Grenze International Journal of Engineering and Technology, June Issue

Automated Stone Detection using CNN and YOLOv5 Approach with Deep Learning Techniques

Divyadharshini J[1], Yogesh Kumar S[2], Srinidhi S[3] and Manikandan K[4]
[1-4]Department of Computational Intelligence, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India
Email: dharshinijai2408@gmail.com, syogeshk23@gmail.com, srinidhi2829@gmail.com, kmanikandan@vit.ac.in

***Abstract*— Kidney diseases, including stones, cysts, and tumours, represent a serious global health issue, requiring precise and timely diagnosis. Traditional diagnostic methods relying on manual image analysis are often time-consuming and prone to errors. This study presents an automated deep learning-based system for kidney disease and stone detection. The proposed framework frame-work employs Convolutional Neural Networks (CNNs) for feature extraction and classification, while YOLOv5 is utilized for real-time Kidney Stone detection in CT and MRI images. Image preprocessing is conducted using OpenCV-Python, with model development facilitated by PyTorch and interpretability enhanced through Grad-CAM. Experimental results demonstrate high accuracy and efficiency, minimizing human intervention and improving diagnostic reliability. Future work aims to integrate the system into clinical workflows and expand the dataset to enhance its generalization capability.**

***Index Terms*—Kidney Disease, Deep Learning, CNN, YOLOv5, Medical Imaging, Multi-class Classification, Grad-CAM.**

I. INTRODUCTION

Urinary calculi development, commonly referred to as Renal stones or Kidney Stones, is a severe and incapacitating issue that affects all communities globally. The condition known as Kidney Stones in medicine is referred to as Ureterolithiasis, Urolithiasis, or Nephrolithiasis. More precisely, the medical names for the various locations of the stones in the Kidney, Urinary system, and Ureter. Kidney Stones, solid mineral and salt deposits that form in the Kidneys, it can cause severe pain and complications to the extent of Renal failure if improperly diagnosed and treated. Kidney Stones are usually classified as calcareous stones and non-calcareous stones according to the chemical makeup of the crystals. Based on this criterion they are further classified into Calcium oxalate, Calcium phosphate, Uric acid, Struvite, and Cystine. There are more than 250 kinds of Kidney Stones found in nature. Kidney Stones can vary in size, shape, and color, and can range from the size of a grain of sand to a golf ball. Small stones (up to 3 mm) can pass without causing pain, but stones between 3 and 5 mm can cause pain while passing down the ureter. There are numerous advancements in understanding and treatment modalities, but still the Kidney Stones continue to be a great challenge for patients and medical professionals.
The branch of Artificial Intelligence that truly offers a far superior answer to this problem is Deep Learning. CNNs allow Deep Learning algorithms to automatically extract and learn features of images from various Imaging techniques, potentially leading to accurate and efficient Kidney Stone detection.

Grenze ID: 01.GIJET.11.2.228_25

These models can differentiate between normal Kidney and Kidney Stone forms and accurately identify them because they were trained on very large datasets of annotated medical images.
Using Deep Learning Techniques to identify Kidney Stones from the data set has recently become a very strong approach to the diagnosis part of the medical imaging. Since radiologists manually diagnose conditions using ultrasound and CT scans, which are time-consuming, to avoid these traditional procedures, development of an ML model comes into picture.
Collecting diverse and high-quality datasets is crucial for developing an effective Kidney Stone detection system. Preparing image content involves refining raw images for tasks like visualization, feature extraction, or training machine learning models. The specific steps vary based on the intended use, ensuring the data is optimized for accurate analysis.

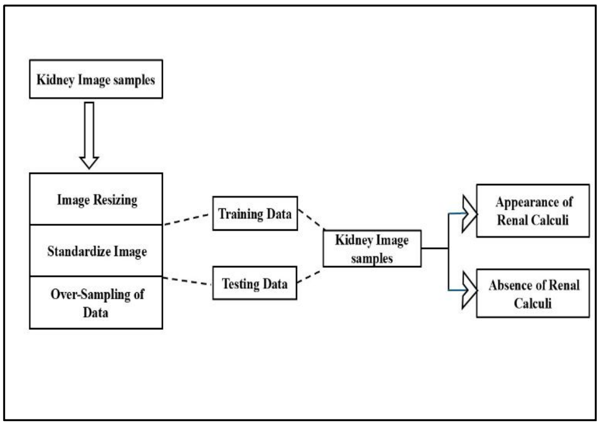
II. Literature Review

Convolutional Neural Networks (CNNs), a Deep Learning approach, have led to an extraordinary improvement in kidney stone detection. and You Only Look Once version 5 (YOLOv5). Traditional methods of Kidney Stone diagnosis, such as ultrasound and CT scans, are likely to be reliant upon human interpretation, which is slow and prone to errors. Automated detection using Deep Learning offers a promising alternative by improving accuracy, reducing diagnosis time, and minimizing the dependency on expert radiologists [3-10].
Convolutional Neural Networks (CNNs) in Kidney Stone Detection: CNNs have become a fundamental tool in medical image analysis due to their ability to extract features and recognize patterns in images. Research studies have demonstrated that CNN architectures, such as VGG16, ResNet, and EfficientNet, can achieve high classification accuracy in Kidney Stone detection. These models work by learning spatial hierarchies of features, enabling them to differentiate between stones, cysts, tumors, and normal kidney structures.
For instance, a study by Salman F. Rabby et al. (2023) explored a CNN-based approach for classifying Kidney Stones in ultrasound images, achieving an accuracy of over 90%. Another study by Khasanov et al. (2024) integrated CNN with transfer learning, utilizing pre-trained models to enhance performance with limited datasets. Such approaches significantly improve the reliability of stone detection and classification, making CNNs a viable option for automated diagnostics.
YOLOv5 for Real-Time Stone Detection: Unlike traditional CNN models that focus on classification, YOLOv5 excels in object detection by identifying and localizing Kidney Stones in real-time. YOLOv5 is known for its speed and efficiency, making it suitable for clinical applications where rapid diagnosis is crucial [5].
It has been recently demonstrated through research that YOLOv5 can accurately identify Kidney Stones in ultrasound scans. For example, Mangala Shetty et al. (2024) developed a YOLOv5-based model that achieved a mean Average Precision (MAP) of over 85% in detecting and localizing Kidney Stones from real-world datasets. The model's ability to process images quickly while maintaining high accuracy makes it a strong candidate for integration into automated diagnostic systems.
Comparison and Integration of CNN and YOLOv5 Approaches: While CNNs are highly effective in classifying Kidney Stones, they lack the ability to localize them precisely within an image[3-5]. On the other hand, YOLOv5 is excellent at detecting and pinpointing the exact location of stones but may not always provide detailed classification. A hybrid approach combining CNNs for classification and YOLOv5 for localization could provide a comprehensive solution for Kidney Stone detection.
Several recent studies have explored this combined approach. For instance, Stefan C et al. (2024) designed a model that first used CNNs to classify Kidney Stones and then employed YOLOv5 to detect and highlight their positions in ultrasound images. This two-step process enhanced overall accuracy and robustness, making automated Kidney Stone detection more reliable and efficient.
Challenges and Future Directions: Despite significant progress, several challenges remain in implementing deep learning-based stone detection in real-world clinical settings. These include the need for large, high-quality annotated datasets, handling variations in ultrasound image quality, and ensuring model generalizability across different patient demographics.
Future research could focus on refining model architectures, incorporating more diverse datasets, and improving interpretability to gain better trust from medical professionals. Additionally, integrating deep learning models into real-time clinical workflows with user-friendly interfaces could accelerate the adoption of automated Kidney Stone detection in hospitals and diagnostic centres.
The use of CNNs and YOLOv5 in Kidney Stone detection presents a powerful advancement in medical imaging. While CNNs offer strong classification capabilities, YOLOv5 excels in real-time detection and localization. A combined approach leveraging both techniques could lead to more accurate and efficient automated diagnostic

systems. With further improvements and clinical validation, deep learning-based Kidney Stone detection has the potential to revolutionize nephrology diagnostics, improving patient outcomes through faster and more reliable detection methods [13].

III. Methodology

Collecting and Preparing Information Image Acquisition: Prepare for training and testing using kidney CT scan pictures. Image preparation entails scaling photos to meet the YOLOv5 model's input dimensions, normalizing pixel values, and performing a range of data augmentation techniques, such as flipping and rotation. Labeling: To precisely identify the area of interest (such as kidney stones), annotate photos with bounding boxes. It will be wise to use tools such as Labeling to help with annotation [7].
Data balancing: Balance and Improve the distribution of the data utilizing SMOTE (Synthetic Minority Over-sampling Technique).
Model Training with YOLOv5 Architecture Selection: Use YOLOv5's built-in configurations (e.g., YOLOv5s, YOLOv5m) based on the complexity of the task. Model Training: Train YOLOv5 on the annotated dataset using a split of 70% training and 30% testing. Loss Functions: Use object localization and classification loss functions for optimizing model accuracy [7-8].



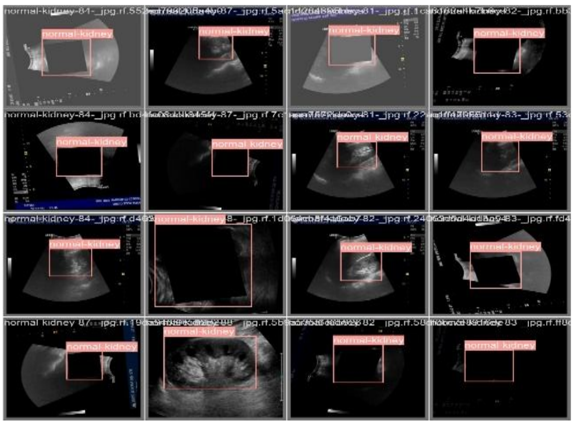


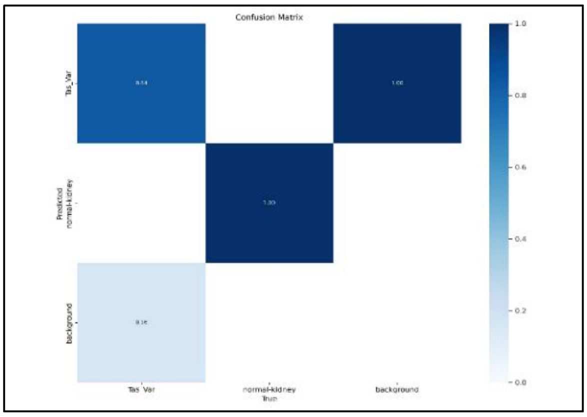
Figure 1. Proposed Model for Data Preprocessing, Splitting and Identification

Figure 2. Normal Kidney Images

Testing and Validation: In order to increase the performance again the model is verified on the testing dataset and the hyperparameters of the model are adjusted so that it is possible to reset the batch size and learning rate too. The metrics like accuracy, recall, F1-score, and MAP (mean average precision) can be used to gain an insight into the performance of a model being considered in evaluation.
After Processing: Use non-maximum suppression to remove duplicate detection which means that only one detection from each position will be retained. A final prediction will be made use of bounding boxes to indicate whether renal calculi are present in an image or not [7].
Integration and Deployment: A trained YOLOv5 model will be utilised to conduct real time detection within clinical environments. Optimise accuracy within the operation at all times so as to ensure that any new data samples are also taken into account and that procedures are continually being upgraded in a bid to catch up with the latest data samples[9]. Fig.2 shows the general kidney structures on CT (Computed Tomography) images, we used the YOLOv5 (You Only Look Once) object detection model. The machine was trained on an annotated corpus of ultrasound images with bounding boxes drawn around normal kidneys. Some of the steps involved in the detecting phase were: Preprocessing: To fulfill the input requirement of YOLOv5, the input CT images were resized and normalized. Model Inference: Post-image processing, the YOLOv5 model generated bounding boxes and confidence scores. Detection and Classification: By indicating the highlighted region within a red bounding box, the model was able to detect the "normal-kidney" region in all images.

IV. Results and Discussions

Figure. 3 Shows how the confusion matrix obtained the performance evaluation of a classification model which consists of three classes: Tas_Var, normal-kidney, and background. The model correctly identified 84% of the Tas_Var cases and 16% of Tas_Var cases were misclassified as background, which could indicate some overlap in features between stones and the background region. This model achieved 100% accuracy in identifying normal kidney stones that indicate normal kidney stones have distinctive features that the model recognizes with high confidence [10].



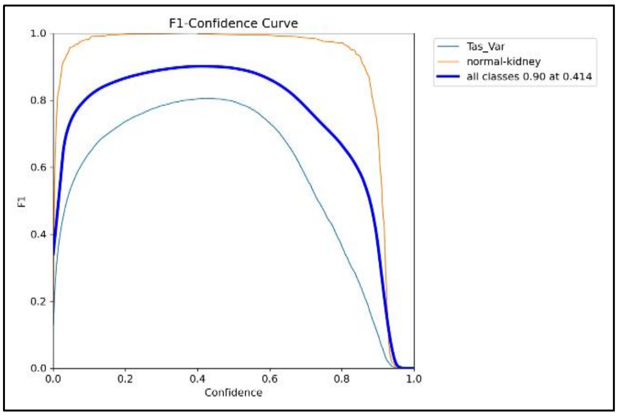
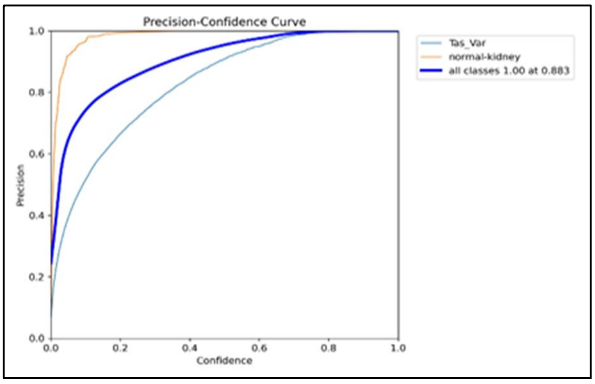


Figure 3. Confusion Matrix for YOLO V5 model

Figure 4. F-1 Confidence Curve

Figure. 4 Shows how the model obtains a maximum F1 value of 0.90 with confidence threshold of 0.414. The normal-kidney class maintains a higher F1 score across most confidence levels, showing that normal kidney stone is easier to classify more accurately. Tas_Var class has a slightly lower F1 curve which is more challenging to classify due to feature overlaps with other classes. As confidence increases beyond 0.8, the F1 score drops sharply, meaning that the model becomes highly confident on its predictions, it sacrifices recall, leading to missed detections. threshold of 0.414 is ideal for maximizing F1 score. Increasing, the confidence threshold reduces false positives but at the cost of more false negatives [10].



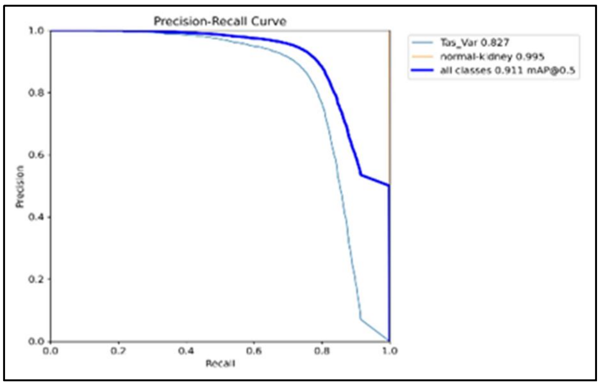


Figure 5. Precision Confidence Curve

Figure 6. Precision-Recall Curve

Figure 5. Shows how the model achieves high confidence levels, meaning that there are almost no false positives in its predictions. The model produces more false positive predictions at lower confidence thresholds because accuracy is reduced. Precision improves at excellent levels above 0.8 confidence as confidence rises. Low recall may result from high precision at 0.883 confidence, which guarantees almost error-free predictions. Given that normal kidney tissues are the simplest to identify, the normal-kidney class has the highest precision value across all confidence levels.

The lower precision curve of the Tas_Var class suggests possible misclassifications when distinguishing kidney abnormalities or stones from normal kidney images[10].

Figure 6. Shows how the precision-recall curves in the one on the left is a distinct class or general performance indicator. Tar Var class The Average Precision (AP) score for this class is 0.827. This class's curve demonstrates that precision begins high and then sharply declines when recall rises over 0.8, suggesting that the model begins to produce more false-positive errors as it attempts to catch more positives.

The AP score of 0.995 for the typical class is nearly flawless. The model is quite successful at accurately recognizing "normal-kidney" instances with little false positives, as indicated by the curve remaining close to the top, which shows very high precision across nearly all recall levels.

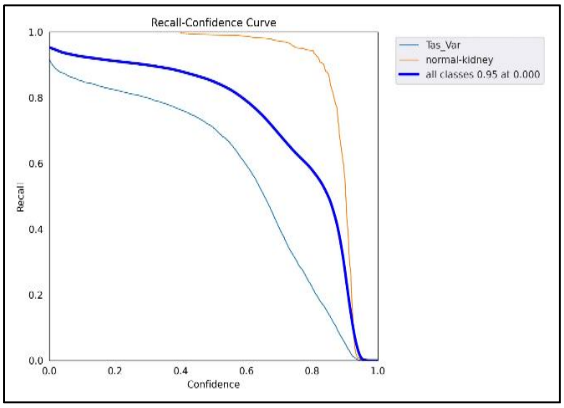


Figure 7. Recall Confidence Curve

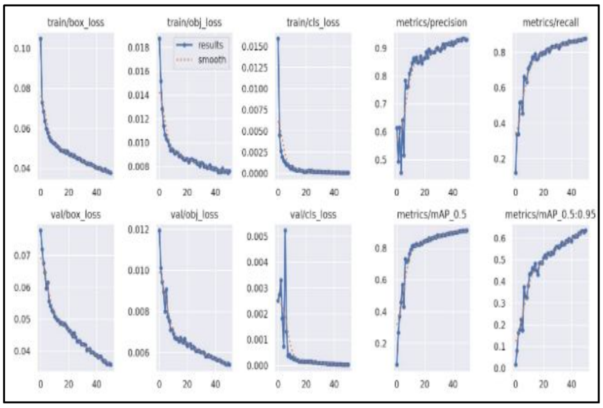


Figure 8. Metrics for YOLOv5 Training and Validation

Figure 7. Shows how This Recall-Confidence Curve depicts a kidney stone detecting model's robustness. Precise classification, in this case, for regular kidneys is captured by the graph of the healthy kidney's reliable recall at the different confidence level. As the confidence increases, the Tas_var and recall for all classes decrease, which means there is increased vagueness in the diagnosis of kidney stones. The general performance of the model at 0.95 confidence level illustrates how recall decreases with increased confidence. A high AUC indicated good detection ability. The sensitivity of stone detection may be increased by modifying the confidence threshold.

Figure 8. shows The YOLOv5 model's training and validation metrics are shown in the above figure with 2x5 grid represents the model's training behavior across about 40 epochs.

Metrics for Training Box Loss: It indicates improved localization accuracy since it reflects the inaccuracy in guessing the bounding box coordinates, which has dropped from 0.10 to 0.04. Objective Loss indicates increased identification as the model's ability to distinguish objects from the background has improved from 0.018 to 0.008 points.

Classification Loss: Indicates that the model rapidly learns to correctly categorize items, as it drops from 0.015 to almost 0.000.

High Accuracy: It also has a strong sensitivity to a reduction in the false positives: 0.15 to 0.95.

Recall: The range of 0.15 to 0.90 shows that the model is effective in locating pertinent objects as well.

Metrics for Validation:Good generalization is shown by the validation box loss, which drops from 0.075 to 0.035 and closely resembles the training box loss.Robust object recognition on unseen data is confirmed by the validation objectness loss, which smoothly decreases from 0.012 to 0.006.The validation classification loss shows robust classification performance on the validation set, which precipitously declines to almost nothing.

mAP@0.5: Demonstrate excellent object detection with lax overlap threshold criteria, increasing from 0.15 to 0.9.

mAP@0.5:0.95: Demonstrate very good performance also under more tight localization criteria and increase from 0.05 to 0.65.

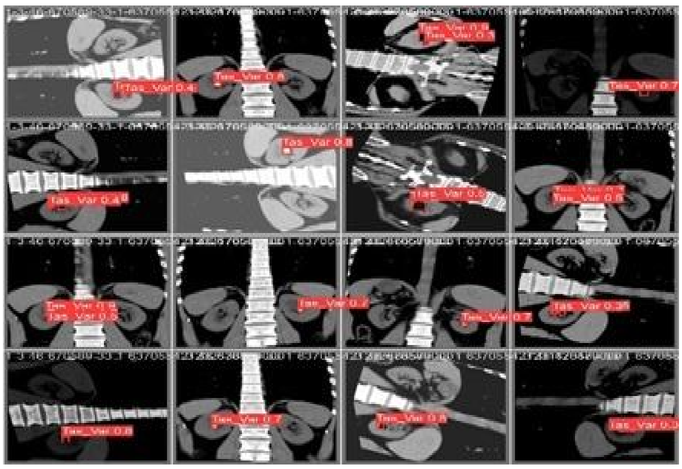


Figure 9. Images of Kidney *with* a Stone

Preprocessing: Computed Tomography Scan images were resized and normalized for YOLOv5 input.
Model Inference: The trained model generated bounding boxes with confidence scores.
Detection & Classification: Kidney stones were highlighted with red bounding boxes labeled as “Tas_Var.”
Post-processing: Detected outputs were visualized to assess accuracy.
This method provides an efficient, real-time kidney stone detection system, aiding in automated medical diagnosis and treatment planning.

V. Authors Contributions

Conceptualization of idea for article was done by Divyadharshini J. The literature survey and data analysis were performed by Divyadharshini J, Srinidhi S and Yogesh Kumar S. Drafting, tabulating and pictures was drafted by Divyadharshini J and Srinidhi S. Coding and Implementation were carried out by Yogesh Kumar S, Divyadharshini J and Srinidhi S. Final manuscript editing and supervision over the work was done by Manikandan K.

VI. Conclusion

It employs CNNs and YOLOv5 and introduces an AI-driven kidney stone diagnosis method to ensure high precision in real-time CT/MRI analysis. By integrating localization with classification algorithms, the approach avoids human effort to enhance diagnostic reliability. Grad-CAM helps physicians by improving the interpretability process.
Future research will focus on accuracy improvement, increasing datasets, and enhancing performance in real time. A significant goal is the integration of friendly interfaces into the healthcare workflow. Federated learning can protect privacy while improving robustness.
For effective deployment in medical contexts, real-world validation and cooperation with healthcare practitioners will be crucial.

Acknowledgment

The authors would like to acknowledge the support of Vellore Institute of Technology, Vellore on providing us the Access over journals and work space accommodation.

References

[1] Murillo Freitas Bouzon, S. Patrício, O. Eduardo, and S. Silva, “Automatic Kidney Stone Detection using Low-cost CNN with Coronal CT Images,” Nov. 2023, doi: https://doi.org/10.5753/wvc.2023.27527.
[2] “Electronics,” doi: https://doi.org/10.3390/electronics.
[3] M. M. Billah, Abdullah Al Rakib, M. I. Haque, A. S. Ahamed, M. S. Hossain, and Kamrun Nahar Borsha, “Real-Time Object Detection in Medical Imaging Using YOLO Models for Kidney Stone Detection,” European Journal of Computer Science and Information Technology, vol. 12, no. 7, pp. 54–65, Jul. 2024, doi: https://doi.org/10.37745/ejcsit.2013/vol12n75465.
[4] Adnin Ramadhani and A. Salam, “Deployment of Web-Based YOLO for CT Scan Kidney Stone Detection,” SinkrOn, vol. 8, no. 3, pp. 1357–1368, Jul. 2024, doi: https://doi.org/10.33395/sinkron.v8i3.13744.
[5] M. A. Islam et al., “Comprehensive Analysis of CNN and YOLOv5 Object Detection Model to Classify Phytomedicine Tree’s Leaf Disease,” Research Square (Research Square), Nov. 2022, doi: https://doi.org/10.21203/rs.3.rs-2099534/v2.
[6] C. Albuquerque, R. Henriques, and M. Castelli, “Deep learning-based object detection algorithms in medical imaging: Systematic review,” Heliyon, vol. 11, no. 1, p. e41137, Jan. 2025, doi: https://doi.org/10.1016/j.heliyon.2024.e41137.
[7] Krishna Sowjanya K, Bindu Madavi K P, Neha Patwari, and S. D. A, “DeepKidney: Multiclass Classification of Kidney Stones, Cysts, Tumors, and Normal Cases Using Convolutional Neural Networks,” Dec. 2023, doi: https://doi.org/10.1109/c2i659362.2023.10431043.
[8] A. Soni and A. Rai, “Kidney Stone Recognition and Extraction using Directional Emboss & SVM from Computed Tomography Images,” Dec. 2020, doi: https://doi.org/10.1109/mpcit51588.2020.9350388.
[9] T. Rahman and Mohammad Shorif Uddin, “Speckle noise reduction and segmentation of kidney regions from ultrasound image,” May 2013, doi: https://doi.org/10.1109/iciev.2013.6572601.
[10] Sabri Koçer, O. Mohamed, and Özgür Dündar, “Disease Detection in Abdominal CT Images Using the YOLOv5 Algorithm: A Deep Learning Approach,” 2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST), pp. 1–4, Jul. 2024, doi: https://doi.org/10.1109/icest62335.2024.10639613.

[11] A. Martinez et al., “Towards an automated classification method for ureteroscopic kidney stone images using ensemble learning,” HAL (Le Centre pour la Communication Scientifique Directe), Jul. 2020, doi: https://doi.org/10.1109/embc44109.2020.9176121.
[12] M. Akshaya, R. Nithushaa, N.Sri Madhava Raja, and S. Padmapriya, “Kidney Stone Detection Using Neural Networks,” 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Jul. 2020, doi: https://doi.org/10.1109/icscan49426.2020.9262335.
[13] H. Zhang, Y. Xu, and J. Liu, “Research on Medical Image Object Detection Based on Swin-Transformer and YOLOv5,” 2022 IEEE 4th International Conference on Power, Intelligent Computing and Systems (ICPICS), pp. 1069–1074, Jul. 2023, doi: https://doi.org/10.1109/icpics58376.2023.10235682.