Enhanced Kidney Tumor Detection using hybrid Deep MSFPT Transformer in computed Tomography images

Sachin Dattatraya Shingade[1], Midhun Chakkaravarthy[2], Dimitrios A. Karras[3] , Sachin S[4], Komal M Masal[5]
[1,2]LUCM PetaLing Jaya Malaysia; [3] LUCM, NKUA Athens Greece and EUT Albania; [4] PICT Pune, IITP India, [5] PICT , DOT SPPU Pune India
[1]pdf.sachin@lincoln.edu.my ; [2] midhun@lincoln.edu.my; [3] dimitrios.karras@gmail.com; [4]sachin_pa2503mth305@iitp.ac.in; [5]kmmasal@pict.edu

Abstract: Kidney tumors (KTs) are a significant global health concern and are ranked among the leading cancers worldwide. Early detection is crucial, as it dramatically improves treatment outcomes, prevents disease progression, and lowers mortality rates. However, many current methods struggle with precisely identifying and localizing tumors within the kidney. This research addresses these challenges by proposing a novel, lightweight detection framework tailored for medical imaging. To handle the limitations of a small, imbalanced dataset, we employ data augmentation techniques to expand and normalize the data. We introduce a MSFPT Transformer to strengthen kidney tumor detection by capturing richer spatial details. To further boost accuracy, a lightweight attention module is integrated to precisely evaluate the post-neck layer features, leading to more reliable identification of affected regions. Our approach not only enhances diagnostic accuracy but also maintains low computational demands, making it suitable for real-time clinical environments. Experimental results demonstrate that our model outperforms existing methods, validating the effectiveness of transformer-based architectures for medical imaging and early cancer diagnosis. The model achieved excellent performance with a 98.20% IoU, 97.20 % precision, 96.90 % recall, 97.10 % F1-score, 97.40 % mAP, ..

Keywords: Kidney Tumor (KT), MSFPT Transformer , lightweight attention module, post-neck layer, tumor localization.

Introduction

Kidney tumors (KTs) are a significant global health threat, making their early detection paramount for improving patient outcomes and reducing mortality, as traditional diagnostic methods are often tedious and labor-intensive [1]. Recent advancements in medical imaging and computational modeling, particularly with deep learning (DL), offer a more efficient and accurate alternative [2]. A wide range of machine learning (ML) and DL architectures have been applied to analyze radiological images, such as CT, MRI, and ultrasound, for the automated detection and classification of kidney abnormalities [3-6]. Various frameworks, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown strong potential for this task [7, 8]. More advanced models, like the adaptive and attentive residual DenseNet with gated recurrent units (AA-RD-GRU) [9] and hybrid approaches combining segmentation and classification [10], have been developed to differentiate between tumor types and improve diagnostic accuracy. The performance of these systems is often evaluated using large-scale datasets, ensuring their robustness and accuracy [11,12]. The continuous evolution of these automated systems is critical for providing reliable diagnostic support in clinical environments.

The primary contributions of this study includes , the development of a new deep learning framework for kidney tumor (KT) detection. This framework is designed not only to identify the presence of a tumor but also to pinpoint its exact location and size, providing crucial information for effective treatment planning. Incorporated a Multi-Scale Feature Pyramid Transformer (MSFPT) to significantly improve the representation of spatial features. This architectural enhancement leads to a more accurate and robust detection of kidney tumors, even in complex cases. To further enhance detection precision, integrated a lightweight attention module. This module efficiently analyzes the outputs of the post-neck layer, allowing the model to focus on the most critical features within tumor-affected regions and filter out irrelevant noise.

Related work

The accurate and early identification of kidney tumors (KT) from CT images remains a significant challenge due to the subtle and complex nature of the features involved. Sachin et al.[16] Expalined various techniques used for Kidney tumour detection. Shingade et al.[17] illustrates methods often struggle with data variability and may lack the precision required for clinical decision-making. Kadhim et al. [18] explored the use of conventional machine learning models such as SVM and MLP, employing texture-based feature extraction and noise filtering to detect KT; however, these models demonstrated limited robustness. Another study by Taha et al. [19] utilized a large dataset to evaluate multiple deep learning architectures, identifying DenseNet-201 as the most effective model for achieving high accuracy in cancer detection, though the work did not introduce a novel methodology. Furthermore, a hybrid approach combining machine learning, deep learning, and fuzzy logic was proposed to enhance CT image contrast and leverage pre-trained networks like DenseNet121 and ResNet101 for comprehensive feature extraction [20].

In the context of total kidney volume (TKV) measurement, which traditionally relies on time-consuming manual segmentation, Sheng et al. [21] introduced a Single Shot Detector (SSD) model to automate the process, incorporating image preprocessing techniques to improve detection on both contrast and non-contrast CT scans. Concurrently, research has focused on the application of transfer learning to address data scarcity, although a key challenge lies in the assumption of consistent feature distances between source and target domains, a limitation highlighted in a comprehensive review by Habchi et al. [22]. While these prior works have contributed to the field, many traditional methods still face limitations: they often require laborious feature engineering, lack generalizability, or are computationally intensive, hindering real-time analysis. The proposed method is designed to overcome these specific issues, offering a solution that prioritizes both high accuracy and computational efficiency.

Motivation

Kidney tumors (KTs) pose a significant global health threat, making their early detection paramount for improving patient prognosis and survival rates. Unfortunately, the disease's often asymptomatic early stages frequently lead to late-stage diagnosis after metastasis has occurred. This necessitates the development of accurate and efficient diagnostic tools, such as automated CT scan analysis, to facilitate timely clinical intervention and effective treatment planning.

Proposed Methodology

Given the significant global health threat posed by kidney tumors, their timely and precise detection is crucial for improving patient outcomes. Our framework addresses the computational challenges associated with this task by introducing a lightweight deep learning approach designed for high accuracy and efficiency. The methodology begins with data augmentation. This is a crucial technique in deep learning that artificially expands the training set by applying systematic transformations to existing images. By creating various versions of the same image, the model's robustness is enhanced, and its tendency to overfit is significantly reduced. Specifically, this work employs horizontal and vertical flips. A horizontal flip mirrors an image along its vertical axis, while a vertical flip mirrors it along the horizontal axis, exposing the model to different orientations of the same tumor and thereby improving its generalization capability during training.

The core of our system is a MSFPT Transformer which leverages a backbone network for initial feature extraction. This architecture creates a feature pyramid that is highly effective at recognizing tumors of varying sizes and locations. To further enhance precision, a compact attention module is integrated into the post-neck layer outputs, enabling the model to concentrate on and emphasize the most critical features within tumor-affected regions. The refined features are then fused using an Inject-and-Fusion mechanism to create a more comprehensive representation. Finally, these consolidated features are passed to an IFE-Head, which produces the final detection output, clearly outlining the tumor's location and extent for effective clinical analysis and treatment planning.

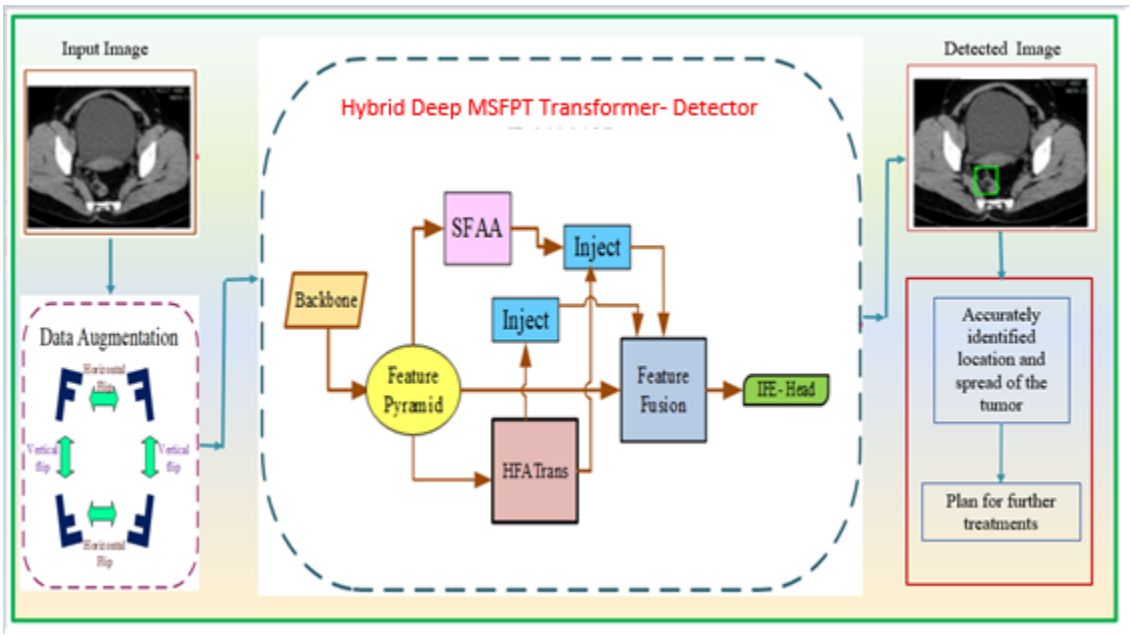


Figure 1. Architecture of Enhanced Kidney Tumor Detection using Hybrid Deep MSFPT Transformer

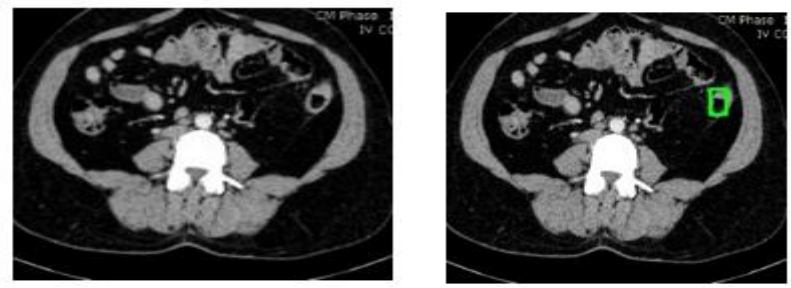
In order to overcome the constraint of a limited dataset of kidney tumor (KT) images, this study utilizes data augmentation. This is a crucial technique in deep learning that artificially expands the training set by

applying systematic transformations to existing images. By creating various versions of the same image, the model's robustness is enhanced, and its tendency to overfit is significantly reduced. Specifically, this work employs horizontal and vertical flips. A horizontal flip mirrors an image along its vertical axis, while a vertical flip mirrors it along the horizontal axis, exposing the model to different orientations of the same tumor and thereby improving its generalization capability during training.The methodology refines feature maps through a multi-stage pipeline, beginning with a module that extracts and groups channel information to optimize computation. An attention-based transformer then enhances spatial details and captures long-range dependencies. Finally, the enriched features are fused in a top-down network for efficient and accurate tumor detection.

The Multi-Level Feature Channel Extractor (ML-FCE), a central component of the SFAA module, enhances feature representation by using group convolution to learn multiscale spatial information. It models global context with a minimal parameter count and uses an attention mechanism with normalization to highlight important features. This integration of spatial and channel-wise extraction significantly boosts performance for tasks like object detection.The HFATrans module integrates convolutional and transformer architectures to enhance global and local feature extraction for object detection. It employs Multi-Scale Dilated Attention to capture multi-scale semantic information and Multi-Head Attention to model long-range pixel dependencies. This design provides a balanced approach to accuracy, computational efficiency, and effective feature interaction. The IFE-Head is a key component designed to enhance the model's ability to extract and interact with global feature information. It employs a two-pronged approach, using a feature interaction extractor to enrich multi-level feature maps from the FPN and a feature extractor to dynamically generate task-specific features. This structure ensures comprehensive feature analysis, leading to more accurate and robust localization results.

Simulation results

The proposed method, built with Python, was evaluated against existing models like RetinaNet and Faster R-CNN on an augmented kidney tumor dataset of 4,820 images. The performance was assessed using metrics such as IoU, mAP, and F1-Score to ensure a fair comparison. The model was trained with the Adam optimizer for 300 epochs, using a 0.01 learning rate and a batch size of 32.



(a) (b)

Figure 2. Simulation results (a) Input and (b) Detected output for kidney tumor detection

Performance analysis

This section presents a comparative analysis of proposed kidney tumor (KT) detection framework against existing methods, with performance measured across several key metrics. ML and Deep learning algorithms are Evaluated on key metrics like recall, precision, Mean Average Precision (mAP), training and testing loss curves, Intersection over Union (IoU), F1-Score and time complexity [13,14].This metrics are crucial in model performance [15]. This comprehensive evaluation provides a robust understanding of approach's effectiveness relative to state-of-the-art solutions.

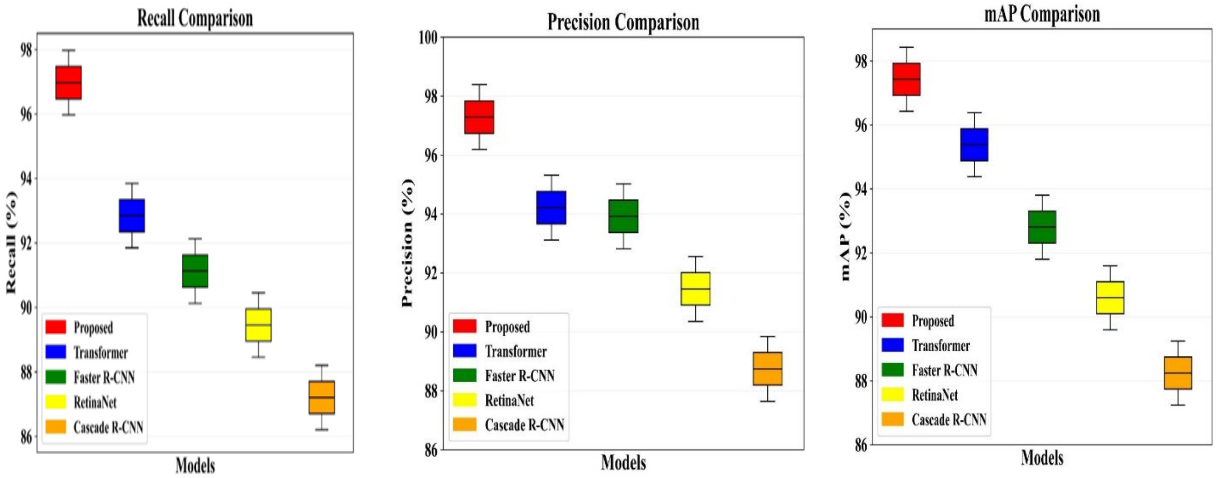
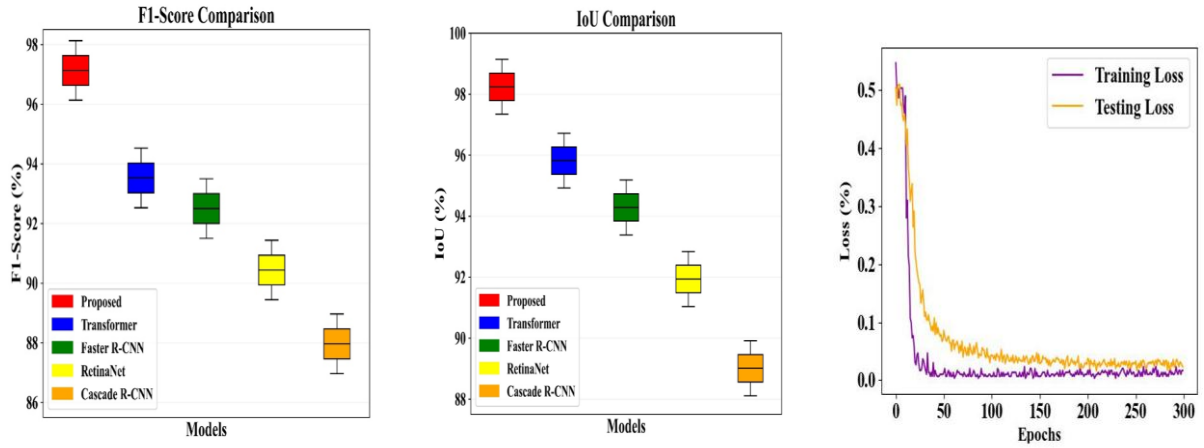


Figure 3. comparative analysis of results (a) Recall, (b) Precision and (c) mAP

The fig. 3 (a) provides a comparative analysis of the recall performance of several object detection models. The proposed model consistently exhibits the highest recall, achieving between 96.90% with low variability, which indicates its superior and stable ability to correctly identify a high proportion of kidney tumors. This performance significantly surpasses the other models, including the Transformer (93%), Faster R-CNN (91%), and RetinaNet (88%), with Cascade R-CNN recording the lowest recall at 86%. This demonstrates that the proposed framework is more effective than the established baselines. The fig. 3 (b) compares the precision of various object detection models, which measures the accuracy of positive predictions. Our proposed model consistently achieves the highest precision, 97.20%, demonstrating exceptional accuracy with minimal false positives. The Transformer model performs next best at 94%, followed by Faster R-CNN (92%), RetinaNet (89%), and Cascade R-CNN (88%). This confirms the superior performance of proposed framework in achieving precise object detection. The fig. 3 (c) illustrates the comparative performance of various object detection models based on mean Average Precision (mAP), a key metric for overall accuracy. The proposed framework demonstrates superior and consistent performance, achieving an mAP 97.40%, which is notably higher than the Transformer model (95%). This result confirms its significant advantage over existing approaches in terms of both detection and classification.



(a) (b) (c)

Figure 4. comparative analysis of results (a) F1-score, (b) IoU and (c) Train Test Loss

The fig.4(a) presents an F1-score comparison across different object detection models. The F1-score, as the harmonic mean of precision and recall, measures a model's overall balance and effectiveness. Proposed model achieves the highest F1-score, 97.10%, demonstrating superior performance and reliability. The Transformer model follows with scores of 94%, while Faster R-CNN, RetinaNet, and Cascade R-CNN show progressively lower performance, confirming proposed model's robustness and balanced performance. The fig.4(b) compares the Intersection over Union (IoU) of various object detection models, a metric that quantifies the overlap between predicted and true bounding boxes. The proposed model achieves the highest IoU, 98.20%, which demonstrates its superior performance and consistency in precisely localizing kidney tumors. This performance is notably higher than the Transformer model (94%), Faster R-CNN (93%), RetinaNet (91%), and Cascade R-CNN (89%), confirming the significant advantages of proposed framework. The fig.4(c) illustrates the training and testing loss curves over 300 epochs. Both curves initially decline sharply, signifying rapid learning, before stabilizing with a minimal gap. This confirms effective model convergence and strong generalization, indicating successful learning from the dataset with minimal overfitting.

Conclusion

This study successfully addresses the challenges of kidney tumor (KT) detection by developing a lightweight and efficient medical image analysis algorithm. To mitigate the limitations of a small dataset, data augmentation was applied to enhance the model's generalization capabilities. The core of the framework, a Multi-Scale Feature Pyramid Transformer (MSFPT), was instrumental in improving spatial feature learning, while a compact attention mechanism refined the detection process by focusing on critical post-neck outputs. The proposed model demonstrated superior performance with an Intersection over Union (IoU) of 98.20%, a precision of 97.20%, a recall of 96.90%, and a mean Average Precision (mAP) of 97.40%. These results confirm the effectiveness of our transformer-based architecture in medical imaging, highlighting its potential for real-time clinical applications. Future work will focus on expanding the framework to include an advanced classifier for categorizing various kidney diseases .

References

1. Usha, M. G., M. S. Shreya, S. Supreeth, and G. Shruthi. "Kidney Tumor Detection Using MLflow, DVC and Deep Learning." In *2024 Second International Conference on Advances in Information Technology (ICAIT)*, vol. 1, pp. 1-7. IEEE, 2024. Doi: 10.1109/ICAIT61638.2024.10690537.
2. Gharaibeh, Maha, Dalia Alzu'bi, Malak Abdullah, Ismail Hmeidi, Mohammad Rustom Al Nasar, Laith Abualigah, and Amir H. Gandomi. "Radiology imaging scans for early diagnosis of kidney tumors: a review of data analytics-based machine learning and deep learning approaches." *Big Data and Cognitive Computing* 6, no. 1 (2022): 29. Doi: 10.3390/bdcc6010029.
3. Abdelrahman, Abubaker, and Serestina Viriri. "Kidney tumor semantic segmentation using deep learning: A survey of state-of-the-art." *Journal of imaging* 8, no. 3 (2022): 55.Doi: 10.3390/jimaging8030055.
4. Mahmud, Sakib, Tariq O. Abbas, Adam Mushtak, Johayra Prithula, and Muhammad EH Chowdhury. "Kidney cancer diagnosis and surgery selection by machine learning from CT scans combined with clinical metadata." *Cancers* 15, no. 12 (2023): 3189. Doi: 10.3390/cancers15123189.
5. Shon, Ho Sun, Erdenebileg Batbaatar, Kyoung Ok Kim, Eun Jong Cha, and Kyung-Ah Kim. "Classification of kidney cancer data using cost-sensitive hybrid deep learning approach." *Symmetry* 12, no. 1 (2020): 154. Doi : 10.3390/sym12010154.
6. Kumar, Yogesh, Tejinder Pal Singh Brar, Chhinder Kaur, and Chamkaur Singh. "A comprehensive study of deep learning methods for kidney tumor, cyst, and stone diagnostics and detection using CT images." *Archives of Computational Methods in Engineering* 31, no. 7 (2024): 4163-4188, Doi: 10.1007/s11831-024-10112-8.
7. Usha, M. G., M. S. Shreya, S. Supreeth, and G. Shruthi. "Kidney Tumor Detection Using MLflow, DVC and Deep Learning." In *2024 Second International Conference on Advances in Information Technology (ICAIT)*, vol. 1, pp. 1-7. IEEE, 2024. Doi: 10.1109/ICAIT61638.2024.10690537
8. Zhang, Meng, Zheng Ye, Enyu Yuan, Xinyang Lv, Yiteng Zhang, Yuqi Tan, Chunchao Xia, Jing Tang, Jin Huang, and Zhenlin Li. "Imaging-based deep learning in kidney diseases: recent progress and future prospects." *Insights into imaging* 15, no. 1 (2024): 50. Doi: 10.1186/s13244-024-01636-5
9. Suharsono, Judi, and Sulis Candra. "Murabaha In Sharia Added Value, An Effort To Increase Probolinggo Shallot Farmers' Economic Scale And Spirituality." *Available at SSRN 2596062* (2013). Doi: 10.2139/ssrn.2596062
10. Patel, Vinitkumar Vasantbhai, Arvind R. Yadav, Prateek Jain, and Linga Reddy Cenkeramaddi. "A systematic kidney tumour segmentation and classification framework using adaptive and attentive-based deep learning networks with improved crayfish optimization algorithm." *IEEE Access* 12 (2024): 85635-85660. Doi : 10.1109/ACCESS.2024.3410833
11. Shingade, S. D., Rohini Mudhalwadkar, and K. M. Masal. "Random forest machine learning classifier for seed recommendation." In *2022 International Conference on Edge Computing and Applications (ICECAA)*, pp. 1385-1390. IEEE, 2022. Doi: 10.1109/ICECAA55415.2022.9936120
12. Masal, K. M., Shripad Bhatlawande, and Sachin D. S. "An integrated region proposal and spatial information guided convolution network based object recognition for visually impaired persons'

indoor assistive navigation." *The Imaging Science Journal* 72, no. 7 (2024): 884-897. Doi: 10.1080/13682199.2023.2230419

13. Shingade, S. D., Rohini Mudhalwadkar, and K. M. Masal. "Random Forest, DT and SVM Machine Learning Classifiers for Seed with Advanced WSN Sensor Node." In *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, pp. 321-326. IEEE, 2022. Doi: 10.1109/ICACRS55517.2022.10029310.
14. Shingade, S. D., and Rohini Prashant Mudhalwadkar. "Hybrid extreme learning machine based bidirectional long short-term memory for crop prediction." *Concurrency and Computation: Practice and Experience* 35, no. 2 (2023): e7482. Doi: 10.1002/cpe.7482.
15. Masal, K. M., Shripad Bhatlawande, and Sachin D. S.. "Deep Learning Attentional Dense based Indoor Object Recognition for Visually Impaired People." In *2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 658-663. IEEE, 2023. Doi : 10.1109/ICECA58529.2023.10394723
16. Chakkaravarthy, Midhun, Dimitrios A. Karras, and Komal M. "Advancing Deep Learning Techniques for Early Detection and Classification of Renal Cell Carcinoma: A review." *SGS-Engineering & Sciences* 1, no. 1 (2025). Doi : https://spast.org/techrep/article/view/5265
17. Chakkaravarthy, Midhun, Dimitrios A. Karras, and Komal M . "-Hybrid Deep Pyramid Convolutional Coordinate Attentional Residual Autoencoder Network for Kidney Tumor Diagnosis from CT Scans:-." *SGS-Engineering & Sciences* 1, no. 2 (2025). Doi: https://spast.org/ techrep/ article/view/5467.
18. Kadhim, Dhuha Abdalredha, and Mazin Abed Mohammed. "Advanced machine learning models for accurate kidney cancer classification using CT images." *Mesopotamian Journal of Big Data* 2025 (2025): 1-25. Doi: 10.58496/MJBD/2025/001
19. Taha, E. T. E. M., and T. E. K. E. Mustafa. "Enhanced deep learning based decision support system for kidney tumour detection." *BenchCouncil Transactions on Benchmarks, Standards and Evaluations* 4, no. 2 (2024): 100174. Doi: 10.1016/j.tbench.2024.100174
20. Akkasaligar, Prema T., Santosh Pattar, Adarsh Uppin, Khushi Kori, and Achyut Kulakarni. "Kidney tumor detection using computed tomography scan images through CNN." In *2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT)*, pp. 1-6. IEEE, 2024. Doi: 10.1109/ICDCOT61034.2024.10516254
21. Sheng, Ting-Wen, Djeane Debora Onthoni, Pushpanjali Gupta, Tsong-Hai Lee, and Prasan Kumar Sahoo. "Segmentation of ADPKD Computed Tomography Images with Deep Learning Approach for Predicting Total Kidney Volume." *Biomedicines* 13, no. 2 (2025): 263. Doi: 10.3390/biomedicines13020263
22. Habchi, Yassine, Hamza Kheddar, Yassine Himeur, Mohamed Chahine Ghanem, Abdelkrim Boukabou, Shadi Atalla, Wathiq Mansoor, and Hussain Al-Ahmad. "Deep transfer learning for kidney cancer diagnosis." *arXiv preprint arXiv:2408.04318* (2024). Doi: 10.48550/arXiv.2408.04318.