2025-04

Detection-guided kidney tumor segmentation in 3d from abdominal computed tomography images

Faruk, Joyita

IUB

DETECTION-GUIDED KIDNEY TUMOR SEGMENTATION IN 3D FROM ABDOMINAL COMPUTED TOMOGRAPHY IMAGES

Joyita Faruk, ID: 1911041

SK Sadia Tasnim Elma, ID: 1930921

Supervised by

Dr. Saadia Binte Alam

Associate Professor and Head

Department of Computer Science and Engineering

Attestation

I understand the nature of plagiarism, and I am aware of the university's strict policy on the matter. I certify that this is an original work by me. However, others' written work or software used in any part of this project is properly cited following internationally accepted academic guideline.

Signature:

I. Joyita Faruk: _____

II. SK Sadia Tasnim Elma: _____

Evaluation Committee

Supervision Panel

……	……
Supervisor	Co-supervisor

Examiners

……	……
Examiner 1	Examiner 2
……	
Examiner 3	

Office Use

……	……
Program Coordinator	Head of the Department

…….

Director Graduate Studies, Research and Industry Relations

……

Library

Acknowledgment

First and foremost, we are deeply grateful to our supervisor, Dr. Saadia Binte Alam, Associate Professor and the Head of the Computer Science and Engineering Department of Independent University, Bangladesh, for her invaluable guidance throughout the course of this research. Without her, this study would not have been possible.

We would also like to extend our sincere gratitude to Dr. Rashedur Rahman Sir, UoH fellow from the University of Hyogo, Japan, and an Associate Professor at Independent University, Bangladesh, who has supervised us from the beginning, equipping us with the knowledge to embark on this study. His expertise and insightful advice have guided us through critical stages of our research.

Furthermore, we would like to thank Sefatul Wasi, R&D Officer, at Independent University, Bangladesh, for aiding us with designing our research and improving our work with his constant support.

Lastly, we are grateful to Siam Tahsin Bhuiyan, for his insightful technical suggestions related to our code.

List of Publications:

[1] **J. Faruk**, S. B. Alam, **SK S. T. Elma**, S. Wasi, R. Rahman, and S. Kobashi, "*Kidney Abnormality Detection Using Segmentation-Guided Classification on Computed Tomography Images*", accepted at the International Conference on Machine Learning and Cybernetics (ICMLC), 2024; accepted on July 29, 2024; presented on September 22, 2024

[2] **J. Faruk**, S. B. Alam, **SK S. T. Elma**, S. Wasi, R. Rahman, and S. Kobashi, "*Kidney Abnormal Region Localization on 3D CT Images Using Detection-guided Segmentation*", in progress for a journal

Abstract

Kidney tumors are one of the predominant causes of kidney cancer, posing a significant threat to human health. Early detection plays a crucial role in improving treatment outcomes for patients. Radiology imaging techniques such as ultrasound, MRI, CT, and X-ray are employed for kidney tumor detection. However, specialists have to manually locate the tumors and cysts from CT scans and with the high incidence rate of kidney cancers, this time-consuming approach puts pressure, leading to increased medical errors. To improve the diagnostic procedure, this thesis study proposes an automated detection-guided segmentation method. The two-step approach uses the YOLOv8 object detection model to detect the kidney region and the U-Net architecture to segment the abnormal region composed of tumor, cyst, or both. This method is evaluated in a 3D subject-wise evaluation technique for abnormal region segmentation. Four comparative methods are studied. Among these four, two are baseline methods, baseline U-Net and baseline YOLO, which are compared with the proposed method to show that guided frameworks perform better than single task frameworks. Two segmentation-guided methods, segmentation-guided classification and segmentation-guided segmentation, are compared to show how proposed method outperforms in the more relevant metric of recall, making it more suited to be a diagnostic method for detecting kidney abnormal regions. The results show that for all IoU thresholds ranging from 10% to 50%, the average recall for the proposed method is 1.000 across all thresholds. When comparing the approaches of the proposed method and the comparative methods, the proposed detection-guided segmentation method proves to be the most suitable as an automated method for kidney tumor diagnosis.

Table of Contents

List of Figures

List of Tables

1. Introduction

Kidneys are a pair of symmetrical organs found in the back of the abdomen in the human body, and are indispensable to the functioning of the human body. They are responsible for filtering out toxins, waste, and excess substances such as water and electrolytes from the blood, forming urine. Kidneys help the human body regulate blood pressure by controlling the amount of water and minerals. They are also responsible for production of the hormone erythropoietin, which stimulates the bone marrow to create red blood cells, which are responsible for transporting oxygen to the body. Hence, kidney diseases and abnormalities in the kidney can be highly detrimental to renal function and health of a person. Two major abnormalities of kidney cancer are tumors and cysts. Tumors are the abnormal growth of cells into a mass of tissue. They can be benign or malignant, the latter of which can lead to cancer. Another structure called cysts, which are fluid-filled sacs, can develop in or around the kidney. Cysts may also be benign or cancerous.

1.1. Background and Context

As per 2022 GLOBOCAN estimates of incidence and fatality rate worldwide for 36 cancers in 185 countries, kidney cancer has the 14th highest incidence rate, with an estimated 434,419 new cases and 155,702 deaths [1]. Furthermore, according to the American Cancer Society's research on US cancer statistics, the incidence of kidney cancers has continued to increase per year by 1.5% [2]. Renal cell carcinoma, the most common type of kidney cancer, alone makes up about 2% of all cancer diagnoses and deaths worldwide [3]. The spreading of cancer from the area it primitively formed in is known as metastasis. Among 1/3rd of kidney cancer cases, the cancer will already have spread to other organs and metastasized by the time of diagnosis [4]. The five-year survival rate of localized cancer, i.e. within the kidney, is 93%, and regional cancer, i.e. surrounding and near the kidney, is 72.5%, whereas metastasized kidney cancer, i.e. the cancer has outspread to distant organs, has a very low five-year survival rate of 12% [5].

The main method of detecting kidney tumors is through radiology imaging, such as ultrasound (US), magnetic resonance imaging (MRI), and computed tomography (CT). Typically, computed tomography (CT), is widely used for detecting kidney diseases [6]. Both types of kidney tumor

masses, cystic or solid, are best detected by contrast-enhanced, triple-phase CT [7]. CT images are 3-dimensional (3D) images taken using X-ray techniques that provide higher dimensional detail than 2-dimensional (2D) images to medical specialists. In the current medical field, the kidney tumor region is detected manually by specialists, although a universally accepted screening strategy has yet to be established [8]. This manual process is time-consuming, relying on the radiologists and nephrologist's expertise to determine tumors or cysts [9]. The ratio of practicing nephrologists, kidney specialists, to the number of cancer patients pose additional pressure. In the US, the number of practicing nephrologists was estimated to be around 10,370 to 12,939 [10]. Compared to that, an estimated 628,355 people were living with kidney and renal pelvis cancer in the US in 2020 [11].

The time-consuming method of detecting kidney tumors and the high prevalence of kidney cancer causes increased medical errors. Nearly 1/4th of nephrologists in a sample size of 457 from the US reported burnout, which is characterized by chronic fatigue [12], [13]. 55.9% of responders in a countrywide survey by the Polish Society of Nephrology stated that work overload leading to rush was the most aggravating factor [14]. Burnout has been linked to declining empathy, increased medical errors, and poorer patient outcomes among physicians, particularly in fields like oncology, which specializes in cancer treatment [15]. Therefore, an automated process would be beneficial, to improve the diagnostic process and reduce the rate of misdiagnosis. Early diagnosis can monumentally improve treatment options for kidney tumor patients and improve mortality rates. Computer-aided diagnostic (CAD) methods are the popular approach to automate diagnoses. Specifically, CAD methods employing deep learning, a subset of machine learning that utilizes deep neural networks, have shown promise.

1.2. Objective

The objective of this study is to improve kidney abnormality detection and localization on 3D CT images of the normal kidney areas and abnormal kidney areas that are tumorous or cystic, through the development of a fully automated method.

1.3. Scope and Limitations

Automated methods for kidney tumor segmentation have the potential to revolutionize medical imaging analysis, provided that existing limitations are appropriately addressed. Manual segmentation is time-consuming and requires specialist knowledge, so the primary challenge is collecting large amounts of annotated data. In this study, the KiTS21 dataset from the 2021 Kidney Tumor Segmentation Challenge, is utilized [16]. It contains 300 anonymized cases of kidney CT data collected between 2010 and 2020. The KiTS21 dataset does not contain any 3D CT volume of only healthy kidneys, as at least one kidney has an abnormality in each case. The KiTS21 dataset is sufficient for the transfer learning technique used for deep learning models, which this study utilizes.

This study explores guided frameworks for kidney tumor segmentation, addressing the limitations of standalone imaging tasks: classification, detection, and segmentation. State-of-the-art models from current literature are utilized for transfer learning and a detection-guided segmentation method is proposed. The proposed method is compared to four comparative methods, two baseline methods and two segmentation-guided frameworks. Axial slice images are utilized for model training. With the proposed method, this study aims develop an automated kidney tumor detection method that improves the current diagnostic process.

1.4. Contributions

In this study, a novel deep learning-based approach for kidney abnormality detection and localization in 3D CT images is proposed. Furthermore, the proposed method is compared against four other comparative methods. The method proposed is a detection-guided segmentation technique, utilizing the KiTS21 dataset [16], the successor challenge to KiTS19 [17]. The detection-guided segmentation approach introduces a novel two-step pipeline combining YOLOv8 [18] for automatic kidney detection and U-Net [19] for abnormal region segmentation. This approach automates the preprocessing steps by employing detection-guided cropping, making it a fully automated and efficient solution. Another significant contribution of this paper is the evaluation carried out for the detection-guided segmentation approach, in which predictions in the form of segmented 2D slices are reconstructed into 3D CT images, allowing for direct comparison

with ground truth 3D masks. The criteria used for reconstruction in the 3D evaluation framework will guide future research in 3D CT image analysis.

1.5. Outline of Thesis

This research is divided into 9 parts. In the first part, the introduction of this thesis is outlined, including the background and context, the objectives, the scope and limitations, and the contributions. In the second part, the literature review conducted is documented. In the third part, the dataset description is given for KiTS21. In the fourth part, the methodology is discussed, in two main subsections, proposed method and comparative methods. In the fifth part, the evaluation methods are explained. In the sixth part, the results are outlined and discussed, in relevant metrics for each method. In the seventh part, a conclusion summarizes the study, discussing the future directions.

2. Literature Review

In the current literature, three notable computer vision tasks are employed for creating CAD methods for kidney tumor diagnosis: classification, detection and segmentation. Each task has its advantages and limitations. Classification refers to classifying an image as a class, such as 'kidney' or 'abnormal', indicating the presence of kidneys or abnormalities. Detection refers to the localized detection of an object, such as a kidney or an abnormality like a tumor or cyst in the image. Typically, the object is localized by bounding boxes. Segmentation refers to the division of an image into the segments of image regions such as kidneys or abnormalities like tumors and cysts, rendering a pixel-wise mask which is more precise localization than detection.

Before the advent of deep learning, several studies have achieved automation through traditional machine learning techniques. A study utilizing the region-growing method, in which pixels from a seed location attempt to merge neighboring pixels, yielded 85% accuracy [20]. A fuzzy C-Means clustering algorithm has also been used to segment kidney tumors in another study [21]. Aslam et. al combined the Sobel method with the image-dependent thresholding method for brain tumor segmentation [22]. The Sobel operator uses convolution to compute gradient magnitudes and detects edges from differences in intensity. Rudra et. al proposed a kidney segmentation algorithm using graph cuts and pixel connectivity for low-contrast MRI data, obtaining a mean Dice coefficient value of $98.60 \pm 0.50\%$ on 25 datasets [23].

However, these image-processing methods are superseded by machine learning algorithms, which are capable of scaling along with and learning directly from the data fed into them. Various supervised and unsupervised machine learning algorithms have been used for segmenting, detecting, and classifying kidney and kidney tumors. Jagga used supervised learning algorithms, J48, Random Forest, SMO, and Naïve Bayes, to classify the stages of renal cell carcinoma (RCC) based on gene expression profiles [24]. The highest scores achieved were 89%, 77%, and 0.8 for sensitivity, accuracy, and an area under receiver operating characteristic (ROC) curve, respectively. Li et. al proposed a method for finding an automatic contour of the kidney based on a support vector machine (SVM), which is a machine learning algorithm used for classification

[25]. The study used an MR data set of eight patients' lower abdomen and the minimum value for the Dice similarity index for the study is approximately 0.9423. Tuncer et. al proposed a decision support system for detecting renal cell cancer using 130 abdominal images of healthy and renal cell cancer tissues from the image archiving system of the Radiodiagnostic Department of Fırat University Medical Faculty [26]. They used the K-Means algorithm to segment the kidney area and then used SVM for classification. The Dice coefficient they achieved for segmentation was 89.3%. Classification of the stage of clear cell renal cell carcinoma (ccRCC) was conducted by Talaat et. al [27]. It is a semi-automated model trained on the TCGA-KRIC dataset, which contains 228 CT scans of ccRCC patients. The watershed algorithm was used to extract tumors from the kidney. Extraction of the feature vectors from the segmented images was conducted through the gray-level co-occurrence matrix (GLCM) method. Random forest (RF) and SVM classifiers were used to classify the stages, yielding overall accuracies of 98.25% and 99.12% respectively.

Recent advancements in the field of machine learning have resulted in the revolutionary development of deep learning, which is capable of direct feature extraction, and eliminates the need for manual feature engineering. Through training a large number of layers of artificial neural networks, deep learning has enhanced the capabilities of computer vision and medical imaging [28]. As medical images feature complex information, deep learning is the leading choice for automatic CAD methods using medical imaging [29]. Currently, deep learning has been on level with medical experts in diagnosing conditions from medical imaging [30]. Particularly, convolutional neural networks (CNN) are highly effective in the field of radiology imaging [31] and are used in the majority of medical image diagnostic methods [32] as they are exceptionally capable of handling image and video data.

However, for effective model training, deep learning requires a substantial amount of data. Acquiring such a number of medical images that have been annotated by medical professionals is a difficult process. Transfer learning technique is often used to overcome this lack of data [32],[33]. Transfer learning involves taking a pre-trained deep learning model, initially trained on a comprehensive dataset called ImageNet for a classification task, and then fine-tuning its final layers using a dataset specific to the task at hand. When fine-tuned with a new dataset, the model is capable of expanding on the previous knowledge base and learning the relevant features from the new dataset. It allows for reduced training time and is advantageous in cases where the dataset

is sufficiently large to train a deep CNN.

Deep learning techniques have accelerated research for medical imaging on various anatomical structures other than kidneys, such as the work on skin lesion diagnosis by Lopez-Labraca et al. [34], and the study by Qian et al. on pituitary tumor detection using MRIs [35]. Dabiri et. al presented an algorithm for automated localization and segmentation of skeletal muscle, subcutaneous adipose tissue, visceral adipose tissue, and intermuscular adipose tissue in CT scan volumes [36]. The algorithm includes a localization network and a segmentation network, both trained on large datasets, achieving high accuracy in slice localization and tissue segmentation. It achieved a mean slice error of 0.87±2.54 for L3 slice localization and mean Jaccard scores of 97% for SM and VAT segmentation and 98% for SAT segmentation. The algorithm's performance is evaluated on three datasets, demonstrating its potential for fully automated body composition analysis with high accuracy. A paper utilizing the LiTS-2017 dataset for the Liver Tumor Segmentation Challenge proposed a 3D deep learning semantic segmentation model, the 3D-DenseUNet-569 [37]. The model combines the advantages of dense connections for improved information flow and the UNet links between the encoding and decoding layers for improved resolution flow. The model received a Dice of 96.2% and a Dice global score of 96.7% for liver segmentation and a Dice of 69.6% and a Dice global score of 80.7% for tumor segmentation.

In 2019, the Kidney Tumor Segmentation (KiTS19) grand challenge was introduced, inviting pioneering researchers to compete and find the most suitable solutions by incorporating various methods of deep learning. It featured a dataset called the KiTS19 and was the first in a challenge series [17]. Subsequent challenge in the series with the respective dataset is KiTS21 [16]. The KiTS challenge prompted several studies. Kim et. al proposed a method for kidney segmentation utilizing active learning and a cascaded 3D U-Net architecture for coarse region detection [38]. They enhanced accuracy by manually correcting preliminary segmentation and called it ‘convolutional neural network (CNN)-corrected segmentation’, which proved to be more effective when compared to manual segmentation. Hsiao et. al proposed an encoder-decoder architecture for kidney segmentation, using EfficientNet-B5 as the encoder and a feature pyramid network as the decoder [39]. The experimental results showed that configuration parameters such as window range, loss function, and data augmentation significantly influenced the model's performance. It achieved a Dice score of 0.969 on the KiTS19 dataset. The paper proposed by Kittipongdaja and

Siriborvornratanakul focuses on automatic kidney segmentation using 2.5D ResUNet and 2.5D DenseUNet for analyzing the malignant potential of complex renal cysts based on CT images [40]. The Bosniak classification is used to distinguish between simple and complex cysts. Although the model was trained on KiTS19, the test data came from a retrospective cross-sectional study of 200 patients which was provided by the Department of Radiology and Department of Pathology, Faculty of Medicine, Ramathibodi Hospital, Bangkok, Thailand. 2.5D DenseUNet yielded higher Dice scores than 2.5D ResUNet, with the highest scores for KiTS train, KiTS val, and Thai patients being 0.9801, 0.9595, and 0.8760 respectively. Causey et. al used an ensemble of U-Net models. One of them, the "K/T" model predicts the kidney and tumor masks separately in two output channels [41]. The other "KT/T" model predicts both kidney and tumor in one channel and only the tumor region in the second channel. This allowed for reducing errors and a post-processing phase refined the resulting masks. The model attained test earned Dice scores of 0.9470 for kidney and tumor segmentation and 0.6099 for tumor segmentation. However, these scores are for the axial slices, as the coronal and sagittal planes did not perform as well. Li et al. used a radiomics-based approach, to automatically classify kidney tumors and normal kidney tissues on the KiTS19 dataset [42]. In total, 837 radiomics features were extracted and 217 were identified after univariate screening. From these, 3 features were selected to establish the binary logistic regression model. The area under the receiver operating characteristic (AUROC) curve for the train set and test set were 0.9798 and 0.9841, respectively. A multi-scale supervised 3D U-Net called the MSS U-Net was developed by Zhao et al. utilizing the KiTS19 dataset [43]. They combined deep supervision with exponential logarithmic loss to increase the 3D U-Net training efficiency, achieving a Dice coefficient of 0.969 for the kidney and 0.805 for the tumor. Rao et al. proposed a weight pruning (WP)-UNet model, which utilized the KiTS19 dataset [44]. It was designed to be a lightweight deep learning model, that involves few parameters, a quick assumption time, and a low floating-point computational complexity. For the kidney tumor region, they obtained a dice score of 0.9799 for the training set and 0.9599 for the validation set. Wasi et. al used an active contour algorithm to reduce the background, achieving a recall and precision of 0.9302 and 0.8045, respectively [4 5].

Outside of the KiTS challenge, significant research has been conducted with independent datasets. A multi-branch feature-sharing generative adversarial network (MB-FSGAN) was presented by Ruan et. al for the simultaneous segmentation and quantification of kidney tumors on CT images [46]. On a dataset of 113 manually annotated kidney tumor patients, the method achieved a pixel

accuracy of 95.7% and high quantification performance with the R2 coefficient of tumor diameter being 0.904. A deep learning model for kidney tumor segmentation based on the FR2PAttU-Net architecture was proposed by Sun et. al [47]. The model employed a cascaded network approach, where the first model coarsely segmented the kidney and tumor regions, followed by a second model that refined the segmentation specifically for the tumor in CT images. After evaluation on a dataset of 210 scans, the resulting kidney dice coefficient was 0.948, the tumor dice coefficient was 0.911, and the composite score was 0.930. Song et. al proposed a cross-modal transfer learning approach for kidney segmentation [48]. Their method utilizes a cycle generative adversarial network (CycleGAN) to generate ultrasound images from CT data. Mainstream CNNs such as U-Net, U-Res, PSPNet, and DeepLab v3+ are trained on the transition dataset and then transferred to real ultrasound images. They achieved Dice similarity coefficient reaching 0.853 for U-Net, 0.850 for U-Res, 0.826 for PSPNet, and 0.827 for DeepLab v3+ on the cross-site test set, thereby showing the effectiveness of transfer learning, particularly when training data is limited. Alzu'bi et. al conducted a study using 2D CNNs for a novel dataset from the King Abdullah University Hospital (KAUH) consisting of 8,400 images of 120 adult patients [49]. A six-layer 2D CNN (CNN-6), a ResNet50 with 50 layers, and a VGG16 with 16 layers were used as detection models, and a four-layer 2D CNN (CNN-4) was used for classification. Kidney tumor detection and multi-classification for 4 labels were done in two phases. In the first phase, normal kidney cases and tumor cases were classified. In the second phase, the detected tumors were further classified into types of benign tumors and malignant tumors. They arrived at accuracy results of 97% for CNN-6, 96% for ResNet50, 60% for VGG16, and 92% for CNN-4. Bolocan et al. conducted a study for the segmentation and classification of RCC, to distinguish malignant and benign tissues and identify the likely subtype [50]. The dataset consists of four-phase contrast-enhanced CT - pre-contrast, arterial, venous, and excretory phases. They used the European Deep-Health toolkit's Convolutional Neural Network. The mean Dice score for kidney segmentation was 0.84, and tumor segmentation was 0.675, while the RCC classification accuracy was 0.885.

Each computer vision task comes with limitations. Classification cannot localize the kidney or abnormality. With detection, there is bounding box ambiguity, which means non-target pixels may appear within the bounding box. Segmentation faces difficulties in identifying small structures in large backgrounds. A guided framework integrating two tasks, where one task guides the other, is necessary for result improvement and for the development of a fully automated CAD system.

This paper proposes a novel 2-step detection-guided segmentation method. The study utilizes transfer learning of two CNN models for automatic abnormal kidney region localization on 3D CT images in a detection-guided segmentation technique. Firstly, the fast and efficient real-time detection of YOLOv8 is used for automatic kidney detection on 2D PNG axial slice images generated from 3D CT images from the KiTS21 dataset. The predicted kidney bounding boxes are used to crop to the kidney region, which is the detection-guided cropping technique. Then, abnormal region segmentation is conducted on the cropped kidney 2D PNG axial slice images by the U-Net architecture, which excels at image segmentation tasks. The study focuses on axial slice images for two reasons. Firstly, axial slice images are preferred by medical specialists as they provide a clear, cross-sectional view of the human body. Secondly, this will allow for more relevant contextual comparison with other studies as most studies in the current literature utilize axial slice images as well.

The evaluation is carried out subject-wise by reconstructing the abnormal region mask as a 3D CT from the segmented abnormal region 2D PNG axial slice images. This allows for direct 3D comparison of the ground truth 3D CT abnormal region to the predictions. Following evaluation, the proposed method is compared with four comparative methods. Two baseline methods for YOLOv8 and U-Net respectively are compared to show that the detection-guided cropping technique is a significant improvement from direct methods. Two segmentation-guided methods are compared with the proposed method to showcase the advantages of detection-guided segmentation over segmentation-guided methods. The results show that for all IoU thresholds from 10% to 50%, the average recall for the proposed method is 1.000.

3. Dataset Description

For the objectives of this study, the KiTS21 dataset from the 2021 Kidney Tumor Segmentation Challenge is used [16]. The dataset contains CT images from 300 patients who went through partial or radical nephrectomy for suspected malignancy in the renal area. The dataset was collected from M Health Fairview and Cleveland Clinic Medical Center. There was a challenge preceding the 2021 challenge, the 2019 Grand Challenge, which included the same 300 cases [17]. However, the 2021 version has added 'cyst' as a semantic class, resulting in 3 total classes of 'cyst',' tumor', and 'kidney'.

The KiTS21 dataset follows a case-wise structure with each case having a folder. Each folder contains the respective 'imaging' NIFTI extension file as well as three aggregated masks. The CT images are in the format of 'imaging' NIFTI (Neuroimaging Informatics Technology Initiative) extension files. They have three channels, corresponding to the axial, coronal, and sagittal planes. The coronal and sagittal planes consist of 512-pixel channels, while the axial plane is variable, ranging from 29 to 1059. The three aggregated segmented masks are derived using AND, OR, and major set operations. It is in these masks that the semantic classes have been segmented. For this study's use, the 'aggregated_OR_seg' is used. 0, 1, 2, and 3 correspond to the background, the normal kidney area, the tumor area, and the cyst area respectively.

There is a mean number of 544 slices per CT 'imaging' file, with a standard deviation of about 473.9. The thickness of the slices varies from 0.5-5mm and the resolution varies from 0.437-1.041mm per pixel. The mean slice thickness is 2.75 mm, with a standard deviation of approximately 1.61 mm. The mean slice resolution is 0.74 mm per pixel with a standard deviation of about 0.26 mm per pixel.

4. Methodology

4.1. Proposed Detection-guided Segmentation Method

The proposed method is a 2-step detection-guided segmentation (DGS) method to segment abnormal kidney areas that are tumorous or cystic. For our objective of detecting and segmenting kidney tumors, we consider two classes: ‘kidney’ and ‘abnormal’. The kidney area without any tumors or cysts constitutes the ‘kidney’ class. The regions of the kidney that include either a tumor or cyst or both are considered the ‘abnormal’ class. Firstly, YOLOv8 kidney detection is conducted on 2D axial plane images, and using the predicted bounding boxes the images are cropped to kidney region. Secondly, abnormal region segmentation is conducted using U-Net model with ResNet50 encoder. For evaluation, the 2D axial images are combined for each case to construct a 3D CT abnormal region mask and compared against ground truth mask. The entire process is visualized in Figure 1.

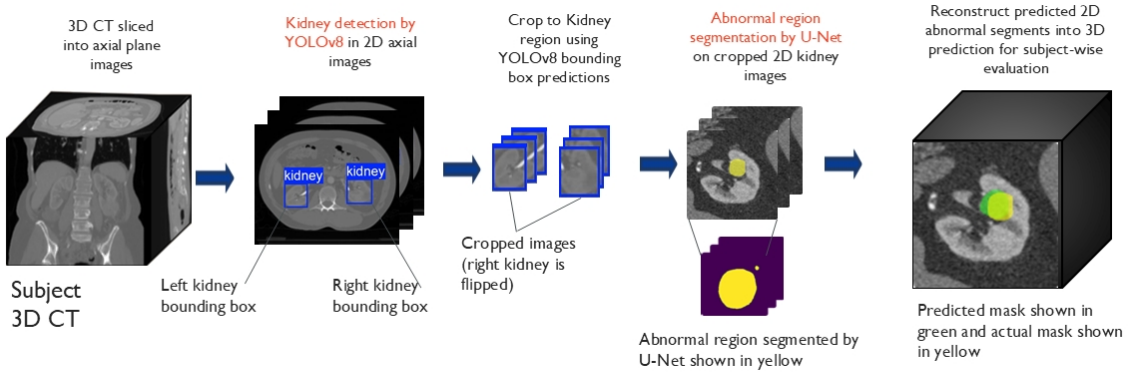

FIGURE 4.1 Block diagram of proposed detection-guided segmentation method

Firstly, 2D PNG axial slice images are generated from the 300 cases of the KiTS21 dataset. The 300 cases are then randomly divided into two halves. From here, 70 cases total are reserved by combining cases from each half for testing. 110 cases are utilized for the first task of training YOLOv8 for detection of the individual kidneys. 120 cases are reserved for the second task of segmentation of the abnormal region.

For detection-guided segmentation, the first task is kidney detection and the second task is abnormal region segmentation. For the first task, the object detection model YOLOv8 is trained on the 110 cases reserved for detection. Then, this trained YOLOv8 model is used to predict the bounding boxes of both kidney regions on the 120 cases reserved for abnormal region segmentation. Using these bounding boxes, the kidney region is cropped out. The detection-guided cropping process is visualized in Figure 2. After that, for the second task, the abnormal region is segmented from the cropped kidney images using the U-Net segmentation model. That concludes the second task. In the end, subject-wise evaluation is carried out on the 70 test cases.

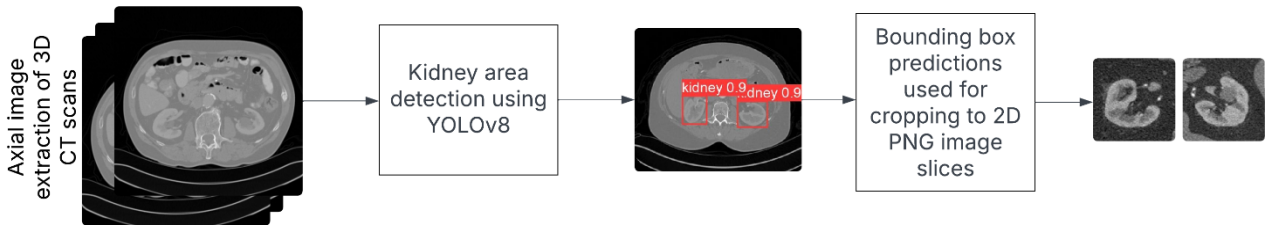

FIGURE 4.2 Detection-guided cropping using YOLOv8

4.1.1 Preprocessing

For the preprocessing, the CT images are sliced into 2D PNG (Portable Network Graphic) extension images in the axial plane. As one of the tasks for detection-guided segmentation is to detect the kidney region using the YOLOv8 model, corresponding YOLOv8 bounding box coordinates are needed. Thus, as the first task, YOLOv8 must be trained to predict the bounding boxes of individual kidneys. In the case of the detection training data of 110 cases, two components are utilized, firstly 2D CT images of 512x512 resolution and kidney masks corresponding to the

cases in the KiTS21 dataset. Here, a simple pre-processing code is used to read the kidney regions from the kidney masks using OpenCV's ‘findContours’ followed by the ‘boundingRect’ function to calculate bounding box rectangles for each kidney [51]. They are then converted to YOLOv8 format coordinates and saved as text files for each image. These text files with the YOLOv8 bounding box coordinates along with the corresponding CT images were the input for the YOLOv8 model. This input was essential for training the model to detect the kidney regions accurately during the kidney detection task. After processing all the images, the dataset was split into 100 cases for training with 5,305 images, 40 cases for test with 2,262 images, and 10 cases for validation with 610 images. Figure 5 shows the bounding boxes generated by YOLOv8.

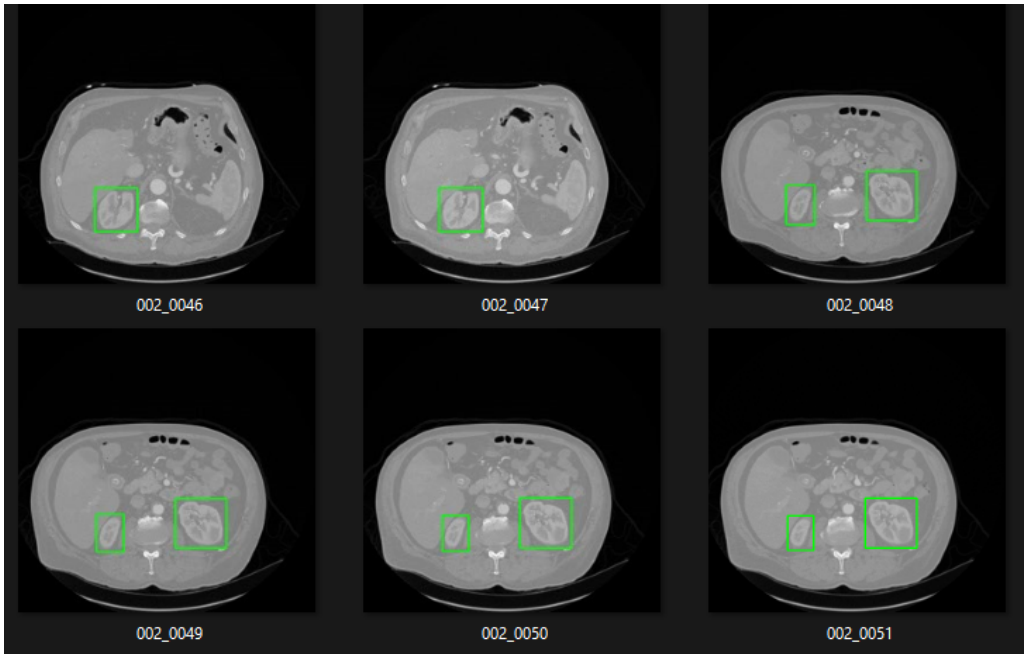

FIGURE 4.3 YOLOv8 kidney detection on axial image slices

For the second task, the previously trained YOLOv8 in the first task is utilized, followed by U-Net segmentation with ResNet50 as encoder. Having trained YOLOv8 for kidney detection, bounding box predictions for individual kidneys are found using YOLOv8. To avoid the error of the edges of the kidney being cut off, the bounding box coordinates are expanded by 20%. Using these bounding boxes, the individual kidney regions, both left and right, are cropped for each slice

where kidney is present. Since kidneys are nearly symmetric anatomical structures, the right kidney images were flipped horizontally. This allowed for increasing the data while ensuring the segmentation model was focusing on segmentation during training. The dataset then fed into the segmentation model, U-Net. For the 120 cases reserved for the segmentation task, 100 cases are used for training and 20 cases are used for validation. There are 8,234 images and abnormal region masks from 100 cases for training and 1,154 images and abnormal region masks from 20 cases for validation.

4.1.2 YOLOv8

The YOLOv8 object detection model is used in both tasks. It is trained in the kidney detection task to be able to generate the bounding box predictions of individual kidneys. Then, the YOLOv8 model trained to detect kidneys is run on the segmentation task data. The bounding box predictions generated are utilized for cropping the kidney region. By using YOLOv8 for cropping before segmentation, this paper aims to highlight the benefits and potential necessity of using a dedicated object detection model in segmentation tasks. For this purpose, a real-time object detection model YOLOv8 has been selected. The You Only Look Once (YOLO) is a series of object detection models that was first developed in 2015 by Joseph Redmon [52]. The algorithm works by dividing the input image into a grid of cells. Within each cell, it tries to predict an object's presence, the bounding box coordinates of the object, and the class of the object. The algorithm of YOLOv8 allows for the identification and location of objects within an image in a single pass through the neural network. Since the release of the original, subsequent versions have been constantly updated. YOLOv8 is the latest version put forth by Ultralytics [18].

The architecture of YOLOv8 is composed of a convolutional neural network that has two main parts: the 'backbone' and the 'head'. The 'backbone' extracts features from the input image at different resolutions. YOLOv8 uses a modified CSPDarknet53 architecture, which has 53 convolutional layers. The model includes partial connections across stages, so there is better channeling of information between the layers. The features learned in the 'backbone' are then passed through a 'neck' that aggregates features from different resolutions for the 'head'. The 'head' consists of a series of convolutional and fully connected layers followed by activation

functions and output layers to make predictions of the objects. The output predictions are in the form of bounding boxes, object presence scores, and class probabilities of the object. Various loss metrics such as box loss, class loss, and object loss help optimize the training.

4.1.3 U-Net Architecture with ResNet50 encoder

In this paper, the U-Net architecture is used for the segmentation of abnormal regions in the segmentation task. The U-Net is a convolutional neural network created for biomedical segmentation, primarily named for its shape. It has an encoder model where the input size is down-sampled for hierarchical feature maps and a decoder model where the input size is up-sampled for accurate pixel location. The output is an image, with a predicted segmentation of the kidney region. The encoder part of the U-Net continuously reduces the spatial dimensions while increasing the number of feature mappings by downsampling the input images through a series of convolutional and pooling layers. This allows the network to learn abstract representations of the input data, which are necessary for precise segmentation.

The encoder model used for this paper is the CNN model ResNet50 [53] of the Keras library [54]. ResNet is a CNN architecture that utilizes skip connections, which enables the network to learn residual functions rather than fitting the intended underlying mapping directly. These skip connections facilitate the training of extremely deep networks which makes it suitable for complex tasks such as kidney abnormal area segmentation. As the name suggests, ResNet50 is a variation of ResNet with a 50-layer deep structure, made up of several residual blocks, which are building blocks that each contain multiple convolutional layers as well as activation functions.

4.2 Comparative Methods

There are four comparative methods. Firstly, a ‘baseline U-Net’ method is conducted with direct U-Net abnormal region segmentation on uncropped CT images from the KiTS21 dataset. Secondly, a ‘baseline YOLOv8’ method is conducted with direct YOLOv8 multi-object detection on uncropped CT images for kidney objects and abnormal region objects. These two baseline

methods are compared to show that the detection-guided cropping technique is a significant improvement from direct methods.

Thirdly, a segmentation-guided classification (SGC) technique is compared with the proposed method. In this technique, the kidney region is segmented from abdominal CT using TotalSegmentator (TS) [55], which is a 3D organ segmentation model. Using a segmentation-guided cropping technique, bounding boxes of each 3D TS kidney segment is calculated using minimum and maximum coordinates of the TS kidney segments in the x, y, and z planes. These bounding box coordinates are used to crop to the kidney region and generate axial slices. Using this segmentation-guided cropping method, the background area is reduced. Following cropping, the ResNet101 [53] CNN is used as the classification model to classify the cropped kidney images into ‘normal’ and ‘abnormal’. Fourthly, a segmentation-guided segmentation (SGS) method is compared with the proposed method. Segmentation-guided cropping technique is used employing 3D TS kidney segments like in the previous method. For segmentation-guided segmentation, this cropping is followed up by U-Net segmentation using ResNet50 as the encoder. These two segmentation-guided methods are compared with the proposed method to showcase the advantages of detection-guided segmentation over segmentation-guided methods.

4.2.1 Preprocessing

Aside from SGC, which is still at the preliminary stage of testing, all other comparative methods have the same case split of 100 train and 20 valid. For SGC, 150 cases were randomized and split for train, test, and validation at 80%, 10%, and 10% respectively.

For baseline U-Net and baseline YOLO, the only preprocessing step is 2D axial slice generation from 3D PNGS. These images are fed, uncropped, into U-Net model and YOLOv8 model respectively. Baseline U-Net segments kidney regions and abnormal regions. Baseline YOLO detects kidney objects and abnormal objects with multi-class multi-object classification.

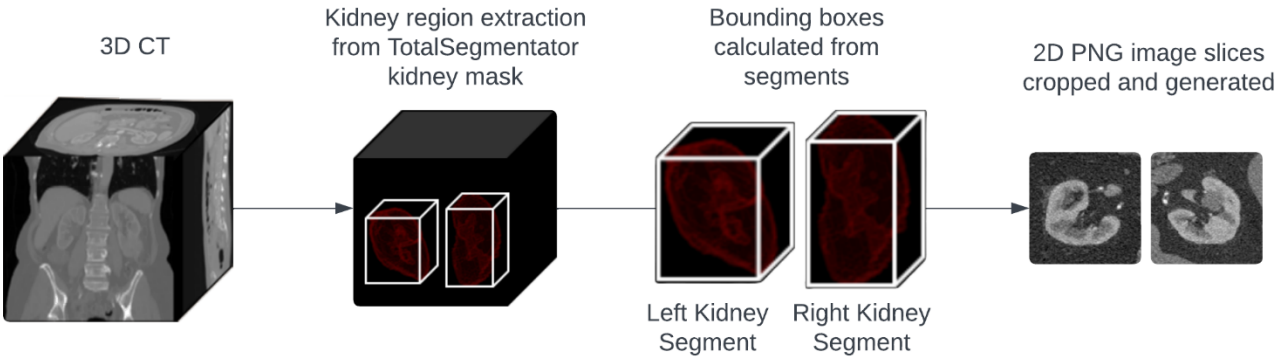

FIGURE 4.4 Segmentation-guided cropping using TotalSegmentator

For SGS and SGC, a segmentation guided cropping technique using TS is used to remove the background as much as possible for each respective dataset. This cropping technique is visualized in Figure 4. Using TS, bounding boxes are calculated for individual kidneys. TS model can segment 104 anatomical structures in 3D, of which the 'kidney_left' and the 'kidney_right' are used in this study. TotalSegmentator can compute high-resolution segmentations of 3mm. However, that can take additional time and high computational power. To facilitate the need for computational costs to be low and the speed to be fast, the 'fast' lower resolution of 1.5mm was used to compute kidney segments. Using these segmentations, a singular 3-dimensional bounding box in the x, y, and z planes is calculated to find the kidney region voxel for each case in the dataset. These coordinates serve as the bounding box coordinates for the kidney. The bounding box coordinates are expanded by 20%. The dataset is cropped according to these coordinates. To increase the pool of data and better guide the focus of the model's training, the right kidney images are flipped. This is possible as kidneys are almost symmetrical organs. All images are resized to 224x224.

For SGS, TS segmentation-guided cropping technique is used first, followed by U-Net segmentation model with ResNet50 encoder for abnormal region segmentation. There are 14,172 images and abnormal region masks from 100 cases for training, and 1,918 images and abnormal region masks from 20 cases for validation.

For SGC, TS segmentation-guided cropping technique is used first, followed by classification. Using the information from the 'aggregated_OR_seg' mask from KiTS21 dataset, the images are separated into two overarching classes: 'normal' and 'abnormal'. If pixel values 2 or 3 is found in the corresponding mask of an image, then the image is considered an 'abnormal' class as 2 represents tumor and 3 represents cyst area. Other images are classified as 'normal'. The train, test, and validation split are 80%, 10%, and 10% respectively, with the corresponding image counts being 15970, 1995, and 1995. In train set, the abnormal image count is 4741 and, in test, and validation sets, the abnormal image counts are 592 each. The standard deviation of images per case is 133.67, with range from 25 to 508.

4.2.2 TotalSegmentator

The TotalSegmentator Segmentation model [55] is selected for comparative analysis of the proposed model. Similar to the proposed detection-guided YOLOv8 cropping, TS is used for a segmentation-guided cropping technique. TS is a deep learning-based segmentation model that excels at segmenting anatomical structures from 3D medical images such as CT scans. It can segment 104 anatomical structures in a CT scan in a single pass. It is based on the no-new U-Net (nnU-Net) architecture, which is a self-adapting model capable of configuring hyperparameters to different dataset characteristics [56]. The nnU-Net is an automated framework designed to handle medical segmentation tasks without manual intervention. TS utilizes a nnU-Net algorithm that was trained on 1204 CTs, then a further 4004 CTs, spanning different years, ages, abnormalities, etc.

Although TS is capable of rendering segmentations of 3mm resolution, to enhance speed and keep computational costs low, the 'fast' method is used, which renders 1.5mm resolution kidney masks. The 'kidney_left' and the 'kidney_right' masks are used to calculate the bounding boxes of the individual kidney regions. The bounding boxes are then used to crop the generated axial slices to each kidney region. For each kidney of each case, a 3D bounding box is calculated from the TotalSegmentator kidney masks by finding the maximum and minimum coordinates of the kidney area in the 3 planes, x, y, and z. Then, to avoid the error of kidney edges being cut off, a 20% expansion is applied to these coordinates to achieve the cropped region coordinates. Using these

coordinates, the slices are cropped, resulting in two images of each kidney.

4.2.3 ResNet101

In this paper, the ResNet101 [53] model from the TensorFlow Keras applications library [54] is utilized for the classification part in SGC comparative method. ResNet is a deep learning CNN architecture developed by Kaiming He et. al. ResNet101, a variation of the ResNet architecture, is extensively utilized in computer vision tasks. ResNet101 contains 101 layers. Two main components of ResNet are the residual block, that learns the difference between the input and the output, and the skip connections, that reduce the vanishing gradient problem. The residual blocks follow a bottleneck design, which consists of three convolutional layers with kernel size of: 1×1, 3×3, and 1×1.

5. Evaluation

5.1 Subject-wise Evaluation Method

The evaluation is carried out for 3D performance in a subject-wise approach, where each CT is considered a subject. A test dataset of 70 cases is used for all methods except for preliminary testing for SGC. The metrics to evaluate performance are namely: intersection over union (IoU), recall, and precision. The IoU measures the overlapped region between the predicted 3D image and the ground truth 3D image. The intersection refers to the intersection between the predicted target region voxels and ground truth target region voxels and the union refers to the union between them. Division of the intersection by the union will give a score that can represent how closely the two targeted regions overlap, which provides insight into how successful the prediction is.

For the detection task of proposed DGS, the more closely the reconstructed 3D CT resembles the kidney in the ground truth CT, the better the prediction. The axial slice images that YOLOv8 made bounding box predictions for are used to reconstruct a 3D image of the kidney region. To do this, all the slice images pertaining to one subject are ordered according to the slice image number and combined together, thus adding another dimension and constructing a 3D NIFTI file. This is done for all 70 subjects for each dataset. The IoU is calculated for each kidney of each subject, and then the values are averaged for each subject.

For the segmentation task of proposed DGS and for SGS, the more closely the reconstructed 3D CT can locate the abnormal region, the better the prediction. The 3D evaluation technique used to judge 3D abnormal mask predictions is visualized in Figure 5. For each dataset, the following steps are conducted. Firstly, the 2D PNG axial slice images that have had the abnormal region segmented by U-Net are used to reconstruct a 3D mask image of the abnormal region. For all 70 subjects for each dataset, the slice images pertaining to one subject are ordered and combined according to the slice image number, thus adding another dimension and constructing a 3D NIFTI file. The predicted abnormal region 3D mask is then compared to the respective subject’s ground truth CT image.

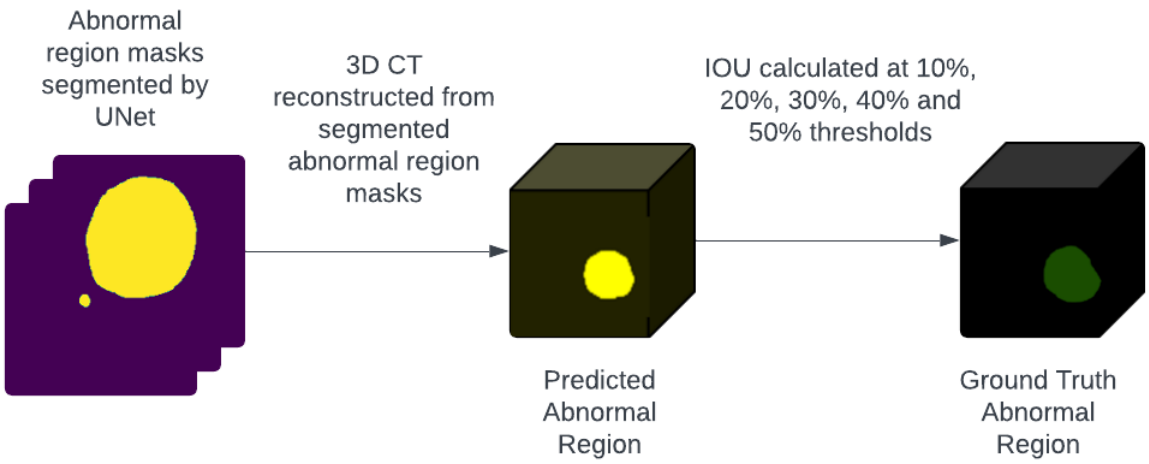

FIGURE 5.1 Subject-wise evaluation technique

To calculate the recall and precision, the true positive (TP), the true negative (TN), the false positive (FP), and the false negative (FN) numbers must first be calculated. The TP, TN, FP, and FN values are defined as follows:

- TP: occurs when both the predicted and ground truth masks indicate an abnormal region, i.e. both are 1
- TN: occurs when both masks indicate no abnormal region, i.e. both are 0
- FP: occurs when the model predicts an abnormal region but there is not one in the ground truth mask, i.e. prediction is 1 and ground truth is 0
- FN: occurs when the model misses an abnormal region that is present in the ground truth, i.e. prediction is 0 and ground truth is 1

Recall is the rate of the correctly predicted positive instances among all the ground truth positives. Precision is the accuracy of the positive predictions. Precision and Recall are calculated using the subject-wise TP, TN, FP, and FN with equations (1)-(2).

$$Precision = \frac{TP}{(TP + FP)} \quad (1)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (2)$$

An IoU threshold sets the minimum overlap required between the predicted and ground truth bounding boxes for a prediction to be considered a true positive. A range of thresholds is applied to determine if a given IoU score is considered a TP or not. If an IoU is below a certain percentage (50%/40%/30%/20%/10%) and it is TP, then the TP becomes FP, i.e. 1 is set to 0 and the threshold value is set to 1. For instance, if a prediction is a TP at a 50% threshold but has an IoU below the 50% threshold when reassessed at a lower threshold (e.g., 40%, 30%), the TP becomes an FP. This reclassification is conducted because the overlap is no longer sufficient at the lower threshold.

The TP, TN, FP, and FN are calculated for each case of the test dataset at threshold values at threshold values 50%, 40%, 30%, 20%, and 10%. Then for each threshold percentage, the values for the left and right kidney of each case are averaged.

5.2 Evaluation Method for SGC

Currently, preliminary testing of SGC has been conducted, image-wise, on a 1995 test image dataset. Precision and recall are calculated using the same equations as before, except the calculation is image-wise instead of subject-wise. For SGC, additional metrics for evaluation are confusion matrix and area under the receiver operating curve (AUROC) for the classification results. A confusion matrix is a tool used when calculating and visualizing a classification model. It is a table where the primary diagonal shows the TP and TN, while the secondary diagonal shows the FP and FN. An AUROC curve is a graphical plot that measures how well a classification model can distinguish between positive and negative examples.

6. Results and Discussion

6.1 Proposed Detection-Guided Segmentation

The results of the detection pipeline and the abnormal region segmentation pipeline show promising outcomes for the proposed detection-guided segmentation method. These are discussed in detail in the subsections.

6.1.1 Kidney Detection

The YOLOv8 trained for kidney detection task of proposed DGS was able to detect the kidney highly accurately, with an average IoU score of 0.9206 for 70 subjects. This is averaged from the left and right kidney, with IoU scores of 0.9210 and 0.9203 respectively, as outlined in Table 1. As the IoU score measured the overlap of the detected kidney region to the ground truth kidney region, the detection can be determined to be highly accurate. The results are also consistent across a relatively large sample size of 70 subjects. This suggests that the pipeline performs well on different individual cases and is not overfitted to a specific dataset. The slight difference of 0.0007 in IoU scores between the left and right kidneys shows that the pipeline has consistent symmetrical performance for both kidneys, as it performed almost equally well on both sides. TS achieved lower IoU, achieving an average score of 0.8577 across both kidneys.

TABLE 6.1 IoU for Kidney Detection

	Left Kidney	Right Kidney	Average IoU
YOLOv8	**0.9210**	**0.9203**	**0.9206**
TS	0.8518	0.8637	0.8577

6.1.2 Tumor Segmentation

6.1.2.1 Recall

Table 2, Table 3, and Figure 6 show that for average recall, the proposed method performs the best among all compared, achieving 1.000 for every threshold. Recall calculates the ratio of TP to sum of TP and FN, as shown in equation (2). Thus, the recall is the ratio of correctly classified foreground pixels to the total number of true foreground pixels in the image. Recall is the primary metric for evaluation, as high recall means that the method is able to accurately identify as many abnormal images as possible. Medical specialists consider recall metrics closely, as a missed abnormality will pose as poor diagnosis and is not as applicable practically. However, a CAD method with high recall but not as high precision is applicable as after screening abnormalities correctly, specialists can further screen out the false positive predictions. The perfect recall values across all thresholds considered show that the proposed DGS has captured most of the abnormal region in the segmentation, meaning it has no false negatives. The graph in Figure 8 shows the constant recall value across the 1.000 on the y-axis. The same score across all thresholds also indicates the consistent performance of proposed DGS method. Table 4 shows the recall value averaged from individual kidneys of each subject in the test dataset, emphasizing its consistent performance across each kidney.

TABLE 6.2 Average Recall at Various IoU thresholds

IoU Threshold (%)	Proposed DGS	SGS	Baseline U-Net	Baseline YOLO
10	**1.000**	**0.9714**	**0.7308**	**0.6162**
20	**1.000**	0.9677	0.6818	0.6150
30	**1.000**	0.9600	0.6500	0.6125
40	**1.000**	0.9523	0.5882	0.6095
50	**1.000**	0.9412	0.4815	0.6058

TABLE 6.3 Average Recall for SGC

Metric	Score
Recall	**0.9626**

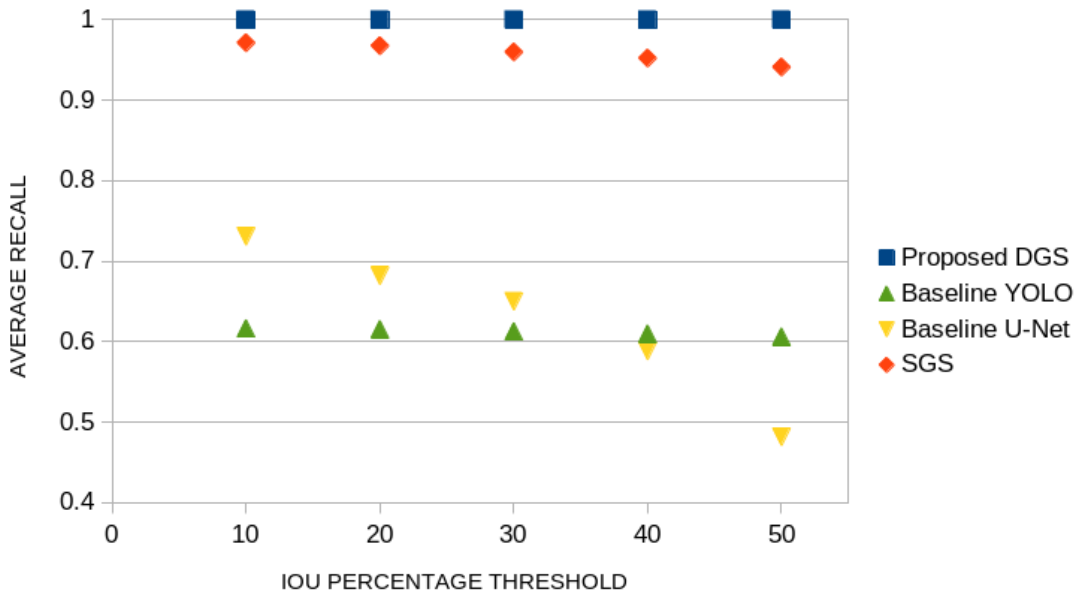

FIGURE 6.1 Average Recall for Proposed DGS, SGS, baseline U-Net, and baseline YOLO

TABLE 6.4 Recall of Individual Kidney for Proposed DGS and SGS at Various IoU Thresholds

IoU Threshold (%)	Proposed DGS		SGS	
	Left Kidney	Right Kidney	Left Kidney	Right Kidney
10	**1**	**1**	**1**	**0.9429**
20	1	1	1	0.9355
30	1	1	1	0.9200
40	1	1	1	0.9048
50	1	1	1	0.8823

6.1.2.2 Precision

As shown in equation (1), precision measures the proportion of true positive predictions among all predictions made by the model. Table 5, and Figure 7 show that the proposed method outperformed SGS at thresholds 40% and 50%, but performed lower at lower thresholds. Furthermore, the preliminary results for image-wise test of SGC shows the highest precision score at 0.8670 as shown in Table 6. The highest precision score for proposed DGS is 0.7075 at 10% threshold. The precision falls slightly as the threshold increases. This shows YOLO's ability to locate the general abnormal region. However, it starts struggling when the required precision increases such as when the region of overlap is increased. Table 7 shows the precision value averaged from individual kidneys of each subject in the test dataset.

TABLE 6.5 Average Precision at Various IoU thresholds

IoU Threshold (%)	YOLO	TS
10	0.7075	**0.7130**
20	0.6699	**0.6873**
30	0.6508	**0.6465**
40	**0.6213**	0.6152
50	**0.6000**	0.5624

TABLE 6.6 Average Precision for SGC

Metric	Score
Precision	**0.8670**

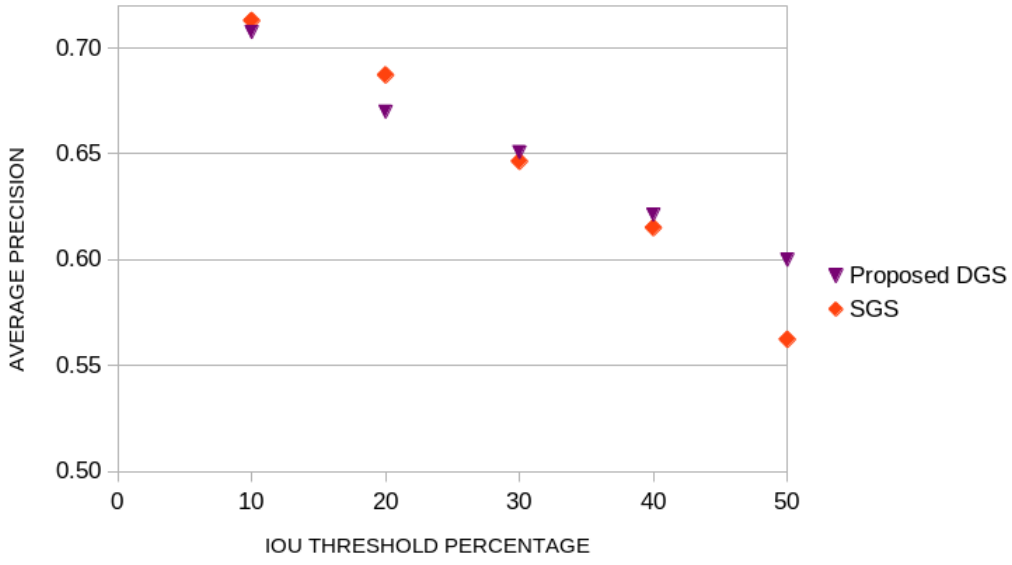

FIGURE 6.2 Average Precision for Proposed DGS and SGS

TABLE 6.7 Precision of Individual Kidney for Proposed DGS and SGS at Various IoU Thresholds

IoU Threshold (%)	Proposed DGS		SGS	
	Left Kidney	Right Kidney	Left Kidney	Right Kidney
10	0.7167	**0.6984**	**0.7660**	0.6600
20	0.6731	**0.6667**	**0.7442**	0.6304
30	0.6600	**0.6415**	**0.7180**	0.5750
40	0.6304	**0.6122**	**0.7027**	0.5278
50	0.6222	**0.5778**	**0.6562**	0.4688

6.2 SGC Classification Results

For SGC, additional metrics for preliminary evaluation are confusion matrix and AUROC curve. A total of 515 abnormal TP predictions were made as shown in the confusion matrix in Table 8. The AUROC for ResNet101 is 0.986 and 0.984 for abnormal and normal respectively, shown in Figure 8. The curve shows that the false positive rate for the model is very low and the true positive rate is very high, close to the ideal AUROC score of 1. The minuscule difference in the scores for abnormal and normal also shows the model's consistency in performance when classifying both classes.

TABLE 6.8 Confusion Matrix of SGC

Actual Labels	Predicted Labels	
	Abnormal	Normal
Abnormal	**515**	79
Normal	20	1381

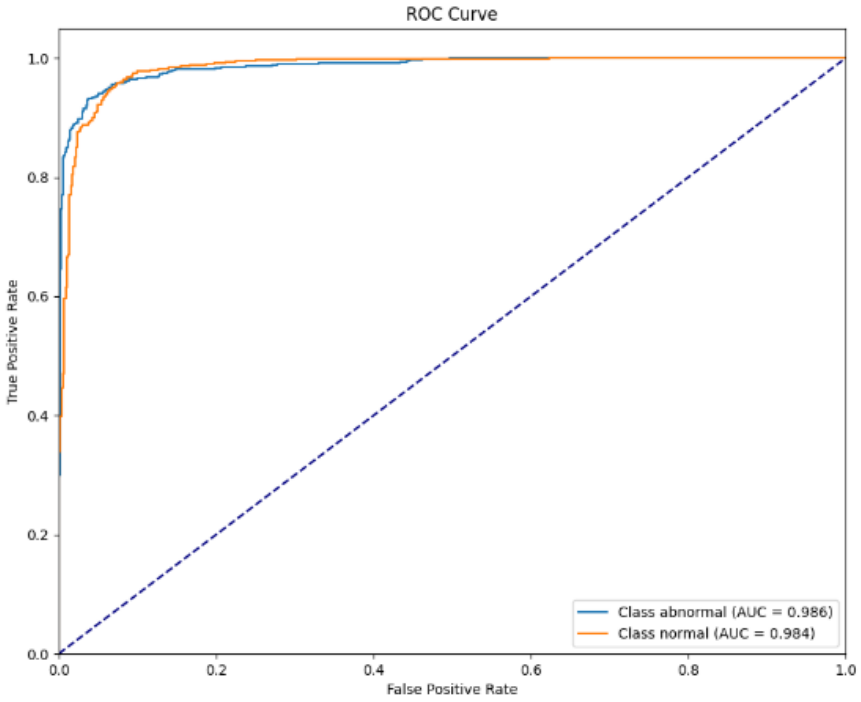

FIGURE 6.3 SGC AUROC Curve

6.3 Comparison

6.3.1 Comparison of Proposed DGS to baseline methods

One of the hypotheses of proposing a DGS method is that using a guided framework can improve results. Thus, the proposed DGS method is compared with baseline U-Net and baseline YOLO, which deal with uncropped data. Table 2 and Figure 6 show that the proposed DGS method surpasses the baseline methods at recall, achieving an average of 1.000 across all thresholds whereas baseline U-Net achieved the highest recall value of 0.7308 and baseline YOLO achieved the highest recall value of 0.6162 at 10% IoU. The other two guided methods, SGS and SGC, also achieved a

much higher score than baseline methods. SGS achieved the highest recall value of 0.9714 at 10% IoU and the image-wise preliminary result of SGC is 0.9626. Thus, the results validate the improvement achieved by cropping the kidneys using guided cropping techniques.

6.3.2 Comparison of Proposed DGS to Segmentation-guided methods SGS and SGC

One of the hypotheses of proposing DGS is that using a dedicated object detection model before segmentation can improve segmentation results and can be used to develop a more applicable CAD method than segmentation-guided techniques. SGS and SGC are compared to the proposed DGS. In Table 1, for the kidney detection task, YOLOv8 performs significantly better at locating the kidney compared to TS. The highest IoU score of TS is 0.8637 and the average score is 0.8577, which are weaker compared to YOLO's highest IoU score of 0.9210 and average score of 0.9209.

Regarding the recall results shown in Table 2, proposed DGS achieved superior results. This indicates the lesser performance of SGS and SGC at learning the details of the abnormal region. The average recall scores for individual kidneys are shown in Table 4. SGS achieves a recall score of 1.000 for the left kidney at each threshold but the highest recall score for the right kidney is 0.9429 at 10% IoU. The right kidney images were flipped during dataset creation to increase testing data, which may have caused this disparity. Proposed DGS recall scores are 1.000 for both kidneys at every threshold, which indicates it is better adept at capturing slight diversity in data.

Precision is an area where proposed DGS performs better at one aspect and worse at the other. The highest precision score of proposed DGS is lower at 0.7075 compared to SGS score of 0.7130 and the image-wise score of 0.8670 for SGC. For the thresholds 10% to 30%, DGS and SGS perform so closely that the graph curves shown in Figure 7 intersect, only diverging at upwards of 40%, where DGS has a higher precision. The improving performance for precision as the abnormal region grows shows that DGS model has a decreasing rate of false positives. It also suggests that DGS is superior at capturing larger abnormal regions. Both DGS and SGS results fall as the threshold increases, which shows their ability to locate the general abnormal region but suggests that they struggle with accuracy when the overlap requirements are higher. However, DGS is more consistent, as its precision falls at a lower rate. However, considering how all guided methods achieved high recall

scores compared to precision suggests a common factor. This could be due to data factors such as class imbalance, as the number of normal class images are significantly higher than abnormal class images. Additionally, medical specialists prefer to consider recall metrics, as a missed abnormality will pose danger if not diagnosed. Detecting a normal region as an abnormality is more acceptable than detecting an abnormality as a normal region. There, a CAD method with higher recall but lower precision is more applicable practically than a CAD method with lower recall and higher precision. This is because after all abnormalities are screened correctly, specialists can further screen out the false positive predictions. Thus, the proposed DGS method is more suited for practical medical applications.

To find the optimal path forward that can result in an automated CAD system medical professionals can use, deep-learning solutions also need to consider certain criteria such as the computational complexity, the capacity for explainability, and the capacity for interpretability of the approach. Computational complexity directly affects the cost, time, and efficiency of a system. Lower computational complexity without compromising results is the ideal balance. Explainability will help nephrologists understand which regions the model is focusing on when making decisions. Interpretability is the degree of understanding of the cause of a model's decision. The higher the interpretability, the more comprehendible the approach is for nephrologists.

When it comes to computational complexity, DGS is less computationally complex. The most computationally complex step, segmentation, takes into account every voxel. In segmentation-guided methods like SGS and SGC, this takes place over a whole 3D CT image. This is a significantly large volume with many other organs and boundary lines. Compared to that, in DGS, the segmentation takes place over a much smaller, relevant region because the kidney area has already been cropped using the detection-guided cropping technique.

DGS is the most explainable, followed by SGS and SGC. DGS mirrors human reasoning by locating general area of target first, i.e. detect bounding boxes, followed by more precise location, i.e. segment masks, making it the easiest to explain. SGS is similar but it goes from coarse-to-fine segmentation, which is a bit less explainable than DGS. For the classification step in SGC, it is difficult to relate the prediction to the input image without advanced explainability tools.

Explainability tools such as Gradient-Weighted Class Activation Mapping (Grad-CAM) can help identify the strongest points of interest for decision-making.

DGS is more interpretable than SGS and SGC. Bounding boxes are easily interpretable to humans. Refining the localization with segmentation is also easy to interpret. Whereas, creating a segmentation mask requires the model to make complex decisions about pixel relationships and shapes, which makes SGS and SGC harder to interpret intuitively. Furthermore, the classification step of SGC also clouds the interpretability as it is not immediately understandable how the model came to a certain prediction, as class labels are less visually tied to the image.

Given that the proposed DGS method scores the highest in the more relevant metric of recall, is less computationally complex, is more interpretable and is more explainable, the method is most suited to be the basis for a CAD system for detecting kidney abnormal regions.

7. Conclusion and Future Work

This study proposed a fully automated method for the detection of abnormal kidney areas that are tumorous or cystic in 3D CT images. To achieve this, a 2-step detection-guided segmentation technique is proposed and then compared. The first task is kidney detection and the second task is abnormal region segmentation. The detection model chosen is YOLOv8 and the segmentation architecture used is U-Net with ResNet50 as the encoder. From the results, the proposed approach performed highly for recall with a consistent score of 1.000 for each threshold and performed adequately for precision with the highest score being 0.7075 at 10% IoU. Four comparative methods are conducted in total. Two baseline methods, baseline U-Net and baseline YOLO, are conducted to show that guided frameworks perform better than single task frameworks. Two segmentation-guided methods, segmentation-guided segmentation and segmentation-guided detection, are compared to show how proposed method outperforms in the more relevant metric of recall, and is most suited to be the basis for a CAD system for detecting kidney abnormal regions.

There are certain limitations we aim to address in the future. The scope of this research is based around the processing of axial plane kidney images generated from 3D CT images. This was done as medical specialists prefer axial plane images as it provides a cross-sectional view of the body and because relevant literature focuses on axial plane which makes it easier for comparison. This limitation will be addressed in the future through the inclusion of sagittal and coronal plane images. Additionally, this study did not incorporate explainability such as with visualization techniques like Grad-CAM (Gradient-weighted Class Activation Mapping), saliency maps, or LRP (Layer-wise Relevance Propagation). These would provide insight into which aspect the model considered most when generating the predictions and help explain the model's behavior. This aspect will be added for future work. Precision results for proposed DGS is considerably lower than the exceptionally high recall results. This means that the current method is predicting a notable rate of false positives. This disparity will be addressed in future work. As this disparity exists for SGS and SGC as well, the data might be a factor. The dataset contains significantly more normal images than abnormal, which creates a class imbalance. Further data augmentation techniques will be considered. For SGC, only preliminary image-wise testing results were included. In the future, subject-wise evaluation will be conducted for this approach for more reliable comparison. For DGS, SGS, and SGC, there is a need to find a dataset where there are whole CT images without

abnormal regions in either kidney so that the proposed method's precision results can be compared to the uncropped dataset. There are also other objectives to extend the work in the future. Firstly, other real-time and non-real-time object detection models should be compared to YOLOv8 to find which model is most optimal. In the same vein, different encoders should be tested against the use of ResNet50 in the U-Net segmentation to find the optimal encoder model.

References

[1] Freddie. Bray, Mathieu. Laversanne, Hyuna. Sung, Jacques. Ferlay, Rebecca. Siegel, Isabelle. Soerjomataram, and Ahmedin. Jemal, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", CA: A Cancer Journal for Clinicians, Vol. 74, No. 3, pp. 229-263, April 2024.

[2] Rebecca. L. Siegel, Angela. N. Giaquinto, and Ahmedin. Jemal, "Cancer statistics, 2024", CA: A Cancer Journal for Clinicians, Vol. 74, No. 1, pp. 12-49, January 2024.

[3] Padala, SA, Barsouk, A, Thandra, KC, Saginala, K, Mohammed, A, Vakiti, A, Rawla, P and Barsouk, A. (2020). Epidemiology of Renal Cell Carcinoma. World J Oncol. 79-87.

[4] American Cancer Society. (n.d.). Kidney cancer surgery. Retrieved April 6, 2024, from https://www.cancer.org/cancer/types/kidney-cancer/treating/surgery.html

[5] Howlader, N, Noone, AM, Krapcho, M, Miller, D, Brest, A, Yu, M, Ruhl, J, Tatalovich, Z, Mariotto, A, Lewis, DR, Chen, HS, Feuer, EJ and Cronin, KA (eds). SEER Cancer Statistics Review, 1975-2016, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2016/, based on November 2018 SEER data submission, posted to the SEER website, April 2019.

[6] W. Bai, Y. Fadil, A. Chadli, M. Dakir, A. Debbagh, and R. Aboutaeib, "Correlation between CT and anatomopathological staging of kidney cancer", International Journal of Surgery Case Reports, Vol. 80, Art. No. 105687, March 2021.

[7] Gray, RE and Harris, GT. (2019). Renal Cell Carcinoma: Diagnosis and Management. Am Fam Physician. 179-184.

[8] Diaz de Leon, A, Pirasteh, A, Costa, DN, Kapur, P, Hammers, H, Brugarolas, J and Pedrosa, I. (2019). Current Challenges in Diagnosis and Assessment of the Response of Locally Advanced and Metastatic Renal Cell Carcinoma. Radiographics. 998-1016.

[9] Alnazer, P. Bourdon, T. Urruty, O. Falou, M. Khalil, A. Shahin, and C. Fernandez-Maloigne, "Recent advances in medical image processing for the evaluation of chronic kidney disease," Medical Image Analysis, vol. 69, p. 101960, 2021, doi: 10.1016/j.media.2021.101960.

[10] Ibrahim, T. (2022). Top 10 predictions about US nephrologists in 2035. ASN Executive vice president's update. Kidney News. https://www.kidneynews.org/view/journals/kidney-news/14/7/article-p8_3.xml

[11] Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Data, 8 Registries, Nov 2021 Sub (1975-2020) - Linked To County Attributes - Time Dependent (1990-2020) Income/Rurality, 1969-2020 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2023, based on the November 2022 submission.

[12] D. Nair et al., "Burnout Among Nephrologists in the United States: A Survey Study," Kidney Medicine, vol. 4, no. 3, p. 100407, 2022, doi: 10.1016/j.xkme.2022.100407.

[13] F. J. Hlubocky, A. L. Back, and T. D. Shanafelt, "Addressing Burnout in Oncology: Why Cancer Care Clinicians Are At Risk, What Individuals Can Do, and How Organizations Can Respond," American Society of Clinical Oncology Educational Book, vol. 36, pp. 271–279, 2016, doi: 10.1200/EDBK_156120.

[14] E. Pawłowicz and M. Nowicki, "Burnout syndrome among nephrologists - a burning issue - results of the countrywide survey by the Polish Society of Nephrology," BMC Nephrology, vol. 21, no. 1, p. 177, May 2020, doi: 10.1186/s12882-020-01829-2.

[15] T. Shanafelt and L. Dyrbye, "Oncologist burnout: causes, consequences, and responses," Journal of Clinical Oncology, vol. 30, no. 11, pp. 1235–1241, Apr. 2012, doi: 10.1200/JCO.2011.39.7380.

[16] Nicholas. Heller et al., “The KiTS21 Challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase CT”, arXiv:2307.01984, July 2023.

[17] Nicholas. Heller et al., "The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes," arXiv:1904.00445, 2020.

[18] G. Jocher, A. Chaurasia, and J. Qiu, Ultralytics YOLOv8, version 8.0.0, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics.

[19] Ronneberger, O, Fischer, P, Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, 234-241.

[20] Kim, DY and Park, JW. (2004) Computer-aided detection of kidney tumor on abdominal computed tomography scans. Acta Radiol. doi: 10.1080/02841850410001312. PMID: 15624525.

[21] Shah, B., Sawla, C., Bhanushali, S. and Bhogale, P., (2017). Kidney tumor segmentation and classification on abdominal CT scans. International Journal of Computer Applications, **164(9)**, pp.1-5.

[22] Aslam, A., Khan, E. and Beg, M.S., (2015). Improved edge detection algorithm for brain tumor segmentation. Procedia Computer Science, **58**, pp.430-437.

[23] Rudra, A.K., Chowdhury, A.S., Elnakib, A., Khalifa, F., Soliman, A., Beache, G. and El-Baz, A., (2013). Kidney segmentation using graph cuts and pixel connectivity. Pattern Recognition Letters, **34(13)**, pp.1470-1475.

[24] Zeenia. Jagga, "Classification models for clear cell renal carcinoma stage progression, based on tumor RNAseq expression trained supervised machine learning algorithms", Proceeding of BMC Proceedings 8, Cincinnati, May 2014.

[25] Li, B., Xue, J., Wang, P., Gu, S., Chen, W., Qian, P. and Jiang, Y., (2021). An Auto-Contouring Method for Kidney based on SVM. In Journal of Physics: Conference Series Vol. 1944, No. 1, p. 012016.

[26] Tuncer, S.A. and Alkan, A., (2018). A decision support system for detection of the renal cell cancer in the kidney. Measurement, **123**, pp.298-303.

[27] Talaat, D., Zada, F. and Kadry, R., (2020). Staging of clear cell renal cell carcinoma using random forest and support vector machine. In Journal of Physics: Conference Series (Vol. 1447, No. 1, p. 012012).

[28] Esteva, A., Chou, K., Yeung, S. *et al.* (2021). Deep learning-enabled medical computer vision. npj Digit. Med. **4**, 5. https://doi.org/10.1038/s41746-020-00376-2

[29] Zakaria. Rguibi, Hajami. Abdelmajid, and Dya. Zitouni, "Deep Learning In Medical Imaging: A Review", in Applications of Machine intelligence in Engineering, 1st ed., J. K. Mandal, S. Misra, J. S. Banerjee and S. Nayak, Eds. Boca Raton (Florida), USA: CRC Press, 2022., Ch. 15.

[30] Liu, X. et al. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. The Lancet Digital Health

[31] Yamashita, R, Nishio, M, Do, RKG and Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights Imaging. **(4)**:611-629.

[32] Ahmad. W. Salehi, Shakir. Khan, Gaurav. Gupta, Bayan. I. Alabduallah, Abrar. Almjally, Hadeel. Alsolai, Tamanna. Siddiqui, and Adel. Mellit, "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope", Sustainability, Vol. 15, No. 7, Art. No. 5930, March 2023.

[33] Hosna, A., Merry, E., Gyalmo, J. et al. Transfer learning: a friendly introduction. J Big Data 9, 102 (2022). https://doi.org/10.1186/s40537-022-00652-w

[34] Javier. López-Labraca, Iván. González-Díaz, Fernando. Díaz-de-María, and Alejandro. Fueyo-Casado, “An interpretable CNN-based CAD system for skin lesion diagnosis”, Artificial Intelligence in Medicine, Vol. 132, August 2022.

[35] Qian Y, Qiu Y, Li CC, Wang ZY, Cao BW, Huang HX, Ni YH, Chen LL, Sun JY. A novel diagnostic method for pituitary adenoma based on magnetic resonance imaging using a convolutional neural network. Pituitary. 2020 Jun;23(3):246-252. doi: 10.1007/s11102-020-01032-4. PMID: 32062801.

[36] Dabiri, S, Popuri, K, Ma, C, Chow, V, Feliciano, EMC, Caan, BJ, Baracos, VE and Beg, MF. (2020) Deep learning method for localization and segmentation of abdominal CT. Comput Med Imaging Graph.

[37] Alalwan, N, Abozeid, A, ElHabshy, AA and Alzahrani, A. (2021) Efficient 3D deep learning model for medical image semantic segmentation. Alexandria Engineering Journal.

[38] Kim, T, Lee, KH, Ham, S, Park, B, Lee, S, Hong, D, Kim, GB, Kyung, YS, Kim, CS & Kim N. (2022) Active learning for accuracy enhancement of semantic segmentation with CNN-corrected label curations: Evaluation on kidney segmentation in abdominal CT. Sci Rep.

[39] Hsiao, Chiu-Han, et al. (2022). A deep learning-based precision and automatic kidney segmentation system using efficient feature pyramid networks in computed tomography images. Computer Methods and Programs in Biomedicine 221

[40] Kittipongdaja, P and Siriborvornratanakul, T. (2022) Automatic kidney segmentation using 2.5D ResUNet and 2.5D DenseUNet for malignant potential analysis in complex renal cyst based on CT images. EURASIP J Image Video Process.

[41] Causey, J, Stubblefield, J, Qualls, J, Fowler, J, Cai, L, Walker, K, Guan, Y and Huang, X. (2022) An Ensemble of U-Net Models for Kidney Tumor Segmentation With CT Images. IEEE/ACM Trans Comput Biol Bioinform. Yunfei.

[42] Li, Xinrui. Gao, Xuemei. Tang, Sheng. Lin, and Haowen. Pang, “Research on automatic classification technology of kidney tumor and normal kidney tissue based on computed tomography radiomics”, Frontiers in Oncology, Vol. 13, February 2023.

[43] Wenshuai. Zhao, Dihong. Jiang, Jorge. P. Queralta, and Tomi. Westerlund, “MSS U-Net: 3D segmentation of kidneys and tumors from CT images with a multi-scale supervised U-Net”, Informatics in Medicine Unlocked, Vol. 19, June 2020.

[44] Patike. K. Rao, Subarna. Chatterjee, and Sreedhar. Sharma, “Weight Pruning-UNet: Weight Pruning UNet with Depth-wise Separable Convolutions for Semantic Segmentation of Kidney Tumors”, Journal of Medical Signals and Sensors, Vol. 12, No. 2, pp. 108-113, May 2022.

[45] Sefatul. Wasi, Saadia. B. Alam, Rashedur. Rahman, M. A. Amin, and Syoji. Kobashi, “Kidney Tumor Recognition from Abdominal CT Images using Transfer Learning”, Proceeding of 2023 IEEE 53rd International Symposium on Multiple-Valued Logic, Matsue, May 2023.

[46] Ruan, Y, Li, D, Marshall, H, Miao, T, Cossetto, T, Chan, I, Daher, O, Accorsi, F, Goela, A and Li, S. (2020) MB-FSGAN: Joint segmentation and quantification of kidney tumor on CT by the multi-branch feature sharing generative adversarial network. Med Image Anal.

[47] Sun, P, Mo, Z, Hu, F, Liu, F, Mo, T, Zhang, Y and Chen, Z. (2022) Kidney Tumor Segmentation Based on FR2PAttU-Net Model. Front Oncol.

[48] Song, Y, Zheng, J, Lei, L, Ni, Z, Zhao, B and Hu, Y. (2022) CT2US: Cross-modal transfer learning for kidney segmentation in ultrasound images with synthesized data. Ultrasonics.

[49] Alzu'bi, D., Abdullah, M., Hmeidi, I., AlAzab, R., Gharaibeh, M., El-Heis, M., Almotairi, K. H., Forestiero, A., Hussein, A. M., & Abualigah, L. (2022). Kidney tumor detection and classification based on Deep Learning Approaches: A new dataset in CT scans. Journal of Healthcare Engineering, 2022, 1-22. https://doi.org/10.1155/2022/3861161

[50] Vlad-Octavian. Bolocan, Mihaela. Secareanu, Elena. Sava, Cosmin. Medar, Loredana. S. C. Manolescu, Alexandru-Stefan. C. Rasuc, Maria. G. Costache, George. D. Radavoi, Robert-Andrei. Dobran, and Viorel. Jinga, “Convolutional Neural Network Model for Segmentation and Classification of Clear Cell Renal Cell Carcinoma Based on Multiphase CT Images”, Journal of Imaging, Vol. 9, No. 12, Art. No. 280, December 2023.

[51] Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.

[52] Redmon, J, Divvala, S, Girshick, R and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016. 779-788.

[53] Kaiming. He, Xiangyu. Zhang, Shaoqing. Ren, and Jian. Sun, “Deep Residual Learning for Image Recognition”, Proceeding of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, June 2016.

[54] F. Chollet et al., Keras, 2015. [Online]. Available: https://keras.io.

[55] Wasserthal, J., Breit, H.-C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D., Cyriac, J., Yang, S., Bach, M. and Segeroth, M., (2023). TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. Radiology: Artificial Intelligence. https://doi.org/10.1148/ryai.230024

[56] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, **18(2)**, 203-211.