Kidney and Kidney Tumor Segmentation, 2019 (KiTS-19)

Author

Ramsha Abbasi

Regn Number

00000320845

Supervisor

Dr. Syed Omer Gilani

DEPARTMENT OF BIOMEDICAL ENGINEERING AND SCIENCES

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD, PAKISTAN

September, 2022

Kidney and Kidney Tumor Segmentation, 2019 (KiTS-19)

Author

Ramsha Abbasi

Regn Number

00000320845

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Biomedical Sciences

Thesis Supervisor:

Dr. Syed Omer Gilani

Thesis Supervisor's Signature:

_____

DEPARTMENT OF BIOMEDICAL ENGINEERING AND SCIENCES

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD, PAKISTAN

September, 2022

FORM TH-4

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: **Ramsha Abbasi** Regn No. 00000320845

Titled: **Kidney and Kidney Tumor Segmentation, 2019 (KiTS-19)** be accepted in partial fulfillment of the requirements for the award of **MS Biomedical Sciences** degree.

Examination Committee Members

1. Name: Dr. Asim Waris Signature:_____

2. Name: Dr. Adeeb Shahzad Signature:_____

3. Name: Dr. Amer Sohail Kashif Signature:_____

Supervisor's name: Dr. Syed Omer Gilani Signature:_____

Date:_____

_____

Head of Department _____

Date

COUNTERSINGED

Date:_____ Dean/Principal _____

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by NS Ramsha Abbasi (Registration No. 00000320845), of School of Mechanical and Manufacturing Engineering (SMME) has been vetted by undersigned, found complete in all respects as per NUST Statues/Regulations, is within the similarity indices limit and is accepted as partial fulfillment for the award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor: Dr. Syed Omer Gilani

Date: _____

Signature (HoD): _____

Date: _____

Signature (Dean/Principal): _____

Date: _____

Declaration

I certify that this research work titled **"*Kidney and Kidney Tumor Segmentation, 2019 (KiTS-19)*"** is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student _____

Ramsha Abbasi

Regn No. 320845

MS Biomedical Sciences

Proposed Certificate for Plagiarism

It is certified that MS Thesis Titled Kidney and Kidney Tumor Segmentation, 2019 (KiTS-19) by Ramsha Abbasi has been examined by us. We undertake the follows:

a. Thesis has significant new work/knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
b. The work presented is original and own work of the author (i.e., there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.
c. There is no fabrication of data or results which have been compiled/analyzed.
d. There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
e. The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

Name & Signature of Supervisor

Dr. Syed Omer Gilani

Signature**:** _____

Copyright Statement

Dedication

Dedicated to my grandfather **'Baba Abbasi'** for making my childhood memorable. I Love You!

Acknowledgments

My sincere thanks to my supervisor **Dr. Syed Omer Gilani**, for his useful guidance and constructive criticism needed in the completion of this research project.

I would like to thank my GCE members Dr. Asim Waris, Dr. Adeeb Shahzad, and Dr. Amer Sohail Kashif for guiding me throughout the whole research project. I am particularly obliged to Mr. Abdullah Kamran, for helping me in the completion of this project and particularly for listening to my problems and guiding me through them.

I also want to thank Mr. Mansoor Chishti for all his support, guidance and valuable discussions that helped me throughout this process.

I am also grateful to all my lab fellows and friends, Fariha, Summiya, Ujala, Zarqa, Adnan Habib, Omer Salman Khan who made my research work experience a memorable one. Special thanks to my parents and Nani Ami for their constant moral boosting, encouragement, and a keen interest in my academic achievements. I am deeply thankful to my late Nana Baba for his prayers and eternal love, you will always be in my heart.

Abstract

Computed Tomography (CT) is the most widely used imaging procedure for locating and diagnosing kidney tumors. The standard treatment for kidney tumors is surgical removal. It is important to accurately segment the kidney and its tumor for effective surgical planning. The manual segmentation of kidney tumors is time-consuming and subject to variability between different radiologists. Therefore, automatic semantic segmentation of kidney tumors using deep learning networks has become increasingly popular in the past few years. Automatic segmentation of kidney tumors is a very challenging task due to their morphological heterogenicity. This work provides the application of 3D UNet and 3D SegResNet on KiTS19 challenge data for accurate segmentation of kidney and kidney tumors. An ensembling operation was added in the end to average the predictions of all models. The proposed method is based on the MONAI framework and focuses more on training procedure rather than complex architectural modifications. The models were trained using KiTS19 training set of 210 cases for which ground truth labels were available. The training data was divided into 190:20, for training and validation respectively. We evaluated the performance of our network on KiTS19 official test set and obtained mean dice of 0.8964, 0.9724 kidney dice, and 0.8204. Our approach outperforms many submissions in terms of kidney segmentation and gives promising results for tumor segmentation. We also used a local test set of 90 cases from KiTS21 challenge to check how well our method adopts to a new dataset. It scored a mean dice of 0.9160, kidney dice of 0.9771, and 0.8550 tumor dice. The obtained results on KiTS19 official test set and local test set show that our approach is effective and can be used for organ segmentation.

Keywords: Computed Tomography, Kidney Segmentation, Tumor Segmentation, KiTS19, KiTS21, MONAI

Table of Contents

List of Figures

List of Tables

Abbreviations

CNN	Convolutional Neural Network
CT	Computed Tomography
FCN	Fully Convolutional Network
HU	Hounsfield Units
KiTS19	Kidney and Tumor Segmentation 2019
KiTS21	Kidney and Tumor Segmentation 2021
MONAI	Medical Open Network for AI
PN	Partial Nephrectomy
PPM	Pyramid Pooling Module
RN	Radical Nephrectomy
ROI	Region of Interest

Chapter 1 Introduction

1.1 Kidney and Kidney Tumor

Kidneys are 2 bean-shaped organs, located below the ribcage in right and left retroperitoneal space are an essential part of urinary system (Stevens et al., 2010, Zheng et al., 2021). The function of the kidneys is to filter blood and remove metabolic waste from the body and produce urine (Finco, 1997, Choudhary et al., 2017). Figure 1 below shows the anatomy of kidney (External and Internal view). The weight of each kidney is between 130g -150g. The kidney on the right side is located at a slightly lower position than the left kidney. Kidneys are part of upper urinary tract. The length of each kidney is approximately 12cm and the width s nearly 6cm. The blood enters the kidney by renal arteries and is filtered in glomerulus. The filtered blood leaves the kidney via left and right renal veins (Mahadevan, 2019).

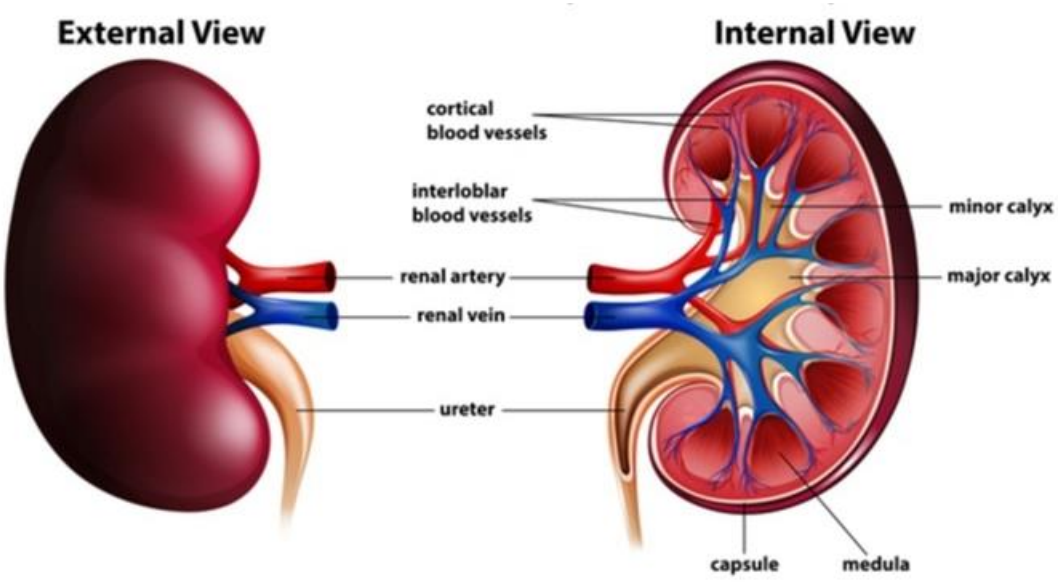


Figure 1: This figure shows anatomy of kidney (internal and external view). The indented medial surface of kidney is called the hilum. Blood vessels, autonomic nerves and lymphatics enter and exit the kidney at the hilum. The hilum is also the point of emergence of the ureter from the kidney.

Kidney cancer or tumor is the uncontrolled growth of masses in kidney. Some of kidney masses are benign and some are malignant (L.B. da Cruz et al., 2022). Figure 2 shows a healthy kidney and a effected kidney. The figure clearly shows the growth of unwanted cells in effected kidney.

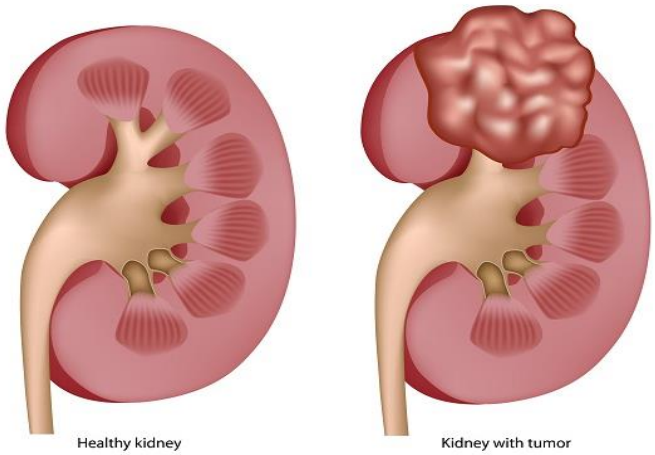


Figure 2: This figure shows the difference between a healthy kidney and a kidney with tumor. Effected kidney shows a clear growth of unwanted cells which form a mass called tumor

1.2 Incidence of Kidney Cancer

In 2020, the annual worldwide prevalence of new kidney cancer cases was reported to be more than 430, 000 causing approximately 179, 000 deaths(Sung et al., 2021). It is the 10th most common cancer diagnosed in women and the 6th among men. Men are at a greater risk of being diagnosed with high-grade large kidney tumors having worst oncological outcomes than women (Capitanio and Montorsi, 2016; Mancini et al., 2020). The established risk factors of kidney cancer include obesity, smoking, hypertension, and chronic kidney diseases (Scelo and Larose, 2018). The incidence of kidney tumors, also known as Renal Cell Carcinoma (RCC) in medical terminology, has shown an increasing trend within the last decade due to improved imaging techniques resulting in early-stage tumor detection (Kowalewski et al., 2022). However, more than 50 percent of renal tumors are diagnosed incidentally. The development of renal cancers in the human body (early to the advanced stage) is determined by general symptoms such as blood pressure, weight

loss, and anemia. This is the reason that majority of times kidney tumors are diagnosed by chance when abdominal imaging is being performed for other medical disorders (Capitanio and Montorsi, 2016; Capitanio et al., 2019; Cinque et al., 2021).

1.3 Diagnosis and Treatment of Kidney Tumors

Although renal tumors are frequently detected incidentally with abdominal ultrasound imaging, the gold standard for renal tumor diagnosis is Computed Tomography (CT) scans (van Oostenbrugge et al., 2018). This incidental diagnosis of renal masses has contributed to an increased survival rate of the disease as the tumors diagnosed are often small and localized at the time of treatment (Heller et al., 2019). The standard treatment for kidney tumors is surgical removal. There are two surgical approaches: Partial Nephrectomy (PN) and Radical Nephrectomy (RN). Nephrectomy means the removal of the kidney. RN is considered the traditional approach which involves the removal of both, the tumor, and the affected kidney whereas PN refers to the removal of the part affected by the tumor.

1.4 Role of Deep Learning in Diagnosis and Treatment of Kidney Tumors

With advancements in imaging technologies and the increasing incidence of early diagnosis of localized small tumors, PN is now being considered a standard surgical treatment for small renal tumors however, RN is still a common approach for large tumors. PN is less invasive as compared to RN, and it also preserves renal function (Mir et al., 2017; Graham-Knight et al., 2019; Heller et al., 2019). A complete understanding of the extent to which the kidney is affected, and the exact location, size, and shape of the tumor is required for effective planning (RN or PN) and evaluation process of surgery (Santini et al., 2019; Kumaraswamy et al., 2020). In this context, accurate detection of renal tumors and their morphological characteristics are of great importance to a radiologist for effective preoperative planning (Hou et al., 2020). Currently, medical images are evaluated manually by physicians. The manual examination of CT images is very tiring and time-consuming. Also, the manual annotation of tumors varies from one radiologist to another due to its morphological heterogenicity. Therefore, automatic computer aided semantic segmentation

of kidney tumors is essential to reduce this subjective disparity and workload to make an accurate diagnosis in less time (Kumaraswamy et al., 2020; Rajendran et al., 2022; Hsiao et al., 2022).

1.4 Kidney and Kidney Tumor Segmentation Challenge 2019 (KiTS19)

Automatic segmentation of renal tumors is a challenging task due to the high variability in the tumor's morphological properties and location. Semantic segmentation is a prominent research field in biomedical image analysis, but its performance depends upon the availability of a large well-annotated dataset (Isensee and Maier-Hein, 2019). In 2019, to overcome this deficiency of publicly available data for renal tumors, the Kidney and Kidney Tumor Segmentation Challenge (KiTS19) was proposed. The purpose of this challenge was to stimulate the development of accurate kidney and kidney tumor segmentation algorithms. Recent advances in Convolutional Neural Networks (CNNs) have shown phenomenal performance in the field of biomedical image classification and segmentation. CNNs have shown better performance in comparison to traditional methods (Rao et al., 2015; Guo et al., 2019; Zhang et al., 2020).

1.6 Research Objective

Despite the great success of deep learning frameworks in segmentation methods, their application to relevant image analysis tasks for end-users is quite limited. There is a huge number of published papers that proposed novel architectural modifications and extensions that improve performance. However, this requires a high level of understanding and experience and really complicates things for a non-expert and sometimes it even becomes hard for an expert to evaluate these studies (Litjens et al., 2017; Isensee et al., 2019). Therefore, we followed a relatively simple approach without any architectural variations and focused more on the training workflow and achieved promising results.

Chapter 2 Related Work

In the literature, many researchers have proposed methods for kidney segmentation. Yang et al., 2018 was the first one to use CNNs for the segmentation of kidney and tumor and achieved a dice score of 0.931 for kidney and 0.820 for tumor. Before KiTS19 challenge, there was no proper publicly available labeled dataset for kidney tumors. 100 teams from five continents participated in KiTS19 challenge. Most submissions used Deep 3D CNNs. Teams had an average kidney dice of 0.915 and average tumor dice of 0.580 (Heller et al., 2021).

Isensee and Maier-Hein (2019) scored first position on challenge leader board. They used KiTS19 training data to perform 5-fold cross validation on 3D UNet and its variants. The 3 networks trained in this work were i) plain 3D U-Net, ii) Residual 3D U-Net and iii) pre-activation Residual 3D U-Net. A input patch size of 80x160x160 was used and models were trained at 1000 epochs. They applied strong data augmentation techniques using batch generators. Each model took five days to complete training. Residual 3D U-Net gave the best results than the other two models and even better than the ensemble of three. As a result, 3D Residual U-Net was used to make final submission on leaderboard. They scored a mean dice of 0.912, kidney Dice 0.974 and 0.851 tumor dice.

Hou et al. (2019) won the second place in this competition. They used a cascaded approach. Their pipeline consisted of 3 stages. In first stage, lightweight 3D UNet was used to perform coarse localization of kidneys. In second stage high-resolution 3D Unet is used to crop VOI and get accurate localization of kidneys (kidney and tumor treated as same label). In final stage, Fully Convolutional Net is used to segment both kidney and tumor. They applied post processing to fill holes and remove false positives. They obtained a kidney dice of 0.967, 0.845 tumor dice. The mean dice was 0.906.

Mu et al. (2019) obtained third position in this challenge. To automatically segment kidney and kidney tumor, they proposed a multi-resolution 3D V-Net Network. They customized the V-Net model using two resolutions and termed it as VB-Net. In coarse resolution, VB-

Net can robustly localize the organ whereas in fine resolution, it can refine the boundaries of each organ accurately. They used PyTorch framework to train their model. Connected component analysis was used as post-processing step to remove false positives. Their submission scored a mean dice of 0.903, kidney dice 0.973 and tumor dice 0.832.

Zhang et al. (2019) won the 4th place in this competition. They used 2-stage approach to segment kidney and kidney tumor. In first stage they used 3D UNet to perform coarse localization of kidney. After localization, 3D volume patches were cropped and passed to second stage for fine segmentation. In second stage kidney and tumor were segmented and ensembling was performed. They used connected component algorithm as a pos-processing step. They scored a kidney dice of 0.974, 0.831 tumor Dice and a mean dice of 0.902.

Ma, 2019 used 3D UNet as the main network to segment the kidney and crop ROI and then in the second stage another 3D UNet is used to segment kidney and kidney tumor from cropped ROIs. Test time augmentations were applied, and predictions were averaged. They used a heuristic algorithm to remove any false positives. They achieved 0.973 and 0.825 kidney and tumor dice respectively. The mean dice was calculated to be 0.899

Figure 3 shows an overview of the top five submissions in KiTS19 challenge. All five submission did not make use of any external data other than KiTS19 training set of 210 cases and evaluated their models on 90 test cases.

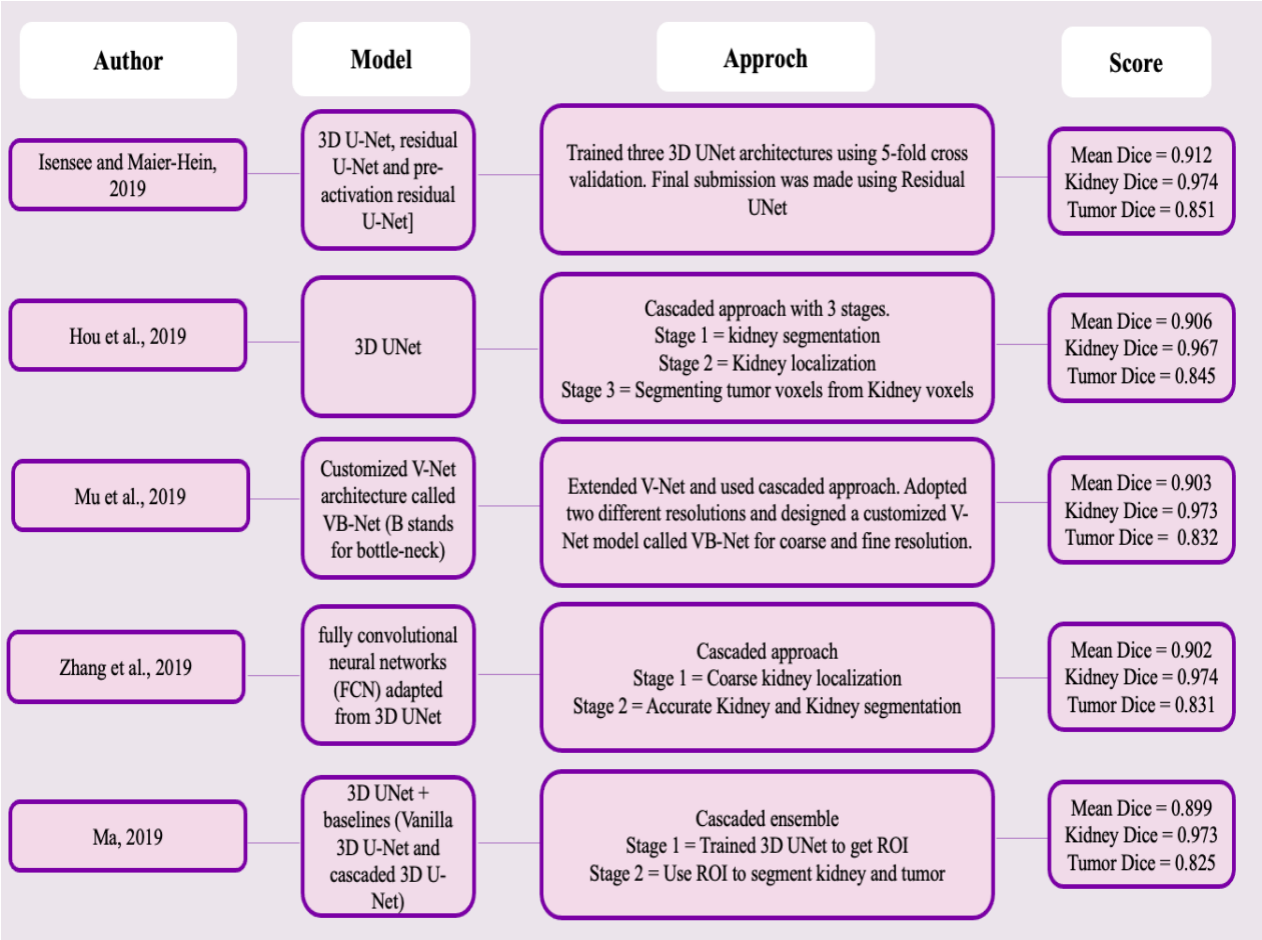


Figure 3: Overview of top five submissions of KiTS19 challenge phase

Chapter 3 Methodology

Our method is based on Medical Open Network for AI (MONAI) framework. MONAI is an open-source, user-friendly, PyTorch-based framework developed for deep learning in medical image analysis. It is an easy-to-use API interface. Figure 4 shows the general workflow of our proposed method.

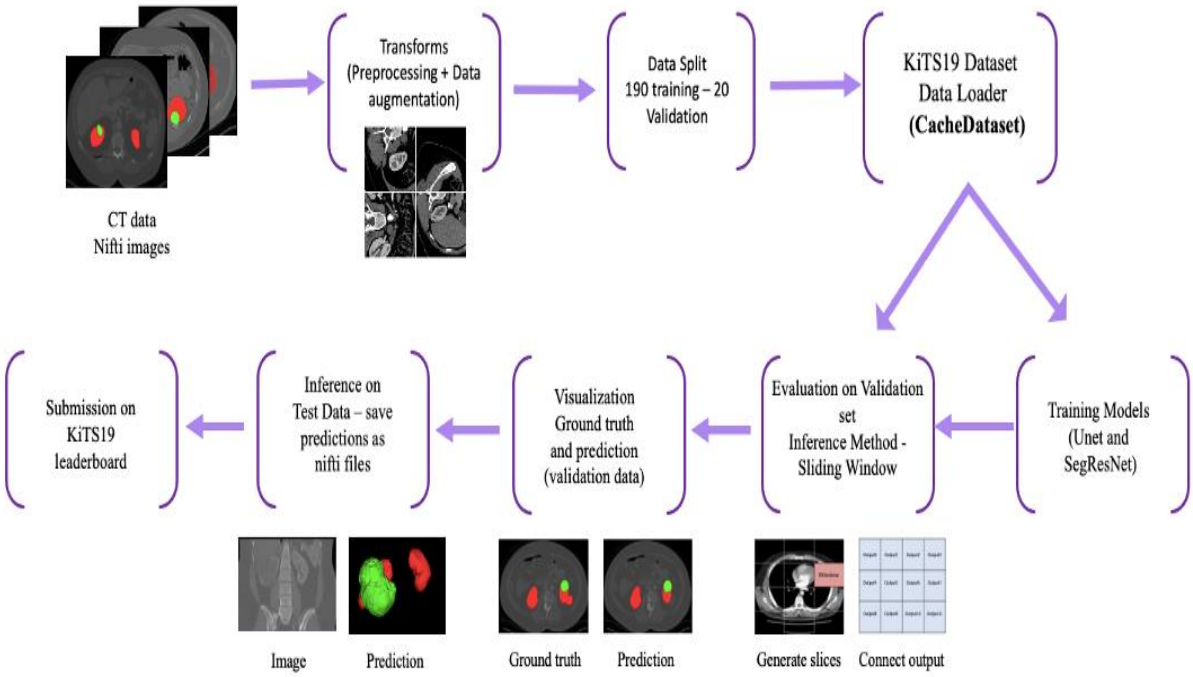


Figure 4: General workflow of proposed methodology. CT data is preprocessed and divided in to training and validation set. Data is loaded using CacheDataset

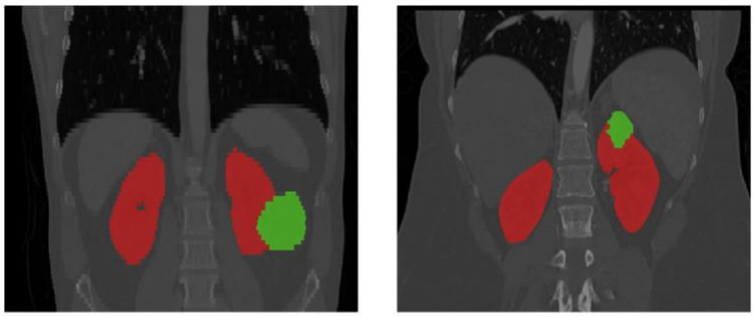
3.1 Dataset

This section describes the datasets used in this research for training and testing.

3.1.1 KiTS19 Dataset

In this research work, we have used the official KiTS19 challenge database for training and validation purposes. The dataset consists of 300 CT scans, 210 of which are publicly available with their corresponding ground truth labels. The remaining 90 CT scans, for which ground truth labels are kept private, are used for objective model evaluation. The

data was downloaded from the official KiTS19 repository. All the data (imaging and ground truth) is provided in NIFTI format. The original resolution of CT scans is 512x512 and the number of slices varies from 29 to 1059. Figure 5 shows scans of two patients from KiTS19 training dataset (red represents kidney and green represents tumor).



Case_00017 Case_00059

Figure 5: Scans of two patients from KiTS19 dataset with their corresponding ground truth labels. Red represents Kidney and green represents tumor

We divided the training dataset of 210 CT scans into 190 and 20 for training and validation respectively. The datatypes of imaging.nii and segmentation.nii are np.float32 and np.uint8 respectively. In the segmentation labels, the background is labeled as 0, Kidney as 1, and tumor as 2. We evaluated our model on KiTS19 official test set of 90 scans.

3.2.2 KiTS21 Dataset

KiTS21 is the second version of the Kidney Tumor Segmentation challenge held in 2021. The dataset is publicly available. The training set consists of 300 CT scans with their corresponding ground truth labels. The semantic classes in KiTS21 are Kidney, Tumor (defined in the same way as KiTS19), and cyst (Kidney masses if available). We only used the labels defined in the same way as in KiTS19. We randomly selected 150 cases from KiTS21 dataset. 90 cases were used as a local test set for model evaluation and the remaining 60 cases were combined with KiTS19 training data to check the effect of increasing training data on the model's performance. Aggregated_MAJ based mask were used for training. Figure 5 shows cases from the KiTS21 dataset with kidney and tumor labels.

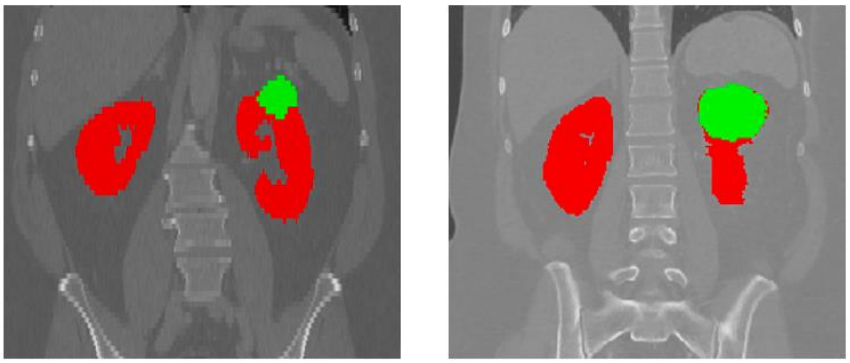
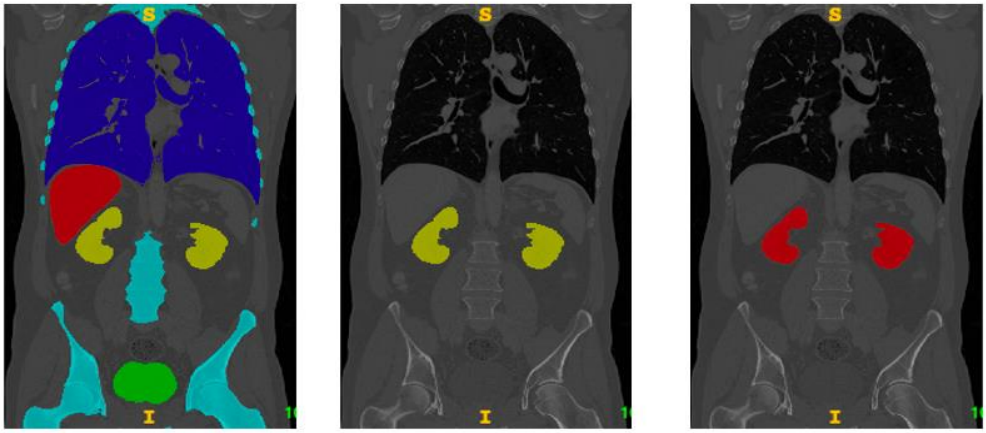


Figure 6: Scans of patients from KiTS21 dataset overlapped with their corresponding ground truth segmentation mask

3.1.3 CT-ORG Multi Organ Dataset

CT-ORG (Rister et al., 2019) is a publicly available multi-organ labeled dataset. It consists of 140 CT scans with 5 organs (bladder, liver, bones, lungs, and kidneys). We randomly selected 20 cases and only used kidney labels to evaluate how well our model segments kidneys from unseen data. ITK-SNAP label editor was used to remove the other 4 labels. Figure 6 shows a kidney label from a CT scan of CT-ORG dataset.



(a) (b) (c)

Figure 7: Kidney label from CT-ORG dataset, (a) shows CT scan with all five labeled organs in CT-ORG dataset. (b) shows the kidney label after removal of other four labels using ITK-SNAP editor. (c) shows the final kidney label in red color as KiTS19.

3.2 Preprocessing

We changed the orientation of CT scans from SAR (Superior, Inferior), (Anterior, Posterior), (Right, Left) to RAS (Right, Left), (Anterior, Posterior), (Superior, Inferior). It is common for publicly available large datasets to have different voxel spacings. As it is difficult for CNNs to understand the voxel spacings natively, we resampled all the input data to common voxel spacing of 1.62x1.62x2 mm (x, y, and z direction). The reason behind setting a lower voxel value on the z- axis is to ensure that a greater number of training slices are available per patient to increase generalizability. Figure 8 shows the effect of these applied transforms.

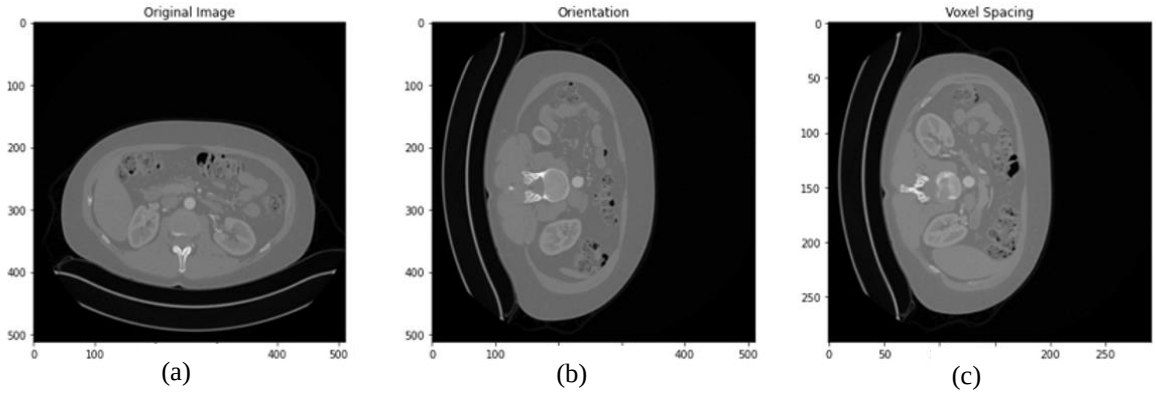
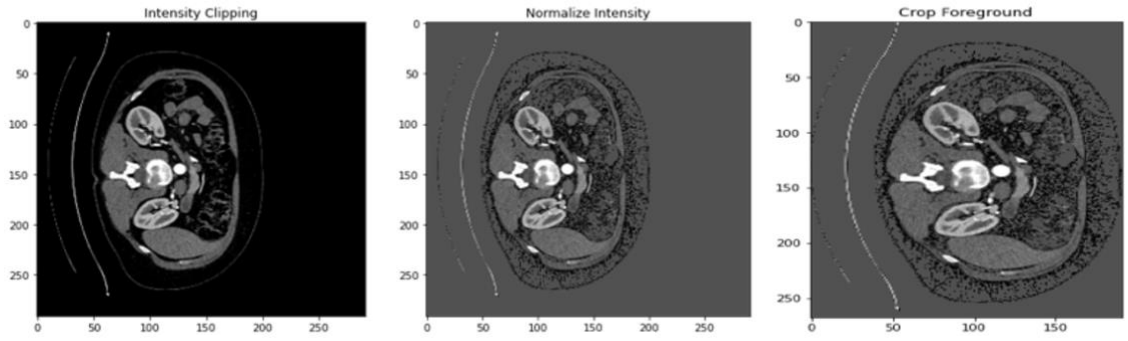


Figure 8: (a) shows the original CT scan from KiTS19 dataset, (b) shows the effect of changing orientation from SAR to RAS, (c) shows the same scan after resampling to same voxel size

The intensity values vary significantly in CT scans, to enhance the contrast of soft tissues and to remove the fat regions that surround the kidney, we clipped the intensity values in the range [-100, 300] HU and rescaled them between 0 and 1. The clipped images were then normalized in the range [-1, 1]. Input images were further cleaned by cropping the foreground. The cropping is based on the value of 0. The area of the image having a value of 0 (no organ present) is cropped. Figure 9 shows the effect of intensity clipping, normalization, and cropping foreground on the same slice.



(a) (b) (c)

Figure 9: (a) Intensity clipping, (b) Normalize Intensity and (c) Crop Foreground

3.3 Data Augmentation

Manual annotation of large 3D images is a time-consuming and difficult process due to which the availability of labeled data is often limited. Therefore, we used strong data augmentation to increase the variation in our data and avoid overfitting during training.

The original resolution of CT scans was 512x512xn (n is the number of slices), which was not suitable as an input size due to the limitation of computational resources. Therefore, we randomly cropped the data into 4 patches of 96x96x64 to capture high contextual information and reduce GPU consumption. This was only done for training data, in validation, original resolution images were fed into the model. Figure 10 shows random cropped patches of the input image from KiTS19 training dataset.

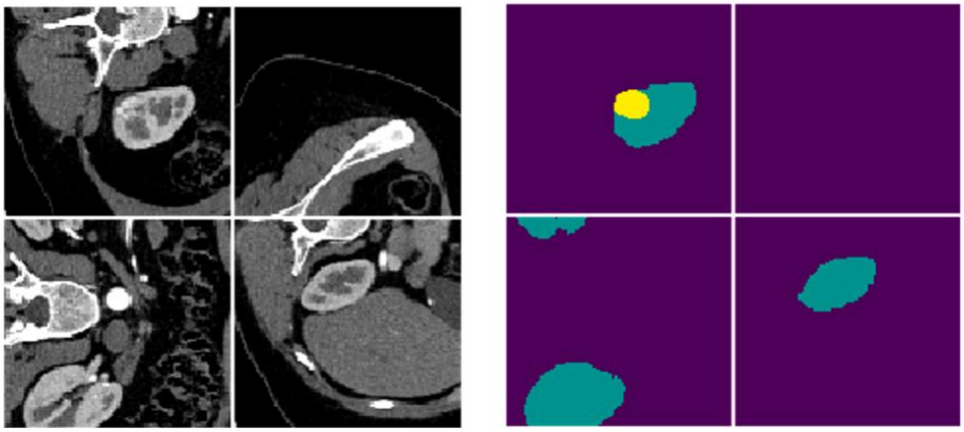
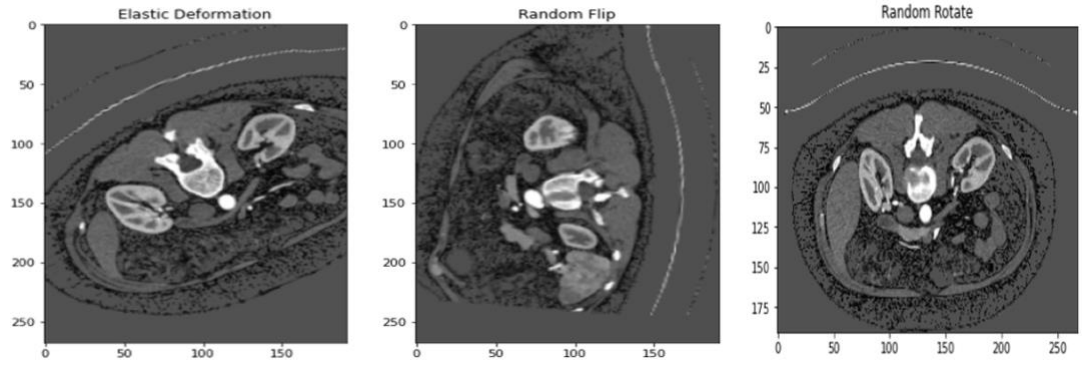


Figure 10: This figure shows the randomly cropped four patches of the input CT image and its corresponding ground truth segmentation mask. The size of each cropped patch was 96x96x64 for all the models

We used 3D Rand Elastic deformation (with 0.5 probability), Rand Affine, RandRotate90, and Rand Flip (images were randomly flipped on all three axis) to introduce spatial anatomical variations in data shape while preserving the spatial information. For variations in intensities of CT scans, we applied RandScaleIntensity with the probability of 0.25 and RandShiftIntensity setting the maximum offset value to 0.1. We also introduced random gaussian noise augmentations. Figure 11 shows data augmentation transforms applied on a single image and Figure 12 shows the effect of all applied transforms (preprocessing and data augmentation) mentioned above on a single case.



(a) (b) (c)

Figure 11: This figure shows the effect of data augmentation transforms. (a) Elastic deformation, (b) Random flip and (c) Random rotate

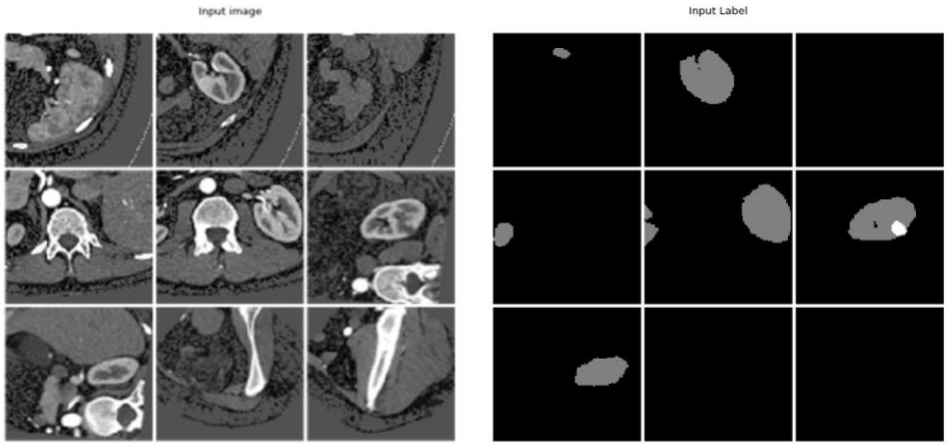


Figure 12: This figure shows the effect of all applied preprocessing and data augmentation transforms on single image

3.4 Training Network

KiTS19 dataset was used for training purposes. To accurately segment the kidney and its tumor we performed a segmentation model ensemble. The two models used are described below:

3.4.1 3D UNet

We used Monai's enhanced UNet architecture. It supports residual units that are implemented with residual units class. The purpose of using a convolution in residual part is to match the input dimensions with the output dimensions. Our network has five layers each with a encode decode path with a skip connection between them. A stride value of two is used for each middle layer. In encode path, data is downsampled using strided convolutions and is then upsampled in the decode path using strided transpose convolution. All encoding and decoding blocks use a kernel size of 3x3x3. We used batch normalization. The number of feature maps are 32, 64, 128, 256, and 512. Figure 13 shows the 3D UNet architecture used in our work. The total trainable parameters of 3D UNet are 19223664.

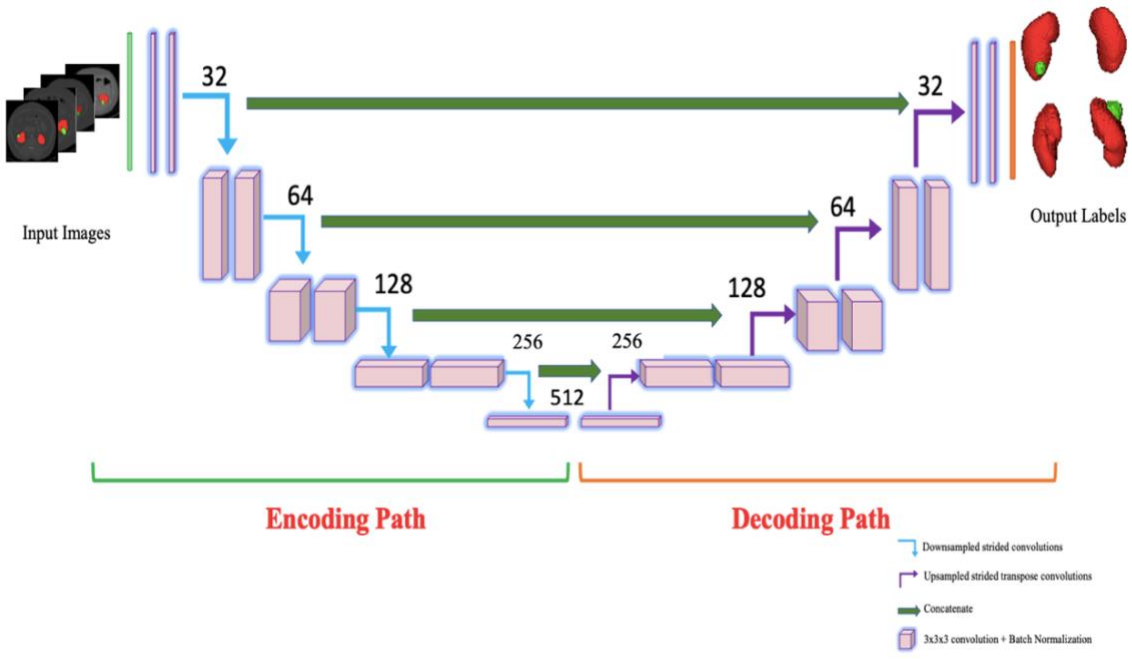


Figure 13: The figure shows 3D UNet architecture used in proposed work. The number of feature maps are 32, 64, 128, 256 and 512. Data is downsampled in encode path, using strided convolutions and is then upsampled in the decode path using strided transpose convolution

3.4.2 3D SegResNet

The second model used is SegResNet (Myronenko, 2018). It is an encoder-decoder-based CNN architecture. In the encoder part, ResNet (He et al., 2016) blocks are used. Each block consists of two convolutions with normalization and ReLU, followed by an additive identity skip connection. We used batch normalization. A regular CNN approach of progressive downsizing image dimensions by 2 and simultaneously increasing feature size by 2 was used as shown in Figure 14. The decoder part is almost like the encoder part, the only difference is that it uses a single block per each spatial level. In the end, the decoder has the same spatial size as the original data, with the same number of features as the initial input feature size followed by convolutions into a single channel and softmax function. The model has 4700931 trainable parameters.

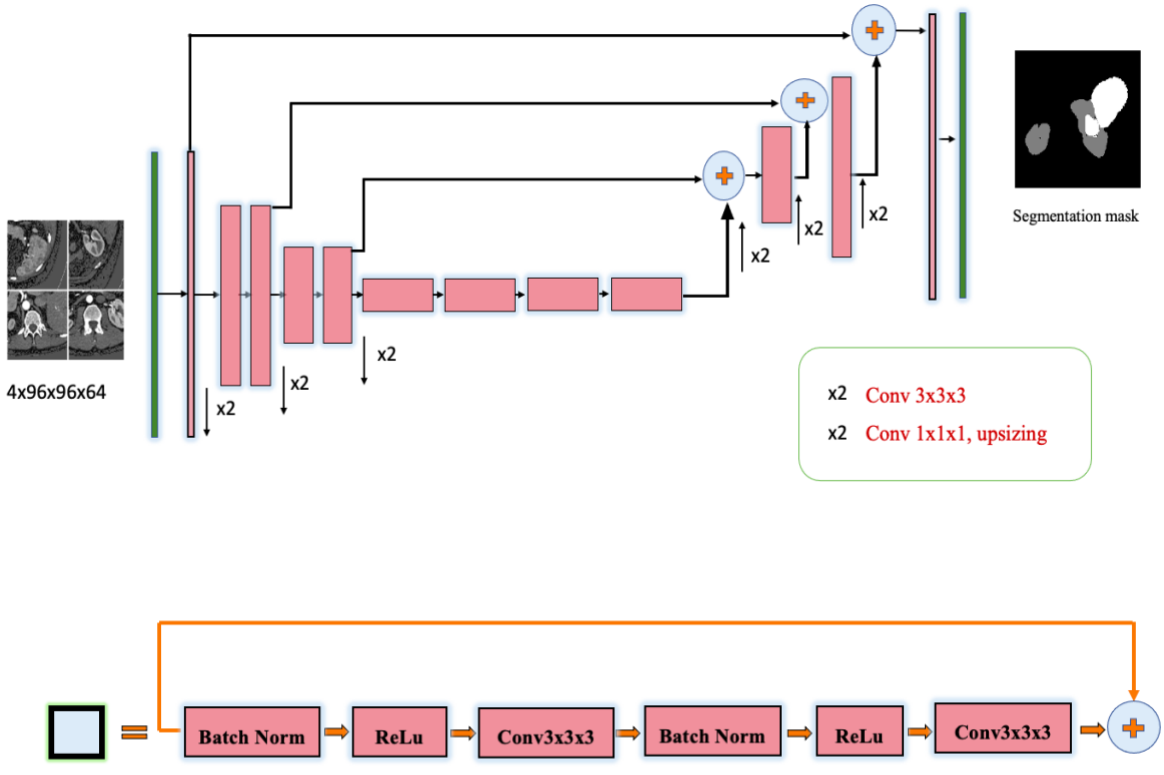


Figure 14: The figure shows the 3D SegResNet architecture used in proposed work

3.5. Training Protocol

We performed a segmentation ensemble of two different models mentioned above. To achieve better results and avoid overfitting we trained UNet and SegResNet models three times but with a separate validation set each time. Figure 15 shows the general training protocol. Pre-processing of medical images requires additional detailed parameters, so we used MONAI's transforms to convert CT scans from KiTS19 in python dictionaries. We used LoadNiftid transform to load the CT scans and applied various pre-processing transforms such as Orientationd, spacingd, Clipping and Normalization. We used extensive data augmentation transforms such as RandCropByPosNegLabeld to crop input image in 4 patches of 96x96x64. Other transforms include Elastic Deformation, Flip, Rotate, addition of Gaussian noise and adjust Contrast. MONAI's CacheDataset was used to load data, it allows multi-threaded processing. We used a cache rate of 1. The transformed data is cached before training which greatly reduces the time in loading data. Data was shuffled before each training and was split randomly into 190:20 (190 training and 20 validation). The models were trained for 600 epochs using a batch size of 2. The activation function used was Softmax. The predictions generated from validation set are compared with their ground truth labels and dice scores are computed. After 600 epochs, when the training is completed a .pth file of best model is generated which is used for inference on test data. The details of inference, evaluation metrics and implementation details are discussed below.

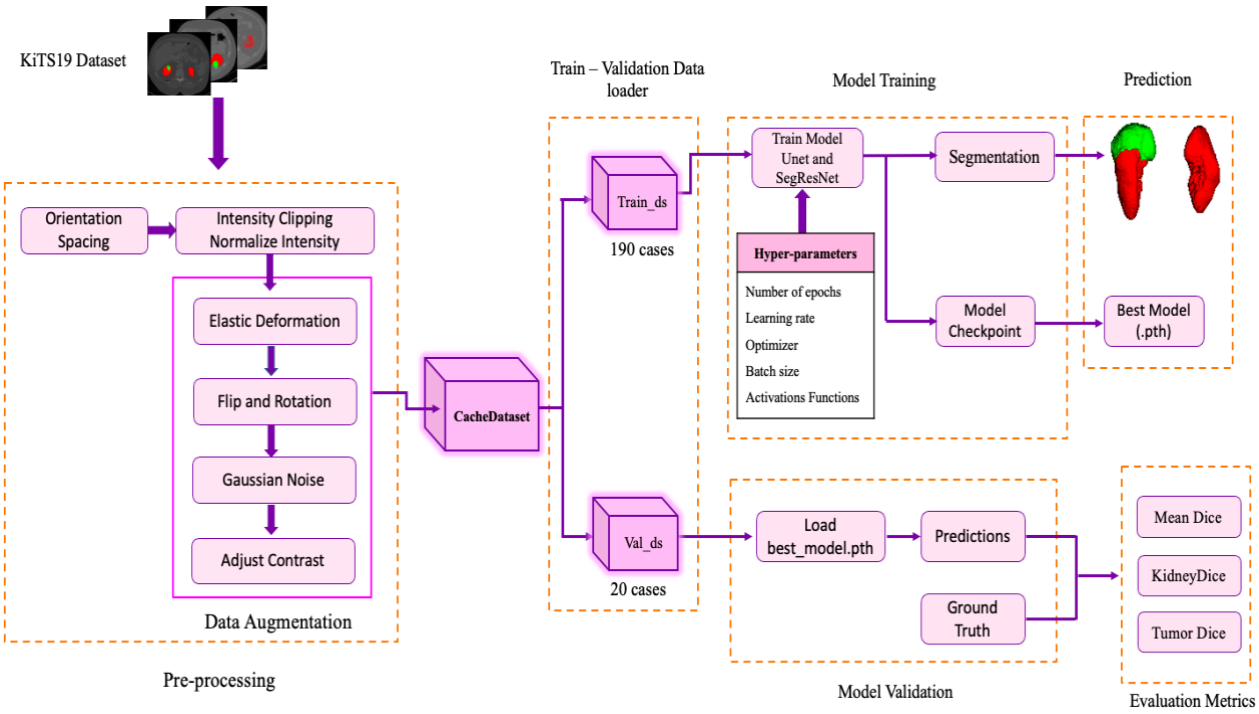


Figure 15: Overview of Training protocol

3.6. Implementation Details

All models have been implemented in MONAI open-source framework. The output of the models is 3 channels corresponding to 3 classes (background, kidney, and tumor). We used DiceCELoss which is a combination of two loss functions, Dice loss, and Cross Entropy Loss. A learning rate of 10e-4 was used for training both models with AdamW optimizer. The Models were trained for 600 epochs using a batch size of 2 with NVIDIA Tesla T4 GPU with 16GB memory and RAM. For UNet, each epoch took 2.5 minutes and training was completed in 26 hours. In the case of SegResNet, each epoch took 3.5 minutes, and the model completed 600 epochs in 37 hours. At the end of each epoch, validation was performed using sliding window inference with batch size 1, and the dice metric was computed. Table 1 shows the details of specifications of the environment used.

Table 1: Specifications of the environment

Programming language	Python 3.8.8
Ubuntu version	20.04.4 LTS
RAM	32GB
GPU	16GB
CUDA version	11.4
Deep learning Framework	PyTorch 1.9.0 and MONAI 0.8

3.7 Evaluation Metrics

The evaluation metric used in the KiTS19 challenge was the average Sorensen Dice coefficient between kidney and its tumor on 90 test cases for which ground truth labels are kept private. The kidney dice score was computed by treating kidney and tumor labels as foreground and everything else as background. For tumor dice, only the tumor label was used to compute the dice score.

$$Dicecoefficient = \frac{2TP}{2TP + FN + FP}$$

TP = True Positive
FP = False Positive
FN = False Negative

3.8. Inference on Test Data

After training, the best model is used to generate predictions of 90 test cases. For inference, we used sliding window inference with an overlap of 0.8 and batch size 4. In post-processing, the 3-channel output is converted back to a single channel as the original image, and the voxel spacing is restored to the original spacing using inverse transforms. Lastly, the SaveImage transform was used to save the predicted mask in NIFTI (.nii.gz) format. For evaluation, we submitted our predictions on the KiTS19 challenge leader board. Figure 16 and Figure 17 below describes the inference method. The process shown in Figure 16

was performed for both networks individually. Predictions of all six models were generated using KiTS19 test set. The score of each model's predictions was checked on KiTS19 leaderboard. We used Monai's MeanEnsemble and EnsembleEvaluator to average the predictions for 3 UNet and 3 SegResNet models and check their score on leaderboard. Finally, MeanEnsemble was used to take an average of predictions by all 6 models as shown in Figure 17. The results are described in the next section.

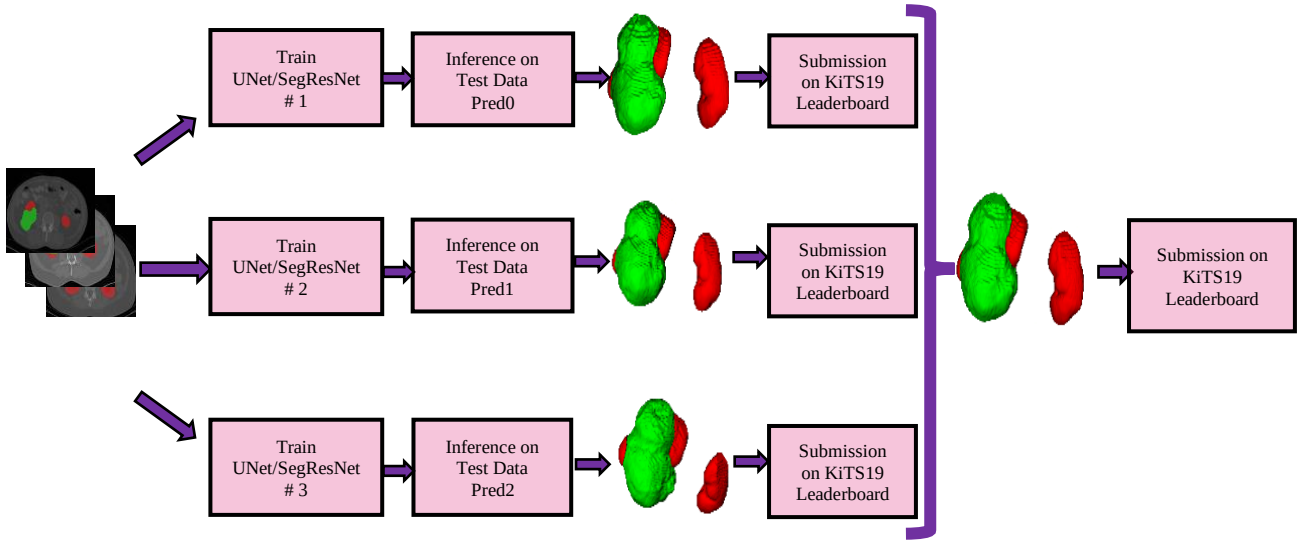


Figure 16: This figure shows the Inference and submission method used in this work. The process shown in this figure is first performed on UNet and then SegResNet. All six models are used to generate predictions on 90 test cases. Predictions for each model are submitted on KiTS19 leaderboard. Finally, MeanEnsemble is used to average predictions of 3 UNets and 3 SegResNets and their score is obtained by submitting predictions on KiTS19 leaderboard

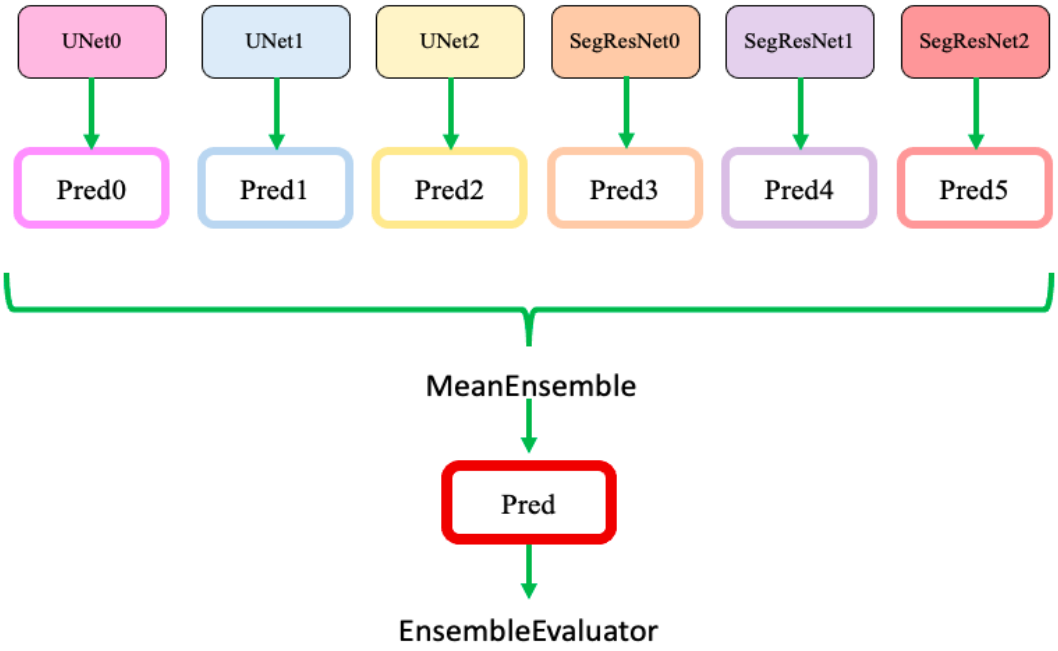


Figure 17: This Figure shows the process of averaging predictions of all six models using Monai's MeanEnsemble. The average of all 6 models was used to make a final submission on leaderboard

Chapter 4 Results

As the deadline has passed, the challenge has entered an open leader board phase. Models were trained on 190 cases and validated on the remaining 20 cases. Figure 18 shows the predictions of validation cases compared with their ground truths. Finally, submissions were made on KiTS19 open phase leader board to evaluate our model's performance on unseen test data. The metrics for evaluation are Mean Dice, Kidney Dice, and Tumor Dice.

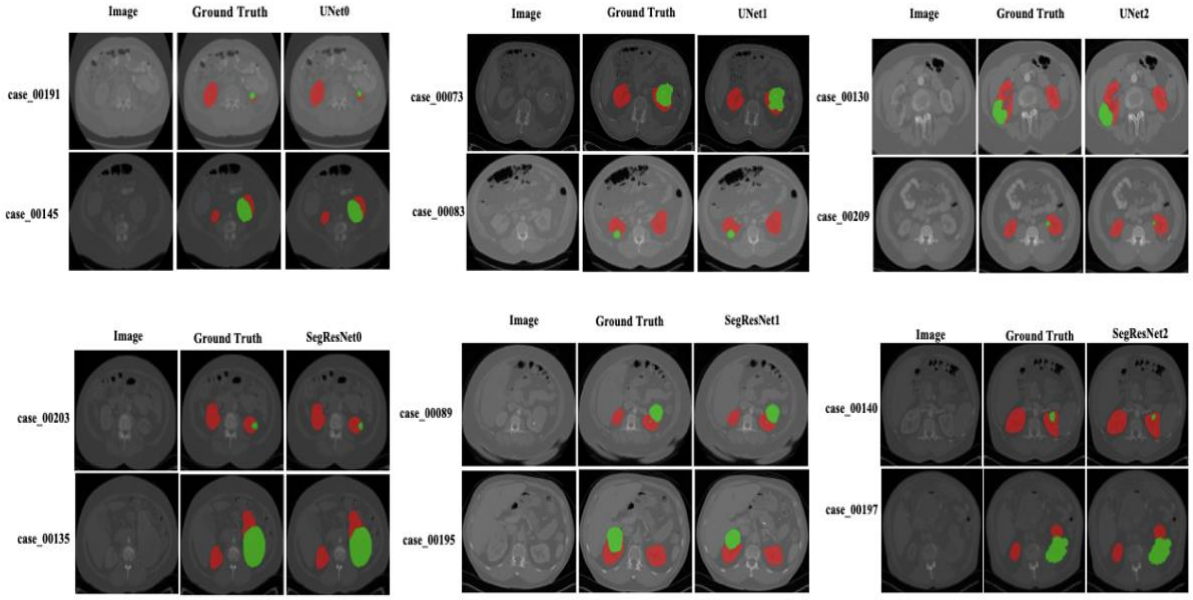


Figure 18: This figure shows the comparison of predictions of validation set generated from all six models with their corresponding ground truth segmentation labels

4.1. Results on KiTS19 Test set

This section provides results of our approach on KiTS19 official test set of 90 cases.

4.1.1. 3D Unet

The results of UNet on test data are shown in Table 2. It can be observed that tumor dice have significantly improved after combining predictions of all 3 UNet models. Figure 19 shows the predictions generated by all three UNets and their ensemble.

Table 2: Results of UNet on KiTS19 Test Set

Models	Mean Dice	Kidney Dice	Tumor Dice
UNet0	0.8769	0.9675	0.7863
UNet1	0.8735	0.9577	0.7892
UNet2	0.8771	0.9644	0.7898
Ensemble	**0.8913**	**0.9690**	**0.8136**

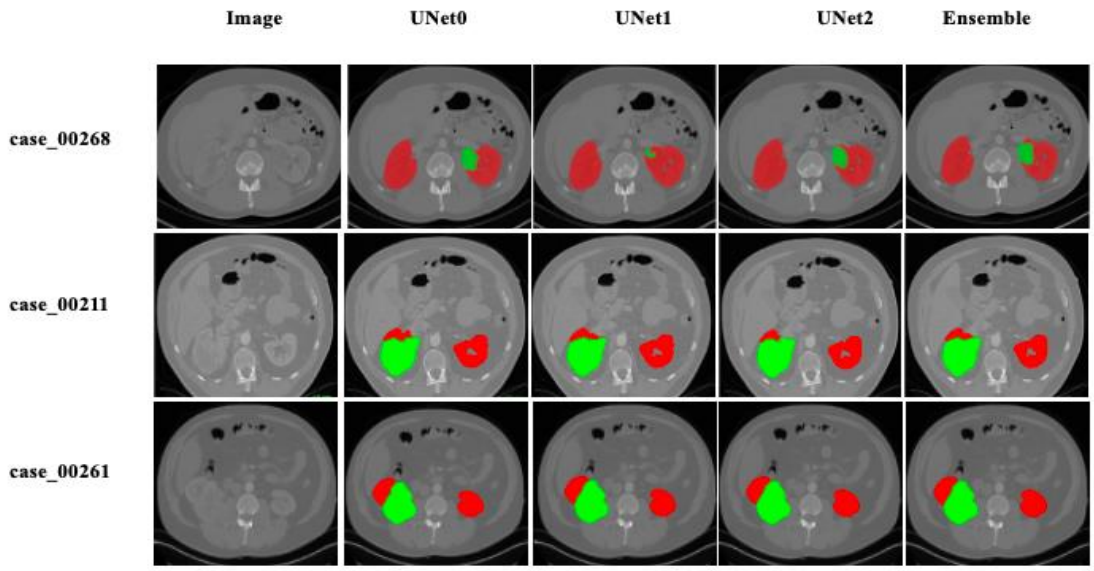


Figure 19: This figure shows the comparison of Predictions generated by all three UNet models and their ensemble from KiTS19 official test data

4.1.2. 3D SegResNet

The results of SegResNet on 90 test cases are shown in Table 3. SegResNet0 and SegResNet1 give a significantly better dice score for tumors but the score of the ensemble is affected by the performance of SegResNet2. Figure 20 shows the comparison of predictions generated by SegResNet models.

Table 3: Results of SegResNet on KiTS19 Test Set

Models	Mean Dice	Kidney Dice	Tumor Dice
SegResNet0	0.8841	0.9692	0.7989
SegResNet 1	0.8804	0.9679	0.7929
SegResNet 2	0.8619	0.9607	0.7631
Ensemble	**0.8919**	**0.9707**	**0.8130**

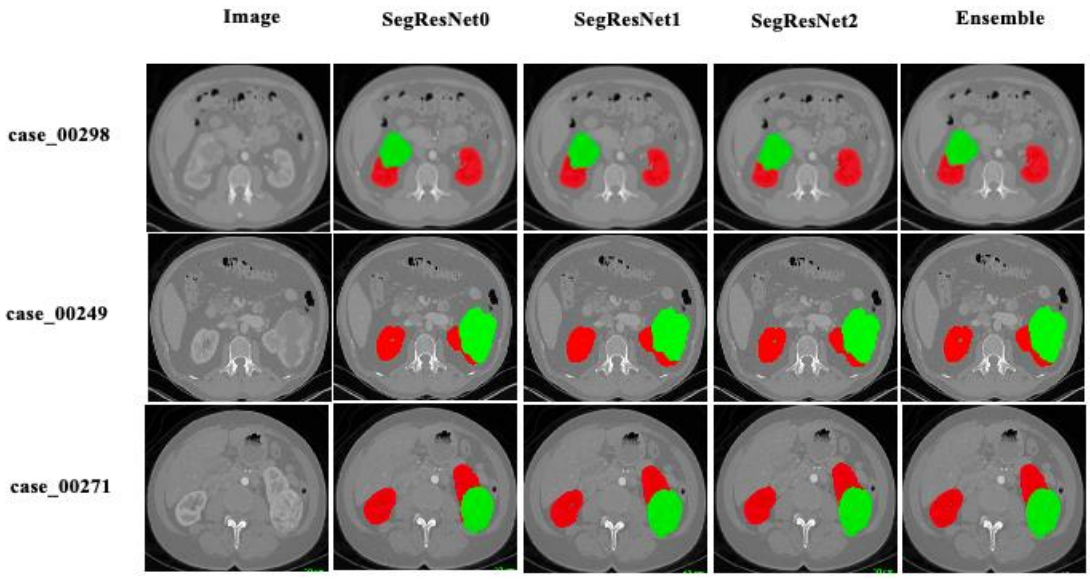


Figure 20: This figure shows the comparison of Predictions generated by all three SegResNet models and their ensemble from KiTS19 official test data

4.1.3 Ensemble of UNet and SegResNet

Table 4 shows the results of our final submission on the KiTS19 leader board. The final submission included the predictions generated by an ensemble of 3 UNet and 3 SegResNet models. Our submission on the online leader board scored a mean dice of 0.8964. The score for kidney and tumor was 0.9724 and 0.8204 respectively. Figure 21 shows the comparison of predictions generated by mean of UNet, SegResNet and their ensemble. Figure 22 shows the training Dice loss of all six models trained in this work.

Table 4: Results of Ensemble on KiTS19 Test Set

Models	Mean Dice	Kidney Dice	Tumor Dice
UNet	0.8913	0.9690	0.8136
SegResNet	0.8919	0.9707	0.8130
Ensemble	**0.8964**	**0.9724**	**0.8204**

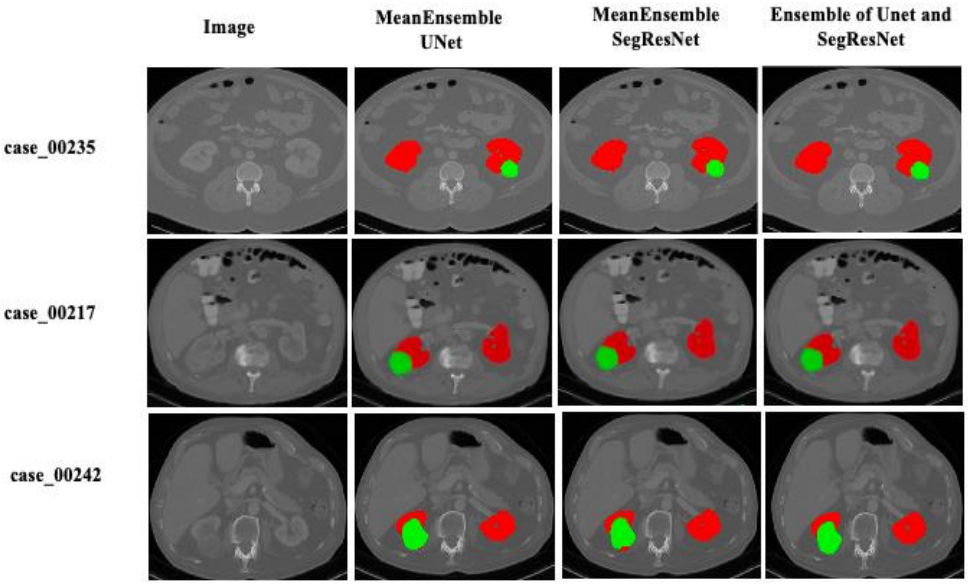


Figure 21: Predictions of Ensemble of UNet and SegResNet

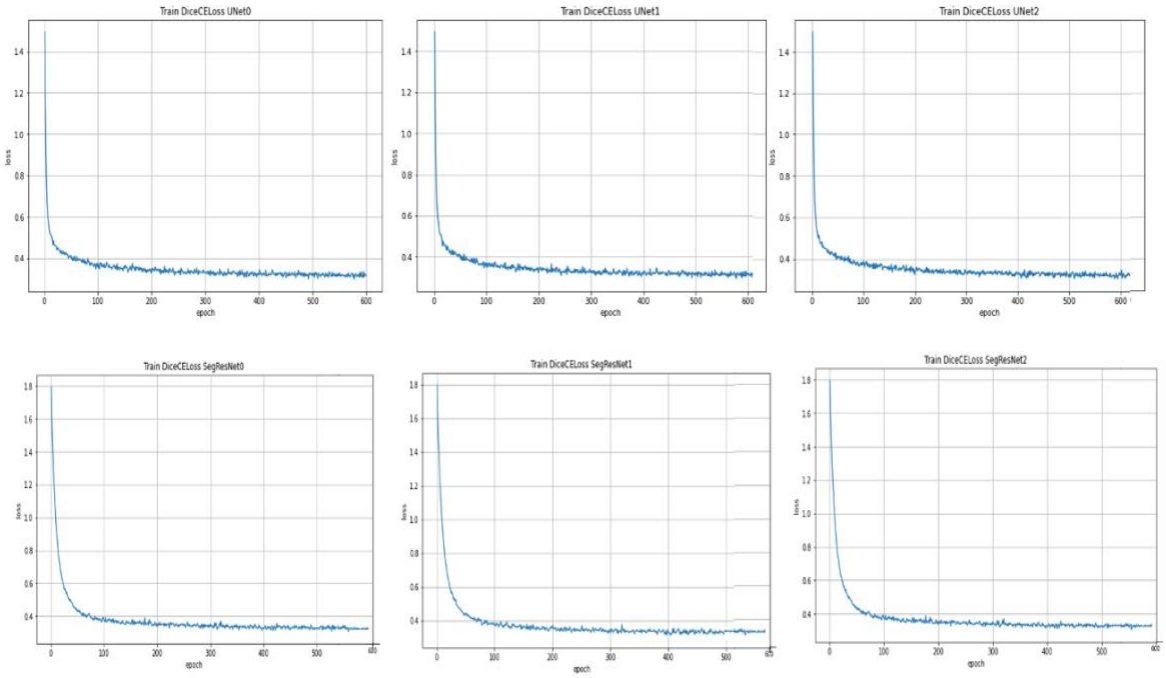


Figure 22: Training loss of all six models

Table 5 shows a general comparison of dice scores obtained from submissions made to the KiTS19 challenge phase leader board on 90 test cases with our study. It can be observed that our tumor dice outperform other submissions with a large margin. We also compared our dice scores with the model that won the competition. Although our tumor dice score is less than the winning team, our ensemble model showed a stable performance for kidney segmentation. They scored a kidney dice of 0.9737 whereas our ensemble model obtained a promising dice score of 0.9724.

Tsai and Sun (2019) used a coarse to fine semantic segmentation approach. They used ResUNet for coarse kidney segmentation and captured the Region of Interest (ROI). For fine segmentation of kidneys and tumors, they trained DenseUNet. The input image size for both models was 512x512xn. The models were trained using 4 NVIDIA Tesla V100 32GB GPU memory each using a batch size of 32. They achieved 0.9639 kidney dice, 0.7533 tumor dice, and a mean dice of 0.8586.

Santini et al. (2019) used a multistage 2.5D deep learning framework which was based upon Residual UNet. They used a 3-stage approach. In the first stage they roughly segmented ROI. Using the ROI of the first stage, kidney and cancerous tissue was segmented in the second stage. In the third stage, final segmentation was performed using the ensembling approach.They used NVIDIA GTX 1080 11GB GPU. They achieved a score of 0.825 mean dice, 0.9627 kidney dice, and 0.7424 tumor dice.

Myronenko and Hatamizadeh (2019) used an encoder- decoder-based 3D framework that was equipped with a boundary stream. It was designed to process the edge information separately and was supervised by edge-aware loss. They used an image input size of 176x176x176. Their final submission used an ensemble of 5 models with Test Time Augmentations (TTA). The model was trained using 8 NVIDIA Tesla V100 with 16GB memory each with a batch size of 8. Their approach scored kidney dice of 0.9742, tumor dice of 0.8103, and mean dice of 0.8923.

Zhao et al. (2020) used a Multi-Scale Supervised 3D UNet approach. They used the framework of a classical 3D UNet but was designed to make predictions from different layers in the decoder path, unlike the basic 3D UNet which only gives prediction in the final layer. The patch size used for training was 192x192x48 with a batch size of 8. 2 Tesla GPUs each with 32GB memory were used for training. The final predictions were improved by postprocessing. They achieved a score of 0.8961 mean dice, 0.9741 kidney dice, and 0.8181 tumor dice.

Table 5: Comparison of Kidney and Tumor Segmentation Methods on KiTS19 Test Set

Authors	Network	Mean Dice	Kidney Dice	Tumor Dice
Tsai and Sun, 2019	ResUNet and DenseUNet	0.8586	0.9639	0.7533
Santini et al., 2019	Multistage 2.5D deep learning using Residual UNet	0.8525	0.9627	0.7424
Myronenko and Hatemizadeh, 2019	Encoder Decoder with boundary stream + TTA	0.8923	**0.9742**	0.8103
Wenshuai Zhao et al., 2020	MSS-UNet	0.8961	0.9741	0.8181
Our approach	Ensemble of UNet and SegResNet	**0.8964**	0.9724	**0.8204**

4.2. Results on KiTS21 Local Test Set

We randomly selected 90 cases from the KiTS21 challenge and used it as a local test set to evaluate our model's perfor mance. We removed the cyst label from the KiTS21 dataset as it was not defined in the KiTS19 dataset. We used the same metrics defined for the KiTS19 challenge. Table 6 shows the dice scores of UNet on local test set and Figure 23 shows the comparison of predictions with ground truth labels.

Table 6: Results of UNet on Local Test Set

Models	**Mean Dice**	**Kidney Dice**	**Tumor Dice**
UNet0	0.9075	0.9750	0.8400
UNet1	0.9008	0.9571	0.8445
UNet2	0.9065	0.9727	0.8403
Ensemble	**0.9124**	**0.9701**	**0.8547**

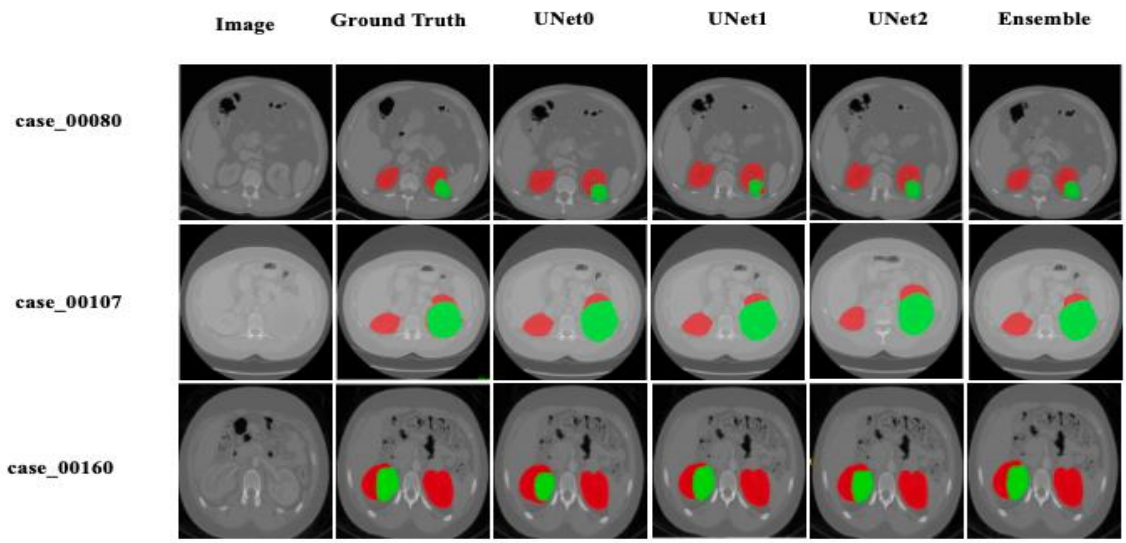


Figure 23: Comparison of Ground Truth labels and Predictions generated by UNet

Table 7 shows the scores obtained from SegResNet and Figure 24 shows the predictions generated by SegResNet and their comparison with ground truth labels.

Table 7: Results of SegResNet on Local Test Set

Models	Mean Dice	Kidney Dice	Tumor Dice
SegResNet0	0.9107	0.9784	0.8430
SegResNet 1	0.9039	0.9664	0.8414
SegResNet 2	0.8912	0.9443	0.8382
Ensemble	**0.9105**	**0.9754**	**0.8450**

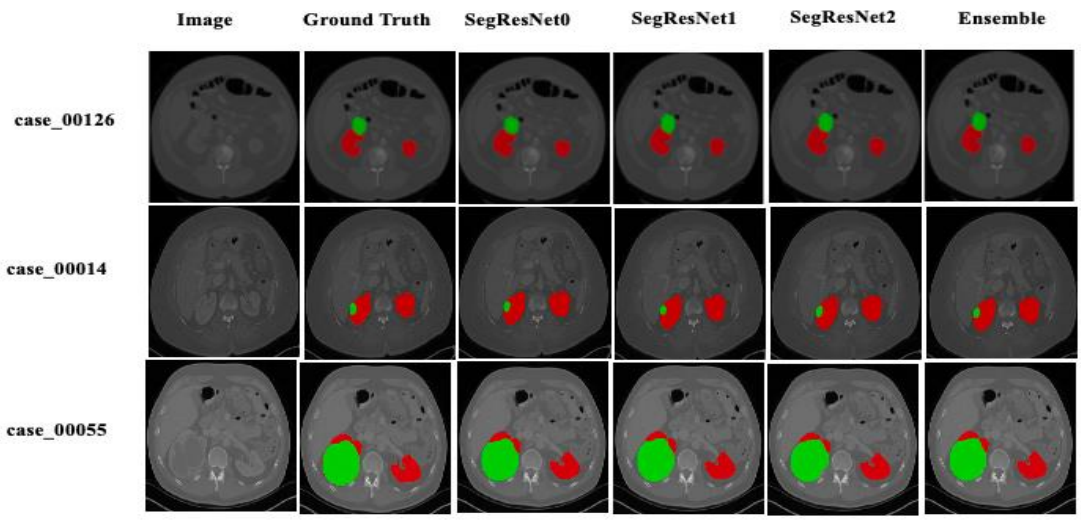


Figure 24: Comparison of ground truth and Predictions of local test set generated by SegResNet

Table 8 shows the results of the ensemble of 3 UNet and 3 SegResNet models on our local test data and Figure 25 shows the predictions and their comparison.

Table 8: Results of Ensemble on Local Test Set

Models	Mean Dice	Kidney Dice	Tumor Dice
UNet	0.9124	0.9701	0.8547
SegResNet	0.9105	0.9754	0.8450
Ensemble	**0.9160**	**0.9771**	**0.8550**

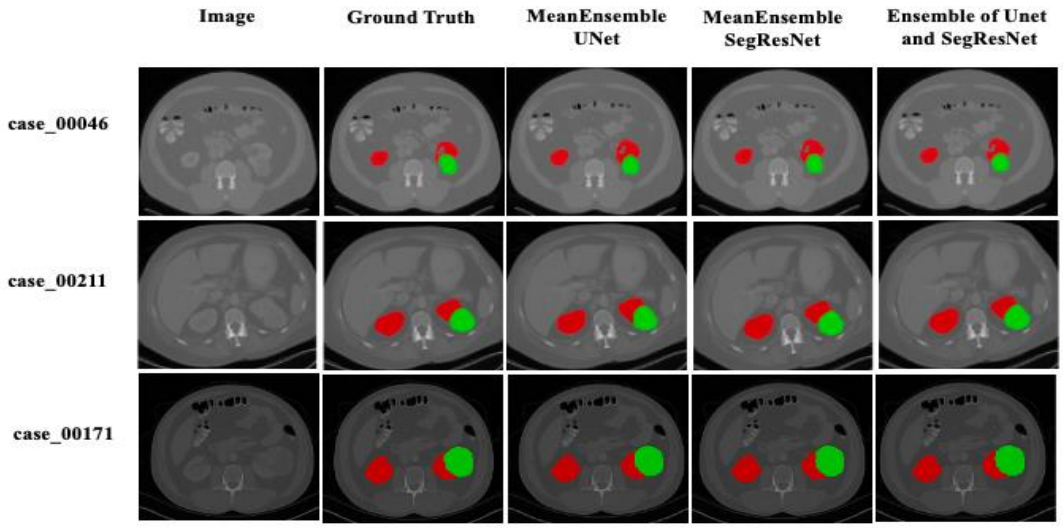


Figure 25: Comparison of Predictions with their Ground truth labels on local test set

Table 9 shows the comparison of our dice scores on the local test dataset with results of kidney and tumor segmentation on datasets other than the KiTS19 official test set of 90 cases present in the literature.

Yang et al. (2018) presented a Fully Convolutional Network (FCN) which was combined with Pyramid Pooling Module (PPM). The dataset consists of a total of 140 CT images of which 90 were used for training purpose and the remaining 50 images were kept for model evaluation. In the first step, ROI extraction was performed using a multi-atlas-based approach. The extracted ROIs were then fed as an input in 3D FCN with PPM. They obtained 0.931 and 0.820 kidney and tumor dice respectively.

Tuncer and Alkan (2018) used the decision support method for the detection of kidney tumors. The dataset consists of 130 total images of which 100 were used for testing purpose. They used the K-Means algorithm to accurately segment kidneys from abdominal CT scans. The segmented kidneys were then used to train the Support Vector Machine (SVM) for the classification of renal tumors. They achieved kidney dice of 0.893.

Mu¨ller and Kramer (2021) proposed a newly developed framework for segmentation of medical images called A Framework for Medical Image Segmentation with Convolutional

Neural Networks and Deep Learning (MIScnn). They randomly selected 120 cases from the KiTS19 training database for which ground truth segmentations were available and divided it into 80 for training and validation and the remaining 40 were used for testing purpose. They performed 3-fold cross-validation on 80 randomly selected cases and obtained a kidney dice of 0.9319 and a tumor dice of 0.6750 on a test set of 40 cases.

da Cruz et al. (2022) proposed a 2.5D network for balancing memory consumption and complexity of the model for kidney tumor segmentation. They performed initial segmentation using DeepLabv3 + 2.5D model with DPN-131 encoder. They used image processing techniques to remove false positives. They randomly selected 31 cases from the KiTS19 train set of 210 CT scans as a local test set, the remaining 179 were used for training and validation purpose. They achieved a tumor dice of 0.8517 on the local test set of 31 cases.

Table 9: Comparison of Kidney and Tumor Segmentation Methods on Local Test Set

Authors	Network	Dataset	Mean Dice	Kidney Dice	Tumor Dice
Yang et al., 2018	3D FCN-PPM	140 cases (90 Training - 50 Testing)	-	0.931	0.820
Tuncer and Alkan, 2018	Decision Support	130 cases (Training 30 and 100 Testing)	-	0.893	-
Muller and Kramer, 2021	MIScnn	120 cases from KiTS19 (80 Training and 40 Testing)	-	0.9319	0.6750
L.B. da Cruz et al., 2022	DeepLabv3+ 2.5D and technique for removing false positives	KiTS19 Dataset – 179 training and validation and remaining 31 as local test set	-	-	0.8517
Our approach	Ensemble of UNet and SegResNet	190 Training cases from KiTS19 and 90 cases from KiTS21(local test set)	**0.9160**	**0.9771**	**0.8550**

4.3. CT-ORG Multi Organ Dataset

We only used the kidney label to evaluate how well our model segments kidney. The results are shown in Table 10, 11 and 12.

Table 10: UNet Kidney Dice on CT-ORG Dataset

Models	Kidney Dice
UNet0	0.9318
UNet1	0.9451
UNet2	0.9362
Ensemble	**0.9477**

Table 11: SegResNet Kidney Dice on CT-ORG Dataset

Models	Kidney Dice
SegResNet0	0.9471
SegResNet1	0.9451
SegResNet2	0.9362
Ensemble	**0.9477**

Table 12: Ensemble Kidney Dice on CT-ORG Dataset

Models	Kidney Dice
UNet	0.9477
SegResNet	0.9499
Ensemble	**0.9501**

Figure 26 below shows the comparison of kidney's predictions generated by all six models and their segmentation ensemble with its ground truth labels.

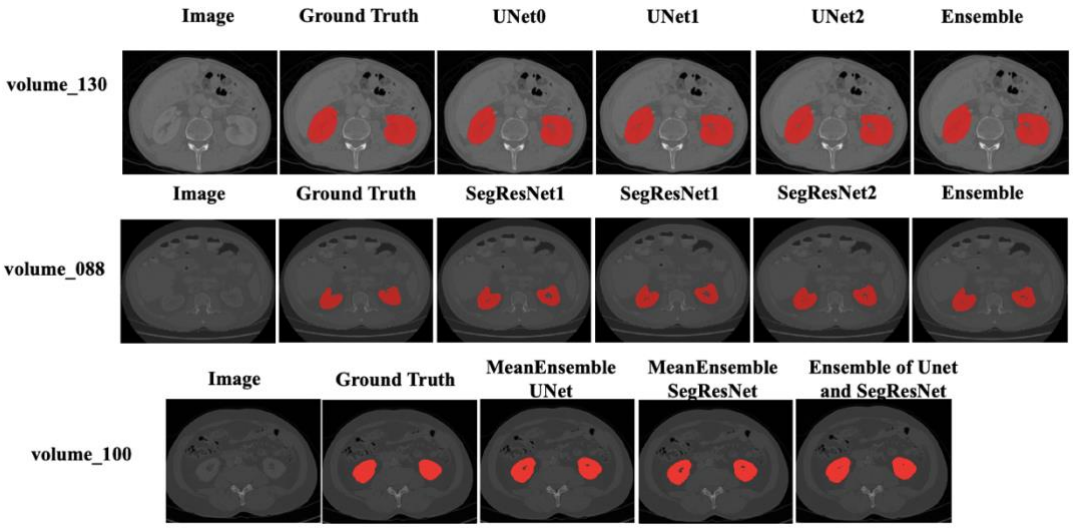


Figure 26: Comparison of kidney's predictions with its ground truth labels from CT-ORG dataset

Table 13 shows a comparison of kidney segmentation methods and their dice scores in literature with our approach to the CT-ORG dataset. Our method outperforms all other methods.

Rister et al. (2020) they released the labeled CT-ORG dataset for multi-organ segmentation consisting of 140 CT scans. They performed a technical validation to check the utility of their dataset for which they trained an FCN on 119 cases and evaluated its performance on 21 test cases. They obtained a dice score of 0.918 for the kidney.

Drees et al. (2022) designed a method, Octree, for semi-automatic segmentation of large 3D multiclass segmentation dataset. The aim was to overcome restrictions in the random walker method. They evaluated the proposed methodology on the CT-ORG dataset for multi-organ segmentation. The score for the kidney was 0.915.

Li et al. (2022) proposed a novel attention module called Large Kernel (LK) which was incorporated into UNet network. They used the CT-ORG dataset to train and test their proposed network for multi-organ segmentation. The dice score for the kidney was 0.9226.

Table 13: Comparison of Kidney Segmentation Methods

Authors	Network	Kidney Dice
Rister et al., 2020	FCN	0.918
Drees et al., 2022	Octree	0.915
Li et al., 2022	LK Attention based UNet	0.923
Our approach	Ensemble of UNet and SegResNet	**0.950**

4.4 Increasing the Dataset

The availability of well-annotated datasets for segmentation tasks is often limited because manual annotation is tedious and error prone task. As a result, the model tends to overfit while training. We used 60 randomly selected cases from the KiTS21 database to compare the results of our network's performance if the training data is increased. 20 random cases from the KiTS19 dataset were used for validation purpose and the remaining 190 were combined with 60 cases from KiTS21. The total number of training cases was 250. We trained a single UNet and SegResNet with this new training dataset using the same preprocessing and training parameters for 600 epochs and evaluated its performance on the KiTS19 official test set on the challenge leader board. The results are shown in Table 14. It can be observed that a single UNet and SegResNet when trained with increased data show more promising results than 3 UNet and 3 SegResNet models trained with 190 cases. Figure 27 below shows the predictions of test cases from KiTS19 test set.

Table 14: Results on KiTS19 Test Data

Model	Mean Dice	Kidney Dice	Tumor Dice
UNet	0.8814	0.9637	0.7991
SegResNet	0.8887	0.9695	0.8080
Ensemble	0.8935	0.9705	0.8164

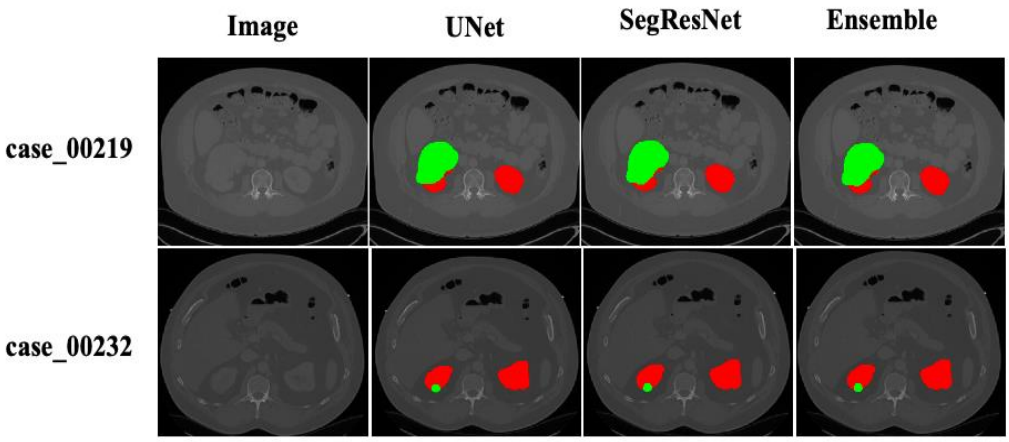


Figure 27: Predictions from KiTS19 test set generated from model trained with increased data

Chapter 5 Discussion

Segmentation is an essential task in the field of medical image analysis. Over the past few years, the use of deep learning-based frameworks for the precise segmentation of organs and tumors has increased drastically. Most of the research is focused on proposing architectures with novel modifications and extensions that can produce better results, but this requires a high level of understanding and experience leading to the increased complexity of the model and complicating things for a layperson as well as for the experts (Isensee et al., 2019). As the model becomes complex it requires more resources for training and inference (GPU and RAM). Therefore, we followed the simple approach of ensemble models and focused more on training procedures rather than architectural modifications and complexity. We used MONAI's UNet and SegResNet models and achieved promising results by focusing on preprocessing, data augmentation, and training protocol. All the models were trained from scratch. Our approach scored better than many submissions which used complex models and more computational resources as shown in Table 5. The competition was won by Isensee and Maier-Hein (2019) with 0.9737 and 0.8505 kidney and Tumor Dice respectively. Our final submission on the KiTS19 leader board obtained a kidney dice of 0.9724 and a tumor dice of 0.8204. Ensembling the outputs of all the models proved to give better dice scores compared to individual model performance. We trained all the models for 600 epochs whereas the winning team trained their models for 1000 epochs. We evaluated the performance of our network on KiTS19 official test set as well as on a local test set of 90 randomly selected cases from KiTS21 database. The obtained dice scores on both test sets (official and local) and the CT-ORG dataset for kidney, shows that our kidney segmentation method is efficient. Our approach outperforms many submissions in terms of kidney segmentation. Segmenting tumors is always a difficult process because of their morphological heterogenicity. The proposed work shows better performance than many other submissions which can be improved further by implementing different techniques such as ROI extraction as the number of slices vary significantly in each CT scan and not all slices contain a Kidney and its tumor andapplying post-processing techniques to reduce false positive.

Chapter 6 Conclusion

In this study, we proposed an ensembling approach without many architectural modifications. We used MONAI's 3D UNet and 3D SegResNet. The results on the KiTS19 leader board show that our approach is accurate for kidney segmentation but falls short on tumor dice. Although deep learning networks have achieved the state of the art performances in kidney segmentation, tumor segmentation still needs improvement. The main reason for this is the lack of publicly available well-annotated datasets containing tumors. The results in the Table 9 show that simply increasing the number of training cases executes better performance for tumor segmentation without any change in training protocol. For future work, we plan to merge the training sets of the KiTS19 and KiTS21 databases and train the models with this data. We also plan to focus more on the segmentation of tumors by incorporating an additional stage for detecting small tumors. We hope our approach can be a leading step in the early detection of kidney tumors for better diagnosis and treatment planning.

Chapter 7 Reference

1. Capitanio, U., & Montorsi, F. (2016). Renal cancer. The Lancet, 387(10021), 894-906.
2. Capitanio, U., Bensalah, K., Bex, A., Boorjian, S. A., Bray, F., Coleman, J., ... & Russo, P. (2019). Epidemiology of renal cell carcinoma. European urology, 75(1), 74-84.
3. Choudhary, U., Kumar, S., Jee, K., Singh, A., & Bharti, P. (2017). A cadaveric study on anatomical variations of kidney and ureter in India. *Int J Res Med Sci*, *5*, 2358-61.
4. Cinque, A., Vago, R., & Trevisani, F. (2021). Circulating RNA in kidney cancer: What we know and what we still suppose. Genes, 12(6), 835.
5. Corr, M. P., & Maxwell, A. P. (2022). Maintain a high index of suspicion for kidney cancer. The Practitioner, 266(1857), 21-24.
6. da Cruz, L. B., Júnior, D. A. D., Diniz, J. O. B., Silva, A. C., de Almeida, J. D. S., de Paiva, A. C., & Gattass, M. (2022). Kidney tumor segmentation from computed tomography images using DeepLabv3+ 2.5 D model. Expert Systems with Applications, 192, 116270.
7. Drees, D., Eilers, F., & Jiang, X. (2022). Hierarchical Random Walker Segmentation for Large Volumetric Biomedical Images. IEEE Transactions on Image Processing, 31, 4431-4446.
8. Finco, D. R. (1997). Kidney function. In *Clinical biochemistry of domestic animals* (pp. 441-484). Academic Press.
9. Graham-Knight, J. B., Djavadifar, A., Lasserre, D. P., & Najjaran, H. (2019). Applying NnU-Net to the KiTS19 Grand Challenge. Univ. Minn. Libr, 1-7.
10. Guo, Z., Li, X., Huang, H., Guo, N., & Li, Q. (2019). Deep learning-based image segmentation on multimodal medical imaging. IEEE Transactions on Radiation and Plasma Medical Sciences, 3(2), 162-169.
11. He, K., Zhang, X., Ren, S., & Sun, J. (2016, October). Identity mappings in deep residual networks. In European conference on computer vision (pp. 630-645). Springer, Cham.

12. Heller, N., Isensee, F., Maier-Hein, K. H., Hou, X., Xie, C., Li, F., ... & Weight, C. (2021). The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. Medical image analysis, 67, 101821.
13. Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., ... & Weight, C. (2019). The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445
14. Hou, X., Xie, C., Li, F., Nan, Y., 2019. Cascaded semantic segmentation for kidney and tumor, in: Submissions to the 2019 Kidney Tumor Segmentation Challenge – KiTS19.
15. Hou, X., Xie, C., Li, F., Wang, J., Lv, C., Xie, G., & Nan, Y. (2020, April). A triple-stage self-guided network for kidney tumor segmentation. In 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI) (pp. 341-344). IEEE.
16. Hsiao, C. H., Sun, T. L., Lin, P. C., Peng, T. Y., Chen, Y. H., Cheng, C. Y., ... & Huang, Y. (2022). A deep learning-based precision volume calculation approach for kidney and tumor segmentation on computed tomography images. Computer methods and programs in biomedicine, 221, 106861.
17. Isensee, F., & Maier-Hein, K. H. (2019). An attempt at beating the 3D U-Net. arXiv preprint arXiv:1908.02182.
18. Isensee, F., Jäger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2019). Automated design of deep learning methods for biomedical image segmentation. arXiv preprint arXiv:1904.08128.
19. Jiao, M., & Liu, H. (2019). 2019 Kidney Tumor Segmentation Challenge Method Manuscript.
20. Kowalewski, K. F., Egen, L., Fischetti, C. E., Puliatti, S., Rivas, J. G., Taratkin, M., ... & Cacciamani, G. (2022). Artificial intelligence for renal cancer: From imaging to histology and beyond. Asian Journal of Urology.
21. Kumaraswamy, A. K., & Patil, C. (2020). A Cascaded U-net for Kidney and Tumor Segmentation from CT volumes. Artificial Intelligence in Oncology, 2(1), 004-008.

22. Li, H., Nan, Y., Del Ser, J., & Yang, G. (2022). Large-Kernel Attention for 3D Medical Image Segmentation. arXiv preprint arXiv:2207.11225.
23. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.
24. Ma, J., 2019. Solution to the kidney tumor segmentation challenge 2019, in: Submissions to the 2019 Kidney Tumor Segmentation Challenge – KiTS19
25. Mahadevan, V. (2019). Anatomy of the kidney and ureter. Surgery (Oxford), 37(7), 359-364.
26. Mancini, M., Righetto, M., & Baggio, G. (2020). Gender-related approach to kidney cancer management: Moving forward. International journal of molecular sciences, 21(9), 3378.
27. Medina-Rico, M., Ramos, H. L., Lobo, M., Romo, J., & Prada, J. G. (2018). Epidemiology of renal cancer in developing countries: Review of the literature. Canadian Urological Association Journal, 12(3), E154.
28. Mir, M. C., Derweesh, I., Porpiglia, F., Zargar, H., Mottrie, A., & Autorino, R. (2017). Partial nephrectomy versus radical nephrectomy for clinical T1b and T2 renal tumors: a systematic review and meta-analysis of comparative studies. European urology, 71(4), 606-617.
29. Mu, G., Lin, Z., Han, M., Yao, G., & Gao, Y. (2019). Segmentation of kidney tumor by multi-resolution VB-nets.
30. Müller, D., & Kramer, F. (2021). MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. BMC medical imaging, 21(1), 1-11.
31. Myronenko, A. (2018, September). 3D MRI brain tumor segmentation using autoencoder regularization. In International MICCAI Brainlesion Workshop (pp. 311-320). Springer, Cham. (segresnet)
32. Myronenko, A., & Hatamizadeh, A. (2019). 3d kidneys and kidney tumor semantic segmentation using boundary-aware networks. arXiv preprint arXiv:1909.06684.

33. Rajendran, R., KM, S. K., Panetta, K., & Agaian, S. (2022, May). KTNet: Towards automated 2D kidney and tumor segmentation. In Multimodal Image Exploitation and Learning 2022 (Vol. 12100, pp. 262-271). SPIE.
34. Rao, V., Sarabi, M. S., & Jaiswal, A. (2015). Brain tumor segmentation with deep learning. MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS), 59, 1-4.
35. Rister, B., Shivakumar, K., Nobashi, T., Rubin, D.L., 2019. Ct-org: Ct volumes
36. with multiple organ segmentations. The Cancer Imaging Archive
37. Rister, B., Yi, D., Shivakumar, K., Nobashi, T., & Rubin, D. L. (2020). CT-ORG, a new dataset for multiple organ segmentation in computed tomography. Scientific Data, 7(1), 1-9.
38. Santini, G., Moreau, N., & Rubeaux, M. (2019). Kidney tumor segmentation using an ensembling multi-stage deep learning approach. A contribution to the KiTS19 challenge. arXiv preprint arXiv:1909.00735.
39. Scelo, G., & Larose, T. L. (2018). Epidemiology and risk factors for kidney cancer. Journal of Clinical Oncology, 36(36), 3574.
40. Stevens, L. M., Lynm, C., & Glass, R. M. (2010). Kidney failure. *JAMA*, *304*(2), 228-228.
41. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 71(3), 209-249.
42. Tsai, Y. C., & Sun, Y. N. (2019). KiTS19 challenge segmentation.
43. Tuncer, S. A., & Alkan, A. (2018). A decision support system for detection of the renal cell cancer in the kidney. Measurement, 123, 298-303.
44. van Oostenbrugge, T. J., Fütterer, J. J., & Mulders, P. F. (2018). Diagnostic imaging for solid renal tumors: a pictorial review. Kidney Cancer, 2(2), 79-93.
45. Yang, G., Li, G., Pan, T., Kong, Y., Wu, J., Shu, H., ... & Zhu, X. (2018, August). Automatic segmentation of kidney and renal tumor in ct images based on 3d fully convolutional neural network with pyramid pooling module. In 2018 24th International Conference on Pattern Recognition (ICPR) (pp. 3790-3795). IEEE.

46. Yu, Q., Shi, Y., Sun, J., Gao, Y., Dai, Y., & Zhu, J. (2018). Crossbar-net: A novel convolutional network for kidney tumor segmentation in ct images. arXiv preprint arXiv:1804.10484.
47. Zhang, L., Wang, X., Yang, D., Sanford, T., Harmon, S., Turkbey, B., ... & Xu, Z. (2020). Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. IEEE transactions on medical imaging, 39(7), 2531-2540.
48. Zheng, Z., Geng, J., Jiang, Y., Zhang, M., Yang, R., Ge, G., ... & Zhang, X. (2021). Kidney Diseases. In *Clinical Molecular Diagnostics* (pp. 553-582). Springer, Singapore.
49. Zhou, B., & Chen, L. (2016, October). Atlas-based semi-automatic kidney tumor detection and segmentation in CT images. In 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (pp. 1397-1401). IEEE.