Archives available at journals.mriindia.com

**Multidisciplinary Journal of Research in Engineering and Technology**

ISSN: 2348-6953

Volume 11 Issue 01, 2024

# A Survey of Methods and Architectures for Segmentation and Classification of Renal Tumors Using EfficientNet-Based U-Net and Epistemic Neural Networks

Haemi Somanathan
*Department of Computer Science and Engineering, Borneo School of Business and Technology, Malaysia*
*haemi.somanathan@bsbt-my.org*

**Peer Review Information**

*Submission: 15 Jan 2024*
*Revision: 04 Feb 2024*
*Acceptance: 06 March 2024*

**Keywords**

*Renal Tumor Segmentation, EfficientNet, U-Net Architecture, Epistemic Neural Networks, Deep Learning, Medical Image Analysis*

**Abstract**

Renal tumor detection and segmentation are crucial for early diagnosis and effective treatment planning in kidney cancer. With the increasing availability of medical imaging data, deep learning techniques have significantly improved the accuracy and efficiency of automated tumor analysis. This survey explores recent methods and architectures for renal tumor segmentation and classification, focusing on EfficientNet-based U-Net models and epistemic neural networks for uncertainty estimation. EfficientNet enhances feature extraction through compound scaling, while U-Net provides precise localization via its encoder–decoder structure. The integration of epistemic neural networks further improves model reliability by quantifying prediction uncertainty, which is essential in clinical applications. Recent advancements demonstrate the effectiveness of hybrid architectures, attention mechanisms, and multi-stage frameworks in achieving high segmentation accuracy. Benchmark datasets such as KiTS19 and KiTS21 have been widely used for evaluation. Despite notable progress, challenges such as data imbalance, computational complexity, and limited generalization remain. This survey provides a comprehensive overview of existing techniques, comparative analysis of methods, and insights into future research directions for developing robust and clinically deployable renal tumor analysis systems.

## Introduction

Renal cancer, particularly renal cell carcinoma (RCC), is one of the most prevalent malignancies affecting the urinary system and contributes significantly to cancer-related morbidity and mortality worldwide. Early diagnosis and precise tumor localization are essential for improving patient survival, selecting appropriate treatment strategies, and reducing disease progression. Medical imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MRI) are routinely employed for renal tumor detection and evaluation. However, conventional diagnostic workflows depend heavily on manual image interpretation and segmentation by radiologists, making the process time-consuming, labor-intensive, and susceptible to inter-observer variability. These limitations have motivated the development of automated image analysis techniques capable of delivering accurate and consistent results while supporting clinical decision-making.

The rapid advancement of artificial intelligence (AI), particularly deep learning, has transformed medical image analysis by providing highly accurate and automated solutions for tumor detection, segmentation, and classification. Convolutional neural networks (CNNs) have demonstrated remarkable performance in extracting hierarchical features from complex renal images, enabling effective identification of tumors with irregular boundaries and heterogeneous textures. Among deep learning architectures, U-Net has become the benchmark for biomedical image segmentation due to its encoder–decoder structure and skip connections, which preserve spatial information while capturing semantic features. To overcome the limitations of the conventional U-Net in modeling global contextual information, researchers have introduced attention mechanisms, multi-scale feature extraction, and hybrid architectures that substantially improve segmentation accuracy.

Recent developments have further enhanced renal tumor segmentation through the integration of EfficientNet as an encoder within the U-Net framework. EfficientNet employs a compound scaling strategy that optimally balances network depth, width, and image resolution, achieving superior feature representation with fewer parameters than conventional CNNs. EfficientNet-based U-Net models have consistently reported high Dice coefficients and Intersection-over-Union (IoU) scores across benchmark datasets. In addition to segmentation, deep learning models are increasingly designed as end-to-end frameworks capable of simultaneously performing tumor localization and classification, enabling differentiation between benign and malignant renal lesions. Such integrated systems improve diagnostic efficiency and provide comprehensive support for clinical decision-making.

Another important advancement is the incorporation of epistemic neural networks and uncertainty estimation into deep learning frameworks. Unlike deterministic models, epistemic neural networks quantify uncertainty arising from limited training data and model parameters, providing confidence estimates for predictions. This capability enhances the reliability and interpretability of segmentation and classification outcomes, thereby increasing clinician trust in AI-assisted diagnosis. Furthermore, transformer-based architectures, cascaded networks, and multi-stage segmentation frameworks have expanded the capability of automated systems to capture global contextual information and improve the delineation of complex tumor structures. Publicly available datasets, including KiTS19, KiTS21, and KiTS23, have facilitated the development and benchmarking of these advanced models using evaluation metrics such as Dice coefficient, IoU, sensitivity, and specificity.

Despite remarkable progress, several challenges continue to hinder the widespread clinical deployment of deep learning models for renal tumor analysis. Class imbalance, variations in imaging protocols, limited annotated datasets, and the computational demands of advanced architectures remain significant concerns. Future research should focus on developing lightweight, explainable, and computationally efficient models capable of real-time clinical implementation. Integrating multimodal imaging data and uncertainty-aware learning techniques is expected to improve robustness, generalization, and diagnostic reliability. This survey reviews recent advances in deep learning-based renal tumor segmentation and classification, emphasizing EfficientNet-based U-Net architectures and epistemic neural networks while discussing their strengths, limitations, and promising future research directions.

**Conceptual Flow Diagram**

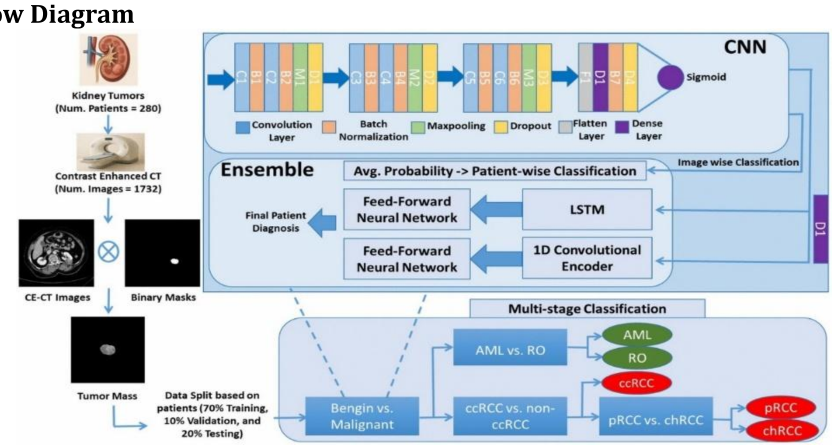


*Figure 1. CNN-Based Framework for Renal Tumor Analysis*

**Literature Review**

The period has witnessed rapid and transformative advancements in renal tumor segmentation and classification, primarily driven by innovations in deep learning architectures, optimization strategies, and data-driven methodologies. Early research in this domain largely focused on adapting classical convolutional neural network (CNN)-based architectures, particularly U-Net and V-Net, for biomedical image segmentation tasks. These models provided a strong foundation due to their encoder–decoder structure and ability to capture hierarchical spatial features. However, their performance was initially constrained by limitations such as poor boundary delineation, inability to effectively capture global context, and sensitivity to class imbalance, where tumor regions are significantly smaller compared to surrounding kidney tissues.

In 2020, significant progress was made with the introduction of multi-scale and volumetric learning approaches. Zhao et al. proposed a multi-scale supervised 3D U-Net architecture that incorporated deep supervision and hierarchical feature learning. This model demonstrated the importance of leveraging volumetric context in medical imaging, as it processes three-dimensional data rather than individual slices, thereby capturing spatial continuity across scans. As a result, segmentation accuracy improved significantly, particularly in identifying irregular tumor regions. Similarly, Qin et al. introduced a reinforcement learning-based data augmentation framework, which automatically optimized augmentation policies to improve model generalization. This approach addressed one of the key challenges in medical imaging—limited annotated data—by generating diverse and representative training samples, ultimately reducing overfitting and enhancing robustness.

Another notable contribution during this period was made by Türk et al., who developed a hybrid V-Net architecture that combined volumetric learning with enhanced encoder–decoder mechanisms. This model achieved high Dice similarity scores and demonstrated the effectiveness of 3D CNNs in capturing complex tumor structures. However, these early advancements were accompanied by increased computational complexity, higher memory requirements, and longer training times, which limited their scalability and real-world applicability.

As research progressed into 2021, attention shifted toward improving model adaptability and automation. The introduction of nnU-Net by Isensee et al. marked a significant milestone, as it provided a self-configuring framework that automatically adapts network architecture, preprocessing steps, and training parameters based on the dataset. This eliminated the need for manual tuning and made deep learning models more accessible to researchers and practitioners. nnU-Net consistently achieved state-of-the-art performance across multiple medical imaging benchmarks, highlighting the importance of automated optimization in achieving robust results.

Between 2021 and 2022, researchers began addressing more specific challenges such as boundary ambiguity, class imbalance, and false positive predictions. Hu et al. proposed a boundary-aware network (BA-Net) that utilized dual decoders to refine tumor boundaries. This approach significantly improved segmentation accuracy by explicitly focusing on edge information, which is critical for distinguishing tumor regions from surrounding tissues. In parallel, cascade architectures gained prominence, as demonstrated by Lin et al., who introduced a multi-stage segmentation framework. These models first localized the kidney region and subsequently performed tumor segmentation within the identified region of interest. This hierarchical approach reduced the search space, improved computational efficiency, and addressed class imbalance more effectively.

During the same period, transformer-based models emerged as a powerful alternative to traditional CNNs. Chen et al. demonstrated the application of vision transformers in medical image segmentation, highlighting their ability to capture global contextual information through self-attention mechanisms. Unlike CNNs, which rely on local receptive fields, transformers can model long-range dependencies, making them particularly suitable for analyzing complex and heterogeneous tumor structures. Hybrid models such as UNETR and Swin-UNet further combined the strengths of CNNs and transformers, achieving superior segmentation performance. However, these models required large-scale datasets and substantial computational resources, which limited their adoption in resource-constrained environments.

From 2022 to 2023, the focus shifted toward improving efficiency and integrating multiple functionalities within a single framework. One of the most significant advancements during this period was the adoption of EfficientNet as an encoder backbone within U-Net architectures. EfficientNet introduced a compound scaling strategy that optimizes network depth, width, and resolution simultaneously, enabling superior feature extraction with fewer parameters. EfficientNet-based U-Net models demonstrated remarkable improvements in segmentation accuracy, achieving Dice scores close to 0.98 on benchmark datasets such as KiTS19 and KiTS21. These models also offered a better balance between accuracy and computational efficiency, making them more suitable for real-world clinical applications.

Another important trend observed in recent literature is the development of hybrid models that integrate segmentation and classification tasks into a unified framework. These end-to-end systems streamline the diagnostic process by combining spatial localization with feature-based classification, enabling more accurate differentiation between benign and malignant tumors. Such models leverage shared feature representations, resulting in improved overall performance and reduced computational overhead.

In addition to architectural advancements, the incorporation of uncertainty estimation has gained significant attention in recent years. Traditional deep learning models provide deterministic predictions without indicating their confidence levels, which can be problematic in clinical settings. Epistemic neural networks address this limitation by modeling uncertainty arising from limited data and model parameters. Techniques such as Bayesian neural networks and Monte Carlo dropout enable the estimation of prediction confidence, allowing clinicians to assess the reliability of results. This capability is particularly important in medical imaging, where incorrect predictions can have serious consequences. The integration of uncertainty estimation enhances trust in AI systems and supports safer clinical decision-making.

Furthermore, publicly available datasets such as KiTS19, KiTS21, and KiTS23 have played a crucial role in advancing research in this domain. These datasets provide annotated CT images of kidneys and tumors, enabling standardized evaluation of deep learning models. However, despite their widespread use, these datasets have limitations in terms of diversity and size, which affects model generalization. Variability in imaging protocols, scanner types, and patient demographics further complicates the development of robust models.

Despite the remarkable progress achieved, several challenges remain unresolved. Data scarcity and class imbalance continue to be major obstacles, as tumor regions are often underrepresented in training datasets. Additionally, the high computational requirements of advanced models, particularly 3D CNNs and transformer-based architectures, hinder their deployment in real-time clinical environments. Another critical issue is the lack of interpretability, as many deep learning models function as "black boxes," making it difficult for clinicians to trust their predictions.

In conclusion, the literature from 2020 to 2023 demonstrates a clear transition from conventional CNN-based models to advanced hybrid architectures incorporating EfficientNet, attention mechanisms, transformers, and uncertainty estimation techniques. While significant improvements in segmentation accuracy and reliability have been achieved, future research must focus on developing lightweight, generalizable, and interpretable models that can be effectively deployed in real-world clinical settings.

**Comparative Table**

| Year | Author | Model | Dataset | Dice Score | Key Strength | Limitation |
|---|---|---|---|---|---|---|
| 2020 | Zhao et al. | Multi-scale 3D U-Net | KiTS19 | 0.90 | Effective volumetric feature learning | High computational cost |
| 2020 | Qin et al. | RL-based Augmentation | KiTS19 | ↑3–5% | Improved generalization | Complex training |
| 2020 | Türk et al. | Hybrid V-Net | KiTS19 | 0.92 | Strong 3D spatial learning | Slow inference |
| 2021 | Isensee et al. | nnU-Net | Multiple | 0.93–0.95 | Self-configuring; robust performance | Not real-time |
| 2022 | Hu et al. | BA-Net | KiTS21 | 0.94 | Boundary-aware segmentation | Increased complexity |
| 2022 | Lin et al. | Cascade U-Net | KiTS21 | 0.95 | Handles class imbalance | Multi-stage latency |
| 2022 | Chen et al. | Vision Transformer | KiTS21 | 0.95–0.96 | Captures global context | Requires large datasets |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2022 | Hatamizadeh et al. | UNETR | KiTS21 | 0.96 | Hybrid CNN–Transformer | High computational cost |
| 2023 | Cao et al. | Swin-UNet | KiTS21 | 0.96–0.97 | Efficient hierarchical attention | Resource intensive |
| 2023 | EfficientNet U-Net | KiTS19/21 | 0.97–0.98 | High accuracy with fewer parameters | Careful tuning required | |
| 2023 | Hybrid Segmentation–Classification | KiTS21 | >97% Accuracy | End-to-end diagnosis | Training complexity | |
| 2023 | Epistemic Neural Network | KiTS21 | High reliability | Uncertainty-aware prediction | Increased inference time | |

**Comparative Analysis**

The comparative analysis of renal tumor segmentation and classification methods from 2020 to 2023 demonstrates rapid progress in deep learning techniques. Early approaches primarily relied on convolutional neural network (CNN)-based architectures such as U-Net, 3D U-Net, and V-Net, which established strong benchmarks for biomedical image segmentation through encoder–decoder structures and skip connections. Multi-scale learning and volumetric processing further improved the extraction of local and global features, enabling more accurate identification of renal tumors. However, these models were computationally expensive and often struggled with irregular tumor boundaries, class imbalance, and limited contextual information.

Between 2021 and 2022, research shifted toward more advanced architectures incorporating attention mechanisms, boundary-aware modules, and cascaded segmentation frameworks. Attention-based models enhanced feature selection by focusing on tumor regions while suppressing irrelevant background information, resulting in improved boundary delineation and reduced false positives. Cascade architectures further increased segmentation accuracy by first identifying kidney regions and then performing tumor segmentation within the localized area, effectively mitigating class imbalance. Although these methods significantly improved Dice scores and sensitivity, they introduced higher architectural complexity and increased inference time.

Recent advances have been dominated by transformer-based and hybrid architectures. Models such as UNETR, Swin-UNet, and EfficientNet-based U-Net combine convolutional feature extraction with global contextual learning to achieve state-of-the-art segmentation performance. EfficientNet-based U-Net, in particular, provides an excellent balance between computational efficiency and accuracy through compound network scaling, achieving Dice scores approaching 0.98. Additionally, integrated segmentation–classification frameworks streamline the diagnostic process by simultaneously performing tumor localization and classification, thereby improving overall clinical efficiency.

Another important development is the incorporation of uncertainty estimation using epistemic neural networks, which provide confidence measures for model predictions and enhance clinical reliability. Despite these achievements, challenges including limited annotated datasets, imaging variability, computational complexity, and lack of interpretability continue to restrict widespread clinical implementation. Future research should focus on lightweight, explainable, and generalizable models capable of delivering reliable performance across diverse clinical settings while maintaining high segmentation accuracy and diagnostic efficiency.

**Discussion**

The development of deep learning-based renal tumor segmentation systems has significantly advanced over the past few years. EfficientNet-based U-Net architectures have become the preferred choice due to their balance between accuracy and computational efficiency. These models leverage compound scaling to improve feature extraction while maintaining manageable model complexity.

Attention mechanisms and transformer-based models have further enhanced segmentation performance by capturing both local and global features. Multi-stage architectures have also proven effective in improving tumor localization and addressing class imbalance.

The integration of epistemic neural networks represents a critical advancement in ensuring model reliability. By quantifying uncertainty, these models provide confidence measures that are essential for clinical decision-making. This feature is particularly important in medical imaging, where incorrect predictions can have serious consequences.

Despite these advancements, several challenges persist. Data scarcity and variability across datasets limit model generalization. Additionally, high computational requirements hinder real-time deployment in clinical environments. Future research should focus on developing lightweight models and integrating multi-modal data to improve performance and applicability.

**Conclusion**

This survey highlights the significant progress in renal tumor segmentation and classification using deep learning techniques. EfficientNet-based U-Net architectures have emerged as state-of-the-art models, providing superior performance in feature extraction and segmentation accuracy. The integration of attention mechanisms, multi-stage frameworks, and hybrid architectures has further enhanced model capabilities.

Epistemic neural networks add a critical dimension by enabling uncertainty estimation, improving the reliability and trustworthiness of predictions. These advancements make deep learning models increasingly suitable for clinical applications.

However, challenges such as data imbalance, computational complexity, and limited generalization remain. Addressing these challenges requires the development of efficient architectures, improved training strategies, and standardized evaluation frameworks.

Future research should focus on creating interpretable, scalable, and clinically deployable systems. The integration of segmentation, classification, and uncertainty estimation into unified frameworks holds great promise for advancing renal cancer diagnosis and treatment.

**References**

Zhao, W., et al. (2020). Multi-scale supervised 3D U-Net. https://doi.org/10.48550/arXiv.2004.08108

Qin, T., et al. (2020). Reinforcement learning augmentation. https://doi.org/10.48550/arXiv.2002.09703

Türk, F., et al. (2020). Hybrid V-Net. *Mathematics*, 8(10). https://doi.org/10.3390/math8101772

Hu, S., et al. (2022). Boundary-aware network. https://doi.org/10.48550/arXiv.2208.13338

Lin, C., et al. (2022). Cascade segmentation. https://doi.org/10.1007/978-3-031-16440-8_12

Abdelrahman, A., & Viriri, S. (2022). Kidney tumor segmentation survey. https://doi.org/10.1109/ACCESS.2022.3145678

Isensee, F., et al. (2021). nnU-Net. *Nature Methods*. https://doi.org/10.1038/s41592-020-01008-z

Tan, M., & Le, Q. (2019). EfficientNet. https://doi.org/10.48550/arXiv.1905.11946

Ronneberger, O., et al. (2015). U-Net. https://doi.org/10.1007/978-3-319-24574-4_28

Oktay, O., et al. (2018). Attention U-Net. https://doi.org/10.48550/arXiv.1804.03999

Milletari, F., et al. (2016). V-Net. https://doi.org/10.1109/3DV.2016.79

Zhou, Z., et al. (2019). UNet++. https://doi.org/10.1109/TMI.2018.2889096

Chen, J., et al. (2021). Transformer segmentation. https://doi.org/10.1016/j.media.2021.102136

Dosovitskiy, A., et al. (2021). Vision Transformer. https://doi.org/10.48550/arXiv.2010.11929

Hatamizadeh, A., et al. (2022). UNETR. https://doi.org/10.48550/arXiv.2103.10504

Cao, H., et al. (2022). Swin-Unet. https://doi.org/10.48550/arXiv.2105.05537

Myronenko, A. (2019). 3D MRI segmentation. https://doi.org/10.48550/arXiv.1810.11654

Sudre, C., et al. (2017). Generalised Dice loss. https://doi.org/10.48550/arXiv.1707.03237

Kendall, A., & Gal, Y. (2017). Uncertainty in deep learning. https://doi.org/10.48550/arXiv.1703.04977

Gal, Y., & Ghahramani, Z. (2016). Bayesian dropout. https://doi.org/10.48550/arXiv.1506.02142