

Optimizing Kidney Stone Detection: Exploring YOLOv8 Variants for Computational Efficiency and Enhanced Accuracy

Amol Satsangi*

Department of Electrical Engineering, Faculty of Engineering, Dayalbagh Educational Institute, Agra – 282005, India

E-mail: amolsatsangi2104218@dei.ac.in

ORCID iD: <https://orcid.org/0009-0005-8611-5654>

*Corresponding author

Shaurya Jain

Department of Electrical Engineering, Faculty of Engineering, Dayalbagh Educational Institute, Agra – 282005, India

E-mail: shauryajain2104187@dei.ac.in

ORCID iD: <https://orcid.org/0009-0008-0548-3167>

Subho Upadhyay

Department of Electrical Engineering, Faculty of Engineering, Dayalbagh Educational Institute, Agra – 282005, India

E-mail: subhome2k10@dei.ac.in

Received: 02 June 2025; Revised: 28 July 2025; Accepted: 26 August 2025; Published: 08 October 2025

Abstract: Kidney stones are solid mineral and salt deposits formed within the kidneys, causing excruciating discomfort and pain when they obstruct the urinary tract. The presence of speckle noise in CT-scan images, coupled with the limitations of manual interpretation, makes kidney stone detection challenging and highlighting the need for precise and efficient diagnosis. This research investigates the efficacy of YOLOv8 models for kidney stone detection, aiming to strike a balance between computational efficiency and detection accuracy. This study's novel evaluation framework and practical deployment considerations underscore its contributions to advance kidney stone detection technologies. It evaluates five YOLOv8 variants (nano, small, medium, large, and extra-large) using standard metrics such as precision, recall, F1-score, and mAP@50, alongside computational resources like training time, power consumption, and memory usage. The comprehensive evaluation reveals that while YOLOv8s and YOLOv8e demonstrate superior performance in traditional metrics, YOLOv8s emerges as the optimal model, offering a harmonious balance with its high precision (0.917), highest mAP@50 (0.918), moderate power consumption (150W), and efficient memory usage. Graphical analyses further elucidate the behaviour of each model across different confidence thresholds, confirming the robustness of YOLOv8s. Additionally, this research explores the impact of model size and complexity on inference speed, demonstrating that smaller YOLOv8 variants achieve real-time performance with minimal latency. The study also introduces a method for model scalability, allowing for adjustments in accuracy and computational demand based on specific clinical or resource constraints. These contributions further emphasize the importance of holistic model assessment for real-world medical applications.

Index Terms: Kidney Stone Detection, Computer Aided Diagnosis (CAD), YOLOv8, Computational Efficiency, Deep Learning.

1. Introduction

Kidney stones, scientifically known as renal calculi or nephrolithiasis or urolithiasis, are salt deposits and solid mineral formed within the kidneys. It provides an excruciating discomfort and pain when it obstructs the urinary tract. It is an increasing urological disorder of human health, affecting about 12% of the world population [1]. The numerous types of kidney stones include calcium oxalate stones, uric acid stones, calcium phosphate stones and struvite stones [2]. Each of these exhibits distinguishable characteristics and require specific treatments.

The primary issue in kidney stones is the lack of early prediction. This happens because, in most cases, the stones

continue to grow and become painful as they increase in size. Due to this delay in detection and diagnosis, it can lead to severe pain, urinary tract infections, and ultimately kidney damage. This disease affects people irrespective of age, gender, and race. Recent research of Zhu Wang *et al.* [3] indicated that kidney stone cases are reported increasing worldwide with 5.8% among Chinese adults affected which about 1 in 17 adults currently have disease. Hence accurate and early becomes important.

To address this problem, medical imaging methods are used as the first step of examination. Many medical imaging types, such as X-ray imaging, computed tomography (CT), and ultrasound, have become significant for diagnosis. These methods require a trained practitioner for accurate detection, which can be time-consuming and lead to delays in treatment. Additionally, the potential for human error can result in inaccurate treatment, and the need for human expertise makes it a costly process. Therefore, this research proposes a computer vision-based algorithm that can be implemented into CAD systems. This approach contemplates to achieve better performance in detecting kidney stones and can be deployed for real-time tasks to assist in diagnosis.

YOLO (You Only Look Once) is a cutting-edge deep learning technique for disease detection in medical imaging [4-7]. Unlike traditional CNNs (Convolutional Neural Network), YOLO is optimized for speed and efficiency, making it ideal for real-time detection. It works by dividing the image into a grid, where each cell predicts bounding boxes and associated class probabilities, allowing for rapid and accurate identification of objects within medical images.

For this study, YOLOv8 was chosen for its improved efficiency, accuracy, and speed, making it ideal for real-time kidney stone detection. Its lightweight design allows deployment on low-resource hardware, ensuring wider clinical use. This study investigates the best YOLOv8 version for the detection of kidney stones using CT-scan images, achieving an optimal balance between computational efficiency and accuracy of prediction. This ensures that the system can be embedded in hardware with lower computational capabilities, making it accessible and practical for widespread clinical use. The main contributions of the research are as follows:

- Investigation of YOLOv8, providing insights into its capabilities for kidney stone detection within medical imaging applications.
- Comparison between all variants of YOLOv8 facilitates a nuanced understanding of each model's unique properties.
- The study introduces a novel evaluation framework that not only incorporates traditional performance metrics like precision and recall but also integrates considerations of computational resources such as training time, power consumption, and memory usage.
- The study has achieved high precision (0.917), high mAP@50(0.918), moderate power consumption and efficient memory usage, making it suitable for real time clinical use.
- The study also demonstrated the potential of smaller YOLOv8 models to achieve real time kidney stone detection with minimal latency, making it application in hardware-constrained clinical environments.

The organization of the research paper is as follows: Section 2 reviews prior work in kidney stone detection. Section 3 covers the YOLO framework and its versions. Section 4 details the methodology, including the dataset, model versions, experiment setup, training, and evaluation. Section 5 presents results and graphical analysis. Section 6 provides conclusion and future research directions have been highlighted in Section 7.

2. Related Works

In the field of kidney stone detection, various studies have utilized deep learning and machine learning techniques. Deep learning offers algorithms that are effective for image segmentation [8-9], detection [10-11], and classification [12-13]. Deep learning is becoming increasingly valuable in diagnosis system [9, 11, 13, 14-15]. Among the pertinent works is the early detection of chronic kidney disease using machine learning techniques [16].

Shi *et al.* [17] presented a segmentation based on fully automatic for kidney images, employing pre-trained CNN to extract relevant features. In [18], the new deep learning model with adaptive weighting score is used to predict a chronic kidney disease with 0.912 AUC. Abdullah *et al.* [16] focused on early detection of chronic kidney disease based on artificial neural networks (ANN, 99.75%) and support vector machines (SVM, 97.75%) techniques. Baygin *et al.* [19] developed ExDark19 for an automated kidney stone detection with 99% accuracy using CT images.

Ahmet Furkan Bayram *et al.* [20] proposed SOTA one stage detector YOLOv7 for detecting kidney disease with CT image. They experimented on both YOLOv7 and YOLOv7 tiny with results 84.9% and 70.3% mAP50 respectively. Authors [21] obtained a kidney segmentation IOU of 91.4% and kidney stone detection accuracies of 99.5% for YOLOv7 and 98.7% for YOLOv5.

The study in [22] enhances kidney stone detection in Kidney, Ureter and Bladder (KUB) X-Ray images by combining super resolution techniques with the redesigned YOLOv7, achieving a sensitivity of 91.7% and a precision of 97.3%. Another study [23] on YOLOv7, proposes an effective kidney stone prediction method with integration of Energy Valley optimizer for segmentation and Pulse Couple Neural Network (PCNN) for classification. This combination achieved superior accuracy of 98.88% compared to traditional techniques. The research [24] develops a kidney stone detection system using YOLOv5 achieving high performance metrics (average mAP of 95.71% and

accuracy of 93.06%) and integrated it into a flask web interface. Table 1 combines the limitation of all the research.

Table 1. Summary of previous research works on kidney stone detection

Year	Author	Critique	Reference
2019	Abdullah <i>et al.</i>	The use relatively small dataset of 400 patients might affect the generalisability of the machine learning model to larger, more diverse population.	16
2020	Shi <i>et al.</i>	This paper struggles with highly variable or noisy ultrasound images that differs significantly form the training data, limiting its robustness in real-world clinical settings.	17
2018	V. B. Kolachalama <i>et al.</i>	This study didn't account for the impact of treatments such as immunosuppressive drugs, on renal survival. This could have influence outcome, as treatment may have improved survival beyond what the machine learning model predicted.	18
2022	M. Baygin <i>et al.</i>	The proposed ExDark19 model was only tested on specific datasets, which may limit its generalisability to diverse clinical senecios.	19
2022	A. F. Bayram <i>et al.</i>	The study only compares YOLOv7 and YOLOv7 Tiny on specific kidney-related conditions and no analysis was provided on the computational efficiency or real-time performance of these models in clinical settings.	20
2023	S. Rabby <i>et al.</i>	The dataset used in the study was relatively small, potentially limiting the generalizability of the results to larger and more diverse datasets.	21
2023	M. T. P. Nguyen <i>et al.</i>	The study focuses solely on KUB X-ray images, which may not generalize well to other imaging modalities like CT or ultrasound.	22
2024	B. Reuben <i>et al.</i>	A limitation of this study is the complexity of the proposed model, making it computationally intensive and less suitable for real-time clinical use.	23
2024	A. Ramadhani	A limitation of this study is the lack of validation on a larger and more diverse dataset, which could affect the generalizability of the model to different imaging conditions and patient demographics.	24

Previous research has demonstrated the effectiveness of deep learning methodologies in various medical imaging tasks. While integration of diverse techniques has shown promising results, there is still a gap in achieving higher accuracy for kidney stone detection. This research findings contribute to the ongoing progress in leveraging artificial intelligence to enhance clinical applications and improve diagnostic accuracy in nephrology.

3. YOLO

This section delves into the You Only Look Once (YOLO) model, a cutting-edge methodology for detecting real-time object with exceptional speed and accuracy. The rest of the section delves into the evolution of YOLO through its various iterations, from YOLOv1 to the most recent advancements.

3.1. You Only Look Once (YOLO)

Joseph Redmon *et al.* first proposed the YOLO framework in 2016 [25], revolutionizing object detection with a real-time end-to-end approach. Unlike previous methods that required either exhaustive sliding windows or a two-step process involving region proposals followed by classification, YOLO streamlined the task by accomplishing detection in a single network pass. This innovation drastically improved efficiency, as the network only "looks once" per image to predict objects. YOLO's architecture employs a straightforward regression-based output to predict bounding boxes and class probabilities simultaneously, contrasting with methods like Fast R-CNN [26], which use separate outputs for classification and bounding box regression. By integrating these predictions, YOLO achieves rapid and accurate detection, making it a valuable for real-time object detection.

3.2. Versions of YOLO

YOLOv1 [25] was launched in 2016, introducing the innovative single-pass detection framework. This was followed by YOLOv2 [27] in 2017, which improved accuracy and speed through batch normalization and high-resolution classifiers. YOLOv3 [28] brought multi-scale predictions and a deeper architecture with residual connections. YOLOv4 [29] further enhanced performance with Cross-Stage Partial connections and the Mish activation function. YOLOv5, YOLOv6 and YOLOv7 focused on real-time performance and architectural innovations [30-32]. YOLOv8 [33] introduced more complex models with increased layers, advanced data augmentation, and improved training strategies. Each version has progressively improved in terms of the number of classes they can classify, the complexity of the model architecture, and overall detection accuracy and speed. Recently release version YOLOv9 and YOLOv10 are also continuing this trend [34-35].

4. Methodology

This section describes the methodology used in this study. It delves into the details of the dataset used, architecture and properties of YOLOv8, the variants of YOLOv8. Following the information about the experiment setup and the

model training process has also been reported. Finally, delineate the evaluation metrics employed to compare the performance of trained models.

4.1. Dataset Description

The dataset utilized in the research is collected from Roboflow (it is a large repository of datasets prepared for training YOLO models). The dataset [36] comprises a total of 1428 medical CT-scan images, annotated with a total of 2951 bounding boxes, indicating the presence and location of kidney stones. Some samples images from the dataset is shows in Figure 1. All the images are resized to a uniform resolution of 640x640 pixels. On average, there are approximately 2 annotations per image. The annotations are distributed across two main classes, "Tas_Var" (stone) and "normal-kidney" (normal). Table 2 shows the distribution of annotation in each set.

Table 2. Annotations in each class

Class	Total Images	Train Images	Test Images	Valid Images
Tas_Var	2837	2281	281	265
Normal	124	99	14	11

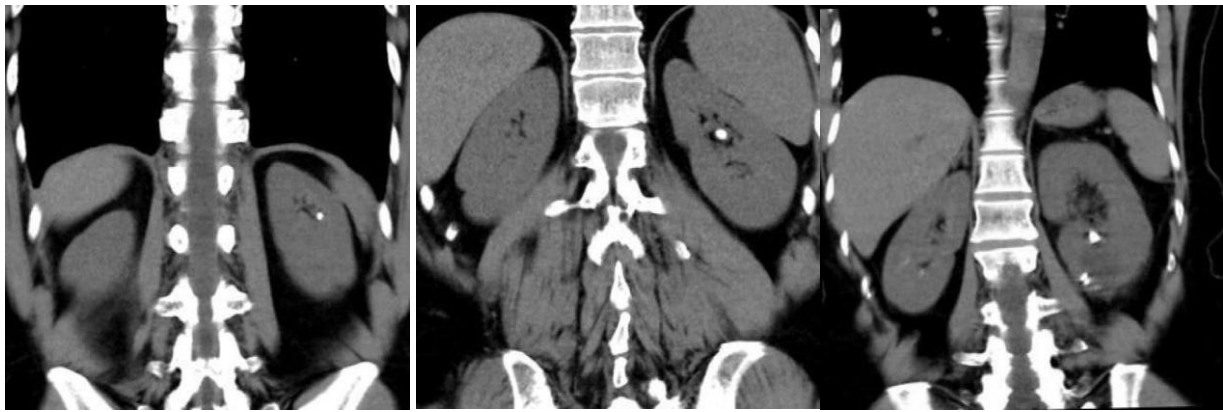


Fig.1. Sample CT-Scan images from the dataset [36]

4.2. YOLOv8

YOLOv8 [29], the latest version released by Ultralytics [37] in January 2023, boasts higher accuracy and faster speed, exemplified by the YOLOv8 (medium) model achieving a 50.2 mAP score at 1.83 milliseconds on the COCO dataset which has 80 classes, using A100 TensorRT. Although it lacks an official paper, YOLOv8 features a Python package and CLI-based implementation, enhancing ease of use and development. This versatile model can perform tasks such as detection, segmentation, classification, pose estimation, and oriented bounding boxes (OBB). The research focuses on detection using YOLOv8, which is equipped with advanced backbone and neck architectures for superior feature extraction and object detection performance. The model's anchor-free split Ultralytics head improves accuracy and efficiency, and its optimized accuracy-speed trade-off makes it ideal for real-time detection tasks. Additionally, YOLOv8 provides a variety of pre-trained models to suit different tasks and performance needs, simplifying the selection of an appropriate model for specific applications.

4.3. Versions of YOLOv8

YOLOv8 [33] is available in five variants, each differing in the number of parameters: nano (n), small (s), medium (m), large (l), and extra-large (x). All these variants are suitable for detection tasks. Details such as the number of parameters and FLOPs for each variant are provided in Table 3. This research aims to experiment with all 5 YOLOv8 variants to determine the most suitable model that balances computational efficiency and detection accuracy. By comparing these variants across multiple performance metrics such as precision, mAP (mean Average Precision) and resource consumptions, the study aims to identify the optimal YOLOv8 variant for kidney stone detection, ensuring a robust combination of precision and efficiency.

4.4. Experiment Setup and Model Training

The experiment was conducted using the Kaggle platform, leveraging a Tesla-P100 GPU with 16GB of VRAM to efficiently handle the computational demands of training YOLOv8 models. Study have used the Ultralytics 8.2.26 package to implement and train the models. Additionally, the research has utilized weights and biases tool for investigation of various evaluation and computational parameters.

Table 3. Comparison of version of YOLOv8

Model	Speed (ms)		Params (M)	FLOPs (B)
	<i>CPU ONNX</i>	<i>A100 TensorRT</i>		
nano	80.4	0.99	3.2	8.7
small	128.4	1.20	11.2	28.6
medium	234.7	1.83	25.9	78.9
large	375.2	2.39	43.7	165.2
extra- large	479.1	3.53	68.2	257.8

The model training process was carried out over 100 epochs, with a batch size of 16 to balance between training time and convergence stability. The research has utilized pre-trained weights for enhancing the initial performance and speeding up the convergence. The AdamW optimizer was employed, known for its ability to handle the complexities of training deep learning models. The learning rate was set to 0.001667, chosen to ensure gradual and stable convergence, while a momentum of 0.9 was used to accelerate the optimizer by dampening oscillations.

4.5. Evaluation Metrics

The models' accuracy, robustness, and generalization capabilities were assessed using standard evaluation metrics, encompassing precision, recall, f1-score, and mean Average Precision (mAP) at IoU thresholds of 0.5 (mAP@0.5). In addition to evaluation metrics, YOLOv8 variants also evaluated on the basis of GPU power usage in Watts, GPU memory allocated in bytes, Memory written on disk during training in mega-bytes, and total training time. These metrics facilitated a comprehensive evaluation of the models' detection performance.

5. Result and Discussion

This section presents the results of this study on kidney stone detection using various YOLOv8 models. The study provides a detailed evaluation of the models' performance through standard evaluation metrics, interpret the models' behaviour through graphical analysis, and identify the most computationally efficient model.

5.1. Performance Evaluation Using Evaluation Metrics

The comprehensive evaluation of the performance of each YOLOv8 variant model is presented in Table 4. It offers valuable insights into their capabilities in detecting kidney stones.

Table 4. Result of YOLOv8 versions

Model	Class	Precision	Recall	F1-score	mAP@50
nano	All	0.896	0.861	0.878	0.897
	Stone	0.825	0.722	0.770	0.799
	Normal	0.967	1.000	0.983	0.995
small	All	0.917	0.889	0.903	0.918
	Stone	0.883	0.778	0.827	0.840
	Normal	0.952	1.000	0.975	0.995
medium	All	0.910	0.875	0.892	0.902
	Stone	0.870	0.751	0.806	0.808
	Normal	0.950	1.000	0.974	0.995
large	All	0.858	0.894	0.876	0.902
	Stone	0.841	0.790	0.815	0.817
	Normal	0.875	0.997	0.932	0.986
extra-large	All	0.917	0.85	0.915	0.904
	Stone	0.880	0.730	0.798	0.813
	Normal	0.954	1.000	0.976	0.995

A notable observation from the table is that both YOLOv8l and YOLOv8n exhibit performance below the threshold of 0.9 across all evaluated metrics (except mAP@50 in case of YOLOv8l), indicating areas for improvement in their detection capabilities. In contrast, YOLOv8m showcases commendable performance, achieving a precision of 0.910 and a mAP@50 of 0.902, demonstrating its effectiveness in detecting kidney stones.

However, among the models, YOLOv8e emerges as a strong competitor, closely rivalling YOLOv8s. Both models exhibit the highest precision of 0.917. While YOLOv8s achieves the highest mAP@50 of 0.918, YOLOv8e achieves

the highest F1-score of 0.915. Notably, both models fall short in recall compared to YOLOv8l, with YOLOv8s slightly outperforming YOLOv8e with a recall value of 0.889.

These findings suggest that both YOLOv8s and YOLOv8e are equally efficient in kidney stone detection. However, to determine the best model for deployment, other factors such as computational requirements and real-world applicability need to be carefully considered.

5.2. Model Interpretation Through Graphical Analysis

This subsection provides a graphical interpretation of the models' performance to gain deeper insights into their detection capabilities. Figures 2, 3, 4 and 5 illustrate the performance evaluation curves. These graphs are essential for understanding the behaviour of the YOLOv8 models across different confidence thresholds and metrics.

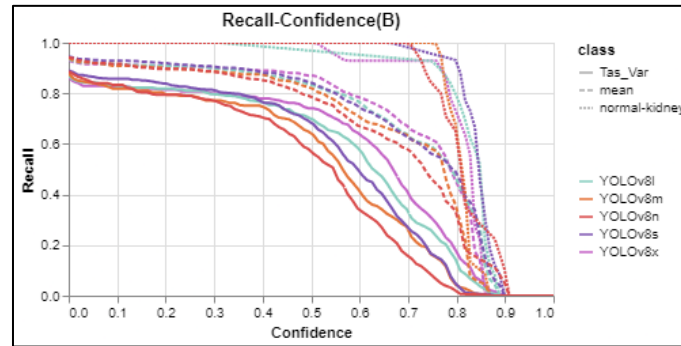


Fig.2. Recall confidence curve

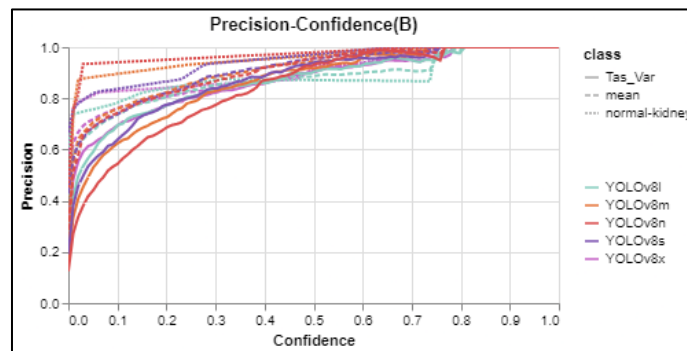


Fig.3. Precision confidence curve

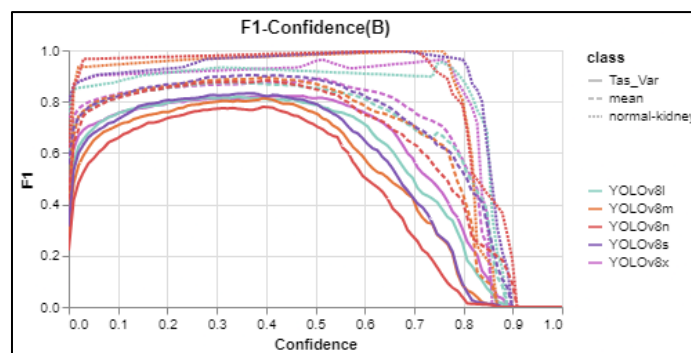


Fig.4. F1 confidence curve

Figure 2 (Recall-Confidence curve) illustrates the model's ability to identify all positive instances, changes as the confidence threshold increases. Higher recall at lower confidence thresholds suggests the model captures more true positives but may also include false positives. In Figure 3, the curve for YOLOv8s exhibits a more optimal shape, closely approaching the ideal curve, indicating superior performance compared to the other models. Figure 2 (Precision-Confidence curve) demonstrates the model's accuracy in identifying only relevant instances, is affected by increasing the confidence threshold. A higher precision at higher confidence thresholds means the model is more selective, reducing false positives while potentially missing some true positives. YOLOv8n seems to be better performing in the graph, approaching to ideal Precision-Confidence curve. Figure 4 (F1-Confidence curve) combines precision and recall into a single metric (F1-score), balancing the trade-off between the two. This curve helps assess the overall performance

of the models as the confidence threshold varies, offering a clearer picture of where the balance between precision and recall lies. In the curve, it can be seen that both YOLOv8s and YOLOv8l are performing better than other models showcased its efficiency in maintaining the trade-off between precision and recall. Figure 5 (Precision-Recall curve), on the other hand, provides insight into the interplay between precision and recall across all confidence thresholds, highlighting the model's ability to maintain high precision while maximizing recall. For the normal class all the models showcased almost near the ideal curve but when looking the other two curves (mean and Tas_Var), it is evident that YOLOv8s are performing better than other models.

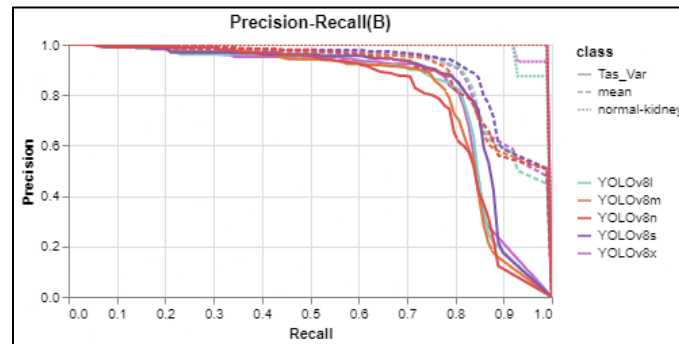


Fig.5. P-R curve

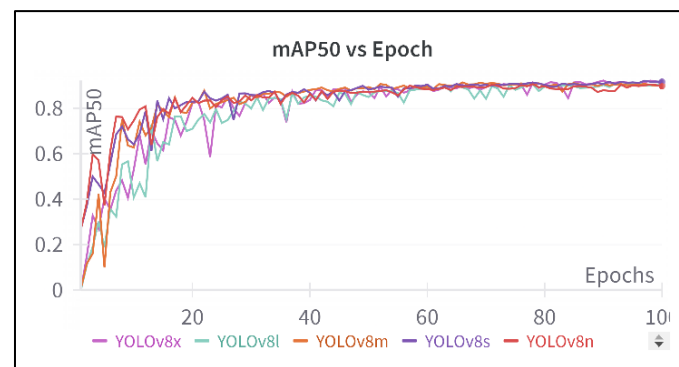


Fig.6. mAP50 vs epoch curve

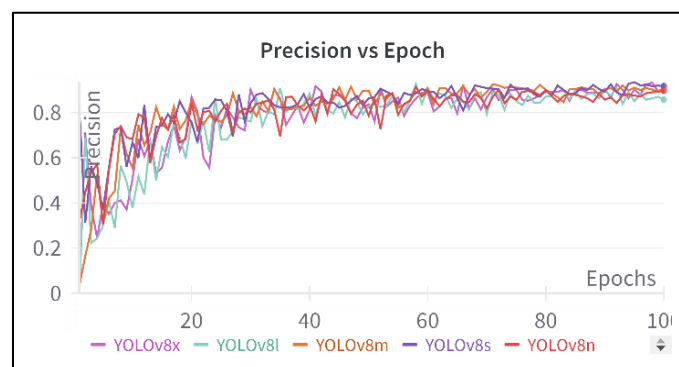


Fig.7. Precision vs epoch curve

Figures 6, 7, and 8 present the progression of mAP50, precision, and recall values across epochs during the training of the models, respectively, offering critical insights into the models' learning behaviour over time. Figure 6 (mAP50 vs Epoch curve) shows the mAP at an IoU threshold of 50% as it evolves throughout the training process. This curve provides a summary of overall model performance in terms of detection accuracy, indicating how well the models learn to balance precision and recall. Examining the curves, it can be seen that the YOLOv8s curve ends slightly higher than other model's curve. Figure 7 (Precision vs Epoch curve) tracks the change in precision across epochs, indicating how the models improve in terms of reducing false positives as training progresses. Precision helps in understanding the ability of the model to make correct detections without sacrificing specificity. YOLOv8s stands out here by achieving the highest precision by the end of the training, indicating that it consistently makes fewer false-positive detections. Figure 8 (Recall vs Epoch curve) provides a view of how recall improves with each epoch, highlighting the model's ability to detect true positives over time. A higher recall throughout training reflects the model's strength in capturing

relevant detections, although it may also result in more false positives at certain points.

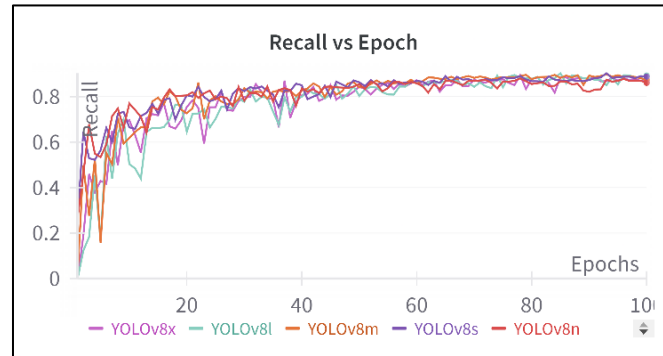


Fig.8. Recall vs epoch curve

Interestingly, despite initial differences, all models show convergence in performance by the end of training, reaching similar mAP, precision, and recall values. However, the Precision vs Epoch curve reveals that YOLOv8s achieves the highest precision, surpassing the other models, while YOLOv8l performs slightly lower in terms of precision compared to the other variants. These observations provide valuable insights into how the models progress and stabilize during training, helping guide model selection and optimization strategies based on the desired trade-offs between precision, recall, and overall accuracy.

5.3. Identification of Computationally Efficient Model

Table 5 encompasses the total and per-epoch training time for each model. It is evident that the extra-large model required significantly more time to train compared to the nano model, approximately seven times longer. This discrepancy is expected due to the extra-large model's complex architecture and exponentially larger number of parameters.

Table 5. Time taken by models to train for 100 epochs

Model	Total Time (min)	Time/Epoch (minute: second)
nano	69	0:42
small	110	1:50
medium	222	2:14
large	333	3:20
extra-large	480	4:48

Evaluation of computationally efficient algorithms must not solely be based on the training time; it can be misleading as other external factors can be influential. Factors such as memory usage and power consumption also play critical roles in the real-world scenario. Therefore, a comprehensive analysis considering multiple factors is essential for making informed decisions regarding model selection and deployment strategies.

Figure 9 illustrates the power consumption of GPUs during training across YOLOv8 models. YOLOv8n notably consumes the least power at an average of 120W, while YOLOv8x exhibits the highest consumption at 190W due to its extensive architecture. YOLOv8s stands out with commendable performance, averaging 150W during training, showcasing efficiency. YOLOv8l and YOLOv8m demonstrate slightly lower power usage compared to YOLOv8x, around 180W. This analysis emphasizes the importance of considering power consumption alongside other metrics for model selection and deployment.

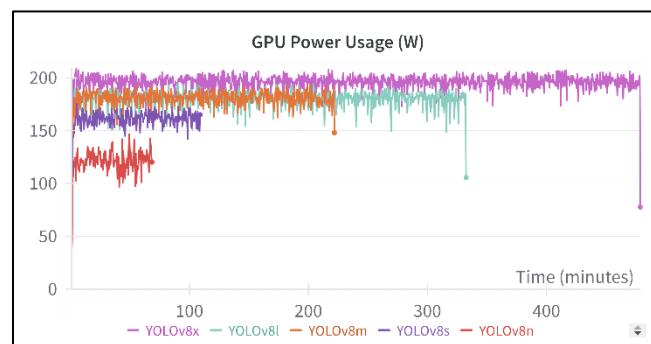


Fig.9. GPU power usage in watt throughout the training

Figures 10 and 11 illustrate the GPU and disk memory allocation throughout the training process. These parameters are crucial as they define the computational expense of the models and the necessary memory requirements for running these models. The graphical representation clearly shows that YOLOv8x is the most computationally expensive, requiring 13.6 billion bytes of GPU memory. In contrast, YOLOv8n is much less demanding, needing only 2.86 billion bytes of GPU memory.

Regarding disk memory requirements, it is directly proportional to the training duration. YOLOv8x, being the most resource-intensive, requires almost 100,000 megabytes, while YOLOv8n needs about 4,500 megabytes—22 times less. The other models, such as YOLOv8s, YOLOv8m, and YOLOv8l, have memory usage that falls between the extremes of YOLOv8n and YOLOv8x. This indicates a balanced computational expense and memory requirement for these models, making them potentially more suitable for practical applications where resources are limited.

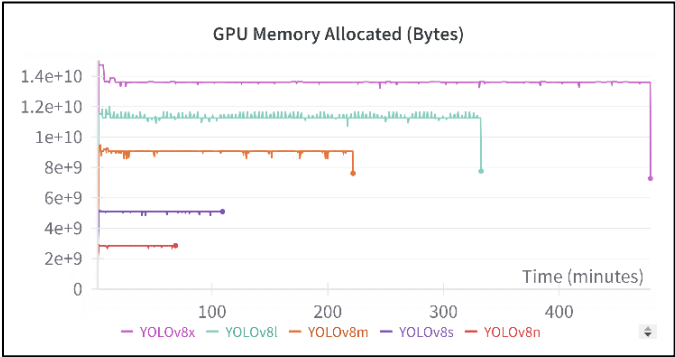


Fig.10. GPU memory allocated in bytes throughout training

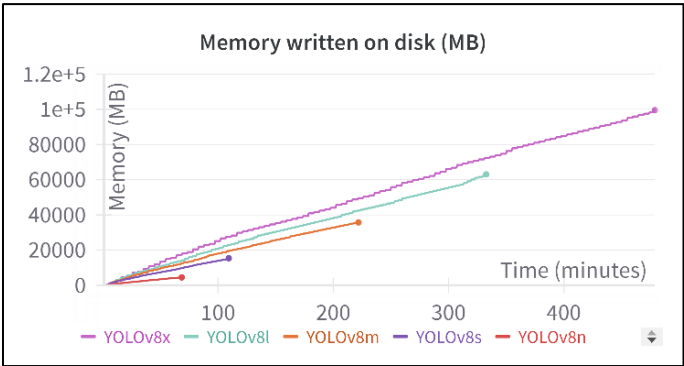


Fig.11. Memory written on disk throughout the training

These findings reveal that each model exhibits distinct characteristics: some excel in evaluation metrics, while others are more computationally efficient, requiring significantly less memory for training. This comprehensive analysis enables us to draw a nuanced conclusion, identifying the optimal model that strikes a balance between computational efficiency and robust evaluation metrics. The synthesis of these insights will guide the selection of a model that is not only resource-efficient but also demonstrates superior performance in detecting kidney stones.

Table 6. Comparison with previous researched

Reference	Authors	Employed Methodology	Evaluation Metrics
[38]	M. Gulane <i>et al.</i> (2024)	Improved Deep Neural Network	Accuracy = 90% Recall = 90% F1-Score = 89.5% Precision = 89%
[39]	Tsitsiflis <i>et al.</i> (2022)	ANN	Accuracy = 81.43%
[40]	D. C. Elotn <i>et al.</i> (2022)	Deep Learning	Sensitivity = 86%
[41]	S. Rabby <i>et al.</i> (2023)	YOLO	IOU = 91.4%
[42]	A. F. Bayram <i>et al.</i> (2022)	YOLOv7	mAP@50 = 85% precision = 88.2% sensitivity = 82.9% F1 score = 85.4%
Proposed		YOLOv8s	Precision = 91.7% Recall = 88.9% F1-Score = 90.3% mAP@50 = 91.8%

5.4. Comparison with Previous Research on Kidney Stone Detection

This section presents a comparative analysis of the proposed methodology against previous works. It highlights how the current study addresses the limitations and weaknesses of earlier approaches, as well as the challenges associated with real-world kidney detection. Table 6 presents the comprehensive table of existing research on kidney stone detection. The table encompasses the methodology employed by the research and their presented results.

6. Conclusions

The study addresses the critical balance between computational efficiency and detection accuracy in kidney stone detection. Rather than solely pursuing the most accurate model or the least computationally expensive one, the research's aim is to identify a model that strikes a balance between these two aspects. Throughout the investigation, the study can conclude the following:

- The research observed that YOLOv8s and YOLOv8x exhibit outstanding performance across evaluation metrics.
- YOLOv8x is significantly more computationally expensive than YOLOv8s in terms of disk space, GPU utilization, and power usage.
- Study recommend YOLOv8s as the optimal model, effectively balancing computational efficiency and performance evaluation.
- The experiment showcased that YOLOv8 achieved the highest metrics values of precision (0.917) and mAP@50 (0.918).
- Moreover, YOLOv8s demonstrates efficient resource utilization, requiring only 150W for training, 5 billion bytes of GPU memory, and 150,000 megabytes of disk space.
- This research highlights the importance of considering both computational efficiency and performance metrics when selecting a model for kidney stone detection.

These characteristics position YOLOv8s as the top performer among all YOLOv8 variants. This research highlights the importance of considering both computational efficiency and performance metrics when selecting a model for kidney stone detection. YOLOv8s emerges as the optimal choice, offering exceptional accuracy while minimizing computational resource requirements.

7. Future Scope

The future scope of this research includes investigating other YOLO models to compare their performance and effectiveness in detecting kidney stones. Additionally, testing the trained models on real-world datasets will be crucial to validate their robustness and applicability in practical medical settings as a CAD tool. Furthermore, validating the trained models on real-world datasets will be essential to ensure their robustness and practical applicability in medical settings. Expanding the focus to include the detection of other urological conditions could further enhance the models' clinical utility and improve overall patient outcomes.

References

- [1] T. Alelign and B. Petros, "Kidney Stone Disease: An update on current concepts," *Advances in Urology*, vol. 2018, pp. 1–12, 2018.
- [2] S. R. Khan, M. S. Pearle, W. G. Robertson, G. Gambaro, B. K. Canales, S. Doizi, O. Traxer, and H.-G. Tiselius, "Kidney stones," *Nature Reviews Disease Primers*, vol. 2, no. 1, Feb. 2016.
- [3] Z. Wang, Y. Zhang, J. Zhang, Q. Deng, and H. Liang, "Recent advances on the mechanisms of Kidney Stone Formation (review)," *International Journal of Molecular Medicine*, vol. 48, no. 2, Jun. 2021.
- [4] B. Aldughayfiq, F. Ashfaq, N. Z. Jhanjhi, and M. Humayun, "YOLO-Based Deep Learning Model for pressure Ulcer detection and Classification," *Healthcare*, vol. 11, no. 9, p. 1222, Apr. 2023, doi: 10.3390/healthcare11091222.
- [5] M. Mousavi, F. S. Mirshafiee, E. Shahbazi, M. Safi, and A. Mosavi, "YOLO for Lung Disease Detection from CT Scans. 2023. doi: 10.1109/sisy60376.2023.10417904.
- [6] F. Prinzi, M. Insalaco, A. Orlando, S. Gaglio, and S. Vitabile, "A Yolo-Based model for breast cancer detection in mammograms," *Cognitive Computation*, vol. 16, no. 1, pp. 107–120, Aug. 2023, doi: 10.1007/s12559-023-10189-6.
- [7] N. Aishwarya, K. M. Prabhakaran, F. T. Debebe, M. S. S. A. Reddy, and P. Pranavee, "Skin Cancer diagnosis with Yolo Deep Neural Network," *Procedia Computer Science*, vol. 220, pp. 651–658, Jan. 2023, doi: 10.1016/j.procs.2023.03.083.
- [8] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "NNU-net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Dec. 2020.
- [9] Z. Jiang, H. Zhang, Y. Wang, and S.-B. Ko, "Retinal blood vessel segmentation using fully convolutional network with transfer learning," *Computerized Medical Imaging and Graphics*, vol. 68, pp. 1–15, Sep. 2018, doi: <https://doi.org/10.1016/j.compmedimag.2018.04.005>.
- [10] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in Deep Learning for Object Detection," *Neurocomputing*, vol. 396, pp. 39–64, Jul. 2020.

- [11] A. Haghanifar, M. M. Majdabadi, Y. Choi, S. Deivalakshmi, and S. Ko, "COVID-CXNet: Detecting COVID-19 in frontal chest X-ray images using deep learning," *Multimedia Tools and Applications*, Apr. 2022, doi: <https://doi.org/10.1007/s11042-022-12156-z>.
- [12] J. Zhang, Y. Xie, Q. Wu, and Y. Xia, "Medical Image Classification using synergic Deep learning," *Medical Image Analysis*, vol. 54, pp. 10–19, May 2019.
- [13] Y. Wang, E. J. Choi, Y. Choi, H. Zhang, G. Y. Jin, and S.-B. Ko, "Breast Cancer Classification in Automated Breast Ultrasound Using Multiview Convolutional Neural Network with Transfer Learning," *Ultrasound in Medicine & Biology*, vol. 46, no. 5, pp. 1119–1132, May 2020, doi: <https://doi.org/10.1016/j.ultrasmedbio.2020.01.001>.
- [14] A. Satsangi, K. Srinivas, and A. Charan Kumari, "Investigating transfer learning models for lung cancer detection from CT scans: A comparative evaluation," *Artificial Intelligence and Sustainable Computing*, pp. 231–242, 2024.
- [15] S. Jain, K. Srinivas, and A. C. Kumari, "Inspecting transfer learning models for diabetic retinopathy detection," 2023 3rd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET), Dec. 2023.
- [16] N. A. Almansour, H. F. Syed, N. R. Khayat, R. K. Altheeb, R. E. Juri, J. Alhiyafi, S. Alrashed, and S. O. Olatunji, "Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study," *Computers in Biology and Medicine*, vol. 109, pp. 101–111, Jun. 2019.
- [17] S. Yin, Q. Peng, H. Li, Z. Zhang, X. You, K. Fischer, S. L. Furth, G. E. Tasian, and Y. Fan, "Automatic Kidney Segmentation in ultrasound images using subsequent boundary distance regression and Pixelwise Classification Networks," *Medical Image Analysis*, vol. 60, p. 101602, Feb. 2020.
- [18] V. B. Kolachalama, P. Singh, C. Q. Lin, D. Mun, M. E. Belghasem, J. M. Henderson, J. M. Francis, D. J. Salant, and V. C. Chitalia, "Association of pathological fibrosis with renal survival using Deep Neural Networks," *Kidney International Reports*, vol. 3, no. 2, pp. 464–475, Mar. 2018.
- [19] M. Baygin, O. Yaman, P. D. Barua, S. Dogan, T. Tuncer, and U. R. Acharya, "Exemplar DARKNET19 feature generation technique for automated kidney stone detection with coronal CT images," *Artificial Intelligence in Medicine*, vol. 127, p. 102274, May 2022.
- [20] A. F. Bayram, C. Gurkan, A. Budak, and H. Karatas, "A Detection and Prediction Model Based on Deep Learning Assisted by Explainable Artificial Intelligence for Kidney Diseases," *European Journal of Science and Technology*, Sep. 2022.
- [21] S. Rabby, F. Hossain, S. Das, I. Rahman, S. Das, J. Soeb, and Md. F. Jubaye, "An automated approach for the kidney segmentation and detection of kidney stones on computed tomography using YOLO algorithms," *Journal of Ideas in Health*, vol. 6, no. 4, pp. 963–970, Nov. 2023.
- [22] M. Tai, Viet Tuan Le, Huu Thanh Duong, and Vinh Truong Hoang, "Detection of Kidney Stone Based on Super Resolution Techniques and YOLOv7 Under Limited Training Samples," *Lecture notes on data engineering and communications technologies*, pp. 23–33, Jan. 2023, doi: https://doi.org/10.1007/978-3-031-46749-3_3.
- [23] Reuben, B., & Narmadha, C., "Effective Kidney Stone Prediction Based on Optimized Yolov7 Segmentation and Deep Learning Classification," *International Journal of Intelligent Systems and Applications in Engineering*, 12(1), 183–192 (2023).
- [24] Adnin Ramadhani, & Abu Salam. (2024). Deployment of Web-Based YOLO for CT Scan Kidney Stone Detection. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 8(3), 1357-1368. <https://doi.org/10.33395/sinkron.v8i3.13744>
- [25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016.
- [26] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Dec. 2015.
- [27] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017.
- [28] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv (Cornell University)*, Jan. 2018, doi: 10.48550/arxiv.1804.02767. Available: <https://arxiv.org/abs/1804.02767>
- [29] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv (Cornell University)*, Jan. 2020, doi: 10.48550/arxiv.2004.10934. Available: <https://arxiv.org/abs/2004.10934>
- [30] G. Jocher, "YOLOv5 by Ultralytics." <https://github.com/ultralytics/yolov5>, 2020. Accessed: March 1, 2024.
- [31] C. Li et al., "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications," *arXiv (Cornell University)*, Jan. 2022, doi: 10.48550/arxiv.2209.02976. Available: <https://arxiv.org/abs/2209.02976>.
- [32] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2023.
- [33] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics." <https://github.com/ultralytics/ultralytics>, 2023. Accessed: March 1, 2024.
- [34] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," *arXiv (Cornell University)*, Feb. 2024, doi: 10.48550/arxiv.2402.13616. Available: <https://arxiv.org/abs/2402.13616>
- [35] A. Wang et al., "YOLOv10: Real-Time End-to-End Object Detection," *arXiv (Cornell University)*, May 2024, doi: 10.48550/arxiv.2405.14458. Available: <https://arxiv.org/abs/2405.14458>.
- [36] "Kidney Stone Detection Image Dataset" <https://universe.roboflow.com/selam-h8tid/kidney-stone-detection-fwubk/dataset/1>. Accessed: January 12, 2024.
- [37] "ultralytics" <https://www.ultralytics.com/>. Accessed: January 12, 2024.
- [38] M. Gulhane et al., "Integrative approach for efficient detection of kidney stones based on improved deep neural network architecture," *SLAS TECHNOLOGY*, vol. 29, no. 4, p. 100159, Jun. 2024, doi: 10.1016/j.slant.2024.100159.
- [39] A. Tsitsiflis et al., "The use of an artificial neural network in the evaluation of the extracorporeal shockwave lithotripsy as a treatment of choice for urinary lithiasis," *Asian Journal of Urology*, vol. 9, no. 2, pp. 132–138, Sep. 2021, doi: 10.1016/j.ajur.2021.09.005.
- [40] D. C. Elton, E. B. Turkbey, P. J. Pickhardt, and R. M. Summers, "A deep learning system for automated kidney stone detection and volumetric segmentation on noncontrast CT scans," *Medical Physics*, vol. 49, no. 4, pp. 2545–2554, Feb. 2022, doi: 10.1002/mp.15518.

- [41] S. Rabby et al., "An automated approach for the kidney segmentation and detection of kidney stones on computed tomography using YOLO algorithms," *Journal of Ideas in Health*, vol. 6, no. 4, pp. 963–970, Nov. 2023, doi: 10.47108/jidhealth.vol6.iss4.313.
- [42] A. F. Bayram, C. Gurkan, A. Budak, and H. Karataş, "A Detection and Prediction Model Based on Deep Learning Assisted by Explainable Artificial Intelligence for Kidney Diseases," *European Journal of Science and Technology*, Sep. 2022, doi: 10.31590/ejosat.1171777.

Authors' Profiles



Amol Satsangi is pursuing a Bachelor of Technology in Electrical Engineering at the Faculty of Engineering, Dayalbagh Educational Institute, Agra, India. He has been actively engaged in research in the field of computer vision algorithms and has undertaken research at the University of Saskatchewan, Canada, focusing on health informatics. His research interests encompass pattern recognition, image processing, and deep learning methodologies.



Shaurya Jain is currently pursuing bachelor's degree in electrical engineering with a specialization in Computer Science from Dayalbagh Educational Institute, Agra, India. She is currently engaged in research as a student, focusing on the fields of deep learning and health informatics. Her research interests encompass a broad range of topics, including computer vision, machine learning and deep learning.



Dr. Subho Upadhyay is an Assistant Professor in the Department of Electrical Engineering at Dayalbagh Educational Institute, Agra, Uttar Pradesh, India. He received his ME degree in Electrical and Electronics Engineering from the Birla Institute of Technology, Mesra, India in 2012 and his Ph.D. from the Indian Institute of Technology, Roorkee, India in 2017. His research interests include Integrated Renewable Energy Systems, Power Quality, Power Electronics Applications in Renewable Energy and Electric Vehicles, Artificial Intelligence techniques, and Optimization. He has edited two books "Renewable Energy Systems: Modeling, Optimization and Applications" by Scrivener publication and "Artificial Intelligence and Machine Learning Applications for

Sustainable Development, CRC Press, Taylor and Francis. He has published more than 30 research papers in various journals and conferences. He is an active reviewer and a member of multiple international journals and conferences including Elsevier, Springer, and IET. He is also working as an NSS Program Officer in his institute.

How to cite this paper: Amol Satsangi, Shaurya Jain, Subho Upadhyay, "Optimizing Kidney Stone Detection: Exploring YOLOv8 Variants for Computational Efficiency and Enhanced Accuracy", *International Journal of Intelligent Systems and Applications(IJISA)*, Vol.17, No.5, pp.1-12, 2025. DOI:10.5815/ijisa.2025.05.01