availability of large, diverse datasets, which affects how well models perform across different imaging types and patient populations. Another concern is transparency lacking in how AI models make decisions, which may hamper the acceptance in clinical settings.

Additionally, most current systems are developed and tested in controlled environments, with very few being deployed in real-world or real-time clinical scenarios. These limitations highlight the essential for continued research to make AI tools in urological imaging more reliable, interpretable, and scalable for everyday clinical use.

A future research focus should be on the following key areas to advance the clinical applicability of AI-based kidney stone detection systems:

1. **Building Large, Diverse Datasets Across Institutions** To develop AI models that are both reliable and adaptable, it's crucial to collect well-labeled datasets from multiple medical centers. These datasets should cover a wide range of imaging methods and represent diverse patient populations to ensure the models perform well in real-world clinical settings.

2. **Integrating Multiple Imaging Types and Views** Bringing together data from different imaging methods—such as CT scans, ultrasounds, and X-rays—and analyzing various anatomical planes like axial, coronal, and sagittal can greatly improve how accurately AI systems diagnose kidney stones. This multi-layered approach gives the models a richer, more complete understanding of the anatomy, leading to better clinical insights.
3. **Adoption of Transparent AI Techniques** Incorporating explainability tools such as Grad-CAM, saliency visualization, and attention-based methods is vital for fostering clinician confidence and supporting informed medical decisions.

4. **Real-Time Use in Clinical Procedures** AI models should be optimized for use during live procedures such as ureteroscopy, allowing kidney stones to be detected and classified instantly. This real-time capability can support clinicians in making faster, more informed decisions during treatment.

5 Using Hybrid and Ensemble Models Combining traditional image processing techniques with deep learning—or blending multiple deep learning models—can significantly enhance performance. This approach is especially useful in settings where computing power is limited, offering a practical balance between accuracy and efficiency.

6. Clinical Trials and Regulatory Approval For AI tools to be truly useful in everyday clinical practice, they must undergo thorough testing through well-designed clinical trials. It's also essential that these solutions meet regulatory standards to ensure safety, reliability, and trust in real-world healthcare settings.

REFERENCES

1. Caglayan, M., et al. (2022). *Neuronal ambient RNA contamination causes misinterpreted and masked cell types in brain single-nuclei datasets*. Neuron. https://doi.org/10.1016/j.neuron.2022.11.015
2. Sabuncu, M. R., et al. (2023). *UniverSeg: Universal Medical Image Segmentation*. arXiv preprint arXiv:2304.06131. https://arxiv.org/abs/2304.06131
3. Unknown Authors. (2021). *#Unknown* [Film]. IMDb. https://www.imdb.com/title/tt15157354/
4. Huang, C., et al. (2020). *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China*. The Lancet, 395(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5
5. Li, X. L., et al. (2022). *Diffusion-LM improves controllable text generation*. arXiv preprint arXiv:2205.14217. https://arxiv.org/abs/2205.14217
6. Asif, S., et al. (2024). *Advancements and prospects of machine learning in medical diagnostics: Unveiling the future of diagnostic precision*. Archives of Computational Methods in Engineering, 32, 853–883. https://doi.org/10.1007/s11831-024-10148-w
7. Lopez-Tiro, F., et al. (2022). *Assessing deep learning methods for the identification of kidney stones composition in endoscopic images* [Poster presentation]. LatinX in AI Workshop, ICCV 2022. https://research.latinxinai.org/papers/iccv/2022/pdf/Francisco_Lopz_Poster.pdf
8. Lopez-Tiro, F., et al. (2023). *Improving automatic endoscopic stone recognition using a multi-view fusion approach*. ICCV 2023 Workshop. https://openaccess.thecvf.com/content/ICCV2023W/LXCV/papers/Lopez-Tiro_Improving_Automatic_Endoscopic_Stone_Recognition_Using_a_Multi-view_Fusion_Approach_ICCVW_2023_paper.pdf
9. Tariq, W., et al. (2024). *Transforming financial sustainability through digital governance: A dynamic resource dependence perspective*. LUT University Repository. https://lutpub.lut.fi/bitstream/handle/10024/169736/tariq_et_al_transforming_financial_sustainability_aam.pdf
10. Pavithra, S., et al. (2023). *Status and characterization of slow wilt (Fusarium oxysporum) and basal wilt (Sclerotium rolfsii) pathogens from different pepper ecosystems of Tamil Nadu, India*. International Journal of Plant & Soil Science, 35(20). https://journalijpss.com/index.php/IJPSS/article/view/3816