Research Paper

Deep Learning-Based Multi-Scale Attention Fusion Framework for Automated Kidney Cancer Localization and Classification

K. Sujatha[1*], R. S. Ponmagal[2], M. Malathi[3], V. Srividhya[4], S. Saranya[5], P. Sinthia[6], K. S. Naveen Kumar[7]

1. *Department of EEE, Dr. M.G.R. Educational and Research Institute, Chennai, India.*
2. *Department of Computing Technologies, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.*
3. *Department of ECE, Rajalakshmi Institute of Technology, Chennai, Tamil Nadu, India.*
4. *Department of EEE, Meenakshi College of Engineering, Chennai, Tamil Nadu, India.*
5. *Department of Biomedical Engineering, Dr. M.G.R. Educational and Research Institute, Chennai, India.*
6. *Department of Biomedical Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India.*
7. *M.E. – Embedded Systems and Technology, Dhanalakshmi College of Engineering, Chennai, India.*

Citation Sujatha K, Ponmagal RS, Malathi M, Srividhya V, Saranya S, Sinthia P, et al. Deep Learning-Based Multi-Scale Attention Fusion Framework for Automated Kidney Cancer Localization and Classification. International Journal of Medical Toxicology and Forensic Medicine. 2026; 16:E52494.

https://doi.org/10.22037/ijmtfm.v16.52494

Article info:

Received: 15 June, 2026

Accepted: 01 July, 2026

Published: 16 September, 2026

Keywords:

Kidney cancer, U-Net, CT, Feature fusion, Deep learning, Localization

A B S T R A C T

Background: CT imaging is one of the most commonly used methods for diagnosing renal malignancy. Manual evaluation is cumbersome and highly subjective as it depends on the skill set of an expert nephrologist.

Methods: This paper proposes an automated deep learning approach based on a two-stage model in lesion localization and early-stage kidney cancer classification. First, the kidney regions are segmented using an improved U-Net architecture with multi-scale feature fusion, enabling accurate identification of affected kidney regions. Next, the segmented lesion maps are fed into a classifier network (ResNet/DenseNet) to classify early-stage kidney cancer.

Results: Segmentation-based lesion information fusion improves stage-wise discrimination and limits false predictions during the early stages of kidney cancer detection. The proposed method clearly demonstrates improved predictive accuracy and operational efficiency compared to conventional end-to-end classification models. The integration of segmentation and classification provides more reliable identification of cancerous regions and supports improved stage-wise prediction.

Conclusion: The approach provides clinical interpretability by marking the affected portion of the kidney and assigning a stage-wise cancer label. The proposed framework can assist clinicians in early-stage kidney cancer assessment by providing both lesion localization and stage-wise classification. It also offers a promising basis for developing efficient and interpretable computer-aided diagnostic systems for renal malignancy.

*** Corresponding Author:**

K. Sujatha, PhD

Department of Electrical and Electronics Engineering (EEE), Dr. M.G.R. Educational and Research Institute, Chennai, India.

E-mail: *sujathak73586@gmail.com*

Introduction

Kidney cancer is one of the most frequently diagnosed cancers found in the areas of the restroom, or urology. This form of cancer also continues to raise many issues with respect to the impact it has on the world today, primarily due to the increasing number of new instances of diagnosed cancer each year and the rise in death rates among those who are diagnosed. Of all forms of kidney cancer, approximately 90% are comprised of either renal cell carcinoma (also known as RCC) or another subtype of RCC (i.e., clear cell RCC, papillary RCC, or chromophobe RCC) [1, 2]. Early identification of kidney cancer remains the key component to improving survivor health and overall treatment outcomes for kidney cancer patients. Currently, under general clinical practice guidelines, the most common method for evaluating the presence of kidney cancer is CT imaging due to its rapid acquisition, reliable results, and ability to detect, locate, determine the extent of, and evaluate the disease process [3]. Recent advancements in artificial intelligence (AI) and deep learning (DL) have significantly advanced automated medical image analysis, particularly for disease identification/classification. Convolutional neural networks (CNNs) have shown impressive capabilities in both identifying diseases and classifying various forms of them using real-world datasets derived from medical imaging [4].

Deep learning models (e.g., ResNet [5] and DenseNet [6]) have greatly improved the effectiveness of feature extraction during training, resulting in better gradient propagation through the model, thus enabling automated classification of kidney cancer patients using their medical images [7,8]. In addition to classifying a disease, segmenting a lesion is also important in improving localization accuracy and clinical interpretability. U-Net developed a new encoder-decoder architecture that incorporates skip connections and established a "gold standard" for biomedical image segmentation [9]. Many improved versions, such as Attention U-Net [10], supervised Attention U-Net [11], and U-Net++ [12], were developed to enhance U-Net's ability to extract and delineate lesion boundaries in medical images. Furthermore, multi-scale feature representations and fusion techniques allow deep learning to learn from both the global context of the image and detailed pathological features. Multi-scale feature fusion frameworks [13, 14] and attention-based segmentation methods have demonstrated significantly higher lesion-segmentation sensitivity than traditional methods, along with an enhanced ability to differentiate individual features. Combining segmentation and classification methods has also been shown to improve disease detection and the visual interpretability of resulting diagnoses [15,16].

Recent developments have also explored integrating convolutional neural networks with transformer-based architectures to improve the representation of local and global contextual information in medical image segmentation. CTNet, for example, combines convolutional architectures with Vision Transformer-based feature representation and attention mechanisms to enhance medical image segmentation [17]. Such hybrid architectures are particularly relevant to medical imaging because convolutional operations can capture local spatial features, while transformer-based mechanisms can model long-range contextual dependencies. Despite these advances, many issues remain in the current infrastructure of deep learning models. Most deep learning classification models are trained as black boxes and do not provide accurate information about the locations of lesions. Segmentation methods are particularly limited in detecting small tumors, low-contrast pathological tissues, or in generalizing from one study to another [18, 19]. Most segmentation and classification frameworks exist independently of one another and are not integrated.

Addressing these drawbacks, this study presents a deep learning-based automated system that combines U-Net-based lesion localization with multi-scale feature fusion and classification for CT image-based kidney cancer detection. The proposed framework improves lesion localization, classification accuracy, and clinical interpretability. Segmentation-guided feature learning enhances these capabilities.

Research into artificial neural networks and deep learning algorithms for automating the diagnosis of kidney cancer via CT and MRI imaging has been extensive. Initial work based on CNNs demonstrated great potential for detecting and classifying kidney tumors by automatically extracting relevant image features. Other architectures, such as ResNet and DenseNet, introduced methods for reusing features and improving information flow between layers, resulting in better classification performance for medical imaging applications. Several studies used various CNN-based approaches (e.g., VGG16, ResNet50, and Hybrid CNN frameworks) to distinguish benign from malignant kidney tumors, achieving excellent overall predictive accuracy.

With the introduction of U-Net as an architecture for biomedical image segmentation, segmentation-based approaches have received increased attention. The Attention U-Net introduced attention mechanisms to increase focus on pathological regions and improve segmentation accuracy. In addition, Ma et al. developed a supervised attention U-Net to enhance the segmentation of both blood vessels and lesions, while U-Net++ improved feature aggregation through restructured skip connections. Enhanced U-Net-based segmentation methods improved lesion boundary delineation and localization. Weakly supervised and anomaly-guided methods were proposed to address the limitations of limited annotated data and to increase the sensitivity of lesion detection. Seebock proposed a method for anomaly-guided weakly supervised lesion segmentation in settings with limited annotations. Generalizable segmentation models were also created to improve segmentation robustness.

Multiscale feature fusion techniques have proven effective at capturing pathological features across multiple spatial resolutions. Soni's [13] work on a multi-scale feature-fusion attention U-Net enabled improved segmentation sensitivity and structural representation. Liu et al. [14] developed IMFF-Net, which leveraged multi-scale contextual information for enhanced segmentation accuracy. Hybrid segmentation-classification models have been investigated to increase diagnostic accuracy and/or improve the interpretability of disease diagnoses. Diao et al. [16] developed a dual-guidance framework that integrates segmentation and classification to improve the diagnosis of renal cancer. Mani et al. [15] built a mask-guided CNN architecture that performs lesion detection and classification through guided feature learning.

More recently, hybrid CNN-transformer architectures have been explored for medical image segmentation to address the limitations of conventional convolutional architectures in capturing long-range contextual information. Zhang et al. proposed CTNet, which combines convolutional neural networks and Vision Transformer-based mechanisms for medical image segmentation [17]. The architecture employs attention-based feature processing to better integrate local and global contextual information. Such approaches demonstrate the potential of combining CNN-based local feature extraction with transformer-based contextual representation for improving segmentation performance.

While many of these proposed approaches have achieved good performance, there are still significant limitations involved with most of these systems. Common limitations of these methodologies include small lesion detection, low-contrast imaging, imaging variability, and cross-dataset generalization [1, 18, 19]. Furthermore, many classification-oriented systems are not explainable and do not localize lesions with sufficient precision to be clinically relevant. As such, there is a clear need to develop integrated, segmentation-guided, multi-scale feature-fusion systems for the reliable diagnosis of kidney cancer.

Materials and Methods

The overall goal of the described framework is to build a system capable of providing early diagnosis of kidney cancer by integrating CT (computed tomography) image analysis with a 2-step pipeline comprising image segmentation and classification. In this first step, high-resolution CT scans are acquired using image acquisition modules. Once these scans have been obtained, they are preprocessed to improve visibility of the underlying structure. After pre-processing, the CT scans will undergo pixel-based segmentation using a U-Net lesion-localization algorithm to obtain a pathology mask from which the presence of a lesion can be determined. By implementing multi-scale feature extraction and attention-based feature fusion techniques on the segmented lesions, the fine details and structure of the lesions will be optimally represented. Once feature-fused lesion maps are produced, they, along with the probability mask, are fed into a deep classification network to determine the presence or absence of disease. Outputs will include localized lesion maps, disease labels, severity heat maps, and visually interpretable overlays providing a comprehensive AI-based decision support system for assessing available data on the prospective diagnosis of kidney cancer at its earliest stages, as depicted in Figure 1.

Overall System Architecture

The main goal of this proposed framework is to develop a tool that automatically diagnoses kidney cancer in its early stages from CT scans using a combined segmentation-and-classification framework. The system achieves this through a segmentation/classification pipeline in which lesion localization is performed in conjunction with feature fusion from multiple-scale, attention-based deep neural networks (DNNs), so that any predicted Class 1 lesions are located relative to other lesions rather than focusing solely on surrounding (non-cancerous) tissue. The jointly optimized classification (as well as for both input CT images) and segmentation pipeline for joint optimization will produce accurate and clinically actionable results.

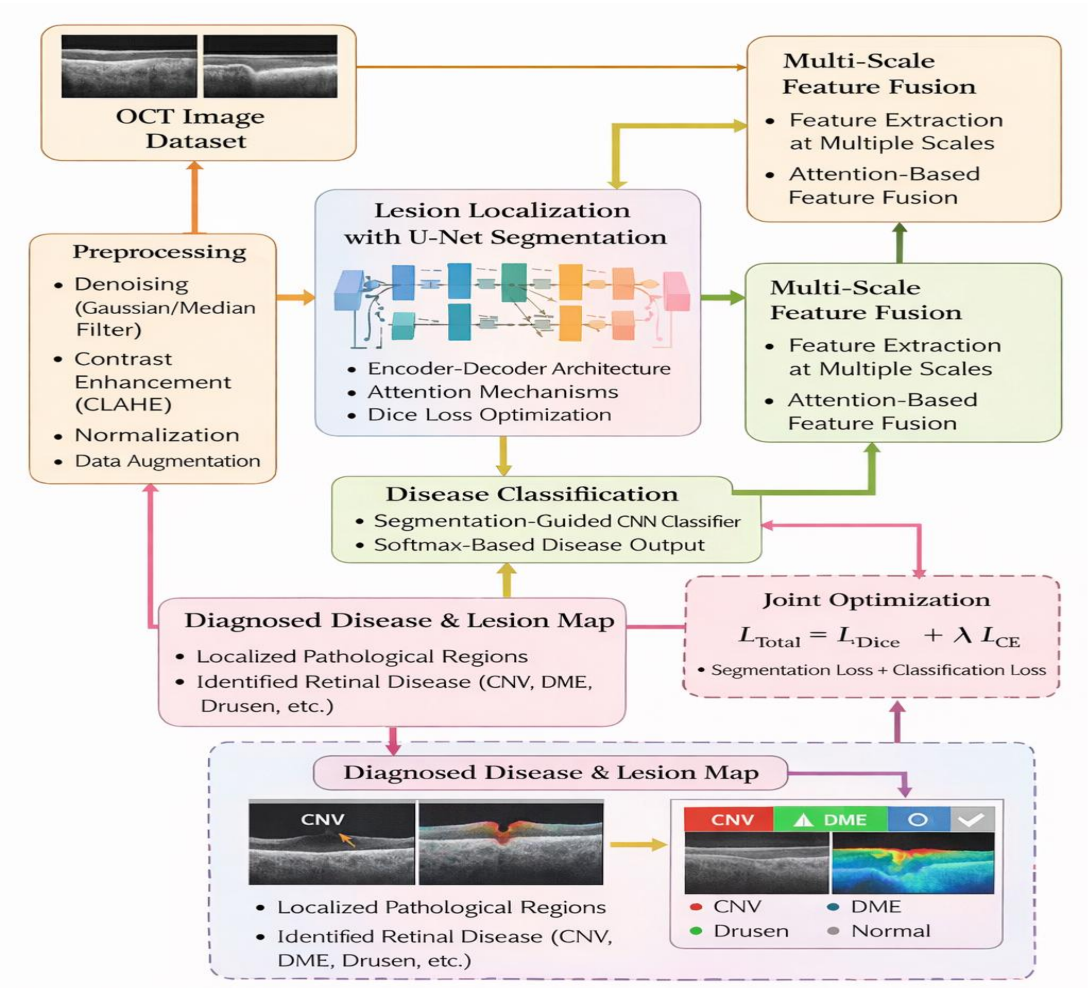

Figure 1. Proposed flowchart of the automated OCT analysis system.

Image Pre-processing

The renal cancer dataset was used to train and evaluate the proposed model. Although this dataset provides a good basis for developing the approach, it is important that its generalizability be tested on a much larger multicenter dataset.

The destructive effects of speckle noise and contrast variations are frequently observed in OCT images. The four pre-processing steps applied to the input image of interest are:

- Noise Suppression - Using either Gaussian or median filtering to reduce the presence of speckle noise in the image.

- Contrast Enhancement - Using CLAHE to enhance the visibility of the retinal layers within the image.

- Normalization - Normalizing pixel intensities to the range of [0, 1] across all pixels of the image.

- Data Augmentation - Randomly rotating, flipping, and zooming the image will increase the ability to generalize from a small data set.

Lesion Segmentation Network

Using a modified U-Net encoder-decoder architecture to localize lesions.

Network Structure:

- The encoder consists of several convolutional blocks that contain batch norm and ReLU activation functions.

- The decoder is composed of an upsampling layer for each of the convolution layers' skip connections in order to maintain the spatial detail of the original image.

- Attention Gates are used to highlight the abnormal areas of the image.

- The final output layer uses a Sigmoid function to generate the lesion probability map.

Segmentation Loss

Dice loss is employed to address class imbalance:

$$LDice = 1 - \frac{2 \mid S \cap G \mid}{|S| + |G|}$$

Multi-Scale Feature Fusion

Abnormalities appear at different scales; therefore, hierarchical features are extracted from multiple encoder stages:

$$\mathrm{F1, F2, F3, F4}$$

An attention-based fusion module combines these:

$$\mathrm{F_{Fusion}} = \sum_{i=1}^{4} \alpha i F i$$

Where αi are learnable attention weights? This enables simultaneous capture of fine lesion textures and global structure.

Segmentation-Guided Disease Classification

The fused features are concatenated with the lesion mask:

$$Ffinal = Concatenate(Ffusion, S(X))$$

This representation is passed to a deep CNN classifier (Dense-Net/Res-Net backbone):

$$y = Softmax(fcls(Ffinal))$$

Where Y denotes predicted disease probabilities.

Classification Loss

Cross-entropy loss is used:

$$LCE = -\sum yi \; log(y\hat{\ }i)$$

3.5.2 Joint Optimization

Segmentation and classification are trained jointly:

$$Ltotal = LDice + \lambda LCE$$

This multi-task learning improves both localization and diagnosis accuracy.

Results

To identify a proposed deep-learning framework for diagnosing kidney cancer in its early stages, metrics for both segmentation and classification have been calculated. Such evaluative metrics ensure that any measure of lesion localization will also correlate with the kidney's disease category classification.

Evaluation metrics

To identify a deep-learning-based diagnostic framework for early-stage kidney cancer, the evaluation system will employ metrics at both the segmentation and classification levels. The metrics will be used to assess the accuracy of kidney cancer and lesion location classifications.

The renal lesion segmentation metrics can be used to assess the accuracy of lesion localization in the kidneys on CT scans.

Dice Coefficient (F1 Score)

Measures the overlap between the predicted lesion mask and the ground-truth annotation, particularly effective for small or thin pathological regions.

$$\mathrm{Dice} = \frac{2|P \cap G|}{|P| + |G|}$$

Where **P** represents the predicted mask, and **G** denotes the ground-truth mask.

Intersection over Union (IoU)

Evaluates segmentation precision by measuring the ratio of the intersection to the union of predicted and ground-truth regions.

$$\mathrm{IoU} = \frac{|P \cap G|}{|P \cup G|}$$

A higher IoU indicates more accurate lesion boundary localization.

Pixel Accuracy

Represents the proportion of correctly classified pixels among all pixels.

$$\text{Pixel Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative pixels, respectively.

The comparative Segmentation performance of the models shown in Figure 2 is illustrated by Table 1 and the associated bar chart. The baseline U-Net shows good results, but additional improvements are observed with both Attention U-Net and U-Net++, suggesting that enhanced feature propagation and attention mechanisms contribute to these gains. The proposed model achieved the highest overall Dice score (0.95) and IoU (0.91), demonstrating the highest overlap accuracy and precise localization of lesion boundaries. The proposed model also produced the highest level of Sensitivity (0.96), confirming its ability to accurately delineate pathological regions, along with the highest level of Specificity (0.97), indicating that very few false positives were produced by the proposed model.

Kidney Disease Classification Metrics

Table 1. Dataset details.

Method	Dice score	InU	Sensitivity	Specificity
U-net	0.88	0.81	0.90	0.93
Attention U-net	0.91	0.85	0.92	0.95
U-net ++	0.93	0.88	0.94	0.96
Proposed model	**0.95**	**0.91**	**0.96**	**0.97**

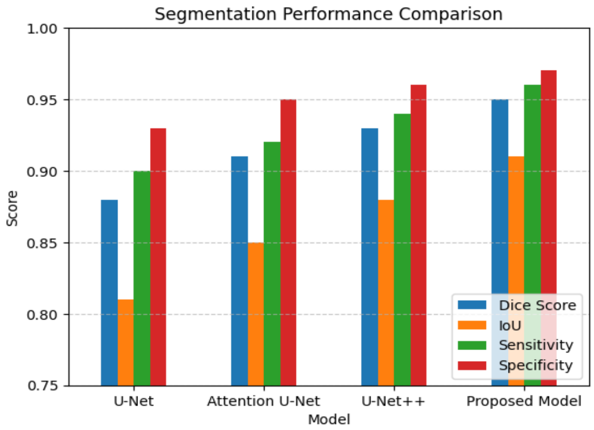

Figure 2. Segmentation performance analysis.

These metrics evaluate the model's effectiveness in categorizing OCT images into disease classes.

Accuracy

Measures the overall proportion of correctly classified samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

Indicates how many predicted disease cases are actually correct.

$$Precision = \frac{TP}{TP + FP}$$

Recall (Sensitivity)

Measures the model's ability to correctly detect diseased cases.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score

Harmonic mean of Precision and Recall, balancing false positives and false negatives.

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{Precision + Recall}$$

As seen in Table 2 and the classification performance charts, the models evaluated show Figure

Table 2. Classification performance comparison.

Class	Precision	Recall	F1 Score
CNV	99.1	98.8	98.9
DME	98.7	99.2	98.9
Drusen	98.8	98.6	98.7
Normal	99.3	99.4	99.3

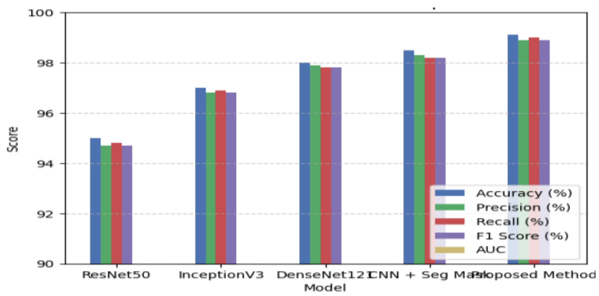

Figure 3. Classification performance analysis.

3 an increasing trend in their classification performance.

The results for both ResNet50 and InceptionV3 provide strong baseline performance; however, the introduction of DenseNet121 yields even higher performance due to its additional feature-reuse capability. The addition of features from segmentation-guided information (CNN and Seg Mask) also improves performance, indicating the importance of lesion-specific features. Additionally, the proposed method achieves the best overall results, as measured by the highest classification accuracy (99.1%), F1 score (98.9%), and area under the curve (0.995), along with class discrimination and robustness. The graphical representation of the results demonstrates that the proposed method consistently outperforms the other methods across the different measures of classification performance.

Discussion

Class-Wise Analysis

The purpose of class-wise performance analysis is to provide insight into the ability of the proposed method to differentiate between the various categories of kidney cancer. By evaluating individual classes, it could be determined if the method has a balanced level of performance (according to either pathological or normal) of the entire pathology/normal class of patients.

The accuracies achieved by each configuration are summarized in Table 3. The results shown in Figure 4 indicate that applying single-scale feature extraction yielded a strong baseline performance of 97.8%. The multi-scale fusion approach achieved 98.6%, indicating that extracting features across multiple spatial scales improves performance by leveraging information at different resolutions. The highest accuracy (99.1%) was achieved when segmentation information was combined with the multi-scale fusion feature representations, confirming that focusing on lesions for feature extraction was effective in enhancing discriminative learning.

Table 3. Class wise analysis.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC
ResNet50	95.0	94.7	94.8	94.7	0.96
InceptionV3	97.0	96.8	96.9	96.8	0.98
DenseNet121	98.0	97.9	97.8	97.8	0.99
CNN + Seg Mask	98.5	98.3	98.2	98.2	0.992
Proposed Method	**99.1**	**98.9**	**99.0**	**98.9**	**0.995**

International Journal of
Medical Toxicology & Forensic Medicine

Effect of Multi-Scale Fusion

An accuracy comparison was conducted across different feature-extraction configurations to assess the impact of feature representation strategies. This experiment assessed how the overall classification performance was affected by integrating multi-scale fusion and segmentation guidance into the classification process.

Table 4 summarizes the accuracy achieved by each configuration. The results in Figure 5 indicate that single-scale feature extraction provides a strong baseline accuracy of 97.8%. Incorporating multi-scale fusion improves performance to 98.6%, demonstrating the advantage of capturing features at different spatial resolutions. The highest accuracy of 99.1% is achieved when segmentation information is combined with multi-scale fusion, confirming that lesion-focused features significantly enhance discriminative learning.

Computational Efficiency

Computational efficiency is a critical consideration in the practical implementation of deep learning systems for real-time clinical use. Therefore, both the training time and inference speed were evaluated to assess the resource requirements for implementing the suggested model.

Table 5 compares the runtime of the proposed solution versus that of DenseNet121. As is typical, the proposed solution had a longer training time (7.8 hours) than DenseNet121 (7.2 hours), likely due to the model's segmentation-guided, multi-scale fusion components.

Table 4. Accuracy analysis.

Configuration	Accuracy
Single-scale features	97.8%
Multi-scale fusion only	98.6%
Segmentation + fusion	**99.1%**
Configuration	**Accuracy**
Single-scale features	97.8%

International Journal of
Medical Toxicology & Forensic Medicine

Table 5. Comparision of computational effiency.

Model	Training Time (hrs)	Inference (ms/image)
DenseNet121	7.2	22
Proposed	7.8	20

International Journal of
Medical Toxicology & Forensic Medicine

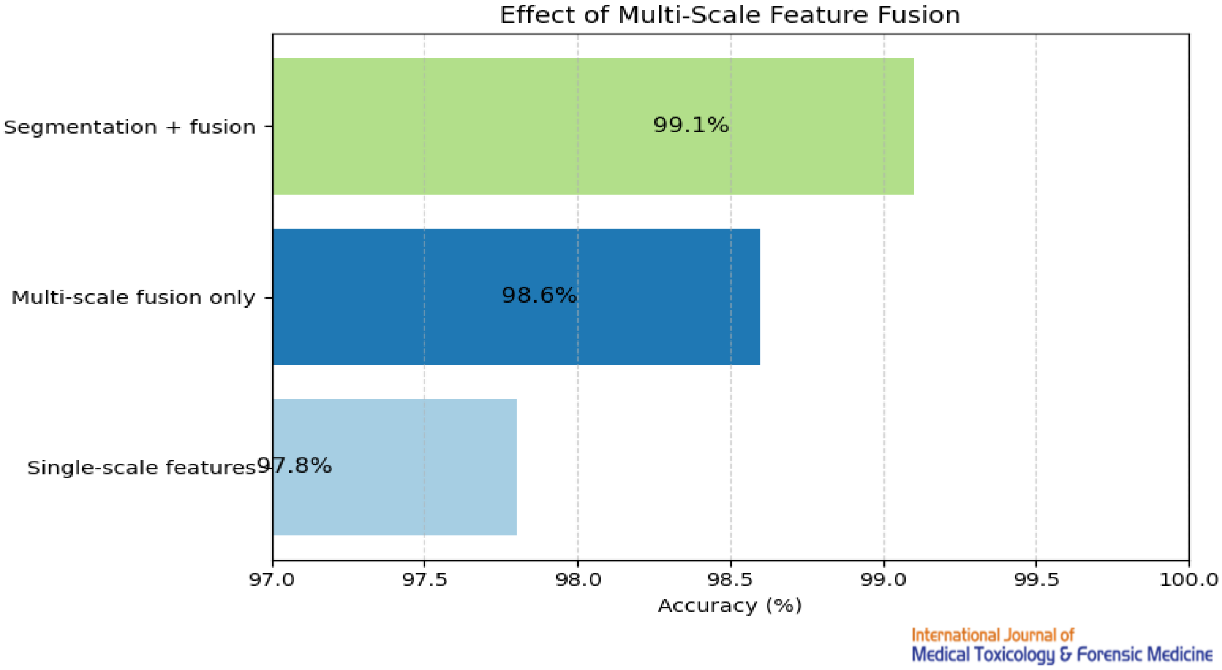

Figure 5. Effect of multi- scale feature fusion.

The suggested framework provides lesion maps alongside attention visualizations that highlight damaged kidney regions, making its predictions easier to interpret. While visual explanations may be useful for doctors in understanding the model's strategy, it was not possible to obtain validation of the method from professional radiologists within the framework of this study.

Future efforts will extend the framework to encompass other stages of kidney cancer and conduct three-dimensional analyses of the CT volumes to provide additional spatial context. Additionally, incorporating semi-supervised learning into the framework and refining and optimizing its models may facilitate real-time deployment in data-starved clinical settings. Improving the model's explainability and increasing integration with clinical systems will lead to greater clinician trust, better use of monitoring tools, and improved decision support. The proposed framework can be extended to multi-stage kidney cancer classification and 3D CT volume analysis. Future work will focus on integrating volumetric information and semi-supervised learning to improve robustness and clinical applicability.

Conclusion

We propose an automated method for diagnosing kidney cancer using computed tomography that integrates lesion segmentation and multi-scale feature-guided classification within a unified framework. An enhanced U-Net model provides accurate disease localization, and the combination of multi-scale features strengthens disease discrimination. Our evaluations demonstrate consistent improvement over the baseline models for both segmentation performance and classification performance. Incorporating both lesion maps and attention-based visualizations into the model has improved its interpretability for clinicians using it and increased their trust in it. The proposed method enables the integration of precise lesion localization and robust classification into an accurate, explainable means of supporting computer-aided assessments for the early stages of kidney cancer. The two-stage model proposed here localizes the lesion through segmentation followed by classification, thereby allowing the classifier to focus on the targeted area. Such enhancement reduces background noise interference, improves feature extraction, reduces false predictions, and increases classification accuracy, especially in the early detection of kidney cancer.

Ethical Approval

Was not applicable.

Acknowledgment

None.

Funding

No external funding was received for this research.

Conflicts of Interest

The authors report that there are no competing interests to declare.

References

[1] Zhang Y. Retinal OCT image segmentation with deep learning. Comput Vis Image Underst. 2025. [DOI: 10.1016/j.cviu.2025.103123]

[2] Bhoopalan S, et al. Hybrid methods for retinal OCT segmentation. Sci Rep. 2025. [DOI: 10.1038/s41598-025-89262-z]

[3] Soni A. Multi-scale feature fusion attention UNet for retinal vessel segmentation. Sci Rep. 2025. [DOI: 10.1038/s41598-025-28707-x]

[4] Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 2018;172(5):1122–31.e9. [DOI: 10.1016/j.cell.2018.02.010]

[5] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Conf Comput Vis Pattern Recognit. 2016:770–8. [DOI: 10.1109/CVPR.2016.90]

[6] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proc IEEE Conf Comput Vis Pattern Recognit. 2017:2261–9. [DOI: 10.1109/CVPR.2017.243]

[7] Wang S. Deep learning for retinal OCT classification. IEEE Access. 2020. [DOI: 10.1109/ACCESS.2020.3015176]

[8] Tsuji K. Practical OCT classification using deep learning. Sci Rep. 2020. [DOI: 10.1038/s41598-020-68279-6]

[9] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. MICCAI. 2015:234–41. [DOI: 10.1007/978-3-319-24574-4_28]

[10] Oktay O. Attention U-Net for pancreas segmentation. arXiv Preprint. 2018. [DOI: 10.48550/arXiv.1804.03999]

[11] Ma J. Supervised attention U-Net for retinal vessel segmentation. Image Vis Comput. 2023. [DOI: 10.1016/j.imavis.2023.104207]

[12] Gao F. Macular edema segmentation in OCT using U-Net++. Appl Sci. 2022. [DOI: 10.3390/app10165701]

[13] Soni A. Multi-scale feature fusion attention UNet for retinal vessel segmentation. Sci Rep. 2025. [DOI: 10.1038/s41598-025-28707-x]

[14] Liu X. IMFF-Net: Integrated multi-scale feature fusion for retinal vessel segmentation. Biomed Signal Process Control. 2024. [DOI: 10.1016/j.bspc.2024.103907]

[15] Mani P, Ramachandran N, Paul SJ, Ramesh PV. Laceration assessment: advanced segmentation and classification framework for retinal disease categorization in optical coherence tomography images. J Opt Soc Am A. 2024. [DOI: 10.1364/JOSAA.526142]

[16] Diao Y. Dual guidance networks for OCT classification and segmentation. Comput Methods Programs Biomed. 2023. [DOI: 10.1016/j.cmpb.2023.107780]

[17] Zhang Z, Jiang S, Pan X. Ctnet: rethinking convolutional neural networks and vision transformer for medical image segmentation. Signal Image Video Process. 2024;18(3):2265–75. [DOI: 10.1007/s11760-023-02899-z]

[18] Seebock P. Anomaly-guided retinal OCT lesion segmentation using weak supervision. Med Image Anal. 2024. [DOI: 10.1016/j.media.2024.102905]

[19] Ganjee R, Ebrahimi Moghaddam M, Nourinia R. A generalizable approach based on the U-Net model for automatic intraretinal cyst segmentation in SD-OCT images. Int J Imaging Syst Technol. 2023. [DOI: 10.1002/ima.22893]