Adaptive Three-Dimensional Identification of Renal Masses and Normal Kidney Tissues in Abdominal CT Using nnU-Net

Sahin Kilic, Buse Yaren Kazangirler, Caner Ozcan*, Ayse Keven, and Ahmet Sukru Alparslan

***Abstract*— The rapid and precise volumetric segmentation of incidentally detected renal lesions in computed tomography (CT) images is crucial for effective diagnosis and treatment planning. This study aims to introduce a three-dimensional (3D) deep learning methodology designed to accurately segment abdominal CT images of the kidneys, solid renal masses, and cystic renal masses, while also evaluating its performance. The study utilized an enhanced dataset comprising 880 cases. The dataset was systematically partitioned utilizing k-fold cross-validation, and a 3D U-Net architecture was trained within the nnU-Net framework, which is designed to automatically optimize for the specific characteristics of the dataset and the segmentation task at hand. A multidisciplinary study is hereby presented on kidney, solid renal mass, and cystic renal mass segmentation using the 3D U-Net, conducted with the participation of radiology and urology specialists. The results indicate a notably high level of performance, with an average accuracy of 95.18%, an average precision of 86.09%, an average recall of 90.90%, and an average F1-score of 87.49%. Specifically, the high accuracy and recall values suggest that the model has achieved a high level of success in identifying relevant lesions. These impressive metrics collectively underscore the robustness and clinical significance of the segmentation methodology employed. These strong results demonstrate the effective application of deep learning to this critical domain and are intended to directly contribute to the enhancement of diagnostic decision-making processes and the improvement of patient care in clinical practice.**

***Keywords*—Artificial Intelligence, Computed Tomography, Renal Mass Segmentation, nn U-Net, 3D U-Net, Medical Image Processing**.

I. Introduction

The kidneys, as vital organs, undertake several essential physiological roles, including the regulation of fluid-electrolyte balance and acid-base homeostasis, the excretion of metabolic waste products, and the modulation of blood pressure [1]. Kidney cancer, is characterized by the uncontrolled proliferation of kidney cells, ultimately leading to the formation of a neoplasm; renal cell carcinoma (RCC) is the leading histological subtype of kidney cancer, accounting for approximately 90% of all diagnosed cases. This form of cancer has gained recognition as one of the mortal urological cancers, underscoring the critical role of RCC in oncological research and therapeutic advancements. Its prevalence draws attention to the need for continued exploration into effective diagnostic and treatment options and a comprehensive understanding of its biological behavior [2]. CT imaging is a commonly used diagnostic tool in the medical field. In particular, life sciences demand the visualization of living tissues and soft matter with high spatial resolution, often in three-dimensional or four-dimensional data formats [2,3]. This imaging technique leverages advanced computerized data processing algorithms to produce cross-sectional images of internal anatomical structures using X-ray technology [4]. The resultant data facilitates the differentiation of tissue densities, thereby illuminating various anatomical structures, including organs, bones, blood vessels, as well as both benign and malignant lesions. CT scans are particularly instrumental in the diagnosis and evaluation of tumors, cysts, infections, and other pathological conditions, particularly within the abdominal cavity.

The causes of kidney cancer are complex and include various risk factors such as obesity, hypertension, tobacco use, genetic predispositions, and the malignant transformation of renal cells. RCC represents the predominant histological subtype of kidney cancer and is categorized as the third most prevalent malignancy within the urological domain, accompanied by the highest mortality rate among these cancers. Typically, individuals diagnosed with non-metastatic RCC are asymptomatic, which complicates early detection efforts. However, with appropriate therapeutic interventions, a complete remission is achievable for those with non-metastatic disease. Conversely, the prognosis for patients with metastatic RCC markedly decline [5]. Once clinical symptoms arise, the likelihood of attaining curative treatment diminishes significantly, often resulting in an expected survival duration that is reduced to months. Moreover, there is no standardized community screening protocol for the early identification of kidney cancer. Approximately 70% of kidney cancer cases are diagnosed incidentally during imaging studies, such as CT scans, frequently conducted for unrelated medical complaints. In contrast, about 10% of patients present at an advanced, incurable stage of disease, primarily seeking medical assistance

Manuscript received 2026. This work is supported in part by Health Institutes of Turkiye under Grant 33756.

S. Kilic is with Department of Urology, Antalya Training and Research Hospital, Antalya, Turkiye

B. Y. Kazangirler is with Department of Computer Engineering, Institute of Graduate Programs, Karabuk University, Karabuk, Turkiye

C. Ozcan is with Department of Software Engineering, Karabuk University, Karabuk, Turkiye

A. Keven is with Department of Radiology, Akdeniz University, Antalya Turkiye

A. S. Alparslan is with Department of Interventional Radiology, Antalya Training and Research Hospital, Antalya, Turkiye

due to symptoms attributable to metastatic kidney cancer, leading to diagnostic examinations at that juncture.

An analysis of the current literature showcases the effective use of different deep learning (DL) techniques for the 3D segmentation of medical images. Particularly, DL approaches that rely exclusively on convolutional neural networks (CNN) have been developed to differentiate and segment normal kidney tissue from masses in abdominal CT scans. These methodologies incorporate multi-channel functional layers, setting them apart from earlier techniques. Within the scope of the current work, the methodology intended for implementation in this study is based on 3D-CNN algorithms, specifically designed for processing 3D data structures [6], [7], [8].

The influence of artificial intelligence (AI) and machine learning (ML) on various aspects of the healthcare industry has been significant. Recent technological advancements have facilitated the analysis of vast data sets in a manner that is both cost-effective and efficient. Research and literature in clinical oncology highlight several advantages brought about by AI. In recent years, sectors such as engineering, communications, manufacturing, and healthcare have all experienced a significant impact from AI and ML [9]. The initial phases of this process can be outlined as follows: first, relevant features must be identified and selected from the dataset. These features may be quantitative and serve to clarify the information conveyed by the data into vectors or arrays. Subsequently, the information is incorporated into more comprehensive predictive models, such as classifiers or regressors, which are tailored to perform specific tasks [10].

II. Related Works

A. Deep Learning in Imaging

The influence of artificial intelligence (AI) and machine learning (ML) on various aspects of the healthcare industry has been significant. Recent technological advancements have facilitated the analysis of vast data sets in a manner that is both cost-effective and efficient. Research and literature in clinical oncology highlight several advantages brought about by AI. In recent years, sectors such as engineering, communications, manufacturing, and healthcare have all experienced a significant impact from AI and ML [9]. The initial phases of this process can be outlined as follows: first, relevant features must be identified and selected from the dataset. These features may be quantitative and serve to clarify the information conveyed by the data into vectors or arrays. Subsequently, the information is incorporated into more comprehensive predictive models, such as classifiers or regressors, which are tailored to perform specific tasks [10].

Recent studies emphasize the significant impact that DL techniques have made in the realm of medical imaging, particularly in enhancing the detection and segmentation of kidney masses. Initially, researchers focused on two-dimensional (2D) CNNs, which laid a foundational understanding of image analysis. However, attention has now shifted to more sophisticated and complex architectures that enable the accurate processing of 3D volumetric data. This progression not only enhances diagnostic accuracy but also fosters a more thorough comprehension of anatomical structures within the kidneys, ultimately leading to improved patient outcomes.

TABLE I: Segmentation and classification performance results on the literature with DL algorithms

Author (Year)	Objective	Approach	Dataset	Performance (%)
Patel et al. [11]	Kidney tumor segmentation	3D TRD U-Net++	KiTS21	Dice Score: 80.5 – 81.6
Uhm et al. [12]	Kidney cancer diagnosis	nnU-Net, ResNet-101	Custom & TCGA	AUC: 88.9, 85.5
He et al. [13]	Kidney tumor classification	3D ResNet-50	Custom	Accuracy: 90.5
Erdim et al. [14]	Kidney tumor classification	Random Forest	Custom	Accuracy: 90.5
Aronson et al. [15]	Kidney tumor segmentation	3D U-Net	Customs	Dice Score: 93.4 (Kidney), 71.1 (Tumor)
Zhao et al. [16]	Kidney tumor segmentation	3D MSS U-Net	KiTS19	Dice Score:96.9 (Kidney), 80.5 (Tumor)

As outlined in Table 1, various models, particularly the 3D U-Net ant the other algorithms, have shown considerable promise in the field of medical imaging; however, the accurate isolation of specific masses and cysts continues to pose a significant challenge. Current research emphasizes the need for supplementary processing steps to enhance diagnostic accuracy. Building on these insights, this study presents innovative advancements through the application of the nnU-Net framework. This advanced approach not only automates the fine-tuning of essential parameters but also improves overall diagnostic precision, facilitating more reliable identification of abnormalities in medical scans. In addition, Table 1 summarizes the focus areas of these studies and the performance outputs they achieved. Aronson et al. [15] performed an impressive accuracy of 93.4% in segmenting kidney tissue using the 3D U-Net architecture. However, their findings indicated a significant drop in performance for tumor segmentation, with an accuracy of just 71.1%. This discrepancy highlights the inherent complexity associated with the diverse morphologies of tumor structures, which greatly complicates the segmentation process. In a comparable study, Zhao et al. [16] achieved an impressive Dice score of 96.9% for kidney segmentation on the KiTS19 dataset by utilizing the advanced 3D MSS U-Net architecture. This remarkable performance underscores the model's effectiveness in accurately delineating kidney structures. However, it is important to note that the model's performance significantly diminished when applied to

tumor segmentation, with the Dice score decreasing to 80.5%. This contrast highlights the inherent challenges associated with accurately segmenting tumors in comparison to healthy kidney tissue.

Patel et al. [11] proposed a novel 3D Trans-Residual Dense U-Net++ (3D TRD U-Net++) network for kidney tumor segmentation, followed by a classification method based on adaptive and attentive residual densenet with gated recurrent unit (AA-RD-GRU). A unique aspect of their approach for segmenting and classifying kidney regions as benign or malignant was the optimization of network parameters using the modified crayfish optimization algorithm (MCOA). The segmentation accuracy was not explicitly stated, but the visual results indicate values ranging from 96% to 97% across the two datasets used. Uhm et al. utilized a diagnosis-focused approach by integrating the nnU-Net framework with ResNet-101 and spatial transformer networks (STNs), resulting in an AUC value of 88.9%. This approach underscores its significant potential for enhancing clinical diagnosis. In contrast, studies conducted by He et al. [13] utilizing 3D residual network-50 (ResNet-50) and Erdim et al. [14] applying Random Forest concentrated on tumor classification rather than segmentation, achieving an accuracy rate of 90.5% on custom datasets.

B. Prior segmentation methods (U-Net and nnUNet)

The U-Net model, originally developed by Ronneberger et al. [17], is predicated on convolutional neural networks and is tailored specifically for segmentation tasks in biomedical imaging applications. The 3D U-Net architecture represents an evolution of the original framework, adapted to address the complexities associated with segmenting 2D images, such as those acquired through CT scans. This adaptation underscores the relevance of CNNs in enhancing diagnostic precision in medical imaging. The architectural framework under examination is based on a sophisticated network structure characterized by multiple convolutional, sampling, and pooling layers. The initial phase involves the processing of 3D images, which are input into the network to undergo feature extraction via convolution and pooling operations. This step effectively transforms the images into feature maps with reduced dimensionality, thereby enhancing the capacity for salient feature extraction. Zhao et al. [16] demonstrated noteworthy advancements in tumor segmentation within renal tissues by employing a multi-scale approach that integrates the 3D U-Net architecture. Nevertheless, they encountered limitations reflected in lower performance metrics specifically pertaining to tumor segmentation in isolation. In a contrasting study, Pandey et al. [18] combined the 3D U-Net architecture with active contouring techniques, revealing that the integration of supplementary pre-processing or post-processing methodologies significantly improves the accuracy of DL algorithms in this domain.

This initial stage is commonly referred to as the encoder path. Subsequently, the integration of low-level and high-level features occurs through a connection path. This stage leverages convolutional and pooling layers to generate higher-resolution feature maps, utilizing the lowest resolution feature maps obtained from the encoder path as a foundational input [11]. The following phase, known as the decoder path, integrates the feature maps derived from both the encoder and connection paths to produce higher-resolution segmentation maps. Concurrently, a backward path is delineated, during which feature merging and resampling operations are executed to yield the segmentation map. This procedure is instrumental in recovering essential details that may have been lost during the initial feature extraction process. Collectively, the encoder, connection, and decoder paths operate synergistically to achieve a more detailed and comprehensive representation in the segmentation process, thereby enhancing the overall efficacy and accuracy of the segmentation outcomes [11, 12].

III. Materials and Methods

A. Dataset and Clinical Annotations

A comprehensive dataset comprising ***880 contrast-enhanced abdominal CT scans***, obtained from Antalya Training and Research Hospital and Akdeniz University Hospital, was curated with a focus on solid and cystic renal masses. This initiative aims to improve our understanding of these medical issues. Following the collection of this valuable data, a thorough and systematic data cleaning process was implemented to ensure the accuracy and reliability of the findings. The dataset utilized in this study has been meticulously assembled in accordance with rigorous quality standards and demonstrates homogeneous characteristics, so no additional preprocessing procedures were made. Previous research has indicated that excessive pre-processing of already clean and standardized datasets can result in a degradation of the original image information, ultimately leading to diminished model performance.

B. Data Annotations

Data labeling is a meticulous process that must be conducted by experts. Contrast-enhanced abdominal CT scan offers extensive information regarding the size, shape, location, and internal structure of the kidneys. The amount of data produced is contingent upon the technical specifications of the imaging device and the scanning procedure employed. Size measurements are typically performed using specifications related to slice and matrix sizes.

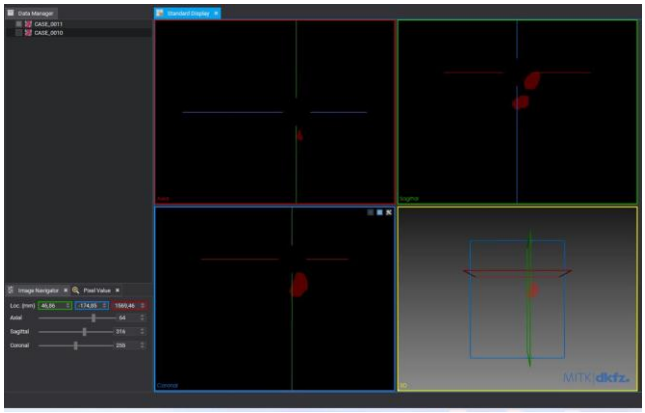

Fig. 1. Annotated sample CT scan on MITK labeling platform.

For 3D cross-sectional data, this involves classifying the cross-sections of 3D objects (such as 2D images) and assigning appropriate labels to each section. In this study, medical

imaging data was utilized, specifically focusing on the abdominal CT images. Each section of these images was carefully labeled by urology and radiology residents with at least three years of experience to identify normal kidney tissue and renal masses, and the labels were reviewed again by expert radiologists.

Kidney scans are captured in one or more slices, with each slice representing a cross-sectional image of the kidneys. The resolution of the data is defined on a pixel basis, which is influenced by the matrix size. To ensure high-quality imaging, high-resolution abdominal CT images was utilized. The CT images were initially obtained in the neuroimaging informatics technology initiative (NIfTI) format, which is ideal for 3D cross-sectional data. To ensure adherence to ethical standards during data usage, a series of anonymization procedures were implemented. The 3D data was segmented into slices, allowing the creation of 2D images. Each slice was evaluated and labeled independently. At this pivotal moment, it is essential to select an appropriate labeling tool for the labeling phase. The medical imaging interaction toolkit (MITK), a user-friendly, open-source image labeling software tool was used. MITK allows users to create annotated datasets and develop AI annotation models for clinical evaluation.

Fig. 1 illustrates the annotated CT scanned data, which comprises labels that delineate between normal renal tissue and renal mass. In addition, the labeled dataset is subsequently utilized within a specialized methodology for object segmentation, facilitating improved analysis and interpretation of the underlying anatomical structures and pathologies. The presented software facilitates both interactive and automatic segmentation of three-dimensional radiologic images through the provision of sample models. This specialized software enables the user to assign labels to 2D slices, thereby conferring distinct classification attributes. The label categories utilized for the identification of objects to be segmented within the dataset encompass "kidneys" and "kidney cysts". Such functionalities are instrumental in enhancing the accuracy and efficiency of image analysis in medical diagnostics.

The dataset is rigorously divided using k-fold cross-validation (CV), with an example showcasing five distinct folds. This method ensures a robust evaluation by allowing each fold to serve as a testing set while the remaining folds contribute to the training process. Prior to the training and validation phases, several essential pre-processing steps are implemented. These steps include normalization, which adjusts the data to a common scale; resampling, which helps to balance the data by adjusting the representation of different classes; and augmentation, which enhances the dataset by creating variations of the existing data. Together, these techniques prepare the dataset for a more effective and accurate modeling process.

C. Normalization and Resampling

The nnU-Net framework integrates preprocessing methodologies that are specifically designed to cater to the unique characteristics of distinct datasets. This tailored approach not only ensures data consistency but also enhances the optimization of model training processes. The implementation of these preprocessing steps plays a crucial role in influencing the generalizability and performance of the resulting models. Normalization is a critical process in medical imaging that standardizes the pixel or voxel values of images to a consistent range. This procedure is essential for mitigating variations in density that may occur due to discrepancies in scanning equipment, imaging protocols, or patient characteristics [12].

D. Model Architecture

CNN algorithms are inherently designed to conduct automatic feature extraction from raw image data. This end-to-end learning framework allows the model to develop its own pre-processing methodologies adaptively. Moreover, it is anticipated that the outputs generated from this study will be integrated into the evaluation process with minimal pre-processing applied to CT images sourced from actual clinical environments. Therefore, implementing excessive enhancement or artificial modification techniques may lead to discrepancies between the model outputs and the real-world data encountered in clinical practice.

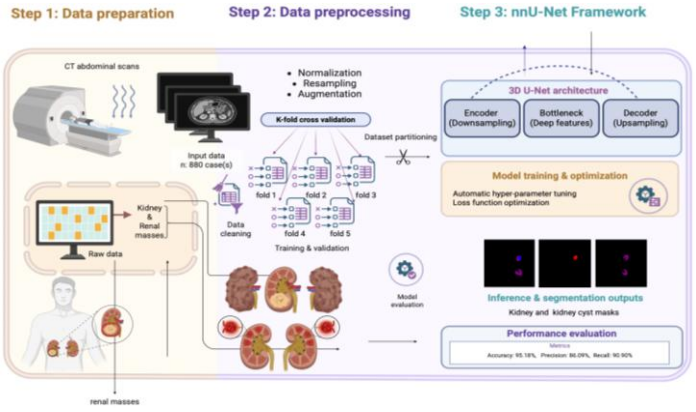

Fig. 2. Overview of the proposed 3D DL methodology for automated segmentation of kidneys and kidney cysts in CT abdominal scans.

In this regard, manual pre-processing interventions could impose limitations on the model's intrinsic learning capacity. At this stage, the collected dataset has been assessed and integrated into the normalization and pre-processing protocols as required by the algorithms.

The aim of this study is to attain a high level of accuracy in the segmentation of 3D images by employing the 3D U-Net [11] architecture, a prominent model within DL algorithms. Specifically, this research focuses on the kidneys as the ROI, which will be semantically distinguished from the surrounding background utilizing the no-new-Net (nn U-Net) [20] framework. As a result, ancillary structures, such as adjacent tissues and osseous elements, which may introduce noise or complexity, will be excluded from subsequent analyses, particularly during the segmentation processes aimed at identifying cysts and masses. The nnU-Net framework represents a significant advancement in automated DL methodologies, specifically designed to optimize architecture, pre-processing, training, and inference parameters tailored to the characteristics of a particular task or dataset. This automation effectively mitigates the need for manual hyper-parameter tuning, thereby not only conserving valuable time but also enhancing overall model performance. The

framework exemplifies a paradigm shift in the application of neural networks, facilitating improved outcomes in various medical imaging and segmentation tasks.

IV. Experiments and Results

A. Training Procedure

The training process for deep learning models was conducted on a workstation equipped with advanced hardware features to process 3D volumetric data, which requires high computational power. The hardware infrastructure, including an Intel (R) Core (TM) i9-14900KF processor and 64 GB RAM, demonstrates the robust setup supporting our work. An NVIDIA GeForce RTX 4090 GPU with 24 GB VRAM was used to accelerate training and handle large data blocks efficiently, ensuring reliable performance.

The dataset, containing a total of 880 cases for data partitioning and validation, was split into two to measure the model's generalization ability. Further, a 5-fold CV strategy was applied to prevent overfitting and ensure a robust evaluation. Before training, basic preprocessing steps, including intensity normalization, resampling, and data augmentation, were applied to the data to make the dataset most suitable for the model. Using the nnU-Net framework for hyperparameter optimization, critical hyperparameters such as learning rate, batch size, and network topology were automatically optimized according to the specific characteristics of the dataset. Throughout the training process, the loss function was optimized to improve segmentation performance. The model was trained using an end-to-end approach to learn the boundaries of renal masses and normal tissues as accurately as possible. Prior to training, basic preprocessing steps, including intensity normalization, resampling, and data augmentation, were applied to the data to make the dataset most suitable for the model [20] .

B. Performance Evaluation

The existing literature illustrates that various performance metrics are employed to assess the effectiveness of DL algorithms. Among the most recognized metrics are accuracy, precision, recall, F1-score, and dice score as in Equation 1, Equation 2, Equation 3, Equation 4, Equation 5 and Equation 6. In classification and segmentation tasks, accuracy is often regarded as a fundamental evaluation metric, defined as the ratio of correctly classified examples. However, relying solely on accuracy can be inadequate and misleading, particularly in scenarios involving imbalanced classifications [21]. To overcome this limitation, it is also crucial to consider other commonly used metrics for biomedical image segmentation. One such metric is the overlap ratio, which evaluates segmentation performance by measuring how closely the predicted mask corresponds to the true mask [22].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - score = 2\,x\,\frac{PrecisionxRecall}{Precision + Recall} \quad (5)$$

$$Dice\ score = \ 2\,x\,\frac{area\ of\ overlap}{total\ area} \quad (6)$$

C. Quantitative and Qualitative Results

A total of 702 cases were designated for training, representing 80% of the overall dataset, while 178 cases (20%) were set aside for testing. The training set is utilized for optimizing the model parameters throughout the learning process, whereas the test set is employed to evaluate the model's generalizability on previously unseen data. This configuration allows for a more realistic measurement of model performance, helping to avoid overfitting to the training data. Additionally, key parameters such as class labels, modality information, and data paths are established, thereby providing all necessary configurations for the nnU-Net training pipeline. Following the training, the 3D U-Net model achieved an impressive overall accuracy of 95.18%, as assessed through the evaluation of 178 patient cases. The total number of voxels, excluding the background, reached 30,106,797. This setup may enhance model performance, attributed to the increased number of cases and voxels.

Table 2 presented below offers a comparative analysis of the performance of the model across three primary segmented classes: kidney, solid renal masses (AML, RCC, oncocytoma) and simple cystic renal masses. Furthermore, it encompasses the overall mean performance metrics for each class. The evaluation employs standard metrics, including accuracy, precision, recall, and F1-score, which enhance the transparency of the assessment. This rigorous approach ensures that the audience gains confidence in the comprehensiveness and validity of the evaluation process. The table presents the metrics of accuracy, precision, recall, and the F1-score for each class, along with the aggregate mean, thereby substantiating the model's substantial capacity to differentiate between diverse lesion pathologies.

TABLE II: Segmentation performance metrics using 3D U-Net within the nnU-Net framework, showcasing the overall average results across the kidney and renal mass classes

Class type	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Kidney	91.28	92.36	95.58	93.94
Solid renal masses	97.06	88.54	78.70	81.88
Simple cystic renal masses	97.20	77.38	98.42	86.64
Average	95.18	86.09	90.90	87.49

Fig. 3 provides a visualization of the dice score (as in Equation 6) values utilized to evaluate expert labels against

actual predictions, emphasizing the significance of thorough assessment in segmentation studies of the developed 3D U-Net model. The dice score is a widely adopted metric in these studies, measuring the overlap ratio between predicted positive pixels and actual positive pixels. In this figure, the left column illustrates the ground-truth labels (GT), the middle column displays the model's predictions, and the right column features the overlay of both masks (red representing GT and blue denoting the predictions). This comparative analysis is fundamental in determining the degree of consonance between the outputs generated by the model and the GT labels meticulously curated by domain experts. Such an evaluation not only underlines the performance of model but also provides insights into its reliability for practical applications. The analysis of Fig. 3 (c) demonstrates that the model exhibits a significant degree of overlap between renal and cystic structures, underscoring a substantial congruence between these anatomical features. The delineation of boundaries appears to be precise, particularly in the context of cystic renal mass segmentation.

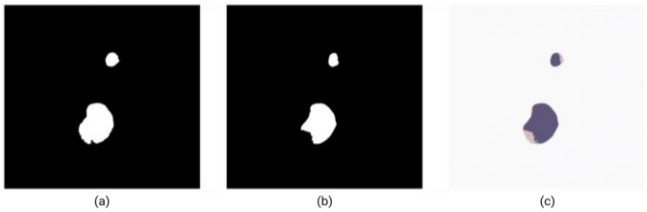

Fig. 3. Visual comparison of expert ground truth labels and the segmentation prediction of model in a representative case. (a) The expert-annotated segmentation mask representing the true boundaries, (b) The segmentation mask output.

Notably, while some minor instances of over-segmentation are observed in specific regions of the estimation mask, indicated by small areas of purple/blue, and under-segmentation is noted in other regions of the GT mask, represented by small areas of pink/red, the overall alignment in segmentation rates suggests that the model maintains a high level of reliability. Furthermore, the model effectively identifies the clinical significance of the lesions, as indicated by the recall metrics. The qualitative assessment aligns with the impressive dice coefficients (e.g., 0.919 for kidney and 0.924 for kidney cyst), thereby visually reinforcing the clinical applicability of the model.

The selected sample sections used to visually assess the model's performance are displayed in Fig. 4s. This visualization includes raw images (a), an overlay view with masks superimposed on the original anatomy (b), and isolated masks with the background noise removed (c). The color coding, organized according to the specifications of clinical experts, represents the following classes:

- Green corresponds to normal renal tissue, indicating healthy and functioning kidney structures.
- Blue corresponds to cystic renal masses, delineating the visual margins between these lesions and the surrounding healthy tissue.
- Red corresponds to solid renal masses, such as AML, facilitating their clear identification and analysis in diagnostic imaging.

The background removing process illustrated in Fig 4 (c) removes the complexities posed by surrounding tissues and bone structures, enabling a concentrated focus on the relevant lesions. Visual analyses reveal the model's recall in accurately defining lesion boundaries and its capability to effectively differentiate between various pathological tissues (cystic versus solid). These qualitative findings support the average accuracy rate of 95.18% obtained and visually confirm the system's applicability in clinical decision support processes.

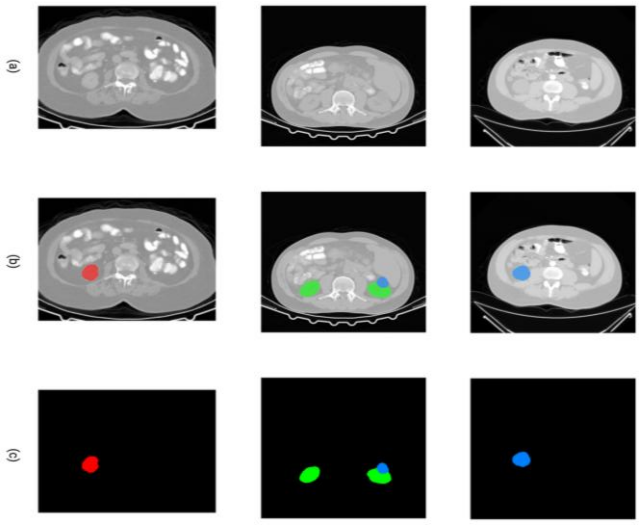

Fig. 4. Qualitative assessment of the segmentation model, (a) original CT images, (b) overlay of predicted segmentation masks, and (c) isolated segmentation masks with background removal.

As demonstrated in Table 1, Aronson et al. [15] reported a Dice score of 93.4% for kidney tissues, while their score for tumors was noticeably lower at 71.1%. Similarly, Zhao et al. [16] achieved a score of 96.9% for kidneys, but only 80.5% for tumors. In contrast, our model attained an impressive accuracy of 97.06% for solid masses and 97.20% for cystic masses, showcasing consistently high performance across various complex pathological categories. A significant strength of this research is its use of a large, curated dataset comprising 880 contrast-enhanced CT scans, significantly surpassing the custom datasets typical of many previous studies. By emphasizing the extensive nature of this dataset, we aim to foster confidence in the audience regarding the model's robustness and reliability across various scenarios. The model exhibits an impressive recall rate of 98.42% for cystic masses, demonstrating the nnU-Net's proficiency in detecting fluid-filled lesions, which are often discovered incidentally. However, a notable limitation is apparent in the recall rate for solid renal masses (78.70%) and the precision rate for cystic renal masses (77.38%). This indicates that while the model seldom overlooks a cyst, it may occasionally over-segment cystic regions or under-segment solid neoplasms, likely due to the overlapping intensity profiles encountered in more complex cases.

V. Conclusion

Building upon the quantitative findings, the qualitative success illustrated in Fig. 4 provides medical professionals with

confidence that the model's outputs consistently align with the anatomical boundaries established by expert radiologists. By accurately distinguishing between solid and cystic masses, the system aids urologists and radiologists in their decision-making processes. The capability to isolate these structures through background removal enhances trust and confidence, allowing for more precise volumetric analysis, which is essential for monitoring lesion growth and planning surgical interventions.

This study presents a methodology grounded in a 3D U-Net architecture, implemented within the nnU-Net framework. The aim of this approach is to enable the automatic detection of normal kidney tissues, solid renal masses, and cystic renal mass lesions in abdominal CT scans. An analysis conducted on a comprehensive and carefully curated dataset, consisting of 880 cases, yielded an average accuracy of 95.18%, a precision of 86.09%, and a recall of 90.90%. These results underscore the clinical significance of the 3D U-Net approach, intended to inspire both healthcare professionals and AI researchers by showcasing its potential for real-world applications. Upon analyzing the model's performance by class, it is evident that the exceptionally high recall rate of 98.42% for cystic renal masses highlights the model's remarkable capability to detect fluid-filled lesions, which should instill confidence among researchers regarding its strengths. Conversely, the lower recall rate for solid renal masses (78.70%) and the reduced specificity rate for renal cystic masses (77.38%) indicate an imbalance in the results. This discrepancy is believed to arise from the complex morphological structures of the masses and the similarities in their tissue densities. To address this imbalance and enhance outcomes in future studies, targeted data augmentation techniques will be implemented, alongside the optimization of loss functions to minimize class imbalances.

ACKNOWLEDGMENT

This study is funded by the Health Institutes of Türkiye under project number 33756.

REFERENCES

[1] R. Dalal, Z. S. Bruss, and J. S. Sehdev, "Physiology, Renal Blood Flow and Filtration," in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2025. Accessed: Dec. 31, 2025. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK482248/

[2] K.-H. Uhm *et al.*, "Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography," *NPJ Precis. Oncol.*, vol. 5, no. 1, p. 54, 2021. https://doi.org/10.1038/s41698-021-00195-y

[3] B. Ljungberg *et al.*, "European Association of Urology guidelines on renal cell carcinoma: the 2022 update," *Eur. Urol.*, vol. 82, no. 4, pp. 399–410, 2022. https://doi.org/10.1016/j.eururo.2022.03.006

[4] S. S. Parvathi and H. Jonnadula, "An efficient and optimal deep learning architecture using custom U-net and mask R-CNN models for kidney tumor semantic segmentation," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, 2022, Accessed: Dec. 04, 2025. [Online]. Available: https://search.proquest.com/openview/ede241576880bbbd936e7ee0de28883c/1?pq-origsite=gscholar&cbl=5444811 https://doi.org/10.14569/IJACSA.2022.0130639

[5] H. Zhang *et al.*, "Development and Validation of a CT-Based Radiomics Nomogram for Predicting Postoperative Progression-Free Survival in Stage I–III Renal Cell Carcinoma," *Front. Oncol.*, vol. 11, p. 742547, 2022. https://doi.org/10.3389/fonc.2021.742547

[6] J. Liu, O. Yildirim, O. Akin, and Y. Tian, "AI-Driven Robust Kidney and Renal Mass Segmentation and Classification on 3D CT Images," *Bioengineering*, vol. 10, no. 1, p. 116, Jan. 2023, doi: 10.3390/bioengineering10010116. https://doi.org/10.3390/bioengineering10010116

[7] N. Bouteldja *et al.*, "Deep Learning–Based Segmentation and Quantification in Experimental Kidney Histopathology," *J. Am. Soc. Nephrol.*, vol. 32, no. 1, p. 52, Jan. 2021, doi: 10.1681/ASN.2020050597. https://doi.org/10.1681/ASN.2020050597

[8] L. Kang, Z. Zhou, J. Huang, and W. Han, "Renal tumors segmentation in abdomen CT Images using 3D-CNN and ConvLSTM," *Biomed. Signal Process. Control*, vol. 72, p. 103334, Feb. 2022, doi: 10.1016/j.bspc.2021.103334. https://doi.org/10.1016/j.bspc.2021.103334

[9] Z. Dlamini, F. Z. Francies, R. Hull, and R. Marima, "Artificial intelligence (AI) and big data in cancer and precision oncology," *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 2300–2311, 2020. https://doi.org/10.1016/j.csbj.2020.08.019

[10] A. Barragán-Montero *et al.*, "Artificial intelligence and machine learning for medical imaging: A technology review," *Phys. Med.*, vol. 83, pp. 242–256, 2021. https://doi.org/10.1016/j.ejmp.2021.04.016

[11] V. V. Patel, A. R. Yadav, P. Jain, and L. R. Cenkeramaddi, "A systematic kidney tumour segmentation and classification framework using adaptive and attentive-based deep learning networks with improved crayfish optimization algorithm," *IEEE Access*, vol. 12, pp. 85635–85660, 2024. https://doi.org/10.1109/ACCESS.2024.3410833

[12] K.-H. Uhm *et al.*, "Deep learning for end-to-end kidney cancer diagnosis on multi-phase abdominal computed tomography," *NPJ Precis. Oncol.*, vol. 5, no. 1, p. 54, 2021. https://doi.org/10.1038/s41698-021-00195-y

[13] Q.-H. He, J.-J. Feng, F.-J. Lv, Q. Jiang, and M.-Z. Xiao, "Deep learning and radiomic feature-based blending ensemble classifier for malignancy risk prediction in cystic renal lesions," *Insights Imaging*, vol. 14, no. 1, p. 6, Jan. 2023, doi: 10.1186/s13244-022-01349-7. https://doi.org/10.1186/s13244-022-01349-7

[14] C. Erdim *et al.*, "Prediction of benign and malignant solid renal masses: machine learning-based CT texture analysis," *Acad. Radiol.*, vol. 27, no. 10, pp. 1422–1429, 2020. https://doi.org/10.1016/j.acra.2019.12.015

[15] L. Aronson, R. N. Massaa, S. J. S. Gardezi, and A. L. Wentland, "Automatic Segmentation of the Kidneys and Cystic Renal Lesions on Non-Contrast CT Using a Convolutional Neural Network," May 14, 2024, *arXiv*: arXiv:2405.08282. doi: 10.48550/arXiv.2405.08282.

[16] W. Zhao, D. Jiang, J. P. Queralta, and T. Westerlund, "Multi-Scale Supervised 3D U-Net for Kidneys and Kidney Tumor Segmentation," *Inform. Med. Unlocked*, vol. 19, p. 100357, 2020, doi: 10.1016/j.imu.2020.100357. https://doi.org/10.1016/j.acra.2019.12.015

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science, vol. 9351. , Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28. https://doi.org/10.1007/978-3-319-24574-4_28

[18] M. Pandey and A. Gupta, "Tumorous kidney segmentation in abdominal CT images using active contour and 3D-UNet," *Ir. J. Med. Sci. 1971 -*, vol. 192, no. 3, pp. 1401–1409, June 2023, doi: 10.1007/s11845-022-03113-8. https://doi.org/10.1007/s11845-022-03113-8

[19] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, vol. 9901, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds., in Lecture Notes in Computer Science, vol. 9901. , Cham: Springer International Publishing, 2016, pp. 424–432. doi: 10.1007/978-3-319-46723-8_49. https://doi.org/10.1007/978-3-319-46723-8_49

[20] F. Isensee *et al.*, "nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation," Sept. 27, 2018, *arXiv*: arXiv:1809.10486. doi: 10.48550/arXiv.1809.10486. https://doi.org/10.1007/978-3-658-25326-4_7

[21] B. Y. Tekin, C. Ozcan, A. Pekince, and Y. Yasa, “An enhanced tooth segmentation and numbering according to FDI notation in bitewing radiographs,” *Comput. Biol. Med.*, vol. 146, p. 105547, 2022. https://doi.org/10.1016/j.compbiomed.2022.105547

[22] C.-H. Hsiao *et al.*, “A deep learning-based precision volume calculation approach for kidney and tumor segmentation on computed tomography images,” *Comput. Methods Programs Biomed.*, vol. 221, p. 106861, 2022. https://doi.org/10.1016/j.cmpb.2022.106861